Pedro Henrique Thibes Forquesato

Essays in Political and Cultural Economics

Tese de Doutorado

Thesis presented to the Programa de Pós–graduação em Economia of the Departamento de Economia, PUC-Rio, as partial fulfillment of the requirements for the degree of Doutor em Economia.

Advisor : Prof. Vinicius do Nascimento Carrasco
Co-Advisor: Prof. Claudio Abramovay Ferraz do Amaral

Rio de Janeiro
April 2016

Pedro Henrique Thibes Forquesato

Essays in Political and Cultural Economics

Thesis presented to the Programa de Pós–graduação em Economia of the Departamento de Economia, PUC-Rio, as partial fulfillment of the requirements for the degree of Doutor em Economia. Approved by the following commission:

Prof. Vinicius do Nascimento Carrasco
Advisor
Departamento de Economia – PUC-Rio

Prof. Claudio Abramovay Ferraz do Amaral
Co-Advisor
Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Prof. Leonardo Bandeira Rezende
Departamento de Economia - PUC-Rio – PUC-Rio

Prof. Juliano Junqueira Assunção
Departamento de Economia - PUC-Rio – PUC-Rio

Prof. João Manoel Pinho de Mello
INSPER – INSPER

Prof. César Zucco Junior
Escola Brasileira de Administração Publica e de Empresas -
FGV – EBAPE/FGV

Prof. Monica Herz
Coordenador Setorial do Centro de Ciências Sociais – PUC-Rio

Rio de Janeiro, April 1st, 2016

Pedro Henrique Thibes Forquesato

Pedro Forquesato holds a B.A. in Economics from the Universidade Estadual de Campinas (Unicamp) and a Ph.D. in Economics from the Catholic University of Rio de Janeiro (PUC-Rio). During his Ph.D. studies, he was visiting scholar at the John F. Kennedy School of Government at Harvard University for a period of one year. His research focus on political and cultural aspects of development and economic theory.

Acknowledgments

First and foremost, I would like to thank my wife, Xinxin Jiang, for the love and company. I would also want to thank my family, especially my parents, Mônica and Edson, and my siblings Camila and Luís.

I am deeply grateful to my advisor, Vinicius Carrasco, for his vital support in all stages of the production of this thesis, support which far exceeded his formal obligation as advisor. I am also thankful to my co-advisor Claudio Ferraz, with whom I wrote the third chapter of this thesis, but who also had a crucial role in the development of the first two. I am indebted to all members of the committee. João Manoel, Leo, Juliano and César, all had significant comments and suggestions before and during the thesis defense; they improved this thesis significantly and also helped shape me as a researcher.

I am grateful to all my professors, during undergraduate and graduate studies, which gave me a solid basis in economics that allowed me to develop the research that led to this thesis. There is something of each of them in what is best in these articles. (Possible shortcomings are entirely my fault.) I am also grateful to the employees of the Departamento de Economia of PUC-Rio for their indispensable support during the five years I stayed in this institution.

I am particularly indebted to Professor Filipe Campante of Harvard University, for having received me during my period as visiting scholar in that institution, and having actively participated in my formation as a researcher, as well as helping develop the research that would eventually become this thesis. I thank all professors and students of Harvard University, for having so kindly embraced me when I was alone in a foreign and unknown country, being always willing to discuss research and help me develop as an academic.

This thesis would have not been possible without the companionship of my graduate colleagues, as well as all my other friends in PUC-Rio and Campinas which stood by my side and wished for my success in the more difficult moments in these five years of intense study and research.

Finally, I am grateful to the financial support of CAPES, both during my PhD in Brazil as well as during my research period in United States, and CNPq.

Abstract

Forquesato, Pedro; Carrasco, Vinicius; Ferraz, Claudio. **Essays in Political and Cultural Economics**. Rio de Janeiro, 2016. 90p. PhD Thesis – Departmento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis is composed of three papers, the first in organizational economics and culture; the last two in political economics. In the first chapter, we model the relation between dissemination of social norms of work ethic and incentives proposed by firms, which we motivate using evidence from three different datasets. In the second chapter, we examine whether neighbors' income affects voting, using data from election results for the 2004-2012 Presidential Elections in Unites States, by precinct and block group. That way, we try to contribute to understanding the reason why there are different demands for income redistribution. As an identification strategy, we use year fixed-effects and tract year dummies; tract is the smallest geographic unit larger than block group (on average, each tract contains 4 block groups). In the third chapter, we study patronage, investigating the effect of a mayoral candidate's victory on the probability that members of his party (or parties in the same presidential coalition) occupy public jobs in the government, or on their income accrued from government, in case they are already public employees. We also analyze the effect of a party's victory over the number of registered members of that party in future years, which would indicate that voters affiliate to political parties as a way to signal support to the office holder. We estimate plausibly causal effects of a party holding mayoral position by comparing municipalities where that party nearly won with places where it nearly lost.

Keywords

Resumo

Forquesato, Pedro; Carrasco, Vinicius; Ferraz, Claudio. **Ensaios em Economia Política e Cultura**. Rio de Janeiro, 2016. 90p. Tese de Doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese é formada por três artigos, o primeiro em economia organizacional e cultura; os dois últimos em economia política. No primeiro capítulo, nós modelamos a relação entre a disseminação de normas sociais de ética do trabalho e incentivos propostos pelas firmas, que motivamos utilizando evidência de três bases de dados diferentes. No segundo capítulo, examinamos se a renda dos vizinhos afeta o voto de eleitores, utilizando dados de resultados de eleições presidenciais (2004 até 2012) nos Estados Unidos, por zona eleitoral e grupo de bairros. Com isso, buscamos contribuir para o entendimento das razões que levam a diferentes níveis de demanda por redistribuição de renda. Como estratégia de identificação, utilizamos efeitos fixos de ano e dummies de trato e ano; trato sendo a menor unidade geográfica maior que o grupo de blocos (em média, um trato contém 4 grupos de blocos). No terceiro capítulo, estudamos patronagem, investigando o efeito da vitória de um candidato a prefeito de um partido na probabilidade de membros deste partido (ou de partidos da mesma coalizão) ocuparem cargos públicos no governo; ou de sua renda advinda do governo aumentar, caso já sejam empregados públicos. Analisamos também o efeito da vitória de um partido sobre o número de registrados a este partido nos anos futuros, o que indicaria um desejo de sinalizar apoio ao candidato eleito. Estimamos o efeito causal de um partido ocupar a prefeitura, comparando municipalidades em que este partido quase ganhou com cidades em que quase perdeu.

Palavras-chave

Economia Política. Economia Cultural. Redistribuição de Renda. Risco Moral. Ética do Trabalho.

Sumário

Lista de figuras

Lista de tabelas

Because of the success of science, there is a kind of a pseudo-science. Social science is an example of a science which is not a science. They follow the forms. You gather data, you do so and so and so forth, but they don't get any laws, they haven't found out anything. They haven't got anywhere – yet. Maybe someday they will, but it's not very well developed...

See, I have the advantage of having found out how hard it is to get to really know something, how careful you have to be about checking your experiments, how easy it is to make mistakes and fool yourself. I know what it means to know something. And therefore, I see how they get their information. And I can't believe that they know when they haven't done the work necessary, they haven't done the checks necessary, they haven't done the care necessary.

Richard Feynman, *Interview to BBC, 1981.*

1
Social Norms of Work Ethic and Incentives in Organizations

"Seest thou a man diligent in his business? He shall stand before kings."(Proverbs, 22-29.)

1.1
Introduction

For a long time, sociologists and classical economists have defended the importance of cultural values of work effort (henceforth, work ethic) in explaining differences in development ((03)), and a recent body of literature ((15), (16)) uses historical evidence to emphasize the importance of these explanations. That literature, however, does not clearly grasp how these differences in cultural values emerge and spread in different societies.[1]

In this paper, I hypothesize that differences in the dissemination of work ethic emerge endogenously from multiple equilibria in cultural transmission decisions made by parents and incentive choices made by firms. I build a model that shows how this process would occur, and I substantiate it using anecdotal and empirical evidence that firm incentives are an important determinant of (intergenerational) cultural transmission.

In my model, parents altruistically choose whether to transmit work ethic to their offspring, taking into account the effect it will have on their utility. Where incentives are high-powered (that is, where payment is more dependent on success), parents expect that their children will work harder, and work ethic will be more rewarding for them. In these societies, parents have stronger incentives to transmit work ethic, and work ethic slowly disseminates. Incentive steepness (as in how high-powered the incentives are) has an effect on work ethic.

When there is complementarity of effort between workers, having an employee with a strong work ethic (who, in equilibrium, will exert higher effort) improves the productivity of his peers. Having more productive workers, the firm optimally chooses to offer high-powered incentives in order to take

---

[1]The literature is not unanimous, for example, on Weber's own explanation for the emergence of these cultural norms. See (20) and (19). I discuss how my explanation relates to Weber's in Section 5.

advantage of the higher productivity of its employees by inducing them to work harder. If worker effort is complementary, then in societies in which the proportion of individuals with work ethic is higher, firms choose to offer steeper incentives and demand more effort.

Consequently, the model has two equilibria. When a large proportion of the society's workforce has a strong work ethic, firms offer steeper incentives even to employees without a work ethic, as having high-effort coworkers makes them more productive. If firms choose high-powered incentives, then agents work hard and cultural values that stimulate effort are particularly useful to them. If a small proportion of the workforce has a work ethic, then workers shirk more, making their coworkers less productive and causing firms to choose less high-powered incentives and demand less effort, even from workers with a strong work ethic. As individuals work less, work ethic is less useful, and parents have weaker incentives to transmit it.

Having established the main argument, I augment the benchmark model by allowing firms to choose their production technologies, adopting either a complementary technology in which an agent's productivity is highly dependent on his coworkers' or a separable technology in which interdependencies among workers are minimal. I interpret this choice as one between modern production processes, such as Fordism or lean manufacturing—more efficient ways of organizing production in which a worker's productivity is highly dependent on his peers' effort—and traditional technologies, such as the putting-out system, which is less efficient overall but creates less dependency between coworkers.[2]

This extended model also has two equilibria, one in which firms use traditional technologies and workers do not have a strong work ethic and another in which firms adopt modern production processes and a strong work ethic is disseminated throughout the society. This pattern matches stylized historical facts and helps explain why countries in which labor is unproductive are reluctant to adopt new technologies and organizational structures from developed countries. A similar argument is made in (22). Here, I contribute by explicitly modeling the worker's effort and the firm's incentive choice. Thus, my model also clarifies why managers do not employ different incentive schemes in firms whose labor productivity is low and explains the evidence provided in (15)

---

[2]Fordism is a system of mass production pioneered in the early 20th century by the Ford Motor Company based, *inter alia*, on the use of assembly lines. Lean manufacturing is a management philosophy derived mostly from the Toyota Production System, which focuses on reducing waste by having the "right tools in the right place at the right time."At the opposite end of the spectrum from these very complex and complementary processes is the putting-out system wherein a central agent subcontracts production to off-site facilities (usually the worker's own home).

that effort, even more than skill, differentiates regions in terms of productivity.

To show the importance of the relation between incentive steepness and work ethic beyond the use of historical examples, I examine 3 different datasets and show that wider dissemination of work ethic is correlated with measures of incentive steepness and effort at work. First, using data from the International Social Survey Programme, I show that people are more likely to arrive from home exhausted, a measure of effort, when living in countries in which the proportion of people who believe that work is a person's most important activity is high. I then use data from the European Values Study and the European Social Survey to show that in regions in which more people believe that hard work is an important characteristic that should be taught at home, wages are more likely to depend on effort, a measure of incentive steepness. Finally, I go beyond self-reported measures by examining the correlation between work ethic and executive compensation, using actual data from Standard & Poor's Execucomp for the United States. The results of these empirical exercises are consistent with the model: workers in areas with wider dissemination of work ethic are more likely to receive high-powered incentives and to work harder, on average, indicating the importance of cultural norms to understanding the behaviors of workers and firms.

This paper is embedded in a new and growing literature on the intergenerational transmission of cultural traits, which is reviewed in (13). In particular, in my model, parents are imperfectly altruistic when choosing how to socialize their offspring, as in (01, 02). My cultural transmission process, however, includes endogenous cultural intolerance (that is, the benefit of socializing an individual to value a trait depends on its dissemination throughout the society rather than on exogenous factors), which is determined by the average incentive scheme chosen by firms. This scheme is a function of the proportion of agents with a strong work ethic in the society. In fact, as parents will want to instill a strong work ethic when most individuals have it, the model exhibits strategic complementarity in the socialization process (as in, for example, (17)).[3]

Closely related to work ethic, a growing literature investigates cultural beliefs about the returns to effort, although its focus has been on explaining redistribution choices and social protection (as in (05) and (04)). While these papers provide an important mechanism by which differences in effort and cultural norms of work exist among countries, namely, the political choices of social insurance and redistribution, they are less able to explain why these differences might exist within countries (as between southern and northern

---

[3]See (13) for other references with endogenous cultural intolerance.

Italy), where redistribution levels are roughly similar, or in non-democratic societies, where redistribution levels are not decided by majority voting.

Furthermore, my model yields various policy implications. If multiple equilibria emerge as a consequence of different levels of redistribution and beliefs about returns to effort, then an exogenous change in taxes and transfers would cause a change of equilibrium. In my model, however, that is not necessarily true. A reduction in taxes in a region with a poorly disseminated work ethic might not be sufficient to change the equilibrium, as it might still be optimal for parents to not instill their offspring with work ethic because firms employ low-powered incentives. In this sense, I contribute to the literature by offering an alternative (and complementary) mechanism that can explain differences in effort and does not depend on political mechanisms. Moreover, the fact that the empirical evidence shows that incentive steepness is correlated with measures of work ethic suggests that my mechanism is relevant to explaining differences in work behavior across regions.

Perhaps the closest paper to ours is (21). They develop a model in which the middle class chooses to transmit a work ethic while the nobility prefers to transmit preferences for leisure. With the Industrial Revolution, a work ethic becomes particularly advantageous, and the middle class takes over and becomes the new entrepreneurs and the dominant economic class. The main difference between their model and mine is that I endogenize the choice of incentives made by firms and focus my analysis on effort exertion rather than on labor supply. This specification is important both in order to explain the anecdotal evidence that worker effort is an important determinant of productivity and to explain the empirical evidence that work ethic is correlated with incentive steepness, a facet that can only be meaningfully analyzed in agency models.[4]

The rest of this paper is organized as follows. In Section 3, I present empirical evidence that suggests that worker effort and incentive steepness vary by region and that this variation is robustly correlated to regional differences in work ethic. In Section 4, I present a model of intergenerational transmission of work ethic and incentives in organizations, and I argue that the model can explain a set of stylized facts. In Section 5, I discuss some assumptions and results of the model, while in Section 6, I extend it by considering the implications of firms being able to choose whether their technology is characterized by complementarity. I analyze the comparative dynamics of this

---

[4]Also related are the literatures on the evolutionary selection of human traits ((62), (63)) and on contracting with social norms ((09), (08), (06), (11), (64), (32), (12) (61), and (60)). Other approaches to studying norms of hard work include (10) and (18).

extended model. Finally, I conclude with a discussion of my results and avenues for future research.

1.2
Work Ethic

Since the 19th Century, classical economists and sociologists have observed that regions exhibit very different cultural norms of hard work and affirmed that this difference is a major cause of productivity (and, consequently, of income) inequality between societies. Max Weber, for example, in his book about the Protestant work ethic, wrote:

> "As every employer knows, the lack of *coscienziositá* [diligence] of the labourers of such countries, for instance Italy as compared with Germany, has been, and to a certain extent still is, one of the principal obstacles to their capitalistic development. Capitalism cannot make use of the labour of those who practice the doctrine of undisciplined *liberum arbitrium* [free will] ...."(03, p. 21).

Summarizing the importance of work ethic for economic (capitalistic) development, Weber writes, "This peculiar idea (...) of one's duty in a calling, is what is most characteristic of the social ethic of capitalistic culture"(03, p. 19).

Here, work ethic is understood as a cultural norm that is internalized by agents and makes shirking costlier while providing moral support for high effort, reducing its cost. Examples of this thought abound in historical and religious sources. Benjamin Franklin, for example, in his "Advice to a Young Tradesman, Written by an Old One,"writes:

> "The Sound of your Hammer at Five in the Morning or Nine at Night, heard by a Creditor, makes him easy Six Months longer. But if he sees you at a Billiard Table, or hears your Voice in a Tavern, when you should be at Work, he sends for his Money the next Day."(30).

Another important source of social norms of effort is religion. As Baxter states, "It is for action that God maintaineth us and our activities; work is the moral as well as the natural end of power."(31, p. 375). Moreover, work ethic is not only a social norm that values effort and dedication but also one that condemns leisure and shirking. Baxter (as cited by Weber) dictates, "Waste of time is thus the first and in principle the deadliest of sins. (...) Loss of time through sociability, idle talk, luxury, (...) is worthy of absolute moral condemnation."(03, p. 104).

Therefore, while a strong work ethic is useful in a utilitarian sense for hard working individuals, promising ethical, monetary, or religious compensation for their toil, it is detrimental to individuals who favor leisure, as it imposes guilt for and moral condemnation of their low effort.

1.3
Basic Facts

Recent empirical evidence substantiates Weber's claim that there are sizable differences in work norms and effort among societies. (07), while investigating the behavior of the employees of a large Italian bank, discover that those working in branches in southern Italy were significantly more likely to shirk then employees of northern Italian branches. Moreover, individual preferences for shirking and work explain most of this differential (rather than sorting, peer effects, or firm attributes). Gregory Clark provides anecdotal evidence that also points to significant differences in worker behavior across regions:[5]

> "In fact workers in poorly performing economies simply supply very little actual labor input on the job. Workers in modern cotton textile factories in India, for example, are actually working for as little as fifteen minutes of each hour they are at the workplace."(15, p. 13).

Furthermore, he argues that this low level of effort in factories was well known by managers and firm owners, which raises the question of why managers did not incentivize their employees to work harder.[6] By endogenizing the choice of incentives (and hence, indirectly, work effort), my model can provide an answer to this puzzle.[7]

In the rest of this section, I present quantitative evidence from three different datasets that is consistent with work ethic variation among regions and is correlated with both effort provision and incentive steepness. I provide descriptive statistics and some alternative specifications in Appendix B.

---

[5]For related arguments, see also (16).

[6](15) mentions, for example, that "[t]o partially control this absenteeism some employers used a pass system, under which a worker could leave the mill only with a pass or token from his or her department. Each department was allotted passes equal to 10-25 percent of the staff."

[7]Another possible explanation for this puzzle is collusion between managers and workers. Clark offers evidence that this mechanism was not very important. For example, in 1895, in Bombay, of 55 mill managers, 27 were British, as were 77 of the 190 weaving masters and other managerial positions. It seems implausible that British managers would be more willing to collude with Indian workers than with British ones.

Figura 1.1: Cross-country correlations between work ethic and effort at work (ISSP 1989, 1997).
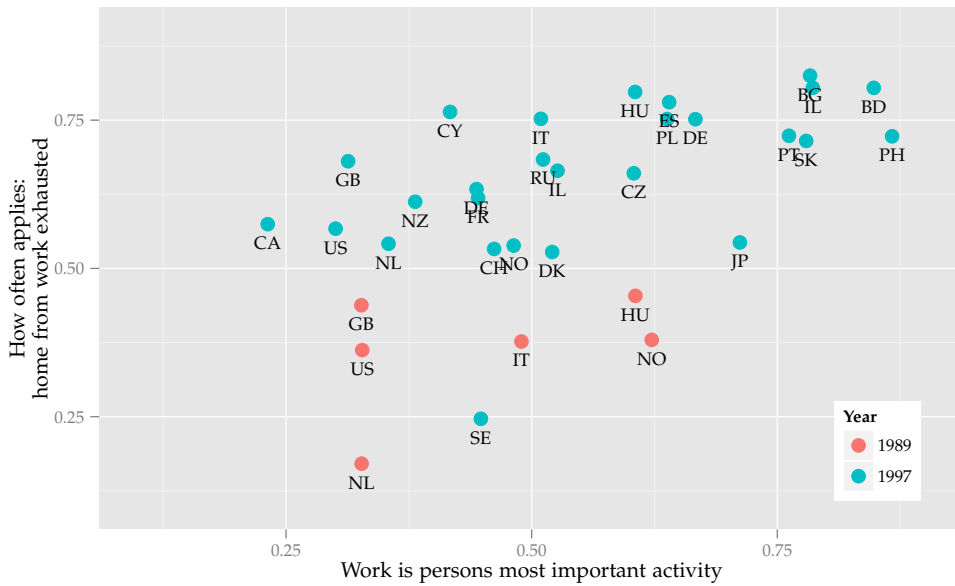
1.3.1
International Social Survey Programme

First, I use data from the International Social Survey Programme: Work Orientations I and II (ISSP 1989 and ISSP 1997) on work ethic and effort.[8] The International Social Survey Programme is a cross-national collaboration for social science surveys. I access its data to build individual- and country-level measures of work ethic and to explore the correlations between these variables and worker effort.

As a measure of work ethic, I analyze respondents' answers to the following question: "How much do you agree with or disagree with, thinking about work in general: Work is a person's most important activity?"To obtain more easily interpretable results, I create a dummy variable that equals one if the respondent answers "Strongly agree"or "Agree"and zero otherwise. As a measure of work effort, I use the respondent's answer to the following question: "How often do you come home from work exhausted?"If the respondent answers "Always"or "Often,"the variable takes the value one; if he replies "Sometimes, Hardly Ever,"or "Never,"it takes the value zero.[9]

The correlation between those two variables is displayed in Figure 1.1.[10]

---

[8]I do not use the 2005 wave because the questionnaire does not contain the questions I am interested in.

[9]I obtain the same qualitative results using the original variables.

[10]The Pearson correlation is 0.56. The countries in Figure 1.1 are Germany (DE), Great Britain (GB), United States (US), Hungary (HU), Italy (IT), Netherlands (NL), Norway (NO), Sweden (SE), Czech Republic (CZ), Slovakia (SK), Poland (PL), Bulgaria (BG), Russia (RU), New Zealand (NZ), Canada (CA), Japan (JP), Spain (ES), France (FR), Cyrus

It shows that both work ethic and effort differ significantly by country in both years, although work ethic seems to vary little from 1989 to 1997 in countries with both values, which is consistent with the view of culture as slow to change. Moreover, there is a correlation between how important people in a country consider work and how much effort, on average, people exert at work. This correlation seems to hold for both years for which there is data. I provide more evidence on this below.

As in the model, I consider the relevant dimension of work ethic to be the average level in the population, as it determines the probability that a worker has coworkers with a strong work ethic. I therefore aggregate the measure of work ethic by country. I also use a number of individual- and country-level controls. Table 1.1 formalizes the findings in Figure 1.1, showing that the correlation is statistically significant and robust to the addition of controls.

The first three columns establish that individuals with a strong work ethic are significantly more likely to return home from work exhausted, which is a measure of how much effort they exert in their jobs. Column 1 presents this result, and Columns 2 and 3 show that it is robust to the addition of individual-level controls and country dummies. Although the point estimates are smaller in these two scenarios, they remain statistically significant at the 5% level.

In Columns 4 to 6, I analyze the effect of country-level work ethic on effort choice. As my model predicts, individuals with a strong work ethic exert more effort, but the culture is even more important in determining their effort. Column 4 presents the basic results, Column 5 uses individual-level controls, and Column 6 uses both individual-level controls and country fixed effects.[11] The estimated effect changes little in Column 5, and it actually increases significantly after I include country fixed effects.

An important *caveat* of this analysis is that the measures of work ethic and effort are self-reported, so it is possible that the results are driven by endogenous measurement error. (If the measurement error is independent, then one would expect at most attenuation bias, which works against my predictions.) Indeed, the negative coefficient for years of education and the strong positive coefficient for female gender point to the flaws of this measure of effort. In the first case, a possible explanation for the negative effect is that more educated workers have white collar jobs, which are less exhausting. In the second case, while women might, on average, exert more effort than men, an

(CY), Portugal (PT), Denmark (DK), Ireland (IE), Israel (IL), Austria (AT), Switzerland (CH), and Philippines (PH).

[11]Since I have only 31 country-level observations, this limits my power to control for time-varying country-level variables.

Tabela 1.1: Correlation between work ethic and effort at work (ISSP 1989, 1997)

	(1)	(2)	(3)	(4)	(5)	(6)
Work important	0.0883	0.0432	0.0167	0.0522	0.0163	0.0128
	(0.0118)***	(0.0126)***	(0.0071)*	(0.0068)***	(0.0062)**	(0.0062)*
Work imp. (country)				0.3480	0.3067	0.5739
				(0.0861)***	(0.1068)**	(0.0600)***
Education (yrs)		-0.0138	-0.0110		-0.0107	-0.0110
		(0.0030)***	(0.0024)***		(0.0028)***	(0.0024)***
Female		0.1047	0.1085		0.1075	0.1091
		(0.0088)***	(0.0083)***		(0.0081)***	(0.0082)***
Year (1997)	0.2762	0.2628	0.2530	0.2578	0.2247	0.2469
	(0.0290)***	(0.0357)***	(0.0478)***	(0.0281)***	(0.0388)***	(0.0295)***
Individual-level controls	N	Y	Y	N	Y	Y
Country dummies	N	N	Y	N	N	Y
R^2	0.058	0.124	0.163	0.069	0.132	0.170
N	40617	32018	32018	40617	32018	32018

The dependent variable is the respondent's answer to "How often do you come home from work exhausted?" I code "Always" or "Often" as one "Sometimes,Hardly Ever," or "Never" as zero. Corrected standard errors are clustered at the country level. P-values are significant at the 0.05 (*), 0.01 (**), and 0.001 (***) levels. Individual-level controls are years of education, sex, age, age squared, and self-reported social class. There are 31 country-level observations.

alternative explanation would be that they are more easily exhausted, and it is impossible to distinguish between these conjectures. Notwithstanding, as long as people with a strong work ethic are not more prone to feeling exhausted (or reporting that they feel exhausted) than those without, my measures will underestimate the total importance of work ethic on effort.

1.3.2
European Social Survey and European Values Study

My second empirical exercise uses data from the European Social Survey: Family, Work and Well Being (ESS2 2004 and ESS5 2010) and the 4th Wave of the European Values Study (EVS 2008) to form a database of measures of work ethic at the regional level and work attitudes at the individual level. The advantage of this database is that the ESS includes questions that are plausible indicators of the steepness of incentives. (The interest is in how high-powered the incentives are, not in how high the wages are, which is more easily measurable.)

From the EVS 2008, I aggregate an indicator of work values at the regional level, namely, if the respondent mentioned "hard work"when asked the following question: "Here is a list of qualities which children can be encouraged to learn at home. Which, if any, do you consider to be especially important? Please choose up to five."I argue that this variable best captures the intergenerational aspect of the cultural transmission of work ethic; nonetheless, my result is robust to the use of other related questions.

I use the ESS data to obtain individual, self-reported measures of the steepness of incentives. I use respondents' answers to the following question: "Please tell me how true each of the following statements is about your current job: my wage or salary depends on the amount of effort I put into my work."To simplify interpretation, again I code the responses as dummy variables.

Figure 1.2 shows the correlation between self-reported incentive steepness and country-level work ethic for European countries in 2004 and 2010.[12] The correlation is strong (0.63); the probability that the respondent considers his wage highly dependent on effort in countries with high work ethic is twice as large as that in countries with low work ethic.

I test this correlation more rigorously in Table 1.2 with individual-level data, showing that it is strong and robust to the inclusion of a multitude of controls. In my most conservative specification, moving from the 1st quartile

---

[12]Country codes are given in footnote 10, with the addition of Ukraine (UA) and Croatia (HR).

Figura 1.2: Cross-country correlation between work ethic and incentive steepness (EVS 2008, ESS 2004, 2010)
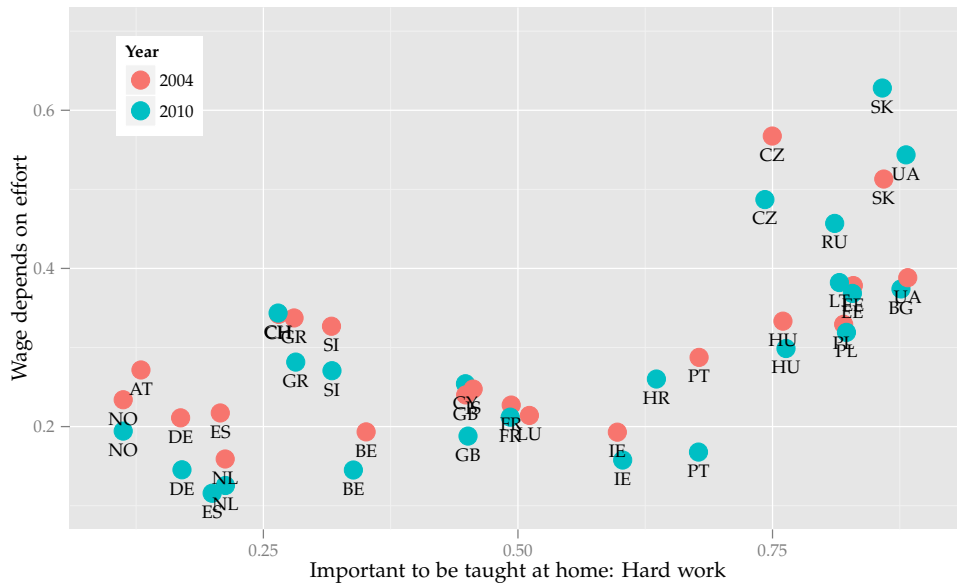
to the 3rd quartile in work ethic makes an individual 25 percent (7 p.p.) more likely to think that his wage depends on his effort.

Figure 1.2 suggests that Eastern European (former Soviet) and Western European countries have different responses for incentive steepness and work ethic. Thus, in all my regressions, I control for being Eastern European (my results are stronger otherwise).[13] The regressions show that while the Soviet heritage is important, it does not drive my results.

Unfortunately, work ethic does not vary enough within countries for us to obtain significant results using country dummies.[14] I believe that this difficulty is due to culture being more widespread than my measure of regions or to the fact that countries in Europe are too small to allow for large within-country variations. This difficulty is also consistent with other country-level variables that cause both higher work ethic and higher-powered incentives. To try to address that possibility, in Columns 3 and 4, I control for a variety of regional- and country-level covariates, including other cultural traits that might influence both dissemination of work ethic and incentives. While many of these covariates have significant effects, they only slightly change my estimates.

1.3.3

[13]While not central to my argument, I conjecture that the higher prevalence of work ethic among Eastern European countries results from the cultural promotion of work values by socialist governments and by the prevalence of incentive wages (piece-rate pay) in industry during the Soviet period. Regressions without these controls are available upon request.

[14]This specification is also available upon request.

Tabela 1.2: European Values Study and European Social Survey (EVS 2008, ESS 2004, 2010)

	(1)	(2)	(3)	(4)
Work imp. (region)	0.1304	0.1430	0.1473	0.1143
	(0.0285)***	(0.0295)***	(0.0349)***	(0.0327)***
Female		-0.0584	-0.0596	-0.0603
		(0.0073)***	(0.0072)***	(0.0072)***
Education (yrs)		0.0002	-0.0009	-0.0007
		(0.0014)	(0.0013)	(0.0013)
Self-employed		0.6070	0.6000	0.5937
		(0.0428)***	(0.0546)***	(0.0637)***
Left-right (region)			-0.0522	-0.0522
			(0.0157)***	(0.0140)***
Bribe just (region)			0.0660	0.0722
			(0.0547)	(0.0524)
GDP pc (region)			-0.0162	-0.0002
			(0.0128)**	(0.0143)
Education (region)			0.0618	0.0447
			(0.0162)***	(0.0157)**
GDP pc (country)				-0.0549
				(0.0314)**
Industry % (country)				0.0022
				(0.0012)
East-west (east)	0.1217	0.1138	0.1242	0.1138
	(0.0174)***	(0.0199)***	(0.0227)***	(0.0224)***
Year (2010)	-0.0286	-0.0214	-0.0240	-0.0313
	(0.0087)*	(0.0125)	(0.0118)*	(0.0120)**
Country-level controls	N	N	N	Y
Region-level controls	N	N	Y	Y
Individual-level controls	N	Y	Y	Y
R^2	0.040	0.085	0.093	0.095
N	34375	16677	16664	16664

 The dependent variable is the respondent's answer to the following question: "Please tell me how true each of the following statements is about your current job: my wage or salary depends on the amount of effort I put into my work."I code "Very True"or "Quite True"as one and zero otherwise. Corrected standard errors are clustered at the regional level. P-values are significant at the 0.05 (*), 0.01 (**), and 0.001 (***) levels. Individual-level controls are religion, gender, age, age squared, years of education, employment relation, NACER2-level industry controls, and whether there is a child at home. Region-level controls are (log) income per capita (PPP), average years of education, average political leaning (left or right), response to whether bribes are sometimes justified, average concern for others, response to whether freedom or equality is more important, and opinion about the government. Country-level controls are (log) GDP per capita and industry participation. There are 30 countries in the sample.

World Values Survey and Execucomp

Finally, in my third empirical exercise (Table 1.3), I test the relation between work ethic and steepness of incentives, without relying on self-reported measures of the latter. To do so, I use Standard and Poor's Execucomp database on executive compensation in the US. I match those data with US regional data from the World Values Survey (WVS). Execucomp provides data for all S&P 1500 firms since 1994. I gather these detailed executive earnings data to create measures of incentive steepness. Obtaining reliable measures of incentive steepness is difficult for the majority of the population, as their contracts usually involve efficiency wages and career concerns. Data for executive compensation, however, is widely available, distinguished by incentive type, and as most executives receive bonuses, more reliable.

I use the WVS waves from 1995, 1999, and 2006. I match the data for each wave to the executive data from the year before the survey, the year of the survey, and the year after the survey (e.g., for the first wave, I use 1994, 1995, and 1996).[15] A difficulty with this dataset is that I only have data on work ethic by region.[16] Still, as I have data for all three waves and for nine years of executive compensation, it is possible to obtain precise estimates, even when controlling for regional fixed effects.

In Columns 1, 2, 4, and 6, I use bonuses as a proportion of bonuses and wages as a measure of incentive steepness. Compared to the base specification in Column 1, increasing the controls in Column 2 increases the point estimates, suggesting that the simple correlations are downward biased. In Columns 3 and 5, I take into account that good executives might try to protect themselves from risk and consider options and other incentive packages. My measure of incentive steepness is then $1 - \frac{\text{wage}}{\text{total payment}}$. These results are weaker but still consistent with my hypothesis. In Column 4, my preferred specification, an increase from the first to the third quartile in regional work ethic increases the proportion of bonuses in the compensation contract by 1.5 p.p., a five percent increase over the mean.

I use very specific industry-level controls to address the possibility that I am capturing different compensation behaviors from different industries that are spatially sorted. Furthermore, the analysis considers within-country variation therefore controlling for country-level institutional features, but it also controls for regional fixed effects using temporal variation. Columns 4 and

---

[15]The results are robust to using executive data only for the year of the wave and using only some of the waves.

[16]Ten regions are included in the data: East South Central, West South Central, New England, South Atlantic, East North Central, West North Central, Rocky Mountain States, Middle Atlantic States, Northwest and California.

Tabela 1.3: WVS for USA and Execucomp (WVS 1995, 1999, 2005, Execucomp 1994-2006)

	(1)	(2)	(3)	(4)	(5)	(6)
Work imp. (region)	0.1386	0.2156	0.1620	0.1884	0.0982	0.0975
	(0.0584)*	(0.0912)*	(0.0747)*	(0.0625)**	(0.0622)	(0.0445)*
Age		0.0082	0.0201	0.0084	0.0200	0.1609
		(0.0017)***	(0.0014)***	(0.0016)***	(0.0015)***	(0.1596)
Age squared		-0.0001	-0.0002	-0.0001	-0.0002	-0.0015
		(0.0000)***	(0.0000)***	(0.0000)***	(0.0000)***	(0.0015)
Interlock		-0.0507	-0.0671	-0.0507	-0.0687	
		(0.0206)*	(0.0198)***	(0.0214)*	(0.0030)***	
CEO (no)		-0.0326	-0.0641	-0.0329	-0.0647	0.1241
		(0.0025)***	(0.0032)***	(0.0026)***	(0.0035)†	(0.2597)
Year dummies	Y	Y	Y	Y	Y	Y
Region FE	N	N	N	Y	Y	N
Executive controls	N	Y	Y	Y	Y	*
R^2	0.189	0.345	0.181	0.350	0.186	0.819
N	60518	29492	29319	29492	29319	30

The dependent variable in Columns 1, 2, 4, and 6 is $\frac{\text{SALARY}}{\text{TOTAL WAGE}}$. For Columns 3 and 5, the dependent variable is $1 - \frac{\text{BONUS}}{\text{BONUS+SALARY}}$. P-values are significant at the 0.05 (*), 0.01 (**), and 0.001 (***) levels. For each wave t of the WVS, I use executive compensation data for the years $\{t-1, t, t+1\}$. Executive-level controls are age, age squared, gender, being the CEO, being the CFO, whether the board is interlocked, and detailed SIC industry codes. All standard errors are clustered at the regional level. The last column aggregates the variables by region and WVS year. Given the small number of observations, I removed gender, interlock, and SIC codes. There are 10 regions in the sample.

5 show that controlling for regional fixed characteristics has small effects on the estimated coefficients.

The coarse nature of the regional data might call into question the robustness of the results. In column 6, I apply the most challenging test possible: I aggregate the observations by WVS wave and region, obtaining 30 observations (three waves for each of the ten regions). Surprisingly, the correlation still holds. (In this specification, I have to drop the interlock and gender variables to maintain statistical power.)[17] I believe that this supports the robustness of the empirical results of this exercise.

Finally, using executive compensation raises the question of how the results generalize to the rest of the working population. I hope that future research will use better data and be able to test the implications of my model more carefully, focusing on causal identification and external validity.

1.4
Model

In my model, there is a population of measure two consisting of overlapping generations of agents that live two periods. In the first period, they acquire by a cultural transmission process a binary trait $k \in \{W, N\}$ that indicates the presence (W) or absence (N) of a work ethic. In the second period, they are employed (and remunerated) by a firm and choose whether to instill a work ethic in their offspring. For simplicity, I assume that reproduction is asexual and that each individual has exactly one child. Hence, the measure of adults in the population is one, and I denote the proportion of adults with a work ethic as x^W.

I assume that the work-related decisions and the cultural transmission choice are separable and consider them separately. I start by analyzing the firm problem.

1.4.1
Moral Hazard in Teams Game

I model employment in firms as a moral hazard in teams game. There is a continuum of measure one of firms, each being matched with two randomly chosen individuals from the population to engage in production each period.[18] I assume that each firm employs its two agents in a project with value ρ, and

---

[17] These variables suffer from excessively high homogeneity. Only 5% of executives in my sample are female, and 1.7% of boards are interlocked. I also drop the SIC industry controls given their number.

[18] Consider, for example, that firm f is matched with agents $f/2$ and $(f+1)/2$, which is an unique matching almost everywhere.

the probability of success is given by $\theta a_i^\alpha a_j^\alpha$, with $\alpha \in (0,\ 1)$, $a_i,\ a_j \in [0,\ 1]$ and $\theta \in (0,\ 1]$. In this equation, a_i (a_j) represents the effort exerted by agent i (j), and θ represents the efficiency of effort of agents i and j. With some abuse of notation, I will use subscripts i and j to refer to both agents and agent types.

This functional form entails that the probability of success is supermodular, representing a production process in which the effort of workers is complementary. An example is an assembly-line production plant, where the productivity of a worker is directly related to the speed at which he receives production inputs, which are the outputs produced by the previous workers in that line.

As employee effort is non-observable, the firm must choose a compensation scheme that is incentive compatible. I assume that the firm has complete information about whether the agent has work ethic.[19] Importantly, I also assume that there are institutional constraints that forbid a negative wage (i.e., limited liability). Limited liability is essential to my model, as otherwise the firm would always equal the workers' expected payoff to the utility gained by their outside option, irrespective of type. Given these assumptions, since the project outcome is binary, the firm can restrict its attention to linear incentive schemes formed by a bonus b for each combination of types, *viz.*, $\{b_{ij}\}_{i,j \in \{W,\ N\}}$.[20] Both workers and firms are risk neutral.

Agents have a quadratic cost of effort that depends on the traits they acquired during their childhoods. Those with a work ethic have a lower marginal cost of effort $\kappa_W a_i$ than workers without it $(\kappa_N a_i)$—namely, $\kappa_W < \kappa_N$—representing the moral justification for effort that eases their toil. However, work ethic entails costs beyond those related to work. For example, in religious thought, while the promise of eternal recompense for hard work (as expressed in the epigraph) can ease the cost of effort, it cannot be dissociated from the ascetic view that leisure and comfort are worthy of "absolute moral condemnation,"as Baxter claims. To account for this, I assume that agents with work ethic suffer a cost K_W that is exogenous to their work behavior, indicating an inability to enjoy free time with family and friends.[21]

Accordingly, the utility of an agent of type i with a coworker of type j (defining $\mathbb{I}_A$ as the indicator function of A) is:

---

[19] I discuss this and other main assumptions of my model in the next section.

[20] Restricting the compensation scheme to a bonus without a base wage is without loss of generality given limited liability if the outside option is not too high (e.g., if it is zero).

[21] Since $\kappa_W < \kappa_N$, without the cost K_W, it would always be advantageous to acquire a work ethic (in my specification of cost linear in types). This is the simplest specification, which implies that for low effort, having a work ethic is worse, while for high effort it is advantageous; this result fits the concept of work ethic as motivated in Section 2. Intuitively, this fixed cost K_W represents the costs of having a work ethic that are unrelated to work behavior, such as the lower utility derived from consumption and leisure time (e.g., vacations) or worse social relations.

$$\left(\theta a_i^\alpha a_j^\alpha\right) b_{ij} - \mathbb{I}_{\{i=W\}} K_W - \kappa_i \frac{a_i^2}{2}. \tag{1-1}$$

Moreover, throughout the paper I assume that the outside option is redundant given limited liability, (*viz.*, the outside option $\bar{u} \leq 0$). Adding a binding outside option (as long as it is not high enough to make the model trivial) would not change the results.

The firm chooses the optimal incentive scheme to maximize its profit, knowing that the agents choose their effort in an incentive compatible way given the proposed scheme. Here, $a(b;\ \kappa_i, a_j)$ denotes the effort a worker exerts when faced with bonus b, having type κ_i and coworkers with effort a_j, that is:

$$a(b;\ \kappa_i, a_j) = \arg\max_a \left(\theta a^\alpha a_j^\alpha\right) b - \mathbb{I}_{\{i|i=W\}} K_W - \kappa_i \frac{a^2}{2}.$$

Then, the problem of the firm is given by:

$$\max_{b_{ij},\ b_{ji}} \theta a_i a_j \left(\rho - b_{ij} - b_{ji}\right), \text{ such that } a_i = a(b_{ij},\ \kappa_i,\ a_j) \text{ and } a_j = a(b_{ji},\ \kappa_j,\ a_i). \tag{1-2}$$

Since effort is complementary, when an agent has a hard-working coworker, his own effort becomes more productive, and the firm will choose to incentivize him to work harder as well. In fact, defining $a_i^*(\kappa_i, \kappa_j)$ as the equilibrium effort choice of an employee of type i with a coworker of type j, the optimal incentive scheme is such that the effort is given by:

$$a_i^*(\kappa_i, \kappa_j) = \left[\frac{\theta \rho \alpha^2 a_j^*(\kappa_j, \kappa_i)^\alpha}{2\kappa_i}\right]^A = \left[\frac{(\theta \rho \alpha^2)^{1+\alpha A}}{2^{1+\alpha A} \kappa_i \kappa_j^{\alpha A}}\right]^{\frac{A}{1-\alpha^2 A^2}}, \tag{1-3}$$

where $A = \frac{1}{2-\alpha}$. Intuitively, an agent with a higher cost of effort is costlier to incentivize, and in equilibrium, he will work less. More importantly, as work effort is complementary, having a coworker who exerts low effort reduces an agent's productivity, inducing the firm to offer lower bonuses and thus elicit lower effort.

The second part of equation (1-3) presents the optimal effort choice as a function of the agents' traits. Workers without a work ethic and workers whose peers do not have a work ethic work less in equilibrium. Even more interestingly, while all individuals work harder when their peers have a work ethic, this difference is greater for workers who have a work ethic themselves. The same is true for the total compensation received by workers: an individual matched to a peer with a work ethic receives higher compensation than one matched to a coworker without a work ethic, but this effect is greater for individuals who have a work ethic themselves (this can be seen by noting that

the compensation received by an agent is a convex function of his effort).[22]

The main result of this section establishes that the same is true for utilities: it is always beneficial to have coworkers with a strong work ethic, but this benefit is higher for workers who have a work ethic themselves. Symmetrically, having a work ethic is especially advantageous when the individual is likely to enjoy a hardworking coworker. Intuitively, in a production line, the faster the inputs reach a worker's station, the faster that employee can work. However, this is only useful inasmuch as she can produce at least as fast as the inputs arrive (by exerting enough effort). For fast workers, the advantage of being in a fast production line (when earning piece rates) is very large, while for a slow worker, it is negligible.

I now proceed to formally state the main result of this section, which is discussed above. For that, I define the *matching payoff* of a worker of type i when matched with a coworker of type j as the utility she receives under the optimal compensation scheme, given her coworker's type. I denote it as $V(i, j)$.

Using the matching payoff, I can define the *ex ante utility* of an individual of type i ($i \in \{W, N\}$) as:

$$U^i(x^W) \equiv x^W V(i, W) + (1 - x^W)V(i, N), \qquad (1\text{-}4)$$

where x^W represents the proportion of adults with a work ethic in the society. The ex ante utility is the expected utility an individual will receive in the labor market, given his type, but without knowledge of whether his future coworker has a work ethic.

The main result of this section is stated below and proven in Appendix A (together with the other Theorems in the main text).

Proposição 1.1 *The benefit of having a work ethic is greater for workers whose peers also have a work ethic. Formally, the matching payoff function is increasing differences in types, i.e. $V(W, W) - V(N, W) > V(W, N) - V(N, N)$.*

1.4.2
Cultural Transmission Process

Having established the labor market payoff of different cultural traits, I can define the parent's problem: choosing whether to transmit his cultural trait to his offspring. As a work ethic is an internalized social norm that

---

[22]Since the increasing convex transformation of a supermodular function is supermodular. The total compensation received by an agent—the bonus times the probability of success—is presented in Equation A-1, in the Appendix.

directly affects the preferences of individuals, I follow the literature on the intergenerational transmission of cultural norms established by (28), (29), and (01, 02).

Parents of type i choose a socialization effort $d^i \in [0, 1]$ in an attempt to pass their trait on to their offspring. This socialization effort succeeds with probability d^i, in which case the offspring will acquire the parental trait (vertical socialization). If this effort is unsuccessful, the child will acquire the trait of a random element of the population (horizontal/oblique socialization). In that case, the probability that the trait acquired is the same as the father's is x^i, where x^i is the proportion of trait i in the society.

Concisely, the probability that a child of a parent of type i acquires trait i (resp., j) is given by:

$$P^{ii}(d^i;\ x^i) = d^i + (1 - d^i)x^i$$

$$P^{ij}(d^i;\ x^i) = (1 - d^i)(1 - x^i).$$

Parents are altruistic and consider their offspring's expected labor market utility when deciding whether to transmit a work ethic. As shown in the previous section, the moral hazard in teams problem entails that the ex ante utility of having a work ethic is a function of the (beliefs about the future) proportion of individuals in the population with a work ethic. I assume, however, that parents are myopic: they choose their socialization effort based on the dissemination of work ethic they observe themselves in the society. This can be seen as a form of imperfect altruism, in the sense that parents observe their own environment and payoff when considering their children's utility.

This form of imperfection is not the same as that in (01). There, they assume exogenous cultural intolerance, and they propose imperfect altruism in the form of a parental view that their own traits always produce higher utility for their offspring. (Intuitively, it represents a case in which parents cannot teach what they do not believe.) I consider my assumption a generalization of theirs to a scenario wherein the dissemination of a trait influences the payoffs of the agents directly.

Moreover, in Bisin & Verdier (2000), it is natural to impose the condition that parents can only socialize children to value their own traits, as they would never want to do otherwise. In my model, I impose the same condition for simplicity of exposition, with the intuition that a parent cannot pass on a belief that she does not share herself. In my model, a deeply religious parent cannot teach her offspring to disregard religion, as it would go against her beliefs. She can, however, exert little effort in socializing her offspring, as it

will not be advantageous for her child to be religious given the cultural values of the society.[23]

Given the cultural transmission process, parents of type i maximize with regards to socialization effort d^i:

$$P^{ii}(d^i; \ x^i)U^i(x^W) + P^{ij}(d^i; \ x^i)U^j(x^W) - \frac{(d^i)^2}{2},$$

where U^i is the ex ante utility of the individual of type i introduced in the last section. The first order conditions imply that $d^W = \max\{(1 - x^W)\Delta U(x^W), \ 0\}$. As in the literature, I call $\Delta U(x^W) \equiv U^W(x^W) - U^N(x^W)$ the cultural intolerance of work ethic. As is clear in equation (1-4), cultural intolerance is endogenous: it depends on the proportion of the population with a work ethic. Moreover, Proposition 1.1 states that individuals with a work ethic gain more by having a peer with work ethic than do individuals without that trait. Hence, this is a game of strategic complements, and my cultural transmission process is a process of cultural conformity (see (13)). I formalize this assertion in Proposition 1.2 below.

Proposição 1.2 *Assume that parameters are not such that it is always better or always worse for an agent to have a work ethic, independently of who he works with. Then, there exists $x^* \in (0, \ 1)$ such that if the dissemination of work ethic at a moment t, x_t^W, is less than x^*, the proportion of individuals with a work ethic converges to zero, and if x_t^W is larger than x^*, then it converges to one.*

Figure 1.3 below illustrates Proposition 1.2 for a set of example parameters. In societies with low dissemination of work ethic, parents expect that their children are more likely to be employed with low-effort coworkers and be paid low-powered wages. As such, they find it in their offspring's best interest to transmit cultural traits that oppose having a work ethic—that is, that promote leisure and the importance of personal life rather than that of a career. However, in societies with high dissemination of work ethic, the average compensation scheme is high-powered, so the returns to high effort are large. In that case, parents will altruistically choose to disseminate work ethic, propagating beliefs that work is important and that it provides a purpose in life (a *calling*).

---

[23]Nonetheless, I note that this has no effect on the results (except to change the rates of convergence of different dynamic equilibria). Indeed, one would obtain similar results by substituting this cultural transmission process with properly defined replicator dynamics. I refrain from doing so, however, because this model provides a more intuitive representation of the intergenerational transmission of cultural values.

Figura 1.3: Dynamics of work ethic for an example parameter profile. ($\alpha = 0.7$, $\theta = 0.25$, $\rho = 10$, $\kappa_N = 0.8$, $\kappa_W = 0.4$, $K_W = 0.6$)
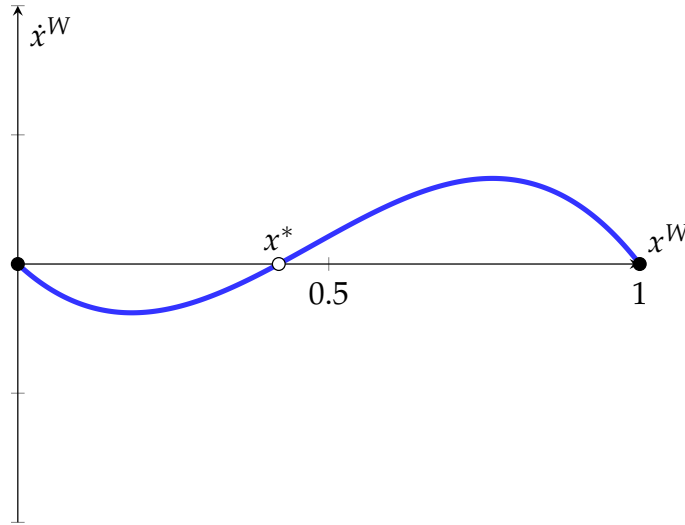
Figure 1.3 also highlights two of the main features of my model: history matters in the sense that the dynamic equilibria are path dependent and very similar initial conditions can lead to drastically different equilibrium cultural outcomes. Together, these characteristics imply that cultural and economic shocks in the past can have, through cultural propagation, effects on current economic outcomes. In particular, they explain what Weber called the "ghost of dead religious beliefs."Protestantism triggered the dissemination of a strong work ethic across Northern European societies, which continued to propagate long after the view of work as a means to salvation had lost importance in Western life. Although the work ethic was initially disseminated through a religious shock, as individuals became more hard-working, firms became more productive and started employing steeper incentives, making the transmission of work ethic beneficial even after it lost its religious meaning.

Furthermore, as I argue in Section 6.1, parental socialization and firm technological choices are related: after the Protestant Reformation, firms chose production processes and managerial techniques focused on complementarity between workers to increase the productivity of a now more hard-working population, which in turn made the dissemination of a culture that emphasizes hard work more valuable.

In this Section, I presented a model that predicts multiple equilibria in effort, incentives, and dissemination of work values in society. This model is consistent with empirical findings that a strong work ethic is correlated with measures of effort and incentive steepness, with previous literature emphasizing the importance of personal background to explain differences in work effort

across regions and stylized facts in the literature pointing to the significant importance of effort exertion in explaining productivity differentials across nations and periods. In the next section, I discuss some assumptions of my model, and in Section 6, I extend it to include the firm technological choice.

1.5
Discussion

In this section, I discuss some of the assumptions and results of my model. First, I argue that the assumption that the firm knows the workers' types and can discriminate among them when devising the incentive scheme is reasonable. Then, I argue that incentive schemes based only on bonuses imply no loss of generality, even though only a small fraction of the work force actually receives incentive wages. Finally, I discuss complementarity of effort and welfare consequences of the model.

To motivate common knowledge of types, imagine that the firm can observe the output y of an agent in each period but not the effort. If the effort is constant, as in my model, then by observing the output a sufficient number of times, the firm can infer the type of the agent, with arbitrarily small degree of uncertainty. As a result, if the contractual relationship is sufficiently long, it is with no loss of generality to suppose that the firm knows the type of agent, even if it cannot observe the effort the agent supplies at any given time.

Moreover, if the firm not only pays bonuses when the output is realized but also adjusts their wages (for example, through raises), then over time, it can make the payment of different types of agents converge to their optimal values by choosing an appropriate adjustment rule, even if direct discrimination is forbidden or discouraged. Therefore, there is no loss of generality in assuming that firms can freely choose incentive schemes contingent on workers' types.

Another assumption of the model is that firms rely solely on incentive wages to elicit effort, while in actuality, many other types of incentives are available, most commonly, career concerns and efficiency wages. While my model does not explicitly account for these other forms of incentives, the intuition is the same, and the generalization would be straightforward. As a result, although my model considers only piece-rate workers, my argument is general.

Importantly for the interpretation of the model, the incentive scheme paid by the firm to an agent of type i with coworkers of type j is:

$$b_{ij}(\kappa_i, \; \kappa_j) = \frac{\kappa_i}{\theta\alpha} \left[\frac{(\theta\rho\alpha^2)^{1+\alpha A}}{2^{1+\alpha A}\kappa_i\kappa_j^{\alpha A}} \right]^{\frac{(2-\alpha)A}{1-\alpha^2 A^2}} \left[\frac{(\theta\rho\alpha^2)^{1+\alpha A}}{2^{1+\alpha A}\kappa_j\kappa_i^{\alpha A}} \right]^{-\frac{\alpha A}{1-\alpha^2 A^2}} \tag{1-5}$$

Regarding the effect of having a coworker with a strong work ethic, there are two opposing forces at work. On the one hand, since the marginal cost of effort is lower, for a given level of effort, the firm needs to compensate the worker less, reducing the bonus. On the other hand, the firm wants to elicit more effort from a now more productive worker and thus needs to pay him a larger bonus for that effort. Equation (1-5) shows that the second effect is stronger: a lower cost of effort for a worker and for his coworker results in higher bonuses, that is, higher-powered incentives. In this sense, my model predicts that societies with a more broadly disseminated work ethic have higher-powered incentives.

A hallmark of the model is that work effort is complementary. Complementarity is a topic that has received increasing attention in the literature on organizational economics (e.g., (33), (26), (25)). Although observational evidence points to effort complementarity as an important part of modern production processes (for example, Fordism and lean manufacturing), so far, it has not been an important part of models of organizational behavior.

One reason is that incorporating complementarity causes significant losses of tractability in models of moral hazard, and in this regard, another contribution of this paper is to suggest a tractable model of effort complementarity. A second is that obtaining causal evidence on the existence of complementarity is difficult. Observational data that corroborates the existence of sorting on worker diligence among sectors and firms notwithstanding, it is difficult to disentangle the effects of effort complementarity from those of other explanations.

Nevertheless, including complementarity in the model has an appealing implication. In the model, a worker's income depends on the effort chosen by his coworkers, which is consistent with differing wages across firms and sectors and correlated within a firm, a features that one observes empirically. (27), for example, notes that occupation and employer identity explain 90 percent of the observed variation in wages.

The model has well-defined implications for welfare, although welfare comparisons in models with changing preferences are notoriously problematic (see (23)). According to Proposition 1.2, there are only two asymptotically stable dynamic equilibria: one with a fully disseminated work ethic and one in which there is no work ethic in the society. Welfare analysis is thus reduced to comparing these two equilibria.

Indeed, in the model, the dynamic equilibria (of Proposition 1.2) are ranked in terms of welfare, as all workers (and firms) prefer to have coworkers with a strong work ethic. In a more basic sense, having coworkers with a work ethic results in a higher probability of success, even if the worker's own effort is constant. More subtly, it makes the agent more productive, and more productive agents work harder, receive larger shares of the surplus, and have higher utilities. My model, therefore, represents a world in which effort pays: individuals who work harder have higher utilities.[24] Hence, for any parameter profile for which it is not always worse to have a work ethic, the equilibrium with full dissemination of a strong work ethic is more efficient, based on the argument above and Proposition 1.1, $V(W, W) > V(N, W) > V(N, N)$.

1.6
Extensions

In this section, I analyze possible extensions to my benchmark model. A question raised by the previous analysis is why firms in societies with limited dissemination of work ethic would choose complementary technologies, as this causes a decrease in productivity. In Section 6.1, I augment the base model to account for the firm technology choice, and I show that firms in low work ethic societies indeed prefer non-complementary technologies. Moreover, if strongly complementary technologies are interpreted as being more technologically advanced than non-complementary ones (as in Fordism or lean manufacture versus more traditional forms of production), then this extension provides a rationale (akin to (22)) for the differences in technological adoption that one observes empirically.[25] In this extended model, I characterize the phase diagram induced by the firm technology choice and the parental socialization process. Finally, in Section 6.2, I show how the dynamic equilibria of this model respond to changes in the underlying parameters to better comprehend why some regions have a widely disseminated work ethic while others do not.

---

[24]It is important to note that this is in fact a general property of moral hazard problems with limited liability. To see this, consider a general moral hazard problem with a worker who has utility function $by(a) - c(a)$ and works for a firm with profit $y(a)(1 - b)$, where a is effort and b a bonus for success. Then, note that the total derivative of the agent's utility in regards to effort (repaid through optimal incentives) is:

$$\frac{\mathrm{d}u}{\mathrm{d}a}(a) = \frac{c''(a)y'(a) - c'(a)y''(a)}{y'(a)^2}y(a),$$

which represents how much better off an agent is by working (in equilibrium) more, for example, because of higher productivity. It is positive under the very standard assumptions that y is increasing and concave and c is increasing and convex.

[25]In this section, my argument is similar to (22). However, besides my focus on culture, I study the effort choice of the workers and endogenize the compensation choice of the firms, while their model is based on sorting on skill.

1.6.1
Firm Selection

In the extended model, firms can choose a technology $q \in \{C, S\}$, where C stands for the complementary technology presented in the base model, and S stands for a separable technology in which the project is divided in two separate tasks, each with a value $\gamma\rho$, and employing a single worker with success probability θa_i^α, where γ is the relative productivity of the complementary and separable sectors of the economy. If one portrays these sectors as representing modern and traditional industries, respectively, then γ corresponds to their relative productivity. Therefore, one would expect that events such as the Industrial Revolution change γ (namely, they decrease it), affecting firm (and indirectly parental) choices.

I assume that there is a lump sum cost of complementary technologies ζ, as these technologies are more modern and expensive production processes (both to acquire and to maintain).

In a firm with separable technology, the utility that agent i receives is independent of the effort (or type) of his coworker and is given by:

$$(\theta a_i^\alpha)\, b_i - \kappa_i \frac{a_i^2}{2} \tag{1-6}$$

Furthermore, as opposed to when employed by firms with complementarity, the effort chosen by an agent of type i when faced with the separable technology firm's incentive scheme depends only on his own type. It is given by:

$$a_i^*(\kappa_i) = \left[\frac{\rho\gamma\theta\alpha^2}{2\kappa_i}\right]^A \tag{1-7}$$

I denote the proportion of firms with complementary technology in the economy by β^C, and I assume that competition among firms causes low-achieving ones to be replaced by new entrants. Specifically, I consider the firms' fitness to be proportional to their revenue relative to the mean revenue in the society.[26] I model the selection process as a (discrete) replicator dynamic with a speed-scaling parameter s. Denoting by $Y(\beta_t^C,\, x_t^W;\, q)$ the expected revenue (net of complementary process cost ζ) of a firm with technology $q \in \{C, S\}$ in period t, the proportion of firms with complementary technology in period $t+1$ is then given by:

$$\beta_{t+1}^C - \beta_t^C = s\beta_t^C \left[Y(\beta_t^C,\, x_t^W;\, C) - \sum_{q \in \{C,\, S\}} \beta_t^q\, Y(\beta_t^C,\, x_t^W;\, k) \right] \tag{1-8}$$

---

[26]I use revenue instead of profits as a selection mechanism because it is more tractable, as profits are a highly nonlinear function of the agents' types, and clarifies the interpretation.

Moreover, the existence of firms with separable technology affects the parental socialization decision. Now, parents must take into account that with probability β^C their offspring will work in firms with complementary technology and receive the payoff given by (1-4) and that with probability $1 - \beta^C$ they will work in a firm without complementarity and receive $V^S(i)$, regardless of their coworker's type, where $V^S(i)$ is the utility of the working offspring given by (1-6) and evaluated at the optimal incentive scheme.

Consequently, the agent's ex ante labor market utility is given by:[27]

$$U^k(x^W,\ \beta^C) = \beta^C \left[x^W V^C(k,\ W) + (1 - x^W)V^C(k,\ N) \right] + (1 - \beta^C)V^S(k)$$
$$(1\text{-}9)$$

Figura 1.4: Dynamic system of x^W and β^C for an example parameter profile. ($\theta = 0.25$, $\rho = 10$, $\alpha = 0.7$, $\kappa_W = 0.4$, $\kappa_N = 0.8$, $\gamma = 0.55$, $\zeta = 0.8$, and $K_W = 0.45$)
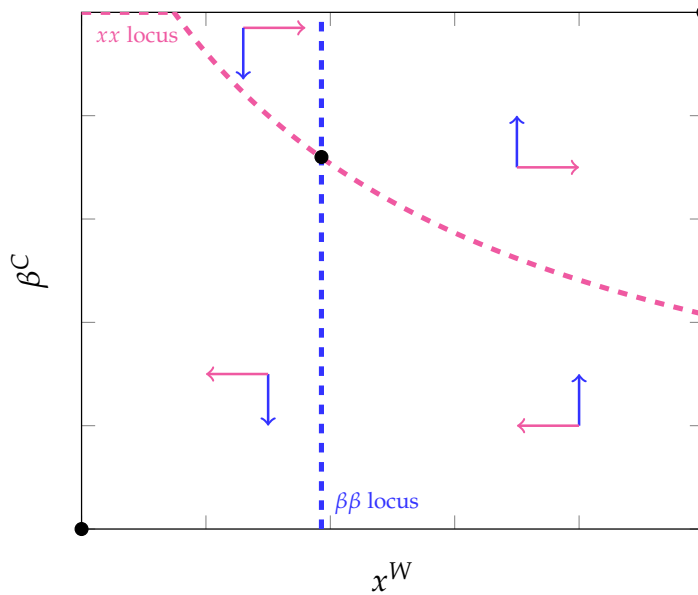
Equation (1-8) and the cultural transmission process implied by (1-9) together generate a two-dimensional (discrete) dynamic system through which I can study the continuous approximation by use of a phase space. This is done in Figure 1.4. The xx-locus is the set of points $(x^W,\ \beta^C)$ for which the proportion of agents with work ethic (x^W) is constant. Analogously, the $\beta\beta$-locus is the set of points in the plane for which the proportion of firms with complementary technology β^C is constant. Notice that the loci include the borders, as neither the replicator dynamics nor the cultural transmission process incorporate mutations. The intersections between the loci represent

[27]Throughout this section, I refer to the matching payoff defined in Section 4, $V(i,\ j)$, as $V^C(i,\ j)$ to emphasize that it is the payoff of an agent i (with coworker j) working in a firm with complementary technology.

stationary points in the dynamic system, of which (0, 0) and (1, 1) are stable, and the interior stationary point is a saddle point.

Intuitively, when the proportion of agents with a strong work ethic in the society is low, firms expect to employ few hard-working agents. With modern complementary technologies, shirkers hinder the productivity of their peers, and firms avoid this by splitting them up in separable production processes. When work ethic is highly disseminated, firms benefit from combining workers' effort in complementary tasks, as the probability of having hard-working employees is high. As Henry Ford learned in Fordlândia, while a production system wherein each worker's effort is highly dependent of his peers' might result in large gains of productivity when employees have a strong work ethic, in the opposite environment, such a system can be disastrous.

I formalize this discussion in Proposition 1.3. It states the condition under which the behavior of the extended model is well represented by Figure 1.4 above, that is, when I have two stable equilibria, one with low dissemination of work ethic and separable technology and the other with high dissemination of work ethic and complementary technology.

Proposição 1.3 *Regarding the dynamic behavior of the model with firm technology selection:*

(i) The $\beta\beta$-locus is a vertical line in the $x^W \times \beta^C$ plane such that for lower levels of x^W, the proportion of firms with complementary technology β^C is decreasing, and for higher levels of x^W, it is increasing.

(iia) If $V^S(W) > V^S(N)$, then the only asymptotically stable dynamic equilibrium is the full dissemination of a strong work ethic.

(iib) Otherwise, the xx-locus is downward sloping and crosses the $\beta\beta$-locus at most once on an interior point. If that intersection exists, then there are two asymptotically stable equilibria, one with a fully disseminated work ethic in which all firms choose complementary technologies and one wherein no individual has a strong work ethic and all firms prefer the separable technology. There is another stationary point (at their intersection), which is a saddle point. (If the intersection is not interior, then the two first points are still stationary, but only one is asymptotically stable.)

The necessary condition for the behavior of this system to resemble Figure 1.4 is that $V^S(W) < V^S(N)$. However, if it is indeed superior to have a strong work ethic even when the technology is separable (i.e., $V^S(W) > V^S(N)$), then the dissemination of work ethic will always be high, with the change by firms to complementary technologies only amplifying this effect. Therefore, the phase space would not resemble Figure 1.4. Nevertheless, this

would be contrary to the argument (presented in Section 2) that a work ethic was absent in Medieval Europe, having emerged instead during the Protestant Reformation and Industrial Revolution.

1.6.2
Comparative Dynamics

In this section, I discuss how the dynamic equilibria are affected by changes in the basic parameters of the model. Because the cultural transmission process and the replicator dynamics of firm selection have no mutations, in any asymptotically stable equilibrium (namely, (0, 0) and (1, 1)), parameter changes do not have any effect.[28] I can, however, study how changes in parameters affect the size of the basins of attraction of the dynamic equilibria. These comparative dynamics for the extended model presented in Section 6.1 are summarized in Proposition 1.4 below.

Proposição 1.4 *Assume a phase space like Figure 1.4 (see Proposition 1.3). Then, an increase in human capital (θ) and in the value of project (ρ) and a decrease in the fixed cost of having a work ethic (K_W) and in the relative productivity of the separable sector (γ) increase the basin of attraction of the equilibrium with high dissemination of a strong work ethic and a high proportion of firms with complementary technology (and* vice versa*).*

Intuitively, when workers are more productive (higher θ) or the project they work on is more valuable (higher ρ), firms optimally choose to offer larger bonuses, eliciting higher effort. As higher effort is less costly for individuals with a work ethic than for those without, transmitting a strong work ethic becomes more valuable for parents. Furthermore, as more productive workers exert more effort, it also becomes advantageous for the firm to choose modern complementary technologies, as the gains from complementarity are increased by higher effort. In this scenario, parents want to strongly instill a work ethic in their children and firms choose more complementary technologies, and the equilibrium with a widely disseminated work ethic and complementary technology becomes more likely (as in there are more initial states that converge to that equilibrium).

Moreover, a decrease in the cost of having work ethic K_W (e.g., the moral condemnation of leisure) makes a strong work ethic more desirable, while a reduction in the relative productivity of the separable sector γ (for example, technological advancements that rely on complementarity, such as assembly

---

[28]The model can be extended to incorporate random shocks to the proportion of workers with a strong work ethic. In that case, changes in the parameters would alter the probability that the equilibrium changes. The intuition of the model would remain unchanged, however.

Figura 1.5: Dynamic system of x^W and β^C ($\theta = 0.25$, $\rho = 10$, $\alpha = 0.7$, $\kappa_W = 0.4$, $\kappa_N = 0.8$, $\gamma = 0.55$, $\zeta = 0.8$, and $K_W = 0.4$)
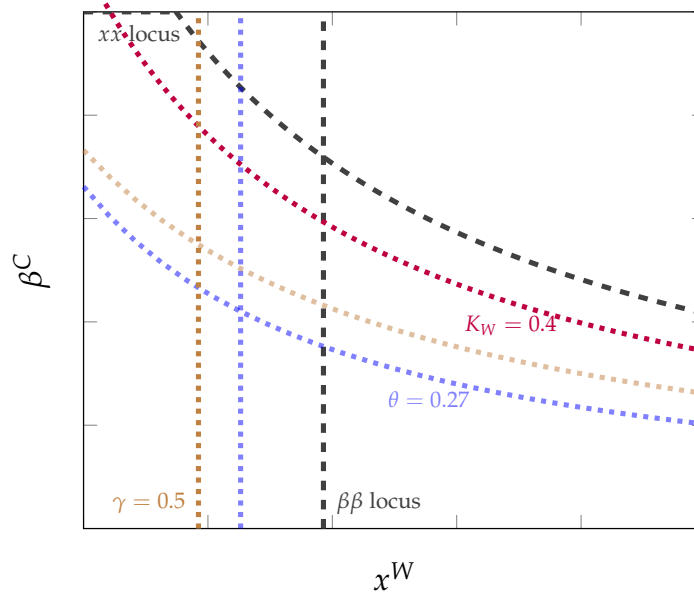
lines, or that disproportionately benefit more modern sectors, such as steam engines and computers) induce firms to adopt complementary technologies. Although the decrease in the cost of having a work ethic affects only the parental socialization choice directly, as the complementary technology is more attractive when the proportion of workers with a strong work ethic is higher, it also indirectly affects firm choice. Analogously, while changes in the relative productivity of the two technologies directly affect only the firms decision, workers with a work ethic prefer to work for firms with complementary technologies, hence they indirectly affect their choice.

Proposition 1.4 is illustrated in Figure 1.5 in which I consecutively apply changes to the parameters of Figure 1.4 (in gray). First, a decrease in the cost of having work ethic K_W has the effect of lowering the xx-locus, increasing the basin of attraction of the hard-working equilibrium. Then, an increase in productivity θ has the effect of moving the xx-locus down and shifting the $\beta\beta$-locus to the left, increasing the area of dissemination of both work ethic and firms with complementarity that leads to the equilibrium with full dissemination of both. Finally, lowering the relative productivity of the separable sector γ moves the $\beta\beta$-locus left, making the hard-working equilibrium more likely.[29] While these changes are purely illustrative, I believe that they accurately represent the transformations that occurred during the 18th and 19th centuries in industrialized economies, leading to the dissemination of cultu-

---

[29] A reduction of γ also has an effect of moving the xx-locus *up*, as it is a analogous to a decrease in ρ for the separable technology.

ral norms of work values and the diffusion of production processes focused on complementarity, which one currently observes.

These comparative dynamics explain the impact of the Protestant Reformation, which disseminated an ascetic view and stressed the importance of hard work for (religious) salvation, lowering the cost of effort and raising the cost of leisure. Consequently, it enlarged the basin of attraction of the equilibrium with hard-working agents, making the dissemination of a strong work ethic more likely. While it only affected the cultural transmission process directly, by allowing the dissemination of a work ethic, the Protestant Reformation also permitted firms to adopt more advanced technologies, using the new hard-working population in more complementary production processes. Analogously, the Industrial Revolution, which favored modern production processes, induced the adoption of new technologies, which by making effort more complementary, also made a work ethic more desirable, facilitating its dissemination. These two changes, therefore, had mutually reinforcing effects, creating a new equilibrium with a broadly disseminated work ethic and complementary technologies.

1.7
Conclusion

Cultural norms of work effort have been studied in economics, but the focus so far has been mostly on the interaction between work ethic and political redistribution. In this paper, I presented a model of work ethic dissemination and firm incentive choice. I showed results that are consistent with this model in three separate empirical exercises.

While my model is not the only relevant one for understanding the cultural determinants of work behavior, it provides new insights and different policy implications from those that have been established in the literature thus far. For example, if one considers multiple equilibria in work ethic to be a purely political issue, then an exogenous change in the tax rate would change the dissemination of work ethic in the society and move it from one equilibrium to the other. I argue in this paper that firms' incentive steepness is related to work ethic; hence, the previous assertion does not have to be true. There might be societies with the same level of redistribution and different levels of work ethic because of differences in (endogenously chosen) firm incentives.

I further show that if I allow firms to choose their technology and I interpret the complementary technology as more modern, then my model provides an explanation for why firms in third-world economies do not adopt more modern practices from advanced economies: given the mean level of effort

in the economy, technologies in which the production of one worker is highly dependent of the production of peers (e.g., when many workers manage a ring spinning textile machine) are less profitable than technologies in which worker production is independent (as in the putting-out system). The same argument can be made for why technological and productive advancement flourished in 18th century northern Europe, reaching southern Europe only at the end of the 19th century.

I explore this point more carefully by examining how the basins of attraction of the different equilibrium levels of work ethic change with changes in economic parameters. I find that an increase in human capital and a decrease in the cost of a work ethic, for example, from religious support for an ascetic lifestyle, facilitate the expansion of both a strong work ethic and the modern sector. I believe that these comparative dynamics provide insights into the impact of higher literacy and the Protestant work ethic on the Industrial Revolution and on overall economic development in 18th century northern Europe.

I hope that future empirical research will further study the correlation between work ethic and firm compensation schemes. It would be particularly interesting to analyze causal effects between these variables, searching for quasi-exogenous variations in both culture and incentives and examining their long-term effects.

2
Neighbors' Income and Demand for Redistribution

"The problem is that sometimes when I'm 'riching' around town, I'll accidentally catch glimpse of a non-rich person. Then I get a funny feeling in my heart. Apparently I have an allergy to non-wealthy people called... empathy". Stephen Colbert in The Colbert Report (July 28, 2014).

2.1
Introduction

A major question in Political Economy is what determines demand for redistribution in modern democracies, and why it differs so markedly between societies, e.g. Europe and United States. In the recent decades a large literature has risen, mainly focused on explaining these differences in terms of different perceived or real views on returns to effort and social mobility.

If perceptions of social mobility and the distribution of income are important in determining demand for redistribution, then a person's environment (and in particular the economic condition of his neighbors) should be an important determinant for how much people want to redistribute. Unfortunately, this has not been sufficiently researched within the literature of Political Economy so far.

In this paper I try to provide a first step in answering that question by using precinct-level election data for United States 2004-2012 Presidential elections and Census block group level socioeconomic variables to analyze whether having high or low income neighbors affects votes for Democrats, which I argue is a good measure for demand for redistribution.

While my evidence is not (quasi-)experimental, I employ a variety of controls, as well as fixed effects and tract-year dummies, in order to account for possible endogeneity of housing decisions and political leaning. Our main identification hypothesis is that within an election cycle and a given tract (each tract has, on average, four precincts—our unit of analysis), housing decisions are random.

Given that my empirical strategy involves the use of very geographically fine spatial dummies and maximum likelihood-based spatial models, which are computationally cumbersome, for most of the paper I restrict attention to the state of Ohio, for which I have data for the 2004 and 2008 elections. In the Appendix I provide results for the entire sample, and I show that the results we obtain for Ohio, using our very demanding empirical strategy, are qualitatively and quantitatively similar to what we find for the entire country, using less computationally demanding (but also less convincingly identified) specifications. The entire dataset comprises of 31 states and three election years (for more details see the Appendix).

The state of Ohio is the best choice for this kind of analysis because it is not only a significant state, being the 7th most populated and 10th most densely populated in United States, but it is also a swing state and a bellwether in American elections, and therefore should be representative of the entire country. (And my results show that it indeed is.)

I find evidence that when faced with low income neighbors, rich (and in some specifications middle income) individuals increase their vote for Democrats, while the opposite is true for poor voters. Similarly, having high income neighbors is associated with more votes for Democrats by high income individuals and lower voter share in low income precincts.

These results are consistent with neighborhood income affecting demand for redistribution. In particular, they are consistent with two hypothesis in the literature. First, the result that high income individuals are more likely to vote for Democrat when their neighbors are rich is consistent with them updating their beliefs about their position in the overall income distribution by examining their position within the neighborhood distribution (and analogously for low income individuals with poor neighbors). Second, the finding that high income precincts have higher share of votes for Democrats when neighboring precincts are poor is consistent with their residents updating their beliefs about the social mobility in the society, and therefore wishing for a higher social insurance against perceived (or real) higher chances of downward mobility. Analogously, low income individuals vote more for Republicans when their neighboring precincts are rich potentially because they perceive a higher chance of upward mobility, and therefore desire less transfers.

Our paper contributes to an already large literature on redistributional choice. The major early contribution comes from (34), who argue that the demand for redistribution in a democracy is proportional to the difference between the median voter's and the mean society income. Trying to explain why societies redistribute less than earlier theories would predict, (35) forma-

lize the "prospect of upward mobility"(POUM) hypothesis, where they show that if the social mobility process is concave, then it is possible for a majority to have future expected utility above the mean, even though the median is below the mean in every point in time. A different argument is presented by (41), who argue that when relative position in the society matters (as for example when individuals care about non-market interactions), redistribution might act to hide information about personal value. That will be more beneficial to the middle class (the pivotal voters in their model) the farther they are from the poor class and closer they are to the rich.

Much empirical evidence has been presented on this topic, most of it using survey responses. (42) try to discern between what they term the "homo economicus effect", that is that people vote for redistribution taking into account the effect in their (current or expected) income, the "public values effect", in which people have different beliefs about the returns to effort in the society, and the "social rivalry effect", in which individuals care about their relative standing in society.[1] We believe we provide a good setting in which to further explore these issues, by investigating how voters respond to their neighbors' income.

(39) also show evidence that voters perceptions of future mobility affect demand for redistribution and that voters seem to want to redistribute more when they believe the society to have more equal opportunities, and (40) complement this by showing that individuals tend to report lower happiness when inequality is high. Finally, in a recent paper, Karadja et al (2014) study the effects of informing voters their true position in the income distribution, and find that most voters (in Sweden) believe themselves to be poorer than they actually are (relative to the rest of the society). Furthermore, informing them of that causes them to become more conservative and demand less redistribution.

Another branch of the literature investigates the causes of the difference in demand for redistribution between Europe and United States.[2] (38) argues that as individuals observe their income histories and develop beliefs about the returns to effort, different income paths might generate distinct persistent beliefs. (05) contribute by observing that different beliefs of returns to effort might generate disparate demands for redistribution, which in turn might self-fulfill those beliefs. Finally, (04) model beliefs as being shaped by a (rational) cognitive dissonance, in an attempt (inter alia) of explaining why Americans

---

[1]Another evidence in this direction comes from (37), who finds evidence that individuals' happiness is related to their relative income in the community. See also (36).

[2]For a book-length study on this topic, see (43).

believe returns to effort are higher than Europeans even though there is little evidence that is true.

The rest of the paper is organized as follows. In Section 2 I present our data, followed by Section 3 where I present our empirical strategy and Section 4 where I establish the main empirical results of our paper (further specifications and robustness checks are left to the Appendix). I then discuss, in Section 5, these results and the potential pitfalls of our research, and finally I conclude.

2.2
Data

I use data from the Precinct-Level Election Data Project, by Stephen Ansolabehere and Jonathan A. Rodden. This database provides voting data at the precinct level, allowing sufficient spatial specificity to analyze the effect of neighborhood income on voting behavior. Indeed, in our data, the median precinct has 1300 people, which in highly urbanized areas comprises of often a single block, while in more suburban areas usually only a few[3]. Having data disaggregated to groups of few blocks makes our empirical strategy particularly plausible, as we are able to compare, within a neighborhood, voters who live next to rich/poor individuals with those who do not.

I focus on presidential elections as they more closely embody redistributional motives, as well as to make policy issues national as opposed to localized. Since I control for policy differences (and other unobservables) at the tract level, only policy focused on different precincts within a tract could be damaging to our analysis, which is particularly unplausible in a presidential election. As a robustness check, I also present, in Table 2.4, results for elections to the United States House of Representatives. The results are qualitatively similar.

The database covers the 2004 and 2008 Presidential Elections for the state of Ohio. I chose Ohio because it is an important electoral state, being considered both a swing state and a bellwether in national elections. Furthermore, it is the biggest state (in number of precincts) for which we have this data.[4] I merge the voting data with American Community Surveys (5-years) 2005-09 for 2008 voting data and Census 2000 for the 2004 election, in order to gather income data and other socio-economic covariates at the precinct-level (using block group Census data).

My data is illustrated in Figure 2.1 below, where blue areas vote predominantly for Democrats, red areas for Republicans and in yellow areas both

---

[3]See descriptive statistics on Table 3.1.

[4]We had to restrict attention to a single state for computing feasibility. In the Appendix I show some results for all states for which the data exists, finding consistent results.

parties receive similar amount of votes. Consistent with being a swing state, we see most precincts are in yellow. Areas in white are the ones for which data is missing for some of the variables I use, and I drop them from the analysis (they account for 8% of total number of Ohio precincts, and are mostly unhabited areas).
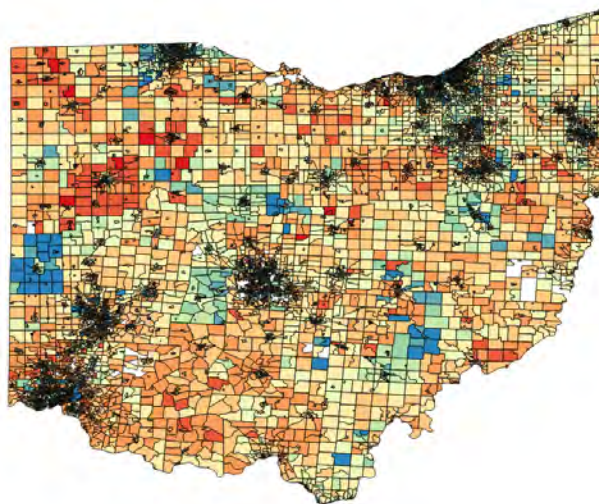
Figura 2.1: Votes for Democrats (blue) as opposed to Republicans (red) for 2008 Presidential Election (Ohio).

In our analysis we are interested not only in voting data, income, and covariates for the precincts themselves, but also for their neighbors. So I use spatial data to calculate a weighted contiguity matrix W, that points for each precinct its neighboring ones. This matrix is weighted to be row-stochastic, and so for each variable X_t, WX_t is its average value for the precincts' neighbors.

2.3
Empirical Strategy

My main empirical strategy uses two forms of identification: to remove long-lasting precint-level characteristics (like culture and precinct amenities), I use a fixed effects model, and to address the potential auto-selection problem, I use time-varying dummies at the tract level. Therefore, my identification assumption is that inside each tract, which comprises on average of four precincts, the assignment is random. This randomization can plausibly be caused by random availability of housing. In a given neighborhood, only some houses (with a given set of characteristics) are available to rent or buy in a given time frame, and therefore individuals do not have discretion over the exact precinct they will live, only the overall neighborhood (tract).

The main purpose of this paper is to estimate the following equation, where W is the weighted contiguity matrix, v_t^D is the share of votes for Democrats in each precinct at each period t, y_t is a vector of precinct income at time t, and X_t and X_t^N are matrices of covariates (with $X^N \subset X$):

$$v_t^D = \alpha_S y_t + \alpha_N W y_t + \alpha_I y_t W y_t + \beta X_t + \beta_N W X_t^N + W v_t^D + u_t$$

The error u is divided in fixed precinct-level unobservables ξ_p, time-varying tract-level unobservable characteristics $\zeta_{c,t}$ and a time-varying precinct-level error $\varepsilon_{p,t}$, that is, $u_{p,t} = \xi_p + \zeta_{c,t} + \varepsilon_{p,t}$. While I cannot address the latter, I control for the first two sources of heterogeneity using a within regression with tract-year dummies. Therefore, our main identification assumption is that $\varepsilon_{p,t}$ is uncorrelated with the independent variables; that inside a tract the choice of precinct is essentially random. As stated, the fact that there is a limited amount of residences with given characteristics available on the housing market at any given time makes this assumption plausible. I show evidence consistent with it in Section 5.

We also have a variety of controls X_t, including race, education, and population density for own precinct and neighbors, age, proportion of houses on mortgage, employment sector, self-employed, and proportion of foreigners. The main measures for income I use in this paper are the proportion of habitants with low and high income, defined as the proportion with less than 30 thousand dollars a year and more than 100 thousand dollars a year, respectively. Table 1 below shows the descriptive statistics for the main variables in our empirical exercise.

A potential issue with this data is that because of its spatial nature, the dependent variable and errors can be spatially correlated. Indeed, in Figure 2.2 below I show that this concern is valid, namely that share of votes for Democrats is significantly spatially correlated. Consequently, I follow the spatial econometrics literature and estimate, besides linear fixed-effects regressions (with errors clustered at the county level to account for spatial autocorrelation), a spatial autoregressive panel model, including a spatial lag of the dependent variable (share of votes for Democrats) to account for unobserved factors that are spatially correlated (as in the estimation formula above). I estimate this model by Maximum Likelihood.[5]

[5]I do not use neighbors' votes for Democrats in the linear model because it is (by construction) endogenous, and it would bias the results. As I will discuss in the next section, the coefficient of spatial correlation—given our controls—is actually small, which explains why both specifications give very similar results.

Tabela 2.1: Descriptive Statistics.

	Min	1st Quartile	Median	Mean	3rd Quartile	Max	Std. Deviation
Year: 2004							
Population	117.0000	1009.3051	1289.9995	1602.3226	1792.9938	8839.0000	1020.5331
Area	0.0233	0.3558	1.4370	6.1135	7.1229	87.6120	10.0384
Share votes (D)	0.0644	0.3581	0.4620	0.4947	0.5952	1.0000	0.1851
High income	0.0000	0.1058	0.1798	0.2204	0.2923	0.9429	0.1561
Low income	0.0000	0.1569	0.2456	0.2718	0.3619	0.9118	0.1541
Avg income	9.3501	10.6392	10.8372	10.8695	11.0591	12.9531	0.3620
Black	0.0000	0.0043	0.0142	0.0981	0.0584	0.9936	0.2146
Hispanic	0.0000	0.0056	0.0094	0.0171	0.0160	0.6070	0.0334
Foreign-born	0.0000	0.0083	0.0183	0.0281	0.0374	0.2945	0.0295
Undergraduate	0.0000	0.0662	0.1122	0.1404	0.1984	0.5623	0.0969
High school	0.1020	0.5673	0.6458	0.6192	0.7014	0.8641	0.1164
Graduate	0.0000	0.0209	0.0405	0.0597	0.0805	0.5229	0.0563
Year: 2008							
Population	83.0000	971.8124	1318.5300	1723.5085	1930.6641	10096.0000	1298.2103
Area	0.0233	0.3558	1.3912	5.9101	6.7030	87.6120	9.8598
Share votes (D)	0.0978	0.3938	0.4950	0.5288	0.6230	1.0000	0.1846
High income	0.0000	0.0675	0.1342	0.1754	0.2404	1.0000	0.1483
Low income	0.0000	0.1756	0.2673	0.2943	0.3897	0.9816	0.1604
Avg income	9.1190	10.7497	10.9829	11.0002	11.2263	13.0267	0.4020
Black	0.0000	0.0000	0.0173	0.1118	0.0876	1.0000	0.2220
Hispanic	0.0000	0.0002	0.0090	0.0231	0.0264	0.5936	0.0441
Foreign-born	0.0000	0.0083	0.0217	0.0343	0.0473	0.4208	0.0376
Undergraduate	0.0000	0.0720	0.1250	0.1530	0.2163	0.6312	0.1048
High school	0.0282	0.5655	0.6637	0.6313	0.7267	0.9116	0.1359
Graduate	0.0000	0.0264	0.0530	0.0723	0.0985	0.5131	0.0647

High income is proportion of precinct population earning more than 100 thousand dollars a year, and low income the proportion earning less than 30 thousand dollars per year.
Share of votes for Democrats is $\frac{\text{votes (D)}}{\text{votes (D)}+\text{votes (R)}}$. There are 10,119 observations per year.
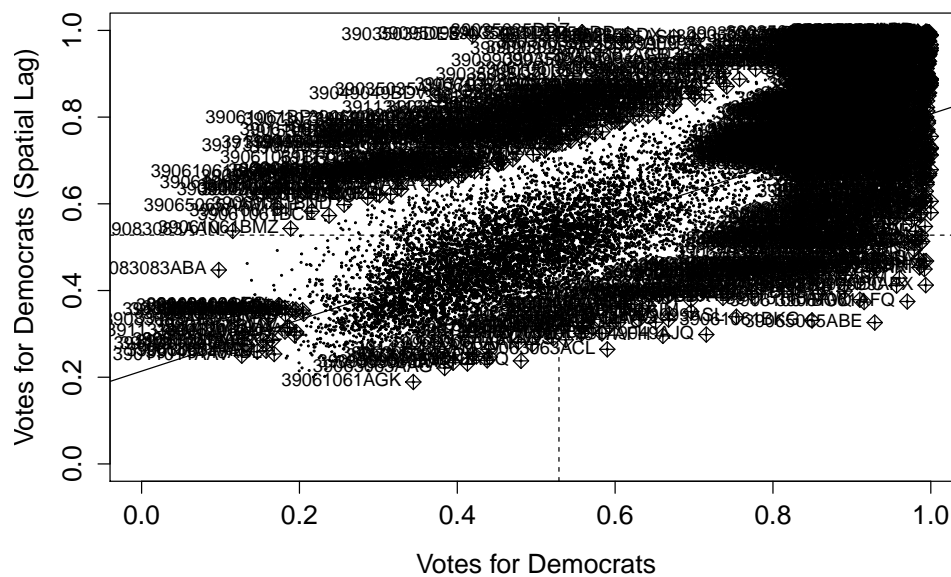
Figura 2.2: Moran graph for share of votes for Democrats (Moran's I: 0.59, p-value 0.00).

2.4
Results

The main results of the paper are presented in Tables 2.2 and 2.3 below. In Table 2.2, while I do not use a spatial regression, the errors are clustered at the county level, which is the sparsest geographical division smaller than the state, in order to account for spatial correlation of observations. Columns 1 and 2 present a naive specification without fixed effects nor tract-year dummies, and its bias indicates that there indeed exists a sorting of Democrat-leaning individuals to precincts with low income neighbors. The most important columns are 7 and 8, which include both precinct-level fixed effects and tract-year dummies, and so correspond to our empirical strategy as exposed in the previous section.

The results show that while on average neighbors' income does not determine voting, the effect is heterogeneous: for low income precincts, having low income neighbors and high income neighbors (as opposed to middle income ones) is associated with less votes for Democrats, and the opposite is true for high income precincts. This effect is statistically and economically significant: for a precinct with double the mean income of the state, having its neighbors pass from the 1st to the 3rd quartile in proportion of low income residents increases the vote for Democrats in 2 percentage points, half of a home state advantage ((44)).

However, this specification does not take into account the spatial nature

Tabela 2.2: Neighbors' Income and Votes for Democrats (Ohio, 2004-08).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High income	-0.164	-0.177	-0.139	-0.136	-0.106	-0.0862	-0.0881	-0.0505
	(0.0167)***	(0.0289)***	(0.0188)***	(0.0288)***	(0.0135)***	(0.0219)***	(0.0144)***	(0.0194)**
Low income	0.109	0.118	0.0874	0.0846	0.0884	0.0757	0.0910	0.0670
	(0.0235)***	(0.0215)***	(0.0182)***	(0.0160)***	(0.0165)***	(0.0215)***	(0.0126)***	(0.0181)***
NB High income	-0.00959	-0.0127	0.00763	0.00350	0.0135	0.0113	0.00694	0.00307
	(0.0122)	(0.0120)	(0.0129)	(0.0125)	(0.0109)	(0.0109)	(0.0138)	(0.0134)
NB Low income	0.0176	0.0168	-0.0122	-0.0133	0.00977	0.00844	-0.000926	-0.00276
	(0.0108)	(0.0109)	(0.0165)	(0.0166)	(0.0117)	(0.0116)	(0.0156)	(0.0158)
Avg. Income (norm.)		-0.0124		-0.0218		-0.0173		-0.0275
		(0.0106)		(0.00766)***		(0.00760)**		(0.00761)***
Avg. Inc. * NB high		0.0316		0.0396		0.0191		0.0279
		(0.0202)		(0.0158)**		(0.0116)*		(0.0111)**
Avg. Inc. * NB low		0.0345		0.0436		0.0265		0.0379
		(0.0165)**		(0.0129)***		(0.0122)**		(0.0110)***
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Fixed Effects	N	N	Y	Y	N	N	Y	Y
Tract-year Dummies	N	N	N	N	Y	Y	Y	Y
R^2	0.836	0.836	0.796	0.796	0.914	0.914	0.882	0.882
N	20,238	20,238	20,238	20,238	20,238	20,238	20,238	20,238

Dependent variable is the share of votes for Democrats over total votes. "High income" is the proportion of residents in the top 3 deciles of income distribution, "Low income" is the proportion of residents in the bottom 3 deciles of income distribution, and "NB" stands for the neighboring precincts' average of these variables. Errors are clustered at the county level. The symbol * denotes p-value smaller than 0.1, ** for 0.05 and *** for 0.01. Controls include sex, age, race, proportion of foreign-born, education, population density, sector of activity, and neighboring precincts' race, education, and population density.

of the data and how voter ideology might spread from a precinct to its neighbors. In Table 2.3 I use spatial panel data methods to account for this. In columns 1 and 2 I use a spatial regression without fixed effects using tract-year dummies and in columns 3 and 4 a spatial fixed effects model without these spatial controls. Given that the spatial models are solved by maximum likelihood methods, which are more demanding computationally than linear programming problems, I could not compute the model with both fixed effects and tract-year dummies.

Tabela 2.3: Neighbors' Income and Votes for Democrats (Ohio, 2004-08), spatial models.

	(1)	(2)	(3)	(4)
High income	-0.1156	-0.0964	-0.1795	-0.1873
	(0.0089)***	(0.0110)***	(0.0095)***	(0.0106)***
Low income	0.0686	0.0562	0.0697	0.0789
	(0.0074)***	(0.0086)***	(0.0080)***	(0.0090)***
NB High income	0.0142	0.0122	-0.0189	-0.0181
	(0.0141)	(0.0108)	(0.0158)	(0.0161)
NB Low income	0.0095	0.0083	0.0346	0.0338
	(0.0100)	(0.0100)	(0.0145)**	(0.0145)*
Avg. Inc (norm.)		-0.0169		-0.0179
		(0.0037)***		(0.0041)***
Avg. Inc * NB High		0.0183		0.0342
		(0.0068)***		(0.0079)***
Avg. Inc * NB Low		0.0262		0.0507
		(0.0067)***		(0.0079)***
ρ	0.0252	0.0256	0.0204	0.0157
	(0.0067)***	(0.0125)**	(0.0104)*	(0.0105)
Controls	Y	Y	Y	Y
Fixed Effects	N	N	Y	Y
Tract-year Dummies	Y	Y	N	N
N	20,238	20,238	20,238	20,238

Dependent variable is the share of votes for Democrats over total votes. ρ is the spatial correlation coefficient of votes for Democrats. Errors are assumed spatially and serially correlated, following (46). The symbol * denotes p-value smaller than 0.1, ** for 0.05 and *** for 0.01. Controls include sex, age, race, proportion of foreign-born, education, population density, sector of activity, and neighboring precincts' race, education, and population density. Estimated using (45).

Interestingly, the results are qualitatively and quantitatively similar to those in Table 2.2, which indicates that the spatial nature of the data does not affect our results significantly when controlling for tract-year dummies.

Indeed, the spatial correlation coefficient ρ is close to zero (0.02) in columns 1 and 2, although statistically significant, and in the spatial fixed effects model we cannot even statistically claim there is spatial interdependence (controlling for the fixed effect and covariates). This result gives us confidence that the other results in this paper without the spatial models suffer from very little bias.

So far the dependent variable analyzed has been the proportion of votes for Democrats over the total of votes in each precinct. It is not clear, however, if the effect is caused by changes in turnover by Democrats, by Republicans, or if voters changed their choice. Moreover, it is interesting to investigate whether the effect we find is also present in elections for other positions besides President. In Table 2.4 I employ the same empirical strategy as in my main specification, but using as dependent variable the proportion of votes over the precinct population. In columns 3 and 4 I do a similar exercise, now analyzing the effect of neighbors' income on votes for the US House of Representatives.

Table 2.4 indicates that most of the effect comes from changes in Democratic turnover; while the Republican regressions' coefficients have the correct sign, they are small in magnitude and not statistically significant (except for share of low income individuals, which causes smaller Republican turnout). Furthermore, the results for the US House of Representatives in columns 3 and 4 are similar to the results from US Presidential Elections (columns 2 and 3), which is consistent with our interpretation, as many of the decisions regarding income redistribution are taken in the Legislative.

2.5
Discussion

In the previous section I have shown evidence that high income precincts vote more for Democrats the higher the proportion of high income and low income neighbors, and the opposite is true for low income precincts. This evidence is consistent with two theories of the effect of neighbors on demand for redistribution. First, if individuals consider themselves poorer than they actually are when they have daily contact with others richer than themselves (and *vice-versa*), then we should see a pattern that high income individuals vote for more redistribution (i.e., for Democrats) when neighboring precincts have higher proportion of high income people, and vote for less redistribution (Republicans) when neighboring precincts have a higher share of low income and middle income individuals. Analogously, poor precincts would vote for less redistribution when their neighbors are also poor (as some are presumably poorer), and desire less redistribution when their neighbors are richer.

Tabela 2.4: Neighbors' Income and Votes for Democrats (Ohio, 2004-08), per party and US House of Representatives.

	(1)	(2)	(3)	(4)
Dependent variable:	*President (D)*	*President (R)*	*US House (D)*	*US House (R)*
High income	-0.0318	-0.0275	-0.0476	-0.0172
	(0.0337)	(0.0183)	(0.0317)	(0.0179)
Low income	-0.00119	-0.0340**	0.00137	-0.0411***
	(0.0213)	(0.0141)	(0.0215)	(0.0137)
NB high income	0.0162	0.0278	0.0121	0.0355
	(0.0313)	(0.0275)	(0.0268)	(0.0264)
NB low income	0.00294	0.00148	-0.00100	0.0101
	(0.0223)	(0.0208)	(0.0211)	(0.0217)
Avg. Income (norm.)	-0.0354***	0.00330	-0.0316***	0.00549
	(0.0101)	(0.00677)	(0.00857)	(0.00749)
Avg. Inc. * NB high	0.0408**	-0.00806	0.0351*	-0.0120
	(0.0200)	(0.00970)	(0.0202)	(0.00964)
Avg. Inc. * NB low	0.0573***	-0.00907	0.0524***	-0.0113
	(0.0200)	(0.0107)	(0.0197)	(0.0108)
R^2	0.685	0.601	0.698	0.668
N	20,238	20,238	20,238	20,238

Dependent variable is the share of votes over the precinct population, in columns (1) and (2) for US Presidential Elections (Democrat and Republican, respectively), and in columns (3) and (4) US House of Representatives. The symbol * denotes p-value smaller than 0.1, ** for 0.05 and *** for 0.01. Controls include sex, age, race, proportion of foreign-born, education, population density, sector of activity, and neighboring precincts' race, education, and population density. All regressions include tract-year dummies and precinct fixed effects.

Second, if we assume that voters consider not only the expected value of redistribution, but also social insurance, and if redistribution choices are long-lasting, then the optimal level of redistribution for them depends on their beliefs about social mobility. If the rich believe that there is very little chance of becoming poor, they want less social insurance, as do the poor if they believe chances of upward mobility are high. In that scenario, we would observe that when poor individuals observe higher income neighbors, they update their beliefs about upward mobility, and vote for less redistribution. Similarly, when the rich observe poorer neighbors, they consider that their chance of moving down the income ladder is higher, and therefore demand more redistribution. Together, these two reasonable hypothesis about voter behavior imply empirical patterns similar to what we observe in the data.

Another, related, theory that is consistent with our results is that voters

might make inferences about social mobility by observing the size of the tails of income distribution in their neighborhood. Observing more high income and low income residents as opposed to middle income ones updates their prior about how unequal the society is. As in the theory above, this has differential effect depending on the voter's income. A poor voter might perceive increased inequality and social mobility as opportunity for upward mobility, and so request less redistribution. On the other hand, high income voters see their wealth as less persistent, and demand more social insurance.

One plausible concern for this paper's empirical strategy is identification. Although I employ precinct fixed effects and tract-year dummies (each tract has on average four precincts), it is possible that individuals sort within tract and within an election cycle in such a way that more liberal high income individuals prefer to live closer to low income and high income individuals, while Republican-leaning low income voters attempt to move closer to high income and low income precincts. Even though it is not clear why such a specific sorting pattern should occur, it could still be responsible for our results (or parts of it).

Indeed, our identification hypothesis is that within a tract, housing location is random. Namely, in a given time period, only a small subsets of houses with certain characteristics (number of rooms, size, price range) are available, and therefore the buyer cannot choose exactly which block or precinct he prefers to live, but instead can only decide on a broader geographical area (the tract, or neighborhood).

If this is true, then any sorting (that occurs between the election cycle, so that it is not captured by the fixed effects) should be captured by our tract-year dummies, and it would not bias our results. (In terms of the estimation equation, $\mathbb{E}\left[\varepsilon_{t,p} y_t\right] = 0$.) If this assumption is false, however, then the results are potentially biased. I try to investigate this possibility in Table 2.5 below.

Table 2.5 analyzes whether neighbors' characteristics have impact in property desirability, which reflects on its value. Specifically, if potential buyers prefer to buy houses near high income precincts—even within the tract—, then those locations should have a higher cost than comparable houses in nearby precincts. The data, however, shows little evidence of that. Although the coefficient for neighbor's proportion of high income households is marginally significant, the coefficient is small: a change from the 1st to 3rd quartile in neighboring precincts' share of high income households is related to a 0.9 percent increase in house value. All other coefficients are economically and statistically insignificant.

Tabela 2.5: Neighbors' Income and House Values (Ohio, 2004-08).

	Log house value
High income	0.317
	(0.0443)***
Low income	-0.0507
	(0.0602)
NB high income	0.0479
	(0.0265)*
NB low income	0.0100
	(0.0382)
Avg. Income (norm.)	0.168
	(0.0149)***
Avg. Inc * NB high	-0.0170
	(0.0395)
Avg. Inc. * NB low	-0.00856
	(0.0324)
Black	-0.120
	(0.0513)**
Hispanic	-0.206
	(0.142)
Bachelor degree	0.633
	(0.0491)***
Graduate school	0.615
	(0.0957)***
R^2	0.919
N	19,914

Dependent variable is the log of precinct median house value. The symbol * denotes p-value smaller than 0.1, ** for 0.05 and *** for 0.01. Controls include sex, age, race, proportion of foreign-born, education, population density, sector of activity, and neighboring precincts' race, education, and population density.

Nonetheless, even though our evidence does not point to sorting as the driver of our results, it must be noted that we cannot rule out that this effect exists and that it is the source of the stylized facts I present. The fact that our evidence is not quasi-experimental, however, does not make our results uninteresting, as our analysis provides an early "big picture"view of an important question. We welcome future evidence that would provide a causal effect that substantiates or questions our results, although I remark that if the main issue is sorting and if it is believed that sorting happens quickly and

sharply enough to drive our results, then it poses a very difficult problem even for quasi-experimental data.

Another potential issue with our empirical analysis is that it might include a social multiplier effect ((47)); the effects I capture might be intensified because of multiplying effects of neighbors inside the precinct. Specifically, if voters tend to adopt ideologies close to those their neighbors adopt, the effect of neighboring precincts' income is intensified by the effect of each voter's ideology change on the other voters within the precinct.

This is related to the issue of ecological regression, namely that it is difficult to capture individual behavior using aggregate data. Although we can analyze how a precinct's votes change in relation to the income of neighboring precincts, it is impossible to tell from which individuals this effect is originated. Even when saying that high income precincts respond differently than poor income ones, it cannot be ruled out the—perhaps unlikely—possibility that all the effect comes from poor individuals, and the heterogenous effect depends on where they live. These drawbacks of our analysis are the unavoidable consequence of trying to investigate voter behavior using hard data (as opposed to self-reported values), given the secret nature of elections.

2.6
Conclusion

In this paper I presented new empirical evidence that points to the importance of a voter's environment in determining his demand for redistribution. I used precinct-level election data to analyze very localized voting patterns, and I found that having rich neighbors is correlated with high income individuals voting for Democrats and low income individuals for Republicans. Similarly, having poor neighbors is correlated with high income individuals being left-leaning, while low income individuals being right-leaning.

Our results are consistent with findings in the literature (vide Alesina et al. (2004)) that in United States rich people's happiness is particularly harmed by inequality. Our theory is that inequality signals that social mobility is higher, which has negative implications in terms of expected income and its variance that are particularly detrimental to richer individuals.

The finding that voter behavior responds to neighbors' income is important, and it has significant implications to urban design and spatial inequality. Indeed, our results suggest that an increase in income segregation might cause voters to become more politically polarized, possibly because they wrongly underestimate the possibility of moving up or down in the income distribution.

Interesting future research would be to try to examine how this behavior

generalizes to other societies and cultures. Alesina et al. (2004) find that in Europe the effect of inequality on happiness is stronger on the poor. Would this result also be valid in our setting, namely is the effect of having low income neighbors stronger for poorer precincts than for richer precincts in Europe? Evidence to this effect would be an interesting complement to our research. I also hope that in the future new quasi-experimental research will be able to add to our results.

3
Party Registration and Mayoral Elections

3.1
Introduction

An essential factor in promoting growth in developing countries is increasing the productivity of the public sector. For that purpose, it is important that the public bureaucracy is formed by qualified workers, and not merely used as a means of inefficient targeted redistribution by politicians towards their supporters. Unfortunately, political parties often use their power to appoint the public bureaucracy to distribute jobs among their supporters as patronage, in order to gather assistance in elections and offer advantages to them. These advantages can be significant, especially in countries where there are large rents in working for the government, like Brazil.

In an environment with patronage, we would find that supporters of a political party would receive large economic benefits, in the form of offers of work in the public sector and higher wages from government, when that party assumes office. This should be especially true for formal supporters of the party, in particular registered members. Additionally, since it is common for parties to demand contributions from affiliates, giving them public jobs is an indirect way of using public funds to finance their campaigns.

In this paper we investigate the presence of patronage in municipalities in the Northeast of Brazil, after the 2008 mayoral elections. We empirically estimate the causal effect of a party winning a mayoral election on the proportion of its affiliates working for the government in the following year, as well as their wage accrued from government. If patronage is big, then a party's victory would have significant effect on economic outcomes of its supporters, in particular the registered members of these parties. We also investigate whether a party winning a municipal election affects the number of affiliations to that party in the upcoming years, as a way for individuals to signal their support for that party, possibly obtaining rents from this political connection.

By analyzing this second issue, our research is also related to the question of why voters register to political parties. One possible answer to this question is that voters register to political parties in order to signal their support for

that party, hoping to receive economic benefits in return. In actuality, as we argue in Section 3, most political affiliations in Brazilian municipalities are motivated by plans of candidacy.

Our data comprises of all the registered members of political parties for all years from 1994 to 2014, and socioeconomic data for all formal employees of firms or government, for the years 2008 and 2009, in the Northeast of Brazil. From the merged dataset we gather the affiliated member's wage accrued from government and the sector of employment. The Brazilian Northeast is a particularly suitable region to analyze because anecdotal evidence points to it as the most corrupt and clientelistic region in the country.

In our empirical strategy, we obtain plausibly causal estimates of the effect of a party's victory on the probability of a registered member working in the government and his wage accrued from the public sector by employing a Regression in Discontinuity (RD) design, comparing elections where the margin of victory or defeat was small. If city and party characteristics are continuous in the proportion of votes received by the party, and if candidates and voters cannot precisely control the amount of votes that a party receives, then the RD design can identify the (causal) treatment effect around the cutoff ((48)).

In both empirical exercises, we do not find consistent evidence that being affiliated to a political party has any positive economic value to individuals (patronage). Members of a political party do not have any higher chance of working for the government after the election, nor do public employees receive any higher wage when a politician of their party or coalition wins municipal elections. Moreover, our results indicate that voters do not register to political parties seeking to show support; a party that narrowly wins an election is not any more likely to receive registrations than a party that narrowly loses it.

Our paper contributes to the debate on patronage, and more broadly to the literature on inefficiencies deriving from targeted redistribution, emphasized by (53) and, more recently, (54). Since in our setting the possible objects of patronage (political affiliates) account for almost 5 percent of the total population, the existence of targeted redistribution would cause these inefficiencies to be severe.

In that literature, (55) analyze the reverse of our research question, asking whether control of patronage increases a party's chance of winning elections, using the implementation of civil service reform in United States. More recently, (56) provides evidence on the existence of patronage by showing that the number of employees is higher for elected chief executives than for bureaucrats, and that this difference is higher in election years and where there is a larger public-private wage differential.

(57) present a theoretical model (based in (58)) and empirical evidence that in American cities, public employment is used as a way of redistributing resources. They argue that when quality of projects is unobservable, politicians can hide redistribution from voters as public projects that would be otherwise undesirable. A different argument is made in (59), who maintain that politicians use public jobs as a redistribution device in order to solve a political commitment problem, by tying the voter's continuation value to the politician's.

Also related is a recent and growing literature investigating the effect of political connections in developing countries. Fafchamps & Labonne (2015) investigate whether politicians' relatives move into higher-paying occupations, and find a positive effect. Querubin (2016) studies political dynasties in Philippines, and finds a causal effect of having a relative in office in the chance of running for office in the future, which he explains by access to public funds for clientelistic policies. Political connections in the particular context of low income areas and poverty alleviation programs are studied by (49), (50), and (51).

In the next section we explain the database we use in our empirical analysis, which contains both the universe of registered party members since 1994 and the universe of formal employees. Then, in Section 3 we provide numerous descriptive statistics on party affiliation in Brazil, followed, in Section 4, by a discussion of our empirical strategy of using a RD design to identify the causal effect of being registered to a winning party. In Section 5 we present the main results of this paper, and then we conclude.

3.2
Data

In this paper, we use a public database of registered members of political parties in Brazil, provided by the Tribunal Superior Eleitoral (Superior Electoral Court; TSE), which lists by name all registered members of political parties and the date of registration. To gather socioeconomic data for these individuals, we match this information with private individual data from Relação Anual de Informações Sociais (RAIS) of the Ministério do Trabalho e Emprego (Ministry of Labor). For this version of the paper, we use data for Northeast region (states of Alagoas, Bahia, Ceará, Maranhão, Paraíba, Piauí, Pernambuco, Rio Grande do Norte, and Sergipe), for the years 2008 and 2009. For that reason, we restrict attention to the 2008 mayoral elections.

It is important to note that although the RAIS database is not a sample, it does not include information on all Brazilian citizens. As a database from

the Ministry of Labor, it has information on all Brazilian formal employees; roughly half of the Brazilian workforce (which itself is around half of the total working age population). Therefore, if this distribution is the same for registered members of political parties as for the general population, we would expect that one forth of the registered members would be in our RAIS database. Nonetheless, while this might have implications for external validity of our findings, as this selection happens equally for treated and control groups, it does not affect the internal validity of the empirical analysis. Moreover, as we focus our attention on government employees, all of them are included in our sample.

Since the TSE database only identifies registered members by name, we merge the databases using approximate string matching, to account for possible misspellings of names. Specifically, within each city we select the closest name in RAIS database to each registered voter according to Levenshtein distance, up to a distance of 2. Levenshtein distance measures the minimum number of single-character edits (insertions, deletions or substitutions) required to change one word into the other.

Finally, in order to select our treated and control municipalities, we gather election results for the 2008 Brazilian mayoral elections from TSE. We restrict tha database to only disputes in the first round (in Brazil, only cities with more than 200,000 registered voters can have two rounds of elections), and consider only the top two candidates in votes, defining a candidate's vote share as the share of votes from the top 2 candidates that she received. We remove ties.

We show summary statistics of our sample in Table 3.1. For convenience, we present separate statistics for the control and treatment groups (in our main specification). An individual belongs to treatment group if it is affiliated to a member of the presidential coalition in a city where a party from that coalition won. Analogously, an individual is part of the control group if it is affiliated to a member of the presidential coalition in a city where a member of that coalition lost. As it is expected given that we use the the entire sample (and not only around the cutoff), individuals affiliated to the winning party have higher wage, schooling, are more likely to work in the government and earn higher income when working there.

We also construct summary statistics for the database aggregated by party and city, which is the database we use to analyze registration effects (and it is also informative about our main empirical exercise). We present it in Table 3.2. Again, parties that won the election have more affiliated members before and after the elections, which is to be expected considering that we

Tabela 3.1: Summary statistics (full sample), separated by treatment and control (registered to a member of presidential coalition, where it won—resp. lost—the election).

Affiliated to losing party (any vote share)

	N	Mean	SD	Median	Min	Max
Affiliation date	29031	2001.63	4.84	2003.00	1989	2007.00
Wages from government (2009)	29031	668.23	1360.39	462.47	0	67022.40
Male	25956	0.59	0.49	1.00	0	1.00
Schooling	25956	6.23	2.01	7.00	1	11.00
Age	25956	39.37	10.50	38.00	11	98.00
Nominal average wage (2009)	29031	775.22	1010.07	558.00	0	44319.28
Works in government (2009)	29031	0.57	0.49	1.00	0	1.00
Income from city government (2009)	29031	497.77	907.00	460.83	0	37475.10

Affiliated to winning party (any vote share)

	N	Mean	SD	Median	Min	Max
Affiliation date	32444	2001.94	4.91	2003.00	1989	2007.00
Wages from government (2009)	32444	800.95	1510.96	493.30	0	42858.33
Male	29365	0.59	0.49	1.00	0	1.00
Schooling	29365	6.27	2.02	7.00	1	11.00
Age	29365	39.49	10.70	39.00	12	108.00
Nominal average wage (2009)	32444	862.42	1251.31	580.70	0	51816.30
Works in government (2009)	32444	0.63	0.48	1.00	0	1.00
Income from city government (2009)	32444	635.51	1229.11	465.13	0	42858.33

calculate the summary statistics using the entire database.

3.3
Party Affiliation in Brazil

In this section, we present numerous stylized facts about party affiliation in Brazil, to better comprehend how and why party members differ from the rest of population, and how this might affect the political process. Specifically, we compare socioeconomic characteristics between party affiliates (who are formal employees) and the rest of Brazilian formal employees, we provide the trends of party affiliation in Brazil, and we analyze how party members that are plausibly planning on running for elections differ from the other party affiliates.

At the time of 2008 Municipal Elections, there were 10,898,236 party affiliates in Brazil. (There were 191.8 million habitants in the country). Figure 3.1 displays the number of affiliates of each of the largest 9 parties (in 2012), at 2000, 2004, 2008, and 2012 Municipal Elections. Party affiliation has increased markedly with the deepening of Brazilian democracy, especially in more progressive parties, like PT and PSB, both of which more than tripled their number of affiliates in this period. (This might also be related to PT holding the Brazilian Presidency for most of that period.)

Tabela 3.2: Summary statistics (full sample, aggregated by party-city), separated by treatment and control.

Losing party (any vote share)

	N	Mean	SD	Median	Min	Max
Affiliated before election	1536	144.16	162.40	103	3	2681
Affiliated post-election	1536	25.56	49.89	14	0	1334
Affiliated at victory year	1536	2.62	13.99	0	0	233
Affiliated at mayor's term	1536	24.47	48.57	13	0	1329

Winning party (any vote share)

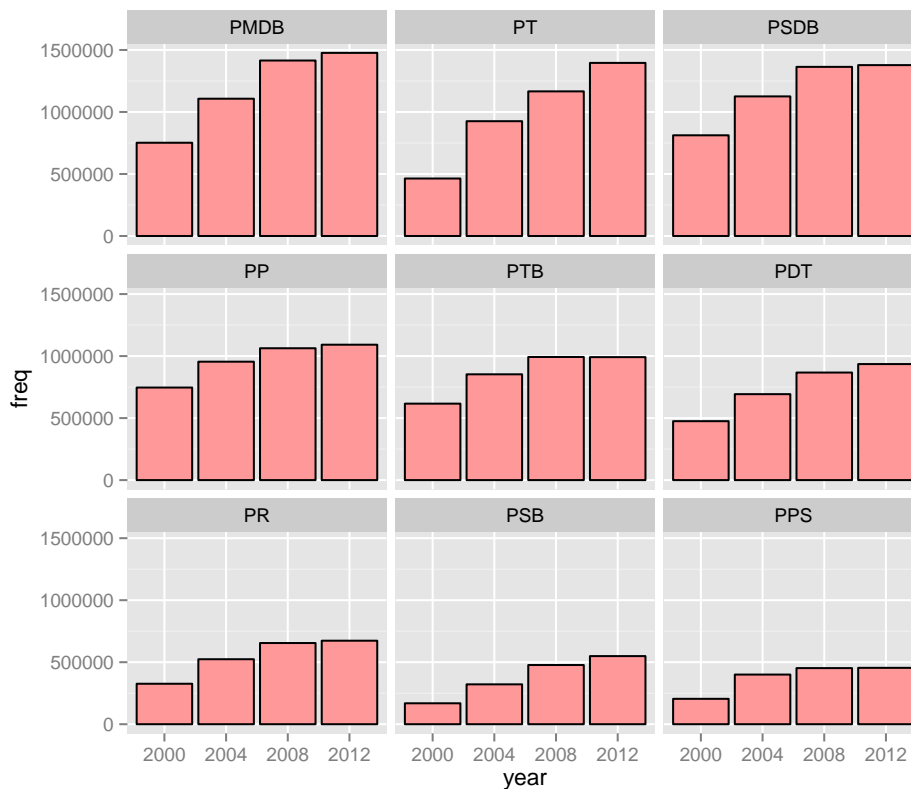|	N	Mean	SD	Median	Min	Max
Affiliated before election	1531	155.23	157.30	113	4	2367
Affiliated post-election	1531	34.61	81.58	17	0	1650
Affiliated at victory year	1531	5.59	54.57	0	0	1542
Affiliated at mayor's term	1531	33.51	80.37	16	0	1650

Figura 3.1: Party affiliates by party and election year (9 largest parties in 2012).

Notwithstanding the growth of parties in the presidential (left-leaning) coalition during the 2000s in Brazil, Figure 3.2 indicates that a large part of the affiliation decision is motivated by local politics; plausibly by plans of candidacy. Indeed, party registrations spike strikingly the year before municipal elections. In particular, registrations for parties in the presidential coalition the year before 2008 Municipal Elections were 5 times higher than the year before that (and 2 times higher than the year after it). This fact indicates that citizens register in political parties as a way to support (or run in) a municipal election—mayoral or legislative.
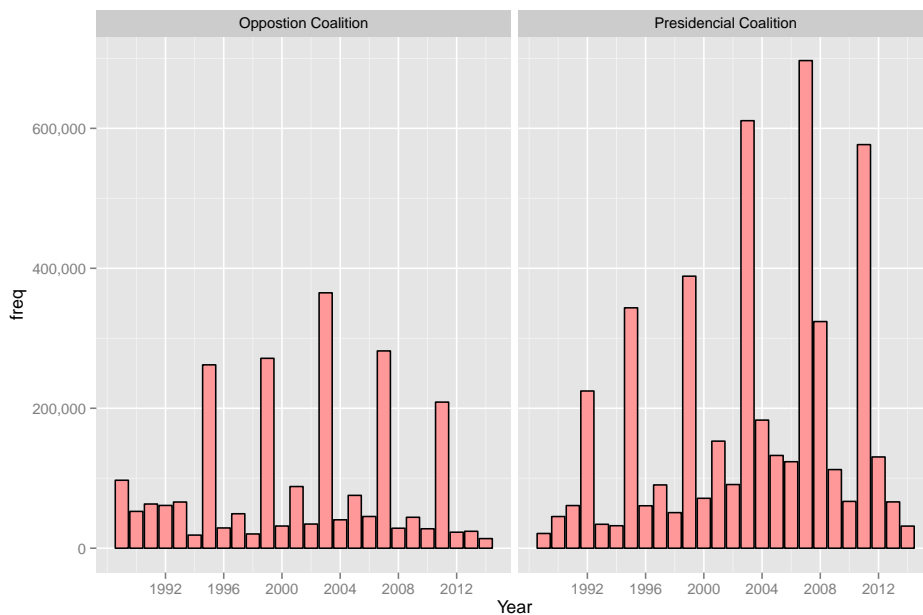
Figura 3.2: Party registrations for parties in presidential and opposition coalitions, per year.

Even within the year before the elections, registrations tend to concentrate on the last weeks before the registration deadline for running in the next year elections. Since this deadline should be irrelevant for those not seeking candidacy, this indicates that a large proportion of a party's affiliations are by people planning on running for (mostly municipal) office. Figure 3.3 documents that fact.[1] In Table 3.3 we compare socioeconomic characteristics between affiliates before and after the registration limit. We argue that this accounts for differences between registered members that intend to run in municipal elections, and those that do not.

---

[1]The plot uses monthly data for clarity, but we have daily data on registrations, and it confirms our hypothesis. (Namely, October registrations are almost entirely concentrated in the days before the candidacy deadline.)
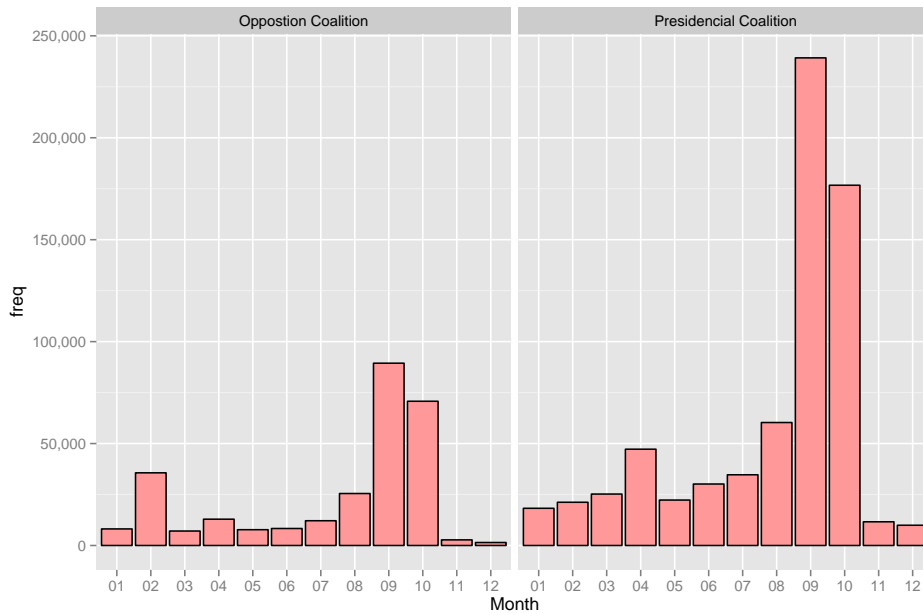
Figura 3.3: Party registrations for parties in presidential and opposition coalitions, monthly (2008).

We find that individuals that are arguably going to run for election in 2008 are more likely to work for the government (as opposed to private firms), are more likely to be male and older, and they have less schooling and total income. This points to the pool of political candidates being worse (in schooling and arguably ability) than the average party supporter. Understanding why individuals intending to be political candidates are more likely to be working for the government deserves further study.

Another interesting question, which our database helps answering, is how political affiliates, which form the pool from where candidates (and consequently office holders) are chosen from, differ from the average member of the population. (In our case, we can only compare the affiliated formal employees with non-affiliated formal employees.) We do so in Table 3.4.

Table 3.4 confirms usual intuitions about party affiliates: on average, they earn significantly more income from government (and especially from the municipality) and less income from firms and social entities. They are more likely to be white and women, and they are older than the average formal employee. Possibly as a consequence of being more likely to work for the public sector, they work in bigger firms and have much longer tenure.

In Columns 7 and 8, we also compare socioeconomic characteristics between individuals affiliated to parties in the presidential coalition with those affiliated to parties in the opposition coalition. We find that differences in work sector are much less pronounced between political parties than between

Tabela 3.3: Summary statistics (before and after registration deadline to run for 2008 elections).

	(1) Before deadline	(2) After deadline	(3) (1) - (2)	(4) p-value
Male	0.60	0.53	0.07	0.00
White	0.15	0.15	0.00	0.95
Wages from Government	483.26	332.25	151.01	0.00
Private Income	249.29	452.24	-202.95	0.00
Avg. Income	1.57	1.73	-0.16	0.00
Schooling	6.13	6.18	-0.05	0.02
Firm size	6.27	5.68	0.59	0.00
Tenure	68.41	62.46	5.95	0.00
Age	36.10	35.76	0.34	0.02
Hours	39.50	40.72	-1.22	0.00
N	55,016	8,827		

This table uses 2008 socioeconomic data for individuals affiliated to political parties in our sample in 2007. Before deadline contains individuals who affiliated less than one month before the registration deadline, and after deadline those that registered after the deadline had already passed. Average income is measured in 2008 minimum wages.

Tabela 3.4: Summary statistics (full RAIS and affiliated).

	(1) RAIS	(2) Affiliated	(3) Presidential Co.	(4) Opposition	(5) (2) - (1)	(6) p-value	(7) (4) - (3)	(8) p-value
Wages from:								
Government	511.455	680.759	743.494	735.500	169.30	0.000	-7.99	0.223
Private Firms	682.876	340.244	337.117	283.561	-342.63	0.000	-53.56	0.000
Social Entities	58.518	25.460	29.730	23.318	-33.06	0.000	-6.41	0.000
Municipality	186.299	543.462	573.969	593.729	357.16	0.000	19.76	0.000
Sex (male)	0.604	0.588	0.599	0.559	-0.02	0.000	-0.04	0.000
White	0.288	0.305	0.312	0.328	0.02	0.000	0.02	0.000
Schooling	6.074	6.071	6.180	6.107	0.00	0.402	-0.07	0.000
Age	35.584	39.398	38.919	40.718	3.81	0.000	1.80	0.000
Avg. Income	2.331	1.857	1.936	1.851	-0.47	0.000	-0.09	0.000
Firm Size	5.780	6.459	6.446	6.689	0.68	0.000	0.24	0.000
Tenure	61.917	81.780	79.760	91.945	19.86	0.000	12.19	0.000
Work hours	40.946	39.713	39.648	39.182	-1.23	0.000	-0.47	0.000

affiliated and non-affiliated. Moreover, in agreement to anecdotal evidence, right-leaning parties are composed of more men, white people, and older individuals.

3.4
Empirical Strategy

In order to obtain causal estimates of the returns to being registered for a party that wins the municipal elections, we need to be able to discern this effect from the plausible confounder that individuals registered to parties that are strong locally are different than individuals who register to parties with small local influence. We do so in this paper by exploiting a Regression in Discontinuity (RD) design, comparing individuals registered to parties that barely won the election with individuals registered to parties that barely lost.

Our main identification assumption is that these confounding forces are continuous in the amount of votes that the candidate of the party the individual is registered to receives. Therefore, if we find any discontinuous jump in the dependent variables at the cutoff in voting share necessary for victory, it must be the causal effect of being registered to a winning party, as opposed to a losing one. As usual with RD strategies, this imposes a cost in external validity: we can only identify the causal effect of being registered to a winning party in municipalities where the elections were highly competitive.

We estimate the following model, at the party (p) and city (c) level:

$$y_{pc} = \beta W_{pc} + f(V_{pc}) + \varepsilon_{pc}$$

Where y_{pc} is the mean outcome variable for individuals registered to party p at city c, W_{pc} is whether party p (or its coalition) won the mayor election at city c and V_{pc} is the voting share of party p (or its coalition) at city c. The error ε_{pc} represents other characteristics of individuals registered to party p at city c. These must not vary discontinuously as a function of V_{pc}. The function f will be estimated using parametric polynomials and non-parametric local polynomials (which differ before and after the cutoff).

We carry out multiple empirical exercises where we restrict our treatment to a single party or coalition of parties, and then compare the registered members of this party in cities where it won by a small margin with municipalities where it lost by a small margin. We do so with the presidential coalition, the opposing (presidential) coalition, and for two of the biggest parties in number of municipalities (in 2008), PMDB (Partido do Movimento Democrático Brasileiro) and PT (Partido dos Trabalhadores). We also analyze a specification where all parties are pooled, and each city participates as treatment and con-

trol (with individuals registered to winning party as treatment and to losing party as control); we do the same when analyzing future party registration.

Tabela 3.5: Balance test. Treatment (resp. control) is registered to a member of presidential coalition, where it won (resp. lost) the election.

	Treatment	Control	Difference	Std. Error	P-value	N
Vote margin: +-5%						
Male	0.5961	0.5937	0.0024	(0.0104)	0.816	23,947
Age	39.691	39.262	0.4290	(0.2882)	0.137	23,947
Race	5.850	5.597	0.2530	(0.1886)	0.181	8,938
Schooling (2008)	6.131	6.107	0.0240	(0.0797)	0.763	23,019
Affiliation year	2001.933	2001.759	0.1732	(0.2389)	0.469	26,804
Number of affiliated in city	152.96	142.15	10.81	(6.686)	0.106	1,491
Vote margin: +-10%						
Male	0.5892	0.5960	-0.0067	(0.0091)	0.459	37,886
Age	39.574	39.329	0.2449	(0.2461)	0.320	37,886
Race	5.878	5.779	0.0992	(0.2187)	0.650	15,157
Schooling (2008)	6.136	6.126	0.0112	(0.0677)	0.868	36,643
Affiliation year	2001.985	2001.667	0.3172	(0.1945)	0.103	42,240
Number of affiliated in city	153.06	146.87	6.19	(6.409)	0.334	2,300

In Table 3.5, we check for manipulation in our RD design by investigating whether non-treated covariates are different between treated and control groups. The only way that these groups would vary in relation to gender, age, race, schooling (before treatment), or affiliation year would be if individuals were aware beforehand of the election results—even for close elections—, or could in some way manipulate these results (precisely). We find no evidence of this kind of manipulation.

Finally, in Figure 3.4 we analyze the density function of the share of votes, to see if there is any discontinuity around the cutoff, which could be indicative of manipulation. In this case, the discontinuity in density is statistically significant (discontinuity coefficient 0.174, standard error 0.067). If we compare the average number of affiliated members of the party in the city between treated and control municipalities (last row of Table 3.5), it shows no statistically significant difference. Therefore, parties in the government coalition are slightly more likely to win in close elections.

3.5
Results

In this section we present the results for the two empirical exercises explained above. First, we investigate whether being registered to the party of the mayor (or to a party in the same presidential coalition) leads to an
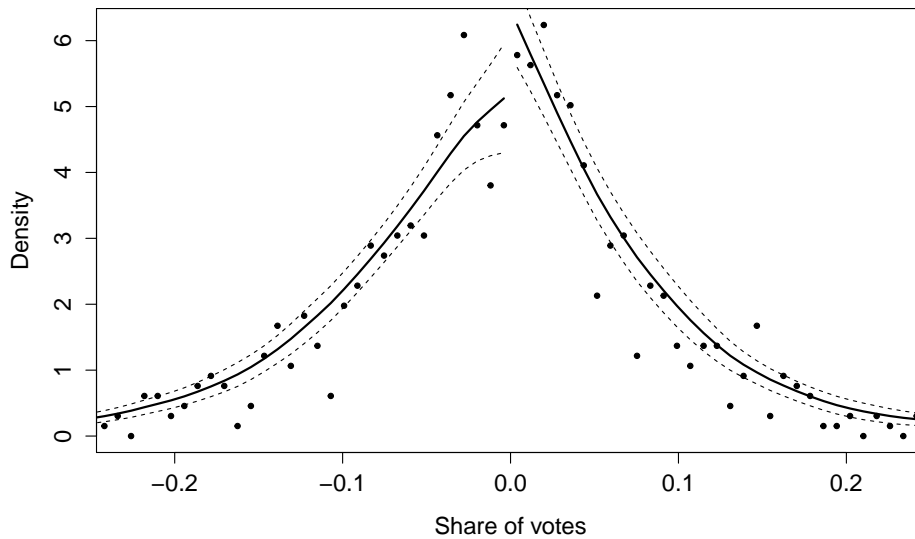
Figura 3.4: Estimated density function around the cutoff (McCrary test).

increased probability of being employed in the public sector or to higher wages for those already employed. We then study whether a party winning a mayoral election increases the number of registrations for that party in the following years, plausibly as a way to signal political support and access economic gains of political connections.

3.5.1
Government Wages

In this section we investigate whether being affiliated to a party connected to the mayor provides political connections that might help the individual enter the bureaucracy and rise in it. We do that by comparing the average wage accrued from and proportion that works in the government between individuals affiliated to the party which won the election and those affiliated to the party which lost the election, in close races.

Before we carry out our RD design, we plot local polynomial regressions of the log of the income accrued from the government (Figure 3.5) and the probability of working in the government (Figure 3.6) on the proportion of votes for a party in the presidential coalition, where the sample is composed of individuals registered to parties in that coalition.

Neither plot shows strong evidence of a discontinuity at the victory threshold. We apply our main empirical specification in Table 3.6, where we estimate the treatment effect of being affiliated to a party in the 2006 presidential
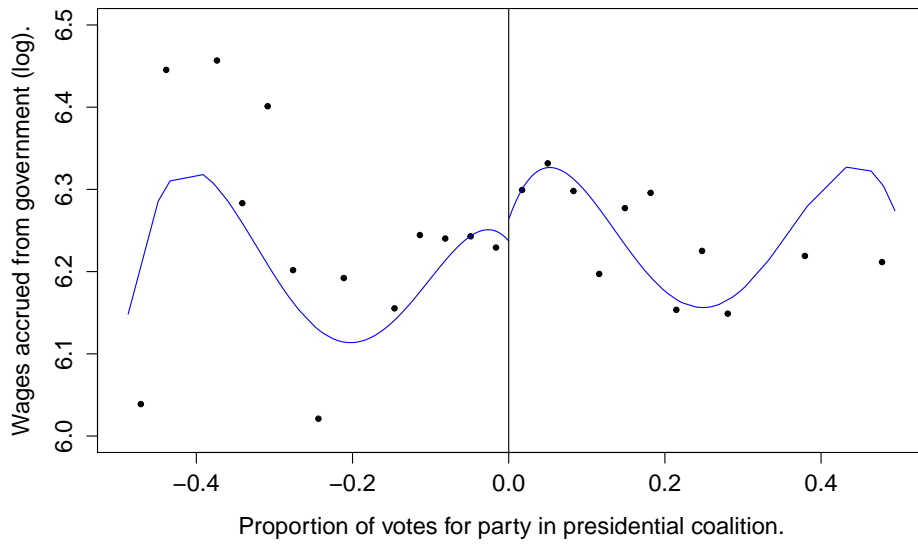
Figura 3.5: Effect of mayoral victory on wages accrued from government, for individuals registered to parties in the presidential coalition.
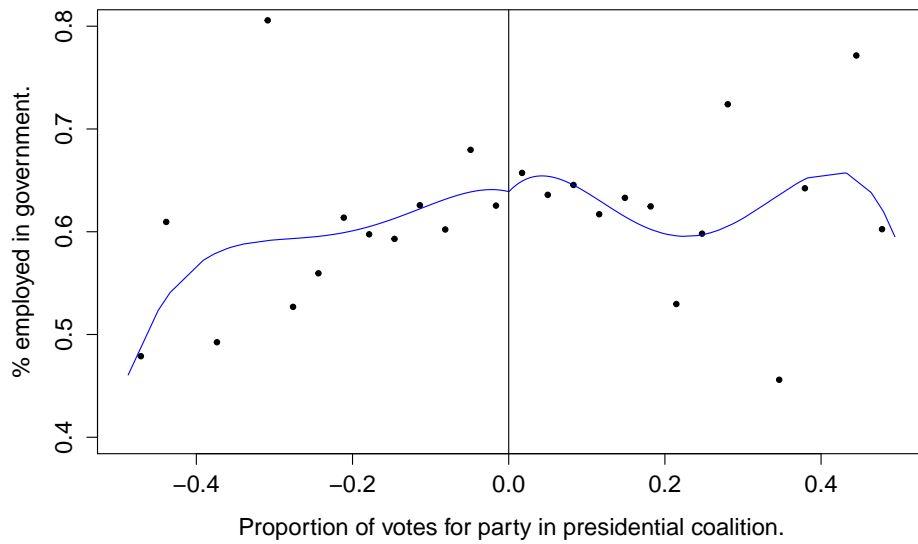

Figura 3.6: Effect of mayoral victory on percent working for the government, for individuals registered to parties in the presidential coalition.

coalition, when a party of that coalition wins the mayoral election, controlling for vote share with parametric and non-parametric (local polynomial) functions.[2] Except for a few particular specifications, most of the coefficients cannot be statistically discerned from zero.

Columns 1 to 6 apply parametric functions of the vote share, using only cities where the margin of victory in the mayoral elections was less than ten and five percent, respectively. Then, in columns 7 and 8 we use nonparametric local linear and local quadratic function, using the entire sample. The results are similar in both exercises.

The fact that members of the presidential coalition are not necessarily allied at the municipal level might introduce measurement error that would bias our results towards zero, attenuating the treatment effect and increasing the variance of the estimative. Therefore, in Table 3.7 we estimate our RD design for two specific parties, PMDB (Partido do Movimento Democrático Brasileiro) and PT (Partido dos Trabalhadores). We choose these parties as they are among the biggest in number of municipalities they controlled in 2008.

Consistently with our previous results, only in particular specifications are the coefficients statistically significant, and they are not stable between different strategies. These results support our conclusion that there is weak evidence of any effect of being affiliated to a political party on the probability of working for the government in the future, or receiving higher earnings when already working for it.

Finally, Table 3.8 presents the results for the regression with all parties pooled. In this case, every city is both in the control group and the treatment group, but with different individuals in each (we take the average value of each group). In this specification, as a check of robustness, we also look at the wages accrued from private firms and restrict government income to wages received from the prefecture. In both cases we still find no results.

3.5.2
Party Registration

If we believe that being affiliated to a political party can bring connections that have economic value, then plausibly when a party wins the mayoral election, these connections become more valuable, and we would expect more people to want to register to these parties. In this section we investigate whether this happens for our sample of Northeastern Brazilian cities for the 2008 mayoral election.

---

[2](52), for example, recommend the use of non-parametric specifications as a complement, rather than substitute, of parametric forms.

Tabela 3.6: Effect of party victory in mayo␣ral elections on numbe␣red individuals working for government, and income from government.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Running variable: share of votes for presidential coalition.								
Government income (log) (Mean: 6.252)	0.0526 (0.0624)	-0.0678 (0.0932)	-0.1911 (0.1196)	-0.0083 (0.0849)	-0.2881 (0.1236)**	0.0191 (0.0662)	-0.0033 (0.0570)	-0.1957 (0.0839)**
Works for government (Mean: 0.6351)	0.0038 (0.0342)	0.0104 (0.0517)	-0.0889 (0.0657)	0.0239 (0.0475)	-0.1610 (0.0673)**	-0.1700 (0.0866)*	0.0005 (0.0293)	-0.0278 (0.0345)
Polynomial	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic	Local Linear CCT Optimal	Local Quadratic CCT Optimal
Margin of victory	10%	10%	10%	5%	5%	5%		
N	2526	2526	2526	1664	1664	1664	3403	3403
Running variable: share of votes for opposing coalition.								
Government income (log) (Mean: 6.224)	0.0738 (0.0878)	0.0527 (0.1307)	-0.0378 (0.1756)	-0.0135 (0.1224)	0.1399 (0.1871)	-0.1677 (0.2495)	0.0735 (0.1031)	-0.0058 (0.1438)
Works for government (Mean: 0.6580)	-0.00313 (0.0436)	0.00644 (0.0657)	0.0223 (0.0891)	0.001 (0.0616)	0.0936 (0.0934)	-0.0117 (0.124)	0.0110 (0.0507)	0.0121 (0.0660)
Polynomial	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic	Local Linear CCT Optimal	Local Quadratic CCT Optimal
Margin of victory	10%	10%	10%	5%	5%	5%		
N	838	838	838	545	545	545	1101	1101

Running variable is the share of votes of top 2 candidates that the candidate received, for a candidate in a party in presidential (resp. opposition) coalition, and for an individual affiliated to a party in the same coaltion. For the parametric specifications, the errors are clustered at the city level. The symbol * denotes p-value smaller than 0.1, ** for 0.05 and *** for 0.01.

Tabela 3.7: Effect of party victory in mayored individuals working for government, and income from government (political parties).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Running variable: share of votes for PMDB.								
Government income (log) (Mean: 6.358)	0.1786 (0.0937)*	0.0462 (0.1394)	0.0050 (0.1768)	0.0886 (0.1291)	-0.0296 (0.1860)	0.1032 (0.2317)	0.0229 (0.1456)	0.0165 (0.1707)
Works for government (Mean: 0.7059)	0.0373 (0.0446)	0.0704 (0.0711)	0.0124 (0.0962)	0.0597 (0.0666)	-0.0272 (0.0997)	0.0500 (0.134)	0.0542 (0.0637)	0.0568 (0.0737)
Polynomial	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic	Local Linear	Local Quadratic
Margin of victory	10%	10%	10%	5%	5%	5%	CCT Optimal	CCT Optimal
N	478	478	478	319	319	319	621	621
Running variable: share of votes for PT.								
Government income (log) (Mean: 6.447)	0.0945 (0.1425)	-0.1993 (0.2005)	-0.2589 (0.2568)	-0.0761 (0.1896)	-0.3628 (0.2646)	-0.3114 (0.3537)	-0.2269 (0.2129)	-0.2691 (0.2344)
Works for government (Mean: 0.6705)	0.0333 (0.0673)	-0.169* (0.102)	-0.137 (0.133)	-0.0973 (0.0947)	-0.241* (0.141)	-0.212 (0.175)	-0.1623 (0.1185)	-0.1820 (0.1321)
Polynomial	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic	Local Linear	Local Quadratic
Margin of victory	10%	10%	10%	5%	5%	5%	CCT Optimal	CCT Optimal
N	188	188	188	112	112	112	269	269

Running variable is the share of votes of top 2 candidates that the candidate received, for a candidate in PMDB (resp. PT) and the voter also affiliated to PMDB (resp. PT). For the parametric specifications, the errors are clustered at the city level. The symbol * denotes p-value smaller than 0.1, ** for 0.05 and *** for 0.01.

Tabela 3.8: Effect of party victory in mayoral elections on probability of registered individuals working for government, and income from government (pooled).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Government wage (log)	0.1914	0.0863	0.0037	0.1234	-0.0307	0.0209	0.0656	0.0106
(Mean: 6.338)	(0.0392)***	(0.0570)	(0.0710)	(0.0540)	(0.0747)	(0.0939)	(0.0657)	(0.0787)
Works for government	0.0567***	0.00563	-0.0221	0.0266	-0.0419	-0.0177	-0.0025	-0.0197
(Mean: 0.6850)	(0.0149)	(0.0221)	(0.0288)	(0.0205)	(0.0302)	(0.0385)	(0.0339)	(0.0391)
Private firm wage (log)	-0.1508	0.0270	0.1961	-0.0100	0.2534	0.0909	0.1039	0.1471
(Mean: 4.958)	(0.0808)	(0.1166)	(0.1481)	(0.1089)	(0.1580)	(0.1972)	(0.1467)	(0.1606)
Works for private firm	-0.0078	0.0015	-0.0304***	-0.00939	0.00604	-0.0256	0.0375	0.0009
(Mean: 0.1811)	(0.0231)	(0.0268)	(0.0117)	(0.0170)	(0.0225)	(0.0159)	(0.0239)	(0.0306)
Prefecture wage (log)	0.240***	0.116**	-0.00515	0.134***	-0.0180	0.0133	0.0439	-0.0043
(Mean: 6.197)	(0.0381)	(0.0540)	(0.0663)	(0.0505)	(0.0689)	(0.0868)	(0.0683)	(0.0791)
Works for prefecture	0.0602***	0.00831	-0.0235	0.0272	-0.0379	-0.0289	-0.0024	-0.0197
(Mean: 0.6568)	(0.0153)	(0.0226)	(0.0297)	(0.0211)	(0.0311)	(0.0400)	(0.0352)	(0.0404)
Margin of victory	Linear 10%	Quadratic 10%	Cubic 10%	Linear 5%	Quadratic 5%	Cubic 5%	Local Linear CCT Optimal	Local Quadratic CCT Optimal
N	2142	2142	2142	1442	1442	1442	2833	2833

Running variable is the share of votes of top 2 candidates that the candidate of the party the individual is affiliated received. For the parametric specifications, the errors are clustered at the city level. The symbol * denotes p-value smaller than 0.1, ** for 0.05 and *** for 0.01.

Before presenting the election results, we plot the local polynomial regression of number of registrations to a party after the election in a given city on the proportion of votes for that party in the mayoral elections. For this plot, we take the logarithm of the number of registrations, for graphical clarity. (In the regressions below, we use the absolute value of registrations, and the results are the same.) As is apparent in Figure 3.7, there seems to be no discontinuity in the number of registrations around the cutoff.
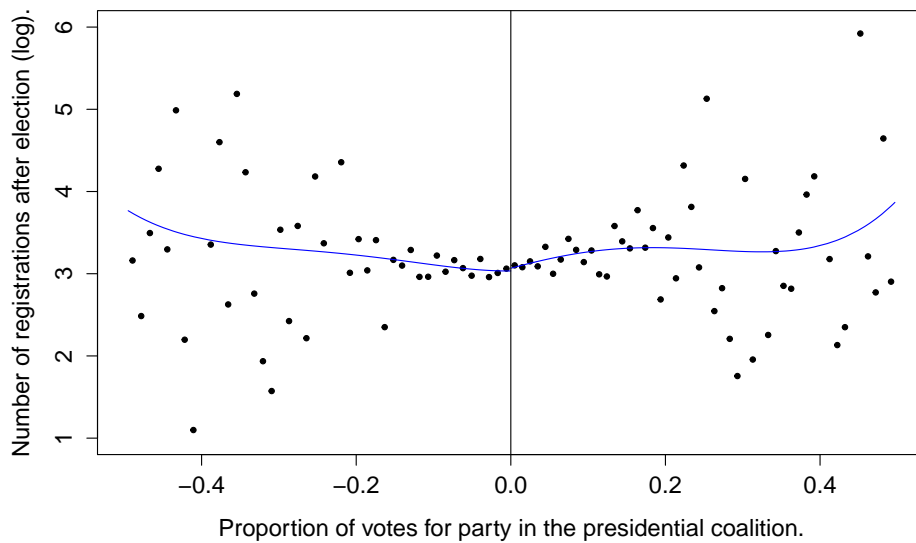
Figura 3.7: Effect of mayoral victory on future party affiliations, for individuals registered to parties in the presidential coalition.

We formalize the procedure of Figure 3.7 in Table 3.9. This table presents the results of parametrical and non-parametrical (local) polynomial RD designs on the number of post-election registrations, the number of registrations in the first year after the elections, and the number of registrations during the victorious mayor's term. In none of these specification are the results statistically significant or consistent between different designs.

We interpret these results as indicating that individuals do not perceive affiliation to the mayor's party as providing any economic benefits, which is consistent with us not finding any evidence of affiliates' salary being any higher than non-affiliates.

3.6
Conclusion

In this paper, we have examined how registered members fo political parties are affected by their party's victory in mayoral elections. We investigated

Tabela 3.9: Effect of party victory in mayoral elections on future party registrations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Running variable: share of votes for party in mayoral election.								
Affiliated to party after election	-7.938	14.401	22.592	20.411	1.370	-24.643	0.0567	-12.552
(Mean: 65.46)	(7.419)	(8.784)	(13.812)	(12.748)	(7.515)	(16.579)	(9.136)	(12.372)
Affiliated to party, 1st year of term	-5.261	10.110	15.507	13.621	1.342	-19.501	-2.465	-2.412
(Mean: 29.78)	(5.345)	(6.134)*	(9.855)	(8.853)	(5.181)	(11.530)*	(6.004)	(6.926)
Affiliated to party, whole term	-8.568	13.373	20.878	19.149	-0.1402	-25.734	-1.229	-13.047
(Mean: 64.26)	(7.363)	(8.695)	(13.689)	(12.648)	(7.363)	(16.391)	(8.985)	(11.884)
Polynomial	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic	Local Linear CCT Optimal	Local Quadratic CCT Optimal
Margin of victory	10%	10%	10%	5%	5%	5%	CCT Optimal	CCT Optimal
N	1401	1401	1401	909	909	909	3625	3625

Running variable is the share of votes of top 2 candidates that the candidate of the party received. The symbol * denotes p-value smaller than 0.1, ** for 0.05 and *** for 0.01. The optimal CCT bandwiths are 0.054 (771/780; linear) and 0.086 (1062/1076; quadratic) for the first dependent variable, 0.069 (918/929) and 0.069 (916/926) for the second, and 0.053 (763/771) and 0.083 (1035/1049) for the third.

the causal effect of the party an individual is affiliated to winning an election on the probability of working for the government, on the wages accrued from the government, and on the number of future affiliations for that party in that city. In none of these cases we were able to find a discernible non-zero effect.

This might indicate that political parties do not have as much leeway for patronage as previously thought, perhaps because of institutional constraints, as the need for public exams into bureaucracy and audits by governmental agencies. Since most political affiliates are already public employees, if they are hired by already specified contracts, the party might not have the power to affect their wages, as they are defined by law. (This might be especially the case when there is a well defined career path, as with teachers or police officers.)

Another possibility is that bureaucracy appointments are slow, and therefore supporters only acquire governmental jobs more than a year after the election, which we would not be able to capture with our current dataset. It is important to note, therefore, that this empirical exercise covers only the 2008 mayoral elections for the Northeast of Brazil, and that we had access only to individual socioeconomic data for the years of 2008 and 2009. We hope, in future research, to increase our database to cover all years for which the RAIS is available, and to extend this analysis to all mayoral elections after the re-democratization. With that enlarged database, we would be able to ascertain more confidently whether our results indicate absence of a treatment effect (as opposed to our inability to capture it), as well as being able to investigate heterogeneous effects according to city, party, and individual characteristics.

Referências Bibliográficas

[01] BISIN, A.; VERDIER, T.. **Beyond the melting pot: Cultural transmission, marriage, and the evolution of ethnic and religious traits**. The Quarterly Journal of Economics, 115(3):955–988, 2000.

[02] BISIN, A.; VERDIER, T.. **The economics of cultural transmission and the dynamics of preferences**. Journal of Economic Theory, 97(2):298–319, 2001.

[03] WEBER, M.. **The Protestant ethic and the spirit of capitalism**. Roxbury Los Angeles, 1998.

[04] BENABOU, R.; TIROLE, J.. **Belief in a just world and redistributive politics**. The Quarterly Journal of Economics, 121(2):699–746, 2006.

[05] ALESINA, A.; ANGELETOS, G.-M.. **Fairness and redistribution**. American Economic Review, p. 960–980, 2005.

[06] BARRON, J. M.; GJERDE, K. P.. **Peer pressure in an agency relationship**. Journal of Labor economics, 15(2):234–54, 2008.

[07] ICHINO, A.; MAGGI, G.. **Work environment and individual background: explaining regional shirking differentials in a large italian firm**. The Quarterly Journal of Economics, 115(3):1057–1090, 2000.

[08] ROTEMBERG, J. J.. **Human relations in the workplace**. Journal of Political Economy, p. 684–717, 1994.

[09] KANDEL, E.; LAZEAR, E. P.. **Peer pressure and partnerships**. Journal of political Economy, p. 801–817, 1992.

[10] LINDBECK, A.; NYBERG, S. ; WEIBULL, J. W.. **Social norms and economic incentives in the welfare state**. The Quarterly Journal of Economics, 114(1):1–35, 1999.

[11] FISCHER, P.; HUDDART, S.. **Optimal contracting with endogenous social norms**. The American Economic Review, 98(4):pp. 1459–1475, 2008.

[12] HUCK, S.; KÜBLER, D. ; WEIBULL, J.. **Social norms and economic incentives in firms.** Journal of Economic Behavior & Organization, 83(2):173–185, 2012.

[13] BISIN, A.; VERDIER, T.. **Chapter 9 - the economics of cultural transmission and socialization.** volumen 1 de **Handbook of Social Economics**, p. 339 – 416. North-Holland, 2011.

[14] WEIBULL, J. W.. **Evolutionary game theory.** MIT press, 1997.

[15] CLARK, G.. **A farewell to alms: a brief economic history of the world.** Princeton University Press, 2008.

[16] CLARK, G.. **Why isn't the whole world developed? lessons from the cotton mills.** The Journal of economic history, 47(01):141–173, 1987.

[17] HAUK, E.; SAEZ-MARTI, M.. **On the cultural transmission of corruption.** Journal of Economic Theory, 107(2):311–335, 2002.

[18] LINDBECK, A.; NYBERG, S.. **Raising children to work hard: altruism, work norms, and social insurance.** The Quarterly Journal of Economics, 121(4):1473–1503, 2006.

[19] BECKER, S. O.; WOESSMANN, L.. **Was weber wrong? a human capital theory of protestant economic history.** The Quarterly Journal of Economics, 124(2):531–596, 2009.

[20] CANTONI, D.. **The economic effects of the protestant reformation: Testing the weber hypothesis in the german lands.** 2013.

[21] DOEPKE, M.; ZILIBOTTI, F.. **Occupational choice and the spirit of capitalism.** The Quarterly Journal of Economics, 123(2):747–793, 2008.

[22] KREMER, M.. **The o-ring theory of economic development.** The Quarterly Journal of Economics, 108(3):551–575, 1993.

[23] STIGLER, G. J.; BECKER, G. S.. **De gustibus non est disputandum.** The American Economic Review, p. 76–90, 1977.

[24] TOPKIS, D. M.. **Supermodularity and complementarity.** Princeton University Press, 1998.

[25] BRYNJOLFSSON, E.; MILGROM, P.. **Complementarity in organizations.** The Handbook of Organizational Economics, 11, 2012.

[26] BRESNAHAN, T. F.; BRYNJOLFSSON, E. ; HITT, L. M.. **Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence.** The Quarterly Journal of Economics, 117(1):339–376, 2002.

[27] GROSHEN, E. L.. **Sources of intra-industry wage dispersion: How much do employers matter?** The Quarterly Journal of Economics, 106(3):869–884, 1991.

[28] CAVALLI-SFORZA, L. L.. **Cultural transmission and evolution: A quantitative approach**. Número 16. Princeton University Press, 1981.

[29] BOYD, R.; RICHERSON, P.. **Culture and the evolutionary process**, 1985.

[30] FRANKLIN, B.. **Advice to a young tradesman, written by an old one**. The Writings of Benjamin Franklin: Philadelphia, 1726-1757, 2007.

[31] BAXTER, R.. **A Christian directory**, volumen I. Soli Deo Gloria Publications, 1996.

[32] HILLER, V.. **Work organization, preferences dynamics and the industrialization process**. European Economic Review, 55(7):1007–1025, 2011.

[33] MILGROM, P.; ROBERTS, J.. **Rationalizability, learning, and equilibrium in games with strategic complementarities**. Econometrica, 58(6):1255–1277, 1990.

[34] MELTZER, A. H.; RICHARD, S. F.. **A rational theory of the size of government**. The Journal of Political Economy, p. 914–927, 1981.

[35] BENABOU, R.; OK, E. A.. **Social mobility and the demand for redistribution: The poum hypothesis**. The Quarterly Journal of Economics, 116(2):447–487, 2001.

[36] FONG, C.. **Social preferences, self-interest, and the demand for redistribution**. Journal of Public Economics, 82(2):225–246, 2001.

[37] LUTTMER, E. F.. **Neighbors as negatives: Relative earnings and well-being***. The Quarterly journal of economics, 120(3):963–1002, 2005.

[38] PIKETTY, T.. **Social mobility and redistributive politics**. The Quarterly journal of economics, p. 551–584, 1995.

[39] ALESINA, A.; LA FERRARA, E.. **Preferences for redistribution in the land of opportunities**. Journal of Public Economics, 89(5):897–931, 2005.

[40] ALESINA, A.; DI TELLA, R. ; MACCULLOCH, R.. **Inequality and happiness: are europeans and americans different?** Journal of Public Economics, 88(9):2009–2042, 2004.

[41] CORNEO, G.; GRÜNER, H. P.. **Social limits to redistribution**. American Economic Review, p. 1491–1507, 2000.

[42] CORNEO, G.; GRÜNER, H. P.. **Individual preferences for political redistribution**. Journal of public Economics, 83(1):83–107, 2002.

[43] ALESINA, A.; GLAESER, E.. **Fighting poverty in the us and europe: A world of difference**. 2004.

[44] MICHAEL S. LEWIS-BECK, T. W. R.. **Localism in presidential elections: The home state advantage**. American Journal of Political Science, 27(3):548–556, 1983.

[45] MILLO, G.; PIRAS, G.. **splm: Spatial panel data models in r**. Journal of Statistical Software, 47(1):1–38, 2012.

[46] KAPOOR, M.; KELEJIAN, H. H. ; PRUCHA, I. R.. **Panel data models with spatially correlated error components**. Journal of Econometrics, 140(1):97 – 130, 2007. Analysis of spatially dependent data.

[47] GLAESER, E. L.; SACERDOTE, B. I. ; SCHEINKMAN, J. A.. **The social multiplier**. Journal of the European Economic Association, 1(2-3):345–353, 2003.

[48] LEE, D. S.. **Randomized experiments from non-random selection in u.s. house elections**. Journal of Econometrics, 142(2):675 – 697, 2008. The regression discontinuity design: Theory and applications.

[49] BESLEY, T.; PANDE, R. ; RAO, V.. **Just rewards? local politics and public resource allocation in south india**. The World Bank Economic Review, 26(2):191–216, 2012.

[50] BET CAEYERS, S. D.. **Political connections and social networks in targeted transfer programs: Evidence from rural ethiopia**. Economic Development and Cultural Change, 60(4):639–675, 2012.

[51] MARKUSSEN, T.; TARP, F.. **Political connections and land-related investment in rural vietnam**. Journal of Development Economics, 110:291 – 302, 2014. Land and Property Rights.

[52] LEE, D. S.; LEMIEUX, T.. **Regression discontinuity designs in economics**. Journal of Economic Literature, 48(2):281–355, 2010.

[53] BUCHANAN, J. M.; TULLOCK, G.. **The calculus of consent**, volumen 3. University of Michigan Press Ann Arbor, 1962.

[54] LIZZERI, A.; PERSICO, N.. **The provision of public goods under alternative electoral incentives**. American Economic Review, p. 225–239, 2001.

[55] FOLKE, O.; HIRANO, S. ; SNYDER, J. M.. **Patronage and elections in us states**. American Political Science Review, 105(03):567–585, 2011.

[56] ENIKOLOPOV, R.. **Politicians, bureaucrats and targeted redistribution**. Journal of Public Economics, 120:74–83, 2014.

[57] ALESINA, A.; BAQIR, R. ; EASTERLY, W.. **Redistributive public employment**. Journal of Urban Economics, 48(2):219–241, 2000.

[58] COATE, S.; MORRIS, S.. **On the form of transfers to special interests**. Journal of political Economy, p. 1210–1235, 1995.

[59] ROBINSON, J. A.; VERDIER, T.. **The political economy of clientelism**. The Scandinavian Journal of Economics, 115(2):260–291, 2013.

[60] HILLER, V.; VERDIER, T.. **Corporate culture and identity investment in an industry equilibrium**. Journal of Economic Behavior & Organization, 103:93–112, 2014.

[61] SÁEZ-MARTI, M.; ZENOU, Y.. **Cultural transmission and discrimination**. Journal of Urban Economics, 72(2):137–146, 2012.

[62] GALOR, O.; MOAV, O.. **Natural selection and the origin of economic growth**. Quarterly Journal of Economics, p. 1133–1191, 2002.

[63] GALOR, O.; MICHALOPOULOS, S.. **Evolution and the growth process: Natural selection of entrepreneurial traits**. Journal of Economic Theory, 147(2):759–780, 2012.

[64] DUR, R.; SOL, J.. **Social interaction, co-worker altruism, and incentives**. Games and Economic Behavior, 69(2):293–301, 2010.

A
Proofs of propositions of Chapter 1

In Appendix A, I prove all the propositions stated in the paper. I ignore equalities throughout, as they would only encumber the presentation without any effect on the results (and they have measure zero in the parameter space).

A.1
Proof of Proposition 1.1

Prova. Given a contract $\{b_{ij}\}_{i,j \in \{W, N\}}$, the problem of an employee i matched with a coworker j is to maximize (1-1), which is a strictly concave problem (for $\alpha < 1$) with unique solution given implicitly by:

$$b_{ij}(a_{ij}, \ a_{ji}) = \frac{\kappa_i a_{ij}^2}{\theta \alpha a_{ij}^\alpha a_{ji}^\alpha} \text{ and } \mu(a_i) = \frac{\kappa_i a_{ij}^2}{\alpha}. \tag{A-1}$$

Given that, the firm maximizes (1-2), which is again a strictly concave problem for $\alpha < 1$ with a unique solution given by (1-3).

Plugging this into the utility function of the agent, I obtain:

$$V(i, \ j) = \kappa_i \left(\frac{1}{\alpha} - \frac{1}{2} \right) \left[\frac{(\theta \rho \alpha^2)^{1+\alpha A}}{2^{1+\alpha A} \kappa_i \kappa_j^{\alpha A}} \right]^{\frac{2A}{1-\alpha^2 A^2}}.$$

All that remains to be shown is that the $V(i, \ j)$ above satisfies increasing differences in types. As is well known, see (24), for C^2 functions increasing differences is equivalent to positive cross derivatives. Indeed:

$$\frac{\partial V(i, \ j)}{\partial \kappa_i \partial \kappa_j} = M \left(-\frac{2A\alpha A}{1-\alpha^2 A^2} \right) \left(1 - \frac{2A}{1-\alpha^2 A^2} \right) \kappa_j^{-\frac{2A\alpha A}{1-\alpha^2 A^2}-1} \kappa_i^{-\frac{2A}{1-\alpha^2 A^2}-1} > 0,$$

where M is a positive value independent of κ_i and κ_j. Remembering that $A \equiv \frac{1}{2-\alpha}$, all parameters are positive, and since $\alpha \in (0, \ 1)$, necessarily $A \in (\frac{1}{2}, \ 1)$. All that remains to be shown is that $\frac{2A}{1-\alpha^2 A^2} > 1$. Clearly, $2A > 1$ and $0 < 1 - \alpha^2 A^2 < 1$.

■

A.2
Proof of Proposition 1.2

Prova. Throughout the proof, I assume that the initial state is interior; otherwise, the problem is trivial. From equation (1-4) and substituting for the probabilities of success in socialization I obtain that a parent of type i will choose $d^i \in [0, 1]$ such that:

$$(1 - x^i) \left[U^i(x^W) - U^j(x^W) \right] \leq d^i,$$

with equality if $d^i > 0$. Clearly, in equilibrium in each period, only one kind of parent exerts positive socialization effort: the kind of parent for whom the ex ante labor market outcome is higher. This happens for W if and only if:

$$\Delta U(x^W) \equiv U^W(x^W) - U^N(x^W) = x^W \left[V(W, W) - V(N, W) \right] +$$

$$(1 - x^W) \left[V(W, N) - V(N, N) \right] > 0$$

According to Proposition 1.1, I know V is increasing differences in types. Therefore:

(a) $V(N, j) > V(W, j)$ for all j, and so $\Delta U(x^W) < 0$ for all x^W;

(b) $V(W, j) > V(N, j)$ for all j, and so $\Delta U(x^W) > 0$ for all x^W;

(c) $V(W, W) > V(N, W)$ and $V(N, N) > V(W, N)$.

Moreover, rewriting $\Delta U(x^W)$, I obtain:

$$\Delta U(x^W) = \left[V(W, N) - V(N, N) \right] +$$

$$x^W \left[(V(W, W) - V(N, W)) + (V(N, N) - V(W, N)) \right]$$

In case (c), $\Delta U(x^W)$ is always increasing in x^W, $\Delta U(0) < 0$ and $\Delta U(1) > 0$. Further, (for interior points):

$$\Delta U(x^W) = 0 \iff$$

$$x = \frac{V(N, N) - V(W, N)}{(V(W, W) - V(N, W)) + (V(N, N) - V(W, N))}$$

Having established the conditions for which $d^i > 0$, only demonstrating the cultural transmission process converging properties remains. Indeed, as

$$x_{t+1}^W = x_t^W P^{WW}(x^W) + (1 - x_t^W) P^{NW}(x^W),$$

I obtain:

$$x_{t+1}^W = x_t^W + x_t^W (1 - x_t^W) \left[d_t^W - d_t^N \right],$$

which can be approximated by the continuous process $\dot{x}^W = x^W(1 - x^W)\left[d^W(x^W) - d^N(x^W)\right]$.[1] Therefore, for interior points, $\dot{x}^W > 0 \iff d^W(x^W) > d^N(x^W) \iff \Delta U(x^W) > 0$. To complete the proof, I show that the only stationary points are 0, x^W, and 1. As stated, $\dot{x}^W = 0$ if and only if $x^W = 0$, $x^W = 1$ (by the equation above), or when $d^W(x^W) = d^N(x^W) \iff \Delta U(x^W) = 0$, which occurs only at x^*. ∎

A.3
Proof of Proposition 1.3

Prova. For ease of exposition, I divide the proof into three steps. Throughout the proof, I assume that the initial state is interior; otherwise, the problem is trivial.

Part 1: First, for simplicity I normalize (with no loss of generality) the profit such that $Y_{WN}^C - Y_{WN}^S = 0$, where Y_{ij}^q is the probability of success of a firm with employees type i and j and technology q. Clearly, as there are no mutations, homogeneous populations are always stable. By equation (1-8), if $\beta_t^C \in (0, 1)$, then:

$$\beta_{t+1}^C > \beta_t^C \iff Y(\beta_t^C, x_t^W, C) > Y(\beta_t^C, x_t^W, S).$$

By my assumptions, this is independent from β_t^C. To show that the $\beta\beta$-locus is a vertical line, with β^C increasing on the right and decreasing on the left, it remains to be shown that $Y(\beta_t^C, x_t^W, C) - Y(\beta_t^C, x_t^W, S) \equiv \Delta Y(\beta_t^C, x_t^W)$ is everywhere strictly increasing in x_t^W. Henceforth, I drop the time subscripts for ease of notation.

Indeed, I have that:

$$\Delta Y(\beta^C, x^W) = (x^W)^2 \left[Y_{WW}^C - Y_{WW}^S\right] +$$

$$2x^W(1 - x^W)\left[Y_{WN}^C - Y_{WN}^S\right] + (1 - x^W)^2 \left[Y_{NN}^C - Y_{NN}^S\right]$$

Denote the derivative of $\Delta Y(x^W)$, shown above, by $D\Delta Y(x^W)$. Then:

$$\frac{1}{2}D\Delta Y(x^W) = \left[Y_{WN}^C - Y_{WN}^S\right] - \left[Y_{NN}^C - Y_{NN}^S\right] +$$

$$x^W \left\{\left[Y_{WW}^C - Y_{WW}^S\right] - 2\left[Y_{WN}^C - Y_{WN}^S\right] + \left[Y_{NN}^C - Y_{NN}^S\right]\right\}$$

(Henceforth I omit the multiplier, as I am only interested in the sign of the derivative.) As the inside of the brackets is a constant (in x^W), the derivative is (generically) either strictly increasing or strictly decreasing in x^W, and it is

---

[1]For details, see (14). This provides a formal justification for Figures 3 to 5.

sufficient to evaluate the equation above at the corners (i.e., with x^W equaling 0 or 1). A necessary and sufficient condition for the above derivative to be positive everywhere is that $\left[Y^C_{WW} - Y^S_{WW}\right] > \left[Y^C_{WN} - Y^S_{WN}\right] > \left[Y^C_{NN} - Y^S_{NN}\right]$, or under my normalization, that:

$$\left[Y^C_{WW} - Y^S_{WW}\right] > 0 > \left[Y^C_{NN} - Y^S_{NN}\right].$$

I start by noting that (under my normalization) $\Delta Y(1) = Y^C_{WW} - Y^S_{WW}$ and $\Delta Y(0) = Y^C_{NN} - Y^S_{NN}$. Note also that $D\Delta Y(1) = Y^C_{WW} - Y^S_{WW}$ and $D\Delta Y(0) = -\left[Y^C_{NN} - Y^S_{NN}\right]$, and that $D\Delta Y(x^W)$ is a linear combinations of both.

There are four possible cases:
(i) $Y^C_{WW} - Y^S_{WW} > 0$ and $Y^C_{NN} - Y^S_{NN} > 0$;
(ii) $Y^C_{WW} - Y^S_{WW} > 0$ and $Y^C_{NN} - Y^S_{NN} < 0$;
(iii) $Y^C_{WW} - Y^S_{WW} < 0$ and $Y^C_{NN} - Y^S_{NN} > 0$; and
(iv) $Y^C_{WW} - Y^S_{WW} < 0$ and $Y^C_{NN} - Y^S_{NN} < 0$.

Thus, I now show that: a) case (iii) is impossible and that b) in cases (i) and (iv) $\Delta Y(x^W)$ has no root.

a) The revenue of the firm with complementary technology is given by:

$$Y(\kappa_i, \ \kappa_j) = \rho\theta \left[\frac{(\theta\rho\alpha^2)^{1+\alpha A}}{2^{1+\alpha A}\kappa_i\kappa_j^{\alpha A}}\right]^{\frac{\alpha A}{1-\alpha^2 A^2}} \left[\frac{(\theta\rho\alpha^2)^{1+\alpha A}}{2^{1+\alpha A}\kappa_j\kappa_i^{\alpha A}}\right]^{\frac{\alpha A}{1-\alpha^2 A^2}}$$

Now, I show that $Y(\kappa_i, \ \kappa_j)$ is increasing differences in κ_i and κ_j, and $Y^S_{ij} = Y^S_i + Y^S_j$ (a trivial consequence of separability). Indeed:

$$\frac{\partial Y^C_{ij}}{\partial i \partial j} = \left[-(1+\alpha A)\frac{\alpha A}{1-\alpha^2 A^2}\right]^2 C\kappa_i^{-(1+\alpha A)\frac{\alpha A}{1-\alpha^2 A^2}-1} \kappa_j^{-(1+\alpha A)\frac{\alpha A}{1-\alpha^2 A^2}-1} > 0,$$

where C is a positive constant (in $(\kappa_i, \ \kappa_j)$).

b) For (i), $D\Delta Y(1) \geq D\Delta Y(x^W)$ for all x^W (as they are a linear combination of $D\Delta Y(1) > 0$ and $D\Delta Y(0) < 0$). By the Fundamental Theorem of Calculus, if there was a y between 0 and 1 such that $\Delta Y(y) = 0$, then:

$$\Delta Y(1) = \Delta Y(1) - \Delta Y(y) = \int_y^1 D\Delta Y(x) \ dx \leq$$

$$\int_y^1 D\Delta Y(1) \ dx = (1-y)\Delta Y(1) < \Delta Y(1),$$

which is a contradiction. (Here, I use $D\Delta Y(1) = \Delta Y(1)$.) Analogously, for (iv), $D\Delta Y(x^W) \leq D\Delta Y(0)$, and thus, if there is $y \in (0,1)$ such that $\Delta Y(y) = 0$,

then one obtains *reductio ad absurdum*:

$$-\Delta Y(0) = \Delta Y(y) - \Delta Y(0) = \int_0^y D\Delta Y(x)\,\mathrm{d}x \geq$$

$$\int_0^y D\Delta Y(0)\,\mathrm{d}x = y\left(-\Delta Y(0)\right) > -\Delta Y(0)$$

(I used $D\Delta Y(0) = -\Delta Y(0)$.) Therefore, it is either case (ii), or it is always superior to choose separable technology or always preferable to choose complementary technology, and the claim is proven.

Part 2: Now I examine the xx-locus. As the ex ante utility is given by (1-9), I have that:

$$\Delta U(x^W; \beta^C) = \left(V_W^S - V_N^S\right) +$$

$$\beta^C \left\{ x^W \left(V_{WW}^C - V_{NW}^C\right) + (1 - x_W)\left(V_{WN}^C - V_{NN}^C\right) - \left(V_W^S - V_N^S\right) \right\}.$$

I now consider the hyperplane $\Delta U(x^W; \beta^C) = 0$ (which must be the case for any interior xx-locus point). First, I note that for any given β^C, $\Delta U(x^W; \beta^C)$ is strictly increasing in x^W (because of Proposition 1.1). Consequently, I obtain that for any given β^C, if $\hat{x}$ is such that $\Delta U(\hat{x}; \beta^C) = 0$, then for any $x^W > \hat{x}$, $\Delta U(x^W; \beta^C) > 0$ (and thus $x_{t+1}^W > x_t^W$; see the proof of Proposition 1.2) and *vice versa*. Furthermore, as the derivative $D\Delta U(x^W, \beta)$ is everywhere positive, I can apply the Implicit Function Theorem on the surface $\Delta U(x^W, \beta^C) = 0$. So, it must be true that:

$$\frac{\partial \beta^C}{\partial x_W} = -\frac{\partial_x \Delta U(x^W, \beta^C)}{\partial_\beta \Delta U(x^W, \beta^C)} =$$

$$-\frac{\beta^C \left[\left(V_{WW}^C - V_{NW}^C\right) - \left(V_{WN}^C - V_{NN}^C\right)\right]}{x^W \left[(V_{WW}^C - V_{NW}^C) - (V_{WN}^C - V_{NN}^C)\right] + (V_{WN}^C - V_{NN}^C) - (V_W^S - V_N^S)}.$$

Because the numerator is always positive, for the surface to be downward sloping I need the denominator to be positive, that is, for:

$$x^W \left[\left(V_{WW}^C - V_{NW}^C\right) - \left(V_{WN}^C - V_{NN}^C\right)\right] + \left(V_{WN}^C - V_{NN}^C\right) - \left(V_W^S - V_N^S\right) > 0.$$

However, the condition above is only relevant where $\Delta U(x^W, \beta^C) = 0$. That happens where, assuming interior β^C and x^W:

$$x^W \left(V_{WW}^C - V_{NW}^C\right) = \frac{\left(1 - \beta^C\right)}{\beta^C}\left(V_N^S - V_W^S\right) + (1 - x^W)\left(V_{NN}^C - V_{WN}^C\right)$$

Ergo, the points x^W (which I denote as $\hat{x}$) for which $\Delta U(x^W, \beta^C) = 0$ as

a function of β^C are:

$$\hat{x} = \frac{\left(V_{NN}^C - V_{WN}^C\right) + \frac{\left(1-\beta^C\right)}{\beta^C}\left(V_N^S - V_W^S\right)}{\left(V_{WW}^C - V_{NW}^C\right) - \left(V_{WN}^C - V_{NN}^C\right)}.$$

(Here, I inverted the $V_{NN}^C - V_{WN}^C$ term.) So, the condition is that for all x satisfying this, that:

$$\frac{\partial \Delta U(x^W; \beta^C)}{\partial \beta^C} = \hat{x}\left[\left(V_{WW}^C - V_{NW}^C\right) - \left(V_{WN}^C - V_{NN}^C\right)\right] +$$

$$\left[\left(V_{WN}^C - V_{NN}^C\right) - \left(V_W^S - V_N^S\right)\right] > 0$$

$$\iff \frac{\left(1 - \beta^C\right)}{\beta^C}\left(V_N^S - V_W^S\right) + \left(V_N^S - V_W^S\right) > 0$$

(Again, some terms were inverted.) This always happens when $V_N^S > V_W^S$, and my proof is complete. ∎

A.4
Proof of Proposition 1.4

Prova.

I show that $\frac{\partial[V(W,\ j) - V(N,\ j)]}{\partial \xi} > 0$, for $\xi \equiv \theta\rho$ (i.e. V is increasing differences in the agent's type and the parameters). The other two assertions are trivial, since they impact directly the relative payoffs of the different types. Indeed:

$$V(W,\ j) - V(N,\ j) = \xi^{(1+\alpha A)\frac{2A}{1-\alpha^2 A^2}}\left[F(\kappa_W) - F(\kappa_N)\right],$$

where

$$F(\kappa_i) \equiv \kappa_i\left(\frac{1}{\alpha} - \frac{1}{2}\right)\left[\frac{(\alpha^2)^{1+\alpha A}}{2^{1+\alpha A}\kappa_i\kappa_j^{\alpha A}}\right]^{\frac{2A}{1-\alpha^2 A^2}}$$

is an everywhere decreasing function of κ_i (as $\frac{2A}{1-\alpha^2 A^2} > 0$, as proved with Proposition 1.1), so the derivative is everywhere positive, as $\kappa_W < \kappa_N$. ∎

B
Further specifications for Chapter 2

As explained in the main text, both main empirical strategies in this paper, namely the use of spatial data models and tract-year dummies, are unfeasible using the entire dataset I have. Therefore, in the main paper I have restricted attention to Ohio, which is a populous and densely populated swing state of United States. In this Appendix I show evidence that Ohio indeed represents the entire United States, by showing results that are consistent with the ones found in the main paper. More specifically, here I employ the same specification as in the main text, but with data for the entire United States, using county-level time-varying dummies (as opposed to tract-level, which are finer).

It is important to note that since precinct-level election data was not provided by all states, I have only 31 states in our sample for 2008 (116,214 precincts). Furthermore, since precincts change over time, I were only able to adequately match precinct-level data for 18 states for 2004 and 2008 elections (56,781), and for 10 states for 2004 through 2012 elections (30,713 precincts).

I use this data to build three balanced databases. I first call D1 the database containing all precincts for 2008. I then build a balanced panel with the precincts for which both 2004 and 2008 data are available, and call it D2. Finally, I do the same for the precincts for which data for all three elections is available (D3). Moreover, as our interpretation of neighboring precincts as neighbors suggests that the effect should be present when precincts are smaller, I restrict the sample for urban areas, as defined by the census, in databases D1 and D2. Since they have less overall observations, I do not do the same for D3, nor for the regressions in the main text (using Ohio).

As it is clear in Table B.1, the results for our entire dataset are qualitatively similar to what we find for Ohio (and quantitatively stronger). This substantiates our choice of Ohio as a representative state of the entire United States, which is consistent with their place as swing state and bellwether of United States Presidential Elections. It also shows that our results have external validity, as they are robust to the choice of particular subset of our entire data we use in our empirical analysis.

Tabela B.1: Neighbors' Income and Votes for Democrats (United States, 2004-2012).

	(1)	(2)	(3)
High income	-0.0385	-0.0240	-0.0393***
	(0.0254)	(0.0381)	(0.00950)
Low income	0.0232	0.0433*	0.0713***
	(0.0280)	(0.0220)	(0.0138)
NB high income	-0.0922***	-0.00970	-0.0586**
	(0.0237)	(0.00757)	(0.0239)
NB low income	0.0637**	-0.0139	0.0406
	(0.0238)	(0.0196)	(0.0255)
Avg. Income (norm.)	-0.0414***	-0.0419***	-0.0304**
	(0.00671)	(0.00615)	(0.0101)
Avg. Inc. * NB high	0.0800***	0.0525***	0.0475**
	(0.0138)	(0.00833)	(0.0193)
Avg. Inc. * NB low	0.0644***	0.0840***	0.0472
	(0.0208)	(0.0156)	(0.0297)
Years	2008	2004-2008	2004-2012
States	32	18	10
Urban Areas	Y	Y	N
R^2	0.789	0.705	0.714
N	59,996	53,043	69,795

Dependent variable is the share of votes over the precinct population. Errors are clustered at the state level. The symbol * denotes p-value smaller than 0.1, ** for 0.05 and *** for 0.01. Controls include sex, age, race, proportion of foreign-born, education, population density, sector of activity, and neighboring precincts' race, education, and population density.