

The London School of Economics and Political Science

Social Capital, Human Capital, and Labour Market Outcomes.

Timothée Carayol

*A thesis submitted to the Department of Economics of the London School of
Economics for the degree of Doctor of Philosophy, London, September 2011*

Declaration

I certify that the thesis I have presented for examination for the PhD degree in Economics of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Chapter 3 of this thesis is the outcome of joint work with Pauline Givord.

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without the prior written consent of the author.

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Abstract

This thesis aims to document several aspects pertaining to the dynamics of human capital, both from a theoretical and an empirical viewpoint.

Chapter 2 studies how informational flows arising from social connections can affect careers and promotions. It aims to achieve identification of this causal pathway by focusing on the careers of bishops in the Catholic church. The range of the data, both in time and in space, makes it possible to infer some types of social connections between bishops (based on geography and careers), which in turn allows for the identification of their effect on careers. I find that being connected to the relevant bishops has a positive and significant effect on the likelihood of promotion to a diocese.

Chapter 3 investigates the transmission of human capital from one generation to the next. While the correlation of parents' educational achievement with that of their children is strong and well documented, there is a scarcity of consensual evidence that this relationship has a causal nature. We use a French reform that increased the duration of compulsory schooling by two years as a natural experiment, providing exogenous variation in parental years of schooling, and study its effect on the children of the affected individuals. We find evidence of a strong effect of paternal education on the educational achievement of children.

Research on employer learning has concentrated on contexts where there is uncertainty only on either the general or the match-specific human capital of the worker.

Chapter 4 develops a model where general and specific human capital coexist, and the uncertainty is on their respective shares in total productivity. The model generates predictions on a number of dimensions, *e.g.* declining worker mobility with experience and increase in wage variance over the lifetime.

Acknowledgments

I wish to thank primarily my doctoral supervisors Alan Manning and Barbara Petrongolo for their availability and constant support during the course of this PhD. This work also benefitted, at various stages, from the advice and comments of Guy Michaels and Steve Pischke, both of whom have consistently provided insightful perspectives and expertise over the years.

I also thank collectively the participants of the CEP (Centre for Economic Performance) labour work-in-progress seminar, as well as the participants of the CEP weekly labour workshops.

Some financial support was provided by an ESRC postgraduate fellowship.

In Summer 2008, I was given the opportunity to work on what would become Chapter 2 of this thesis at IZA (Institute for the Study of Labour) in Bonn, Germany, funded by a Marie Curie Fellowship for Transfer of Knowledge. My thanks go collectively to the IZA researchers for many helpful discussions; more specifically I wish to mention Ana Cardoso, Steffen Altmann, Daniel Hammermesh, Zhaopeng Qu and Jan Stuhler.

The research in chapter 2 also benefitted from the work and patience of two individuals. First and foremost, I am very grateful to Bishop John Arnold, of Westminster, for receiving me and accepting to answer my questions on the Catholic Church and some of the intricacies of bishop careers. Second, I also thank David Cheney for making some of the data on bishop careers from the *Annuario Pontificio*

more readily available on his website *www.catholic-hierarchy.org*.

The research reported in chapter 3 was done in collaboration with Pauline Givord. I am grateful for her time and effort on this joint project, as well as her expertise on the French *Enquête Emploi*.

Contents

1	Introduction	8
2	Social Connections and Promotions	12
2.1	Introduction	12
2.2	Microeconomic model of connections and promotions	16
2.3	Institution and Data	21
2.3.1	The Institutions	21
2.3.2	The Data	28
2.4	Model, estimation and results	30
2.4.1	Model	30
2.4.2	Estimation method	35
2.4.3	Results	37
2.5	Conclusion	49
3	Transmission of Human Capital	51
3.1	Introduction	51
3.2	The Berthoin reform	55
3.3	Data and Descriptive Analysis	60
3.3.1	Data	60
3.3.2	Characteristics of Parents and impact of the reform	61

<i>CONTENTS</i>	7
3.3.3 Education of children	62
3.4 Methodology and Results	64
3.4.1 Identification strategy : RDD estimation	64
3.4.2 Technical choices	68
3.4.3 Two-stage estimation	69
3.4.4 Wald estimation	71
3.5 Conclusion	74
4 Employer learning	76
4.1 Introduction	76
4.2 The three-period model	81
4.2.1 Informational structure and worker mobility	81
4.2.2 Equilibrium wages	87
4.3 Infinite-period model	92
4.3.1 Model	92
4.3.2 Wages in the infinite model	94
4.3.3 Numerical analysis	94
4.3.4 Additive productivity shocks	98
4.4 Discussion and conclusion	99
Appendix to chapter 2	107
Canons relative to the appointment of bishops	108
Appendix to chapter 3	111
The data	112

Chapter 1

Introduction

The last sixty years have witnessed the emergence of the very powerful and ubiquitous concept of human capital (see, for example, Becker, 1962) in economic research. The closest to an exhaustive definition may be as follows: human capital is the set of characteristics specific to a person and which may affect the ability of this person to create value. This includes cognitive skills, such as literacy, mental calculus, general knowledge, as well as non-cognitive skills such as patience, ambition, moral values; it also includes health, both physical and mental.

Another concept, very related to that of human capital to the point of sometimes overlapping, is that of social capital. Loosely defined, social capital describes the value of all the social connections of an individual or group of individuals. This includes, for example, friendships, professional relationships, and family.

This thesis explores the interplay between social capital and human capital, and how these impact an individual's labour market outcomes.

The next chapter, chapter 2, investigates the impact of social connections on the likelihood of future promotions. As social capital in an organization is likely to be correlated with human capital (if not in terms of number of connections, at least in terms of intensity of connections to the most influential persons), isolating

this causal effect requires a cautious identification strategy. Chapter 3 examines intergenerational transmission of human capital, that is, how human capital (more specifically, educational achievement) is passed on from parents to children. Again, correlations are likely to be misleading, as many factors are bound to affect both parental education and child education. Finally, chapter 4 examines how the human capital of an employee, initially assumed to be difficult to observe, becomes progressively known both to the employer and to the employee. The consequences of this learning process in terms of worker turnover and wage dynamics are explored.

Anecdotal evidence abounds to indicate that social connections can play a crucial role in an individual's labour market outcomes. An individual connected to influential decision makers within an organization will likely be treated differently from other candidates to a given vacancy. Such occurrences are seen, in some cases, as beneficial to an organization (hence the existence, in many firms, of sizeable incentives to refer a social connection as a potential employee); in other cases, they are perceived as harmful and are a source of criticism. A famous example of the latter is the tendency that popes once had to offer positions of responsibility to their nephews (sometimes as young as 14), with arguably little regard to whether such appointments were in the interest of the Catholic Church. Identifying the causal effect of social connections on the likelihood of being offered a position, however, is not straightforward. Firstly, social connections are difficult to observe or quantify; secondly, the sources of exogenous variation in social connections that would aid identifications are rare. Chapter 2 attempts to achieve this by focusing on the careers of bishops. Particular institutional features, as well as the availability of exhaustive data, make it possible to construct an identification strategy: distances between dioceses provide a way to infer certain types of connections, while death and retirement of bishops provide a source of arguably exogenous variations. It is found that a connected bishop is up to 40% more likely to be promoted to a larger diocese than an otherwise identical

unconnected bishop.

In France, cohorts born before 1953 faced compulsory schooling laws that required them to study until age 14. For later cohorts, schooling was made compulsory until age 16 as a result of a reform announced in 1959 and known as the Berthoin reform. Compliance to this reform was remarkable; the data shows a clear jump in the proportion of a cohort who studied until age 16 or older. For compliers, this reform is an exogenous source of variation in the number of years of schooling achieved. We use this as a natural experiment in chapter 3, enabling us to study the educational achievement of the children of the cohorts impacted by the reform. This exercise allows us to identify, for these populations, the causal impact of an individual's education on the education of his or her children.

Some forms of human capital are general, in that they may affect the ability to create value in a wide variety of circumstances. The ability to read, for example, is useful in any employment relationship in developed countries, and required in most. Other forms of human capital are much more specific: in-depth knowledge of atomic structure and nuclear reactions, while an absolute prerequisite for a career as nuclear physicist, would be useless for a professional tennis player. This distinction is almost as old as the concept of human capital (see Becker, 1962), yet the question of how these two quantities interact and that of their relative importance in employment relationships has proved difficult to answer fully; in fact, much of the literature has been forced to make strong simplifications, often assuming that all human capital is general. This duality of human capital is explored, from a theoretical point of view, in chapter 4 of this thesis. More specifically, the core assumption of the model developed in this chapter is that, while the production of a worker is easy to observe, it is uncertain how much of this production is due to match-specific factors, and how much would in fact be transferable to other employment relationships. This chapter contributes to the literature on employer learning, the core tenet of which is that

employers learn only progressively about the productivity of each employee. This literature typically assumes that all human capital is either match-specific (as in the seminal paper, Jovanovic, 1979) or general (for example, Lange, 2007). While these assumptions may make sense in some contexts, in reality, human capital is likely to be partly specific (e.g. very specific expertise) and partly transferable to other matches (e.g. ability to read, interpersonal skills).

Chapter 2

Stairway to Heaven: Social Connections and Promotions in the Catholic Church

2.1 Introduction

The focus of much of empirical personnel economics is on profit-maximizing firms in the private sector. An increasing number of studies document careers within a firm, i.e. the dynamics of wages and promotions. See, for example, Lazear (1999), Baker et al. (1994). These studies are mostly descriptive, but do not explain the mechanisms through which, for example, a candidate for a promotion is chosen over another. Belzil and Bognanno (2008) study the promotion dynamics of American executives across hundreds of firms; the panel structure of their data allow them to delve deeper into the dynamics of promotions and careers than earlier research was able to. However, there is yet a lot to learn about the determinants of promotions. In face of their importance, both for firms (as they define their hierachy), and for workers (as they are key to their career and act as a key incentive), further work to

understand promotions seems necessary.

Another literature concerns itself with the interplay between social connections and labour market outcomes. The outcomes considered are usually employment (Bramoullé and Saint-Paul (2004), Cingano and Rosolia (2006)..) or pay (Montgomery (1991)). The effect of social connections on promotions has been overlooked so far, probably due to a lack of readily available data on social connections. This restricts the economic research on social connections to theoretical studies or very specific empirical studies. Indeed, social connections are rarely observed or measured. Even proxies are hard to come by (e.g. numbers of email exchanges in Marmaros and Sacerdote (2006)). To my knowledge, no study so far has attempted to identify the role of social connections on promotions. This paper aims to pioneer this research avenue, using data on the careers of Catholic bishops.

The reasons that make the hierarchy of the Catholic Church lend itself ideally to this exercise are multiple. (1) It is a uniquely large organization, with a homogenous hierarchy over a large part of the world and over an extended period of time. (2) It is a quintessential internal labour market. Individuals are enrolled in this market at a young age, and normally only leave it upon death or retirement. It is therefore possible to observe entire careers with great accuracy. (3) Its local institutions (dioceses) form a geographical web over each country where the Church is present. (4) The local administrators (bishops) are required to meet regularly with their geographical neighbours, owing to institutional features detailed in the next pages. Identifying some connections based of geographical criteria is therefore possible. Because successful identification relies on each of these four features, a similar exercise would be much more perilous in most other organizations.

The (unobserved) actual social connections¹ of bishops are likely to be endoge-

¹Meaning the full knowledge of how well each bishop knows each of his peers, and how highly he thinks of each of them.

nous. Indeed, the unobservable skills that are valued in this labour market may be correlated with social skills, which in turn enable the formation of strong social connections with influential peers. This means that bishops with powerful friends are likely to also be highly skilled and/or ambitious, both of which are unobserved and may cause better career outcomes. However, focusing on the arguably exogenous, purely geographical proxies for social connections that I use in this study addresses this concern at least partially. I am able to back this approach with specification tests involving placebo connections with recently dead bishops.

There are two ways in which one could interpret an effect of social connections on promotions. One way is through signalling under imperfect information. The ability of workers is hard to observe in most labour markets, especially in services. Social connections make it possible to observe characteristics of the worker, some of which may be correlated with productivity and performance. Therefore, a socially connected worker may be more or less likely to be promoted than a less connected colleague, depending on the assessment that his connections have made of his skills. (A very able worker will benefit from social connections, whereas a less able worker would perhaps be better off with a low profile.) If signalling about ability is the only medium through which social connections affect careers, the aggregate effect of social connections can be expected to be positive; this is especially the case if the organization is risk averse, i.e. if a worker known to have an average productivity is preferred to a worker whose productivity is entirely unknown.

The other way is through favouritism. Some social connections result in friendships, and decision makers may prefer to promote a friend rather than someone they have never heard of, even if their expected suitability for the job is similar. A comparable behaviour has been documented in Bandiera et al. (2009), where it is found that managers will tend to favour their friends regardless of ability as long as they

are paid a fixed wage². This phenomenon has a long history in the catholic church, where in past centuries it was usual for popes to incardinate their nephews (hence the name “nepotism”).

Of course, these mechanisms are not mutually exclusive. This study does not attempt to distinguish between them. It is a meaningful distinction to make, however, and a tantalizing subject for further research. The first mechanism is a healthy feature of a meritocracy, as it helps the flow of information within the organization and benefits the hierarchy. The second mechanism, on the contrary, introduces frictions in the process of allocating skills to the jobs that use them best. However, in the data used in this study, both mechanisms are undistinguishable³ and the observed effect is a combination of both phenomena.

My findings show that connections do matter in this labour market. Candidates to a vacant diocese are more likely to be appointed if they have a connection with one or more of the bishops in the province of the vacancy. More precisely, connected bishops with otherwise average characteristics are about 40% more likely to be appointed than otherwise identical, non-connected bishops. I present evidence that this effect is causal, in the form of placebo specifications where past connections with bishops that are now retired or dead are shown to have no significant effect⁴ on the likelihood of promotion.

The rest of this paper is organized as follows. The next section, section 2.2, proposes a simple microeconomic model that illustrates the possible role of connections in promotions. Section 2.3 describes the institution and the data. Section 2.4 describes the estimation strategy and the results. Section 2.5 concludes.

²Under performance pay, however, managers tend to favour high-ability workers regardless of connections.

³We would require, at least, information on worker ability to make the distinction.

⁴With standard errors small enough to rule out an effect of the magnitude found for live bishops.

2.2 Microeconomic model of connections and promotions

This section describes a simple theoretical model on the interplay between connections and promotions.

A vacancy must be filled, for which there are $N + 1$ candidates. For simplicity, only one candidate is connected to the decision maker. Let $i \in \{1, \dots, N\}$ index a candidate with no connection. Let $N + 1$ denote the connected individual.

Regardless of connectedness, every individual has an (i.i.d.) ability parameter $a_i = u_i + \gamma d_i$, where u_i follows a uniform distribution between 0 and 1, and d_i follows a Bernoulli distribution with parameter p . u_i and d_i are independent. γ is a parameter between 0 and 1. This (to my knowledge) unusual distribution of ability is attractive for two reasons. First, depending on the values of the parameters, it can be made to look like a uniform (low γ) or like a thin-tailed distribution (γ around .5). Furthermore, the skewness of the distribution is related to p . A high p will cause a thin lower tail, and a thick upper tail of the ability distribution. Second, these distributions are general enough to carry interesting results, yet simple enough to obtain very tractable formulae.

For the connected candidate, a_i is fully observable; for all others, only the u_i component can be observed, and is only observed if the decision maker decides not to pick his connection⁵. Furthermore, the decision maker receives a personal premium c when he appoints the connected candidate; however, this premium does not translate into actual production. The actual production if candidate i is appointed is simply assumed to be given by a_i . I assume further that the connected candidate is high type⁶, i.e. has $d_{N+1} = 1$ ⁷

⁵This reflects the fact that the decision maker may well not even consider other candidates if his connection is strong enough.

⁶In this section, for convenience, I call “high type” a candidate with $d_i = 1$, and low type a candidate with $d_i = 0$.

⁷If that were not the case the analysis would be less interesting, as the connected would be chosen mostly on the basis of the personal premium c , and connections would be unambiguously

The decision maker will pick the candidate that maximizes $V_i = E(a_i) + c_i$, where $c_i = c$ for the connected candidate and $c_i = 0$ for the others. Hence the value of the candidate to the decision maker will be $V_{N+1} = u_{N+1} + \gamma + c$ for the connected candidate, and $V_i = u_i + \gamma p$ for the non connected. The choice of choosing someone connected or not therefore boils down to the choice between the connected candidate and the non-connected with the highest u_i . Conditional on not observing the individual u_i 's, the decision maker rationally believes the distribution of the highest u_i to be a $\beta(N, 1)$ distribution⁸, with expected value $\frac{N}{N+1}$. This means that the connected candidate will be chosen when:

$$u_{N+1} > \frac{N}{N+1} - \gamma(1-p) - c$$

Ex ante, the probability that the connected candidate be chosen is therefore $\text{Min}\left(1, 1 - \frac{N}{N+1} + \gamma(1-p) + c\right)$. This is increasing in c (the personal value of friendship), decreasing in p and increasing in γ (reflecting the value of the component of the ability d_i , only observable for the connected), and decreasing in N (reflecting the fact that as N increases, the expected quality of the best non-connected increases).

The parameter c introduces a conflict between the personal utility of the decision maker and the social optimum. The assumptions above make it easy to calculate the expected social gain (or cost), denoted ΔW , of choosing the connected candidate rather than the best non-connected. $E(u_{N+1} | \text{connected is chosen}) = \frac{1}{2} \text{Max}\left(1, 1 + \frac{N}{N+1} - \gamma(1-p) - c\right)$, so:

$$E(\Delta W) = \gamma(1-p) - \frac{N}{N+1} + \frac{1}{2} \text{Max}\left(1, 1 + \frac{N}{N+1} - \gamma(1-p) - c\right)$$

socially harmful.

⁸This is a well documented property of the order statistics of a uniform between 0 and 1.

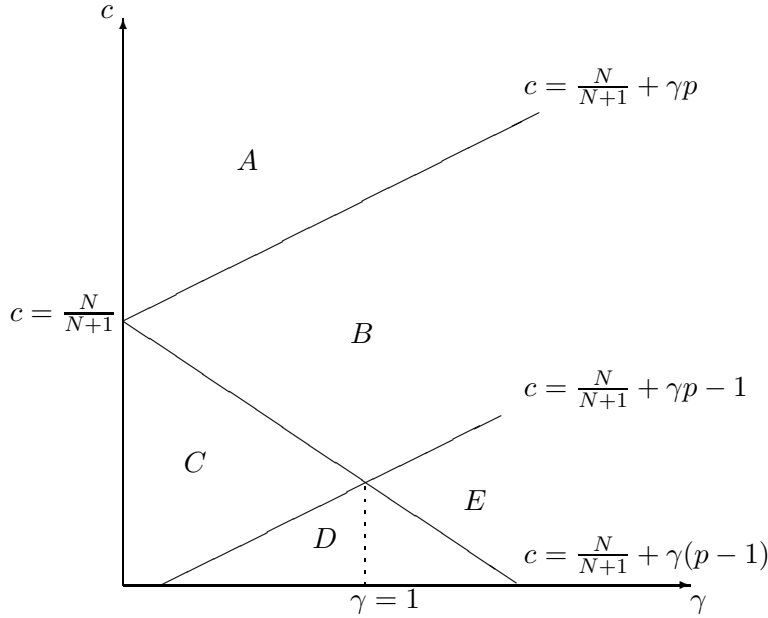
Hence, when the connected candidate is chosen rather than the best non connected, the organization may either gain (low p , high γ , low N , low c) or lose. Note that a low p and a high γ correspond to a situation where the unobserved component of ability is very important, in which case choosing a connected candidate who is known to have $d_i = 1$ is likely preferable. A low N corresponds to a situation where the pool of candidates is small, making it less likely to contain an outstanding non-connected candidate. Finally, where c is 0, the decision maker faces incentives that are well aligned with those of the organization, and therefore the availability of a connected candidate is beneficial. With $c = 0$, $E(\Delta W) = \frac{1}{2} \left(1 + \gamma(1 - p) - \frac{N}{N+1} \right) > 0$.

I now drop the assumption according to which the connected candidate has $d_{N+1} = 1$, and instead assume that d_{N+1} also follow a Bernoulli distribution with probability p . I want to compute and compare the probability that any given candidate be chosen for the vacancy, from the point of view of an external observer who does not know the realization of d_{N+1} . Indeed, this is close to the setting of the empirical part of this paper, where I compare the likelihood of promotion of connected and non-connected candidates without observing the ability of any candidate.

$$\begin{aligned}
& P(\text{connected is chosen}) \\
&= P(d_{N+1} = 1)P(\text{connected is chosen} \mid d_{N+1} = 1) \\
&\quad + P(d_{N+1} = 0)P(\text{connected is chosen} \mid d_{N+1} = 0) \\
&= \begin{cases} 1 & \text{if } (\gamma, c, N, p) \in A \\ p + (1 - p) \left(\frac{1}{N+1} - \gamma p + c \right) & \text{if } (\gamma, c, N, p) \in B \\ \frac{1}{N+1} + c & \text{if } (\gamma, c, N, p) \in C \\ p \left(\frac{1}{N+1} + \gamma(1 - p) + c \right) & \text{if } (\gamma, c, N, p) \in D \\ p & \text{if } (\gamma, c, N, p) \in E \end{cases}
\end{aligned}$$

In the equations above, regions A , B , C , D and E are regions of the (γ, c, N, p)

space defined as below:



In region A , c is high enough that the connected candidate will always be chosen regardless of ability. In region B , he will be chosen regardless of u_{N+1} if $d_{N+1} = 1$ (high type); however if $d_{N+1} = 0$ (low type) the connected candidate will only be chosen if u_{N+1} is sufficiently high. In region C , the choice will depend on u_{N+1} for either value of d_{n+1} . In regions D and E , c is low enough that the connected candidate will never be chosen if $d_{N+1} = 0$; however if $d_{N+1} = 1$, the connected candidate will be chosen for sure (in E) or for sufficiently high values of u_{N+1} (in D).

As long as c is high enough that we are in regions A , B , or C , the likelihood of choosing the connected candidate is higher than $\frac{1}{N+1}$, meaning that he is more likely to be appointed than the average candidate. Note also that the existence of region D depends on whether the intersection point between regions B , C , D , and E is for a valid value of c (i.e. $c \geq 0$). That is the case if and only if $p \geq \frac{1}{N+1}$. In that case, it is also the case that the connected candidate has a greater chance of being

appointed than average in region E . Only in region D is it possibly the case that the connected candidate is less likely than average to be chosen. That will be true if γ is sufficiently low, creating a situation where the informational advantage of the connected candidate is unlikely to play a large role even if he is known to be high type (d_{N+1}). Hence in that situation, either the connected candidate is low type (in which case he won't be chosen, by definition of region D); or he is high type, and will compete mostly on the basis of his u_{N+1} (because γ is low), which is likely to be much lower than the highest u_i in the pool of non-connected candidates.

The previous discussion serves to show that in most cases the connected candidate should be expected to be more likely to be picked than the average non-connected candidate, even in a situation where there is very little favouritism. The informational advantage that he enjoys is sufficient to give him an edge.

A possible extension for this model is to consider a risk-averse organization⁹. Risk-aversion increases the appeal of the connected candidate, and reduces the appeal of the non-connected.

This simple model therefore carries the following insights. (1) Where performance is easy to observe (low γ), an influence of connections on promotions is socially harmful. (2) Connections are more likely to be beneficial for risk-averse organizations than for risk-neutral (or risk-loving) organizations. (3) Whether connections are harmful or beneficial depends largely on their personal value c to the people in a position to make promotion decisions. (4) The connected candidate almost always has a higher probability of being picked. This is true even with no favouritism ($c = 0$). The presence of favouritism further increases this probability.

⁹This is simple to do by using a concave objective function and using the expected utility framework; but the algebra becomes unwieldy.

2.3 Institution and Data

2.3.1 The Institutions¹⁰

The Catholic Church is a global religious organization headed by the Pope from Vatican City, in Rome. The global decisions, e.g. on the doctrine of the Church, are centralized in Rome; but most other decisions are taken at a local level. The local decision-makers are bishops, generally heading a territory known as diocese.

These bishops in charge of a diocese are the focus of this study; however, other individuals hold the title of bishop. First, retired bishops keep their title and are referred to as Bishop Emeritus. They are entitled to a residence in the diocese which they formerly headed. Second, a number of bishops (mostly cardinals) have their career in Vatican City, helping the decision making of the Roman Curia. Thirdly, the Apostolic Nuncios, diplomats for the Vatican in countries with a large catholic community, have the title of bishop, even though their careers are mostly disjoint from that of diocesan bishops (with very little mobility between those two career paths, as the skill and formation required are considered different).

We focus here on the appointment of bishops involved in the management of a diocese. Again, within this category are several titles. First, *ordinary* bishops, the most common, are the individuals at the head of a diocese, meaning that every diocese has exactly one ordinary bishop¹¹. Second, *auxiliary* bishops are appointed in many dioceses to assist the ordinary bishop. This is mostly common in larger dioceses (see table 2.1 for the distribution of the number of auxiliary bishops across dioceses). Third, *coadjutor* bishops can be appointed. The role of a coadjutor bishop is to act as auxiliary bishop until the ordinary bishop retires or die, upon which

¹⁰The sources for this section are Canon Law (1983), Reese (1989) and Reese (1996), as well as a conversation with John Arnold, Auxiliary Bishop of the Diocese of Westminster in London. Relevant canons are quoted in the Appendix.

¹¹Except in the transitory period where a bishop died, retired, or left the diocese. This situation is denoted as *sede vacante*, meaning “vacant see”.

Table 2.1: Number of auxiliary bishops per diocese in 1983: frequency table

Item	Number	Per cent
0	1,867	82.7
1	271	12.0
2	64	2.8
3	27	1.2
4	8	0.4
5	12	0.5
6	5	0.2
7	1	0.0
8	1	0.0
9	1	0.0
Total	2,257	100.0

circumstance the coadjutor will automatically take his succession. In this paper, I treat the appointment as a coadjutor as the appointment to ordinary (since *de facto*, most coadjutors succeed within a year or two of their appointment, meaning that they are deemed ready to be an ordinary). This institution is meant to assist ill or fatigued ordinaries and to ensure a smooth transition.

Every country with a strong Catholic presence is partitioned into dioceses. The size and number of dioceses varies a lot across countries; typically, countries with a long Catholic history and a large Catholic community today have a dense diocese web (e.g. Italy, France); countries with a smaller Catholic community, or a shorter history of catholicism, have a sparser web (United States, Germany..). Diocese boundaries sometimes follow the boundaries of other subdivisions (either current or historical), e.g. regions, states, etc. Dioceses with particular significance, whether because of their larger size or because of a remarkable history, are called archdioceses and headed by an archbishop. This prefix is purely honorific; there are ordinary, auxiliary, and coadjutor archbishops. It is associated with the diocese more than with the individual, even though in a small number of cases, the Vatican can choose to honour

the bishop of a regular diocese with the title of archbishop. Such examples, however, are too rare to be meaningful for this study.

Dioceses are largely independent from each other, and ordinary bishops have, *de facto*, absolute power over their diocese. This includes the finances, catholic institutions (schools, hospitals), personnel management, etc. A bishop is a high level manager, responsible for a workforce often in excess of one thousand priests and members of religious institutes, and for assets of sometimes considerable value. A bishop's compensation is a fixed stipend set at the diocesan level, and appears to be of the same order of magnitude (though higher) than a priest's with similar tenure¹². There is no performance pay, and virtually no chance of being demoted. Work incentives are therefore intrinsic motivation and promotion prospects. A bishop is required by Canon Law to reside in the diocese, and to offer to retire at 75¹³.

Within countries, dioceses are grouped into ecclesiastical provinces. These are clusters of variable size, usually 3 to 6 dioceses (see table 2.2). Each ecclesiastical province has one main diocese, often an archdiocese, called the metropolitan diocese; the others are called suffragan dioceses. This is mostly a historical feature, with little relevance today; in particular, the metropolitan bishop has no meaningful formal authority over his suffragans¹⁴. However, ecclesiastical provinces still play a role where bishop appointments are concerned, as will be explained below.

A bishop vacancy opens in three situations: when a bishop moves, retires or dies; when an ordinary bishop wants an additional auxiliary (which he can only do with a permission from Rome); or, more rarely, when a new diocese is created.

In each of these situations, a process starts which aims to find a priest or an existing bishop who would fit the vacancy. The process is slightly different depending

¹²Accurate information on bishop stipends is hard to obtain. Most dioceses do not make their financial reports available publicly, and those that do have only done so for a few years.

¹³This retirement age was introduced after the Second Vatican Council, in 1966. Before that, bishops were free to work as long as their health allowed.

¹⁴See Canon 436 of the Code of Canon Law for details.

Table 2.2: Number of dioceses per ecclesiastical province: frequency table

Item	Number	Per cent
1	25	4.5
2	25	4.5
3	118	21.2
4	108	19.4
5	109	19.6
6	58	10.4
7	43	7.7
8	31	5.6
9	15	2.7
10	11	2.0
11	3	0.5
12	5	0.9
13	3	0.5
14	1	0.2
18	1	0.2
20	1	0.2
Total	557.0	100.0

on whether the vacancy is for an ordinary or an auxiliary bishop. In the case of an auxiliary bishop, the ordinary bishop of that diocese (i.e. the hierarchical superior of the future appointee) is to choose a list of three candidates, called the *terna*, that he deems able to fulfill the role, which he then sends to the Nuncio. Usually the Nuncio will send those same three names directly to the Vatican for the final choice, though he is formally allowed to amend this list if he finds it necessary.

In the case of an ordinary bishop, the person who defines the *terna* is the apostolic Nuncio to the country of the vacancy. To do this, he has to contact a number of agents, and in particular the bishops in the relevant ecclesiastical province. They can provide him with information about promising priests and bishops whom they consider suitable for the job. He will also approach priests who work with the candidates. His role is to narrow down the list of candidates to three names (the *terna*) and to provide, for each of those three names, an extensive file containing interviews, references, and any relevant information. He then sends this file to the Vatican, where the final choice is to be made.

Once the *terna* reaches the Vatican, it is discussed by the Congregation for Bishops (the ministry of the Vatican in charge of bishop appointments, staffed primarily with Italian cardinals), which makes a recommendation for the Pope. The Pope has the last word. If none of the names in the *terna* is deemed satisfactory, the Vatican contacts the Nuncio (or ordinary bishop, in the case of the appointment of an auxiliary) and asks for three new names to be sent.

The whole process is explained in much more details, in Reese (1989). Reese only describes the situation in the US, but the process is fairly universal, at least *de jure*.

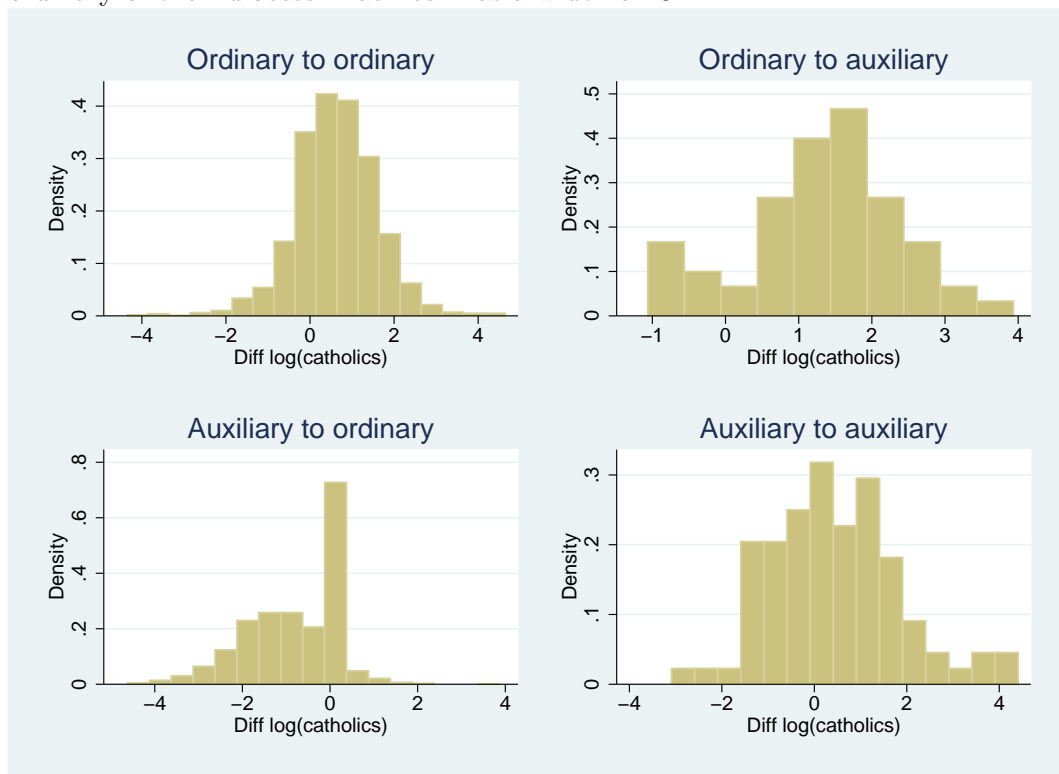
I observe, in my sample, a total of 13195 bishop appointments, which corresponds to all bishop appointments decided by the Congregation for bishops or the Congregation for the Evangelization of People in the Vatican. Excluded are the ones depending on the Congregation for the Oriental Churches, that is to say the appoint-

ments of bishops (or equivalent, e.g. patriarchs) in the Eastern Catholic Churches. For the sake of homogeneity of institutions, I focus on the Latin rite. Out of these 13195, 67% (8828) correspond to appointments of former priests; 34% correspond to promotions of existing bishops. Out of the 8828 appointments of new bishops, 64% (5630) become ordinary bishops; the rest become auxiliary bishops (25%) or coadjutor bishops (11%).

Identifying what constitutes a promotion in this labour market is not trivial. Several types of mobility can be observed, e.g. from auxiliary bishop of an archdiocese to ordinary bishop of a smaller diocese. I denote d_v a vacant diocese and d_b the former diocese of bishop b , and as $\Delta_{cat}(d_b, d_v)$ the growth in number of catholics in between d_b and d_v , i.e. $\Delta_{cat}(d_b, d_v) = \log(catholics_{d_v}) - \log(catholics_{d_b})$. Figure (2.1) gives the distribution of Δ_{cat} for a number of cases, namely (1) *OO*: an ordinary bishop appointed as ordinary bishop of another diocese (2236 cases, 17% of appointments); (2) *OA*: an ordinary bishop appointed as auxiliary bishop of another diocese (91 cases, .7% of appointments); (3) *AO*: an auxiliary bishop appointed as ordinary bishop of another (or the same) diocese (1929 cases, 15% of appointments); and (4) *AA*: an auxiliary bishop appointed as auxiliary bishop of another diocese (111 cases, 1% of appointments). As this graph shows, most *OO* and *OA* appointments correspond to a move to a bigger diocese; *AA* shows no clear pattern; *AO* usually corresponds to a move to a smaller diocese (but more responsibilities), with a mass point at 0 corresponding to auxiliaries promoted to ordinary of the same diocese.

In this study, I will treat as a promotion: (1) any appointment of type *AO*, and (2) any appointment of type *OO* corresponding to a move to a bigger diocese (in terms of number of catholics). In my analysis of promotions, I only focus on appointments as ordinary, because (as seen above) existing bishops are rarely appointed as auxiliary. Given their rarity, including such appointments would add little benefit, yet would complicate the analysis needlessly.

Figure 2.1: Diocese size differential upon bishop promotions. The graph “Auxiliary to Ordinary” has a mass point at 0, corresponding to bishops being appointed as ordinary of their diocese. Each bar has a width of .5.



2.3.2 The Data

The data for this study comes primarily from the *Annuario Pontificio*. These are books, released every year by the Vatican, giving a complete snapshot of the Catholic hierarchy throughout the world. They includes, for every diocese, the names, age and place of birth of the ordinary bishop and, where applicable, of the auxiliary bishops. Also indicated (but not used in this study) are the names of the vicars general, who are the most important priests in the diocese. Diocese characteristics are also given, including population, area, number of catholics, of baptisms, of churches... This data aggregates the answers to a questionnaire sent each year by the Vatican to every ordinary bishop. Although the annuarios have been released since the end of the 19th century, I focus on appointments from 1930 onwards.

A major limitation of this data is that individuals can only be observed after they become bishops. The only indications on their earlier life are the place and date of birth, as well as the date of ordination as a priest. Many priests grow up and work in their diocese of birth, and most in their province of birth, but that is not always the case. Of interest too would be the place of seminary.

Missing from the *Annuario Pontificio* are the geographical coordinates of the dioceses (necessary to compute the distance between a bishop and a vacancy, or between two bishops, which is crucial to the analysis). To obtain these, I downloaded data containing the coordinates (latitude and longitude) of all cities with a population higher than 10,000 from GeoNames, a free geographical database. I then matched the names of all dioceses to this database. The match quality is good overall, with only 7% dioceses with missing coordinates. A map of the dioceses included in the sample can be found as figure 2.2. The figure makes a number of things apparent. First, the geographical size of dioceses varies a lot across countries¹⁵, reflecting differences in

¹⁵Diocese boundaries are not visible, but the average geographical size has an inverse relationship with the spacing between the dots. For example, dioceses in North America cover a larger area on average than in Western Europe.

history and demography. Second, certain regions of the world are totally absent from my sample, e.g. Greece and Russia. This is either because catholicism is (virtually) absent in these regions (e.g. because the Orthodox Church dominates), or because catholicism follows rites other than the latin rite, and hence are outside the scope of this study.

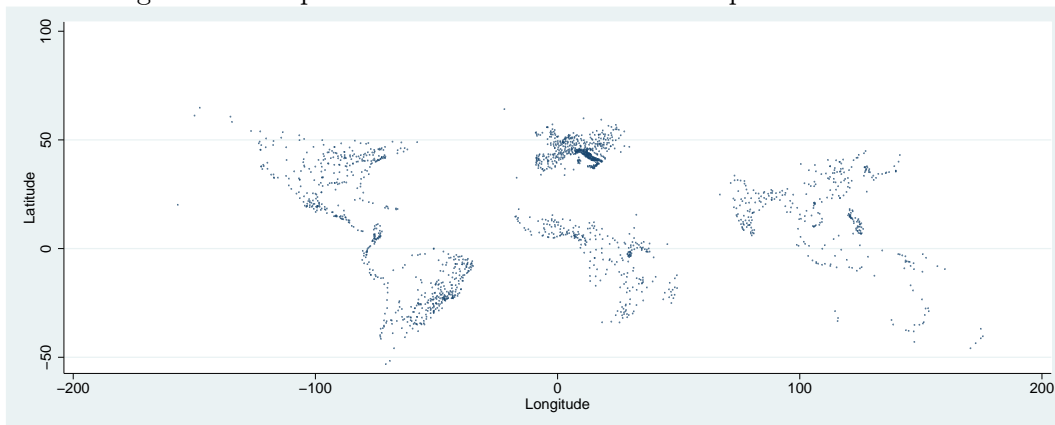
A number of variables come from different sources, and matching of those sources, to this date, is imperfect. The data on bishop promotions is exhaustive; but the data on diocese characteristics (e.g. number of catholics per diocese) comes from year 1983. Depending on how much they vary with time, it is not very clear how accurately these variables describe the situation in 1930 (the start of the sample period). Data on birthplaces is imperfectly matched¹⁶; however unmatched observations (i.e. bishops for which I do not have the diocese of birth¹⁷) are as-if-randomly picked, and hence should not affect my estimation results other than through a loss of efficiency or an understatement of the effect. Data on diocese coordinates is, as said above, well matched overall, although coordinates for a few hundred dioceses are missing. Again, I expect those to be close to randomly picked. Finally, even for 1983, a number of diocese characteristics have missing observations. In the *annuario pontificio*, variables are given as missing both when they are actually missing and when they are zero. I use in my estimations the variables that are the most rarely missing (and the least likely to be zero), namely for example the number of catholics and number of priests.

Catholic influence changed over the studied period, growing in some regions (e.g. Africa, South America) and declining in others (e.g. Western Europe in recent decades). A side effect of these dynamics is that diocese boundaries are not constant. Depending on the evolving needs and demography of each region, dioceses

¹⁶For all bishops, the *Annuario Pontificio* indicates the town and diocese of birth. The data I got from catholic-hierarchy.org, however, only indicates the town of birth, making matching with dioceses non-trivial. Using Optical Character Recognition methods on scans of the *Annuarios*, I was able to match town of birth to dioceses in about 60% of the cases.

¹⁷Meaning that some connections are missing for these bishops

Figure 2.2: Map of dioceses in 1983. Each dot represents a diocese.



may split or merge, or two dioceses may transfer a part of their territory from one to the other. As of now I have no really satisfactory way to incorporate these events in this study. Such changes are rare enough that they can be ignored with no major consequence on the estimates. Hence, I take 1983 as the year of reference for diocese boundaries; I effectively ignore deviations from the 1983 map of dioceses. But a more satisfactory (and tedious) way to proceed would be to systematically track these changes in boundaries.

Table 2.3 gives summary statistics for dioceses as of 1983. The large standard deviations reflect the heterogeneity of dioceses, with the smallest more than 3 orders of magnitude smaller than the biggest (as measured, for example, by number of catholics).

2.4 Model, estimation and results

2.4.1 Model

The unit of observation I use in my estimation is a vacancy. Each time a vacancy appears, a person is chosen to fill it among the population of priests and bishops of

Table 2.3: Diocese summary statistic for year 1983.

Variable	Mean	Std. Dev.	Min.	Max.	N
Area (sq. km)	39,646	(125,634)	1	2,550,000	2,176
Population	1,394,381	(2,839,558)	5,650	40,000,000	2,177
Catholics	357,584	(546,058)	1,000	8,600,000	2,190
Baptisms	8,119	(12,191)	2	165,470	2,075
Parishes	95.8	(143)	1	1,127	2,132
Churches	57.7	(106)	1	998	1,085
Priests	114	(198)	1	2,376	2,160
Seminarians	22.3	(32.4)	1	352	2,044
Auxiliary bishops	0.277	(0.787)	0	9	2,258
Vicars General	0.918	(0.619)	0	6	2,258
% of dioceses in Italy	0.118				
% of dioceses in Europe (other)	0.174				
% of dioceses in Africa	0.147				
% of dioceses in South America	0.183				
% of dioceses in North America	0.131				
% of dioceses in Oceania	0.028				
% of dioceses in Asia	0.175				

the considered country.

My model for the choice of a bishop to fill vacancy v is an Additive Random Utility Model (ARUM), i.e. the candidate providing the highest “utility” is chosen as bishop. The decision-making involves all the individuals who are going to provide the Vatican with information about the candidates. This information is primarily provided by the people in the ecclesiastical province where the vacancy arises, as they are the ones whom Canon law explicitly requires the Apostolic Nuncio to consult. The objective function of interest is that which the Pope attempts to maximize, given all the information brought to him by the local bishops, the Apostolic Nuncio, and the Congregation for Bishops. I will use simply the word “utility” in this section, in the name of consistency with the literature on ARUM.

I consider that the utility function has two components. One describes how good a fit the bishop is for this vacancy given characteristics directly observable by the Vatican, e.g. given his previous appointments, his publicly known achievements, etc. The other component reflects information given to the Vatican by the agents

it consults; this component will be affected by the social connections of the bishop. I assume that these components, as well as the error term, enter additively in the utility function. Neither of these two components can be perfectly observed by the econometrician; but some information at least is available for both.

I write the utility from having bishop b fill vacancy v in year y as:

$$U_{vby} = X_{vby}\beta + \gamma \sum_k^{K_{vby}} v_{kby} + \varepsilon_{vby} \quad (2.1)$$

where: X denotes a set of variables directly observable by the Vatican relative to the suitability of the match; k denotes the individuals who will be contacted by the Vatican if b is considered for v ; v_{kby} reflects the information that k provides to the Vatican about b in year y . (v_{kby} therefore incorporates both the favouritism effect and the information effect mentioned earlier.) ε_{vby} denotes a random component reflecting any element affecting the perceived quality of the match that is not captured by the other variables.

Upon receiving the reports of v_{kby} , it is likely the case that the Vatican is aware that personal relationships can get in the way of objective evaluation of a candidate's suitability for the vacancy. Nonetheless it is clearly not in the Vatican's interest to dismiss entirely reports from the individuals who are known to have interacted most closely with the candidates, as they are also the ones who know the most about them.

It is obvious that the Vatican has access to much more information about each candidate and vacancy than are publicly available or can realistically be obtained. Information about the needs of the diocese, the personality of the previous bishop and of the candidates, the finances of the diocese, and so on, cannot be realistically included as regressors. Also missing in my analysis is information about the pool of priests in the country, since I only have information on individuals who will eventually become bishops.

The proxies I will use for $\sum_k^{K_{vby}} v_{kby}$ are variables that give an indication of how likely bishop b is to be connected to people who will have an influence in the filling of vacancy v . Among those people are the Apostolic Nuncio, and the bishops in place in the province where the vacancy arises. I do not observe connections to the Nuncio¹⁸, and can only infer connections to existing bishops, in the manner described in the next paragraphs. For the rest of the paper, when I have to refer to two connected bishops, I will refer to the one considered for the vacancy as “candidate” and the one possibly helping the former as “incumbent”.

I allow for a connection to exist between two bishops if, at any point in the past, they have been working synchronously in dioceses that are geographically close (as defined, depending on the specification, either by the two dioceses belonging to the same ecclesiastical province, or by the distance between the two dioceses being less than a cutoff value). The connection (or likelihood of connection) is allowed to be decreasingly intense with distance, in space or (in one specification) in time¹⁹. As this is the variable of interest in this study, a few points about its construction are to be emphasized.

I include as candidates for a vacancy all the bishops in the country aged 35 to 70, and who are either auxiliary bishops, or ordinary bishops in a smaller diocese than the vacant one. The connections that I include on the right hand side are those with incumbent bishops in the province of the vacancy, meaning either auxiliary or ordinary bishops who are in a position to affect the process of choosing the next bishop.

I only keep connections that are terminated at the time of the vacancy, with one or the other of the bishops having moved some time before the vacancy arose.

¹⁸Nuncios are overwhelmingly Italians, and are very mobile across countries; therefore in most cases, they are unlikely to have strong connections with the bishops they appoint.

¹⁹That is to say, a connection that ended long ago (e.g. with one of the bishops changing appointments) is allowed to have less effect than a more recent connection

I exclude ongoing connections. Including them is problematic as the result of two considerations: first, they do not generate interesting variations. All the local candidates will have very similar ongoing connections (as the relevant connections are connections to bishops in the ecclesiastical province of the vacancy). Hence ongoing connections typically do not discriminate meaningfully between the likely candidates. Second, local candidates have ongoing connections with all the local bishops, but also differ from faraway candidates in other respects – including how well they know the region, the local priests, etc. These effects are already captured by controls such as distance between candidate and vacancy, but likely not perfectly. Excluding ongoing connections therefore seems preferable²⁰. In practice, I only retain connections that were formed while the incumbent was in a province other than that of the vacancy.

In some specifications, I only take into account connections formed after the first appointment as bishop (I use the shorthand “bishop connections” for these specifications). In others, I also count as connections people who were born in the vicinity of each other (using the shorthand “all connections”), making the strong assumption that people remain in their diocese of birth until they become appointed as bishop. It is generally not exactly accurate; however, a majority of bishops do remain in their province of birth during their priesthood. Given that dioceses within a province are close to each other, and given that my definition of connection is based on distance, mobility within a province is not a major concern.

Some of the more sophisticated specifications allow for a gradual attenuation of the effect of a connection with distance between the connected bishops – however, most specifications here rely on the existence of a cutoff. Distances below the cutoff are counted as allowing a connection; distances above the cutoff are not. I report results for cutoff values of 100 km and 200 km.

²⁰In fact, including them does not change the results substantially. The point estimates barely change; the standard errors are somewhat reduced.

In most specifications, I only retain connections with active incumbents; in others, as a specification test I allow for connections with retired or dead incumbents to have an effect.

I model v_{kby} as

$$v_{kby} = \alpha \exp(-Z_{kby}\gamma) \quad (2.2)$$

where Z can include, for example, a measure of distance between the dioceses when the connection was ongoing, and how long ago the connection was terminated. In most specifications, for simplicity, Z will in fact be zero (or infinity, if k and b are not connected), and all connections will be assumed to have the same effect.

2.4.2 Estimation method

Given the ARUM framework described above, the most natural method of estimation is the *GEV* (Generalized Extreme Value) family of estimators, which includes as special case the conditional logit specification which I use here. Using the vocabulary of multinomial models, a case here is a vacancy and an alternative is a bishop.

Compared to most applications of multinomial models, this setup has a number of features that must be noted. First, even though the decision making is a complex process involving several dozens of individuals, the Pope always has the last word. Over the period considered, the number of Popes is small (seven), and the institutions fairly stable. This lends credibility to the view that the objective function was relatively homogenous throughout, and that its parameters can be safely assumed to be constant. Second, the number of alternatives available for each vacancy is considerable, since any bishop or priest (or even, in certain rare cases, other members of the catholic community) can be chosen to fill the vacancy. This implies that all

the regressors are bound to be alternative-specific, as identification with case-specific regressors relies on each alternative being picked several times²¹. Third, the set of alternatives, or choice set, varies a lot across vacancies. Dead, unborn or foreign bishops cannot be chosen, which means that the set of alternatives varies with time and geography. More specifically, I retain, as my choice set, *{bishops living in the country of vacancy v and aged between 35 and 75}*.

The conditional logit specification is subject to the IIA (independence on irrelevant alternatives) property. IIA in this context means that the relative probability of choosing bishop b_1 over bishop b_2 does not depend on the characteristics or number of the other candidates. This is plausible, and testable. Performing tests based on the Hausman-McFadden approach as described in Hausman and McFadden (1984)²² indicates that the data fails to contradict the IIA, validating the conditional logit specification.

The main concern is that of endogeneity of social connections. If I had access to perfect information on the degree of connectedness between any pair of bishops in the data (including, for example, the belief of each bishop about the ability of each other bishop), and if I built a measure of connectedness of each bishop based on this perfect information, then I would arguably have a more accurate measure of connectedness; but one also, crucially, more crippled with endogeneity. Unobservables exist, such as interpersonal skills, that affect the connectedness of a bishop, and are at the same time valued in this labour market (hence will affect U_{vby} above and beyond the effect of connectedness). If unable to control for ability, the effect of connections would then be overestimated. However, given my coarser proxies for social connections, this is not as much of a concern here. My measure is only a function of the history of the

²¹This is the distinction between conditional logit, including alternative-specific regressors (as here), and multinomial logit, including case-specific regressors

²²This approach relies on the fact that if the IIA holds, then the estimates should be robust to dropping a random subsample of the alternatives.

bishop—where he was born, and where he was appointed in his early career. Since I also control for characteristics of the diocese of previous appointment, exogeneity of my measure of social connections is a more credible assumption than if I had a more accurate measure. Using the language of the instrumental variables literature, what I estimate is ultimately a reduced form. I do not observe the endogenous variable and hence cannot perform an IV estimation; but the reduced form itself is of interest²³.

2.4.3 Results

New or existing bishop?

For descriptive purposes, we focus first on the decision to appoint a new bishop rather than an existing bishop using a binary logit specification. Results are reported in table (2.4). Column (1) shows that in general a priest would less likely be chosen over a bishop for an ordinary vacancy than an auxiliary vacancy. But, as column (2) shows, this no longer holds if we control for the size of the diocese: only big dioceses have auxiliaries, and for small dioceses, one is more likely to appoint someone new as ordinary than an existing bishop. Finally, it seems that the tenure of the apostolic nuncio in the country considered is (slightly) positively correlated with a higher probability of appointing a priest—perhaps because the nuncio has better knowledge of the pool of able priests.

Conditional logit

From now on, all specifications are run on the subset of appointments such that: (1) the appointee is an existing bishop; and (2) the vacancy is for an ordinary bishops (because there are only a couple of hundred cases of an existing bishop being

²³It is arguably the case that the effect of the connections that I infer is not very different from the effect of the connections that I do not observe; i.e. that a bishop who was my neighbour during my early career has a similar effect on my career as a bishop whom I know equally well from different (unobservable) circumstances (e.g. seminary or bishop conference).

Table 2.4: Decision to appoint a new bishop or promote an existing bishop

	(1) Priest	(2) Priest	(3) Priest
<i>Ordinary</i>	-2.462*** (0.0748)	3.678*** (0.959)	4.266*** (1.063)
<i>Catholics</i>		-0.0353 (0.0680)	-0.00737 (0.0749)
<i>Ordinary * Catholics</i>		-0.517*** (0.0711)	-0.550*** (0.0786)
<i>Tenure of nuncio (years)</i>			0.0160*** (0.00553)
Constant	2.754*** (0.0720)	3.315*** (0.924)	2.695*** (1.021)
Observations	13322	11741	8882

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Dependent variable = 1 if chosen candidate is a priest; 0 if already a bishop.

appointed as auxiliary: most auxiliary appointments go to priests).

In these, and in all subsequent conditional logit specifications, reported standard errors are based on the observed information matrix.

Table (2.6) gives results for different sets of controls, excluding any measure of connections (though the regressors “distance” and “Same Province” may be thought of as capturing some effect of connections to local bishops.)

The variable “Density” refers to the inverse of the average of the square root of the areas of dioceses in the country considered.

$$Density = \frac{\#dioceses\ in\ the\ country}{\sum_{d \in dioceses\ in\ the\ country} \sqrt{Area_d}} \quad (2.3)$$

For example, if dioceses in country A have four times the area of dioceses in country B, the density of country A will be half of that in country B.

This variable is meant to account for the fact that distance may play a different role in different countries. In a country with very high density like Italy, distance may start playing a role for very short distances (a few dozen kilometers); in the US, where dioceses are much larger, distances of a few dozen kilometers seem mostly irrelevant. The interaction of distance and density can therefore be interpreted either as a usual interaction term (added effect of distance for countries with a certain density), or as a variable in its own right (holding absolute distance constant): this product, given the definition I take for density, represents the distance counted in terms of the average radius of dioceses in the country. A value of 2, for example, means a distance of two average dioceses.

Table (2.5) gives descriptive statistics for most controls.

Column (1) is a limited set of controls; column (2) is the full set of controls that I use throughout the rest of the study. Local bishops are more likely to be appointed

Variable	N	Mean	Std. Dev.	Min.	Max.
<i>log catholics if OO</i>	1373	11.99245	(1.146985)	7.824046	14.57912
<i>log catholics if AO</i>	1555	13.33079	(1.186076)	7.824046	15.96727
<i>AO</i>	2928	.5310792	(.4991184)	0	1
<i>Distance</i>	2928	371.6179	(495.3426)	0	3975.68
<i>Distance * Density</i>	2928	2.751598	(3.293844)	0	31.24728
<i>Age</i>	2928	55.20506	(7.085446)	35.73443	70.68309
<i>Total tenure as bishop</i>	2928	8.001607	(5.648679)	-18.2642	33.96851
<i>Tenure at previous appointment</i>	2928	7.035798	(4.772059)	.1149897	37.21013
<i>Born in province of vacancy</i>	2928	.1697404	(.3754687)	0	1
<i>Born in diocese of vacancy</i>	2928	.0665984	(.2493677)	0	1
<i>Worked in province of vacancy</i>	2928	.4043716	(.4908538)	0	1

Table 2.5: Descriptive statistics of regressors for successful candidates.

(and distance, as suspected, plays a larger role in dense countries); bishops with a low tenure at their current position, less likely.

The controls for the size (measured in log number of catholics) of the diocese of origin and vacant diocese are difficult to interpret, due to the interaction terms. The flexibility that this specification allows is necessary for a clean identification. A careful interpretation of these size effects is however possible.

Auxiliaries are more likely to be promoted to ordinary if they come from a large diocese. The effect of size of diocese of origin is more intricate for ordinary candidates. If the size of the vacant diocese is small, then ordinaries from small dioceses are favored; for larger vacant dioceses, ordinaries from larger dioceses are favored. For average sized dioceses, the effect is undetermined.

For a large diocese of origin and small vacant diocese, the results from column (2) imply that auxiliary candidates are more likely to be appointed; but for small diocese of origin and large vacant diocese, ordinaries are favoured.

The next regression output table (2.8) introduces the connection variables, and produces estimates based on them. Each of the three columns is a specification including the full set of controls, in addition to the variables indicated on the row of the table. These new variables are defined based on an underlying variable which,

Table 2.6: Conditional logit – no connection controls

	(1)	(2)
$\log(\text{catholics})_{cand} * \log(\text{catholics})_{vac} * OO$	0.173*** (0.0192)	0.175*** (0.0202)
$\log(\text{catholics})_{cand} * \log(\text{catholics})_{vac} * AO$	0.121*** (0.0162)	0.114*** (0.0170)
$\log(\text{catholics})_{cand} * OO$	-2.328*** (0.252)	-2.431*** (0.265)
$\log(\text{catholics})_{cand} * AO$	1.465*** (0.231)	1.384*** (0.243)
<i>AO</i>	-19.03*** (2.939)	-18.00*** (3.112)
<i>Distance</i>	-0.000954*** (0.0000969)	-0.00109*** (0.000134)
<i>Distance * Density</i>	-0.225*** (0.0138)	-0.235*** (0.0208)
$(\text{Distance})^2$		-0.000000280** (9.59e-08)
$(\text{Distance})^2 * \text{Density}$		0.000127*** (0.0000228)
<i>Age</i>	-0.0365*** (0.00363)	
<i>Age * OO</i>		-0.0364*** (0.00503)
<i>Age * AO</i>		-0.0529*** (0.00454)
<i>Total tenure as bishop</i>	-0.0128* (0.00636)	0.0343* (0.0172)
<i>Tenure at previous appointment</i>	0.0447*** (0.00653)	0.852*** (0.0500)
$(\text{Total tenure as bishop})^2$		-0.00198** (0.000656)
$(\text{Tenure at previous appointment})^2$		-0.0937*** (0.00725)
$(\text{Tenure at previous appointment})^3$		0.00383*** (0.000405)
$(\text{Tenure at previous appointment})^4$		-0.0000516*** (0.00000737)
<i>Candidate born in province of vacancy</i>		0.366*** (0.0852)
<i>Candidate born in diocese of vacancy</i>		0.565*** (0.125)
<i>Candidate worked in province of vacancy</i>		0.892*** (0.0591)
Observations	207716	207716

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Dependent variable = 1 if candidate is chosen for the vacancy. Full set of controls included.

Table 2.7: Frequency table of the number of connections for successful candidates

Item	Number	Per cent
0	2,357	80.50
1	419	14.31
2	121	4.13
3	19	0.65
4	9	0.31
5	3	0.1
Total	2,928	100.0

for every pair of incumbent and candidate, takes value one if they are connected (according to the criterion described in the previous section) and value zero otherwise.

- Variable $dConn$ takes value 1 if the candidate considered is connected to at least one incumbent; 0 otherwise. (It is the maximum, for each candidate, of the underlying variable described above).
- Variable $nConn$ takes for value the total number of incumbents that the candidate is connected to. (It is the sum of the underlying connection variable).
- Variable $nConnSq$ is the square of $nConn$, included here to allow for non-constant returns to connectedness.

The frequency table of the resulting $nConn$ variable is given in table (2.7).

I find a strongly significant effect of connections on the likelihood of promotions. Taking the relative probability at the average values of covariates, these results mean that a bishop with one connection is about 40% more likely to be promoted than one with none. The effect of connections show decreasing returns with the number of connections, with an estimated maximum effect at three connections.

The next table (2.9) breaks down the results above by country (only the con-

Table 2.8: Conditional logit results

	(1)	(2)	(3)
<i>dConn</i>	0.321*** (0.0583)		
<i>nConn</i>		0.208*** (0.0416)	0.369*** (0.0813)
<i>nConn</i> ²			-0.0757* (0.0338)
Observations	207716	207716	207716

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Dependent variable = 1 if candidate is chosen for the vacancy. Full set of controls included.

nection variable is allowed to have different effects across countries; all the other coefficients are constrained to be the same). This allows us to study potential variations in the size of the effect of connection across countries. However we note that the results seem remarkably stable over space. It is worth noting that there is no significant effect for countries other than the Americas and Europe; however those are in a minority, and the standard error is too large to interpret with confidence.

I do not report here the results from a similar breakdown over time instead; but the conclusion is the same: there does not seem to be a significant evolution of the size of the effect studied here over the time range of the data (1930 to 2008).

In table (2.10) I relax some of the assumptions that underly the definition of a connection, namely (1) instead of setting the relevant boundary for connections as being that of the ecclesiastical province, I introduce a distance cutoff of either 100km to 200km, and any bishop within this distance is considered a connection; and (2) results are reported where connections are only measured after individuals become bishops (i.e., not using the information on their place of birth). Specifications using this subset of connections are referred to as using “bishop connections”, by opposition to “all connections”.

Table 2.9: Conditional logit results across countries.

	(1)	(2)	(3)	(4)
<i>dConn</i>	0.321*** (0.0583)			
<i>dConn * Other Europe</i>		0.275** (0.119)		
<i>dConn * Italy</i>		0.577*** (0.135)		
<i>dConn * South America</i>		0.394*** (0.114)		
<i>dConn * North America</i>		0.291** (0.113)		
<i>dConn * Other</i>		-0.064 (0.185)		
<i>nConn</i>			0.208*** (0.0416)	
<i>nConn * Other Europe</i>				0.114 (0.0854)
<i>nConn * Italy</i>				0.342*** (0.0877)
<i>nConn * South America</i>				0.297*** (0.0874)
<i>nConn * North America</i>				0.238** (0.0825)
<i>nConn * Other</i>				-0.0056 (0.118)
Observations	207716	207716	207716	207716

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Dependent variable = 1 if candidate is chosen for the vacancy. Full set of controls included.

Table 2.10: Conditional logit results: alternative specifications.

	(1)	(2)	(3)	(4)	(5)	(6)
Connection type	all	all	all	bishop	bishop	bishop
Distance cutoff	province	100km	200km	province	100km	200km
<i>nConn</i>	0.369*** (0.0813)	0.398*** (0.0793)	0.269*** (0.0505)	0.463*** (0.157)	0.465** (0.145)	0.252** (0.0839)
<i>nConn</i> ²	-0.0757* (0.0337)	-0.0732* (0.0313)	-0.0240* (0.0119)	-0.114 (0.0900)	-0.114 (0.0794)	-0.0513 (0.0304)
Observations	207716	207716	207716	207716	207716	207716

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Dependent variable = 1 if candidate is chosen for the vacancy. Columns (1), (2) and (3): all connections; columns (4), (5) and (6): bishop connections; columns (1) and (4): cutoff = province boundary; columns (2) and (5): cutoff = 100km; columns (3) and (6): cutoff = 200km. Full set of controls included.

Results do not seem to vary significantly with the choice of specification; however, dropping the connections made before the first bishop appointment results in a considerable loss of efficiency, as indicated by the soaring standard errors. Hence for the rest of the study, I consider all connections, including those inferred from the place of birth.

“Placebo” specifications: connections to dead bishops

The data allows the researcher to build the connection variables not only for live incumbents, but also for moved, retired, or deceased ones. One would expect the effect of a connection to drop to zero, or at least be considerably reduced, as the connected incumbent dies; and be at least somewhat reduced as he moves to another province or retires. He will not be able to weigh in as actively as incumbents in office; however, a retired or faraway bishop can still be consulted by the nuncio, and a dead bishop may have talked to live incumbents about possible candidates before dying. Such an indirect recommendation could thus still explain some positive correlation.

Table (2.11) features two specifications; in each, I include one actual connection

Table 2.11: Conditional logit results: Placebo specifications.

	(1)	(2)
<i>nConn</i>	0.272*** (0.0780)	0.283*** (0.0652)
<i>Placebo nConn</i>	0.0376 (0.0948)	0.0610 (0.0601)
Observations	207716	207716

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

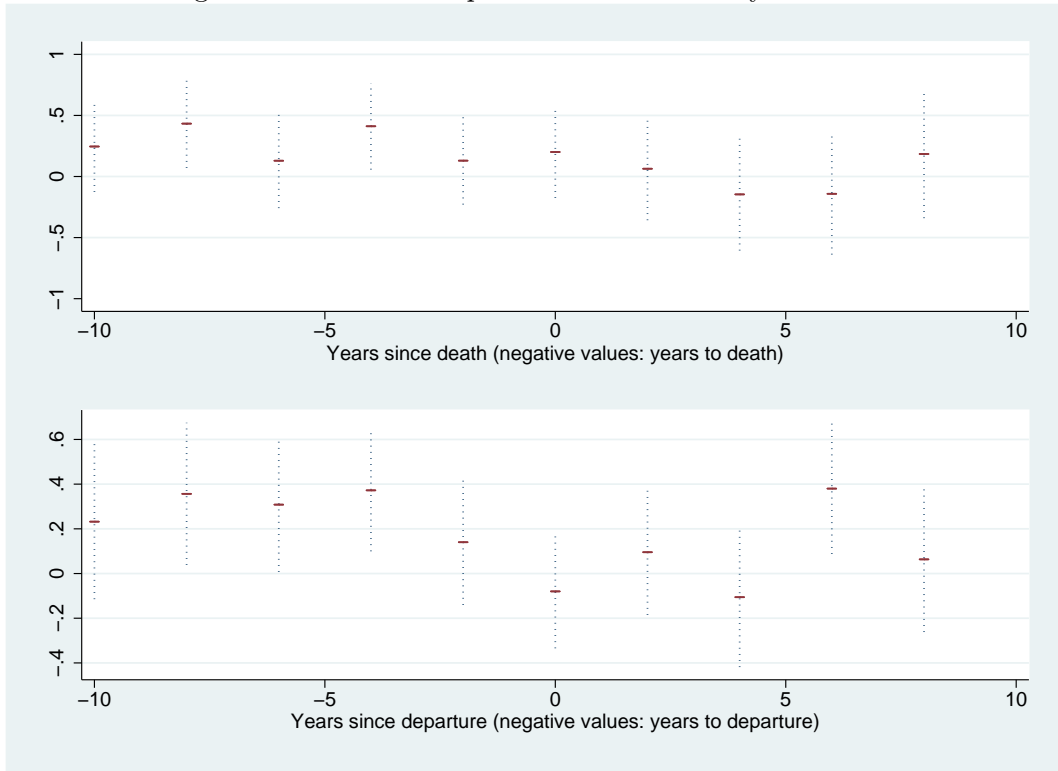
Dependent variable = 1 if candidate is chosen for the vacancy. Column (1): Placebo connections are connections to bishops (dead or alive) who (at the time of the vacancy) are no longer incumbents in the province of the vacancy. Column (2): Placebo connections are past connections to incumbents who are now dead (at the time of the vacancy). Full set of controls included.

control, and one placebo. In column (1), the actual control corresponds to connections with active incumbents, and the placebo to connections with bishops who are no longer in exercise in the province of the vacancy (whatever the cause: mobility, retirement, or death). Column (2) restricts the placebo to dead bishops.

The results validate the identification. Both specifications indicate that placebo connections have no significant effect on the likelihood of promotion. A test of equality of the effects of connections to incumbents and former incumbents (column 1) is rejected with a p-value of .0186; the same test for dead incumbents (column 2) has a p-value of .0685.

In the same spirit, figure (2.3) reports the regression coefficients (along with 95% confidence intervals) corresponding to a specification including (in addition to the full set of controls) a complete set of connection controls, broken down with respect to the date of death (respectively departure) of the incumbent. Points to the left of zero correspond to connections with bishops who are about to move or die; to the

Figure 2.3: Placebo coefficients: negative years are coefficients on connections with bishops still alive or in place; positive years correspond to connections with bishops who are no longer in exercise in the province of the vacancy.



right, bishops who moved or died some time ago. Each regressor covers two years: for example, one of them is “Number of connections with bishops who are going to die in 1 or 2 years”. These graphs indicate a decrease in the effect of connections that is synchronous with the death or mobility of the incumbent (though not very clearly, as this pushes the data to its limits, and the standard error for each coefficient is large).

Continuous attenuation with distance

So far, all specifications constrained the effect of connections to be binary: connections between bishops distant by less than a cutoff value (boundary of the province,

or distance of 100km or 200km) are predicted to have a constant effect; connections between bishops separated by a larger distance are expected either not to exist, or not to have any effect. This is of course at odds with the complexity of reality (even though probably valid as an approximation). In this section, we estimate a slightly more complicated specification, where connections enter the right hand side interacted with a term that features attenuation with distance, and there is no such arbitrary cutoff (connections corresponding to any distance within a country are included).

More precisely, looking again at the functional form for v_{kby} (the term through which connections between b and k affect the objective function of the Vatican) we have:

$$v_{kby} = \alpha \exp(-Z_{kby}\gamma) \quad (2.4)$$

So far Z was set equal to 0; here we include in it variables that are expected to possibly have an effect on the strength of the connection between b and k :

- The duration of the connection, labelled *DurConn*.
- The time elapsed since the connection was terminated, labelled *SinceConn*.
- The distance between the candidate and the incumbent at the time the connection was live, labelled *DistConn*.

All the other controls are unchanged, and still enter the utility function linearly.

The results for these new specifications are given in table (2.12). These results indicate a decrease with distance of the effect of a connection, but no clear pattern for the duration of a connection or the time since connection termination. A distance of 100 kilometers attenuates the effect of the connection by one half; a distance of 200 kilometers, by three quarters.

Table 2.12: Conditional logit results: attenuation with distance.

	(1)	(2)
connection		
<i>nConn</i>	0.245*** (0.0581)	0.504* (0.203)
distance		
<i>DistConn</i>	0.00693* (0.00278)	0.00786*** (0.00276)
<i>SinceConn</i>		.0288 (0.0304)
<i>DurConn</i>		.0246 (0.0166)
Observations	352482	352482

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Dependent variable = 1 if candidate is chosen for the vacancy. Full set of controls included.

2.5 Conclusion

This study shows that connections matter, and matter a lot. In some specifications, a connected bishop's likelihood of being chosen for a given vacancy is predicted to exceed by up to 50% the likelihood for the same, counterfactual, unconnected bishop. This was done using only a small subset of the whole range of connections available to bishops: I only use connections originating from past geographical proximity. In reality, there are many more ways in which bishops can establish connections. People they study with, bishops they work with at the national conference for bishops, friends of friends, and so on and so forth: this study focuses on the observable tip of a much more complex iceberg.

As illustrated in the model of section 2, the social implications of this phenomenon depend on a number of factors. These factors include how accurately the Vatican can assess the ability of the bishops, and how much personal utility a bishop receives when someone he has interacted with in the past is promoted to a nearby diocese.

As these are difficult to quantify without more data, this study cannot conclude as to whether the phenomenon it measures is beneficial or harmful for the Catholic Church.

An important question is whether, and to what extent, the results obtained here for the Catholic Church can be extrapolated to other organizations. Only further research can answer this accurately; nonetheless, despite all the features that make it unique, the Catholic Church at its core is no different from most organizations. The Vatican aims to select the best person for a given diocese, in the same way as a profit-maximizing organization wishes to promote its most able workers. The institutions make it possible for social connections to play a role in this; it is also the case in many firms or organizations in general. Except in rare cases where performance is perfectly observable, the process of measuring an employee's ability is bound to interact with that of creating a social connection. In a sense, the fact that the estimates obtained here seem rather stable over time and space also lends some credibility to their external validity, as it proves that the phenomenon studied is robust to the differences between countries, as well as to the evolutions of the institutions over much of the twentieth century.

The framework developed here to measure the effect of social connections on careers can likely be used in other settings. Its appeal is that it requires no information about actual connections – e.g. no survey asking for the identity of friends. All it requires is complete career history, and a wide range over time and space. A similar approach, for example, could potentially be transposed for military data, or government agencies (where departments or ministries could play the role of dioceses).

Chapter 3

Intergenerational Transmission of Human Capital

3.1 Introduction

It is a widely observed fact, in all developed countries, that more educated parents have more educated children on average. This intergenerational correlation is a central concern for policies dealing with equality of opportunity. It suggests that policies aiming at increasing the average educational level of the population may produce, if indeed there is a causal link, long-lasting effects through spill-overs to the next generations. Several processes can drive this correlation. For example, parents enjoying abilities that help achieve higher levels of schooling are also more likely to have children with the same abilities (ability bias). If this were the only driver of the observed correlation, increasing education of one cohort would not necessarily improve the opportunities of the next generation. Several arguments point to a causal impact of parents education on their children, however. Higher education often induces higher income, which helps raise children in conditions that are more favorable for schooling (Maurin, 2002). Education, more generally, can improve some

parents' skills that are useful for their children's education. Children can also conform to their parents' model.

The last decade has seen an upsurge in the analysis of this causal relationship. Most of this empirical literature is a reduced-form analysis (for one exception see for instance Belzil and Hansen, 2003), and exploits three main sources of exogenous variations of potential innate abilities of parents (for a comparison of these methods see Holmlund et al., 2010). First, exploiting differences in education between monozygotic twins allows researchers to control for genetic endowment (see for a recent example Behrman and Rosenzweig, 2002). Second, some studies exploit the difference in genetic endowment between adoptive parents and children to avoid ability bias (Plug, 2004). Finally, several empirical studies rely on natural experiments to mimic experimental conditions, by using, for instance, changes in educational policy as a source of exogenous variations in parental education. Numerous studies, for instance, have exploited modifications in compulsory schooling laws. Such studies notably include Oreopoulos and Page (2006) for the United States, Black et al. (2005) for Norway, and Oreopoulos (2006) for the United Kingdom.

Despite this renewed interest, no consensus emerges. Results differ markedly across countries. The conclusions also differ on whether intergenerational transmission of education is more related to mother or father. For instance Black et al. (2005) find no significant impact of father's education on their children's schooling but a positive impact of mothers' schooling (mostly on their sons' schooling). On the contrary neither Plug (2004) nor Holmlund et al. (2010) find any significant impact of mothers' education on their children's schooling, while fathers' education is found to have a significant impact. The main channel of this possible causal relationship is also undetermined.

This paper adds the French case to the existing literature. We use the most recent increase in compulsory education, known as the Berthoin reform, as a natural

experiment. This reform was voted in France in the late 1950s, and affected all the individuals born in 1953 or later. It increased the school leaving age from 14 to 16, bringing the duration of compulsory education from eight to ten years. One specificity of this reform is that it was carefully prepared, applied to the whole targeted generation, and thus resulted in remarkable compliance. Another specificity is that this new schooling age does not correspond to any specific schooling graduation. As shown by Grenet (2010), this explains why this policy did not lead to any significant increase in income, contrary to the UK analogue. This provides us with a natural experiment that introduces an exogenous change in parental education but not in parental income or career. Therefore, the effect that we set out to identify is the causal effect of parental education, abstracted from the changes in parental outcomes that would normally themselves be caused by an increase in education. In fact, at least three causal pathways can be at play between parental education and child education: (1) the direct pathway, e.g. parents with more education understand better how to teach their children to perform better in the educational system; (2) the income pathway, e.g. parental education causes more parental income, which allows the parents to create a familial environment more favourable to child educational achievement; (3) the social pathway, e.g. parents with more education form social links, including marriage, with individuals that are also more educated, which causes better child education. The effect that we are able to identify in this paper is likely a combination of the first and third of these causal pathways; but the evidence from Grenet (2010) makes a convincing case that the second is negligible.

In order to estimate the impact of parental education on their children's schooling attainment, we use a regression discontinuity design around the time of the enforcement of the compulsory schooling reform. This method has been widely used in empirical economics since the late 1990s. It exploits a given exogenous threshold determining the "treatment", which, here, is the fact of having achieved more than 10

years of education (meaning leaving school after the age of 16, as French compulsory schooling starts at 6). The Berthoin reform introduced an exogenous discontinuity in the probability that one leaves school after 16, depending on birth cohort. Parents born before 1953 face a compulsory school leaving age of 14, while those born later face a compulsory school leaving age of 16. Using the common terminology in this empirical literature, it should be noted that the design is “fuzzy” and not sharp. A significant portion of the cohorts affected by the Berthoin reform already stay at school after 16. The reform only introduced a discontinuity in the probability of studying up to 16 years, rather than directly in the years of education of the individuals. As shown by Hahn et al., 2001, identification can be achieved in this setting under some assumptions. Estimation relies on a two-step procedure.

We use data from the French Labour Force Survey (also known as *Enquête Emploi*). The *Enquête Emploi* is a large household survey providing information on all individuals aged 15 and over (for uses of this data on related topics see for instance Grenet, 2010 or Maurin and McNally, 2008). As all household members are interviewed, one can reliably link the education of one parent with schooling attainment of his (still at home) teenage child. We focus on the educational achievement of children aged 16 at the date of the survey, more precisely whether a child is in sixth form of high school (*Seconde*) at that age. The sample size is large enough to target precisely the cohort affected by the Berthoin reform and to use the discontinuity it has created.

Another study, Albouy and Lequien (2009), similarly uses the Berthoin reform as a natural experiment; however it focuses on the health outcomes of the impacted individuals, and uses data from the *Echantillon Démographique Permanent* rather than the *Enquête Emploi*. The authors are unable to find a significant effect of an exogenous change in education on subsequent mortality, in spite of a vast literature suggesting that such an effect does, in fact, exist. The reason for this may be related

to the points raised previously, namely that the Berthoin reform did not cause the changes in income or lifestyle that could themselves have had an impact on health outcome.

We confirm the observations of previous studies (Albouy and Lequien, 2009 and Grenet, 2010) that the Berthoin reform had an effective impact on the school leaving age of the affected cohorts. At the year of the discontinuity, about 10% more parents leave school at age 16 or above than the year before, jumping from 70% to 80% of a cohort. This effect is entirely concentrated on the year of the discontinuity; on both sides of the discontinuity, a stable linear trend towards more education can be observed. This qualifies the reform as a valid candidate for fuzzy regression discontinuity design.

Our estimates indicate a positive impact of fathers' education on their children's schooling, but an insignificant effect (the point estimate being negative in some specifications) of mother education on their children. For father education, the effect appears to be particularly present on daughter's education; it is found that complying to the Berthoin reform raises by as much as 30% to 40% the probability that the child does not repeat a grade by age 16.

The paper is organized as follows. The next section presents the Berthoin reform and its impact on education level of the targeted cohorts. The third part presents the data and some descriptive statistics. Finally, the fourth part presents our identification strategy and our results.

3.2 The Berthoin reform

As in many other Western countries during that period, compulsory schooling in France experienced several reforms during the 20th century. Their purpose was to reduce poverty by providing universal access to basic education, thereby (for

example) reducing illiteracy.

Following the seminal reform initiated by Jules Ferry in 1882, schooling was compulsory in France for children aged 6 to 13. In 1936, the Zay reform extended the leaving age for compulsory schooling to 14 for all children born after 1923. Finally, the Berthoin reform in 1959 added two years of compulsory schooling. Unlike the previous reform, this reform was announced and scheduled ahead for implementation, and the extension of compulsory schooling applies only to “children over six years of age on January 1, 1959” (for a detailed description see Grenet, 2010). It was thus effective in 1967, for cohorts born from January 1, 1953 onwards.

The reform had a noticeable impact on school leaving age for both men and women (see Figure 3.1). The average school leaving age jumped from 17.46 years in 1952 to 17.79 in 1953. The proportion of one cohort that completed at least ten years of education (i.e. left after age of 16) is 69.1% for the 1952 cohort (last pre-Berthoin cohort) while it is 79.4% for the 1953 cohort (first post-Berthoin cohort). This effect clearly came into play for the generation born in 1953, with no detectable evidence of anticipation nor of progressive implementation in the follow-up years. Figure (3.1) also highlights a secular trend towards longer education in France in the middle of the 20th century. Less than 40% of individuals born in 1930 were still at school at age 16; by 1960, almost everyone was. The effect of the Berthoin reform therefore needs to be carefully separated from this secular trend.

Figure 3.2 provides another view on the same data, by reporting year-on-year variations in the proportion of a cohort who left school at age 16 or higher. The effect of the reform is even more striking than in the previous graph. All the effect is concentrated in just one year; on either side of 1953, the year on year increase is remarkably stable. This confirms that the reform was (a) not followed by individuals from older cohorts in anticipation of the new law and (b) implemented with sufficient preparation that it achieved all its effect in the first year it was implemented. The

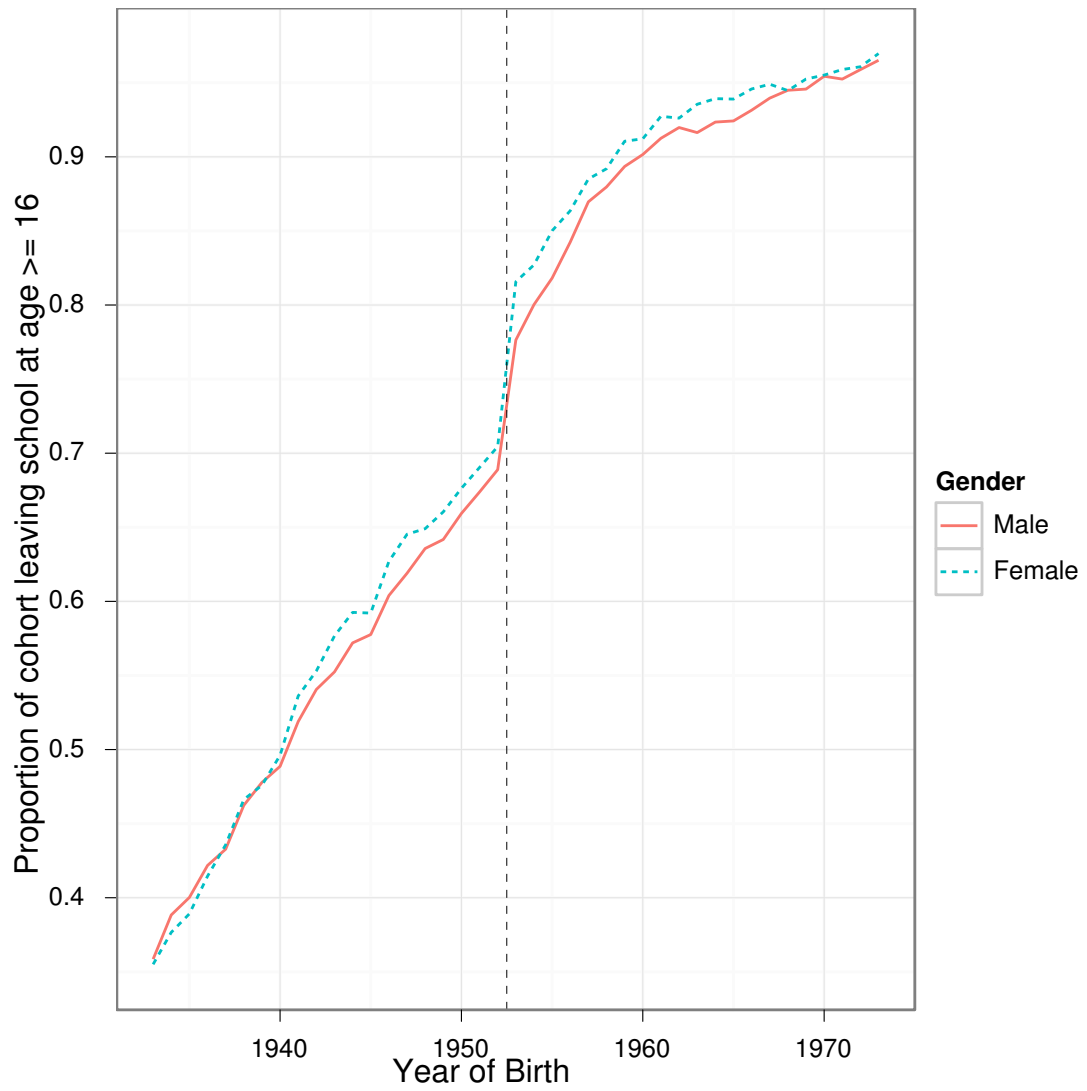


Figure 3.1: First stage on extended sample, by gender

effect of the reform appears to have been slightly stronger for females than for males, possibly because the benefits from non-compliance (e.g. expected income from early entry in the labour market) were perceived to be lower for females.

The impact is much more noticeable for individuals with lower socio-economic backgrounds, as indicated in table 3.1. Children of executive or middle-class fathers were very likely to stay in education after age 16 (up to 97% of children of managers and intellectuals). The reform had a much greater impact on children of manual workers. For the former, the reform was irrelevant, while for the latter, the probability of studying until at least age 16 jumped from 57.5% to 72%.

	1952	1953
Farmers	63.8%	73.4%
Self-employed and business owners	77.9%	85.2%
Managers and intellectuals	97.3%	96.7%
Intermediate professions	87.5%	92.4%
Employees	77.2%	86.2%
Workers	57.5%	72.4%

Table 3.1: Percentage of individuals born in 1952 and 1953 who left school at age 16 or late, by socio-professional category of father

It should be emphasized that while the reform did impact the individuals who did not study beyond age 16, it had no noticeable spillover effect on higher education levels. Indeed, we can observe that, from the year of the reform onwards, a vast number of individuals who would have normally left education at age 14, went on until age 16 instead (see Figure 3.3). By contrast, we do not observe any impact in the proportion of the affected cohorts that continue education beyond compulsory school (meaning a school leaving age higher than 17). This is consistent with the observation that the reform did not result in a noticeable change in the qualification of these cohorts, as previously stressed by Grenet (2010). A school leaving age of 16 does not correspond to any diploma in the French educational system (see Figure

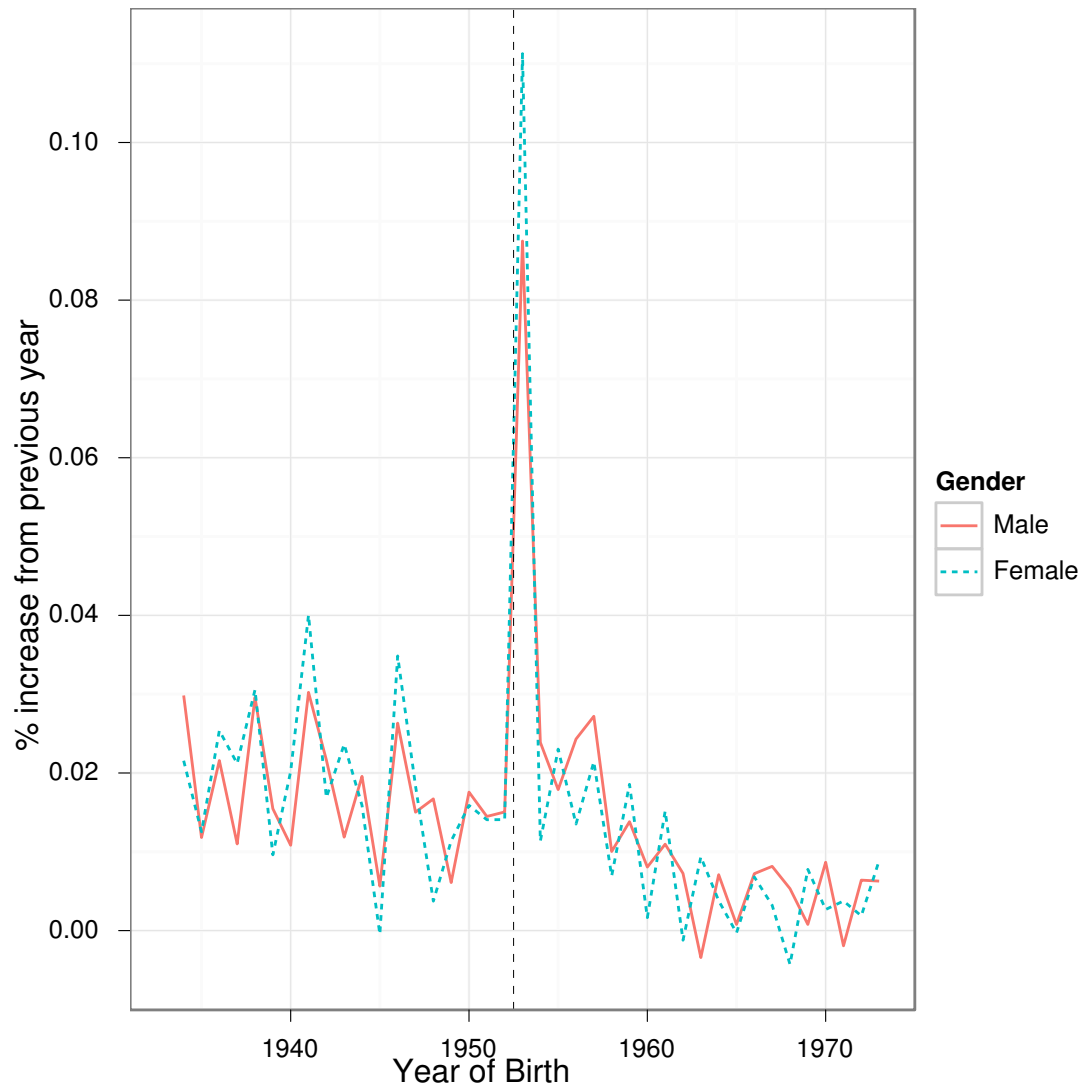


Figure 3.2: Year-on-year increase in proportion of cohort leaving school at age ≥ 16 , by gender

3.4).

As mentioned in our introduction, we feel that this feature is an opportunity for us to measure intergenerational transmission of human capital in a way that is not only causal, but also excludes causal pathways that involve parental income or diploma, contrary to previous studies.

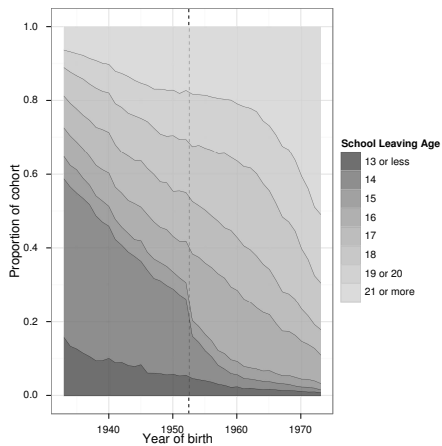


Figure 3.3: School leaving age – both genders

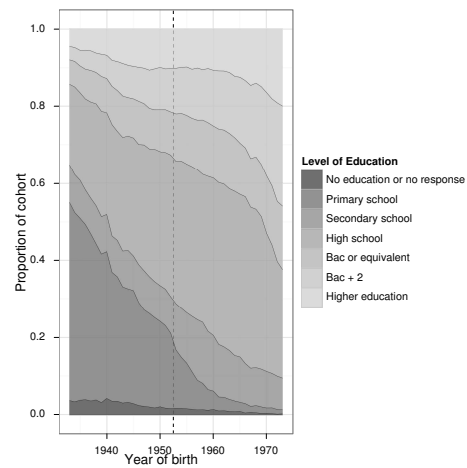


Figure 3.4: Level of education – both genders

3.3 Data and Descriptive Analysis

3.3.1 Data

We use data provided by the French Labor Force Survey (*Enquête Emploi*) from 1982 to 2010. This survey was conducted by the French Office of Statistics (*INSEE*) over a large sample of individuals. It provides detailed information on labor force situation as well as personal information such as education level. All members of each household are registered, and every member older than 15 is interviewed in detail. Children can thus be linked to their parents. We could therefore precisely identify the individuals whose parents were born around 1953, and measure their

school achievement.

From 1982 to 2002 the *Enquête* was conducted each year over a sample of about 100,000 households (1/600 of the French population), and conducted quarterly since 2002 over a sample of 55,000 households. Pooling all these surveys, we obtain a large sample of individuals born around 1953. The sample size is still large even when restricting to households with teenage children. We thus observe at least 4,000 individuals (male and female) for each cohort born between 1943 and 1963 with a child age 16 years at the date of survey. The questionnaire of the LFS was modified twice since 1982, however. As we mostly use fairly basic information, these changes do not affect most of our variables of interest (cohort of birth, gender, school leaving age, socio-professional group...) but induce breaks in some variables measuring schooling levels (see below).

3.3.2 Characteristics of Parents and impact of the reform

Regarding parents, we are able to measure accurately the year of final graduation. We also have information on location of birth of each respondent, on their occupation, as well as the occupation and place of birth of their fathers (i.e. the grandfathers of their children). Table 3.2 present basic statistics for individuals born between 1943 and 1963, irrespective of parental and marital status.

We choose to use as an indicator the fact of having graduated after age 16, which is the outcome that the reform explicitly targets. First estimates confirm the impact of the reform observed in the previous section (Figure 3.1). Table (3.3) shows a very strongly significant effect of the Berthoin reform, even when controlling for a polynomial of degree 2 in the year of birth of the individual.

	Proportion	Freq
Socio-Professional Group		
Employees	0.24	111550
Farmer	0.02	10872
Intermediate professions	0.18	84776
Managers and intellectuals	0.10	46628
Retired	0.01	2417
Self-employed and business owners	0.06	27915
Unemployed or military	0.15	69880
Workers	0.23	108152
School Leaving Age		
13 or less	0.05	21327
14	0.15	68627
15	0.04	20076
16	0.15	67951
17	0.13	62155
18	0.16	75221
19 or 20	0.13	61742
21 or more	0.18	84564
Level of education		
No education or no response	0.02	5050
Primary school	0.15	43980
Secondary school	0.12	34243
High school	0.37	105478
Bac or equivalent	0.12	34887
Bac + 2	0.12	34116
Higher education	0.10	29693

Table 3.2: Descriptive statistics on individuals born between 1943 - 1963

3.3.3 Education of children

Our main indicator of educational attainment for children is to be in sixth form of high school (*seconde*) at age 16. This is the case for children who have never repeated a grade, or who have not repeated more grades than they have skipped grades. Children who both skipped and repeated a grade by age 16 are a small minority, hence our measure can safely be interpreted as indicating children with no history of grade repetition. This choice was driven by several considerations.

Table 3.3: First stage results.

	Males	Females
(Intercept)	0.70088 (0.00348) ***	0.71762 (0.00323) ***
Post-reform	0.07701 (0.00496) ***	0.09719 (0.00458) ***
(Year of birth) ¹	0.01761 (0.00078) ***	0.01533 (0.00072) ***
(Year of birth) ²	0.00071 (0.00014) ***	0.00014 (0.00013)
Number of observations	123083	134883
F-test for excluded instrument	240.592	450.562

Effect of the reform on the affected cohort's education achievement. The dependent variable is an indicator that the individual was at school until age 16 or more. The variable of interest is Post-reform, indicating whether the individual was born in 1953 or later and was therefore affected by the Berthoin reform. The sample is all individuals born within 5 years of the reform and surveyed in the LFS (Enquete Emploi) before 2002.

Significance key: *** = .1%, ** = 1%, * = 5%, . = 10%.

First, being in high school is of primary importance for the considered cohorts in terms of access to jobs and future earnings. It notably conditions the graduation of high school diploma (*baccalauréat*), meaning the first level of diploma that matters in terms of labour market prospects. The first part of French high school, the *college*, also coincides with the award of a diploma, the BEPC or *brevet*, but it is achieved by almost 100% of a cohort and considered as useless in terms of labour market prospects as well as the entrance to senior high school.

Second, several studies indicate that grade-retention is correlated with lower future educational achievement and ultimately earnings (see for instance Brodaty et al., 2008 and Manacorda, 2010). However, this could not be accurately measured before the sixth-form of high school. In France, grade retention is made with the agreement of pupils' parents until that level of schooling. This could create discrepancies for students according to their social origin for instance. By contrast, entrance in senior high school requires the sole agreement of teachers at the end of the junior high school.

Third, we prefer to focus on children sufficiently young to still live at their parents' house, in order to be sure that they are simultaneously interviewed in the survey. This disqualifies for instance the use as an indicator the fact of obtaining a high-

school diploma : the normal age for getting this diploma is 18, an age at which a number of youths do live on their own. We thus avoid the risk of endogenous attrition by focusing on younger individuals.

Finally, this variable appears less sensitive to the changes occurred in the LFS. Figure 3.5 presents the proportion of 16-year-old children in time in *Seconde*. It clearly presents an upward trend over the period. We do observe a dip in 1990, which could be related to a change in the questionnaire of the LFS that year. By contrast, the proportion of 16-year-old children with *brevet*, on the same graph, show major discontinuity in 1990 and 2002 (years at which the LFS underwent methodology changes); it is, besides, not available before 1990. On that same graph, we also report, on a third time-series, individuals who are in *seconde* and have been awarded the *brevet*.

For those reasons, our preferred measure of children educational achievement is defined as being in *seconde* (or a higher grade) by age 16; however we will report some results for the obtention of the *brevet* by that same age, for the sake of robustness.

Table 3.4 reports results obtained by linear regression of child education (defined as being in *seconde* at age 16) on parental education and parental year of birth. It indicates that on average and controlling for parental year of birth as well as survey year, children of a parent who stayed in school until at least age 16 are about 25% more likely to be in *seconde* by age 16.

3.4 Methodology and Results

3.4.1 Identification strategy : RDD estimation

Our identification strategy exploits the discontinuity introduced by the Berthoin reform in the completed schooling of cohorts born around 1953. More precisely, we use a Regression on Discontinuity Design, which became standard in the literature

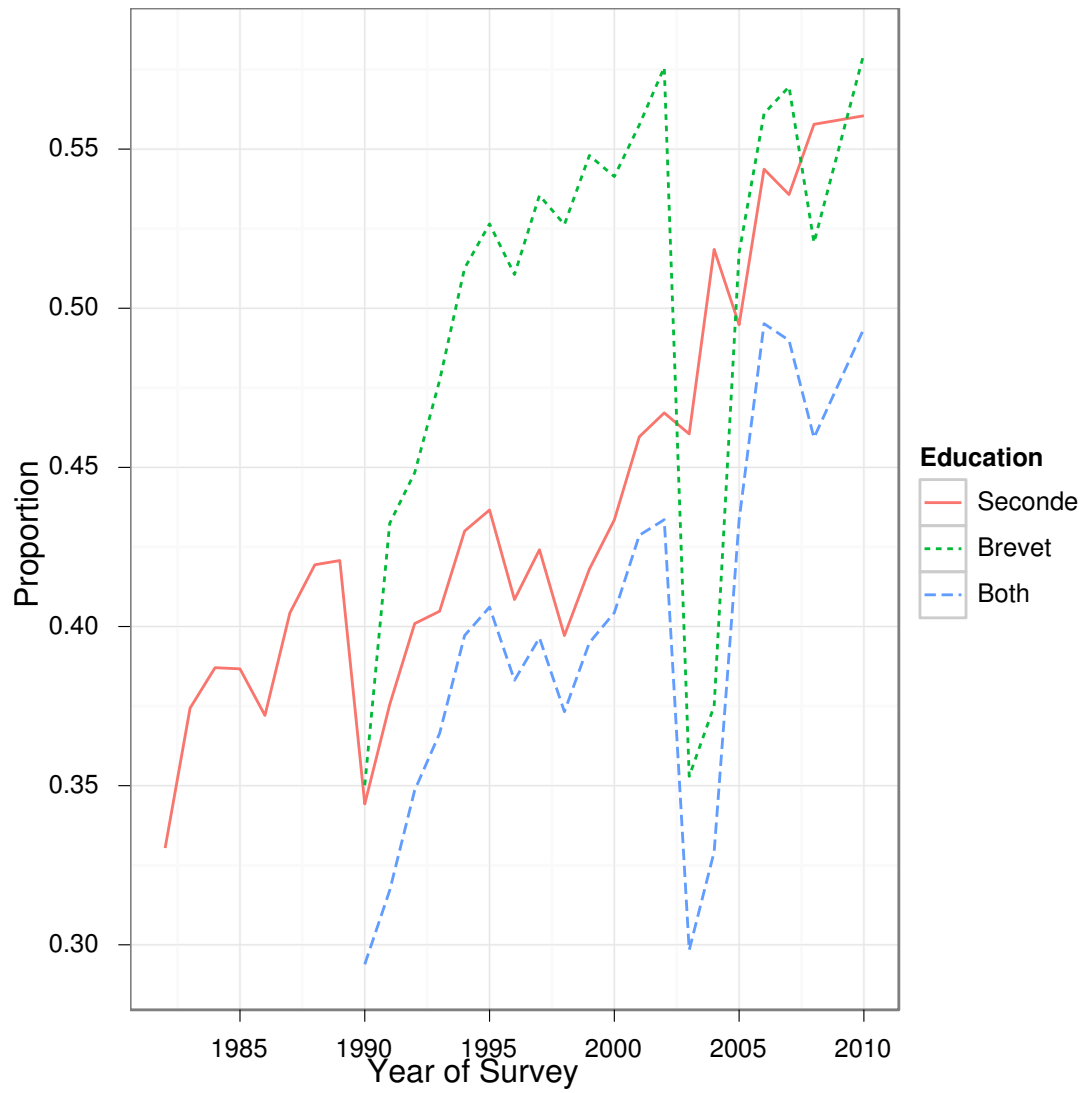


Figure 3.5: Proportion of 16 year old in Seconde or with Brevet, by year of survey

Table 3.4: OLS results.

	Father	Father	Mother	Mother
(Intercept)	0.11 (0.022) ***	0.20 (0.023) ***	0.081 (0.018) ***	0.15 (0.019) ***
Parent SLA \geq 16	0.25 (0.0085) ***	0.26 (0.0089) ***	0.26 (0.0080) ***	0.28 (0.0083) ***
(Parent's year of birth) ¹	-0.012 (0.00091) ***	-0.012 (0.00096) ***	-0.015 (0.00087) ***	-0.016 (0.00091) ***
(Parent's year of birth) ²	0 (0.00013)	-0.00014 (0.00014)	0.00014 (0.00012)	.00005 (0.00013)
(Survey year) ¹	.0008 (0.0026)	0.0045 (0.0027) .	-0.00018 (0.0023)	0.00580 (0.0024) *
(Survey year) ²	0.00039 (.00008) ***	0.00023 (.00008) **	0.00046 (.00008) ***	0.00031 (.00008) ***
Number of observations	16439	15459	19063	18260
Child gender	Son	Daughter	Son	Daughter

OLS regression relating parental schooling and child education. The dependent variable is an indicator that the child was in *seconde* or higher at age 16. The variable of interest is Parent SLA \geq 16, indicating whether the parent left school at age 16 or higher. The sample is all individuals born within 10 years of the reform and surveyed in the LFS (Enquete Emploi) before 2002.

Significance key: *** = .1%, ** = 1%, * = 5%, . = 10%.

Gender	Seconde	Brevet
Female	48%	56%
Male	36%	45%

Table 3.5: Education at age 16 by child gender

on causal effects and treatment effects since (at least) the late 90s (see Imbens and Lemieux, 2008 for a clear and precise presentation of this method). The basic setting of this design is a context where the endogenous variable T partly depends on the value of another continuous value, the selection variable. The key feature of regression discontinuity designs is that this dependency is discontinuous for a known cutoff value of this selection variable. As formally proven by Hahn et al. (2001), under some identification assumptions one can consistently estimate the causal impact of the treatment on the outcome of interest.

In our case, we are interested in the effect of having completed at least ten years of schooling on the schooling attainments of a person's children. As mentioned in the last section, completed schooling is an increasing function of the cohort of birth,

as it has increased all over the century, and “jumped” for cohorts born in 1953 and after (due to the Berthoin reform). Note also that more than half of the people born before 1953 already completed more than 10 years of schooling, while a number of individuals born later still completed less than mandatory schooling. We therefore face a so-called “fuzzy design”. Several assumptions are required for identification of a causal impact.

Let us denote by S the year of birth (our selection variable), and \underline{S} the value of the threshold (1953). We follow the potential outcomes framework, and denote, for an individual i , by Y_{i0} , the potential value of completed schooling if he had not been affected by the Berthoin reform (i.e. birth before 1953); and by Y_{i1} , the potential value if he had been affected (i.e. birth after 1953). As stated by Hahn et al. (2001), the causal impact can be locally identified provided a local continuity assumption (meaning that no other variable discontinuously impacts the outcome at the cut-off point) and a local monotonicity assumption (similar to Angrist et al., 1996) are satisfied¹. More specifically, this latter assumption states that the probability of having performed at least ten years of schooling increases due to the reform for all individuals. This rules out the existence of “defiers” (in the AIR terminology), individuals that would have performed more than ten years of schooling when it was not mandatory and less instead. This seems a reasonably mild assumption in our case.

Under these assumptions, a consistent estimator for the causal impact δ of having parents who completed ten years of schooling on one child’s schooling attainment is given by:

$$\delta = \frac{\lim_{S \searrow \underline{S}} E(Y|S) - \lim_{S \nearrow \underline{S}} E(Y|S)}{\lim_{S \searrow \underline{S}} E(T|S) - \lim_{S \nearrow \underline{S}} E(T|S)} \quad (3.1)$$

¹An alternative would be a somewhat stronger local independence assumption, see again Hahn et al. (2001).

In practice (see Imbens and Lemieux, 2008), semi-parametric estimation in an RDD can be simply performed by local linear regression. More precisely, a consistent estimate of the numerator $\lim_{S \searrow \underline{S}} E(Y|S) - \lim_{S \nearrow \underline{S}} E(Y|S)$ is provided by:

$$\min_{(\delta, b_1, \dots, b_K)} \sum_i 1_{(\underline{S}+h > S_i > \underline{S}-h)} (Y_i - a - \delta 1_{(S_i > \underline{S})} - b_1(S_i - \underline{S}) - b'_1(S_i - \underline{S}) 1_{(S_i > \underline{S})} - \dots)^2$$

where h is the size of the bandwidth and K the degree of polynomial extension. Similar estimate could be obtained for the denominator $\lim_{S \searrow \underline{S}} E(T|S) - \lim_{S \nearrow \underline{S}} E(T|S)$.

Two remarks are in order. First, one should emphasize that the obtained estimator is a LATE (Local Average Treatment Effect). As such, it measures the average impact of parents' schooling for individuals who comply to schooling obligation: they perform eight years of schooling if born before 1953, and ten if born after. It is also restricted to children whose parents are born around 1953.

Second, under some technical choices, an estimate for 3.1 obtained by 2SLS method, using the indicator $1_{(S_i > \underline{S})}$ as the excluded instrument, would be numerically identical to an estimate based on separate estimates of the denominator and numerator (see Imbens and Lemieux, 2008)².

3.4.2 Technical choices

We choose to estimate $\lim_{S \searrow \underline{S}} E(T|S) - \lim_{S \nearrow \underline{S}} E(T|S)$ (which corresponds to the "first stage" in an IV procedure) on all the individuals who were born close to 1953 (one to ten years, depending on specifications), i.e. the cohorts impacted by the reform. We thus use a much larger sample than the one we can use for estimating

²More specifically, using the same bandwidth size and polynomial extension degrees for both denominator and numerator the estimator could be simply obtained by a 2SLS estimation on the subsample of the neighborhood around the cut-off value.

the numerator $\lim_{S \searrow \underline{S}} E(Y|S) - \lim_{S \nearrow \underline{S}} E(Y|S)$ (the “second-stage”), as we do not restrict the sample to individuals who went on to have children. This is not an entirely innocuous procedure, as those individuals may differ slightly from the 2nd stage population – i.e. the individuals who are surveyed when they have a 16-year-old child in the household. However, we feel that the benefits of this approach, in terms of the precision of the first stage we are thus able to estimate, far outweigh any possible bias that could arise, for example, if future parents were differently affected by the reform than the average individual. Note, also, that most individuals in this extended sample *will* be a parent; whether we can use these individuals in the second stage depends on whether we observe the household when a child is 16, which is random – albeit arguably increasing in the number of children in the household.

We first choose to use a non parametric estimator with a small bandwidth (one to three years after the introduction of the Berthoin reform). This corresponds to a Wald estimator. As the sample size is small, we also present the estimates using a larger bandwidth, and controlling for the potential impact of cohort on schooling by polynomial function. We obtain inference by bootstrap.

3.4.3 Two-stage estimation

The tables below describes results that depart slightly from the discontinuity of 1953; rather, they rely on a polynomial parameterization, both of the relationship between parental education and parental year of birth, and of that of children educational achievement and parental year of birth. In our baseline, preferred specification, we allow for a polynomial of degree two to describe these relationships within a window of 10 years on each side of the reform – that is, parents born between 1943 and 1963. While we also report results using polynomials of degree 1 and 3, we feel that degree 1 is too low to capture the possible concavity of the relationships, while degree 3 may be asking too much of the data given the relatively small size of the window.

Fathers

In many of the specifications relative to paternal education, we find a large and significant effect of parental education on a child's early educational achievement. More precisely, in our preferred specification (column (1) in Table (3.6)), we find that the fact that a father who finished school after age 16, rather than earlier, improves by about 25% the likelihood for his offspring to have reached *seconde* by age 16. Similar results are obtained with our other child education variables in columns (4) and (5).

Columns (6) and (7) allow the functional form around the discontinuity to depart from a polynomial of degree 2. (6) and (7) use, respectively, polynomials of degree 1 and 3. While the results in (6) are remarkably similar to the baseline, the measured effect is somewhat weakened in column (7); the p-value for the significance test of the effect, nonetheless, is still relatively low, at about 11%. The magnitude of the point estimate is mostly unchanged in both specifications varying between 25% and 30%.

Column (8) uses a continuous dependent variable (parental school leaving age) rather than the indicator used in the other columns. The estimation is too imprecise for the result to be relied on; nonetheless the point estimate is, again, in the same region as the other specifications. The same applies to column (9), in which we reduce (at the expense of sample size) the size of the window, to consider only 5 years on either side of 1953.

Mothers

Table (3.7) reports the same specifications for the effect of maternal education. Though the precision of our estimates does not allow us to rule out the possibility of a moderate effect, it appears that maternal education has a smaller causal effect on child early education outcomes than paternal education does. In fact, one

specification, in column (4), indicates a significantly negative effect; however, it relies on the measurement of child educational achievement (obtention of “Brevet” by age 16) which we trust the least (see dips in the time series, as seen in figure (3.5)).

3.4.4 Wald estimation

We estimate in this section the causal effect of interest using the Wald estimator, in direct proximity to the 1953 break. This estimator relies very explicitly on the discontinuity introduced by the reform; ideally we would compare individuals born just before 1st January 1953, to individuals born just after. Indeed, as we have seen, the reform was introduced at a time when there was a very clear trend towards longer education; therefore, taking a window of even a few years around 1953 would potentially be comparing individuals facing a different social and education environment (even if the reform had never happened). However, the narrower our window around 1953 is, and the fewer individuals we have in our sample; therefore we need our window not to be too narrow. We choose to make its radius vary between one and three years.

Tables (3.8) and (3.9) present results for these different window sizes, and respectively for fathers and mothers. The Wald estimates are reported separately for boys and for girls, as well as a third specification that does not distinguish child gender. The sample size for these estimations varies between 3555 children (father specification with 1-year window) and 12317 (mother specification with 3-year window).

We again find evidence of a significant, possibly large effect of father education on child educational outcome. In the one-year window specification, our point estimate is particularly large, though with very wide confidence bounds, resulting from the small sample size. The two- and three-year window specification point to a smaller, yet still considerable effect. Going to school for two additional years as a result of the reform is indeed estimated to increase by about 20% (though it varies across

Table 3.6: Second stage results – Fathers.

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(Intercept)	0.09 (0.093)	0.314 (0.123) *	0.24 (0.139) .	0.196 (0.097) *	0 (0.091)	0.093 (0.092)	0.043 (0.158)	-3.65 (5.48)	0.134 (0.145)
$\hat{e}du_{parent}$	0.243 (0.123) *	0.111 (0.164)	0.38 (0.185) *	0.164 (0.129)	0.32 (0.121) **	0.242 (0.123) *	0.306 (0.209)	0.221 (0.308)	0.195 (0.194)
sex	0.125 (0.006) ***			0.117 (0.006) ***	0.118 (0.005) ***	0.125 (0.006) ***	0.125 (0.006) ***	0.125 (0.006) ***	0.122 (0.007) ***
(Year of birth) ¹	-0.003 (0.003)	0 (0.004)	-0.006 (0.004)	-0.001 (0.003)	-0.007 (0.003) *	-0.003 (0.003)	-0.005 (0.006)	-0.018 (0.028)	-0.001 (0.006)
(Year of birth) ²	0 (0)	0 (0)	0 (0)	0 (0) *	0 (0)		0 (0)	0.001 (0.001)	0 (0)
(Year of birth) ³							0 (0)		
Number of observations	30314	15614	14700	24932	30314	30314	30314	30314	18137
Dependent Variable	Seconde	Seconde	Seconde	Brevet	Both	Seconde	Seconde	Seconde	Seconde
Window (years)	+/- 10	+/- 10	+/- 10	+/- 10	+/- 10	+/- 10	+/- 10	+/- 10	+/- 5
Child gender	Both	Boy	Girl	Both	Both	Both	Both	Both	Both
Endogenous variable	SLA >= 16	SLA >= 16	SLA >= 16	SLA >= 16	SLA >= 16	SLA >= 16	SLA >= 16	SLA	SLA >= 16

Second stage

Significance key: *** = .1%, ** = 1%, * = 5%, . = 10%.

Table 3.7: Second stage results – Mothers.

Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(Intercept)	0.229 (0.075) **	0.359 (0.101) ***	0.467 (0.109) ***	0.484 (0.084) ***	0.197 (0.074) **	0.261 (0.072) ***	0.288 (0.107) **	-0.306 (1.71)	0.264 (0.112) *
$\hat{e}du_{parent}$	0.035 (0.096)	0.019 (0.131)	0.046 (0.141)	-0.232 (0.108) *	0.039 (0.096)	0.003 (0.094)	-0.042 (0.138)	0.032 (0.098)	-0.024 (0.143)
sex	0.121 (0.005) ***			0.117 (0.006) ***	0.115 (0.005) ***	0.122 (0.005) ***	0.122 (0.005) ***	0.121 (0.006) ***	0.127 (0.007) ***
(Year of birth) ¹	0.001 (0.002)	0.001 (0.003)	0.001 (0.003)	0.006 (0.002) *	-0.002 (0.002)	0.001 (0.002)	0.004 (0.004)	-0.002 (0.01)	0.003 (0.004)
(Year of birth) ²	0 (0) **	0 (0) **	0 (0)	0 (0) .	0 (0) *		0 (0) *	0 (0) *	0 (0)
(Year of birth) ³							0 (0)		
Number of observations	35402	18105	17297	27392	35402	35402	35402	35402	21211
Dependent Variable	Seconde	Seconde	Seconde	Brevet	Both	Seconde	Seconde	Seconde	Seconde
Window (years)	+/- 10	+/- 10	+/- 10	+/- 10	+/- 10	+/- 10	+/- 10	+/- 10	+/- 5
Child gender	Both	Boy	Girl	Both	Both	Both	Both	Both	Both
Endogenous variable	SLA >= 16	SLA >= 16	SLA >= 16	SLA >= 16	SLA >= 16	SLA >= 16	SLA >= 16	SLA	SLA >= 16

Second stage

Significance key: *** = .1%, ** = 1%, * = 5%, . = 10%.

Table 3.8: Wald estimation results (fathers)

Child gender	Boy	Girl	Any
1 year radius	0.231 (0.173)	0.413 (0.239) .	0.362 (0.151) *
2 year radius	-0.026 (0.195)	0.408 (0.166) *	0.179 (0.12)
3 year radius	0.109 (0.095)	0.228 (0.076) **	0.192 (0.084) *

Wald estimator.

Significance key: *** = .1%, ** = 1%, * = 5%, . = 10%.

Table 3.9: Wald estimation results (mothers)

Child gender	Boy	Girl	Any
1 year radius	-0.11 (0.173)	-0.237 (0.21)	-0.126 (0.111)
2 year radius	-0.052 (0.123)	-0.086 (0.107)	-0.057 (0.093)
3 year radius	0.019 (0.057)	0.025 (0.078)	0.024 (0.066)

Wald estimator.

Significance key: *** = .1%, ** = 1%, * = 5%, . = 10%.

child gender and window specifications) the probability for a man's children to be in Seconde by age 16. The effect seems particularly strong for girls.

For mothers, we find no evidence of any effect. If anything, point estimates are slightly negative for 1- and 2-year windows, but remain far from conventional significance levels.

3.5 Conclusion

We find evidence that paternal education has a strong causal effect on the education of children, considerably affecting the likelihood of grade repetition by age 16. Interestingly, as previous studies have convincingly made the point that the Berthoin reform had a strong effect on years of schooling, yet hardly any on diploma or future earnings, the causal effect that we identify is truly attributable directly to education rather than, for example, to byproducts of an increased education, such as career success.

That we only found a significant effect for fathers, and none for mothers, is difficult to explain. Nonetheless, while maternal education is not found to significantly

affect the education of the child, neither is its effect found to be significantly different from that of paternal education. Therefore, it would likely be unwise to attempt to explain why paternal education could thus affect the child's education in a different way from maternal education.

These results imply that policies such as the Berthoin reform are likely to be more successful in the very long run than early evaluations (based on early career or income) are likely to estimate. Indeed, our study uses recent data (most of our sample was surveyed between 1990 and 2010) to document the effects of a reform that occurred more than forty years ago.

The channels through which this causal effect expresses itself are beyond the scope of this study, but may include such factors as a greater understanding of the role of the educational system and of the type of work that is expected from successful pupils, both of which may be passed on to the child regardless of the parent's diplomas or income. A lot of recent evidence suggests that parental attitudes to their offspring's education and homework, as well as their degree of involvement, may have vast effects on educational outcomes (see, for example, *Avvisati et al., 2010*).

Another channel could be through homogamy, that is, the tendency for individuals to match with partners with similar backgrounds and education. As we treat the effect of parental education completely separately from the spouses, we cannot rule out this causal channel.

Chapter 4

Employer learning in presence of general and match-specific human capital

4.1 Introduction

Information imperfections play a central role in the labour market. In particular, they contribute to matching frictions of at least two types. First, unemployed workers and vacant jobs coexist at any point in time; yet in most cases, unemployed workers are capable of contributing a non-negative production value. Second, observed matches of workers to firms are often not optimal, and a reallocation of workers to firm may increase aggregate output. One interpretation is that productivity is at least partially match-specific and that there is some uncertainty about this match-specific productivity - uncertainty which is, at best, costly to resolve.

This paper explores the assumption that the productivity of a worker in a given match is not entirely specific to this match. In other words, a certain amount of this productivity would be transferred to the new match, should the worker decide

(or be led to) change jobs. I think of this situation as one where productivity has both a general (transferable) and a specific (non-transferable) component. Total productivity on a job is, then, assumed to be an increasing function of each of these two components.

The relative importance of general productivity and match-specific productivity is likely to vary across sectors of the labour market. If the focus is on careers and mobility between relatively substitutable jobs, e.g. lecturer in economics in different institutions, general productivity will be relevant to the skills that are necessary for this type of job in any university (pedagogy, insight, motivation). Specific productivity will likely be narrower, and include for example peer-effects from department-specific colleagues. In this situation, it seems that the general component for productivity will amount to most of the productivity in the match¹. In a way, this “general” productivity is already specific to this segment of the labour market.

However, one can be interested in mobility between occupations requiring skills that overlap less evidently. In that case, the transferable part of human capital is expected to amount to a relatively lower share of the total productivity. One can think of many potential sources of variations of productivity across matches. Different workers may react differently to pressure at work; some other worker characteristics, e.g. interpersonal skills, are valued differently across jobs. Worker intrinsic motivation is also likely to vary across jobs and affect productivity.

The fundamental assumption made in this paper is that the general productivity of the worker is unknown *ex ante*, as well as his job specific productivity in any job available to him. Once on a job, whilst the worker’s total productivity is perfectly observed by all the agents in the economy, it is not known how this productivity is split between its general and specific components. In other words, there is uncertainty

¹In other words, and to be consistent with the framework developed later, the variance of the specific component is small compared to that of the general component.

on the share of productivity which can be transferred to the next job, and that which is specific to the current job. This assumption implies that the only way for the economy to obtain information on a worker's general productivity is to observe him experiment enough jobs to extract this information with sufficient clarity.

Much of the literature on employer learning on the labour market focuses on information on match-specific productivity. The seminal paper is Jovanovic (1979). In this paper, the worker and his firm learn continuously about the quality of the current match until the match is broken. Since all human capital is match-specific and the worker lives forever, the value of quitting does not depend on history, and it is shown that the decision to quit depends only on whether the current productivity is above or below a reservation productivity, which itself is a function only of tenure on the current job. More recent papers featuring uncertainty on job-specific productivity include, for example, Moscarini (2005) and Felli and Harris (1996). In this paper, by contrast, the turnover history of the worker affects his value of quitting by updating the public information on his general human productivity.

Another branch of the learning literature focuses on learning about general productivity. In a working paper, Eeckhout (2006) focuses explicitly on the case where all the human capital is general, that is, the worker is expected to have the same productivity in all jobs. Other papers, such as Farber and Gibbons (1996) or Altonji and Pierret (2001), implicitly study learning about general productivity, in that the worker is endowed with a certain unobserved innate ability, which affects his productivity on all matches. In the case where the worker has better information about his ability than prospective employers, studying uncertainty on general human capital allows for interesting discussions on the value for the worker of signalling his general human capital through education, *à la* Spence (1973). That is the focus of both Farber and Gibbons (1996) and Altonji and Pierret (2001), whereas Eeckhout (2006) concentrates on mobility and competition amongst employers.

In some cases, assuming that all human capital is transferable seems to be a better approximation of reality than the assumption that it is entirely match-specific (e.g., example of academics); the truth, as often, is likely to lie between these two extremes, which is the motivation of this paper².

While this study aims to focus on employer learning about the productivity of the worker (assumed to be constant over time), accumulation of human capital on the job (learning-by-doing) is another important phenomenon that has been the object of much interest in recent years. These two effects are not mutually exclusive³, although for the sake of tractability most analyses tend to focus on either one. An interesting exception is Nagypál (2006), which endeavours to distinguish and measure the two effects. Nagypál builds and estimates a model where (specific) learning-by-doing and learning about match quality coexist. Her estimates suggest that the relative strength of these two effects varies with tenure. Learning-by-doing may be present during the first few months of employment, but the dominating, longer-lasting effect seems to be learning about match quality. Moreover, Nagypál's estimates imply that learning about match quality leads to an increase in output of roughly 30% over a ten years horizon. This is a selection effect: matches that are revealed to be unproductive are discontinued earlier, and only the most productive remain.

My assumption that the total productivity of a worker on a match is immediately observable by all agents may seem strong, as it implies an infinite speed of learning about match quality. However, one way to rationalize this is to interpret each period as long enough for this assumption to be justified. For example (in the three-period version of the model) if the first period is to be interpreted as the entry of the

²This feature (coexistence of general and specific human capital) comes at the cost of strong assumptions on other respects, e.g. that workers themselves do not know their general human capital. Tractability of the results is also limited; in fact, only the three-period model yields closed-form solutions. For the extension of the model to an infinite number of periods, the inquiry is aided by computer simulations.

³In fact, some authors consider employer learning to be itself a form of human capital accumulation; see, for example, Felli and Harris (2004).

young worker on the labour market, and the third period as immediately followed by retirement, then each period can actually be seen as summarizing more than ten years, which is more than enough to learn perfectly about match quality⁴. According to Lange (2007), after three years, the employer's initial expectation errors about match-quality have declined already by 50%. One contribution of this paper is to raise the point that perfect knowledge about the worker's productivity on a given match does not say anything about how much of this productivity is transferable to other potential matches, leading (among others) to potential inefficiencies.

This paper explores some implications of learning on both general and specific human capital, first in a three-period model and then in an infinite-period model. In the three-period case, I first assume that the worker is always paid exactly his productivity. In this case, the worker turnover decisions are efficient, as they optimize his expected productivity, and thus his incentives are aligned with those of a social planner with the same (limited) information set. I derive the turnover decision of the worker in this context, and show that he will be slightly more prone to change job after the first period if his initial information on general productivity is imprecise. I then assume that the firms (all assumed to observe the job history of the worker) make simultaneous wage offers at each period. Although his rewards are modified, the worker's turnover behaviour in each period is unaltered compared to the previous case. Competition between potential new employers guarantees that the outside wage paid to the worker, if he moves, equals his expected productivity. The current employer can try to extract rent from the worker's match-specific productivity, as long as it can afford to pay at least the outside wage.

The case of an infinite number of periods is also explored. Closed form solutions are not tractable, but the analysis lends itself to computer simulations, the results

⁴For the same reason (and for simplicity), it is assumed here that learning-by-doing is negligible; hence it is not included in the model. This is consistent with Nagypál's results.

of which I provide. It is found that many of the insights from the three-period case carry over; in particular, wage inequalities arise across workers that are ex ante identical. Depending on specifications, wage dispersion can widen with age, but at a decreasing pace, which matches empirical observations. It is also found that the rate of mobility decreases over time. Finally, the model predicts that, at any given time (past the first few periods), the correlation between workers' general human capital and match-specific human capital is negative. This is mainly driven by the fact that low ability workers move sufficiently often during early periods to find a match that is comparatively good. By contrast, high-ability workers tend to settle early for jobs that are not as good for them as they believe; they dare not move, for fear of getting a worse match.

4.2 The three-period model

4.2.1 Informational structure and worker mobility

I consider one worker facing an infinity of job vacancies, issued by separate firms competing on the market for labour, indexed by $i \in \{1, \dots, +\infty\}$. The ex ante characteristics of job vacancies are identical. The worker is endowed with a general productivity, denoted as g , and match-specific productivities corresponding to each of the job vacancies, denoted as s_i . These quantities are unobserved, but all agents (worker and firms) have the same beliefs about them. There is no information asymmetry.

Ex ante, g is believed to follow a normal distribution with mean m_g and precision h_g ⁵. Similarly, for all i , s_i is normally distributed with mean 0 and precision h_s ⁶. g and all the s_i are believed to be independent from each other.

⁵Following the convenient convention in Bayesian work, I call "precision" the inverse of the variance

⁶Throughout the paper, I will use the notation h for precisions, and σ^2 for variances.

The prior belief on g can be interpreted as deriving from all the signals available about g before the worker enters his first job - i.e., signals resulting from the education of the worker, interview skills, etc.

At the beginning of each period, the worker is free to choose to work in any of the jobs, but cannot go back to a job he left in the past. Since all the jobs are equivalent *ex ante*, this means that in fact, his choice is between staying on the current match, or changing job. The production x_i of the period is perfectly observed, and is known to satisfy $x_i = g + s_i$. That is, total productivity is the sum of general productivity and match-specific productivity.

I assume, for now, that the wage paid to the worker is equal to his total productivity. This ignores the fact that profit-maximizing firms will try to retain at least part of the rent, especially if the match-specific (non transferable) component is believed to be high. I will depart from this benchmark later, when I allow for strategic wage-setting.

Studying the turnover behaviour of a worker being paid his full productivity is a way to find the efficient turnover, maximizing social surplus, of a more general model where the rent is somehow shared between the firm and the worker. The worker, in this benchmark, internalizes all the effect of his turnover. He gets paid his expected total productivity at the beginning of each period ⁷.

Upon changing jobs, a new s_i is drawn, and x_i , its sum with g , is perfectly observed. This allows the worker and the firms to update their belief on s_i and (more crucially maybe) on g .

The worker's objective is to maximize his intertemporal utility, which is assumed

⁷Assuming, as I do, that all agents are risk neutral, this last assumption is equivalent to the alternative where the worker is paid at the end of the period, once production has been observed - or even to an intermediate possibility (not developed here) where productivity within a match is learnt progressively during each period, and wages follow the evolution of beliefs on total productivity. However for this last context to be consistent with the other assumptions about learning between periods, it still needs to be the case that total productivity is perfectly known by the time one period ends.

to be the sum of his discounted wages. The interest rate is denoted as r .

The worker lives only for three periods. He can change jobs after the first and second periods.

I henceforth use subscripts to denote the period. For all t in $\{1, 2, 3\}$, the worker's productivity (perfectly observed) is denoted as x_t and the two components of his productivity are believed to be normally distributed with means μ_t^g and μ_t^s and precision h_t ⁸.

Theorem 4.2.1 *If the economy's belief on g is normally distributed with mean μ_t^g and precision h_t and the worker starts job j where his productivity is revealed to be x , then the economy's updated belief on g is normally distributed with mean $\frac{\mu_t^g h_t + x h_s}{h_t + h_s}$ and precision $h_s + h_t$.*

Proof It is known that if x is normally distributed with known precision h but unknown mean μ , and the prior for μ is itself normally distributed with mean μ_0 and precision h_0 , then the posterior for μ upon observing x is normally distributed with mean $\frac{\mu_0 h_0 + h x}{h_0 + h}$ and precision $h_0 + h$.

When the worker starts a new job where his productivity (distributed normally with mean g and precision h_s) is revealed to be x , the posterior distribution for g therefore becomes $\frac{\mu_t^g h_t + x h_s}{h_t + h_s}$ with precision $h_s + h_t$. \square

This theorem, applied as many times as necessary, gives us the distribution of the economy's belief once past matches are observed and accounted for. A corollary is:

Corollary 4.2.2 *The economy's belief on g after n matches have been experienced is normally distributed with precision $H_n = h_g + n h_s$, or equivalently variance $S_n = \frac{1}{H(n)}$.*

⁸Since $x_t = \mu_t^g + \mu_t^s$ and x_t is observed, it needs indeed be the case that beliefs on μ_t^g and μ_t^s have the same variance.

Behaviour in the second period

The worker gets payoff x_2 in the second period and decides whether to change jobs or not for his third period. He will change jobs if his expected payoff from doing so is greater than x_2 , which he can get for sure by not breaking the current match. That is, the worker moves if $\mu_2^g > x_2$, or $0 > \mu_2^s$. An expression for μ_2^s is $\frac{(x_2 - \mu_1^g)\sigma_s^2}{\sigma_s^2 + S_1^2}$.

The present discounted value of the payoff in the second period is therefore a function of x_2 which I shall denote as $\Pi_2(x_2)$:

$$\begin{aligned}\Pi_2(x_2) &= x_2 + \frac{1}{1+r} \text{Max} \left(x_2, \frac{\mu_1^g \sigma_s^2 + x_2 S_1^2}{\sigma_s^2 + S_1^2} \right) \\ &= x_2 \left(1 + \frac{1}{1+r} \right) + \frac{1}{1+r} \text{Max} \left(0, \frac{(\mu_1^g - x_2)\sigma_s^2}{\sigma_s^2 + S_1^2} \right)\end{aligned}\tag{4.1}$$

Behaviour in the first period

In the first period, the worker gets payoff x_1 and decides whether to change jobs or not. This decision depends on how his expected discounted payoffs compare in both cases. If he does not change jobs, he receives payoff $(1 + \frac{1}{1+r} + \frac{1}{(1+r)^2})x_1$ ⁹. If he changes jobs, he receives payoff $x_1 + \frac{1}{1+r}E_1(\Pi_2(x_2))$, where E_1 denotes expected value given the information available (i.e. beliefs held) at the end of period 1.

⁹It is easy to show that if the worker does not move today, he will not move tomorrow either. This is because the expected reward for moving is larger after period 1 than after period 2. Indeed, in the case of mobility after period 2, the expected return from moving is simply the difference between the current productivity and the expected productivity on the new match. On the contrary, in the case of mobility after period 1, if the new match is a good match, it can be enjoyed for two periods; if it is a bad match, it is possible to move again in period three. That is, the possibility of moving again after period 2 makes moving after period 1 more appealing.

Theorem 4.2.3 *The worker decides to change jobs in the first period if and only if*

$$\mu_1^s \leq \frac{1}{2+r} \frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}} \quad (4.2)$$

or equivalently

$$x_1 - m_g \leq \frac{1}{2+r} \frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2 + \sigma_g^2}{\sqrt{\sigma_s^2 + S_1^2}} \quad (4.3)$$

Proof Given that $E_1(x_2) = \mu_1^g$ and that $\mu_1^g - x_2$ is believed to be normally distributed with mean 0 and variance $S_1^2 + \sigma_s^2$, it can be proved¹⁰ that $E_1(\Pi_2(x_2)) = \left(1 + \frac{1}{1+r}\right) \mu_1^g + \frac{1}{1+r} \frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}}$. Therefore, the worker's expected payoff if he moves is believed to be $x_1 + \frac{1}{1+r} \left(\left(1 + \frac{1}{1+r}\right) \mu_1^g + \frac{1}{1+r} \frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}} \right)$.

Comparing this expected return with the return of staying with the current employer straightforwardly yields the theorem. \square

The right-hand side in 4.3 can be shown to be always increasing in σ_g^2 . Its variations with respect to σ_s^2 are non-monotonic. It decreases until $\sigma_s^2 \leq (\sqrt{3} - 1)\sigma_g^2$, and increases ever after.

In other words, *ceteris paribus* (in particular, for a given x_1 higher than m_g ¹¹), I learn three things about the worker's mobility *given his first draw* x_1 . (1) The worker tends to move more the higher the initial uncertainty on his general productivity is. (2) If the draws of match-specific productivity are very precise, then reducing this precision tends to make the worker move less. This second effect is more pronounced when there is high initial uncertainty on general productivity. And (3), if the draws of match-specific productivity are rather imprecise, then reducing this precision tends to make the worker move more.

¹⁰Using the property that, if z is a standard normal, $E(z|z > 0) = \frac{2}{\sqrt{2\pi}}$, which implies $E(\text{Max}(0, z)) = \frac{1}{\sqrt{2\pi}}$.

¹¹If $x_1 < m_g$ the worker unambiguously moves, as his μ_1^g is greater than x_1 (being a weighted average of x_1 and m_g), which means that he believes his s_1 to be negative.

A few comments may be in order regarding each of these effects. (1) is simply a consequence of the fact that if there is high initial uncertainty on general productivity, then the first match has a large influence on the way the economy's belief is updated. Remember that I restrict this discussion to the case where x_1 is larger than m_g , which means that the belief on general productivity is updated upwards. At the limit, when σ_g^2 goes to infinity, the new belief approaches x_1 , and then the worker will want to move no matter how x_1 compares to m_g . (2) works through a similar mechanism. If $x_1 \geq m_g$, and σ_g^2 is quite high compared to the initial σ_s^2 , then when σ_s^2 increases, the updated belief on general productivity (μ_1^g) shifts away from x_1 and towards m_g , and thus decreases. Then the current match is perceived as a better match than before the σ_s^2 increase, and the worker's willingness to move decreases. This is a rather small effect, but it can dominate over (3) in certain situations. The intuition behind (3) is as follows: if $x_1 \geq m_g$, and σ_g^2 is relatively low compared to the initial σ_s^2 , then when σ_s^2 increases, the most important effect is that the distribution of future draws becomes more variable. This means that exceptionally high draws become more common. (As do exceptionally low draws, of course, but at this point the worker knows that there will be one more opportunity to move in the future - so a low draw will only be imposed on the worker for one period, whereas a high draw will be enjoyed twice.) Therefore when σ_s^2 increases, the worker's incentives to move go up.

The previous discussion was all conditional on the first draw x_1 . It is quite straightforward to show that prior to this first draw, the probability of the worker moving after the first period is $\Phi\left(\frac{1}{2+r} \frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2 + \sigma_g^2}{\sigma_s \sqrt{\sigma_s^2 + 2\sigma_g^2}}\right)$, which is everywhere decreasing in σ_s^2 and increasing in σ_g^2 . That is, the worker's unconditional probability of moving increases when his initial information on general human capital becomes more imprecise, or when the signals he receives on each new job become more precise - ie, when the speed of learning increases. Effect (2) now dominates over (3) everywhere.

Let us now condition the probability of moving on the general human capital of the worker g (e.g. to study mobility of high human capital workers, keeping in mind that they do not know that their general human capital is high). I then obtain a probability of moving after the end of the first period equal to $\Phi\left(\frac{1}{2+r} \frac{1}{\sqrt{2\pi}} \frac{(\sigma_s^2 + \sigma_g^2)^{\frac{3}{2}}}{\sigma_s^2 \sqrt{\sigma_s^2 + 2\sigma_g^2}} - \frac{g - m_g}{\sigma_s}\right)$. This probability is increasing in σ_g^2 and varies ambiguously with σ_s^2 . Besides, it is obviously decreasing in g : workers with high human capital will tend to attribute part of their good productivity in their first job to the match-specific component of their productivity, thus over-estimating the quality of their current match and under-estimating the returns to mobility. This prediction does not seem to fit the data very well¹². I believe that this can be solved by introducing asymmetric information (which was not done in this paper, as signalling considerations arise which complicate the tractability considerably). Assuming indeed that the worker is aware of his high human capital, while the firms are not, could increase the incentives to move for high general human capital workers, whilst reducing that of low general human capital workers.

4.2.2 Equilibrium wages

In this subsection, I allow for firms to make wage offers to the worker at the beginning of each period. This contrasts with the previous version of the model, where wages were assumed to equal expected total productivity instead of being set strategically by firms. Firms now compete with each other to attract the worker. I will derive the

¹²Although it is hard to pick up and may call for a careful empirical study, there does not appear to be an important effect of ability on mobility of workers in the data.

equilibrium wage offers for the last two periods¹³.

I shall refer to the firm at which the worker is currently working as the *current* firm, and to the firms at which he could be working after breaking the current match as the *other* firms.

Third period

Between the second and third periods, the other firms will, in equilibrium, offer to the worker a wage of μ_2^g (his expected productivity). If they offer more, they expect to make losses, and if they offer less each of them would have an incentive to deviate and offer slightly more, thus securing that if the worker moves, then he will move to that firm.

The current firm will match that offer if and only if $x_2 > \mu_2^g$. In that case, I assume that the worker will choose to stay with that firm (imagine that the firm pays a very small ϵ on top of μ_2^g to make sure the worker stays). Otherwise it will offer x_2 ¹⁴.

Then the worker moves at the beginning of the third period if and only if $x_2 < \mu_2^g$.

Second period

Theorem 4.2.4 *Between the first and second periods,*

¹³In the first period, all firms make the same wage offer, the algebraic formulation of which too complex to comment. For the curious reader, the said expression is

$$w = m_g + \frac{1}{1+r} \left[(2+r) \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + \sigma_g^2}} \phi \left(\frac{1}{2+r} \frac{1}{\sqrt{2\pi} \sigma_s \sqrt{\sigma_s^2 + 2\sigma_g^2}} \right) - \frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2 \sqrt{\sigma_s^2 + \sigma_g^2}}{\sqrt{\sigma_s^2 + 2\sigma_g^2}} \left(1 - \Phi \left(\frac{1}{2+r} \frac{1}{\sqrt{2\pi} \sigma_s \sqrt{\sigma_s^2 + 2\sigma_g^2}} \right) \right) \right]$$

¹⁴In fact this is only one out of a continuum of possible equilibria. If $x_2 < \mu_2^g$, the current firm offering any wage strictly below μ_2^g is an equilibrium. Likewise, if $x_2 > \mu_2^g$, all that is needed for there to be an equilibrium is that at least one of the other firms offers μ_2^g . The other firms are indifferent between any wage offer below μ_2^g .

- If $(x_1 - \mu_1^g) < \frac{1}{2+r} \frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}}$, then the worker moves. The current firm offers x_1 ¹⁵. The other firms offer $\mu_1^g + \frac{1}{1+r} \left(\frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}} \right)$.
- If $(x_1 - \mu_1^g) \geq \frac{1}{2+r} \frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}}$ then the worker stays. All firms offer¹⁶ $\mu_1^g + \frac{1}{1+r} \left(\frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}} \right)$.

Proof Let us denote as $V_c(x_1)$ and $V_o(x_1)$, respectively, the wages offered in (an) equilibrium by the current firm and the other firms. Then to ensure positive expected profit of the current firm, we need $V_c(x_1) \leq x_1 + \frac{1}{1+r} 1_{x_1 \geq \mu_1^g} (x_1 - \mu_1^g)$. Likewise, we need, for the other firms, $V_o(x_1) \leq \mu_1^g + \frac{1}{1+r} \frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}}$, and because of competition between firms the inequality needs to be an equality for all firms in equilibrium¹⁷.

If the worker stays, he will receive $V_c(x_1) + \frac{1}{1+r} \mu_1^g$. If he moves, he expects to receive $V_o(x_1) + \frac{1}{1+r} E_1(\mu_2^g) = \mu_1^g + \frac{1}{1+r} \left[\mu_1^g + \left(\frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}} \right) \right]$. He will therefore move if and only if $V_c(x_1) < \mu_1^g + \frac{1}{1+r} \left(\frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}} \right)$.

Given this behaviour of the worker, the optimal wage for the current firm to set is $\mu_1^g + \frac{1}{1+r} \left(\frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}} \right)$, but this wage will guarantee positive profits only if it is lower than $x_1 + \frac{1}{1+r} 1_{x_1 \geq \mu_1^g} (x_1 - \mu_1^g)$, which is only possible if $x_1 > \mu_1^g$. Therefore the condition for the worker to stay is indeed $(x_1 - \mu_1^g) \geq \frac{1}{2+r} \frac{1}{\sqrt{2\pi}} \frac{\sigma_s^2}{\sqrt{\sigma_s^2 + S_1^2}}$

□

Efficiency

The analysis above indicates that the turnover decisions of the worker depend on the same conditions in the case of firm setting their wages as on that of the wages being set equal to the productivity. That is to say, the mobility of the worker is efficient even in the case where firms set their wages.

¹⁵For example. Anything below the other firms' offer would give another, equivalent equilibrium.

¹⁶Or at least the current firm and one other firm do. The other firms can make lower offers and yet sustain the equilibrium, as long as at least one offers exactly that wage.

¹⁷At least when the offer of the current firm is lower, i.e. when the worker moves.

Efficiency of turnover between the second and third periods is intuitive. The other firms make offers equal to the worker's expected productivity on the new match, and the current firm only matches this offer if the worker's total productivity on the current job is higher than the outside offers: therefore the worker will move if and only if moving increases his expected output in the third period.

Efficiency between the first and second periods can be explained as follows. Due to competition between other firms, the wage paid to the worker if he leaves his current job is equal to his expected productivity on the new job, plus the expected potential rent that can be extracted in period 3. His current employer can only match offers up to his productivity on the current job, plus the rent it will extract in period 3. Therefore, the incentives faced by the worker are exactly the same as before¹⁸.

Note that this result is sensitive to the assumption that the worker is risk neutral. If we were to make instead the assumption that the worker is risk averse, then he would tend to move less when he is paid his exact productivity at the end of each period than if the (risk neutral) firms can pay him his expected productivity at the beginning of each period. But then his turnover would still be efficient in the case of wage-setting firms. It is in the case of the wage set equal to productivity that turnover would be inefficiently low.

In environments of imperfect but symmetric information like this one, it is not unusual to observe such efficient turnover behaviour when the firms set wages. In Felli and Harris (1996) and Jovanovic (1979), it is indeed the case that the worker's turnover decisions maximize social surplus given the information available in the economy. One paper that differs is Felli and Harris (2004). By introducing firm-specific training as a choice variable of the firms, the authors introduce inefficiencies in

¹⁸There is only a difference in timing, with wages paid in period 2 anticipating productivity in period 3.

their model, which in all variations of their baseline model translate into inefficiently low turnover.

Rent sharing, general and specific productivity

The worker invariably gets all of what is believed to be the return from his general productivity. He also receives a positive bonus corresponding to the rent the firm expects to be able to extract from match-specific productivity in subsequent periods. This result is similar to those in Eeckhout (2006)¹⁹.

However, the particular realization of match-specific productivity is a risk supported entirely by the firm. It may take a high value that no other firm can expect to enjoy, and that the current firm can enjoy for several periods; or a low value which will lead to the immediate break of the match. As a consequence of competition, firms expect to make zero profits.

Let us have a look at how the return from a high general productivity is shared between the firms and the worker. By looking at the wage equations, it becomes clear that high general productivity workers only get rent from their high ability through the economy's belief of it. That is, initially all workers are paid the same wage (first period), after which high productivity workers, in general, will be believed to be more able (in the sense of general productivity) and will get added returns from it. The opposite is true for workers with low general productivity, who can thus enjoy benefits from having their productivity over-estimated by the economy. The latter effect is stronger when h_s is low and h_g is high (that is, when updating is slow). However, since the workers do not have more information than the firms, they cannot adapt their turnover behaviour to this. If they did have more information,

¹⁹However my assumptions differ in many dimensions. Eeckhout assumes that employers have asymmetric information about the worker's general human capital, with the current employer having access to a better signal; moreover, the wage setting rule is a second price auction. Yet Eeckhout finds, like me, that employers bid up to the expected (general) productivity, plus all rents that can be extracted in the future.

we might expect (*ceteris paribus*, i.e. for a given level of total productivity) high productivity workers to move more to signal their type and get the rent from it, and low productivity workers to move less to hide their type and continue enjoying a rent they do not actually deserve²⁰.

4.3 Infinite-period model

4.3.1 Model

In this section I present a more general version of the model, where agents live and work for an infinite number of periods. Given the increased complexity of the problem, it becomes impossible to derive closed form formulae for the wage and the mobility decision; however numerical method can be applied, and I present the results of simulations for a wide variety of parameter values.

The basic assumptions and notations of the model are unchanged. Similarly to the three-period version, the worker is never going to move after not moving at least once in the past. This is because while on a given match, no new information is revealed, and the incentives that the worker faces are static. Hence we should expect every individual to move every period for a certain time, and then stop moving forever.

The present discounted value of future income can be represented by a time-invariant Bellman function, which I denote as Π , that takes as its arguments x (present production), μ^g (present belief on general human capital), and n (number of past moves, indicative of the precision of the present belief on g). With these notations, the present belief on g is $g \sim N(\mu^g, h_g + nh_s)$.

²⁰But this behaviour would, in turn, be informative to firms.

We then have:

$$\Pi(x, \mu^g, n) = x + \beta \text{Max} \left(\frac{x}{1 - \beta}, E_{x^*} V \left(x^*, \frac{\mu^g(h_g + nh_s) + x^*h_s}{(h_g + (n + 1)h_s)}, (n + 1) \right) \right)$$

where x^* is believed to follow a normal $N(\mu^g, \frac{h_t h_s}{h_t + h_s})$.

I compute Π numerically for different values of h_g , h_s (reflecting the relative importance of firm specific and general human capitals, as well as the speed of learning about general human capital) and of $\beta = \frac{1}{1+r}$ (reflecting the rate at which the future is discounted, which can also be interpreted as related to the life expectancy of the individual). The computational method used is as follows.

(1) Expectations are computed using Monte Carlo integration, using Halton sequences for increased efficiency²¹. This requires taking the mean of the function of which we wish to know the expected value, on a quasi-random sample designed in such a way as to cover the space more evenly than a truly random sample would. This results in greater accuracy, for a given computation time.

(2) Π is calculated recursively, iterating until convergence. The computational burden increases exponentially with the number of iterations, as each iteration calls the previous function as many times as the length of the Halton sequence used in the integration (I typically used a length of 100). To avoid this problem, I substitute Π with an approximation²² every two iterations until the function reaches the neighbourhood of its limit; and then every three iterations for the last few iterations.

(3) Iterations continue until the differences between two consecutive iterations become negligible.

Based on this approach, we can, for different values of the parameters, examine the destinies of cohorts of workers. Of particular interest are the evolution of the

²¹See, for example, Train (2009).

²²The approximation is made by computing a 3-dimensional, $20 \times 20 \times 20$ grid of values for x , μ_g and n ; values between grid nodes are approximated by a weighted average of the nearest nodes.

distribution of productivity and of beliefs about human capital over the lifetime, as well as the patterns of mobility.

4.3.2 Wages in the infinite model

I assume for simplicity that the worker is paid a wage equal to his production each period (ensuring efficient worker mobility). By contrast, allowing the firms to make discretionary offers to the worker (as studied in the three-period model) would result in wage offers that are a direct function of the current belief on g , independently of his current match-specific human capital²³.

More sophisticated contractual structures could be introduced (for example employment contracts that bind the firm and / or worker for several periods), but are beyond the scope of this study and are left as a topic for further research.

Numerical simulations make an inquiry into wage dispersion possible. In particular, this model features residual wage dispersion, as per Burdett and Mortensen (1998) and Postel-Vinay and Robin (2002): that is to say, workers that are initially thought identical (same m_g , h_g , h_s , h_r) can have very different careers, depending both on their (unobserved) general human capital g , and on the quality of the random matches that they experience throughout their career.

4.3.3 Numerical analysis

I analyse numerically the effect of the different parameters of the model on outcomes. The parameters under scrutiny are: (1) the precision of the initial belief on general human capital h_g ; (2) the precision of the match-specific productivity h_s ; (3) the precision of the random productivity shocks h_r ; and finally (4) the discount rate β .

The outcomes that I choose to examine are: (1) wage inequality, measured by

²³This is because current match-specific productivity is irrelevant from the point of view of the other firms, which define the outside option of the worker that the current firm has to match

the ratio between the 9th and the 1st decile of the wage distribution; (2) mobility, measured by the proportion of movers during a given period; (3) average productivity; (4) average belief on general human capital; (5) standard deviation of belief on general human capital; (6) number of past moves; (7) difference of number of past moves for population with g above median, to that for population with g below median; (8) the observed correlation between g and s in the matches that exist during any given period.

For each set of parameters that I simulate, I randomly generate a cohort of 100,000 workers whose productivity and mobility decisions are as described by the model. I simulate in this manner 36 (4×3^2) samples, one for each possible combination of (h_g, h_s, β) where the domain for h_g and h_s is $\{.6, .8, 1\}$ and the domain for β is $\{.8, .85, .9, .95\}$. These domains allow for a wide range of relative signal precisions, and a reasonably inclusive array of discount rates.

I examine the average effect of each parameter on each outcome over time, where the effect of parameter p at period t on outcome y is defined as being the coefficient on the parameter from a linear regression of outcome y on the four parameters for the simulations taken at period t .

Results are reported for the twelve first periods only, as that is where most of the learning and mobility take place given our parameter values.

Figure 4.1 shows how the outcomes respond to the parameters over time:

(a) Wage inequality is always fairly high; however it is decreasing over time. This is because low ability (low g) workers are more likely to move than high ability workers, and end up with a higher match-specific productivity than workers with a high g . After a number of moves, this results in a situation where g and s are negatively correlated (see (h)).

(b) The proportion of movers over time appears to be decreasing in most specifications; the speed of decrease is inversely related to the amount of employer learning

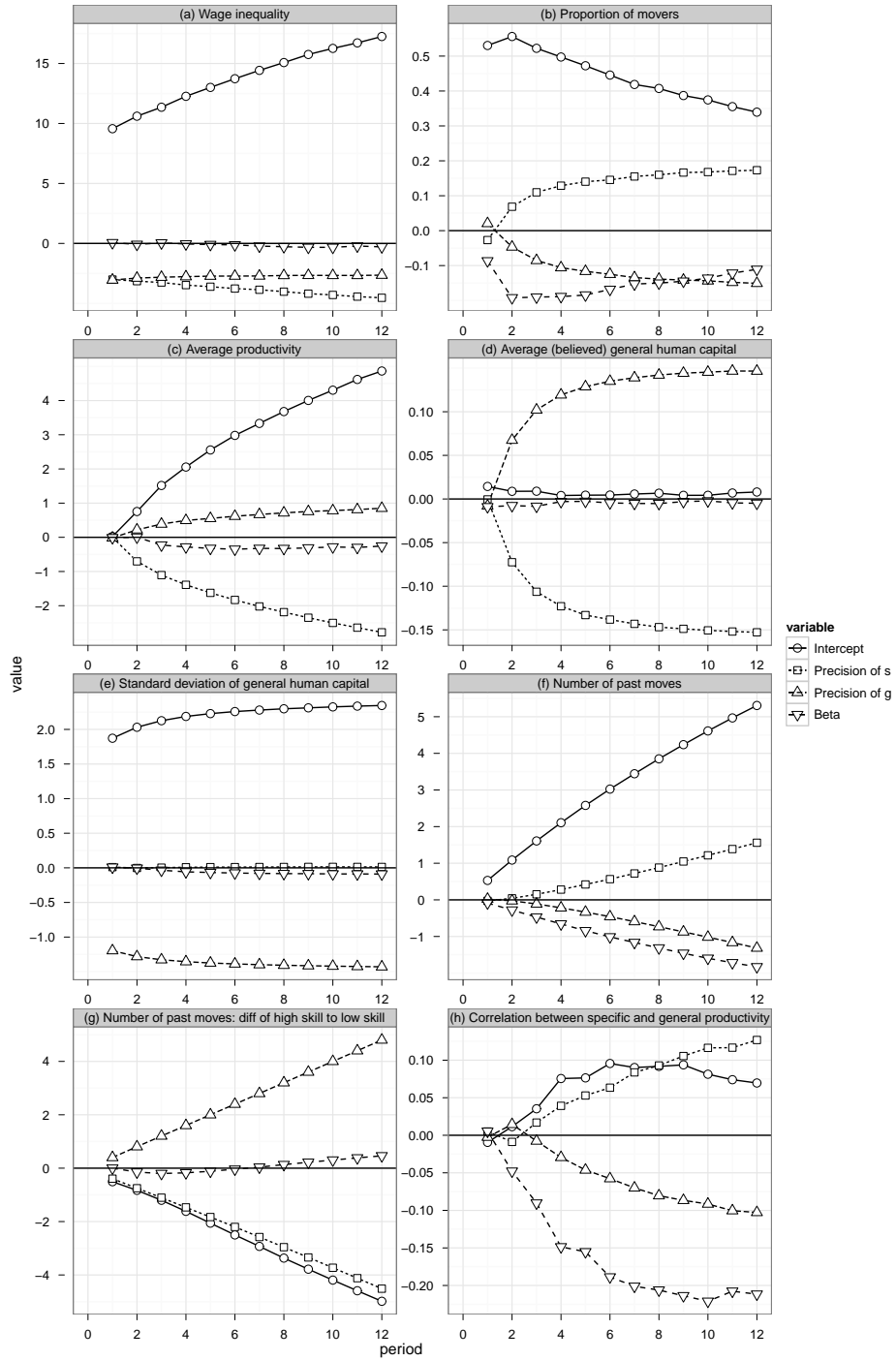


Figure 4.1: Effect of the parameters over time

to be expected in the labour market (i.e. high h_s , low h_g : each new match is very informative about the quality of the match). The proportion of movers in any period is also affected negatively by the discount factor β .

(c) Average productivity is globally increasing over time, though the increase is much slower where s is precise (and g imprecise, to a lesser extent). Where s is imprecise, many workers will benefit from an excellent draw (very productive job match), and keep that match for many periods; by contrast, the effect of terrible draws is exactly one period (and followed by another move).

(d) The average of the belief on general human capital is generally fairly close to 0. There are likely a number of effects at stake, the joint outcome of which is that high presence of learning (low h_g , high h_s) results in slightly worse average belief over general human capital.

(e) While the average belief is very close to zero, this conceals strong variability across individuals, reflected in the standard deviation. The range of beliefs increases somewhat over time, especially for low h_g (i.e. where there is potential for learning).

(f) The average number of past moves is another perspective on the proportion of movers per period, and the same patterns can be observed.

(g) It is found that in general, highly skilled workers (high g) move less, every period, than low skilled workers. This is especially the case where g is imprecise and s precise (meaning more difference between the average high skilled and the average low skilled workers). That is because workers with high g attribute some of their good luck on a probable high s (as they believe their g to be lower than it actually is). Their incentive to move is therefore lower than workers with low g , who, by contrast, will tend to blame their bad luck on a poor match.

(h) At any given period, s and g are found to be negatively correlated with each other. This is driven by the mobility pattern found in (g). Workers with low g move more and learn more about their human capital than workers with high g ; therefore

they become better able to recognize what is a good match for them and stick to it. This is especially the case with high h_g and low h_s , i.e. where learning is slow, and match-specific productivity has a high variance²⁴.

4.3.4 Additive productivity shocks

Numerical approaches bring increased flexibility. It is straightforward, for example, to introduce an additional source of uncertainty to the production of the worker. The aim is to address two inaccuracies of the model as described so far: (1) wage dispersion is predicted to decrease over time, which is at odds with empirical observations (see, for example, Lemieux (2006)); and (2) workers are predicted to move every period until they find a suitable employer, after which they never move anymore. While worker mobility is indeed observed empirically to decrease over the life cycle, this behaviour seems extreme.

In this section, an additive random walk is added to the productivity of the worker on the job every period after the first period of any match. This will be shown to solve both of those empirical inaccuracies.

The innovation of the random walk is denoted as r_t , and assumed to be normally distributed with mean 0 and precision h_r . This does not affect information about general human capital (and therefore, the outside wage offer), but does affect the current employer's reservation wage (and expected future wages at the current employer); therefore it also affects the mobility of the worker. A worker who experiences a lasting negative shock on his present match may now wish to switch jobs, even if the match had been satisfactory until now. This also implies that no match will last forever, as, given enough time, the probability that a random walk goes below any value can be observed to be arbitrarily close to one.

²⁴This increases what can be thought of, somewhat loosely, as the “reservation (believed) match-specific productivity”: indeed, a very good s becomes more common as h_s decreases

Results are reported in the same fashion as for the previous section, namely, regression coefficients are reported for each of the twelve first time periods. h_r is made to vary over the same domain as h_g and h_s , i.e. (h_g, h_s, h_r, β) ; this brings the number of simulated samples from 36 (previous section) to 108 (36×3).

Figure 4.2 shows how the outcomes respond to the parameters over time. Most panels are not remarkably affected by the introduction of the random productivity shocks; in particular the dynamics of learning about general capital are generally unchanged. However, two key features are changed.

First, panel (a) now indicates increasing wage dispersion over the lifecycle; a lower precision h_a (i.e. larger productivity shocks) corresponds to larger wage dispersion at all ages, and larger gradient of wage dispersion with respect to time period.

Second, workers now move more often, as indicated by panels (b) and (f); in particular, the drop in the number of movers over time is much flatter than in the previous section. Indeed, even after any arbitrary amount of time, the random shocks will still cause a number of workers to switch jobs.

4.4 Discussion and conclusion

In this section, I critically assess the model described in this paper, and discuss directions for further research.

On job mobility, I find declining rates of mobility with the life cycle, which is a well documented stylized fact (see, for example, Topel and Ward (1992)). This is mostly driven by the variations on job-specific productivity across jobs: workers keep moving until they find a match that they think is good enough for them.

On rent sharing, I find that the belief that the economy has on a worker's general productivity has an important impact on how much rent its employer can extract. A worker with high general productivity will only enjoy returns from it if all the firms

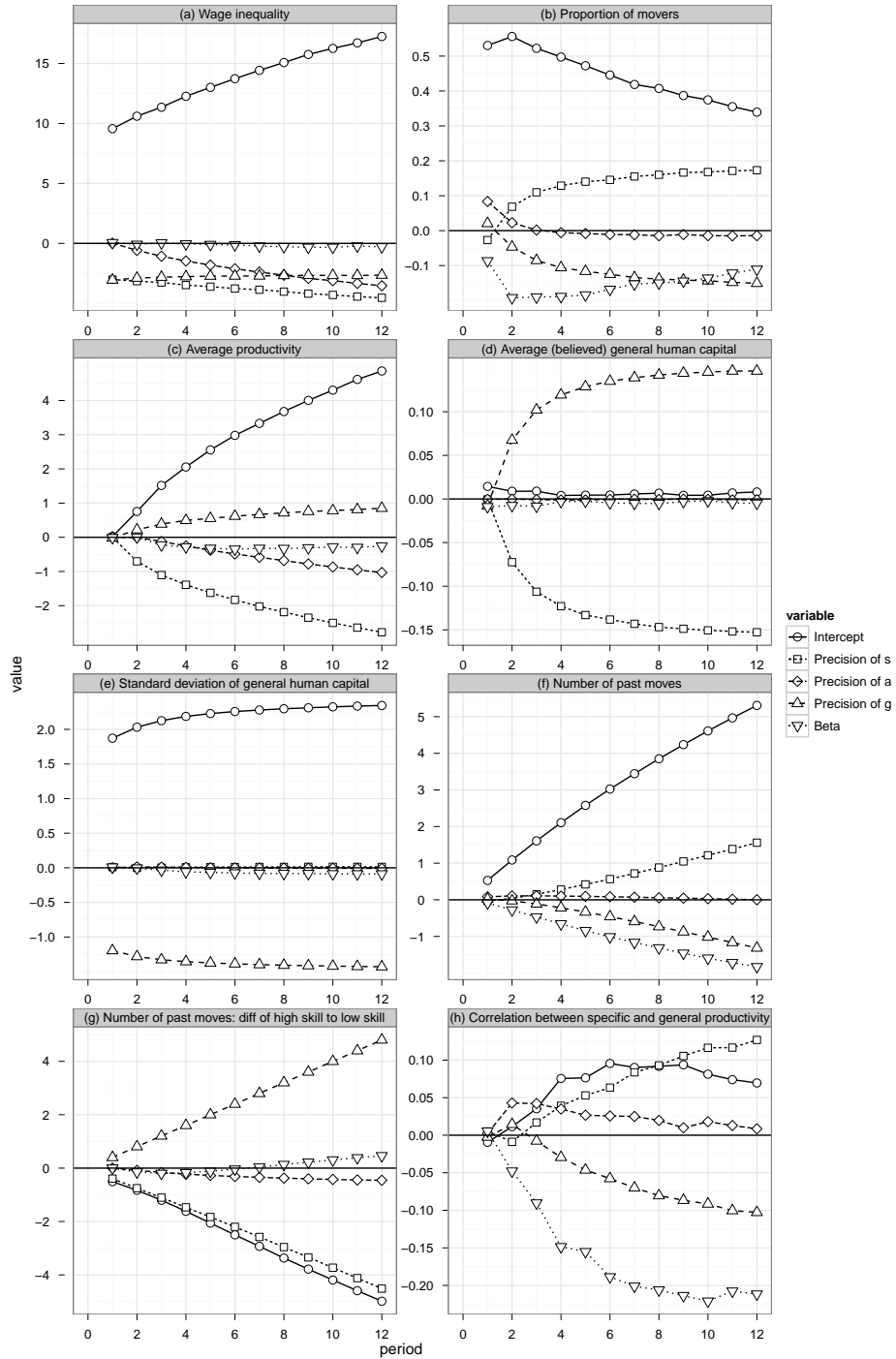


Figure 4.2: Effect of the parameters through time

are aware of this high productivity. This shows how much knowledge about general human capital matters in wage determination, and suggests that workers would be willing to strategically signal or hide their true general human capital if they had private information about it.

It seems reasonable to expect that, in reality, workers have some idea of their general human capital, through experiences that firms find hard to observe (e.g. grades obtained in high school, amount of efforts needed to obtain a degree, etc.). Asymmetric information such as this would likely affect some of the predictions of the model. For example, it would likely no longer be the case that high general productivity workers move less than less able workers. It may actually be the other way round, if mobility can be used as a signal for high general human capital²⁵. I believe that introducing asymmetric information would make the worker's turnover decisions inefficient (relative to his knowledge), precisely because of these signalling issues. High productivity worker's mobility may be inefficiently high, and that of low productivity workers inefficiently low.

Asymmetric information could also allow for endogenous education choices. For example, it may be assumed that education is less costly for high general human capital workers, who could therefore use it as another signalling device. The details of this, of course, are for further research to determine. The current model will provide a useful benchmark for future versions of this work, as it illustrates the extreme situation where information is always perfectly symmetric.

Further research avenues include placing this framework in a wider context, for example confronting it with data on careers (labour surveys, or longitudinal data).

²⁵But this will probably only be possible if the worker's information on his own general human capital is perfect or very precise: otherwise, the signalling effect will compete with the effect of the worker's uncertainty. The former pushes mobility up (the worker wants to signal that his general human capital is likely to be higher than the firms think), but the latter pushes it down (if the worker observes a high productivity, he will be tempted to interpret it partly as job-specific, and thus be reluctant to move).

Additional adjustments may have to be made to allow for (1) differences in education (which likely maps to the initial belief on general human capital, characterized by m_g and σ_g), and (2) on-the-job learning, which I excluded from this model. One simple way to incorporate on-the-job learning without complicating the informational structure would be to assume that every period spent on a match increases the productivity of the match by a constant value.

This model illustrates the need for theories of employer learning that takes into account the dual nature of uncertainty on the worker's productivity. Models that concentrate only on learning on general productivity often fail to feature realistic job mobility and residual wage dispersion; and models that concentrate on learning on match-specific productivity cannot feature signalling on worker general ability. Either way, these models can only offer a very narrow understanding of labour market dynamics.

Bibliography

- Albouy, V. and Lequien, L. (2009), ‘Does compulsory education lower mortality?’, *Journal of Health Economics* **28**(1), 155–168.
- Altonji, J. and Pierret, C. (2001), ‘Employer learning and statistical discrimination’, *The Quarterly Journal of Economics* **116**(1), 313–350.
- Angrist, J., Imbens, G. W. and Rubin, D. (1996), ‘Identification of causal effects using instrumental variables.’, *Journal of the American Statistical Association* **91**(434), 444–455.
- Avvisati, F., Gurgand, M., Guyon, N. and Maurin, E. (2010), Getting parents involved: A field experiment in deprived schools, CEPR Discussion Papers 8020, C.E.P.R. Discussion Papers.
- Baker, G., Gibbs, M. and Holmstrom, B. (1994), ‘The wage policy of a firm’, *The Quarterly Journal of Economics* **109**(4), 921–955.
- Bandiera, O., Barankay, I. and Rasul, I. (2009), ‘Social connections and incentives in the workplace: Evidence from personnel data’, *Econometrica* **77**(4), 1047–1094.
- Becker, G. S. (1962), ‘Investment in human capital: A theoretical analysis’, *Journal of Political Economy* **70**(5), pp. 9–49.

- Behrman, J. R. and Rosenzweig, M. R. (2002), 'Does increasing women's schooling raise the schooling of the next generation?', *American Economic Review* **92**(1), 323–334.
- Belzil, C. and Bognanno, M. (2008), 'Promotions, demotions, halo effects, and the earnings dynamics of american executives', *Journal of Labor Economics* **26**(2), 287–310.
- Belzil, C. and Hansen, J. (2003), 'Structural estimates of the intergenerational education correlation', *Journal of Applied Econometrics* **18**(6), 679–696.
- Black, S. E., Devereux, P. J. and Salvanes, K. G. (2005), 'Why the apple doesn't fall far: Understanding intergenerational transmission of human capital', *American Economic Review* **95**(1), 437–449.
- Bramoullé, Y. and Saint-Paul, G. (2004), Social networks and labour market transitions, CEPR Discussion Papers 4523, C.E.P.R. Discussion Papers.
- Brodaty, T., Gary-Bobo, R. J. and Prieto, A. (2008), Does speed signal ability? the impact of grade repetitions on employment and wages, CEPR Discussion Papers 6832, C.E.P.R. Discussion Papers.
- Burdett, K. and Mortensen, D. T. (1998), 'Wage differentials, employer size, and unemployment', *International Economic Review* **39**(2), 257–273.
- Cingano, F. and Rosolia, A. (2006), People i know: Workplace networks and job search outcomes, Temi di discussione (Economic working papers) 600, Bank of Italy, Economic Research Department.
- Eeckhout, J. (2006), 'Employer learning and general human capital', *Working Paper*

- Farber, H. and Gibbons, R. (1996), 'Learning and wage dynamics', *The Quarterly Journal of Economics* **111**(4), 1007–1047.
- Felli, L. and Harris, C. (1996), 'Learning, wage dynamics, and firm-specific human capital', *Journal of Political Economy* **104**(4), 838–868.
- Felli, L. and Harris, C. (2004), 'Firm-specific training', *Sticerd Discussion Paper* .
- Grenet, J. (2010), Is it enough to increase compulsory education to raise earnings? evidence from french and british compulsory schooling laws, Pse working paper series, PSE.
- Hahn, J., Todd, P. and Van der Klaauw, W. (2001), 'Identification and estimation of treatment effects with a regression-discontinuity design', *Econometrica* **69**(1), 201–09.
- Hausman, J. and McFadden, D. (1984), 'Specification tests for the multinomial logit model', *Econometrica* **52**(5), 1219–40.
- Holmlund, H., Lindahl, M. and Plug, E. (2010), The causal effect of parents' schooling on children's schooling - a comparison of estimation methods, CESifo Working Paper Series 3234, CESifo Group Munich.
- Imbens, G. and Lemieux, T. (2008), 'Special issue editors' introduction: The regression discontinuity design—theory and applications', *Journal of Econometrics* **142**(2), 611–614.
- Jovanovic, B. (1979), 'Job matching and the theory of turnover', *Journal of Political Economy* **87**(5), 972–990.
- Lange, F. (2007), 'The speed of employer learning', *Journal of Labor Economics* **25**(1), 1–35.

- Lazear, E. P. (1999), 'Personnel economics: Past lessons and future directions. presidential address to the society of labor economists, san francisco, may 1, 1998', *Journal of Labor Economics* **17**(2), iv–236.
- Lemieux, T. (2006), 'Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?', *The American Economic Review* **96**(3), pp. 461–498.
- Manacorda, M. (2010), The cost of grade retention, CEPR Discussion Papers 7889, C.E.P.R. Discussion Papers.
- Marmaros, D. and Sacerdote, B. (2006), 'How do friendships form?', *Quarterly Journal of Economics* **121**(1), 79–119.
- Maurin, E. (2002), 'The impact of parental income on early schooling transitions: A re-examination using data over three generations', *Journal of Public Economics* **85**(3), 301–332.
- Maurin, E. and McNally, S. (2008), 'Vive la révolution! long-term educational returns of 1968 to the angry students', *Journal of Labor Economics* **26**, 1–33.
- Montgomery, J. D. (1991), 'Social networks and labor-market outcomes: Toward an economic analysis', *American Economic Review* **81**(5), 1407–18.
- Moscarini, G. (2005), 'Job matching and the wage distribution', *Econometrica* **73**(2), 481–516.
- Nagypál, E. (2006), 'Learning-by-doing versus learning about match quality: Can we tell them apart?', *Working Paper* .
- Oreopoulos, P. (2006), 'Estimating average and local average treatment effects of education when compulsory schooling laws really matter', *American Economic Review* **96**(1), 152–175.

- Oreopoulos, P. and Page, M. E. (2006), 'The intergenerational effects of compulsory schooling', *Journal of Labor Economics* **24**(4), 729–760.
- Plug, E. (2004), 'Estimating the effect of mother's schooling on children's schooling using a sample of adoptees', *American Economic Review* **94**(1), 358–368.
- Postel-Vinay, F. and Robin, J.-M. (2002), 'Equilibrium wage dispersion with worker and employer heterogeneity', *Econometrica* **70**(6), 2295–2350.
- Reese, T. (1989), *Archbishop: Inside the Power Structure of the Catholic Church.*, Harper Collins Publishers.
- Reese, T. (1996), *Inside the Vatican: The Politics and Organization of the Catholic Church.*, The Woodstock Theological Center.
- Spence, M. (1973), 'Job market signaling', *The Quarterly Journal of Economics* **83**(3), 355–374.
- Topel, R. H. and Ward, M. P. (1992), 'Job mobility and the careers of young men', *The Quarterly Journal of Economics* **107**(2), 439–79.
- Train, K. (2009), *Discrete Choice Methods with Simulation (Second Edition)*, Cambridge University Press.
- Vatican (1983), *Code of Canon Law (1983)*, Libreria Editrice Vaticana.

Appendix to chapter 2

Canon Law of 1983²⁶: Canons 375 to 380 relative to the appointment of bishops

Canon 375

1 By divine institution, Bishops succeed the Apostles through the Holy Spirit who is given to them. They are constituted Pastors in the Church, to be the teachers of doctrine, the priests of sacred worship and the ministers of governance.

2 By their episcopal consecration, Bishops receive, together with the office of sanctifying, the offices also of teaching and of ruling, which however, by their nature, can be exercised only in hierarchical communion with the head of the College and its members.

Canon 376

Bishops to whom the care of a given diocese is entrusted are called diocesan Bishops; the others are called titular Bishops.

²⁶The previous code of Canon Law goes back to 1917; but the rules outlined here have been applied *de facto* throughout the period covered by my sample.

Canon 377

1 The Supreme Pontiff freely appoints Bishops or confirms those lawfully elected.

2 At least every three years, the Bishops of an ecclesiastical province or, if circumstances suggest it, of an Episcopal Conference, are to draw up, by common accord and in secret, a list of priests, even of members of institutes of consecrated life, who are suitable for the episcopate; they are to send this list to the Apostolic See. This is without prejudice to the right of every Bishop individually to make known to the Apostolic See the names of priests whom he thinks are worthy and suitable for the episcopal office.

3 Unless it has been lawfully prescribed otherwise, for the appointment of a diocesan Bishop or a coadjutor Bishop, a ternus, as it is called, is to be proposed to the Apostolic See. In the preparation of this list, it is the responsibility of the papal Legate to seek individually the suggestions of the Metropolitan and of the Suffragans of the province to which the diocese in question belongs or with which it is joined in some grouping, as well as the suggestions of the president of the Episcopal Conference. The papal Legate is, moreover, to hear the views of some members of the college of consultors and of the cathedral chapter. If he judges it expedient, he is also to seek individually, and in secret, the opinions of other clerics, both secular and religious, and of lay persons of outstanding wisdom. He is then to send these suggestions, together with his own opinion, to the Apostolic See.

4 Unless it has been lawfully provided otherwise, the diocesan Bishop who judges that his diocese requires an auxiliary Bishop, is to propose to the Apostolic See a list of the names of at least three priests suitable for this office.

5 For the future, no rights or privileges of election, appointment, presentation or designation of Bishops are conceded to civil authorities.

Canon 378

1 To be a suitable candidate for the episcopate, a person must:

- be outstanding in strong faith, good morals, piety, zeal for souls, wisdom, prudence and human virtues, and possess those other gifts which equip him to fulfil the office in question;
- be held in good esteem;
- be at least 35 years old;
- be a priest ordained for at least five years;
- hold a doctorate or at least a licentiate in sacred Scripture, theology or canon law, from an institute of higher studies approved by the Apostolic See, or at least be well versed in these disciplines.

2 The definitive judgement on the suitability of the person to be promoted rests with the Apostolic See.

Canon 379

Unless prevented by a lawful reason, one who is promoted to the episcopate must receive episcopal consecration within three months of receiving the apostolic letters, and in fact before he takes possession of his office.

Canon 380

Before taking canonical possession of his office, he who has been promoted is to make the profession of faith and take the oath of fidelity to the Apostolic See, in accordance

with the formula approved by the same Apostolic See.

Appendix to chapter 3

The data

We use data from the French LFS (*Enquête Emploi*), years 1982 to 2010. Over this period, the survey methodology underwent one major change, as well as a number of smaller changes. We made every effort to ensure continuity in the variables that we require for our estimations; this was made easier by the fact that, as a consequence of the very nature of our methodology, only a very limited number of variables entered our model.

Enquête Emploi en continu

The major break in the survey methodology occurred in 2002, when the LFS transitioned from a yearly schedule to a continuous process, where households can be surveyed at any point during the year.

Up to 2002, the design was to survey every sampled household three times, in March of three consecutive years. One third of the sample gets renewed every year.

From 2002 onwards, every sampled household was surveyed once every quarter, for six consecutive quarters. One sixth of the sample gets renewed every quarter.

While this was accompanied by a change in the definition of some variables, most of the variables of interest in this study (e.g. levels of education for the children) have kept the same definition over the transition. Therefore, the only difficulty originating

from this change was in terms of sampling.

When the data was collected annually, every household was surveyed in March, meaning that a question such as FORM: “What level of education are you currently enrolled in” was unambiguous. However, as households, in recent years, were also asked the same question in other months, ambiguity arose. For example, there is no official guideline to decide what a respondent should answer to this question during summer holiday, between two grades.

To resolve this ambiguity, our strategy to build our sample of 16-years-olds was the following:

- In the annual surveys (pre-2002), sample one child from all the household where one sixteen year old is observed at least one year.
- In the continuous surveys (post-2002), only keep surveys from quarters one and two (January to June), and sample one child from all the households where one sixteen year old is observed at least one (remaining) quarter.

Age and calendar year

In all the surveys, the age metric that we use is directly derived from the year of birth, rather than the age at the time of the survey. More precisely, we use the age as defined by the number of full years that will have elapsed by the end of the survey year (31st of September). Hence, someone born on February 1982 will qualify as 16 years old when they are surveyed in March 1998.

This definition coincides with the modalities of compulsory schooling in France, as children are required to start primary school in September of the calendar year in which they reach age 6 – which means that, if no grade repetition occurs, they will be admitted to *seconde* in the year in which they reach age 15. In March of the year in which they reach age 16, they must therefore be in *seconde*.