Essays on the economics of child labour and child education

by

Markus Ulrich Sauder

Thesis

Submitted to the University of Warwick

to the degree of

Doctor of Philosophy

Department of Economics

July 2007

THE UNIVERSITY OF

WARWICK

Acknowledgments

I would like to thank my supervisors, Prof. Amrita Dhillon and Professor Ian Walker for all their support and help. I would like to specially thank Prof. Amrita Dhillon for her invaluable advice and dedicated assistance during the final part of my PhD, and Prof. Myrna Wooders for her suggestion on the game theoretical part of this PhD. Additionally, I wish to express my deepest gratitude to Professor Manuel Arellano and Professor Stéphane Bonhomme at CEMFI, Madrid. Professor Arellano opened my eyes for advanced econometrics and encouraged me with his careful and didactic approach to deepen my knowledge in this area. I wish also to thank Professor Stéphane Bonhomme, not only for the transmission of his enthusiasm on econometrics, but also for the many occasions in which he took time to explain and assist me with statistical issues, programming and the understanding of econometric concepts. Thank you very much also to all the other staff members at CEMFI, who made my stay pleasant and productive. I would also like to thank Prof. Florian Pelgrin and Prof. Rafael Lalive at Lausanne University for their kind encouragements and endorsement.

I wish to thank my parents and my brother for their continuous encouragement and emotional support. I am also grateful to my parents for their financial help. Danke Mama, Papa und Axel für Alles, ganz speziell für Aufmunterungen und Unterstützung während der gesamten "Produktionsphase" dieses Doktorats. I dedicate this thesis to my parents, my brother and to my grandmother, Sofie Germesin.

A very special thank you goes to Claudia Lozada for all her love, care, patience, encouragements and help during this PhD. Not only did she sit down with me for long and painstaking proof-reading and notation checking, essential to the appearance of this thesis, but she also supported me throughout the whole process through many encouragements and uplifting words. Without her support, it would have been undoubtedly harder to "produce" this thesis. I big thank you to you for your love and for all the time. I acknowledge also Manolo Lozada-Zumaeta and Socorro Lozada-Can for their support, encouragement and care - Muchas gracias de la parte de Pollo.

I finally wish to acknowledge all the rest of my family and friends in Germany, France and the UK for their company and assistance during these four years. It would be impossible to list all the people who supported me during the redaction of this thesis and their respective contributions on a finite number of pages. I wish, however, to thank some people who helped me exceptionally during this period: Rina Kim and family Puch-Thönnes, namely Maya, Elke and Roberto, for their never failing support and encouragements and for the many occasions I was allowed and had the pleasure to stay with them. Thank you very much !

Last but not least, I acknowledge the financial support of the Economic and Social Research Council's grant in the period 2004-2005 and 2006-2007, Number PTA-030-2004-00361 and the one-year doctoral visiting fellowship of the European Research ad Training Network, in the period 2005-2006 "Microdata Methods and Practice".

This thesis was typeset with LaTeX 2_ε[1].

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The style package *wnewthesis* was used.

Declarations

I declare this thesis to be my own work, except where explicitly stated. Chapter four of this thesis was co-authored with Prof. Stéphane Bonhomme, CEMFI, Madrid. This thesis has not been submitted elsewhere.

Abstract

This thesis focuses on the economics of child labour and child education within developing and developed countries.

The first part of the thesis examines child labour and child education in developing countries. It investigates the motivations of parents to send their children to work and analyses the so-called commitment problem of child labour in a dynamic, overlapping generations game theoretical model. As a novelty, this model relaxes the requirement of an observable history of play and models the decision problem as an overlapping generations cyclic game. We show that first-best contracts may me implemented, implying optimal child education and low child labour, if a bequest sanction can be imposed by grandparents. We also discuss the special role that grandparents have within this model.

The second part of the thesis analyses the economics of child education within a developed country context: the transmission of education across generations and the impact of a schooling reform on educational choice and later outcomes. In a first chapter of this second part, we examine specifically the influence of grandparents, as postulated by the model in part one, on the education of grandchildren. A unique dataset on three generations, the National Child Development Survey of the UK, is used. As a special feature, we apply recent econometric techniques to deal with censoring in a semi-parametric setting. The results indicate that it is not education but rather unobservable factors on the parent and grandparent level that affect the educational choice of grandchildren. These unobservable factors may be interpreted as innate ability or parenting skills. In a second chapter within this part, a schooling reform, the introduction of comprehensive schools in the UK and its impact on educational and labour market outcomes is evaluated. We find, using data from the National Child Development Survey and applying a new, quasi-differenced matching estimator, that bias corrected estimates of the reform suggest no effect on the means, but a sizeable effect on the variance of outcomes. We interpret this finding as indicative of a higher risk inherent to the selective education system.

In summary the thesis sheds some new light on the economics of education and child labour, both in a theoretical and an empirical context, and provides a valuable reference and starting point for future research in this area.

Contents

List of Tables

List of Figures

Introduction

This thesis focuses on child labour and child education. The first two chapters are devoted to analysing child labour in developing countries, whereas the last two chapters address empirical questions related to child education in a developed country context. The first chapter reviews the theoretical and empirical evidence on child labour. Using a theoretical model, the second chapter analyses the decision between child labour and child education in a developing country context. The last two chapters are dedicated to the empirical analysis of two important questions in the economics of education: transmission of educational achievement across generations in the light of grandparental influence and the impact of changes in the education system on labour market and educational outcomes of pupils.

This thesis provides in four chapters some new elements in the analysis of both child labour and education. The topics of child labour and child education are linked, if one is willing to assume child labour and education are, to some extent, substitutes. It has been well documented in developing country contexts that excessive child labour affects education achievement adversely and thus, such an assumption seems not implausible. The phenomenon of child labour and child education can thus be analysed from an economic perspective of parental or child choice, although the perspectives are clearly different whether one chooses a developing or a developed country setting. In a developed country setting, child labour will be prohibited by law, but the possibility of part-time work exists typically in adolescence. Furthermore, individuals may leave school at the minimum school leaving age and choose to enter the labour market as early as possible. What distinguishes the two settings is the age at which the child enters the labour market, but the connection is the choice between education and early labour market participation.

In developing countries, children of ages 5-15 often have to conform to their parents choices in terms of education or labour for them. It is often thought that extreme poverty is responsible for parents to choose child labour for their children instead of education. The thesis will insist, as in the previous literature, that there might be important other reasons and not only the obvious poverty argument for parents to choose child labour for their children. Such reasons can be found in so called commitment problems, i.e. the impossibility for children to make binding promises to their parents to repay any educational investment children received. This thesis reviews the evidence on the motivations of parents to send children into work in chapter one, which gives a general overview over the child labour literature and underlines open questions. Chapter two focuses then on commitment problems in an overlapping generations model. In this chapter, it is shown that commitment problems prevent efficient investment in children in a non-cooperative game. It is then argued, that the usual game theoretic solution to such intergenerational games, trigger strategies, cannot solve the commitment problem if the history of play is unobservable. The novelty of chapter two is thus to address commitment problems within this limited information setting and to introduce bequest sanctions to overcome the information constraints. The chapter develops an entirely new model for overlapping generation games by modeling them as cyclic games. This assumes essentially that the history of play is irrelevant, and that individuals are entirely forward looking. While parents cannot solve the commitment problem, we show that grandparents can. The intuition for this result is that grandparents may punish deviators with strategic bequests. They can, as opposed to parents, save over a longer horizon and make bequest resources grow at the same rate as resources in the economy grow. Additionally, grandparents observe their children's actions towards their grandchildren, which limits deviation strategies. Furthermore, grandparents can bequest credibly to grandchildren instead of to their own children. The special role of grandparents is then also empirically assessed in Chapter three, within a developed country context.

Chapter three and four shift the focus of analysis to developed countries. In developed countries, young adults choose the amount of education they acquire after the minimum school leaving age. However, the decision of a young adult might be constrained by many factors, such as social background of the parents, their income and also inherited traits. Indeed, families with low income might be in a similar commitment problem situation as families in developing countries, i.e. they might not receive any compensation from their children for investing into their education, which is assumed costly. The alternative for young adults in developed countries is perhaps not child labour, but at least to drop-out of school at the relevant minimum age and start working. In Chapter three, this thesis analyses UK data, where such a drop-out at the minimum age is alarmingly high. This part of the thesis tries to assess if grandparents do have any influence (strategic or not) on grandchildren's education. A unique three-generational dataset from the UK, the National Child Development Survey, is used to assess this hypothesis. This chapter applies new techniques to deal with both endogeneity and censoring in a parametric and semi-parametric regression context. Instrumental variable techniques are used to assess the intergenerational effect of grandparents' and parents' education on children's education. Furthermore, the paper provides evidence that education effects of parents and grandparents on children are small, but the effect of unobservable factors is significant and large. This is an interesting insight for educational policy, as any such policy should analyse these unobserved components in greater detail to affect educational transmission between generations.

Chapter four, coauthored with Stéphane Bonhomme at CEMFI, Madrid, then addresses a related question of school quality and education in developed countries. As has been argued in the child labour and child education literature, low school quality could prevent individuals from acquiring more education. The English and Welsh education system of selective education was often critisised for perpetuating educational inequality. It was argued that the division of the selective system into high quality, high-ability grammar schools and low-quality, low-ability secondary modern schools did not allow pupils, who failed the admission exam into grammar school, to acquire education beyond the minimum schooling age. Thus, an individuals choice of education and his lifecourse might be affected by the schooling system itself, which prevented him from accessing higher education. A policy of establishing comprehensive schools was introduced in England from 1965 onwards as a potential remedy to this alleged systemic issue. The comprehensive schooling reform, rolled out in the UK between 1965 and 1975, merged high ability grammar and low ability secondary modern schools into a single, so-called comprehensive school. We assess the effects of this reform on various educational and labour market outcomes and present a new econometric methodology to achieve this. Previous studies found a large, negative effect of the comprehensive reform, for example on years stayed in school and test scores. It appears, however, that traditionally used estimators with observational data, such as matching for example, are severely biased in this particular data context. In the face of the inadequacy of the usual econometric techniques we develop a quasi-differencing method to remove the bias from the matching estimand. We find that comprehensive re-organisation has little effect on the mean of outcomes, but a large effect on the variance. We interpret this as the selective system being more risky than the comprehensive one - a reason for parents to prefer the latter.

In summary, this thesis provides some new theoretical insights with regard to investment in and choice of child education in developing and developed countries. It presents an entirely new overlapping generations game theoretic model which uses cyclic games. Furthermore, this thesis presents in-depth, advanced empirical anaylsis on education transmission mechanisms between three generations and assesses a very debated school reform in England and Wales. We derive interesting insights which open up new and challenging research questions in the economics of education, the school quality and the intergenerational mobility literature.

Chapter 1

Economic theories of child labour

Viewed in the light of our advancing civilisation and its greater opportunity for growth and human service, and also its greater demands for preparation for lifes of highest usefulness, I say again, the evils of child labour cannot well be exaggerated. *Samuel McCune Lindsay, Secretary of the National Child Labour Committee, 1903*

1.1 Introduction

The last fifteen years have seen an astonishing growth in research on child labour. The above quote illustrates that child labour is not a new phenomenon. Indeed, child labour was found in presently developed countries, in Europe and in the United States for example, until the middle of the twentieth century. In the nineteenth, and the beginning of the twentieth century, child labour was a key preoccupation for governments and public policy, as historical documents illustrate. As child labour decreased in developed countries, it became relatively under-researched. It was only "re-discovered" at the beginning of the 1990s in a developing country context[1].

What can be expected from this present review and from the extensive reviews conducted by Grootaert and Kanbur (1995), Basu (1999), Bhalotra and Tzannatos (2003), Basu and Tzannatos (2003) and Edmonds (2007) ?

Basu (1999) and Bhalotra and Tzannatos (2003) focus on the disconnected nature of theoretical and empirical research. While this now seems less apparent, Edmonds (2007) highlights the need for more peer-reviewed policy evaluation on child labour. This

[1] One can hypothesize that this rediscovery is due to an increased involvement of international organisations, such as the United Nations, on a global scale. The early 1990's witnessed the end of the bipolar international order, and many international organisations saw their potential field of development policy extend, after decades of relative blockade and inertia within the organisation, Boutros-Ghali (1992). This might explain the renewed interest of policy makers in child labour

present article agrees with this line of argument, and would like to highlight important grey-zones that remain in the theoretical, but also empirical, child labour literature. This article is not meant to be an exhaustive summary of empirical research: for this, see the excellent article of Edmonds (2007). Rather, it intends to illustrate what we understand about the rationale of child work, and which areas of the literature require further scientific investigation. As will appear, despite intensive research, there remain important open questions.

Research opinions both converge and diverge within the child labour literature. As the review will show, convergence of opinions have been achieved in agreeing that child labour cannot effectively be tackled by secondary instruments, such as bans or legal measures. It is rather the root causes of child labour that should be addressed. There is also some convergence in identifying these root causes empirically. Poverty, credit constraints, commitment problems and parental preferences within the household are considered important in explaining child labour. A unique convergence of opinion is also detectable as far as the worst forms of child labour are concerned: clearly, harmful work should be banned on moral grounds. Conversely, a consensus seems to emerge that some forms of child work need not necessarily be harmful.

Divergence, however, arises from the different country studies. These illustrate, to what extent child labour is a cultural and country specific phenomenon. Although some general causes, as stated above, might explain child labour in most country settings, the heterogeneity in particular situations translates into specific policy options adapted to each country context. Extrapolation from one country context and imposing similar policy instruments within another country context is not to be recommended. In addition, child labour has received a very different treatment in other disciplines which also inform policy. Anthropology and development studies, to name only two, tend to analyse the child labour problem from its cultural perspectives, Groves (2004). These disciplines generally voice the concern that Western institutions apply "Western standards" to the problem of child labour, disrespecting local customs and traditions. Nieuwenhuys (2007) speaks even of a "sacralisation of childhood" by Western institutions, refering to policies aimed at reducing child labour. Thus, in summary, there is still no consensus on how to combat and understand the phenomenon of child labour in a cross-disciplinary manner.

The debate about the definition of "child labour" alone is illustrative of the difficulties researchers face when trying to analyse the child labour problem. It is helpful to dwell on this for a while, as it is essential to define child labour adequately in order to discuss research or policy options. Let us exclude the worst forms of child labour from a potential definition. These are clearly extreme cases, harmful to child development, and there exists, as mentioned above, a consensus on banning them. But the classification of other child activities and work is problematic. When exactly is child work harmful? Is child work not also a way to learn additional skills? One could agree that one should call the activity 'child labour' if the activity is harmful to the future development of the child. Again, it seems difficult to define a valid counterfactual to this situation, Edmonds (2007), i.e. what a child would have done if it had not worked. As the most important part of child labour occurs within the household, the definition of "market work" is clearly also inadequate. Therefore, in this chapter, we opt for a wide definition of child labour, which includes market and domestic work, and includes also light (and potentially non-harmful) work. In the following, we will alternate between our definition of child labour and the official definition of the International Labour Organisation (ILO in the following), which does not take into account domestic work. The ILO defines child labour as child work that is harmful to development - with the above mentioned limitations.

Child labour, as measured by the ILO, is a global phenomenon that primarily affects developing countries, though it exists as well, to a very limited extent, in developed countries. Although the numbers of child labourers have declined steadily over the last 40 years, labour market participation rates of children between 5-14 years are still high in some parts of the world. The report of Hagemann et al. (2006) reveals that a total of 190 million children between 5-14 years work, down from 210 million in 2000, ILO (2002). As one can notice from the statistics in Table 1.1, child labour participation is high in Asia, Africa and Latin America. The Asian continent has most child workers in absolute value, whereas the participation rates are highest in Africa. Absolute numbers of child labourers have been on the increase in Africa, reversing the downward trend in all other regions. These statistics should be considered with caution. First of all, the definition of child labour used here is limited to market work or near-market work.

All estimates of the number of working children are thus an underestimate, since, as discussed, most child work occurs in the family farm or in the informal sector of the economy, activities which are inherently problematic to measure.

Region	**2000**		**2004**	
	('000s)	(%)	('000s)	(%)
Asia and the Pacific	127.3	19.4	122.3	18.8
Latin America and the Carribbean	17.4	5.7	16.1	5.1
Sub-saharan Africa	48.0	28.8	49.3	26.4
Other regions	18.3	6.8	13.4	5.2
Total	211	17.6	190.7	15.8

Table 1.1: Regional estimates of economically active children (5-14) in 2000 and 2004, Source: ILO (2006)

As mentioned before, despite convergence in identifying the root causes of child labour and a wealth of country studies, there are *lacunae* to be filled. Eight years ago, Basu (1999) argued that theory and applied research are widely disconnected in the field. And still today, a closer link of applied research with theory must be encouraged. The main body of research in theory and applied work has produced an astonishing variety of results that often contradict themselves and lack, as discussed, a consistent and common theoretical approach, Bhalotra and Tzannatos (2003). A more methodological empirical approach to child labour is also needed [2]. Empirical research seems to have taken the lead on theory, often providing results without a solid theoretical foundation. Thus welfare analysis is not possible and policy implications are difficult to assess when abstracting from general equilibrium effects. Data limitations were certainly a valid excuse for a disconnection between theory and applied research, but this will hopefully be a thing of the past with some excellent datasets. [3]

In addition to this disconnectedness between theory and applied work, some theoretical problems are still unresolved. To give a few examples: the demand side of child labour has been neglected and the commitment problem of child labour has not been analysed within a dynamic framework, although admittedly, a static model cannot do

[2] For an excellent overview of potential pitfalls of empirical approaches and a detailed summary of all empirical studies, see Bhalotra and Heady (2003) or Edmonds (2007)

[3] See Edmonds (2007), Bhalotra and Heady (2003) or http://www.ucw-project.org/ and http://www.ilo.org/ipec/index.htm for examples of information on datasets

the problem justice. Moreover, the literature has assumed (except, for example, in the contributions of Goldin and Parsons (1989), Manacorda (2006) and Bhalotra (2007)) that parents are altruistic. Thus the role of parental preferences is an under-researched topic. In the same vein, the role of children's preferences towards child work seems also an interesting, untouched field: are children really on their own supply curve?

The present contribution links theory and applied work in a modest and concise way. It attempts to outline unresolved problems and research ideas in the context of child labour. It is a difficult task to partition the child labour literature into meaningful subgroups. The following is a tentative outline, in order to give the review some structure. First of all, a historical overview seems appropriate: intuitively, one might think that the conditions of child labour since the period of proto-industrialisation and the so-called "Industrial Revolution" would be very similar to the situation in today's developing countries. Returning to the present, one can distinguish between the supply and the demand of child labour, OECD (2003). These two broad categories can then be subdivided into thematic subgroups. On the demand side of child labour, we will look at the following subgroups: (i) child-specific skills and productive technology, (ii) international trade, and (iii) household production

Similarily on the supply side, we will examine four subgroups: (i) poverty and parental preferences, (ii) credit constraints, shocks and commitment, (iii) multiple equilibria, and finally a small section on (iv) norms and cultural factors. For every section, a paragraph on theory and applied research will be developed. The sections are thus self-contained and can be read separately. Having thus examined the past and the present of child labour, we venture to take a look into the future by describing potentially successful policy: the last summary section illustrates important interconnections and includes a discussion on policy options. Throughout the review, we provide some illustrative graphs from a country panel data set of child labour, drawn from the World Development Indicators (2003).

1.2 A historical perspective

Apprenticeships for young adults and children were common in medieval Europe, and involved early labour market participation. Children enrolled in such contracts were considered lucky, since learning a trade would yield high returns, especially if one rose within the hierarchy of the trade guilds, Becci and Julia (1998). Various arrangements governed the apprenticeship contracts, which can be thought of as "learning by doing" training, Becci and Julia (1998). Parents had to *pay* a fixed sum of money, depending on the trade to be learned by the child. The so-called "pauper apprentices" were a subclass of apprenticeships, and no parental payment was required. Anecdotal evidence indicates that chimney sweeps, for example, belonged to this category of child workers. Often they were recruited in poor areas where parents had to fear for the subsistence of their families. Parents were happy to send their children away, knowing that they would at least survive, Becci and Julia (1998). Pauper apprenticeships were very common towards the end of the eighteenth century in Europe. Thus, there were essentially two forms of early child labour, and child labour would not necessarily be harmful in some instances. Apprenticeships offered an opportunity to learn a trade, involved early child labour in order to acquire necessary skills, but also resulted in higher wage compensations later in life.

An interesting hypothesis is that industrialisation may have distorted these established mechanisms by lowering the returns to the established trades. The Industrial Revolution demanded cheap, unskilled labour. In addition, non-manual skills became relatively more valuable during this transition, Rosenzweig (1995). Learning by doing was the orthodox way parents tried to raise the human capital of their children. However, this form of human capital investment lost importance compared to schooling. In the post-Industrial Revolution era, early child labour was a disadvantage in skill accumulation, whereas before it had been a way to gain significant returns. Different human capital was necessary to yield returns and it was no longer exclusively transmitted through apprenticeships, but also through schools. One might hypothesize that parents needed to adapt to this new setting, where schooling in early childhood became relatively more important.

Early theoretical ideas on child labour, however, from contemporary economists focus on income need and poverty as a rationale for child labour. Basu (1999) cites Marx (1867) who invoked the pure necessity to *make ends meet*.

Many applied studies have examined the situation of child labour in the nineteenth and early twentieth century. For example, Goldin and Parsons (1989) focus on parent-child relationships within the family. They find some evidence for a weak altruistic link between parents and children, and put it as "parents did not have strong (economic) altruistic concerns for their children". Goldin and Parsons (1989) suggest as well that it was rather technological progress, than parental altruism, which brought about a change in the prevalence of child labour. Manacorda (2006) identifies adult and sibling labour supply responses to changes in the child labour opportunities of children in US data from the early twentieth century. He uses different age discontinuity thresholds implied by state child labour laws in the US. This very original work of identification tests, therefore, the extent to which parents reduce their labour supply when children work. This would implicitly be consistent with a selfish attitude towards children. The data seems, however, not to support the selfishness hypothesis and, in this context, it seems rather poverty and not the exploitative preferences of parents that lies behind child labour prevalence. In view of such diverse evidence, parental preferences and child labour seem an intriguing topic, with still a wealth of research opportunities.

Another branch of the economic history literature argues that child labour would only be used within simple, primitive technologies. Thus, technological progress could play a key role in eliminating child labour. Nardinelli (1980) claims that it was technological change, and not legislation such as the Factories Act, which caused a decline in child labour. In his view, children performed secondary tasks of assistance to adults. Technological advances, such as the self-acting spinning mule for example, made these tasks redundant. He presents some data on the employment of children and argues that child labour laws only accelerated a trend of withdrawing children from the labour force. Nardinelli's (1980) data shows large fluctuations in the textile sector child labour force, both before and after the legislation. This seems to relativise the reported decline as a consequence of legislation, since it could also be re-interpreted as a temporary decline. Nardinelli's (1980) claim that children were employed mostly in the textile industry is certainly accurate, but other important employment opportunities existed in England, especially, for example, children working as domestic assistants, Kirby (2005). Extensive data from Belgium, as in Cunningham and Viazzo (1996), produces evidence on the proportions of working children in several industries compared to the total number of workers. In the Ghent textile sector, participation rates were as high as 15-20% whereas they reached 26% in paper mills and 33% in tobacco factories. Hence children had ample employment opportunities, and it is, in general, difficult to explain the decline accurately with the limited sectoral data available.

Nardinelli's (1980) analysis is also supported by Moehling (1999) who argues that child labour laws only "codified" a decline. Nevertheless, the issue is still debated, since Merkeerk (2004) underlines the importance of assessing the laws beyond their immediate context in the late 19th century, and taking into account the accelerated demographic transition and the wealth effect of colonisation in England. These economic and demographic factors could potentially drive child labour down faster. The decline could thus be plausibly unrelated to the child labour legislation.

The main question remains to know the causes for the growth and the later decline of child labour activities, as this knowledge might have direct applications in todays developing country experiences. Horrel and Humphries (1995) claim that economic historians agree that the Industrial Revolution pushed the number of child labourers up. It is speculated that the employment opportunities for children caused an increase in child labour. From home production many families switched to industrial production/employment and at the same time geographically from rural to urban locations[4]. Horrel and Humphries (1995) identify increased male wages and changes in technology as the main theoretical driving forces in the reduction of child labour. Their data analysis suggests that older children left home earlier and shifted the burden of work increasingly to their younger siblings. Horrel and Humphries (1995) argue that child labour was stable during the Industrial Revolution compared to the period of proto-industrialisation. Although the child labour force increased in *absolute* numbers, its relative proportions remained unchanged. Interestingly, Horrel and Humphries (1995) find contradictory

[4]Note as well that children became "unemployed" on the countryside, which incited families to move to the labour-absorbing cities

income and wage effects on child labour participation rates. For example in mining families, the wage effect on child labour participation was negative whereas it was positive in the case of factory workers. Here the poverty or "need" hypothesis holds only for some occupational groups, a phenomenon which would require further investigation in a modern setting.

Historical analysis could gain from a more detailed dataset. The planned *Victorian Panel study* in the UK (1850-1901)[5], could potentially constitute a rich source to retrace family living conditions in the late nineteenth and early twentieth century in relation to child labour.

This section on historical child labour studies underlines the importance of understanding what led to the increase and subsequent decrease of child labour in historical times. Policy makers today are especially keen to know which factors led child labour to decline in industrialised countries at the end of nineteenth and at the beginning of the twentieth century. The debate about the effectiveness of legal measures, but also about the role of parental preferences, is still lively in today's policy context. After having examined historical accounts of child labour, it seems now appropriate to turn the focus of our analysis back to the present. The next section outlines the main factors thought to influence the demand for child labour.

1.3 The demand for child labour

The demand for child labour has not received as much attention as the supply side. As a large part of child labour occurs as domestic work or in the family farm, Edmonds and Pavcnik (2004), the household supplies and demands child labour at the same time. Thus, all household analysis considers child labour demand, but there are only a few contributions to the literature that address the role of household production technology or the productivity of a child directly, in order to explain child labour. An important question, concerning household and industry demand for child labour, is whether children have specific skills or characteristics making child work more desirable? This aspect of child labour demand has been neglected in most of the studies, yet is one of the most

[5]http://www.data-archive.ac.uk/randd/vps.asp

interesting research questions on the demand side. Bhalotra and Heady (2003), for example, underline the fact that child employment resolves agency problems within the household. In addition, one might hypothesize, children might be more docile and easier to discipline in addition to having lower reservation wages than adults.

1.3.1 Child-specific skills and technology

Theory The "nimble fingers" hypothesis is a classic buzzword to justify the existence of child employment. Children are supposed to have specific abilities or a higher productivity at certain activities than adults. This would contradict the now well established theoretical substitution axiom between adult and child labour of Basu (1999).

There is some evidence that children perform different tasks over their child "lifetime" - hence some specialisation by age, OECD (2003). This would imply that children are employed in specific, age-specialized occupations. On the other hand, and in terms of technology, children are seen to perform tasks which can, supposedly, be replaced by machines. Hence more modern technology drives the demand for child labour down, a conjecture we already touched upon in the historical section, Nardinelli (1980). Given the informal nature of children's employment and the small tasks they perform (e.g. in the service sector), there is today no study that has found an effect of technological improvement on child labour. However, as Figure 1.1 shows, there seems to be a clear negative correlation between agricultural productivity and child labour in a panel of countries observed in 1960-2000[6]. This strengthens the evidence that there is some productivity argument in relation to child labour, be it within the household or within agricultural environments.

The research of Dessy and Pallage (2005), for example, examines the relation between productive technology and child labour on a theoretical level. A coordination failure between companies and parents may exist in developing countries: because parents do not invest in higher human capital for their children, firms do not invest in better technologies which require more human capital from workers, and *vice versa*. This could induce a vicious circle, perpetuating poverty. This would be a valuable starting point

[6]There are sometimes no datapoints for the year 1960 in this panel data drawn from the World Development Indicators

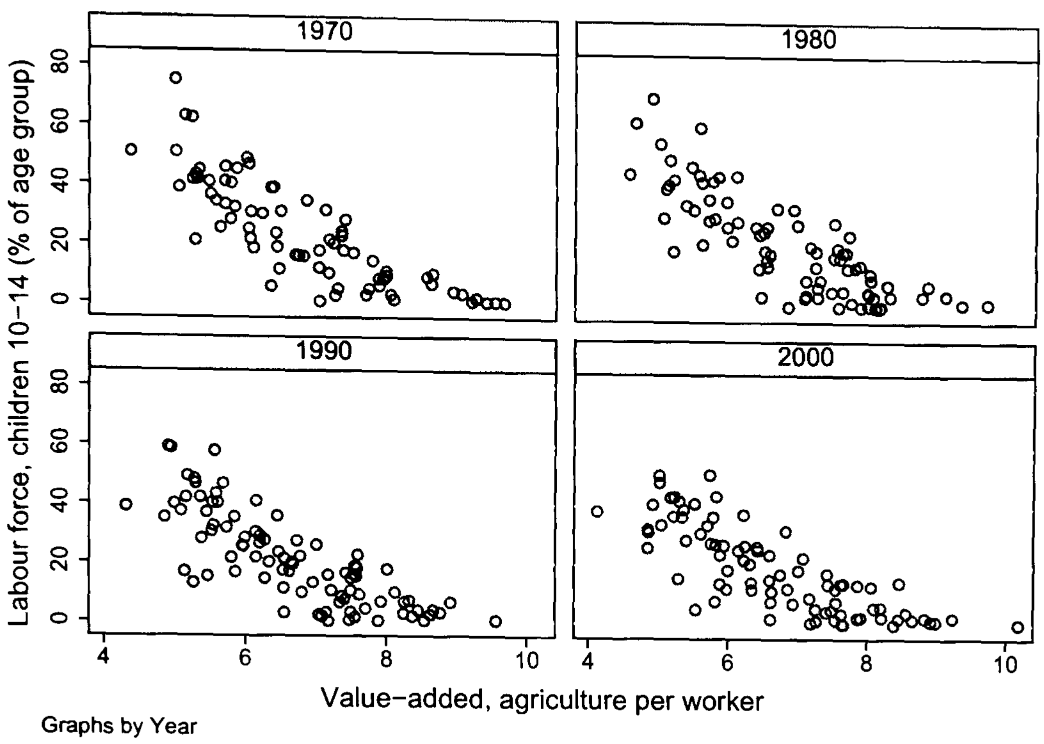


Figure 1.1: Value added in the agricultural sector and labour market participation rates of 10-14 year olds in developing countries.

for an empirical analysis.

Davies (2005) examines another interesting demand-side argument which we will briefly discuss. It involves labelling initiatives by producers, and it can be thought of as a demand-side intervention. The demand for child labour by employers could be influenced by raising the costs of using child labour inputs. Labelling non-child labour products and banning products that have been made with the use of child labour has been endorsed by the United States in the past, Basu (1999). The argument is that labelled products should discourage consumers in developed countries from buying products made by children. Employers would then not use child labour. Davies (2005) shows, in an original and simple model, that labelling does not necessarily reduce child labour use to zero. In his model of Bertrand competition, (i) cost differences between adult and child labour and (ii) price-sensitive consumers make child labour use attractive. Price sensitive consumers do not care whether products have been made using child or adult labour. Hence some employers will still continue to use child labour intensive production lines.

Social labelling, which aims to inform the consumer about the cirumstances of production, especially about labour standards, has received increased attention in recent years, Basu et al. (2006). In a variation on the work by Davies (2005), Basu et al. (2006) examine the impact of such labelling and associated trade sanctions. The general conclusion of Davies (2005) is not changed; however, Basu et al. (2006) discuss also the implementation of such labelling programs. Interestingly, the threat of trade sanctions on all products, regardless of whether they are produced with child labour or not, discourages the affected countries from maintaining a credible labelling system.

Another argument against labeling which reduces child labour opportunities is a welfare argument. If labelling decreases child labour, children (and parents) may be worse off if they are not compensated for the foregone child wages, Basu et al. (2006).

Applied work and discussion In spite of the theoretical contributions discussed above, there is not much empirical work on child labour demand.

Anker et al. (1996) attempted to test the nimble finger hypothesis. They did not find any evidence that children weave carpets faster than adults in a dataset of carpet production in India. The data they collected is however possibly subject to measurement error, since age is not reported reliably and the production of carpets by children is not well documented. The authors rely on the very small sample sizes of high quality carpets, and the recorded intensity of work of children, to infer about their weaving capacity. However, it seems difficult to disentangle the effects of child and adult production in this context. Thus any conclusion should be drawn with care, and further studies in this area would certainly be of great interest.

In relation to technology, an interesting study of cocoa producing farms in Côte d'Ivoire, using child labour, documents that children work less in higher productivity environments, Nkamleu and Kielland (2006). This could be due to an income effect. Higher productivity might imply higher wages for adult workers, which in turn induces them to replace children. This would also fit into the argument of multiple equilibria on the labour market, Basu and Van (1998). The effect could also plausibly be due to the use of modern technology, such as machines and fertiliser, which suppresses child labour opportunities. Unfortunately, the study of Nkamleu and Kielland (2006) does not allow us to distinguish between these alternative explanations, as they would have very different policy implications.

The hypothesis of child-specific skills has received surprisingly little empirical attention. This is probably because the hypothesis of child-specific skills itself has been used by advocates of child labour, and thus does not fit into the discourse or research agenda of international organisations. It would be of prime importance for policy to establish a clear connection between child labour and child skills or productive technology. Appropriate data collection seems still a priority in this context.

1.3.2 Household production

There is a wide-ranging literature on fertility, child labour productivity and schooling in developing countries, Rosenzweig and Evenson (1977), Levy (1985). However, in all recent surveys, little attention has been dedicated to the topic of household production. Household production effects are mostly addressed within gender or sibling effects, Edmonds (2006b). Most studies that examine the effects of siblings or birth-order on child labour touch upon the fascinating topic of household production when they examine, for example, comparative advantage between siblings or the effect of work experience of siblings in market/domestic work.

As the bulk of child labour occurs on the family farm or within households Heady and Bhalotra (2000), it seems crucial to revive the interest in evaluating shadow wages and income contributions of children to household or farm production. The productive contribution of children is often not well understood, as they may, for example, perform complementary tasks. Such complementary tasks, such as water and firewood fetching, cattle guarding, without which adult production would not be possible at all, is difficult to value. The unavailability of high quality data, such as time use surveys, but also the methodological problems of measuring child or adult shadow wages, are clearly detrimental to further work on the issue.

A detailed assessment of home production would plausibly allow us: (i) to test the substitution axiom, (ii) to evaluate the impact of productive technology on child labour and (iii) to test the specific skills hypothesis.

Applied work It is relatively well documented that children contribute significantly to household work, Psacharopoulos (1997). There is also evidence on comparative advantage between siblings, documented by Edmonds (2006b). In Edmonds's data from Nepal, older girls seem to work more than their younger siblings, which suggests the importance of work experience. In addition, increasing numbers of younger siblings, and longer birth intervals between them, increase the workload of older girls. Thus home production, so Edmonds (2006b) argues, seems to generate comparative advantage: girls spending more time in domestic work. Interestingly, adding a younger brother to households with a girl generates more market and domestic work for girls, whereas adding a younger brother to a household with a boy results in augmenting the boys' market work only. Edmonds (2006b) expands slightly on the differential returns to education for boys and girls, which will affect their schooling or their involvment in domestic work. But how productive are younger children, girls or boys, in general in domestic work? This is an essential question which should be addressed, also in view of comparing their productivity to the returns from education.

Applied work on the productivity of children has been done by Rosenzweig and Evenson (1977), Levy (1985), Cockburn (2001), Jacoby (1993) and Skoufias (1994). They have attempted to evaluate shadow child wages and productivity. These studies find that children are indeed quite productive, less so than adults though. Cockburn (2001) for example finds that children are to some extent substitues for adult labourers, and that they are about one third to one half as productive than adults. They contribute only a small fraction (5%) to household income. Rosenzweig and Evenson (1977) and Levy (1985), but also Cockburn (2001), find a connection between produtive assets used, or mechanisation, and a reduction of child labour. These findings are very interesting from a policy point of view, since development of the agricultural sector could thus be a way of curbing child farm labour.

The problem in this literature is the lack of available data on inputs, but more specifically it lies in the assumption of a separation of consumption from income. Rosenzweig and Evenson (1977) and Levy (1985) rely on structural models to identify the effects of child productivity on fertility, and Cockburn (2001) and Skoufias (1994) use an IV strategy to explore the contribution of children to household income. While there has certainly been some research on local labour markets, child time allocation and occupation of children, Edmonds (2007), very few research papers have attempted to quantify child productivity. In this area, there are still many open questions. These obviously extend to the relative returns from child labour and education and other facets of child labour theory, and require a deeper understanding, also for a successful policy to reduce child labour.

1.3.3 International Trade

Theory International trade does not immediately connect with a demand-side argument, and thus might be misplaced in a demand for child labour section. A standard argument from trade theory implies, however, that increased openness raises the return to the relatively abundant factor. In many developing countries, unskilled labour can be seen as the abundant factor, and child labour opportunities should thus become more prevalent. But increased openness might have ambiguous effects on child labour demand. Raising the return to child labour activities might lead to a replacement of children by adult workers. It might also lead to increased child labour, depending on the size of the wage increase and the characteristics of the country's labour market.

There are now more theoretical publications that examine the relationship between trade and child labour. Jafarey and Lahiri (2002) construct a model examining trade sanctions. Trade sanctions lower the price of the product of the unskilled industry. In the presence of borrowing constraints, this feeds through into household's discount rates. Raising household discount rates leads then to lower investment in child education. In his model, trade openness, having the opposite effect as the sanctions, is principally good. Openness may lower the interest rate to world levels, and foster educational investment if households are not credit constrained. This analysis abstracts from the problem of commitment, which will be discussed in the next section. Chaudhuri and Gupta (2004) show, to the contrary, that trade liberalisation need not have a positive effect on child labour and that the effect depends on the relative factor intensities in the export and import sectors. A controversy remains thus on the effects of trade openness.

In terms of trade sanctions, Ranjan (2001) establishes that bans on child labour products often run into enforcement problems. The welfare of the poorest households might be reduced by such a ban. More recent work by Grossmann and Michaelis (2007) discusses the effect of trade sanctions in an imperfect competition model. Only firm-specific bans would be effective in such a setting, with a tariff rate varying with the amount of child labour incorporated into the good. Bans on trade of child labour products might have unintended welfare consequences. We refer the interested reader to Section 1.4.3 below and the model of Basu and Van (1998), where the usefulness of bans is discussed in general terms.

Applied work and discussion Edmonds and Pavcnik (2004) showed in a cross-section of countries that trade openness does not explain high levels of child labour, once one accounts for the income differences between countries. The statistically insignificant elasticity of child labour with respect to trade openness asserts thus no effect. Cigno et al. (2002) come to the same conclusion. Judging from these two studies alone, trade sanctions might not be an effective instrument in curbing child labour. This might be explained by the low prevalence of child labour in the exporting sector of some countries. Neumayer and Soysa (2005) examine the effect of foreign direct investment on child labour and find that countries more open to foreign direct investment tend to have lower child labour participation rates. This would support theories of openness that argue for a positive effect, for example through higher adult wages, meaning that children will be withdrawn from the labour force, or through better technology that displaces child labour. In summary, the link between trade and child labour is also empirically "under-investigated", and present studies do not provide enough corroborating evidence on the overall effect of trade openness or trade sanctions on child labour.

1.4 The supply of child labour

1.4.1 Poverty, parental preferences and household decision-making

Theory A theoretical approach to child labour and poverty can be expressed in terms of the "luxury axiom", Basu and Van (1998): children's leisure time is a luxury good. Only households with very low non-child labour income will send their children to work. This implies that parents are altruistic towards their children, and take into account the value of children's time. Thus the "luxury axiom", and to some extent the poverty hypothesis, has, as its antagonist, the beneficial parental preference hypothesis. The necessity argument above invalidates the reasoning that parents send their children to work because they have preferences for child work, Lopez-Calva (2001). In the media and in the public image, child labour is often seen as a substitute for parents' leisure time, Basu and Van (1998). In the theoretical literature, such non-altruistic parents have not received much attention. As it is plausible that altruism need not necessarily be assumed, this is would be a valuable extension to pursue. Some evidence of non-altruistic parents can also be found in the preceding section on history. Moreover, the poverty hypothesis associated with the hypothesis of altruistic preferences has been challenged by some researchers, Bhalotra (2004, 2007). This is justified, insofar as many studies in developed countries found only weak evidence of altruism (Bhalotra (2004) and references therein). An interesting case of a rejection of altruism in a developed country is Blow et al. (2004), documenting that parents direct child benefits relatively more towards alcohol purchases than towards purchases for children. Also, as Bhalotra (2004) rightly states, many programmes in developing countries assume either that parents are non-altruistic, or that parents ignore what is best for their children, in developing transfers and programs conditional on parents' actions.

Many researchers, Edmonds and Pavcnik (2004), maintain that poverty is the main determinant of child labour. However, some applied work, discussed below, shows the link poverty-child labour is debatable at the household level, given that we do not observe parental preferences. This has important consequences for policy recommendations.

Another extension of the child labour literature, which follows the questions on parental preferences is to assess who makes the decisions in the household and whose preferences determine child work. This is the non-unitary approach to household decision-making or bargaining. Following Basu (1999), the bargaining literature on child labour can be divided into two areas: intra-household bargaining and employer-parent bargaining over the child. In all these models, one does not assume parental altruism. Children are modelled as goods over which the bargaining process takes place.

The employer-parent bargaining approach assumes that parents bargain with an employer over the wage w to be paid to the child. The threat point of this bargaining game is the no-agreement or stand-alone income that employer and parents earn. Gupta (2000) reasons that rises in adult wages implicitly lower the employer's threat point. Consequently child wages rise in line with adult wage increases. Parents can however only influence child wages if they can coordinate their actions towards an employer. As the potential market of unskilled labour is relatively large in developing countries and child labour is mobile and substitutable, the influence of parents on wages seems debatable, and thus employer-parent bargaining models have not received much attention.

Intra-household bargaining models have been more popular in analysing child labour supply by the household. These barganining models revoke the assumption of unitary household decision making, McElroy (1990). Conflicting interests and views in the decisions on how to share and allocate household income are allowed for, McElroy and Horney (1981). More recently, Moehling (1995), Moehling (2003) and Blundell et al. (1994) have developed and used such models in relation to decisions on child time allocation.

A typical model assumes that household members have a different weight in the decision-making process. For example the weight could depend, as in Basu (1999), on the share of income that a household member contributes to household income. Thus, there is a collective problem as opposed to a unitary problem, denoted by:

$$\max_{x_1,x_2} \alpha(y_1,y_2)u_1(x_1,x_2) + [1-\alpha(y_1,y_2)]u_2(x_1,x_2)$$

where α is an index of power in the household. Suppose the x_i are consumption vectors and the y_i incomes. Then α could be thought of as depending on the wealth a household member brings to the household (bequests in a marriage from the side of the women for example, education of the household members, and cultural factors). Here one can see the effect of households income y_i on the maximisation problem directly through the household power factor α.

Having examined the theoretical approaches to poverty, parental preferences and household decision-making, we turn now to the empirical results of this part of the child labour literature.

Applied research Rosenzweig and Evenson (1977) analyse child labour in Indian villages. In this excellent article, the authors find that families are likely to experience poverty, high fertility and high child labour at the same time. This corroborates a possible motive of the household to increase child labour income through increased fertility. Other studies have confirmed a strong link between poverty and child labour. Skyt-Nielsen and Dubey (2002) test the subsistence hypothesis alongside four other hypothesis. Like Grootaert and Kanbur (1995) and Ray (2002), Skyt-Nielsen and Dubey (2002) find a strong correlation between poverty and child labour. The authors of the various studies emphasise that in their view, a general consensus has emerged to associate child labour with lack of income. As a variant on this theme, Skyt-Nielsen (2001) argues that high school costs, approximately one month of their annual income, prevent parents from sending their children to school. Grootaert and Patrinos (1998) argue that it is not the high cost of school but the implicit opportunity cost of lost (child-)wages that prevents child schooling. On the individual household level, the poverty effect remains debated. Ray (2002) for example finds a strong link between poverty and child labour is absent in urban and semi-urban areas whereas it is present in rural areas. In an original way, Bhalotra and Heady (2003) provide evidence of how land ownership can increase child labour. In this model, children are more productive than hired workers. Hired workers are more difficult to monitor and children will work in their own interest, i.e. to inherit the family farm. Thus, wealthier land-owners might actually prefer to use their own children as (child-) workers than to rely on hired labour.

Bhalotra (2007) tests the poverty hypothesis by deriving the implications of the hypothesis from a theoretical household model. The test is based on examining the wage elasticity of a child's labour supply. Bhalotra's (2007) idea is the following: if a household is very poor, then children work towards a subsistence income target. The elasticity of a child's labour supply should then be negative. Indeed, as the child wage rises, children should work fewer hours to meet the subsistence income target. Bhalotra (2007) finds that boys seem to work out of poverty compulsion, since their child wage elasticity is negative, whereas girls' labour supply seems not to respond to changes in child wages; the girls' wage elasticity being insignificantly different from zero. This raises interesting questions about parental preferences, but also, as in Edmonds (2006b), about differential returns to education and work for boys and girls. Finally, it challenges the effectiveness of poverty- alleviating policies to reduce girls' child labour activities.

Bhalotra (2004) finds evidence of altruistic preferences within the family, examining adult and child consumption responses. Interestingly, however, the altruistic link seems weaker when households spending on tobacco is examined. Thus, there could be evidence of some preference heterogeneity over child consumption between different households. The endogeneity problem encountered here, due to the self-selection of smokers, could however not be resolved due to the lack of credible instruments, Bhalotra (2004).

With reference to the household barganing theory[7], Duflo (2000) tested the unitary household model using a study of the South African old age pension benefit. The pension benefit may provide a 'natural experiment' setting, since persons/households that are just eligible for the pension benefit may be compared with those that are not yet eligible but are otherwise similar. The setting of Duflo (2000) is interesting, since it allows us to distinguish the use of benefit money, depending on the recipient within the household. Duflo (2000) finds that pension income received by women was likely to be redirected to female grandchildren. This was reflected in increases of height-for-age and weight-for-age Z-scores of children living in the same household as the pension recipient. There was no such increase for children living with male pension recipients and also no effects for boys living with their grandmothers. Child outcomes may vary, depending whether the money goes to "the wallet or the purse", Duflo (2000) and depending on child gender.

There seems to be evidence, although mixed, that the status of children plays a role in household decision-making, Basu and Van (1998), Udry (1996) and Bowning (1992)). Moehling (2003) showed that children have more "power" within the household if they contribute to household income. Moehling (2003) analyses in a recent working paper how much power, or influence, children have when the decision to buy a particular

[7]The collective model of the household has come under increased criticism. In empirical work it seems very difficult to justify measurement of power in the household. Duncan et al. (2002) refer to this problem as the "Achilles heel" of applied work: income seems to be a good proxy for household power, however, one can often not disentangle if the observed income is exogeneous to household decisions or if it depends on past decsions, say past labour supply decisions. Hence, only exogeneous variations in income of one of the recipients can give good predictions. Consequently, one needs to be careful to apply this setting within appropriate context. However, it can yield interesting insights, as in Duncan et al. (2002) or Duflo (2000).

good is made. Such psychological factors might thus play a role in the children's decision to accept work.

A more general idea would thus be to ask whether children are *on* their labour supply curve or not. This is definitely a question of central importance, especially for targeting policy interventions. Satz (2003) is the only research article that addresses this question from a normative point of view. If children are happy supplying child labour, and this translates into increases in status and well-being, and furthermore if child labour does not interfere with health and human capital accumulation, policies of banning child labour seem misdirected.

The bargaining models, Duflo (2000), but also the household production models, Edmonds (2006b), point to an important gender effect of child labour, which seems worth a short mention within the research work on parental preferences and parental bargaining. This gender effect is apparent in the statistics of Hagemann et al. (2006), see Table 1.2. Slightly more boys than girls are engaged in market work in the age group 5-11; however, this gender bias becomes very pronounced in higher age groups. The explanation for this might be increased non-market/household work, or preferences for girls to be educated. Accounting for unobserved household work, girls appear more likely than boys to work and miss school, OECD (2003).

The data from a country panel data set seems to confirm the poverty hypothesis in Figure 1.2, since over the five decades, child labour participation rates are higher in countries where GDP per capita is lower.

Although there seems to be a consensus on the relative importance of poverty in the explanation of child labour, poverty seems not the unique, nor the most important determinant. Poverty might mask other factors at work at the household or country level, such as parental preferences and commitment problems (see section below).

Swinnerton and Rogers (1999) relate child labour to the unequal distribution of income, not to poverty alone. Such an approach can be compared to the political economy models at the end of the present chapter. In addition, child labour is traditionally associated with developing countries, which are low income countries. But child labour has also been reported, although to a lesser extent, from developed countries. Pettit (1998) gives some insights into the case of the UK. Thus, other driving forces of

Age group	Child Labour - Gender distribution			
	2000		**2004**	
	('000s)	(%)	('000s)	(%)
5-11	109,700		107,647	
Boys	56,300	51.3	53,103	49.3
Girls	53,400	48.7	54,544	50.7
12-14	76,600		58,105	
Boys	41,500	54.2	31,848	54.8
Girls	35,100	45.8	26,257	45.2
Total 5-14	186,300		165,752	
Boys	97,800	52.5	84,951	51.3
Girls	88,500	47.5	80,801	48.7
15-17	59,200		51,911	
Boys	34,000	57.4	32,250	62.1
Girls	24,800	42.6	19,661	37.9
Total 5-17	245,500		217,663	
Boys	132,200	53.8	117,201	53.8
Girls	113,300	46.2	100,462	46.2

Table 1.2: Child labour and its gender distribution

child labour than income poverty should exist and need to be investigated. The next section gives an important other rationale for parents to choose child work. Credit constraints and commitment problems have also been analysed within the literature on the economics of education as factors preventing optimal schooling investment in children.

1.4.2 Commitment problems, shocks and credit constraints

Theory Parents have incentives to send their children to school and invest in their human capital, instead of sending them to work, if such an educational investment increases the total income of the children, and the family above the income attainable under a full child labour and no schooling regime. If we can believe that education means capital formation and has not a pure signaling role, as in Chevallier et al. (2004), and translates, through a human capital production function, Ben-Porath (1967), into higher incomes, then parents could be better off by sending their children to school. This would enable them to earn, as a family, higher wages through their educated children.

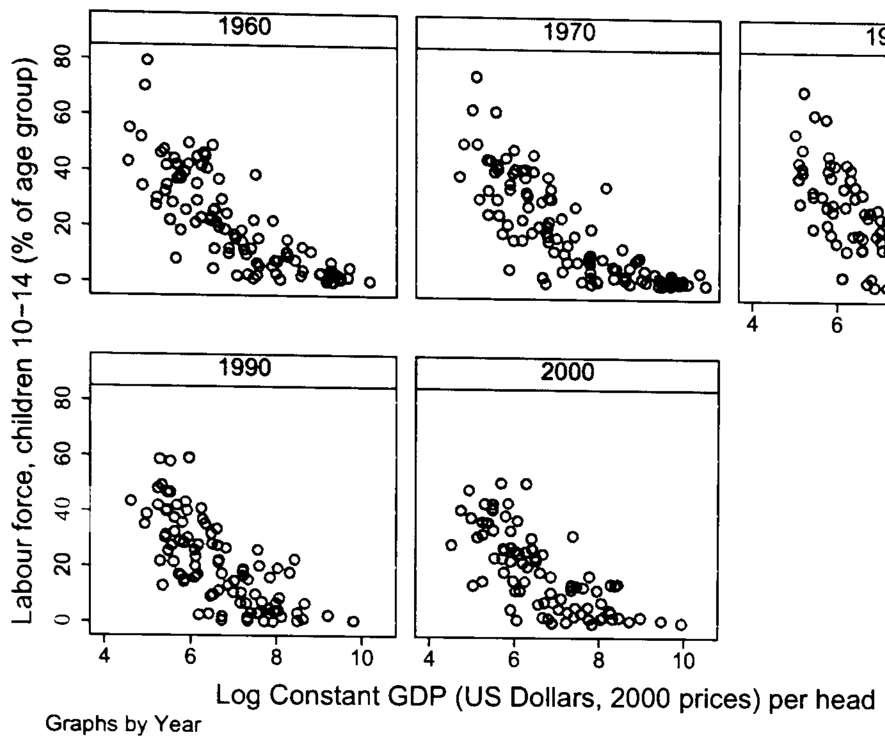


Figure 1.2: Log of constant dollars per capita GDP and labour market participation rates of 10-14 year olds in developing countries.

For completeness, it should be mentioned here, that parents have also to take into account returns to other investments, and notably investments in the capital market, when making a decision between child education and child labour. But supposing the returns to education were high and thus the most profitable investment. Then parents should rationally invest in child education.

The problem is, however, that the children will earn the higher wages, not the parents. Thus, unless transfer arrangements or some form of reverse altruism between children and parents exist, there is no guarantee that the parents will reap the benefits of an investment in child education, while the parents are certain to incur the costs of the investment (lost child wages and schooling costs). This is the intuition behind the commitment problem within the household. Children might not repay the parents' expenses, nor allow a sharing of resources across the family. This is perhaps the single most interesting research topic besides the poverty hypothesis. Since a very influential article of Baland and Robinson (2000), the topic has been classified as resolved on a theoretical basis. Although an excellent model, the intergenerational connections of commitment constraints are not well understood in the Baland and Robinson (2000) model, which lacks dynamics and possible strategic interactions between family members. There is some important research to be done in this area of intergenerational contracts and child labour.

Another factor, besides commitment problems, are imperfect credit markets which prevent parents from investing optimally: parents cannot lend or borrow against the child's future income. Along the same lines, the lack of insurance and credit markets in developing countries imply that a child's labour capacity might be used as an old-age insurance or to smooth out shocks. Several papers have addressed this strand of literature. Ranjan (2001), Dehejia and Gatti (2002) and Rammohan (2001) claim, and produce evidence, that children's work increases with negative income shocks to the household and in the presence of borrowing constraints. Thus, children's work earnings can be seen as a consumption smoothing device, that help households to maintain their permanent income over the lifecycle, Dehejia and Gatti (2002).

Baland and Robinson (2000) consider not the role of incomplete capital markets, but also the role of bequests. Because, as discussed above, future income (after investment) is realised by the children and not the parents, there is a commitment problem even with perfect credit markets. Perfect markets alone do not solve the commitment problem of child labour, a key insight of Baland and Robinson (2000). The role of bequests is as follows in this context: by reducing bequests to their children, parents could compensate themselves for the foregone consumption and/or for the amount of money they invested in their children. However when bequests are zero, then parents simply run out of resources. As children cannot commit to repay any expenses, child labour is, as a result, inefficiently high. Inefficiently high means, in this context, that society as a whole could be better off if children were educated.

As other research has shown, Wagener (2002), repayment by children is not common in developed countries. In some instances, remittance payments to parents or the wider family are observed in developing country contexts. In particular immigrants to developed countries transmit to their families in developing or transition countries. The literature on remittances argues that such transfers are sustained partly by strategic motivations, Stark (1995), Cox (1987). Although we observe such child-parent, child-family transfers, the commitment problem potentially exists because repayment only depends on some form of reverse altruism between children and parents, or on strategic transfer arrangements. Underlying cultural norms or rules could enable societies to sustain transfer payments that solve such commitment problems. Even if such culturally encoded strategic transfer arrangements exist, they can be perturbed by various factors of uncertainty and must depend on family characteristics, i.e. outside options or available punishments for deviators from the norm. Also, it is not clear why child labour is so prevalent if such a contractual incompleteness as the commitment problem, can be overcome by simple intergenerational trigger strategies, Lopez-Calva (2003),Rangel (2003).W̃hile trigger strategy set-ups usually do not imply a punishment, except reverting to the one-shot Nash equilibrium of the game, it seems very important to realise that such trigger strategy set-ups require ample information on past family members' play. Such a perfect information setting, having information on all relevant actions of past generations, seems perhaps not the most natural. This part of the child labour literature, related to commitment problems and strategic exchange within generations, requires still closer scrutiny. The following chapter of this thesis intends to provide some answers to these open issues.

Applied work In terms of the consumption smoothing hypothesis, Skyt-Nielsen (2001) argues that parents take *a loan on the human capital market* instead of in the formal credit market. This is a very original way of analysing the problem. Parents might take these decisions in order to maintain their consumption levels, which are often close to the subsistence level and below the poverty line. Skyt Nielsen finds returns to schooling of 6-7% p.a. in Zambia, which should make an investment worthwhile. Thus other immediate constraints (such as a high preference for immediate consumption or lack of collateral) or commitment problems might prevent parents from realising these gains.

Rammohan (2001) showed that the development of financial markets and the increase in financial rates of return can lead to higher investment in schooling, a corresponding decrease in child labour, and also plays an important role in fertility decisions. The main channel through which financial markets affect the households in the model is through lower borrowing rates, brought about by more competitive financial markets.

Edmonds (2006a) studied the effects of an anticipated increase in income on child labour and schooling in a very detailed way, and within the same data context as Duflo (2000) above. He finds that the largely anticipated increase in pension income in South Africa leads to a significant decrease in child labour. This finding is consistent with liquidity constraints, but also with commitment problems and altruistic parental preferences. Indeed higher income will yield higher investment in child education through altruism, if we assume parents were credit-constrained initially. However, this is still consistent with commitment problems, since one could possibly conceive that informal credit market arrangements exist in South Africa and would allow even poor households to borrow. Thus the evidence is also consistent with parents not wanting to borrow because of commitment problems. It is very difficult to show the existence of such commitment problems, since they seem observationally equivalent to credit constraints. Edmonds (2006a) illustrates however, that credit constraints or commitment problems are a large determinant of child labour, and not only poverty or high returns to child labour, as often stated.

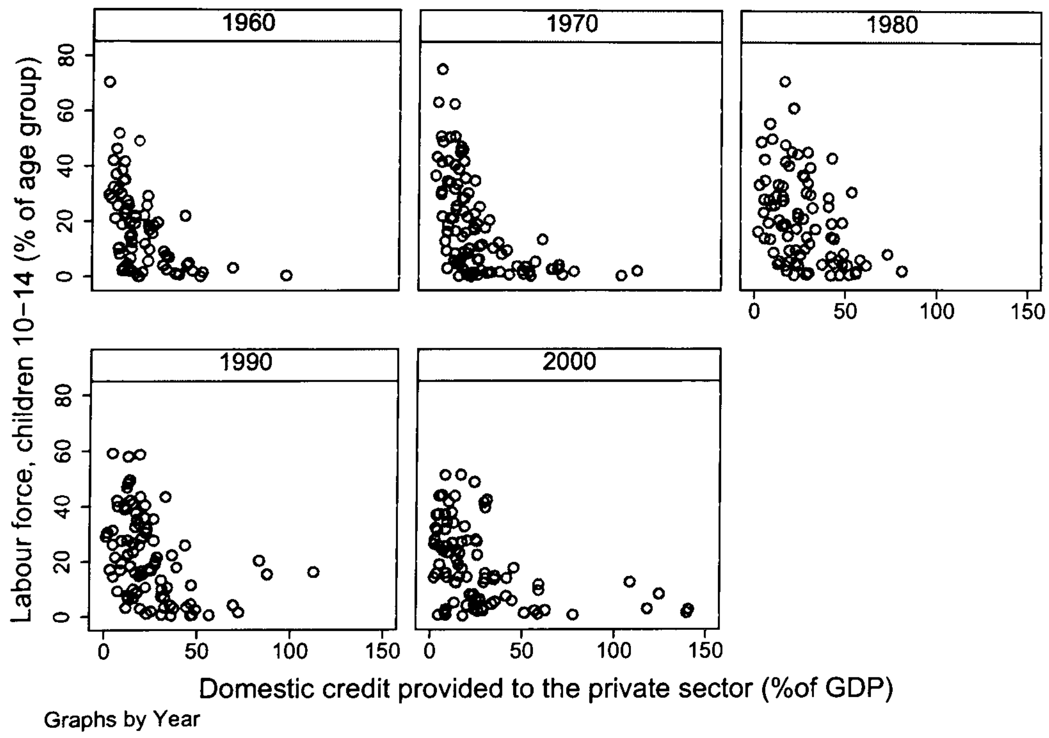


Figure 1.3: Domestic credit provided to the private sector (% of GDP) and labour market participation rates of 10-14 year olds.

Credit market imperfections seem to play a weaker role on aggregated data. As an illustration, we plot in Figure 1.3 a proxy of credit market existence (domestic credit as percentage of GDP) against child labour participation rates. It seems that there is a steep association in the past, for example in the 1960-1980's but this has become less distinguishable over time. This could be explained by the poor proxy measures of the selected variables. Another explanation could be the reliance of households on other means of credit than those provided by formal institutions, or also, of course, the existence of the aforementioned commitment problem.

1.4.3 Multiple equilibria

Theory Basu and Van (1998) and Basu (1999) have shown that child labour can arise as a consequence of multiple equilibria in the labour market. Multiple equilibria can be obtained by combining the already discussed "luxury axiom" with an assumption that child and adult labour are partially substitutable, the "substitution axiom".

The argument proceeds as follows. Suppose the adult wage is w, and the child wage is γw, a fraction of the adult wage, $0 \leq \gamma \leq 1$. For a given demand for labour in a *competitive* labour market, it depends on the prevailing adult wage rate if the children are sent to work or not. Suppose the wage is below some lower limit $\underline{W}$, which we will call the subsistence threshold. Then children are sent to work. This is implied by the "subsistence axiom": parents will send children to work only if income falls below this threshold. Note again that consequently parental altruism is assumed here.

In the market, there exists an equilibrium where all children and the parents work if the wage rate is at or below the subsistence threshold $\underline{W}$. If the wage were to rise, children are withdrawn from the labour market, because they earn only a fraction γW of the total wage, and adult income rises over the threshold. It is better that parents supply all the labour to reap the full benefits of the wage increase. At a sufficiently high adult wage, no children will work and all labour demand is met by adults. Hence, there are two stable equilibria. In the low wage equilibrium both children and parents supply labour, in the high wage equilibrium only adults supply labour. The most fascinating analysis of Basu and Van (1998) is related to the welfare effects of a child labour ban. A ban on child labour might increase the adult labour wage rate, if we suppose that there is no excess supply of adult labour on the market. Reducing child labour thus increases the demand for adult labour and the associated wage. Interestingly, a ban on child labour only improves welfare if the post-ban adult wage is sufficiently high to compensate a family for the lost child wages. The problem of the welfare effects of a ban resurfaces in many settings, for example as explained in the section on trade bans or labelling initiatives, policy needs to determine the welfare effects of such interventions. Basu and Van's (1998) model provides an elegant illustration that widely advocated bans might have unintended welfare consequences, i.e. when the adult wage does not raise sufficiently in response to the ban, the family might be worse off.

Another explanation for child labour can be found in unemployment, Basu and Van (1998). Assume the same model as above, with the additional feature that a fraction λ of workers own the firms in the economy and receive also profit income, Basu (1999). Then a fraction of workers $(1-\lambda)$ will earn only wage income.

The dividend-earning workers will never send their children to work. Thus, only poor workers will send their children to work: again, a necessity argument. If we consider now such an equilibrium where poor families (i.e only non-profit earning workers) send their children to work, unemployment will foster child labour. An intuitive explanation without going through all the details of the model is that the less likely it is that parents find work, the more they will rely on their childrens' wages to generate family income.

Another branch of the child labour literature, examining the role of child labour in growth models, has also given evidence of such multiple equilibria and development traps. For example, the theoretical contribution of Strulik (2004) relates child mortality to child quality investments. Suppose child mortality depends on economy-wide income. If child mortality is high and quality "unaffordable" (in the sense that it does not pay to invest in quality), parents will rationally not invest in quality and instead increase child labour. However this leads the economy as a whole into stagnation, since generations do not acquire human capital at all. High mortality remains prevalent because of stagnant income at the macroeconomic level. The economy can only take off if a new technology is introduced and makes income rise for some time. Then, less children die and parents invest gradually more in child quality, which in turn then also raises the human capital of the economy, and per capita income.

Another model of fertility and child labour by Hazan and Berdugo (2002) uses a similar argument. Under some restrictions, an economy might be in a poverty trap. High child labour and high fertility coexist. If the wage differential between parental and child wages increases, the value of child labour is reduced. This is essentially due to technological progress as well. Hazan and Berdugo (2002) come therefore to very similar results in a dynamic framework as Basu and Van (1998) in a static version. The wage differential between adult and child labour, or the adult wage alone, has to rise sufficiently. Moreover, the technological progress element underlying the work of both Strulik (2004) and Hazan and Berdugo (2002) can be compared to the approach to the topic of technological change on the demand side of child labour. This technological progress motivation is not yet well understood in applied work on the demand side, and leaves room for more investigation.

Applied work and discussion There has been no explicit test of the theory of multiple equilibria in the labour market. This seems to be an obvious empirical *lacuna* to be filled, and could be done by analysing legal measures against child labour and their general equilibrium effects.

However, there has been an empirical test of the growth models incorporating child labour and human capital formation by Hussain and Maskus (2003). The authors set up a very simple model of child labour where child labour exists, if the marginal benefit of child production exceeds the marginal benefit of education. Thus, Hussain and Maskus (2003) analyse the determinants of child labour and child education in a cross-section of countries. They estimate a system of equations as follows:

Human capital depends on past human capital, quality of schooling and past child labour:

$$\mathrm{HK}_t = \alpha_2 + \beta_4 \mathrm{HK}_{t-1} + \beta_5 \mathrm{COST}_t + \beta_6 \mathrm{QUAL}_t + \beta_7 \mathrm{CHILDLAB}_t + \epsilon_{2t}.$$

Child labour at time t depends on past human capital, cost of education and quality of schooling:

$$\mathrm{CHILDLAB}_t = \alpha_1 + \beta_1 \mathrm{HK}_{t-1} + \beta_2 \mathrm{COST}_t + \beta_3 \mathrm{QUAL}_t + \epsilon_{1t}.$$

Output depends on past human capital:

$$\mathrm{OUTPUT}_t = \alpha_3 + \beta_8 \mathrm{HK}_{t-1} + \epsilon_{3t}.$$

These recursive specifications are needed to identify the system uniquely. Hussain and Maskus (2003) employ the seemingly unrelated regression technique to allow for possible correlations between the error terms of the equations. One problem is that the exclusion restrictions used are derived from an underlying structural model, which might simplify the process of child labour accumulation significantly, i.e. why should present human capital not enter the child labour equation and why does child labour not enter the output equation? Although one can critisize the system approach, their model provides plausible estimates: past human capital influences the stock of child labour negatively (inter-generational effect) and lower school density (proxy for high costs) induces more child labour. The model is also a theory-based estimation, a precious species in this literature. It gives additional evidence that parents who face low school quality and experience high child labour will send their children as well into child labour - a poverty trap situation.

Another very interesting article in the literature on child labour and poverty traps is Souza and Emerson (2003). Here, the authors estimate the intergenerational persistence of child labour by examining parents and grandparents of child labourers. It appears that children are more likely to start work earlier if their parents were themselves child labourers or have lower education attainment. Also grandparents have an influence, since they decided on the human capital level of the parents. Souza and Emerson (2003) interpret this finding as a persistence effect, which goes beyond income and poverty alone and would point to norms and other forms of influence. However, it is likely that the father's or mother's child labour participation captures a lower lifecycle income of parents and is a good proxy for lower income. This would then be a reconciliation with the poverty hypothesis. On the other hand, it appears that returns to education are high in Brazil and thus children can potentially expect a much higher income. This is consistent with the view in the multiple equilibria literature, that there is a 'good' equilibrium of higher incomes without child labour.

As a matter of discussion, the results in the multiple equilibria framework seem plausible, especially the theoretical conclusions on labour markets and growth models. Empirical tests still lag behind and should be encouraged.

1.5 Norms, cultural factors and the political economy of child labour

Norms and traditions are often invoked in theorectical as well as in empirical work, Skyt-Nielsen and Dubey (2002), Skyt-Nielsen (2001), Souza and Emerson (2003). Norms and traditions can play a role in transfer behaviour within the family. This could be directly applied to a model similar to the one of Baland and Robinson (2000), where transfers are impossible due to commitment problems. Norms or agreed sanction mechanisms could overcome these shortcomings, Lopez-Calva (2001).

In terms of cultural factors, Lopez-Calva (2003) proposed a social stigma model of child labour. In very simplified terms, suppose that the social stigma cost $\theta(n)$ of child labour depends negatively on the number of persons who send their children to work. Consequently the social stigma cost is low if child labour is widespread and accepted as 'normal'. If there is however an increased awareness of the (potentially) negative effects of child labour or if there is a change in perception, some households might withdraw their children from the labour force. This increases social stigma costs to all other households in the model. Eventually, no child labour exists if social stigma costs rise sufficiently.

The pattern in Figure 1.4 illustrating child labour in rural and urban areas (symmetric graph) could well be the consequence of different norms in these two areas. It is clear that it could also reflect the different employment opportunities of children, or alternatively, different stages of development. This is open to discussion, but child labour in a country dominated by rural regions seems much higher than in an urbanized country.

Political economy contributions examine the interactions between citizens, voting behaviour and child labour. The enforceability of legislative bans on the demand side of child labour is analysed. Would citizens vote for such a ban? Families, employers, exporters and also the government might have similar and compatible incentives to maintain child labour, Anker et al. (1996). Tanaka (2003) and Doepke and Zilibotti (2003) develop models where restrictive child labour laws are not voted for or supported, because they induce welfare losses for the very poor, and because the poor are the

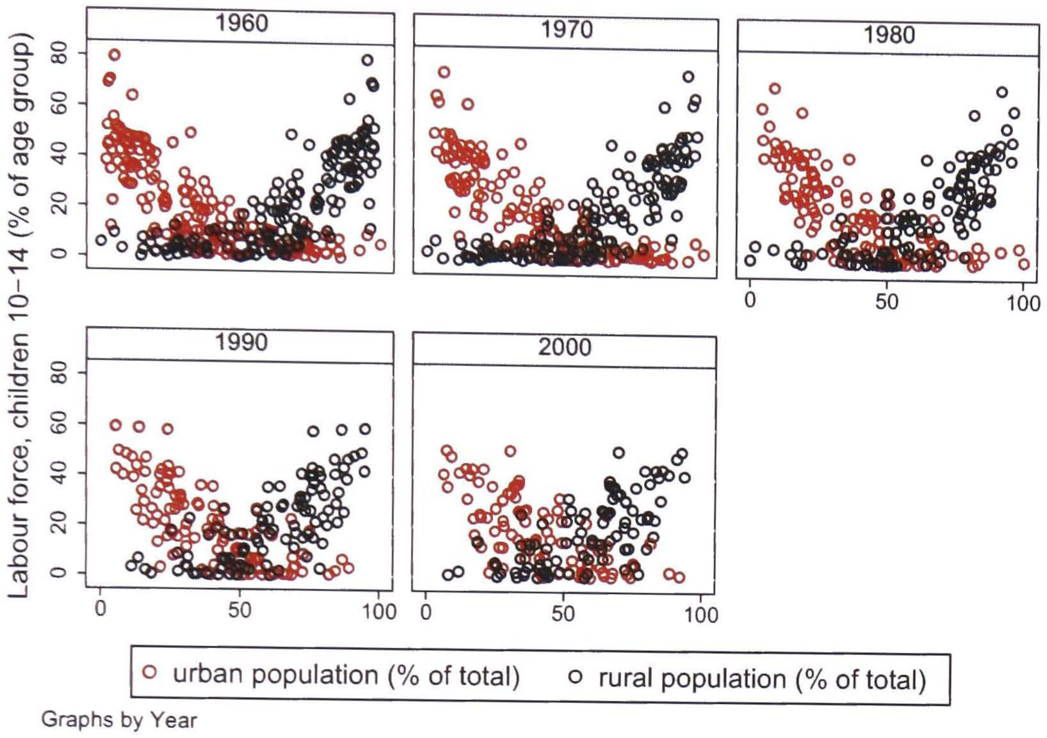


Figure 1.4: Urbanisation and labour market participation rates of 10-14 year olds in developing countries.

largest population fraction. This has important feedback effects on the demand and supply for child labour, since propositions to ban child labour might not be enacted into legislation. Even if bans are voted for or if the government is autocratic and dictates a ban, such measures subsequently run into enforcement problems. It is well known that most child labour activities take place in the informal sector, where child labour laws and other labour standards are generally not well respected. Thus, the beneficairies of informal sector production might have endogeneous reasons not to control or enforce such standards. Political economy contributions tend to emphasize that parts of the society have an interest in not voting for or not enforcing bans on child labour and maintaining the status quo. A slightly modified version of the model with an empirical test is provided by Maffei et al. (2006): political repression by the elite seems to impact child labour levels.

Summary and policy approaches Considering the weak empirical evidence on demand side interventions, additional research would be necessary into the role of technological progress, the hypothesis of child-specific skills and household production. In terms of possible policies let us first consider a ban on child labour or a ban on products made with child labour (trade-ban, labelling).

Theoretical research seems to agree that a trade ban or trade-liberalisation might have ambiguous effects on child labour. Labelling initiatives may be beneficial in preventing children from work, but they do not solve the immediate welfare effect on poor households which will forego this child income, Basu et al. (2006).

Kis-Katos and Schulze (2005) defend, along the lines of theoretical work on trade sanctions but also on the political economy of child labour, the argument that child labour regulation is unable to address the root causes of child labour, and that domestic policies are more suitable. As discussed in the section above on trade, sanctions are not targeted enough towards child labour intensive production, and penalise a whole export sector. Often, they are also perceived as the revival of protectionist reflexes by the developed North. In terms of the political economy of legal measures, two main features permeate from the theoretical literature: 1. Child labour may be desired by the elite, Maffei et al. (2006) or by a majority of the population if it entails major gains, Tanaka (2003) 2. Legal intervention will only be effective if it is well enforced, Basu (1999), Grootaert and Patrinos (1998), which again depends on the preferences of the population in point 1. In general, policy options on the demand side need further empirical research as to their effects. The labour supply model of Basu and Van (1998) demonstrates that, additionally, the welfare effects of a ban must be analysed with care. Basu and Van (1998) argue for the employment of bans only if they constitue a "benign" intervention, since a new, stable equilibrium and welfare improving equilibrium needs to be reached.

Thus the topic of a ban as a policy option links directly to a multiple equilibria setting. Bans can be effective, if there exist multiple equilibria. Which other policy options are suggested by multiple equilibria models? As seen above, growth models suggest interventions such as "education for one generation" or redistribution policies to solve the multiple equilibria problem in one generation. This then translates automatically into benefits for all following generations. The main conclusion of the literature on bans and multiple equilibria is that households, and whole economies, might be trapped in poverty. This implies high child labour and low growth in the future. This broad idea is then easily added to poverty as an additional determinant of child labour. It is surprising that there seems to be relatively little empirical evidence on the effects of bans on child labour. This seems an important issue to pursue in terms of policy options on the demand side.

We have reviewed the main evidence on the supply side of child labour. Poverty, the issue of insufficient credit markets, compounded by commitment problems, and cultural factors are all very important to explain child labour. What are the policy options on the supply side?

Clearly, fostering growth would enhance the general per capita income level and hence reduce child labour. However, growth is not always equitable. Therefore one could also imagine redistribution policies. A general discussion is which incentive structures should be created within redistributive policies. If poverty is a main determinant of child labour, direct household poverty relief might be a policy instrument. If on the other hand poverty is not important but relative returns to child labour and education, then the quality of education should be increased. Skyt-Nielsen (2001) argues that policies which provide education free of charge and also free meals to school-attending children have been a success. These have to be associated however with educational subsidies that replace the child's earnings from work to be effective, argue Grootaert and Patrinos (1998). Ray (2002) cites employment generation schemes in poor areas that were pushing poor households above the poverty line as successful in curbing child labour.

Addressing the market imperfections implied by the lack of credit and insurance markets seems another valuable approach. Although formal credit markets are incomplete in developing countries, informal credit markets are filling this gap. Hence, we would expect parents to be able to borrow, Udry and Pranab (1999). Especially in rural regions, where child labour is most prevalent, Heady and Bhalotra (2000), such markets exist in order to smooth out agricultural shocks. Hence, there ought to be other forces at work which could explain the high incidence of child labour, and a credit market expansion does not seem to be the most urgent option.

If commitment problems within the household are dominant, measures to remedy market imperfections will not be helpful. It seems difficult to empirically disentangle commitment problem effects from borrowing constraints. In the context of commitment problems, a credit market liberalisation might not bring about the desired outcome, since commitment constraints still bind.

Possible incentive structures to resolve commitment problems and credit constraints could be composed of transfer payments to parents conditional on school attendance of their children, raising the return to education through better quality of teaching and better schools. Conditional cash transfer programs seem to have had success, for example in Mexico with Progresa, OECD (2003). However, do we really understand the dynamic impact of such programs? The commitment problem is a dynamic problem, as it affects inter-generational transfers of both children and parents. Will a temporary transfer be enough to lift parents out of a bad equilibrium with no intra-family transfers to a better equilibrium with transfers? These questions still remain unanswered, yet answers are crucial for a well designed policy.

Various models have tried to assess whether parents have gender-dependent preferences over child quality, Edmonds (2006b). Also there is a growing literature that brings into question the unitary model of the household, Basu (1999). The collective model of the household points in both theoretical and applied work to important gender effects. These can be observed in the distribution of child labour and child schooling. Moreover, parents or grandparents' gender is thought to influence resource allocation, Duflo (2000). This could mean that traditional poverty alleviation could function more efficiently if transfers were made to the appropriate gender. It turns out that, empirically, transfers to women result in better outcomes in terms of child health. Another more theoretical policy recommendation within the bargaining literature appears to be parental education programmes. Additional education can rise the 'threat points' in the bargaining game and make women more influential. Especially the 'empowerment' of women could then prove a successful measure to fight poor child "quality" outcomes (poor health, malnutrition, no school attendance).

What has been done, what was succesful? Edmonds (2007) groups all possible programs directly affecting child work, and not being general development programs, into six categories: information campaigns, income replacement campaigns, flexible schooling programs, reintegration projects, restrictions on employment and conditional cash transfers. While Edmonds (2007) critisizes the lack of of peer-reviewed evaluations of most of the programs, except for the last two programs mentioned, another dramatic aspect of child labour policies is apparent. Not only are theory and empirical research often disconnected, but also with respect to policy, the implementation and cross-disciplinary social science research seem to be separated. To give only one example, Groves (2004) mentions only very seldomly the actual outcomes of the program implementing ILO Convention 188 in Honduras, but managerial aspects are in the forefront of the discussion. Certainly, the managerial and communication approach is crucial for a successful program, but it should be undertaken with consultation of other social scientists and, finally, proper evaluation. It seems that a cross-disciplinary approach to child labour is not yet established, and that, as Edmonds (2007) explains, scientific evaluation criteria and research are secondary when it comes to the practical implementation of programmes.

We have cited already some policy interventions that have worked. Although there is some modern evidence Dayiolu (2005) that schooling legislation can help to curb child labour, there has not been much other work on legal measures. It appears that the success of legal measures depends most likely on the degree of enforcement, Basu (1999).

Most programmes implemented are either school quality or conditional cash transfer programmes. Let us examine these in turn.

Educational quality also influences the return to work, thus school quality measures have been discussed to make child schooling more attractive. Figure 1.5 shows an association between school quality and child work in our panel data. For example, using a simulation exercise, Jafarey and Lahiri (2005) compare the option of facilitating credit to families to the option of an increase in school quality. The best policy option is dependent on the relative elasticity of credit supply. If school quality is increased, parents need to be able to forego income and send their children to better schools. Thus, a "food for education" program might be better if credit supply is nearly inelastic. Unfortunately, the authors brush aside the commitment problem. With commitment problems, school quality improvements will most likely not lead to an optimal increase in school enrolment in this model. Conditional cash transfer programs avoid the commitment problem in one generation, but do not address the commitment problem in future generations. The commitment problem is a key issue behind the effectiveness of many policy interventions.

Conditional cash transfers programs seem to work. They have also received proper evaluation. Famous examples are Bolsa Escola in Brazil, Bourguignon et al. (2003) or Progresa in Mexico, Schultz (2004). Schultz (2004) estimates that the Progresa program, which affects school enrolment, reduces household and market work by 4.3% points for girls at the secondary school level. The effect for boys at the secondary school level is slightly smaller, but an overall effect of the program is nevertheless a reduction of 2.4% in market and household work for boys. Schultz (2004) extrapolates that the program has increased baseline schooling from an average of 6.8 years, by 0.66 years. There was no effect of the program on the fertility of poor women.

The conditional cash transfer programme seems to work, because it directly affects the returns to schooling and work and avoids commitment problems. It is however difficult to distinguish the income effect of Progresa from the effect of the program on schooling costs. Similarily, it is not guaranteed that conditional cash transfers are the most appropriate policy, given the school quality improvement, legal ban, compulsory schooling or any demand side policy option.

Todd and Wolpin (2002) extend the evaluation of Progresa, by proposing an original method of validating a structural estimation model by a policy experiment. Todd and Wolpin (2002) then, at least in a partial equilibrium framework, simulate the effects of other policies on child labour. Interestingly, they mention the possibility that because more people attend school, the wage on the child labour market might rise. This would somehow mitigate the effect of school attendance, since the returns structure would be affected in the opposite way, i.e. by child work becoming more attrative. Simulating an increase in child wages by 25% associated with the Progresa program raises girls' enrolment in school only by 89% of the original treatment baseline and boys reach only 69% of the original baseline effect. Thus, this paper provides evidence of further effects of such cash transfer programs. The unknown response of child labour demand and the short time impact of the program might paint a too rosy picture. The program additionally does not really address commitment problems in a longer term or intergenerational setting, since the program is working only on a temporary base. Thus, for a successful policy, it seems crucial to understand the commitment problem using a dynamic approach. The most promising policy option might need to tackle three potential reasons for child labour: 1) cash transfers reducing the poverty constraints 2) artificially removing the commitment problem from the household by issuing such cash transfers with conditions, and 3) the policy must provide an inter-generational linkage so that current children know that their own children will have better opportunities as well, and that eventual returns from schooling will be shared across generations.

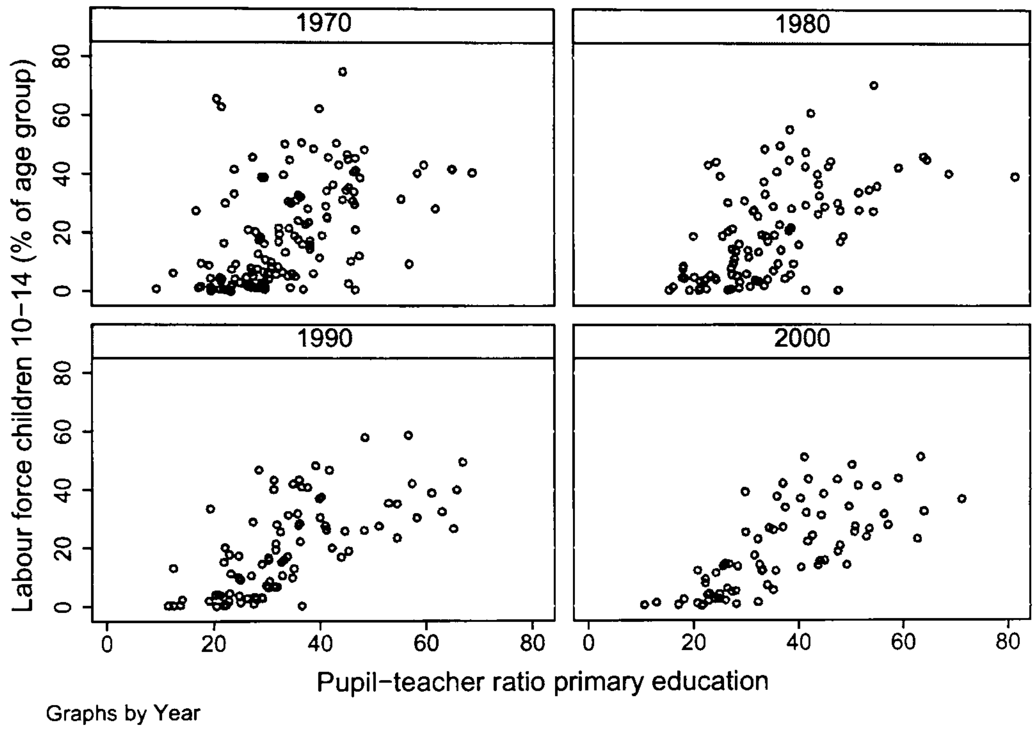


Figure 1.5: School quality (pupil-teacher ratio) and labour market participation rates of 10-14 year olds in developing countries.

Thus, from a policy point of view, conditional cash transfer programs seem succesful. Three challenges at least remain: better research and evaluation of demand side programs (affecting the return to child work, say, by introducing new technologies) seems urgent, as Todd and Wolpin's (2002) point to important (partial) equilibrium effects of increased school enrolment. Secondly, efforts have to be made to evaluate and experiment in general with a more diverse range of policy options. Thirdly, theory and practice have to be linked to implementation in a cross-disciplinary effort. Progresa in Mexico and Bolsa Escola in Brazil are example of a succesful cross-disciplinary and practically oriented approach, but these programs are too few, compared to the dimension of the global child labour problem.

1.6 Conclusion

Child labour is a phenomenon with multiple facets. Empirical research has provided an astonishing amount of possible answers to the economics behind the question of why children work. Some powerful explanations re-emerge in various country settings: Extreme poverty, market failure (lack of credit markets but also commitment problems) and parental preferences seem to be at the heart of the child labour problem. In addition, intergenerational persistence, non-perceived societal or non-realised private returns to education are, together with cultural and gender effects, the most salient theoretical arguments on the supply side of child labour.

A revival of theoretical and empirical interest in the demand side of child labour appears important. Especially the role of technological progress should be investigated in greater detail, but also models should be developed to evaluate the economic contribution and productivity of children.

Some remaining empirical and theoretical challenges need to be addressed. The modelling of the commitment problem of child labour has not been extended to a dynamic framework. Policy design should take into account the intergenerational aspect of commitment problems, as illustrated above, for example, by the choice between Progresa-type cash transfers and alternative policies. An additional question, namely if child laboureres are on their supply curves, remains unanswered.

At the moment of writing, child labour figures are high, and continue to rise in subsaharian Africa. It is not yet very well understood how child labour responds to policy interventions, except in the case of cash transfer programs. Moreover, the disconnection between theoretical and empirical research also seems to extend to a disconnection between practical implementation and research. Unless the root causes of child labour are clearly isolated, we cannot claim full economic understanding of the international child labour issue, and further research and policy evaluation are both necessary.

The next chapter aims to provide some theoretical insights by scrutinizing the commitment problem of child labour within an intergenerational and dynamic framework.

Chapter 2

Cyclic overlapping generation games: Grandparents' bequests and child labour

The children during the tender years of infancy are well fed and properly taken care of, and when they are grown-up, the value of their labour greatly overpays their maintenance. When arrived at maturity, the high price of labour and the low price of land enable them to establish themselves in the same manner as their fathers did.

Adam Smith, Part Second, Causes of the prosperity of New Colonies

2.1 Introduction

The report of Hagemann et al. (2006) estimates that worldwide more than 217 million children of schooling age are working. Child labour occurs primarily in developing countries, and thus poverty should, intuitively, play a role in high child labour participation rates. Children seem indeed to work out of poverty compulsions, but poverty is certainly not the only determinant of child labour. Child labour can be analysed as an alternative to child education, assuming the two goods are substitutes. Returns to education in developing countries are high and schooling seems thus a valuable alternative to child labour. Why do parents send their children to work instead of to school ? This chapter argues that commitment problems lie at the heart of this puzzle of unrealised returns to schooling. The commitment problem means, that once children have received education and earn higher wages, they might not repay the education expenditures that their parents incurred. Children cannot commit to parents to repay[1].

Credit constraints have often been cited, as well in developed countries, as an impediment to invest optimally in children. This is certainly an important aspect. Parents forego income when they send children to school, namely the child labour income, and have to pay eventually additional schooling costs. It is argued that the impossibility to borrow is a factor in the child labour decision, i.e. parents prefer to smooth their

[1]The commitment problem of children to pay back can also be viewed as an incomplete contract problem, as in Maskin and Tirole (1999) and Hart and Moore (1999). The parents have no property rights on the human capital that the child embodies once it is produced, so to say.

consumption through child labour, and not through a loan. In many countries where child labour is prevalent, official credit markets are imperfect, and thus, one rightly suspects missing credit markets as the culprit. On the other hand, and as is well documented by Rosenzweig and Evenson (1977), many informal arrangements of borrowing exist, especially in agricultural contexts. Thus, functioning credit markets might not as much influence the child labour decision as the commitment problem does, i.e. the impossibility for parents to recover debts from children.

Thus one is tempted to focus the attention on the commitment problem, as in Becker and Tomes's (1986) paper. This commitment problem, and not only imperfect markets, might be one of the root causes of child labour as Baland and Robinson (2000) already remarked. The present article extends their work by postulating the commitment problem as an overlapping generations problem.

This contribution intends to analyse the sustainability of child-parent transfers within a cyclic overlapping generations game. This game setting does not, in contrast to other work, require the conditioning of strategies on past behaviour of family members. Usually, strategies are formulated with respect to past actions of players in overlapping generation games, in so called trigger-strategy setups. The present game intends to capture the imperfect information family members might have of their ancestors' actions. Intuitively one can think about it as follows: a young member of a family might receive signals about past actions of family members. He is, however, unable to distinguish cheap talk from true information. We assume therefore, that such a young player does not condition his action on information of past play.

In order to sustain first best contracts of exchange, players commit to bequest a certain fraction of their income to the next generation, if the next generation honors the contractual obligation. We show that it is difficult to generate such bequest equilibria without the possibility to save over a long horizon. Grandparents have thus a special role in the model, since they can save longer than in a two generational setting.

Finally, and perhaps most interestingly, the model favours grandparents for additional reasons. In the model, there is a natural tension between committing credibly to an amount to bequest at the end of a generations life and the designation of an alternative beneficiairie in case of disinheritance. Disinheriting can be thought as costly to the parent, and even more so, if no credible alternative beneficairie exists. Grandparents can however transfer all stored resources credibly to their grandchildren instead of to their children. This re-inforces the contractual position of grandchildren. Grandparents also live long enough to observe if their children execute a *part* of the contract with their grandchildren, before releasing bequest resources. This "locks" parents into the contract and plays also a key role in limiting deviation options of players. The model also explicitly takes into account the transition from an uneducated to an educated generation. This yields the insight that although equilibria exist once a generation is educated, an uneducated generation might never be able to attain these equilibria.

In addition, the present game might provide a tentative answer to the puzzle of bequests vs. inter-vivos transfers. The puzzle consists in the empirical regularity of observing an equal division of bequests among children or a bequest to the oldest child (primogeniture), but unequal lifetime transfers[2]. We provide an explanation for equal bequest rules in developed countries or, to the opposite, primogeniture rules in developing countries .

A literature review section will be followed by a section describing first-best intergenerational transfers and the usual trigger strategy solution, which, as we show, relies on observability of the history of play. We then examine in the next two sections if first-best contracts can be enforced within the environment of an unobserved history of play. In the two-generational model, we establish a strong non-existence results of first-best contracts. First-best contracts can, however, be implemented in the three generational model, using bequest sanctions. Finally, policy discussions and conclusions are presented.

2.2 Some insights from the previous literature

This contribution encompasses different themes: the vast child labour literature, the literature on family contracts and bequests and game theoretic models ecountered in

[2]The model might also apply in developed country contexts. In poor families of developed countries, similar commitment constraints may bind, in addition to a low valuation of schooling by some parents. Thus bequests can be seen as a compensation for transfers by children, which could be in time or money. In general one can show that bequests need only be small compared to transfers from children, which makes this model credible in such contexts.

Country	Year of study	Mean years of schooling	Estimated return	Source
Bolivia	1993	.	10,7	Patrinos (1995)
Brazil	1989	5,3	14,7	Psacharopoulos (1994)
Burkina Faso	1980	.	9,6	Psacharopoulos (1994)
Cote d'Ivoire	1986	6,9	20,1	Psacharopoulos (1994)
Egypt	1997	.	5,2	Lambropoulos et. al. (1999)
Estonia	1994	10,9	5,4	Kroncke (1999)
Ethiopia	1972	6,0	8,0	Psacharopoulos (1994)
Finland	1993	.	8,2	Asplund (1999)
France	1977	6,2	10,0	Psacharopoulos (1994)
Germany	1988	.	7,7	Cohn and Addison (1998)
India	1995	.	10,6	Kingdon (1998)
Jamaica	1989	7,2	28,8	Psacharopoulos (1994)
Kenya	1986	8,0	16,0	Dabalen (1998)
Korea	1986	8,0	13,5	Ryoo et. al. (1993)
Kuwait	1983	8,9	4,5	Psacharopoulos (1994)
Malaysia	1979	15,8	9,4	Psacharopoulos (1994)
Mexico	1992	.	7,6	Psacharopoulos et al. (1996)

Table 2.1: Returns to education estimated from a Mincer type equation, selected countries.

overlapping generation games. This paper is novel on several accounts. Within the child labour literature, we develop a new model of dynamic commitment. In this model strategic bequests have a central role. We model the commitment problem as an overlapping generations cyclic game, relaxing the requirement to observe the history of the game fully. In order to situate this contribution within the wider literature, we give three brief accounts of the child labour literature, the literature on bequests and the overlapping generation games literature.

Child labour literature An extensive survey of the child labour literature can be found in Basu (1999), Bhalotra and Tzannatos (2003) or more recently in Edmonds (2007). Our contribution focuses on the commitment problem and only this part of the literature will be briefly reviewed here. Baland and Robinson's (2000) article mentions two possible causes for child labour: incomplete markets (in the sense of borrowing constraints or commitment problems) and poverty. The approach of Baland and Robinson (2000) assumes parents cannot expect any repayment for the education expenses of their children. This is the famous commitment problem.

In a static setting with altruistic linkage between generations, Baland and Robinson (2000) find that as long as bequests at the time of the child labour decision are positive and credit markets perfect, child labour will be at a socially efficient level. This is indeed a possibility but it assumes altruism from the side of the parent, since parents bequest out of an altruistic motivation.

We model a situation in which parents are selfish and do not care about child utility. Also, children do not care about parental utility. This is perhaps an extreme case, but serves as a benchmark for future modelling. Parents and children might only agree to establish a (mutually) beneficial contract between themselves. Baland and Robinson (2000) assumed a static model, where dynamics are ignored. In the present model, we also consider the dynamics of child labour in the parent-child game.

The commitment problem and its dynamic implications have not received much attention in the child labour literature. Lopez-Calva (2001) addresses the issue, but only to find that commitment problems are easily overcome through intergenerational trigger strategies. This is only possible if the entire history of play is observable or if the information transmission on past play is perfect. To address child labour and education decisions in the context of public policy we ought to understand the commitment problem better. This has to be done in an intergenerational context, taking into account the informational constraints of the intergenerational game.

Bequest literature The theoretic models of bequests or inheritance can be classified into at least four distinct categories, depending on the motive for bequests. The first hypothesis is altruism or in some variations "warm-glow" and "joy of giving" motivations. The testator leaves bequeathable wealth to his children because his utility is directly dependent on the beneficiaries utility or he simply enjoys giving. These models have been widely used in dynastic models in macroeconomics, Barro (1974), or in the economics of the family, Becker and Tomes (1986).

In second place, one can mention the hypothesis of Abel (1985) of precautionary savings and hence accidental bequests. An individual cannot predict his exact date of death but can form expectations about the likelihood of death - if an individual dies earlier than expected, he leaves an "unanticipated" or accidental bequest.

A third explanation for bequests is the motive for exchange or intergenerational reciprocity, Cigno (1993) and Anderberg and Balestrino (2003). Bequests are seen as part of a wider system of exchange within the family. There might be some informally or formally agreed system of exchange that each geneation replicates. For example, a father might want care and attention in his old age from his children, in exchange for the promise of a bequest.

This brings us towards the fourth point made by Bernheim et al. (1985); the strategic bequest motive. Parents or grandparents hold wealth in forms of bequeathable assets because they intend to influence the behaviour of their children. This comes very close to the model presented here. Bernheim et al. (1985) model the exchange of attention from a child as being related to the size of the expected bequest. The paper shows that the problem of an alternative beneficiarie overshadow this motive. Is it credible to disinherit a unique child ? Bernheim et al. (1985) show that this problem is resolved if they are two children and the parents can threaten each one of them with disinheritance. This is still problematic since not beqeathing to one child might still be emotionally costly and create an additional "envy" effect between children. Skipping one generation seems another means how to costlessly and credibly transfer bequest resources to alternative beneficiairies, and this is what grandparents can do in the model of this paper.

Empirical work has usually great difficulties to reconcile the theory of bequests to observed behaviour. The central question why parents leave wealth undistributed until they die remains, if not unanswered, then at least insufficiently clear. If the motivations for bequests were purely altruistic, then certainly it would be rational to transfer wealth *inter vivos* and not delay transfers until death. In addition, there is the so called "unequal transfers- equal bequests" puzzle in observed behaviour. That is, parents do seem to have altruistic motives for transfers during the lifetime of their children, i.e. they transfer more to relatively poor children and less to relatively wealthy children. However, most parents seem to divide bequests equally among siblings, McGarry (1997).

This behaviour is in disagreement with the altruistic bequest motive, since under this approach bequests should be compensating. Stark (1995) explains this phenomenon with interpersonal comparisons that siblings make *post mortem*. Then unequal division of bequests leaves some children relatively more deprived than the equal division of an estate. Other authors such as Bernheim and Severinoro (2000) claim that parents would like to leave a reputation of "fairness" after their death. Inter vivos transfers are likely not to be observable but bequests are observable. Hence by bequeathing an equal amount, parental preferences between children are perceived as equal or fair even though they might not be.

Empirical evidence confirms there being no unique motive for bequests. There seems to be rather a heterogeneous collection of motives with exchange and reciprocity being important. Page (2003) shows that bequests might be intentional, since inter-vivos transfers increase when inheritance, or estate tax increases. This is indirect and weak evidence. On the other hand, Perozek (1998) provides evidence to the opposite conclusion, while commenting on the results of the model of Bernheim et al. (1985). As soon as child characteristics are included in the model of Bernheim et al. (1985) attention given to parents seem not to depend on total bequeathable wealth but rather on the characteristics of the child. Hence, bequests would not be perceived as strategic and are perhaps not intentional. Other contributions have argued that bequests could at the limit be a product of parental indifference - but this is less likely since why would parents take the effort to specify an equal distribution?, Laitner and Ohlsson (2001).

It is also Laitner and Ohlsson (2001) who provide the first evidence that transfer and altruistic motives could be important both in Sweden and the US. The Swedish panel data and the PSID for the US used in this study are small samples, but rich in terms of income and wealth estimates, and allow the authors to exploit the panel dimension. The altruistic model seems to fit the data well, but with far smaller coefficients than theory would suggest. Hence the exchange motive could provide a better explanation for observed behaviour. This is also the quintessential conclusion of Arrondel and Masson (2002). The authors find that intergenerational reciprocity might explain most of the intergenerational transfers in France. Also, there is evidence that backward transfers (parent to child) are much more significant than forward transfers (child to parent). However, some of the forward transfers might be difficult to measure and hence any conclusions should be drawn with special care.

In summary, and following recent research and empirical evidence, bequests are to be intentional in most of the cases and they are usually divided equally among siblings. Bequests obey heterogeneous motives, where altruism and reciprocity appear as the most favoured explanations. The present contribution tries to provide a link between the reciprocity motive and the strategic motive to present a novel treatment of bequests. As will be shown in the cyclic overlapping generation game, bequests replace the information requirements on past play, and are thus strategic, but can also be seen as a reciprocity arrangement.

Related overlapping generations game theoretic models Cigno (1993) and Anderberg and Balestrino (2003), but also Smith (1992) and Rangel (2003) are typical contributions in this literature. The main determinants of intra-family transfers are the intra-family return on an investment in children compared to the capital market return of the same investment. If the former exceeds the latter, self-enforcing transfer contracts may exist.

Such models assume that families can implement trigger strategies. This supposes that a current generation can obtain information of actions/behaviour of past generations and is able to punish deviators by excluding them from the family system. Commitment problems are then resolved. Lopez-Calva (2001) is one example applied to child labour, but it could be that the informational constraints on such trigger strategy models are unrealistically strong.

The unrealistic nature of informational requirements relates to the way generations might transmit information from one to the other. The observability of the entire history of play is the basis of the trigger strategy approach. It is however not clear whether such a perfect information environment holds, since proofs of past play are clearly not obtainable. Only unverifiable messages of past play exist, which can be "cheaply" modified by parents. It is clear then, that children do not have reliable information on past play. This is the problem we address with cyclic games - assuming the child has (and does not need) information on past play. In this context, bequests can be seen as a signal. They ascertain that the next generation has behaved "as expected"[3].

Cyclic overlapping generation (OG in the following) games were thus chosen to model the interaction between family members. Cyclic games model a recurrent situation into which players can enter, and from which they exit after some time. Entry and exit can be unique (i.e. as in this case when we talk about generations) but we could allow multiple exit and entry (i.e. an individual trading in the financial market and entering and exiting it daily). Cyclic OG games can be seen as another representation of an infinitely repeated sequential OG game or also of dynastic OG games. Cyclic OG games require no memory of moves from previous players. Players are entirely forward looking during their lifecourse - they only consider strategies and the associated payoffs of the various possible cycles[4].

[3]This could have also an ambiguous meaning, in the sense that later generations might align the next generation on their preferences and enforce more "conservative" outcomes than without such sanctions. This is an interesting extension to the model presented here.

[4]Appendix A.3 gives a detailed review of the similarities and differences of cyclic OG games and the repeated OG game. It would be premature to present these comparisons here, since we have not formally introduced cyclic OG games yet.

2.3 Intergenerational contracting

In this section we describe the features of a first best intergenerational contract. The actual enforcement of a contract in a non-cooperative setting using trigger strategies is then examined.

We build a model on the work of Baland and Robinson (2000). In addition to Baland and Robinson's (2000) model we allow for child-parent transfers and an infinite horizon overlapping generations model. Unlike Baland and Robinson (2000), we assume that generations are selfish.

There are households composed of a single child and a parent. Parent and child are selfish and they only care about their own utility[5].

An individual lives for two periods. The first period is subdivided into two periods, the child and the young adult period. In the first subperiod, the individual is a child, which can either go to work or to school. In the second subperiod, we call the individual a young adult, which earns a wage and might decide to repay some of the schooling expenses to his parents. The second period of an individual is the parent period. Parents and children (child, then young adult) overlap and any such overlap will be labelled as period t, $t \in \mathbb{N}$. There is no discounting of future payoffs, as in Baland and Robinson (2000), for simplicity.

Parents can send their child to work or to school. If the parents send their child to work, the parents earn child labour income, $l_t w_t^{CH}$. The variable l_t is the labour supply of the child (in hours) in period t and w_t^{CH} represents the hourly wage. If the child is sent to labour, his human capital shrinks.

Let h_t represent the human capital production function at period t. The domain of this function lies on l_t with $l_t \in [0, l_{max}]$. Thus $h_t(l_t)$ represents the human capital attained by an individual in period t. Furthermore, h_t is a decreasing function in l_t. More precisely, $h_t'(l_t) < 0$ and $h_t''(l_t) < 0$. The function $h(l_t)$ has a lower bound $h(l_{max}) = h_0$,

[5]Parental altruism would make later derived cooperation results easier to sustain and the required bequest resources smaller but is not essential to such cooperation. Indeed Swinnerton and Rogers (2004) showed that even with reciprocal altruism, children do not necessarily transfer back to their parents. Once parents income is above a certain threshold, children consider that parents do not need transfers anymore, whereas parents would still like to receive them. A similar problem arises as well in our model. Indeed growth in (educated individuals) wages over time is necessary, otherwise children will not see the need to transfer to the next generation.

which denotes innate human capital. Figure 2.1 shows a hypothetical function h_t with these properties.

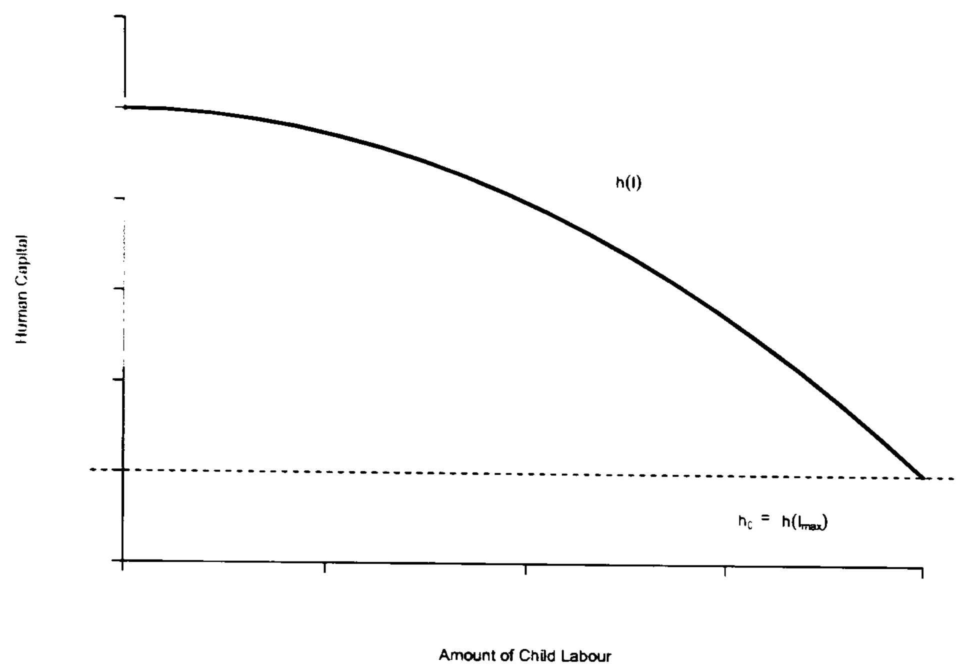


Figure 2.1: An example of a human capital production function and child labour.

Child labour has a cost in lower future human capital of the child. Human capital translates (for convenience) directly into wages. A parent and a young adult receive a wage according to their human capital accumulated as a child. Thus a young adult receives a wage equal to $h_t(l_t)$. A parent receives a wage equal to $h_{t-1}(l_{t-1})$, which is determined by the amount of human capital accumulated in the previous period while working l_{t-1}. The dynamics of this setup are shown in Figure 2.2

If the parent sends his child to school, he foregoes child labour income. We abstract here from school fees for simplicity, so the foregone child labour income is the only cost. In exchange for the foregone income he might expect some transfer, say $\alpha_t h_t(l_t)$, with $\alpha_t \in [0, 1]$, from the young adult in the future. Human capital theory would justify such a young adult-to-parent transfer. An educated young adult should earn significantly more than an uneducated young adult, even in a developing country context, see Table[6] 2.1. Furthermore, parents would like to obtain a compensation for their investment in education. We postulate that parents and children could agree to a

[6] This Table is a shortened version of a complete Table, published in a research paper by Psacharopoulos and Patrinos (2002)

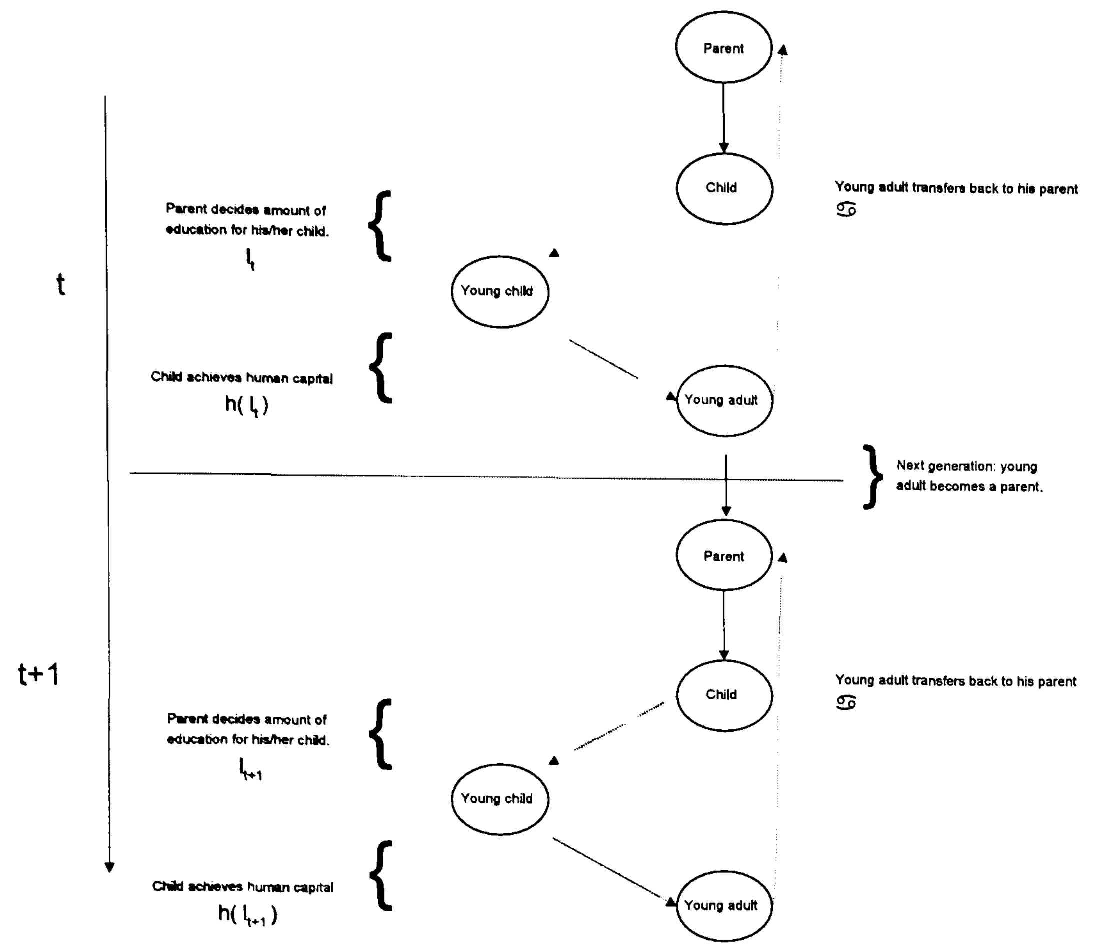


Figure 2.2: The division in child and parent period and actions of children and parents.

contract, in which parents will invest an amount l_t in the human capital of their children and young adults will transfer a fraction α_t out of their current income to their parents.

2.4 The model

Utility of a young adult in period t is given by

$$V_t(h_t(l_t)(1-\alpha_t)).$$

The utility function of a parent in period t is given by

$$U_t(h_{t-1}(l_{t-1}) + l_t w_t^{CH} + \alpha_t h_t(l_t)).$$

We normalise the utility in the child's potential schooling period, the child subperiod, to be zero i.e. $V^{CH} = 0$. The amount of equilibrium transfers α_t^* will depend on the investment l_t^*. Thus we write $\alpha_t(l_t)$ to make this dependency explicit.

We will investigate the first-best level of transfers, α_t^* and l_t^*. This is a joint maximisation problem, involving parents' and children's utility functions. We show in the appendix A.2, that the first best outcome is reached if children choose α^* first and parents chose l^* given α^*, given that a stationarity condition on α is satisfied and a maximum of the individual's utility function with $\alpha > 0$ exists.

2.4.1 The first-best solution

Thus, in the first best case, parents chose l_t and take $\alpha_t(l_t)$ as given. Thus, parents maximise

$$\max_{l_t} \; U_t(h_t(l_{t-1}) + l_t w_t^{CH} + \alpha_t(l_t)h_t(l_t)).$$

Parents chose l_t so that the following condition is satisfied for an interior solution in l_t^*

$$w_t^{CH} l^* + \alpha_t(l^*)h_t(l^*) \geq w_t^{CH} l_{max}, \tag{2.1}$$

that is chosing an $l^* < l_{max}$, parents must be at least as well off receiving (reduced) child labour income $w_t^{CH} l^*$ and the transfer $\alpha_t(l^*)h_t(l^*)$ as under full child labour income, $w_t^{CH} l_{max}$.

A young adult, taking as given l_t^*, chooses his contribution, α_t, as follows. The influence that an individual has on his lifetime utility is through α_t and l_{t+1}. His lifetime utility is given by

$$V_t(h_t(l_t)(1-\alpha_t)) + U_{t+1}(h_t(l_t) + l_{t+1} w_{t+1}^{CH} + \alpha_{t+1} h_{t+1}(l_{t+1})). \tag{2.2}$$

The interpretation of this equation is that an individual looks forward and chooses α_t with respect to his own future wage $h_t(l_t)$, the child labour income $l_{t+1} w_{t+1}^{CH}$ and the transfers he will receive from his children $\alpha_{t+1} h_{t+1}(l_{t+1})$, choosing himself the child labour involvement of his children in the next period, l_{t+1}. In order to solve this problem we assume stationarity, that is

$$\begin{aligned} l_{t+1} &= l_t = l \qquad \forall\ t \in \mathbb{N}, \\ \alpha_{t+1} &= \alpha_t = \alpha \qquad \forall\ t \in \mathbb{N}, \\ V_{t+1} &= V_t = V \qquad \forall\ t \in \mathbb{N}, \\ U_{t+1} &= U_t = U \qquad \forall\ t \in \mathbb{N}. \end{aligned}$$

Furthermore, we introduce an exogeneous wage growth rate g. This implies that wages at time t, here symbolised by human capital, are related to wages at time $t-1$

by a proportional and exogeneously given factor g. We will call this factor the wage growth rate. Thus $h_t(l^*) = gh_{t-1}(l^*)$ and $w_t^{CH} = gw_{t-1}^{CH}$.

We can then write the optimal α as the solution to the maximisation problem given l^*:

$$\max_{\alpha} V(h_t(l^*)(1-\alpha)) + U(h_t(l^*) + l^* g w_t^{CH} + \alpha g h_t(l^*)). \tag{2.3}$$

Notice that the subindex t only concerns the human capital production function and child wages, which are supposed to vary over time t. The first order condition (FOC) for α is then

$$-h_t(l^*)V'(h_t(l^*)(1-\alpha)) + gh_t(l^*)U'(h_t(l^*) + gl^* w_t^{CH} + g\alpha h_t(l^*)), \tag{2.4}$$

and initially, at $\alpha^* = 0$ we assume $V' < U'$, otherwise it would not be possible to improve total family utility by transfering. Thus there would be no positive transfer by definition and no mutually beneficial contract. Since V' is increasing in α and U' is decreasing in α, this will eventually give an upper bound on α^*. The above equation also defines the minimum growth g necessary to maintain the contract. We show now that this growth g must greater than one.

Suppose there is no difference between a child and a parent's utility function, $U = V = W$. Assume that M' is the inverse function of W'. Thus we have

$$h_t(l^*)(1-\alpha) = M'\left(gW'\left(h_t(l^*) + gl^* w_t^{CH} + g\alpha h_t(l^*)\right)\right).$$

Solution for specific functional forms

Lemma 1 *If W is a bijective function of the form $W(x) = \frac{x^{\alpha-1}}{\alpha-1}$ then $W'(x) = x^{\alpha}$, $\alpha \in \mathbb{R}$ but $\alpha \neq 1$. Let M' be the inverse function of W' such that $M'(W'(x)) = x$. Let $k \in \mathbb{R}$ be a scalar, then*

$$M'\left(kW'(x)\right) = M'(k)x,$$

Analogously

$$W'\left(M'(k)x\right) = kW'(x).$$

Proof

$$M'\left(kW'(x)\right) = M'(kx^{\alpha}),$$

with

$$M'\left(kx^{\alpha}\right) = (kx^{\alpha})^{\frac{1}{\alpha}} = k^{\frac{1}{\alpha}}x = M'(k)x.$$

Analogously

$$W'\left(xM'(k)\right) = \left(xM'(k)\right)^{\alpha} = \left(xk^{\frac{1}{\alpha}}\right)^{\alpha} = x^{\alpha}k = W'(x)k.$$

□

The above Lemma helps to solve for the transfer in a more general setting. Note that some of the utility functions comprised are popular choices such as $\ln(C)$, and the constant elasticity of substitution utility function $\frac{C^{1-\gamma}}{1-\gamma}$, and thus a fairly general selection. Using the previous result,

$$\alpha^{*} = \frac{h_t(l^{*}) - M'(g)h_t(l^{*}) - M'(g)gl^{*}w_t^{CH}}{M'(g)gh_t(l^{*}) + h_t(l^{*})}.$$

Thus for α to be positive the numerator of the preceding expression should be greater than zero, since the denominator is always greater than zero, as $M' > 0$, $g > 0$ and $h > 0$. The following condition must hold for a positive transfer to exist:

$$h_t(l^{*}) - M'(g)h_t(l^{*}) - M'(g)gl^{*}w_t^{CH} > 0,$$

and thus

$$M'(g)h_t(l^{*}) + M'(g)gl^{*}w^{CH} < h_t(l^{*}).$$

It follows

$$\frac{M'(g)}{(1 - M'(g))} < \frac{h_{t-1}(l^{*})}{l^{*}w_t^{CH}},$$

$$M'(g) < \frac{h_{t-1}(l^{*})}{l^{*}w_t^{CH} - h_{t-1}(l^{*})}.$$

Re-applying the result from proposition 1, we have

$$g > \frac{W'(h_{t-1}(l^{*}))}{W'(l^{*}w_t^{CH} + h_{t-1}(l^{*}))}.$$

From the properties of W' it follows that $g > 1$. Furthermore g depends on the returns to child labour and human capital at the optimum l^{*}.

In order for future transfers to be possible, growth in wages over generations will be necessary. Otherwise, no positive transfers exist[7].

Proposition 1 *The first-best contract* $C^* = (\alpha^*, l^*)$ *has* $\alpha^* > 0$ *whenever* $g > \frac{W'(h_{t-1}(l^*))}{W'(h_{t-1}(l^*)+l^* w^{CH})}$. *In this contract* $l^* = h'^{-1}\left(\frac{M'(g)w^{CH}}{1+M'(g)}\right)$ *and* $\alpha^* = \frac{h_t(l^*)-M'(g)h_t(l^*)-M'(g)gl^* w_t^{CH}}{M'(g)gh_t(l^*)+h_t(l^*)}$.

If $g < \frac{W'(h_{t-1}(l^*))}{W'(h_{t-1}(l^*)+l^* w^{CH})}$, *no such contract exists with* $\alpha > 0$.

Proof *If Equation (2.1) is not true,*

$$w^{CH} l + g\alpha h(l) < w^{CH} l_{max}.$$

then parents chose l_{max} *and there is no contract* C^*.

If Equation (2.4) is not true,

$$g < \frac{W'(h(l^*))}{W'(h(l^*) + l^* w^{CH})}$$

then $\alpha \leq 0$. *If* $\alpha = 0$, C^* *is of no interest and* $\alpha < 0$ *is not sustainable under the model.* □

The set of a possible contract is well defined by equations (2.1) and (2.4). There is however the well known additional commitment problem, Becker and Tomes (1986). The commitment problem can be summarized as follows: No child has a direct incentive to repay his parents if there are no altruistic links between generations. The child is always better off not repaying his parents. If the parent can however not recover a transfer payment from his child (and if he is sufficiently selfish), parents will not send children to school or make an optimal investment in child education.

Once educated the child will prefer to transfer $\alpha = 0$ and in anticipation of this choice, a parent will chose at worst l_{max}, the highest amount of work hours possible for their child or at best no schooling at all. We will call this situation the *null* contract $C^0 = \{\alpha = 0, l = l_{max}\}$. The literature has so far argued that these problems can be

[7]Since future transfers are absent from the static setup, this is already an interesting extension. If growth prospects are uncertain in developing countries, this might be an important element in tackling child labour from a macro-economic perspective. Such growth over time could imply a similar game theoretic solution as in the usual macroeconomic literature. Every generation transfers their whole wealth to the generation before and makes everybody better off. This is not possible in the cyclic setup, since we would again need strong information requirements on past play and the history of moves. We will discuss the strong information requirements below, but see also the Appendix.

overcome in family or dynasty games, using intergenerational trigger strategies. We outline these trigger strategies below. We then argue, that trigger strategies imply perfect information about past play which is unlikely to hold.

In order to analyse the existence of a beneficial contract between generation, the previous literature will be followed, e.g. Rangel (2003) by assuming generations can play intergenerational trigger strategies.

2.4.2 Equilibria with full observability

Simple trigger strategies Suppose there is a strategy space S for every individual, implying that a young adult picks an $\alpha \in [0,1]$ and a parent picks an $l \in [0, l_{max}]$. A particular strategy of S can be denoted $s \in S$, of which $s^* = (\alpha^*, l^*)$ is a possible element. s^* has each young adult choose α^* and each parent l^*, the first-best contract. Define then, as in Rangel (2003) a history function $\mu(h_t)$, where h_t denotes the history of play of overlap t. The history of play can either be cooperative (C), deviation (D) or punishment (P). Assume the history of play is known, i.e. at least α_{t-1} needs to be observable in this set-up[8].

A simple trigger strategy prescribes the following

$$\begin{aligned} s(h_t) &= (\alpha^*, l^*) \text{ if } \mu(h_t) = C, \\ s(h_t) &= (0, l^*) \text{ if } \mu(h_t) = P. \end{aligned}$$

and μ can be defined recursively as follows:

Definition 1

$$\begin{aligned} \mu(h_t) &= C \textit{ if } \mu(h_{t-1}) = C \textit{ and } \alpha_{t-1} \geq \alpha^*, l_{t-1} \leq l^*, \\ \mu(h_t) &= C \textit{ if } \mu(h_{t-1}) = P \textit{ and } \alpha_{t-1} < \alpha^*, l_{t-1} \leq l^*, \\ \mu(h_t) &= P \textit{ otherwise.} \end{aligned}$$

[8] In any model, children cannot directly observe what parents did when parents were children

The idea is that μ keeps track if the organisation is in a cooperative or a punishment phase, Rangel (2003). Now, other conditions need to be fulfilled in order to be able to sustain such a first best equilibrium. The first condition concerns deviation options, the second condition concerns informational requirements.

The minimum necessary conditions for the existence of a contract are (2.1) and (2.4). The contract needs also to be beneficial to every generation, which is the deviation constraint:

$$W(h_t(l^*))+W(h_t(l^*)+l_{max}w^{CH}) < W(h_t(l^*)(1-\alpha^*))+W(h(l^*)+l^*w^{CH}+g\alpha^*h_t(l^*)) \tag{2.5}$$

We will call this condition the intergenerational return constraint. The intergenerational return constraint (2.5) states that once a child is educated (here symbolised by the educated wage in the parent equation) it must prefer to still apply the contract. Note that this condition is not necessarily satisfied by the optimality conditions, as the optimality conditions apply only to chosing an l_t which is equal for parents and children. The optimality requirements does not allow the choice of l_{max} for children and an l^* for parents, as above.

Proposition 2 *The first-best contract can be implemented using simple trigger strategies if Equation(2.5) holds and under full observability, (Rangel 2003)*

Proof *Consider the simple trigger strategies, $s(h_t)$, as defined above. If the history of play was cooperation, then a player should abide by cooperation, this yields*

$$W(h_t(l^*)(1-\alpha^*)) + W(h(l^*) + gl^*w^{CH} + g\alpha^*h_t(l^*)).$$

If a young adult deviates and sets $\alpha < \alpha^*$, then he should be refused a transfer α^* by the next generation, a punishment phase P is initiated. The young adult will then obtain, choosing l^* for his children:

$$W(h_t(l^*)) + W(h(l^*) + gl^*w^{CH}).$$

This will be, by optimality, less than the outcome of cooperation. The next generation will, under full observability, identify the player, who played a punishment as a "police *men" and future generations will resume cooperation with him. The execution of the punishment is beneficial to the "police" player since*

$$W(h_t(l^*))+W(h(l^*)+gl^*w^{CH}+g\alpha^*h_t(l^*)) > W(h_t(l^*)(1-\alpha^*))+W(h(l^*)+gl^*w^{CH}+g\alpha^*h_t(l^*)).$$

Now the deviating adult could, as a parent, also chose $l = l^{max}$ for his children. This would yield

$$W(h_t(l^*)) + W(h_t(l^*) + gl_{max}w^{CH}),$$

as a worst possible deviation. Thus Equation (2.5) ensures that such a deviation is not profitable, since by Equation (2.5),

$$W(h_t(l^*))+W(h_t(l^*)+gl_{max}w^{CH}) < W(h_t(l^*)(1-\alpha^*))+W(h(l^*)+gl^*w^{CH}+g\alpha^*h_t(l^*)).$$

The same scenario applies now to the parent who cooperated. Suppose a parent deviates after transferring α^* and chooses $l > l^*$. Then, by the simple trigger strategy definition, we enter immediately a punishment phase and parents will not receive a transfer α^* from their children , whereas the children can be identified as "police persons" again - i.e. as carrying out a justified punishment. Thus for $l > l^*$ parents would receive

$$W(h(l^*)(1-\alpha^*)) + W(h(l^*) + glw^{CH}).$$

which is less than the optimal amount under cooperation.

Cooperation in each generation can be enforced. □

Informational requirements and observability The above game of trigger strategies assumes that the history of play is known, at least up to the last two generations.

Bhaskar (1998) and Yoon (2001) give the same example concerning the importance of the assumption of an observable history of play. Consider a trigger strategy in which every player can only remember what the previous generation did. So a player in period t can only remember what a player in period $t-1$ did. Player t must give a transfer to $t-1$ when he is young. If he fails to do so, then he must be punished by player $t+1$. In the above set-up this corresponds to playing P. However, when player

$t+1$ initiates the punishment sequence, player $t+2$ cannot distinguish if the agent justifiably initiated a punishment or if that agent simply deviated. Player $t+2$ cannot verify what player t did, because he does not observe the history of play. Thus, agent $t+1$ will also be punished in any case, since this is the best reply. Thus agent $t+1$ has an incentive to conceal player t's deviation and do as if nothing happened. Clearly then player t will play a deviation - thus there is no other equilibrium than the non-cooperative Nash solution in this context. This has been referred to as the "Anti-Folk Theorem", Bhaskar (1998). Thus the above proposition breaks down, and no trigger strategy equilibrium can exist if the history of play is not observable, at least up to two generations. Cyclic games push this logic to an extreme, since in cyclic games, there is absolutely no memory. Some other equilibria are possible, because cyclic games feature sequential moves and potential sanction threats, such as disinheritance in this model.

Baland and Robinson (2000) invoke the threat of parents not to bequest to children once they are grown up and did not transfer. Such a threat requires enough parental resources, but also credibility. We analyse a stylised model of two generations, in which parents cannot save and compare it to a three generational model, in which parents can save and live one period longer: as grandparents. This addresses the resource part of the constraint. In order to address the credibility part, we suggest that, as grandparents overlap with their grandchildren, they are more credible enforcers of such a bequest equilibrium. In fact, grandparents observe part of the actions of their children towards grandchildren, and can thus condition the bequest on their children's behaviour towards their grandchildren. They can also much more credibly commit to bequest all the reserved wealth to their grandchildren in case their own children deviate from the norm. They might do this by costless transfers at the end of their life, which is not in contradiction to our definition of a bequest. This bequest rule will have the additional effect that grandchildren will be better endowed to enforce contracts with their own children.

Before we continue with the model, we will define bequests as used in this paper.

Definition 2 *Bequests in this model will be assets that are clearly earmarked for transmission to the next generation. An individual cannot consume them once they are* *earmarked. However, they have the property that they can be diverted to the recipient of the individuals' choice at the end of his life and are perfectly observable.*

In summary, this paper does the following: it asks if the first-best contractual situation can be implemented using bequest sanctions in a setting where there is no information on the history of play. The information on past play is the crucial distinction between the game outlined here and trigger strategy games as above. Thus we relax the observability assumption and see what can be sustained in a non-cooperative setting with bequest threats. In order to address the limited observability, we model the decision problem as a cyclic game. Cyclic games have no memory and players enter them without knowledge of the history of play.

We start with a two generational model, where parents and children overlap. We show that in these models, the first best contract cannot be implemented even with bequest sanctions. In a three generational model, we show that the first-best contract may be implemented using a bequest sanction. The bequest shifts the commitment problem from children to grandparents. We assume grandparents can write conditional, legally enforceable, bequest contracts. The bequest constitutes a costly signal, replacing the missing history of play but the bequest also needs to be large enough to favor payoffs of cooperating individuals over individuals that deviated.

The intuitive reasons why the model works in a three generational context are: (i) bequests grow in this model (savings from parent to grandparent period), (ii) grandparents have a credible and costless alternative beneficiairie: grandchildren, (iii) grandparents can observe the actions of parents towards their children, and limit the deviation options of parents. Thus, grandparents solve the credibility problem of commitment and grandparents solve the information problem of the game.

2.5 The extended model: cyclic games

We use cyclic overlapping generation (OG) games to model the decision problem. Cyclic OG games are related to repeated overlapping generations game (OLG-game in the following). In this OLG-game literature, optimal outcomes are often enforced through the use of trigger or grim trigger strategies. As we have seen, the transfer game between parent and child can support first best contracts if we use trigger strategies. These trigger strategies require information on the past behaviour of family members. In a family setting, past behaviour is not verifiable, since most players are not alive yet, when the action(s) took place. The players have thus to rely on messages provided by previous players. Since messages about previous actions can be seen as cheap talk, the inference to be drawn from such messages is limited: every previous family member has an incentive to pretend he was executing the contract. We assume consequently that children cannot base there inference on these messages, since they are costless. Children therefore do not take messages on past behaviour into account. However, bequest resources can replace such noisy information, as they will be transferred only upon cooperation. Cyclic games are perfectly suited for modelling this situation as i) players are entirely forward looking ii) the game is not a simultaneous move game but a dynamic game in which bequest threats conditional on previous moves make sense. Cyclic games allow also a graphical representation of repeated dynamic games. Instead of having an ever expanding tree, one obtains a cyclical pattern. For simplicity, I consider first the two generational cyclic game. Then I extend the game to allow for three generations, or grandparents.

2.5.1 The two-generational model

Assume the child entering the game at overlap t needs to make the same decisions as a child in overlap $t+k$, $k>0$ - their strategies are the same. Every player is a reincarnation of the previous player. The only change in every generation occurs through the exogeneously given growth rate. We simplify the problem by discretizing the strategy space as discussed in the previous section. A cyclic game is defined as in Selten and Wooders (2001)

$$G = (X, Z, E, P, A, \pi, W), \tag{2.6}$$

where a set of decision points, X, is determined and assumed to be finite. There are payoff points in the game, Z, whose set is finite. Exit points are defined as a subset of payoff points, $E \subset Z$.

P is a partition in subsets of $X \cup Z$ such that $P = (X_0, X_1, ...X_n, Z_1, ...Z_n)$, where $\cup X_i = X$ and $\cup Z_i = Z$. Furthermore X_0 is called the set of random points and all other points X_i for $i \neq 0$ are called the set of decision points for player i. Each set Z_i contains a non-empty subset of exit points, E_i.

The function A maps from subsets $X \cup Z$ to subsets $X \cup Z$. Hence A(.) is a directed graph linking decision points with decision points and/or exit points. π describes a probability distribution over the directed graphs of A.

W assigns a payoff to every payoff point.

We change this set-up to adapt it to a cyclic OG context. The following describes the game in an informal way, before introducing it more formally.

We start with the two generation model to examine sustainability of the first best contract. Assume two generations are alive at any one time, a parent and a child. Each parent has a single child for simplicity.

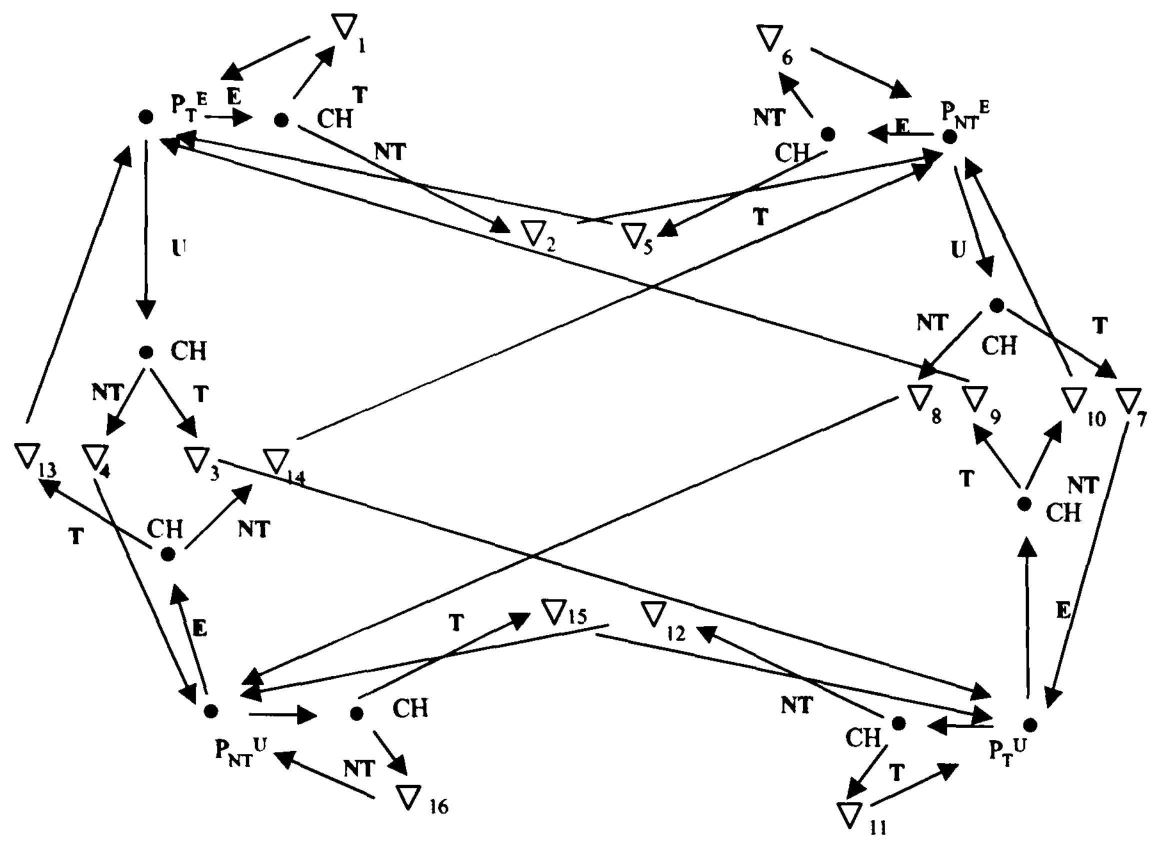


Figure 2.3: A representation of the infinitely repeated game as a cyclic game.

Let us first look at the game in a fairly general way, without describing a particular path. Start the game with a child player that looks forward onto the choices open to him during his lifetime. As a child, he must make a first decision, to transfer T or not to transfer NT. Every decision in a cyclic game is made at a decision point. In Figure 2.3, there are eight child decision points $\bullet^{CH}$. The game can start from any of these points. After making his decision, the child will become a parent and thus reach a parent decision point, $\bullet^{P}$. In Figure 2.3 there are four parent decision points. As a parent the player must decide to educate, symbol E, or not educate, symbol NE, his own children. The player has then made all his decisions in his lifetime. Before he exits and receives a payoff in one of the payoff points, symbol ∇, he must await the decision of his child to transfer to him or not. A parent payoff obviously depends on this choice. Note that there are sixteen payoff points in the game of Figure 2.3 and they have subindex i, $i \in \{1, ..., 16\}$, indexing the values of the different payoffs, *not* the player receiving them. At these payoff points, a parent also exits, thus in this game the sets E and Z contain the same elements, $E = Z$.

Of course, a child player could start the game in any of the eight child decision points of Figure 2.3. Observing the child starting the game on one of these decision points depends on previous moves by the parents. The child ignores this history. He observes only if he was educated or not. Indeed, a child enters the game *before* a parent makes the decision to educate or not to educate. Thus a new child player enters the game when an old, previous parent player exits. A child enters the game at the preceding exit point of a previous parent ∇. The history beyond this first move is irrelevant to the child's entirely forward looking decision. This makes the game very different compared to standard OLG games.

Let us now examine a particular path of the game. Suppose we start the game with a child player index t in the top left corner of Figure 2.3. This child has been educated by his parents $t-1$, symbolised by the arrow, superscripted E and pointing from the parent to the child. The child makes a decision at his decision point. Suppose the child decides to transfer. In this case, the game reaches an exit point for the parent $t-1$ subscripted ∇_i. In this point a parent $t-1$ receives a payoff and exits. Also, a new child $t+1$ is born into the game at this point. Also, in this point the previous child player t becomes a parent $t+1$. As a parent, the player must decide to educate or not educate his children. Suppose the parent decides to educate his children. Then the game will continue with the decision point of his own child. Suppose the child decides to transfer. Then the game reaches the exit point for player t, a parent t where he receives a payoff.

We can denote this payoff as being dependent on the strategies played by players $t-1$, t and $t+1$. Thus, in the particular point we described in the above paragraph, we have payoff point ∇_1, yielding payoff $W(s_{t-1}, s_t, s_{t+1}) = W(E, \underline{T, E}, T)$, depending on the decision of a parent of the previous generation $t-1$, a player of generation t (first a child, then a parent) and a child $t+1$. We have underlined the decision of player t. The exit points are labelled as in Figure 2.3 and the payoffs are represented accordingly in Table 2.2.

Subscripts are omitted in Figure 2.3 for clarity. Cyclic games have the important property, that we can cut out the individual players' graph from entry to exit point of a player and obtain a finite tree. To clarify the time structure of the game more precisely and the property of obtaining a finite path for every player, consider figure 2.4.

A payoff symbolised by $W(E, \underline{T, E}, T)$ would imply that a player was educated by his parents, the player then transferred as a child, educated his children and then received a transfer from his own children. As the education or transfer decisions are binary, there are sixteen (2^4) different exit points for an educated player and in total there are 16 exit points in this model.

In a more forrmal setup and respecting the previous notation we could write:

$$G = (X, Z, E, P, A, W), \tag{2.7}$$

where the $X_i = \{CH_j, P_T^E, P_{NT}^E, P_T^U, P_{NT}^U\}$ $j \in \{1, ..., 8\}$, $Z_i = \{\nabla_1 ... \nabla_{16}\}$. In this game $E_i = Z_i$, P being the partition in decision and exits points of Eq. (2.7) and A being the directed graph depicted in Figure 2.3. W is the payoff function as defined in the payoff table. We assume there to be only pure strategies, thus the probability distribution π is degenerate and equal to 1. The strategy set of every player is defined as

$$S = \{\{T, NT\}, \{E, U\}\}.$$

refering to his possible strategies as a child (first entry) and as a parent (second entry) of the strategy vector. As mentioned, we discretise the strategy space, and assume that a young adult played T when he exactly transfers $\alpha \geq \alpha^*$. Similarily, we assume a parent exactly played E when he made his children work no more than the equilibrium amount of child labour, $l \leq l^*$. In all other cases, we say that the young adult did not transfer NT, or the parent did not educate, NE.

A first task is to characterise the equilibrium in pure strategies for this two generation overlapping generations (OG in the following) cyclic game. The appendix A.1 shows the derivation of this property from the finite normal form representation of two-generation cyclic games and Nash's (1951) symmetry definition.

2.5.2 Equilibrium in the two generation model

Equipped with the tools from symmetric games, we can now search for an equilibrium point in our cyclic 2-generation OG game. The previous section outlined the features of a first best contract $\mathcal{C}^* = (\alpha^*, l^*)$. We will now show that, without any sanctions, the first-best contract is not implementable due to commitment problems. The payoffs of the game at the payoff points are given in table 2.2. The last column of this table duplicates the first entry of each of the payoff functions W in column 3, namely the move of parents $t-1$ of child t. The table is also split in two parts, depending on the decision of parents in $t-1$, in order to emphasise that a player t will make a decision conditional on either being an educated player or being an uneducated player.

Proposition 3 *The two generation cyclic game has a unique equilibrium in pure strategies with each player playing $s_i = (NT, NE)$, each player exiting in exit point ∇_{16}*

Proof *We must consider only symmetric equilibria. Every player could play (T, E), (NT, E), (NT, NE) or (T, NE). (T, E) cannot be an equilibrium, since (NT, E) does better if everybody else was playing (T, E). This can be seen as $W_5(E, NT, E, T) > W_1(E, T, E, T)$. (NT,E) cannot be an equilibrium since (NT,NE) does better from $W_8(E, NT, NE, NT) > W_6(E, NT, E, NT)$. and finally (T, NE) cannot be an equilibrium, since it would pay to deviate to (NT, NE) if everybody else was playing (T, NE) since $W_7(E, NT, NE, T) > W_3(E, T, NE, T)$.*

Hence (NT, NE) is the only symmetric equilibrium. As the ranking of the payoffs in the exit points is preserved the pure strategy equilibrium is not affected by this transformation.

Each player must therefore exit in exit point ∇_{16} since

(i) the same reasoning with respect to strategies applies in the bottom half of the table, exit points $\nabla_9 - \nabla_{16}$, since the ranking of these payoffs is equivalent to the ranking of $\nabla_1 - \nabla_8$.

(ii) no player t would like to educate the first generation of children.

□

Parents have thus an incentive to design contracts punishing their children if children do not transfer, if this would overcome the commitment problem. Parents could, for example, not make a certain transfer to their children: we consider here the bequest transfer as a special case. Thus we are examining if the first-best contract $\mathcal{C}^* = (\alpha^*, l^*)$ can be implemented using a bequest sanction.

It is common knowledge of all players that their path leads inevitably to exit point ∇_{16}. Thus uneducated parents would be willing to forego up to the difference in payoffs received in exit point ∇_{13} and those received in exit point ∇_{16}, that is:

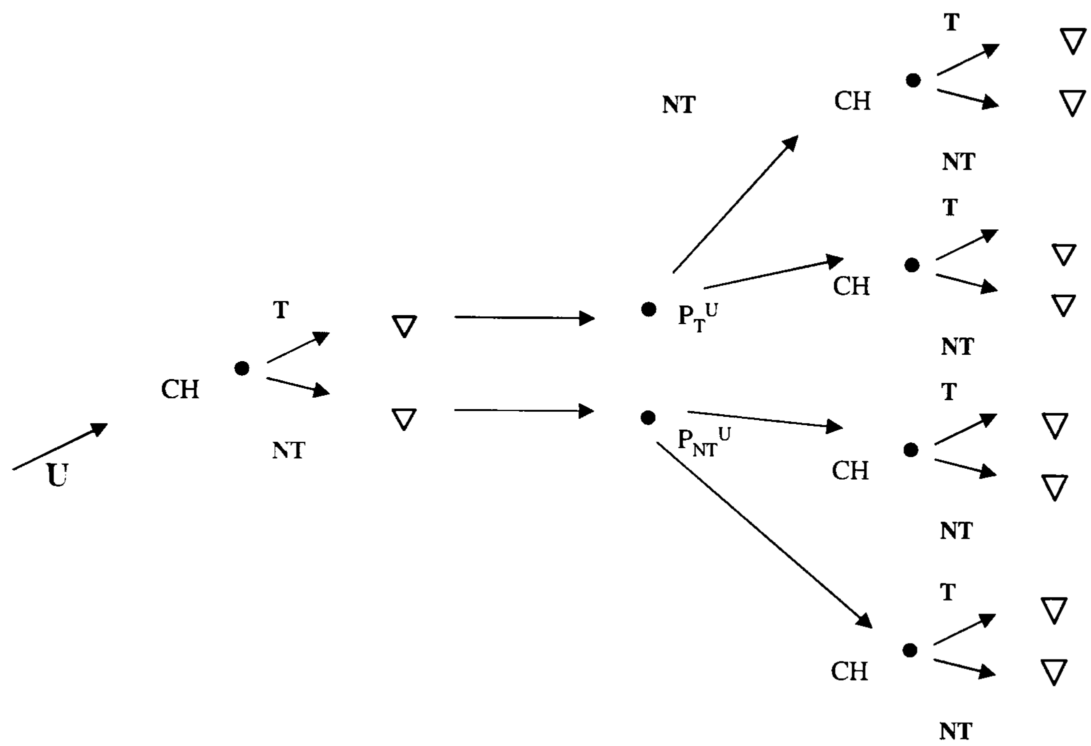
$$W(NE, NT, E, T) - W(NE, NT, NE, NT)) > \delta > 0,$$

if this would help to implement the contract with transfers. Indeed, if they could enforce the contract of transfers by not consuming an amount δ and receive the payoff in ∇_{13}, $W(NE, NT, E, T) - \delta$, they must be better off with a contract that prescribes transfer α^* and education l^* than under the no contract, $\mathcal{C}^0$, situation. This definition can be compared to an individual rationality constraint: an individual cannot be worse off implementing the contract than without any contract.

The following definition formalises how much parents would be willing to give up.

Definition 3 *A parent is willing not to consume a part of his income $\hat{B}_t$, called bequests, and transmit the bequest to his children such that*

$$W_t(NE, NT, E, T) - \hat{B}_t > W_t(NE, NT, NE, NT)$$



In period t, a child enters the game and observes his parent chosing no education (U)	In period t , a child decides to transfer or not to transfer (T/NT)	At the end of period t, a parent of period t-1 exits the game and receives a payoff	A child of period t becomes a parent. He educates or does not educate his children (E,U)	In period t+1 , a child decides to transfer or not to transfer (T/NT)	Parent born in period t exits and receives a payoff

Figure 2.4: Obtaining a finite graph: the graph of an individual that was not educated within the cyclic game, cut on the entry (child) and exit point (parent) of the individual

Exit point	Payoff symbol	Payoff dependent on strategies	Payoffs period t	Own Parents strategy
∇_1	W_1	$W(E,T,E,T)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*))$	E
∇_2	W_2	$W(E,T,E,NT)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH})$	E
∇_3	W_3	$W(E,T,NE,T)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l_{max} w^{CH} + g\alpha^* h_t(l_{max}))$	E
∇_4	W_4	$W(E,T,NE,NT)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l_{max} w^{CH})$	E
∇_5	W_5	$W(E,NT,E,T)$	$W(h_t(l^*)) + W(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*))$	E
∇_6	W_6	$W(E,NT,E,NT)$	$W(h_t(l^*)) + W(h_t(l^*) + l^* w^{CH})$	E
∇_7	W_7	$W(E,NT,NE,T)$	$W(h_t(l^*)) + W(h_t(l^*) + l_{max} w^{CH} + g\alpha^* h_t(l^*))$	E
∇_8	W_8	$W(E,NT,NE,NT)$	$W(h_t(l^*)) + W(h_t(l^*) + l_{max} w^{CH})$	E
∇_9	W_9	$W(NE,T,E,T)$	$W(h_t(l_{max})(1-\alpha^*)) + W(h_t(l_{max}) + l^* w^{CH} + g\alpha^* h_t(l^*))$	NE
∇_{10}	W_{10}	$W(NE,T,E,NT)$	$W(h_t(l_{max})(1-\alpha^*)) + W(h_t(l_{max}) + l^* w^{CH})$	NE
∇_{11}	W_{11}	$W(NE,T,NE,T)$	$W(h_t(l_{max})(1-\alpha^*)) + W(h_t(l_{max}) + l_{max} w^{CH} + g\alpha^* h_t(l^*))$	NE
∇_{12}	W_{12}	$W(NE,T,NE,NT)$	$W(h_t(l_{max})(1-\alpha^*)) + W(h_t(l_{max}) + l_{max} w^{CH})$	NE
∇_{13}	W_{13}	$W(NE,NT,E,T)$	$W(h_t(l_{max})) + W(h_t(l_{max}) + l^* w^{CH} + g\alpha^* h_t(l^*))$	NE
∇_{14}	W_{14}	$W(NE,NT,E,NT)$	$W(h_t(l_{max})) + W(h_t(l_{max}) + l^* w^{CH})$	NE
∇_{15}	W_{15}	$W(NE,NT,NE,T)$	$W(h_t(l_{max})) + W(h_t(l_{max}) + l_{max} w^{CH} + g\alpha^* h_t(l^*))$	NE
∇_{16}	W_{16}	$W(NE,NT,NE,NT)$	$W(h_t(l_{max})) + W(h_t(l_{max}) + l_{max} w^{CH})$	NE

Table 2.2: Representative payoffs for player t in the cyclic game, no bequest move introduced.

Additional notation We adopt the following notation: The consumption of a parent t is denoted by C_t. A parent may chose to bequest a certain fraction of his current consumption to his children, denoted by $\beta \in [0,1]$. Thus the bequest is $B_t = \beta C_t$. For convenience, we denote the consumption under the first-best contract as $C_t^* = h_t(l^*) + l^* u^{CH} + g\alpha^* h_t(l^*)$, and we calculate all bequests with respect to this consumption. An equilibrium contract will be denoted by $\mathcal{C} = (\alpha^*, l^*, \beta^*)$, specifying the fraction α^* to transfer, the amount of child labour l^* and the amount of bequests β^* in every period $t \in \mathbb{N}$.

In addition to this, we have the contractual rule set by parents' commitment to a bequest: every child that transferred receives a bequest and every child that did not transfer receives no bequest. Somewhat contrary to intuition, such a contract might equalise parents' ex-post payoffs, and this suffices to implement the equilibrium. In equilibrium, every parent generation receives a bequest $\beta^* C_{t-1}$ from the previous generation and makes a bequest $\beta^* C_t$. Off the equilibrium path, deviating children (those that did not transfer) receive no bequest.

The equilibrium may exist, as payoffs received in ∇_1 are lowered by a smaller amount than the payoffs received in ∇_5 for every fraction of consumption used as a bequest. Indeed, every cooperating individual receives a bequest in ∇_1, which in turn raises his payoff whereas a deviating agent does not receive such a bequest. Both agents have to commit to the same bequest in order to receive transfers and thus a cooperating agent may be better off, depending on the size of β.

The intuitive reasoning why the contract could be stable is as follows: every educated child knows that the best possible deviation payoff under the contract, ∇_5, implies implementing the proposed contract with the next generation. Thus, even if a child did not transfer, it still needs to commit to a bequest if it wants to implement the contract with the next generation. Suppose a child wants to follow this path, and deviates by playing NT in the young adult stage. The child becomes a parent. How much does the parent need to promise the next child in order to receive the transfer ? This suggests that β^* must be such that $W_5 < W_1$, by which a parent equalises his own payoffs in ∇_5 and ∇_1 respectively.

In the next paragraphs, we present three tables. Table 2.3 reproduces the top part of the previous payoff table. Table 2.4 is the same Table as 2.3 but lists payoffs received, if a previous generation $t-1$ implemented the proposed contract $\mathcal{C}^*$ and the generation t did not make any commitments. Table 2.5 finally presents the payoffs under the contract $\mathcal{C}^*$, including a bequest choice variable in the strategies. The letter B denotes the action of bequeathing to children and the letter NB denotes the action of bequeathing to an *alternative* beneficiairie. We assume here that parents have commited to a bequest fraction β^* *before* they educated their children. Thus, they will always lose this amount of consumption, and hence, it is always subtracted from their payoffs. We assume also that the decision to bequest can be seen as a automatic choice, conditional on actions of the children. This follows from the definition of the bequest - once the amount of income is stored it can only be re-directed, not consumed. We assume thus, that although children cannot credibly commit to repay, parents can credibly commit to bequest, conditional on child actions. We will re-examine the credibility statement later. Note that since this bequest move is modelled as automatic, the number of payoff points does not change.

The next proposition shows the impossibility of implementing the contract in a non-cooperative setting with bequest sanctions.

Exit point	Payoff symbol	Payoff dependent on strategies	Payoffs period t	Own Parents strategy
∇_1	W_1	$W(E,T,E,T)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*))$	E
∇_2	W_2	$W(E,T,E,NT)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH})$	E
∇_3	W_3	$W(E,T,NE,T)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l_{max} w^{CH} + g\alpha^* h_t(l_{max}))$	E
∇_4	W_4	$W(E,T,NE,NT)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l_{max} w^{CH})$	E
∇_5	W_5	$W(E,NT,E,T)$	$W(h_t(l^*)) + W(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*))$	E
∇_6	W_6	$W(E,NT,E,NT)$	$W(h_t(l^*)) + W(h(l^*) + l^* w^{CH})$	E
∇_7	W_7	$W(E,NT,NE,T)$	$W(h_t(l^*)) + W(h_t(l^*) + l_{max} w^{CH} + g\alpha^* h_t(l_{max}))$	E
∇_8	W_8	$W(E,NT,NE,NT)$	$W(h_t(l^*)) + W(h_t(l^*) + l_{max} w^{CH})$	E

Table 2.3: Representative payoffs for player t who had been educated, with player t not committing to bequest B_t^*.

Exit point	Payoff symbol	Payoff dependent on strategies	Payoffs period t	Own Parents strategy
∇_1	W_1^R	$W(E,T,E,T)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*) + \beta^* C_{t-1}^*)$	E
∇_2	W_2^R	$W(E,T,E,NT)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH} + \beta^* C_{t-1}^*)$	E
∇_3	W_3^R	$W(E,T,NE,T)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l_{max} w^{CH} + g\alpha^* h_t(l_{max}) + \beta^* C_{t-1}^*)$	E
∇_4	W_4^R	$W(E,T,NE,NT)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l_{max} w^{CH} + \beta^* C_{t-1}^*)$	E
∇_5	W_5	$W(E,NT,E,T)$	$W(h_t(l^*)) + W(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*))$	E
∇_6	W_6	$W(E,NT,E,NT)$	$W(h_t(l^*)) + W(h(l^*) + l^* w^{CH})$	E
∇_7	W_7	$W(E,NT,NE,T)$	$W(h_t(l^*)) + W(h_t(l^*) + l_{max} w^{CH} + g\alpha^* h_t(l^*))$	E
∇_8	W_8	$W(E,NT,NE,NT,NB)$	$W(h_t(l^*)) + W(h_t(l^*) + l_{max} w^{CH})$	E

Table 2.4: Representative payoffs for player t who had been educated; player $t-1$ complying with the contract, player t receiving B_{t-1}^* upon a transfer but no commitment at the end of player t's life.

Exit point	Payoff symbol	Payoff dependent on strategies	Payoffs period t	Own Parents strategy
∇_1	W_1^B	$W(E, T, E, T, B)$	$W(h_t(l^*)(1-\alpha^*)) + W((h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*)(1-\beta) + \beta C_{t-1}^*)$	E
∇_2	W_2^B	$W(E, T, E, NT, NB)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH} + \beta^* C_{t-1}^*)$	E
∇_3	W_3^B	$W(E, T, NE, T, B)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l_{max} w^{CH} + g\alpha^* h_t(l_{max}) + \beta^* C_t^* - \beta^* C_{t-1})$	E
∇_4	W_4^B	$W(E, T, NE, NT, NB)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l_{max} w^{CH} + \beta^* C_{t-1}^*)$	E
∇_5	W_5^B	$W(E, NT, E, T, B)$	$W(h_t(l^*)) + W((h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*)(1-\beta^*))$	E
∇_6	W_6^B	$W(E, NT, E, NT, NB)$	$W(h_t(l^*)) + W(h(l^*) + l^* w^{CH})$	E
∇_7	W_7^B	$W(E, NT, NE, T, B)$	$W(h_t(l^*)) + W(h_t(l^*) + l_{max} w^{CH} + g\alpha^* h_t(l^*) - \beta^* C_t^*)$	E
∇_8	W_8^B	$W(E, NT, NE, NT, NB)$	$W(h_t(l^*)) + W(h_t(l^*) + l_{max} w^{CH})$	E

Table 2.5: Representative payoffs for player t who had been educated with players t and $t-1$ complying with the contract, receiving B_{t-1}^* upon a transfer and bequest B_t^* at the end of his life conditional on a transfer.

Remark 1 *The bequest is defined as a time-constant fraction β out of total assets under cooperation in time t, C_t^*. Consumption in equilibrium grows at rate g, thus bequests grow at rate g, i.e. $B_{t-1} = \beta C_{t-1}^* = \frac{\beta}{g} C_t^*$.*

Non-existence of equilibrium

Proposition 4 *(i) The fraction of consumption C_t^*, β^*, must be set such that $W_1^B > W_5^B$, $W_1^B > W_8$ and $W_1^B > W_4^R$ for an equilibrium contract $\mathcal{C} = (\alpha^*, l^*, \beta^*)$ to exist. (ii) In the two-generational model there is no solution.*

Proof *To (i): For a solution to exist $W_1^B > W_5^B$, $W_1^B > W_8$ and $W_1^B > W_4^R$. Consider other strategy choices than applying the contract, $\mathcal{C}^*$.*

An educated child that does not transfer can at most expect to receive W_8 if it does not commit to a bequest. The player is worse off than in the proposed equilibrium contract $\mathcal{C}^*$, if β^* can be set in such way that $W_1^B > W_8$.

An educated child that transfers, but does not educate his/her children, receives W_2^R. This would be the best possible deviation, given the child has cooperated. Thus the contract needs to rule out this situation. By setting $W_1^B > W_4^R$, this ensures implicitly that $W_1^B > (W_2^R, W_3^R)$.

Consider now the "partial contract" situation, i.e. a child only implements the contract with generations ahead of him. Suppose a child decides to deviate from the proposed equilibrium; the child does not transfer in the first period, but still intends to implement the proposed equilibrium in the second period (as in W_5^B) by bequeathing an amount $\beta < \beta^*$.

Every next generation child and parent knows that the best possible contract can only be a contract that establishes $W_1^B > W_5^B$, i.e. ex-post indifference between payoffs, and no generation will accept to repay if the bequest is $\beta < \beta^*$. Hence a child that deviated has to subscribe to a contract leaving at least β^* for the next generation to cooperate, but this implies $W_1^B > W_5^B$.

To (ii):

The curves W_5^B *and* W_1^B *cross, if ever, only once in* β^*. *Initially, we have at* $\beta = 0$, $W_5^B > W_1^B$. *Consider the partial derviatives with respect to* β*:*

$$\frac{\partial W_1^B}{\partial \beta} = (\frac{1}{g} - 1)C_t^* W'(C_t^*(1 - \beta + \frac{\beta}{g})),$$

and

$$\frac{\partial W_5^B}{\partial \beta} = -C_t^* W'(C_t^*(1 - \beta)).$$

As $C_t^*(1 - \beta + \frac{\beta}{g}) > C_t^*(1 - \beta)$, *the marginal utilities* $W'(C_t^*(1 - \beta + \frac{\beta}{g})) < W'(C_t^*(1 - \beta))$, *additionally* $\frac{1}{g} >, 0$ *this implies* W_1^B *decreases at a slower rate than* W_5^B. *Thus, the functions may cross. As they are strictly monotonic and decreasing in* β, *they may cross, if ever, only once.*

A trivial feature of a crossing is that $W_1^B > W_5^B$ *beneath the crossing point. If the curves do not cross, we must have still at any* β *that* $W_5^B > W_1^B$. *We can now use the feature of the crossing of the curves to prove that no contract exists.*

Define $\hat{g}$ *as the growth* g *implying a positive transfer* α. *Suppose all the conditions for a beneficial contract are met from the previous section. Setting* $\beta C_t^* = \hat{B}_t = l^* - l^{max} + \hat{g}\alpha^* h(l^*)$, *we have that* $W_8 = W_5^B$. *Thus,* W_5^B *is already at the value of the payoff* W_8. *If we can show that* W_1^B *is below* W_8 *at this point, then the curves did not cross before and no contract exists. Now at* $\hat{B}_t$,

$$W_1^B = W(h_t(l^*)(1 - \alpha^*)) + W(h(l^*)(1 - \alpha) + \frac{(l_{max} - l^*)}{g} w^{CH}),$$

which is strictly below W_8.

□

The intuition for this impossibility result is relatively straightforward. Consumption of every generation grows by the exogeneously given rate g: however, the bequest sanction does not grow at any rate. The bequest is only transmitted from generation to generation. It becomes difficult for an individual to equalise ex-post payoffs of his children in this context. The possibility of savings, or the possibility to let bequest resources grow at an interest rate, changes this fundamentally. We allow grandparents to save their resources and this stylised fact helps them to sustain such contracts. Also, there are the deviating options open to parents, i.e. receiving the bequest and sending their children to full child labour (W_4^R). If grandparents overlap with parents, they can observe the parents and condition the bequest on education of grandchildren.

We illustrate proposition 2.5.2 in the Figure 2.5, where we depict the non-existence of equilibrium in this parametric example (log-utility function) for various values of g and β. As can be seen, the green cooperative utility plane W_1^B is always below deviating utility planes.

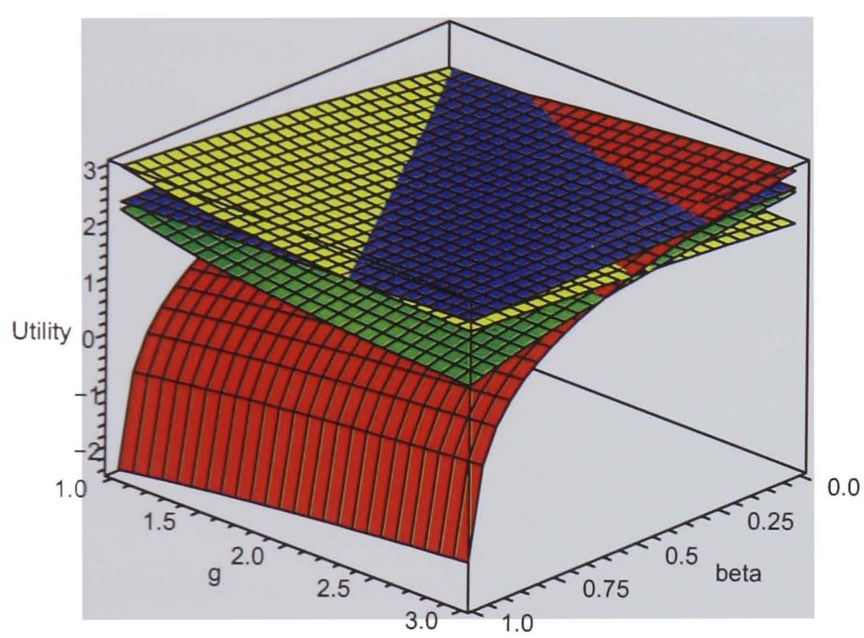


Figure 2.5: The values of W_1^B (green), W_7^B (red), W_8 (green) and W_4^R (yellow) to illustrate non-existence of an equilibrium

Another reason why grandparents are important is related to the credibility of sanctions[9].

Suppose that establishing an alternative beneficiairie for the bequest outside the family is costly, i.e. psychic costs as well as monetary costs. Within the family, only children can be beneficiairies. As mentioned by Bernheim et al. (1985) siblings could be punished separately. However, this would still not allow contracts in our model and might entail psychic costs. If there are costs of disinheriting natural children and in the event children do not cooperate, parents must be worse off bequeathing to the

[9]Bequests to the parents own brothers and sisters are not credible, since they as well might have only a limited time to live when the bequest is made (in our model they do not exist, but if they existed, they would theoretically die at the same time as parents.

alternative beneficiaire rather than giving the money to the children directly. This will be true even though children did not cooperate. To illustrate this, let ϵ_t be such small costs, then

$$U(h(l_{max}) + l^* w^{CH} - B_t^* - \epsilon_t) < U(h(l_{max}) + l^* w^{CH} - B_t^*).$$

That is the utility from bequeathing to an alternative beneficiairie is lower than bequeathing to own children even though they deviated.

Definition 4 *If costs of disinheriting are small, $\epsilon_t > 0$, no generation would like to commit to a bequest B_t^*.*

Clearly the mere existence of a grandparent generation resolves this problem as grandparents overlap with grandchildren. Grandparents can transfer bequests to grandchildren if the parents have not been applying the optimal contract. This seems a credible use of the bequest, since grandchildren will benefit from these measures. Also, the dynasty as a whole might benefit, since resources stay within the dynasty.

The overall intuition why contracts might be enforceable with grandparents is (a) grandparents might be able to accumulate more wealth because they can save over a longer horizon, (b) grandchildren provide a good alternative use of the bequest, (c) grandparents can control their grandchildren's education and reduce deviating options[10].

2.5.3 The extended model: a three generational cyclic OG game

We have established that without the full observability of the history of play and despite having bequest sanctions, the first-best contract cannot be implemented in a two generational setting. In fact, payoffs are strictly decreasing in the received bequest and this determines the result. If we take bequest costs and/or the observability of whether

[10]Of course one could think also of reasons why grandparents might be worse for grandchildren, e.g. grandparents might foster the maintaining of more conservative values than parents. In addition grandparents might consume all the resources for bequest themselves if longer lives go hand in hand with higher medical or care costs. Articles by Edlund and Rahman (2005) point this way - grandparents seem to encourage more investment in what she admits to be public goods such as health of children than parents. But grandparents invest less than parents in private goods with which she thinks of education. One could argue that education in the context of exchange is not a exclusive private good, but parents and grandparents have interests to develop it further due to the transfer that they receive.

grandchildren are educated into account, grandparents seem to be able to take a more prominent role.

Grandparents usually leave resources for their offspring. It is an unresolved question why most leave bequests instead of giving inter-vivos transfers. There is evidence that grandparents might leave such resources for strategic reasons [11] e.g. maximising attention dedicated to them Bernheim et al. (1985) or in order to receive old age care, Wagener (2002).

Introducing grandparents into the model The model of two generations outlined before extends naturally to a model of three generations. We can conceptualise grandparents as long-lived parents. The parents live one period longer, and can save at (gross) interest rate R for their retirement. Their payoff will be (under certain conditions on R) increasing in the received bequest. In addition, grandparents overlap with their grandchildren, and can thus condition the receipt of a bequest on the education decision of parents. Grandparents "lock" parents into the education contract. Also, costs of disinheriting are zero, since grandparents can change beneficairie costlessly to their grandchildren.

The total utility of a player is then given by

$$W_0(h(l_t)(1-\alpha_t)) + W_1(h_t(l) + l_t w^{CH} + g\alpha_t h(l_t) - s_t) + W_2(Rs_t - c),$$

which consists of a child utility, a parent utility and a grandparent utility.

The labelling is unchanged, except that c denotes the minimum consumption that grandparents need to spend for formal sector old age care. R represents the gross

[11] Evolutionary economics and biology also provide explanations why grandparents can have different preferences over investment in grandchildren than parents. Grandparents might be less selfish and more altruistic towards children, since they have "survived" above reprodutive age and therefore would like to favour the survival of others of their dynastyy. They maximise survival of their dynasty and not of their own survival. In a purely biological view this would mean to increase either quantity or the quality of offspring, Robson (2003). Grandparents will thus try to increase the quality of offspring (children and grandchildren). Parents might still be more concerned with their own survival and hence privilege this over their childrens' survival. Grandparents should feel particularly worried about their grandchildren in this biological context, since they are potentially at risk of being exploited by (more "evolutionary" selfish) parents. In this context we do not take these evolutionary arguments into account, but rather focus on the implications that grandparents have through their longer life horizon (savings) and observability of parents' actions.

interest rate. Similarly to the previous section, grandparents are willing to commit to a bequest of at least the difference in payoffs between the outcome without any contract and the outcome with contract.

Now the contract also involves education of grandchildren. The game does not change as to the previously outlined game, since payoffs of an exiting player are affected by the same strategy choices as before. Grandparents exit only after their grandchildren are educated, however, the payoffs of grandparents are unaffected by the moves of the current parent.

The following figure depicts the tree for an educated child. We have simplified the tree in three ways: as before, we take the move of previous parents as given, and analyse the game always conditional on that move. Also, as before, the grandparents commit to a binding bequest contract which establishes automatically their best-response at the end of their life. Thus if the previous generation transfered and educated (!) their children, this triggers a bequest. Any other moves triggers no bequest (NB) and a grandparent simply transfers stored resources to the grandchildren, i.e. disinherits. Given these assumptions, we can write the payoffs as a function of five different strategic moves of players t and $t+1$, that is $W(s_{t-1}, s_t, s_t, s_{t+1}, s_{t+1}, s_t)$. A particular point in this payoff functions could be, for example, $W(E, T, E, T, E, B)$, meaning that this is a payoff given a parent $t-1$ educated (the first entry) and the current player transferred T (second entry) and educated (third entry) his own children; and given that his children transfered (fourth entry) and educated their children (fifth entry) he bequests (sixth entry). Note again that children make their decision conditional on s_{t-1}, thus this is still a two player game, thus previous results from two player cyclic OG games apply.

In order to further simplify the graphical representation, we eliminated strictly dominated strategies from the graph, that is children never transfer if they were not educated. Thus there is no transfer arc after the decision point of a child which has *not* been educated previously.

We also arrange the payoffs for an individual t in a payoff table as before. Define again $B_{t-1} = \beta C^*_{t-1}$, where $C^*_{t-1} = h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*)$. Hence parents choose again to commit to a certain fraction β of their current consumption C^*_t, before they educate their children. They have to transfer the resources as grandparents in this case - thus savings can increase any assets stored and also those received from previous generations.

The educated path of the bequest game

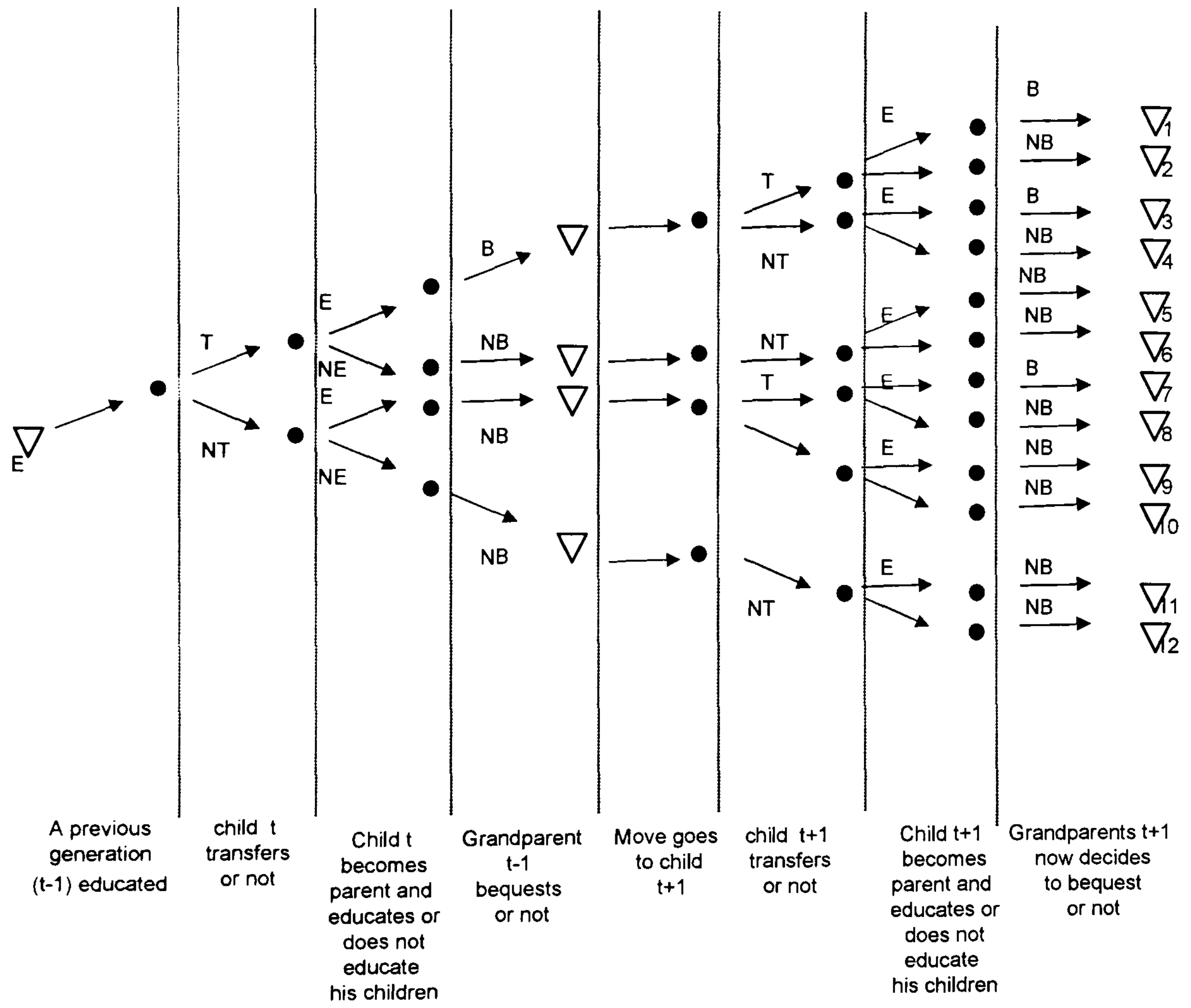


Figure 2.6: Simplified representation of all finite paths for an educated individual going from child entry to grandparent exit point in the cyclic game

Exit point	Payoff symbol	Payoff dependent on strategies	Payoffs period t	Own Parents strategy	Rank of payoff given $B_t^* = B_{t-1} = 0$
∇_1	W_1	$W(T, E, B, T, E, B)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*) + B_{t-1}^* - s_t^{**}) + W(Rs_t^{**} - c - B_t^*)$	E	2
∇_2	W_2	$W(T, E, B, T, NE, NB)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*) + B_{t-1}^* - s_t) + W(Rs_t - c)$	E	2
∇_3	W_3	$W(T, E, B, NT, E, NB)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH} + B_{t-1}^* - s_t) + W(Rs_t - c)$	E	6
∇_4	W_4	$W(T, E, B, NT, NE, NB)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l^* w^{CH} + B_{t-1}^* - s_t) + W(Rs_t - c)$	E	6
∇_5	W_5	$W(T, NE, NB, NT, E, NB)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l_{max} w^{CH} - s_t) + W(Rs_t - c)$	E	4
∇_6	W_6	$W(T, NE, NB, NT, NE, NB)$	$W(h_t(l^*)(1-\alpha^*)) + W(h_t(l^*) + l_{max} w^{CH} - s_t) + W(Rs_t - c)$	E	4
∇_7	W_7	$W(NT, E, NB, T, E, B)$	$W(h_t(l^*)) + W(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*) - s_t^*) + W(Rs_t^* - c - B_t^*)$	E	1
∇_8	W_8	$W(NT, E, NB, T, NE, NB)$	$W(h_t(l^*)) + W(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*) - s_t) + W(Rs_t - c)$	E	1
∇_9	W_9	$W(NT, E, NB, NT, E, NB)$	$W(h_t(l^*)) + W(h_t(l^*) + l^* w^{CH} - s_t) + W(Rs_t - c)$	E	3
∇_{10}	W_{10}	$W(NT, E, NB, NT, NE, NB)$	$W(h_t(l^*)) + W(h_t(l^*) + l^* w^{CH} - s_t) + W(Rs_t - c)$	E	3
∇_{11}	W_{11}	$W(NT, NE, NB, NT, E, NB)$	$W(h_t(l^*)) + W(h_t(l^*) + l_{max} w^{CH} - s_t) + W(Rs_t - c)$	E	5
∇_{12}	W_{12}	$W(NT, NE, NB, NT, NE, NB)$	$W(h_t(l^*)) + W(h_t(l^*) + l_{max} w^{CH} - s_t) + W(Rs_t - c)$	E	5

Table 2.6: Payoffs to player t in the representative 3 period 2-player cyclic game with bequests: educated generation.

For example the payoffs received at ∇_{11} are equal to those received at ∇_{12}. The whole table can thus be represented in a much simpler way still, but this would not yield further insights. However, note that the payoffs have ranks assigned in the rightmost column, imposing $B_{t-1}^{*} = 0$ and $B_{t}^{*} = 0$ and $l^{max} w^{CH} < l^{*} w^{CH} + \alpha^{*} g l^{*} w^{CH}$, thus doing as if there was no bequest move and as if there was a beneficial contract. Equal numbers imply equal rank.

The following Lemma can be deduced as in the two generation case. For this purpose, let us define the grandparent generation as a pure "mechanism".

Lemma 2 *In the absence of any bequest resources from grandparents, that is whenever the bequest of the last generation say B_{t-1} and the bequest of the current generation, B_t, are zero, the equilibrium for every generation is to play (NT, L).*

Proof *In the absence of any commitment mechanism and knowing that the equilibrium must be identical in strategies, (T, E) cannot be an equilibrium. This follows from the fact that a deviation would do better. Hence playing (NT, E) if everybody is playing (T, E) does better which follows from $\nabla_7 > \nabla_1$. But everybody playing (NT, E) cannot be an equilibrium either since (NT, L) does better from $\nabla_{11} > \nabla_5$. Also (NT, L) does better than ∇_1, thus (NT, L) must be the only equilibrium possible.*

The payoff point ∇_{11} is not stationary. If everybody plays the game repeatedly the family will be on the uneducated path and hence the outcome must be $\nabla_{24}(NT, NE, NT, NE)$, which is worse than ∇_1. However, as in the two generation case, no player can prevent this from happening, since contracts are not enforceable and a beneficial deviation exists.

This can be improved upon by introducing a bequest. We proceed as before. The maximum uneducated parents are willing to put aside to enforce cooperation is given by a bequest, $\hat{B}$ that equalises ∇_{19} to ∇_{24} and the most educated parents are willing to give is the difference between ∇_1 and ∇_{11}. Now we need to consider that savings s_t will be different, depending on which path is chosen.

We focus on some key comparisons in the payoff table. Without the bequest, we have $\nabla_7 > \nabla_1 > \nabla_{11}$, taking into account optimal savings.

Now, as in the two generation model, a child knows that he should only be transferring if the parent can threaten to withhold at least an amount βC_t^*, with which the individual could then threaten the next generation effectively with. Hence βC_t^* must again equalise the payoffs of the best possible deviation ∇_7 and the payoff of full cooperation ∇_1. Writing these out, we have

$$\mathcal{W}_1 \equiv W_0\left(h_t(l^*)(1-\alpha^*)\right) + W_1\left(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*) + \frac{\beta}{g} C_t^* - s^{**}\right)$$
$$+ W_2\left(Rs^{**} - c - \beta C_t^*\right)$$

$$\mathcal{W}_7 \equiv W_0\left(h_t(l^*)\right) + W_1\left(h_t(l^*) + l^* w^{CH} + g\alpha^* h_t(l^*) - s^*\right)$$
$$+ W_2\left(Rs^* - c - \beta C_t^*\right)$$

Here s^{**} denote optimal savings under the cooperative equilibrium and s^* denote savings under defection as a child, and trying to re-establish cooperation later. Note that $s^* < s^{**}$, thus the final grandparent marginal utility in ∇_7 is higher than in ∇_1

This leads to a solution in a path of β which equalises payoffs between ∇_7 and ∇_1 in every period t. We can calculate the effect of β now by examining the partial derivatives of the two payoff functions with respect to β. Then we can postulate the following, main result of the paper:

Theorem 1 *If β can be set in such way that $\mathcal{W}_1 > \mathcal{W}_7 > \mathcal{W}_{11}$ or $\mathcal{W}_1 > \mathcal{W}_{11} > \mathcal{W}_7$, then there exists a beneficial contractual solution and a contract $\mathcal{C} = (\alpha^*, l^*, \beta^*)$. A necessary condition is $R > g$ if a non-mutually beneficial contract exists or $R = g$ if a mutually beneficial contract exists.*

Proof *$\mathcal{W}_7$ is initially greater than $\mathcal{W}_1$. The payoff $\mathcal{W}_{11}$ is greater than $\mathcal{W}_1$ if the contract is not mutually beneficial (optimal), $\mathcal{W}_{11}$ is smaller than $\mathcal{W}_1$ if the contract is mutually beneficial (optimal). If $\mathcal{W}_1$ can be shown to be increasing in β and $\mathcal{W}_7$ can be shown to be decreasing in β, then at least one solution must exist where $\mathcal{W}_1 = \mathcal{W}_7$, if a mutually beneficial contract exists or $\mathcal{W}_1 = \mathcal{W}_{11}$ if a non-mutually beneficial contract exists. Note first that $\mathcal{W}_{11}$ is unaffected by β, and thus a constant.*

$\mathcal{W}_7$ is monotically decreasing in β, the bequest fraction made by the current generation:

$$\frac{\partial \mathcal{W}_7}{\partial \beta} = -C_t^* W_2^{(7)'}(Rs' - c - \beta C_t^*) < 0$$

$\mathcal{W}_1$ is monotonically decreasing in the bequest to the current generation, but monotonically increasing in the bequest received by the last generation. We obtain

$$\frac{\partial \mathcal{W}_1}{\partial \beta} = \left(\frac{1}{g} - s'(\beta)\right) W_1' \left(C_t^* \left(1 + \frac{\beta}{g}\right) - s^{**}\right) + (-C_t^* + Rs'(\beta)) W_2^{(2)'} (Rs^{**} - c - \beta C_t^*)$$

Using the FOC in optimal savings, $RW_2'(.) = W_1'(.)$, we have

$$= W_1'(.) \left(\frac{1}{g} C_t^* - s'(\beta) - \frac{C_t^*}{R} + \frac{Rs'(\beta)}{R}\right).$$

Note that $s(\beta)$ is the propensity to save out of a unit of bequest received. It is positive and increasing in β, hence $s'(\beta) > 0$.

Thus for the effect of bequests to increase $\mathcal{W}_1$, me must have

$$C_t^* \left(\frac{1}{g} - \frac{1}{R}\right) > 0,$$

and which implies an positive interest rate and gross rate $R > g$.

Then, a solution consists in the intersection of $\mathcal{W}_1$ with $\mathcal{W}_7$ if a mutually beneficial contract exists and in the intersection of $\mathcal{W}_1$ with $\mathcal{W}_{11}$ if a non-mutually beneficial contract exists. Note that we require $R > g$ if a non-mutually beneficial contract exists, but we require only $R = g$ if a mutually beneficial contract exist, since then $\mathcal{W}_1 > \mathcal{W}_{11}$ and if $\mathcal{W}_1$ stays constant in β and $\mathcal{W}_7$ decreases in β, a contract exists.

The solution obtained will be denoted as the fraction of bequests out of period t consumption, C_t^*, β^* $\forall t \in \mathbb{N}$. Since all consumption grows with rate g in every period, this does not change the equilibrium fraction of bequests.

The contract implements thus a better payoff than any deviation outcome if it can be set at such a β. This follows immediately from the ranking of payoffs and from the arguments of the two-generational model.

Thus, with three generations and the possibility to save, we obtain a solution with bequest sanctions and first-best contracts. This is possible even when ignoring the history of play - which would not have been possible in a classical OLG-game with trigger strategies. The following remarks summarize important other features.

Remark 2 *Note that now as well, contracts which are not first-best are implementable, since we have that the cooperation payoffs are an increasing function of bequests β and thus they can reach the non-cooperative solution, even though the cooperative payoffs are initially smaller.*

Remark 3 *Note that now there are not the same deviation possibilities as in the two-generational setup. The payoffs W_i^R do not exist, since grandparents can condition the transmission of a bequest on grandchildrens' education enrolment - and thus avoid these possible deviations.*

Remark 4 *The payoffs for the uneducated generation can be read off directly from table 2.6, replacing children's and parents' human capital by $h(l^{max})$ and adding 12 to the payoff number.*

Another problem needs now to be addressed after knowing that an equilibrium exists in the educated generation: can this point be reached by uneducated generations? In other words, we need to investigate if the maximum uneducated parents are willing to give up is large enough to enforce the contract with the next generation if $\hat{\beta}_t \geq \beta_t^*$.

Lemma 3 *(i) As long as $R\left(l^* w^{CH} - l_{max} w^{CH} + g\alpha^* h(l^*)\right) > \beta(h(l^*) + g\alpha^* h(l^*) + l^* w^{CH})$, an uneducated generation can reach the educated generations equilibrium and would like to implement the equilibrium. (ii) The bequest that the first uneducated generation is willing to make must be greater than minimum consumption, $U_2(Rs'' - \hat{B}) \geq U(C^{min}.)$*

Proof *Consider the two possible payoffs in period $t-1$, one denoted by $\mathcal{W}_{19}$ and the other one denoted by $\mathcal{W}_{24}$.*

$$\mathcal{W}_{19} = W_0(h(l_{max})) + W_1(h(l_{max}) + g\alpha^* h(l^*) + l^* w^{CH} - s') + W_2(Rs' - c - B_{t-1})$$

$$\mathcal{W}_{24} = W_0(h(l_{max})) + W_1(h(l_{max}) + l_{max} w^{CH} - s'') + W_2(Rs'' - c)$$

The bequest $\hat{B_{t-1}}$, equalising these payoffs is the maximum bequest uneducated parents would like to commit to. As uneducated parents know the equilibrium contract *exists once children are educated, we then need to determine if bequests they are willing to give are sufficient.*

Assuming optimal savings, we need to compare only the last two utility terms of each payoff, $W_2(.)$. They should be equal under optimal savings if the payoffs are equal. Thus, applying again the result from Proposition 1, writing $\kappa = W'^{-1}(.)$, we derive savings to be:

$$s' = \frac{\left(h(l_{max}) + g\alpha^* h(l^*) + l^* w^{CH} + \kappa C + \kappa \hat{B}\right)}{1 + \kappa R},$$

$$s'' = \frac{\left(h(l_{max}) + l_{max} w^{CH} + \kappa C\right)}{1 + \kappa R}.$$

Substituting these into the argument of the two last terms of the payoff functions, W_2, and equalising the arguments, we have:

$$R\frac{\left(h(l_{max}) + g\alpha^* h(l^*) + l^* w^{CH} + \kappa C + \kappa \hat{B}\right)}{1 + \kappa R} - c - \hat{B} = R\frac{\left(h(l_{max}) + l_{max} w^{CH} + \kappa C\right)}{1 + \kappa R} - c.$$

After some manipulation we obtain

$$\hat{B_{t-1}} = R(g\alpha^* h(l^*) + (l^* - l_{max})w^{CH}).$$

The required bequest in the next generation is

$$\frac{\beta}{g}\left(gh(l^*) + g^2\alpha^* h(l^*) + gl^* w^{CH}\right),$$

which gives the result. □

We provide a small calibration example to illustrate the equilibrium.

Calibration example Since it is difficult to evaluate the magnitude of transfers, bequests and child labour in a purely theoretical model, a small numerical example might be helpful. Assume again a single child family. We calculate only the uneven numbered payoffs $\nabla_1 - \nabla_{11}$ for an educated generation, since all even payoffs must be strictly less by the fact that in the case of even payoffs, the last child generation did not transfer.

We know that $h(l^*) > h(l_{max}) > w^{CH}$. We set $h(l^*) \approx 2 \times h(l_{max})$, $h(l_{max}) = 1$, $w^{CH} = 0.4$, $R = 3$ which is a reasonable value for a lifespan of around 30 years and

$c = 1$, $k = 2$. The human capital accumulation function is $h(l) = \kappa - l^2$ for $l > 0$ and $h(l^{max}) = h_0$, where h_0 represents the human capital endowment of the child. Furthermore we assumed the function $W(.)$ to be of the logarithmic type.

The following table shows the payoffs with and without bequest threat [12]

	without contract β^*	with contract β^*
∇_1	2.11423	2.21920
∇_3	2.11423	2.21094
∇_5	1.84388	1.84672
∇_7	2.42602	2.21919
∇_9	1.51502	1.51502
∇_{11}	2.15567	2.15567
β^*	0	0.25599
$B_t = \beta C_t^* = \beta\left(h(l^*) + g\alpha^* h(l^*) + l^* w^{CH}\right)$	.	0.787
$g\alpha^* h(l^*)$	0	1.001
$\hat{B}_{t-1} = R(g\alpha^* h(l^*) + (l^* - l_{max})w^{CH})$		2.102
l	$l_{max} = 1$	$l^* = \frac{gw^{CH}}{2(g+1)} = 0.13$

Table 2.7: Numerical example of payoffs with and without the bequest threat.

The table shows that accepting the bequest contract changes the payoffs in such way that now the cooperative equilibrium does (just) better ($\nabla_1 > \nabla_7$). Also child labour is more than seven times lower under the contract. Note that without the contract, the no-education, no transfer equilibrium is preferred, as ∇_7 is the largest payoff, which leads to chose no transfer as a child. This being anticipated by parents, the no-education, no-transfer equilibrium is the outcome without any contract.

The equilibrium bequest $B_t^* = \beta C_t^*$ is well below the maximum bequest that any previous grandparent would like to commit to, $\hat{B}_{t-1}$. Notice as well that the transfer received is higher than the bequest one needs to commit to before the receipt of the transfer. This is explained by the increase in resources through savings.

In addition to this numerical example, we illustrate the sensitivity of the contract to the wage growth rate g, which we let vary between 1 and 3, and to the child labour wage, where we have first a high child labour wage $\frac{4}{5}h(l_{max})$ and a low child labour wage $\frac{2}{5}h(l_{max})$. The three-dimensional Figure 2.7 depicts the optimal contracts as planes in $\mathbb{R}^3$, where utility is depicted on the vertical axis, and g and β are depicted on the

[12] The Maple code to evaluate these payoffs can be requested from the author

horizontal axes.

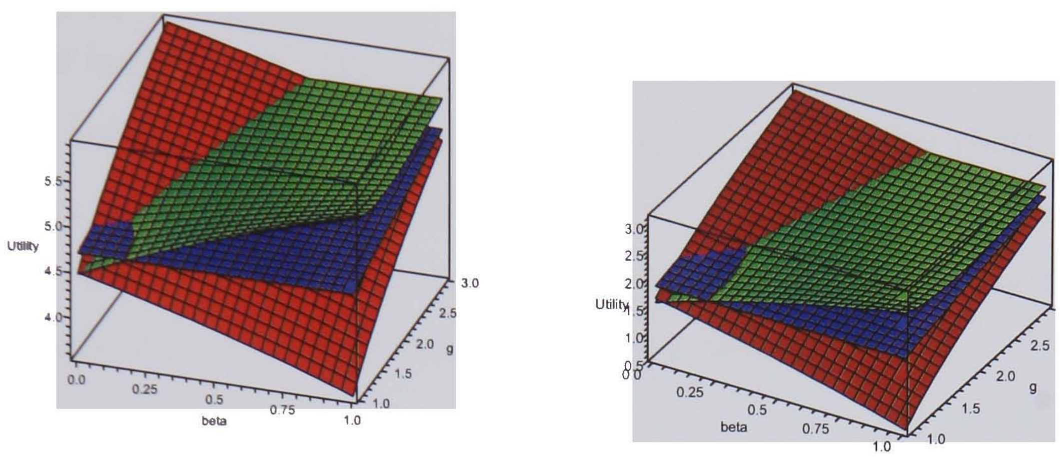


Figure 2.7: The values of the payoffs $\mathcal{W}_1$ (green plane), $\mathcal{W}_7$ (red plane) and $\mathcal{W}_{11}$ (blue plane) for values of g and β, given a high and a low child labour wage.

The first-best contract with bequests can be read off the figures by noting the set or junction of saddle points formed by the blue and the green plane in the left-hand side of the figure and thereafter by the junction of the green and the the red plane. This saddle point will make individuals just better off under cooperation and allows them to reach the green, cooperation plane. Furthermore, we can see how the figure to the right is flatter both in aspect and scale. In the left part of the figure, the returns structure has been altered in favor of higher child wages: one can distinguish that the planes have a steeper inclination. However, the space representative of deviation options (blue surface, surrounded by red and green planes) seems to have increased in the right diagram due to the planes being flatter. The equilibrium contract is not very much affected by this change.

Fertility Suppose we extend the previous model and allow parents to choose the number of children. Clearly, parents can now potentially receive more transfers from children than before. Thus, they could enforce potentially better bequest contracts. However, parents receive only a single bequest from a previous parent generation. As they need to commit to bequest to, say, n^E educated children, their utility is potentially decreasing in the bequests they need to commit to. We can thus infer directly from the conditions on the bequests how many children parents can educate. This is a great deal easier than to re-derive equilibrium conditions with children and these bequest constraint also impose a constraint on the number of educated children.

Suppose parents can chose the number of children $n \in [0, n_{max}]$ where n_{max} is the natural fertility limit. How does this change the allocation of education and the division of bequests between children? We assume parents either chose to educate a child or not to educate it. This yields $n^* = n^E + n^U$, where n^E is the number of educated children and n^U is the number of uneducated children. Only those children that were educated, n^E, and transferred, receive a bequest.

Proposition 5 *The number of educated children is bounded from above by $n^E < \frac{R}{g}$, thus allowing an unequal division of bequests and unequal inter-vivos transfers (some children work, some are educated[13]). We observe high fertility, $n^* = n_{max}$.*

Proof *Using as above the FOC for optimal savings, we obtain:*

$$\frac{\partial W_1}{\partial \beta} = C_t^* \left(\frac{1}{g} - \frac{n^E}{R}) W_1'(C_t^*(1 + \frac{\beta}{g}) \right).$$

Thus for there to be a positive effect

$$\frac{1}{g} > \frac{n^E}{R},$$

Then the term is positive for

$$n^E < \frac{R}{g}.$$

The family will always choose a maximum of uneducated children since utility is increasing in n^U, but only educating n^E, thus $n^U = n_{max} - n^E$.

□

Lemma 4 *If there are direct educational costs and no child labour opportunities, we will observe equal division of bequests, inter-vivos transfers and lower fertility if $n_{max} > \frac{R}{g}$.*

Proof *In case of no child labour activities, parents chose $n^U = 0$ since their utility is decreasing in n^U, but educate $n^E < \frac{R}{g}$ children and divide bequests equally among them.* □

[13] *Note that R is a gross lifetime interest rate and can thus take high values, e.g. > 2*

The additional consideration of fertility gives the interesting prediction that we should see primo-geniture or other unequal bequest rules emerge if one cannot invest in the quality of all children. Additionally, if the number of children is small and they can all be educated we should observe equal division of the estate.

2.6 Discussion and policy implications

The cyclic OG game points to the importance of information on past play and observability of the history of play. The dynamics of the game highlight also the importance of sustained growth in wages and the possibility to save over generations, a point not obvious from the static setting. Grandparents can implement first-best contracts: they can save over a longer horizon, they have credible alternative beneficiairies (grandchildren) and grandparents can condition contracts on the parents' behaviour. In summary grandparents facilitate contracts, which lead to a path of education and transfers. Once families attain the educated path (or are able to commit/ write binding contracts), they are likely to remain there. This would imply, that public policies have considerable intergenerational benefits that can be felt long after they were implemented.

The traditional analysis of commitment had shown that the optimal policy to address child labour problems needs to be linked to the inefficiency arising from incomplete contracts. A family alone might struggle to implement such first best-contracts, if the family has no bequest resources or if no grandparents are alive. This would give a greater role to policy than previously thought.

It seems therefore advisable not only to redistribute resources to the poor. Policies that redistribute income alone could certainly have a positive, short term effect on child labour, although not in our model. In our model, income transfers would need to be sustained for long periods in order to generate the expected increase in income necessary for families to generate sufficient bequest resources and the income transfers need to be in the form of a conditional transfer to parents. Note that conditional cash transfer programs would need to be implemented in every generation in order to yield a lasting decrease in child labour.

A ban on child labour could be beneficial in our model if the reduction in fertility (implied by Lemma 4) does not reduce parents' overall utility. Utility could however be reduced, since less children are working for wages and thus overall utility falls, if the family has many children. Given the relatively small amount of children that can be educated in the model, it would be plausible that only a welfare worsening ban exists in this situation.

Some alternative policies to cash transfers and bans are a credit system for students. Public policy measures could provide credit to students. These could be repaid by the grown-up children as soon as they find an employment (say by taxing the grown-ups). However, such state transfers *do not* give resources to parents. In our model, as well as in Baland and Robinson's (2000), it is parents that forego resources in the first place to educate their children - and it is also parents that decide about child schooling in the age where returns from schooling are usually highest. Hence, as in previous work by Baland and Robinson (2000), an effective credit system on its own will not be sufficient to guarantee that parents will not send their children into child labour, as opposed to the usual perfect market hypothesis.

Now one could think of other transfer arrangements that could enforce intergenerational contracts similar to the ones considered here.

A proposition along these lines is contained in the work of Rangel (2003): a social security system could provide incentives for parents to invest in the education or quality of offspring. Suppose the parents receive social security contributions once they enter grandparent age. The social security contributions are financed by contributions from the current parent generation. This could be done by taxing a fixed proportion of the current parents' generation income. A problem here is that these payments are most likely subject to free riding, since education investment is costly. Every parent would expect the other families to invest in education and, on average, parents will underinvest in their children. Thus, education as a public good is likely to be underprovided and the solution to the child labour commitment problem is incomplete.

Which public policies are then suitable? A combined solution seems necessary. A first step, according to the model presented here, suggests that any assistance be divided into two parts: i) providing resources to parents exclusively encouraging them to accumulate fixed assets and/or life insurance and ii) provide credit to parents to temporarily counterbalance losses from child labour income. This will enable families to implement the discussed sanction system.

Another, more plausible solution, would be a one-to-one third-party contract, where educated individuals are taxed in order to transfer the resources directly to parents. The transfers to parents have to fulfill conditions outlined in the text. In addition, parents receive a short term credit to counterbalance losses from foregone child labour.

So far we have not addressed the associated general equilibrium problems and the macro-economic variables R and g which are important in our model. In terms of general equilibrium considerations, having more children in school should raise the child wage, which in turn improves parent's utility from child work and might well undermine a beneficial contract. Schultz (2004) found such wage effects in Mexico. As far as the macro-economic variables are concerned, it is clear that expectations about both g and R are crucial for contracts to exist and hence stability and a functioning financial sector seem pre-requisities for any public policy to eradicate child labour.

Another problem of the presented model and public policy is that the model implicitly assumes that grandparents do have beneficial preferences (i.e. they care only about the return to education and do not derive utility from child labour or education directly). This might not be entirely correct, since grandparents might force other generations through the bequest tool to align themselves on their own preferences, beneficial or harmful for society as a whole. Assume for example that grandparents have a preference for child labour because in their times child labour was much more common and accepted. Then such bequest contracts could have the opposite effect than intended. Grandparents could worsen child outcomes in terms of education, Edlund and Rahman (2005).

Certainly more empirical investigation is needed to assess the role of grandparents in the commitment problem and the education decision. The present model points in any case to an important shortcoming of family transfer systems when information on past play is imperfect and the problem of families to maintain first-best contracts within such environments. Current policies (credit markets, poverty reduction, health) to address child labour do not resolve the commitment problem and the associated losses for society as a whole, since they do not simultaneously address the problems of poverty and commitment.

2.7 Conclusion

The child labour literature has so far analysed the commitment problem from a static point of view. A dynamic, game-theoretic analysis of the commitment problem, addressing imperfect information transmission between family members as proposed in this paper is thus a significant step forward. This paper has made an effort to model the commitment problem in a novel way using a cyclic overlapping generations game. It derived implications for a contractual equilibrium between selfish families.

Commitment mechanisms are difficult to sustain inside the family once we focus on strategies that are independent of the history of play. Cyclic games do not assume a previous history, are dynamic and thus ideally suited for strategic punishment strategies. Additional punishment mechanisms are indeed necessary to reach a cooperative outcome with commitment within this game. Generations are likely to remain in a cycle of poverty even though credit markets are perfect and returns to education are high if they cannot implement such sanctions.

We show that commitment problems disappear if parents can commit to a bequest, can save over a long enough time horizon and can observe their children's behaviour towards grandchildren, i.e. if parents live long enough as grandparents. Replacing the missing history of play by a monetary bequest sanction solves the commitment problem. Grandparents have in addition other advantages to be the "market-maker" in such contracts: grandparents can costlessly transfer to grandchildren and can write binding and conditional bequest contracts.

The policy implications of the model cast doubt on current approaches to eradicate child labour that focus only on direct poverty reduction through income transfers or pure educational credits. We propose a third party system of intra-family transfers to parents conditional on (educated) children's income combined with a micro-credit to effectively reduce child labour. A topic of further empirical research is the role of grandparents in relation to the commitment problem and education investment.

Chapter 3

Education transmission across three generations

3.1 Introduction

What difference can grandparents make in the educational success of their grandchildren? Previous studies have focused mainly on the effect of parents' education on children's education using two-generational transmission models. The present chapter intends to extend this literature and point out new and interesting directions for future research. The extensive two-generational analyses are often referred to as "intergenerational mobility" studies. The concept of mobility refers to the ease by which members of a given society move up and down the ladder of life outcomes, such as income or education for example, given their parents' position in the outcome distribution. Mobility studies are interesting because they mirror the opportunities a given society offers its members to improve or change their position in the outcome distribution. Being thus influenced by grandparents in addition to parents opens up a new channel of thinking about mobility. Policy implications, however, are not the main interest of this chapter, but rather it aims to empirically assess a potentially three-generational education transmission mechanism.

The research question to be addressed is how much a child's choice of education is determined by her parents' and grandparents' choices. Analysing the child's choice, a researcher cannot observe every variable affecting the child's decision. Ability or parenting skills are typical examples of such unobserved components of a child's decision. If ability were genetically transmitted, for example, then more able parents would acquire better education, but would also have more able children - and these children would acquire better education through this channel. As a researcher does not observe ability directly, it is difficult to disentangle parental education from parental ability effects.

We address this endogeneity concern by using instrumental variable techniques. Schooling reforms implemented in the UK in 1945 and 1973 and mothers' birth-order allow us to have credible instrumental variables to address the endogeneity of parents' and grandparents' education. In short, the present study is the first to have both three generations and an identification strategy addressing unobservables at parents' and grandparents' level.

The question of whether grandparents enter the transmission mechanism is interesting, and not exclusively so because the wider literature has neglected it so far. Most citizens of developed countries benefit from longer lifespans and an increased time with their grandchildren. Transfers of resources from the grandparents' generation, such as gifts, bequests and childcare, can make a difference, even more so to the marginally poor family. One could think of credit constraints being relieved by grandparents. Grandparents' influence through various channels on grandchildren seem an exciting and new research topic which this study addresses at the forefront of the current mobility literature.

The objective of this paper is thus to provide some descriptive evidence of grandparental influence and its channels, and to investigate whether the effects of grandparents' and parents' education on children's education are causal: is it really parents' and grandparents' education that matters for children's education choices or are there other important but perhaps unobserved factors of influence?

A statement of caution is appropriate here. This contribution does not claim to isolate causal effects. Rather, it illustrates with a wide range of econometric techniques the point that censored regression effects can be seen as upper bounds and that unobservable factors play a major role in the transmission of education. Whether such a finding may be extended to other outcome measures is open to debate. The paper illustrates as well that parametric and non-parametric techniques yield different yet complementary results.

The paper is structured as follows: The previous literature is reviewed, thereafter, relevant features of the current data are discussed and a preliminary data analysis is carried out which does not take into account the endogeneity of education. The subsequent section presents the identification strategy, taking unobserved variable bias into account, and the remaining sections feature selected results as well as sensitivity checks.

3.2 The literature on intergenerational transmission

The two papers of Becker and Tomes (1979), Becker and Tomes (1986) and the work of Solon (1992) have spurred widespread interest in assessing the intergenerational correlation in earnings. The wider literature has been particularly concerned with upward mobility: that is, if and how easily children improve upon their fathers' status, Dearden et al. (1997). Additionally, there has been an interest in studying such mobility in greater detail and parametrising the joint distribution of outcomes, such as for example education, Bartolucci and Scaccia (2004). Models used by Chevalier et al. (2005), Chevalier (2004), Oreopoulos et al. (2006) or Carneiro (2005) can be seen as reduced form models of education transmission, where the intergenerational correlation in income or wealth is proxied by a correlation in education. In an extension, these models are augmented in this paper by a grandparent generation.

There exist some contributions taking into account three generations, for example Plug and Vijverberg (2005), Maurin (2002) and Sacerdote (2005). Sacerdote (2005) examines a three-generation transmission mechanism by assessing the impact of slavery on the grandsons of slaves. He provides evidence that intergenerational effects disappear after only three generations. In general, grandparents are not given an active, explicit role in these studies. The present contribution tries to go beyond their findings by presuming an active role for grandparents.

The empirical analysis faces an important statistical problem. Estimating causal effects of parental background variables such as education and income on childrens income is complicated by the largely unknown transmission mechanisms, by measurement error and by unobserved important variables. The econometrician might not observe many of the essential inputs into the production function of child outcomes, Todd and Wolpin (2003). In addition, theory does not give us any clear hint towards the expected transmission mechanisms, Blow et al. (2005). Is it linear or non-linear, and which magnitude is explained? In addition, many studies are affected by measurement error in the outcome variables. In studying income transmission, typically most of the father's income trajectory is in the data, but we observe only the first few years of the son's income profile. This leads to estimates of permanent income which are difficult to compare. Education suffers from a similar problem, since individuals might take additional training programmes during their lifecourse. We only observe the first years of children's education trajectories.

Previous research has relied on distinct approaches to identify the causal effect of schooling and income on the outcomes of the next generation and to address measurement problems. Chevalier (2004) identifies three main approaches.

The first approach is to compare the children of identical twins, Behrman and Rosenzweig (2002). Certainly, the 'nature' background of at least one of the parents is the same and hence any differences might only arise from 'nurture', that is, different family background, care and other inputs. Other approaches have looked at twin-differences: that is, the difference in outcomes between identical twins in relation to parental background. This differences out any unobserved variables in family background and education. Such approaches are only partially valid, since differences in preferences, even between twins, can be accentuated. Family background might not be all that matters for such decisions, but rather (unobservable) individual factors that affect each twin separately. Behrman and Rosenzweig's (2002) coefficients using twins were found to be 0.04-0.0149, and thus much smaller than the usual OLS estimates on mobility, which range between 0.2 and 0.4. Behrman and Rosenzweig (2002) even found negative effects for mothers' schooling which they associate with decreasing home time of mothers for children with increasing education. The results indicate a positive bias in the cross-sectional OLS coefficients.

The second approach is to examine adoptees, Plug and Vijverberg (2005), Plug (2004). Plug and Vijverberg (2005) use extensive data on adoptees to estimate an unconfounded effect of income on schooling outcomes. The authors state that estimates of the impact of income on the education of the next generation are often biased or less credible since one cannot control for correlated parental ability. They argue that at least part of this ability is transmitted genetically. By using adoptees, any such bias by inherited factors is excluded, and Plug and Vijverberg (2005) find significant effects of income on child schooling. However, one cannot assume that they have found evidence of a causal effect, since wealthier parents could also have better child-rearing skills, assignment into "treatment" (adoption) might be selective, or adoptees might be different in other ways from non-birth children. Plug and Vijverberg (2005) provide, however, a large set of evidence that the transmission mechanism of income on education might be causal. In their specifications though a third generation measure, such as years of education of grandmothers/grandfathers never seem statistically significant.

The last approach to evaluating intergenerational effects which one suspects to be subject to unobservable variable bias are instrumental variables. Instrumental variables use an exogeneous variation in explanatory variables, which is independent of unobserved factors, to attempt to identify causal effects. Instruments can be factors that change the decisions of agents, unrelated to their own, unobserved factors. Popular instruments in education or income transmission are factors that raise schooling costs and policy reforms.

The literature has used instrumental variables to estimate the effect of parents' education alone on children's education, using suitable natural experiments, Black et al. (2003), Chevalier (2004), Oreopoulos et al. (2006). The respective authors came to rather different conclusions. Black et al. (2003) found severely reduced effects or even insignificant effects when parental education was assumed to be endogeneous in data from Norway, as Chevalier (2004) states. Chevalier (2004) attributes this to the rather homogeneous society that was analysed in the Norwegian data. On the other hand Chevalier (2004) estimated coefficients on education that nearly double when he instruments education. He identifies this as a local average treatment effect: that is, the reforms that he uses affected only those people who wished to leave school at 15 years. Thus the reform would particularly affect those individuals with a low taste for schooling and thus over-estimate the true effect. Oreopoulos et al. (2006) identifies as well a larger impact of schooling on grade retention when using an IV strategy than under the standard OLS approach. He argues that the larger sample and the greater variation in schooling laws in the US might explain an IV effect above the OLS effect.

A positive OLS bias seems well-established in the adoption and twins literature whereas IV strategies have lead to different estimates. Smaller effects of education in an instrumental variable estimation are consistent with the assumption that unobservable ability, for example, is positively correlated to education. The results obtained in this paper have to be judged based on such previous evidence.

This paper will focus on instrumental variable estimation using policy reforms in a three-generational context. The next section will outline the data and conduct a preliminary analysis before presenting the indentification strategy used.

3.3 Data and preliminary analysis

The NCDS is an ongoing longitudinal survey of a British birth cohort born between March 3^{rd} and March 9^{th} 1958. The initial sample consisted of 17,000 individuals who were resurveyed on six further occasions in order to monitor their changing health, education, social and economic circumstances[1]: - in 1965 (age 7), 1969 (age 11), 1974 (age 16), 1981 (age 23), 1991 (age 33), 1999/2000 (age 42) and 2004 (age 48) [2]. During the 1991 survey, a special sub-survey was undertaken concerning the children of the NCDS. A random draw of one third of the cohort members was administered a detailed questionaire about their children's development, health and education. Tests were administered to different age groups to assess mathematical abilities and literacy.

The NCDS was conceived first as a one-shot birth survey and has since been extended to a longitudinal survey. We are going to base our analysis on a subsample of the NCDS data. We will examine the education of the children of the NCDS. Thus, our sample will only include families that have children who are at least sixteen years or older. We will mainly examine a sample of only the eldest children of these families. In some contexts, we will analyse a sample of all children which are already over sixteen.

Our data is not as representative as the full NCDS sample, as we exclude parents who have children after the age of 32. Thus not all women in the orignal sample have completed their fertility cycle. This is clearly a drawback and affects the initial randomisation of the NCDS, consisting in selecting a birth cohort born in a particular week. Of course, in general, it is also open to discussion if initial randomisation in generation one will still yield random (representative) samples of the population three generations ahead, even if we would wait until the fertility cycle is completed. Despite these more general issues, one has to acknowledge that because not all children have grown up beyond schooling age, the sample retained here is biased. We select a sample of families with rather low levels of educational achievement, since families with lower education tend to have children earlier, and these are some of the children we observe in our sample. This has to be borne in mind during the whole analysis - and the comparison

[1]This and the following descriptions were adapted and shortened from a longer version available at http://www.cls.ioe.ac.uk/studies.asp?section=000100020003

[2]The latest wave of 2004 is now out but is not yet incorporated into this analysis

of tables 3.1 and 3.2 reveal these differences in background variables of the grandparents and also in mean attainment of children of NCDS cohort members.

Furthermore, and as mentioned before, there is censoring in the observed child education variable which affects about 40% of the observations. Thus 40% of the individuals of our sample are still in school at the time of the survey.

Another potential pitfall of the NCDS is that the researcher always has a wealth of information on cohort members, but more limited information on partners. For example, we have the whole schooling history, from schools attended to the number of exams taken and test results for the cohort member, but we have only the years of education and some general measures for the partner. The same phenomenon appears also for grandparents. We have only observations on the cohort member's grandparents, and thus two important figures in our transmission story are missing.

But despite these shortcomings, the NCDS is to our knowledge, together with the CNSLY the only study which provides such a wealth of information. The limitations mentioned can be partly overcome by the use of advanced econometric techniques.

Variable	N	Mean	Std. Dev.	Min	Max
Grandfather's education	6487	4.08154	2.07616	1	12
Grandmother's education	6487	3.99337	1.60905	1	12
Mother's education	8889	4.30745	2.61720	1	32
Father's education	8849	4.22409	2.80246	1	31
Proportion of sample divorced	11372	.11501	.31905	0	1
Log of 2000 CM income	7945	6.82521	.82935	.08	10.63
Household size	11370	3.46763	1.33794	1	10
CM's father died	11056	.41045	.49193	0	1
CM's mother died	10969	.22098	.41493	0	1
Partner's mother died	9135	.22900	.42021	0	1
Partner's father died	9039	.40150	.49275	0	1

Table 3.1: Full sample characteristics NCDS wave 6

Comparing samples reveals that mean education is slightly lower for those parents in the selected sub-sample, Table 3.2, than for the full sample, Table 3.1. This is, interestingly, also observed in grandparents' education, which is lower for the subsample. In other aspects, the subsample and the full sample are surprisingly similar, especially

Variable	N	Mean	Std. Dev.	Min	Max
Child's education (censored and uncensored, first born)	3918	5.07810	1.65710	1	17
Sex of child	6107	.48911	.49992	0	1
Grandfather's education	4391	3.70892	1.53229	1	11
Grandmother's education	4467	3.71566	1.17598	1	11
Mother's education	5769	3.85994	1.90211	1	19
Father's education	5467	3.81690	2.20059	1	18
Proportion of sample divorced	6103	.10732	.30955	0	1
Log of 2000 CM income	4292	6.69396	.82998	1.46	10.59
Household size	6103	3.93331	1.18702	1	10
CM's father died	6019	.41020	.49191	0	1
CM's mother died	6024	.22626	.41844	0	1
Partner's mother died	5264	.24620	.42083	0	1
Partner's father died	5242	.44029	.49646	0	1

Table 3.2: The children of children sample

	Grandfather	Grandmother	Mother	Father	Son	Daughter
Grandfather	1 (.) 1834					
Grandmother	.377 (.00) 1823	1 (.) 1876				
Mother	.134 (.00) 1735	.127 (.00) 1775	1 (.) 2430			
Father	.110 (.00) 1592	.148 (.00) 1629	.191 (.00) 2095	1 (.) 2225		
Son	.037 (.34) 631	.106 (.00) 638	.123 (.00) 834	.122 (.00) 776	1 (.) 871	
Daugther	.007 (.81) 903	−.007 (.82) 925	.085 (.00) 1184	.146 (.00) 1075	. 0	1 (.) 1248

Table 3.3: Raw correlations of years of education, p-values in parenthesis

when it comes to income, divorce or the probability of one of the grandparents having died already. The subsample is only composed of the oldest children. There are many more variables that we could put up for comparison, but we focused on important variables that we will use in the models.

Table 3.3 of raw correlations in years of education shows an interesting difference in the correlation between couples of parents (grandfather-grandmother /father-mother). This suggests that the process of finding a "matching" partner has perhaps changed over time. However, it seems that the raw correlations in education between the grandparents/parents and parents/children have not changed drastically. This would underline an argument of stable education transmission across generations. Thus older generations were not more closely tied to their ancestors.

Interestingly, we observe an effect of grandparents on the grandchildren, although only for grandmothers on grandsons. This suggests a gender effect in the transmission mechanism. The effect of a grandmother on a grandson is relatively large and quite comparable to the effect of the father or the mother. This slightly contradicts the fact that generations have become more mobile. It seems that daughters receive comparatively more attention from fathers and are less tied to grandparents' education.

Finally, in Figure 3.1 we plot the education distributions for all three generations in the data. It is immediately obvious that a large proportion of individuals leave school at the relevant minimum age (14/15 for some grandparents, 15/16 for parents and 16 for children). One can also distinguish the impact of the reform for grandparents and partners, a large proportion of them leaving either at the old or the new minimum school leaving age. Conversely, most cohort members leave education at the minimum school leaving age 16, compulsory already for the 1958 cohort. We can also observe that partners' and cohort members' distributions (and to a small extent grandmothers' and grandfathers') exhibit a second hump at around 20 years. The absence of this slight bimodality in the education distribution of children underlines the fact of censoring: we do not observe highly educated children. Therefore we also estimate probit models of the staying-on decision, as explained in the text.

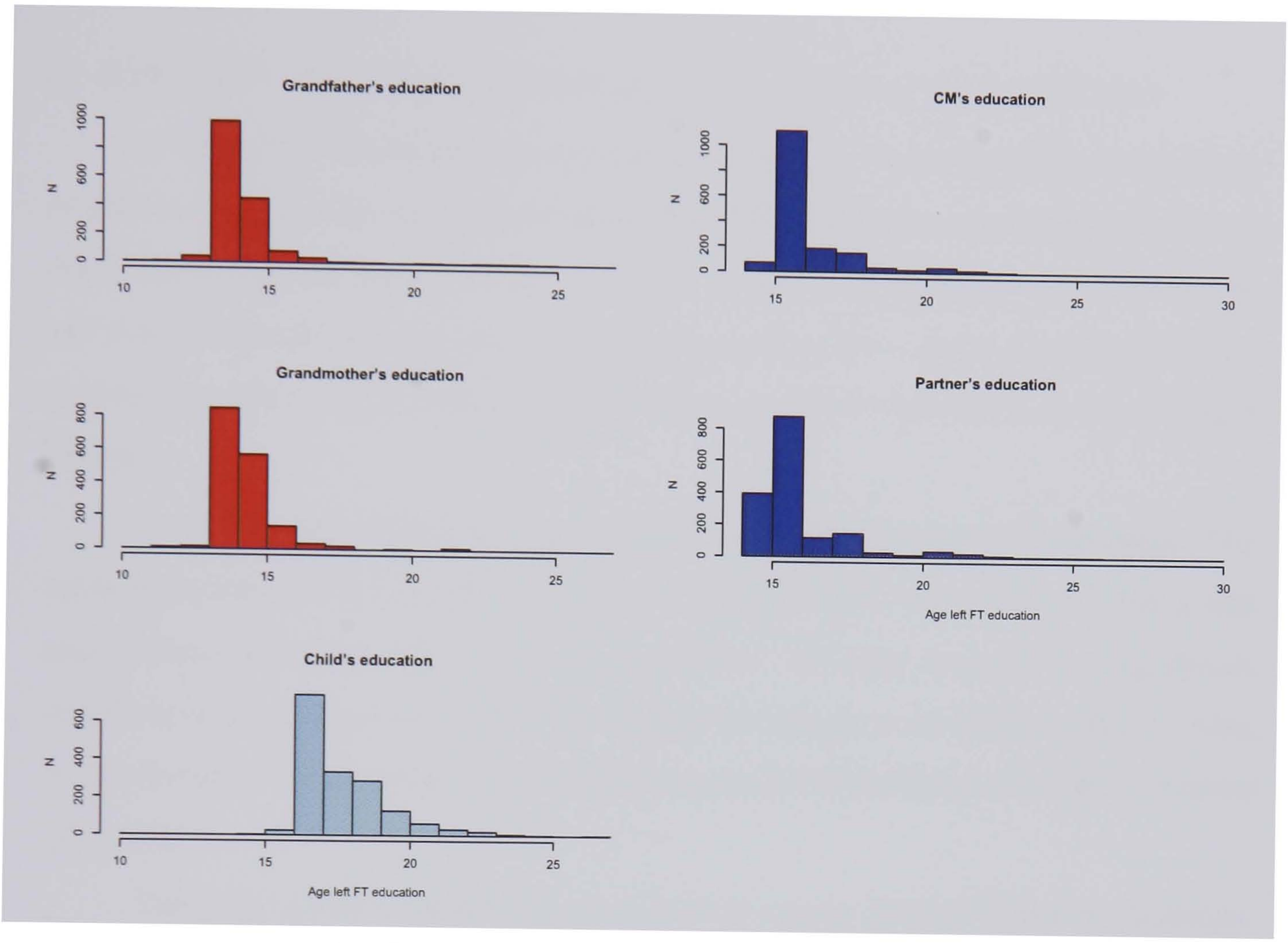


Figure 3.1: The education distributions for all three generations

3.3.1 Some censored regression results

We use here a censored regression model, assuming exogeneity of all the explanatory variables. While aware of potential endogeneity problems, it seems useful for us to examine such rich three-generational data in more detail in a first step.

The censored regression results in Table 3.4 show some evidence, although marginal that grandparents' education matters. This is in line with the first summary table in the data description section, which indicated a positive effect. Noticeably, the coefficient on aggregated parents' education (2) and on mother's education (3) drops when introducing grandparents, which is indicative of an overestimated coefficient when omitting grandparents' characteristics. However, the coefficient on father's education rises, which in turn is indicative of a neglected correlation between the newly-introduced grandparent variables. One standard deviation of the ability factor raises child schooling by 0.27 years, whereas one standard deviation of the social class factor lowers child education by -0.09 years [3]. The social class coefficient is only significant at the 10% level.

In Table 3.5, we present some more detailed results. These will answer the question of what one can do with such detailed data as the NCDS provides, in addition to what one could do having only grandparents' education or even only two-generational data. We will show the results of some censored regression specifications which indicate significant grandparent effects. It is interesting to see through which channels such effects might operate.

The interaction effects portray various channels of grandparental influence. The table illustrates the sequential way by which we proceeded. We estimated first a rich specification in column (1) for the oldest children. We then excluded non-significant variables for grandparents (2),(3). We widened the sample to all children (4)-(7). Also, in this sample of all children, we contrasted parents above and below the median income (6) and (7).

Table 3.5 reveals a particular feature of these results: the coefficient on "Substantial help of grandmother while child at school" is negative and significant in the first three columns of the eldest child sample, whereas the financial or other help variables never seem significant. Also the variable "grandparents gave childcare help" is positive and significant in the all children sample. The grandparent help variables might have a special meaning in this context, since this variable might well pick up the effect of resource-constrained families. This is confirmed in the the last two columns, where grandparent help effects disappear as soon as we consider incomes above the median. In general, grandparents' inheritances seem important for the wealthier sample, whereas having an educated grandmother moderates the negative effect in the lower-income group. The results are relatively robust, since we also include cohort member ability measures which should capture some of the unobserved heterogeneity. However, the variables "Substantial help of grandmother while child at school" and "inheritance" are not ideal, since they might select particular groups of grandparents. Notice as well that ability effects are absent for the lower-income subsample which suggests a non-linear effect of ability.

[3]These factors are normalised and have variance 1. The effects of one standard deviation can thus be directly read from the coefficient estimates

The help variable is certainly the most salient, but also the most controversial. It would be interesting to assess the interplay of working mothers and helping grandmothers. If the mother works, so the hypothesis goes, the grandmother might help out more in the household and have more influence on grandchildren. It seems unlikely that the grandmother replaces the mother figure totally, but an increased grandmother influence on grandchildren is plausible in this setting. We have tried to assess the work substitution motive with NCDS data, including indicators of mother's work, but these were only sparsely available and never significant in the reduced sample sizes examined. With different data, however, these interaction effects would be very interesting to examine and might have consequences for childcare arrangements and, as seen here, for outcomes, Michael (2005).

In Tables 3.6 and 3.7 we split the sample of the eldest children between boys and girls. In the previous Table 3.4 there seemed to be a sizeable difference between boys and girls in means, but in the Tables 3.6 and 3.7 we observe also an overall difference in coefficients. Grandparents' influence is not existent for girls, whereas the effect is very strong and nearly doubled in the case of boys compared to girls. The main effect seems to come from grandfathers' education. Parents' coefficients are significantly lower by the introduction of grandparents in the case of aggregate measures, but parents coefficients exhibit the same asymmetry as before (fathers' coefficient raises, mothers coefficient drops, except when excluding grandmothers). Grandfathers seem to be much more important than grandmothers and exert an influence comparable to parents in the case of boys.

Finally, we introduce dummies for grandparents being alive at the time of the survey, called "CM's father alive" etc. for "cohort member's father alive". We do this only for the aggregated variables, since in the disaggregated ones, the sample sizes are too small to include the necessary cohort dummies. This results in some modified estimates summarized in Table 3.8. Looking at the first column, and thus at the joint sample of boys and girls, we can observe that the effect of having a grandfather or grandmother being alive measured by the dummy variables is positive, although only significant for cohort member's grandparents. However, when we examine the coefficient of the variable "Alive", taking the value one if all of the dummy variables indicate that the grandparents are alive, we can observe a negative interaction effect of this variable with education. This means that being alive has a positive effect on the education of grandchildren, whereas the effect of education of grandparents is lower or very small if they are alive. The effect of grandparents being alive seems similar for girls and boys as far as the on-off effects of the dummy variables "CM's father alive" etc. is concerned. However, the effect of grandparents education is again stronger for boys, and the interaction effect with grandparents being alive is negative for boys and girls, but greater for girls. In the case of girls, this negative interaction between being alive and human capital is not compensated with a greater coefficient on human capital of the grandparents (HKGPAR). In order to control for potential non-randomness of death by grandparents we included new controls which proxy the behaviour of grandparents towards their health. For example, we included controls for grandparents being heavy smokers, which should correlate with mortality. This maintains the strong observed effects. Note that this effect seems to be a substitute for the "inheritance effect", since including inheritance dummies now produces no significant effects.

Thus, it seems presence is important, not only resources. It is nevertheless difficult to distinguish this from a strategic model, since grandparents could provide childcare as a resource to parents, or financial transfers other than bequests.

After these censored regression results we have to address the problem of unobserved variables. The coefficients on education and consequently on all the interactions with education might be severely biased due to unobserved heterogeneity. In general, as outlined in the literature section, researchers have thought about ability as an important omitted variable, but one can also think of parenting attitudes, parenting skills and other unobserved characteristics. These effects might additionally be hidden or enhanced by measurement error.

	Dependent variable: Childs' education, first born (Years of schooling)				
	(1)	(2)	(3)	(4)	(5)
HKPARENTS[a]	.396 (.04)	.354 (.05)			
HKGPAR		.098 (.05)			
Mother's education			.429 (.07)	.397 (.07)	.373 (.06)
Father's education			.129 (.04)	.133 (.04)	.161 (.04)
Grandmother's education				.02 (.07)	
Grandfather's education				.054 (.05)	.092 (.04)
Child female dummy	.665 (.10)	.662 (.10)	.627 (.13)	.655 (.13)	.586 (.13)
Number of siblings	−.222 (.15)	−.046 (.15)	−.273 (.20)	−.192 (.20)	−.032 (.19)
Number of siblings (parents)	−.035 (.02)		−.012 (.03)		
Ability (factor[b])	.256 (.06)	.264 (.06)	.275 (.08)	.258 (.07)	.216 (.07)
Social class (factor[c])		−.092 (.05)			−.125 (.08)
Constant	16.138 (.74)	15.626 (.75)	15.935 (.98)	15.413 (1.02)	16.581 (.396)
N	1760	1685	1114	1089	1139

Table 3.4: Censored regression results

[a]HKPARENTS is equal to $\frac{\text{father's education}+\text{mother's education}}{2}$ and similarily HKGPAR is equal to $\frac{\text{grandfather's education}+\text{grandmother's education}}{2}$. See Section 3.4.1 for more details on the construction and reasoning behind these variables

[b]We extracted a first principal component from test scores of the cohort member up to age 11, see Appendix for details. The social class factor has an inverse scale, higher values reflecting lower social class.

[c]We extracted a first principal component from all the social class indicators using the method of Tipping and Bishop (1999) to recover missing values, see Appendix for details

	Dependent variable: child's education						
	Oldest children only			All children[a]			
						$I \leq$ med(I)	$I >$ med(I)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HKPARENTS	.372	.381	.369	.371	.361	.329	.374
	(.05)	(.05)	(.05)	(.04)	(.04)	(.07)	(.05)
HKGPAR	.021	.036	−.109	.072	−.049	.006	−.097
	(.06)	(.05)	(.08)	(.05)	(.07)	(.10)	(.10)
Child female dummy	.673	.674	.676	.639	.647	.838	.446
	(.10)	(.10)	(.10)	(.08)	(.08)	(.11)	(.12)
Number of siblings	−.246	−.241	.248	−.298	−.302	−.358	−.267
	(.07)	(.07)	(.07)	(.05)	(.05)	(.08)	(.08)
Birth order grandmother	.002	.0007	.003	.011	.013	−.007	.032
	(.02)	(.02)	(.02)	(.01)	(.01)	(.02)	(.02)
Mother worked since child at school	−.089	−.081	−.090	−.129	−.157	−.157	−.111
	(.11)	(.11)	(.11)	(.08)	(.08)	(.12)	(.13)
Mother little interested in child's education	−.193	−.216	−.202	−.172	−.155	−.097	−.206
	(.14)	(.14)	(.15)	(.11)	(.11)	(.16)	(.16)
Grandparents any other help	−.090						
	(.43)						
Grandparents give financial help	−.064						
	(.43)						
Inheritance > £500	.328						
	(.12)						
Inheritance information missing	−.180						
	(.33)						
Grandparents' help intensity (derived)	.024						
	(.05)						
Substantial help of grandparents while child at school	−.205	−.201	−.828	−.107	−.659	−.777	−.429
	(.10)	(.10)	(.41)	(.08)	(.35)	(.50)	(.50)
Grandparents gave childcare help	.157	.208	.20	.296	.289	.306	.218
	(.14)	(.13)	(.13)	(.11)	(.11)	(.15)	(.16)
Ability(factor)	.210	.231	.212	.183	.162	.109	.205
	(.06)	(.06)	(.06)	(.05)	(.05)	(.06)	(.07)
Social class (factor)	−.108	−.100	−.091	−.077	−.071	−.126	−.022
	(.06)	(.05)	(.06)	(.05)	(.05)	(.07)	(.06)
HKGAR x HELP			.172		.151	.125	.142
			(.11)		(.09)	(.13)	(.13)
HKGPAR x INHERIT			.092		.084	.045	.124
			(.03)		(.02)	(.03)	(.03)
Constant	16.79	16.768	17.286	16.458	16.94	16.665	17.381
	(.71)	(.70)	(.75)	(.62)	(.66)	(.81)	(1.181)
N	1707	1707	1707	2297	2297	1129	1168

[a]This means all children above sixteen years in the sample, there can be more than one child per family. Standard errors in parenthesis. Significance levels are $+p < 0.10$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. Additionally, we included variables measuring inter-sibling birth intervals, grandparents' cohort dummies. The standard errors are adjusted for clustering in the all children sample

Table 3.5: Censored regression results and interactions

	Dependent variable: Childs' education (girls)				
HKPARENTS	.385 (.08)	.384 (.08)			
HKGPAR		−.01 (.09)			
Mother's education			.461 (.12)	.506 (.12)	.388 (.10)
Father's education			.12 (.06)	.119 (.06)	.171 (.06)
Grandmother's education				−.026 (.10)	
Grandfather's education				−.072 (.08)	−.063 (.07)
Number of siblings	−.474 (.26)	−.395 (.26)	−.789 (.33)	−.711 (.34)	−.579 (.32)
Ability(factor)	.344 (.09)	.345 (.10)	.418 (.13)	.407 (.13)	.40 (.12)
Social class (factor)	−.104 (.09)	−.118 (.09)	−.102 (.15)	−.140 (.15)	−.201 (.14)
Constant	18.067 (.62)	16.787 (1.341)	18.545 (.883)	18.056 (.95)	18.147 (.62)
N	801	790	511	497	534

Table 3.6: Censored regression results for girls only

	Dependent variable: Childs' education (boys)				
HKPARENTS	.378 (.06)	.337 (.05)			
HKGPAR		.224 (.07)			
Mother's education			.335 (.08)	.244 (.08)	.341 (.08)
Father's education			.133 (.05)	.144 (.05)	.179 (.05)
Grandmother's education				.094 (.09)	
Grandfather's education				.21 (.06)	.21 (.05)
Number of siblings	.05 (.18)	.197 (.19)	.183 (.24)	.361 (.24)	.263 (.24)
Ability(factor)	.182 (.07)	.208 (.07)	.193 (.09)	.201 (.09)	.121 (.09)
Social class (factor)	−.064 (.06)	−.074 (.07)	−.089 (.09)	−.082 (.10)	−.174 (.10)
Constant	16.268 (.84)	15.062 (.89)	15.595 (1.00)	13.859 (1.09)	15.964 (.49)
N	915	920	585	575	620

Table 3.7: Censored regression results for boys only

	Dependent variable: Childs' education		
	Boys and girls	**Girls**	**Boys**
HKPARENTS	.369	.430	.325
	(.05)	(.09)	(.05)
HKGPAR	.176	.100	.272
	(.06)	(.11)	(.09)
Alive x HKGPAR	−.156	−.193	−.137
	(.05)	(.09)	(.07)
CM's father alive	.498	.545	.509
	(.19)	(.33)	(.24)
CM's mother alive	.337	.421	.268
	(.16)	(.27)	(.20)
CM's partner's father alive	−.108	−.134	−.087
	(.10)	(.17)	(.13)
CM's partner's mother alive	.092	.027	.188
	(.11)	(.20)	(.14)
Child female	.687		
	(.10)		
Number of siblings	−.079	−.465	.221
	(.15)	(.26)	(.19)
Ability (factor)	.279	.359	.218
	(.06)	(.10)	(.07)
Social class (factor)	−.098	−.157	−.047
	(.05)	(.09)	(.06)
Constant	15.121	16.179	14.652
	(.78)	(1.3576)	(.95)
N	1643	771	872

Table 3.8: Censored regression results for examining the influence of grandparents being alive

3.4 Identification strategy

The identification strategy seeks to establish whether there exists a causal link between parents, grandparents and child education in a regression context. As outlined in the previous section, a researcher might suspect that some of the covariates might be endogeneous, i.e. correlated with the unobserved components of this equation. Let the outcome of interest be denoted as E for education. Partition the set of covariates X into exogeneous components, X, and endogeneous components X^e. Suppose that we dispose of instruments Z, correlated with the endogeneous regressors but not with the unobservables.

The model to be estimated is a reduced form equation for education in the spirit of Oreopoulos et al. (2006). The instruments used will be described below. We proceed initially as if we had only two-generational data.

$$E_i^{CH} = \gamma_0 + \gamma_1 X_i^e + \lambda X_i + \epsilon_i. \tag{3.1}$$

Expressing the equation in the previous notation:

$$E_i^{CH} = \gamma_0 + \gamma_1 E_i^P + \lambda X_i + \epsilon_i. \tag{3.2}$$

where E_i^{CH} is child education, E_i^P a matrix of mother's and father's education and these are the potentially endogeneous variables X^e, X_i a matrix of other family characteristics.

In general, estimation of an equation such as (3.2) presents several problems. First of all, as mentioned, education of the parent might be correlated with unobserved factors, contained in the error of the equation. Related to this, we might exclude from the education equation other important influence factors, such as income or family characteristics which are all summarized in the same error term.

Supplementing the above equation with income variables requires the use of additional instrumental variables, which are not available for this dataset. We will therefore assume that education is a good proxy for other, favorable outcomes later in life, and thus we shall not need to account for income. Higher income can therefore be seen as another manifestation of (higher) education levels.

The data of the NCDS survey allows us to extend the above equation by including the education of grandparents. Additionally, one can instrument both variables by available schooling reforms. As mentioned before, we have observations on schooling attainment of the child, of the cohort member (parent), and naturally on the cohort member's parents. Notice that this gives data on three generations, as the cohort member's parents are two of the four possible grandparents to the cohort member's children. Thus the equation of interest becomes:

$$E_i^{CH} = \gamma_0 + \gamma_1 E_i^P + \gamma_2 E_i^{GP} + \lambda X_{i\cdot} + \epsilon_i \qquad (3.3)$$

This makes estimation more complex, since we have in our data four possible education variables (mother's education, father's education, grandmothers' and grandfathers' education, E_i^P, E_i^{GP}) which are now in the matrix of endogeneous variables X^e. As will be mentioned below, the NCDS data has very limited information on the partners' variables. Thus, a split in mothers and fathers is only feasible in a subsample of the NCDS, as we will explain below.

This data, we feel, provides us with excellent instruments and therefore a great opportunity. The majority of these instrumental variables have the advantage that they can be seen as natural experiments.

A school leaving age reform which affects grandmothers and grandfathers equally can be used to instrument for grandparents' education. This reform, the raising of the minimum school leaving age in 1947, was unanticipated, and lengthened schooling for one year for all pupils aged 14 or under. It provides the necessary exogeneous variation, since it is unrelated to ability or parenting skills. The law was well enforced, and implemented simultaneously within the UK. There is wide ranging evidence that the average years of schooling rose by up to 0.5 years comparing pre- and post reform cohorts, with the effects being permanent.

At the parent generation level, another similar school leaving reform, the 1973 raising of the minimum school leaving age. This affected the partners of the NCDS cohort members. The school leaving age reform implemented in 1973 is similar and had a comparable effect on the education NCDS cohort member's partners to that on the grandparents, although the effect is somewhat lower. Some of these reforms have been successfully used by Chevalier et al. (2005) and Harmon and Walker (2000).

Recall that the age of NCDS cohort members is constant within surveys, since they were born between March 3 and March 9 1958: thus, we will not be able to use this particular reform instrument for the cohort members. Partners' age however varies and we can thus use the 1973 reform to instrument partners' education.

We present in Table 3.9 the proportion of partners subject to the reform by gender. One can observe that there are potentially more female than male partners who were affected by the reform. This is natural, since we observe that male cohort members tend to have rather younger partners. This means that female partners represent very little variation with respect to the reform since a majority of them has been subject to the reform and a minority has not been exposed to the reform. Thus the reform effect is not well identified for female partners. Since this is a potential problem of identification, in one sample[4] we drop all male cohort members and retain only male partners (for which the reform is well identified). This also has the advantage that we can exploit the detailed information on the cohort member to attempt to identify mothers' effect.[5]

Reform		**Male Partners**	**Female Partners**	**Total**
Subject to reform	ROSLA$==1$	803	1,203	2006
Not subject to reform	ROSLA$==0$	1,399	232	1631

Table 3.9: Partners by gender and reform impact

We use birth order as the instrument for the NCDS cohort members. Clearly, a person's birth order should not be correlated with his or her unobserved or innate factors, since a childs' position in their family's birth order is an event of chance. Previous studies which examined birth order effects, Black et al. (2005) or used birth order as an instrument, Gary-Bobo et al. (2006), put forward certain criticisms as to the use of birth order as a valid instrument. For example, a higher birth order necessarily implies also a

[4]This will be called sample 2 and will be described below

[5]The remaining cohort members are mothers after dropping male cohort members and the associated female partners. Notice that we keep only stable unions and marriages in order not to confound the effects of partner changes with the effect of interest. The effect of divorce would be another interesting question to examine

larger family. There is a sizeable literature on the (potentially negative) effects of family size on outcomes. We include a measure of mother's family size in all equations in order to avoid a confounding effect. Secondly, there might be cohort effects in education for the mother, and later-born mothers might be in different cohorts for educational reasons. In the NCDS, all mothers belong to the same cohort, and thus we cannot include cohort effects. A third criticism is that a researcher should also include mother's age at first birth, since the mothers of first-born children are likely to be younger than the mothers of later-born children. We respond to this criticism by including a proxy age at first birth, so that we can capture such differences in age at first birth [6]

Unfortunately, we cannot use the partner's birthorder or grandfather's birth order as instruments since they are not available in the data. The NCDS was first conceived as a one-shot peri-natal mortality survey (PMS) in 1958, and detailed information was collected on the mother (the granmother in our perspective) and on the child (the cohort member in our perspective). Information on the father (garndfather in our perspective) was collected sparsely through the mother in the PMS. Information on the partner was first collected in 1981, and is also very limited unfortunately.

3.4.1 Implementation

As mentioned before, the education variable is observed for cohort members and partners, which does not correspond to the natural division into mothers and fathers. Our instruments need to be applied to partners and cohort members, and have less power to identify father's and mother's effects. This has essentially to do with the age structure of partners as will be shown below. We propose thus two strategies:

1. Aggregate father's and mother's education into a single variable, called parental human capital (HKPARENTS). We have checked that the weight in parents' education does not influence the results significantly and we have thus allocated a weight of $\frac{1}{2}$ to fathers and a weight of $\frac{1}{2}$ to mothers education. This corresponds to a hypothesis of equal importance and estimate a modified version of (3.2) or (3.3). Previous studies such as Plug and Vijverberg (2005), Oreopoulos et al. (2006) and Duflo (2000) have

[6] We regressed age on the number of children so as to obtain a proxy for age at first birth. Including this measure did not change the results for birth-order in any way

found mothers to be of greater importance than fathers. Thus one may think that this is also the case in this study. As we have no guidance how to aggregate the variables in a weighted way, we decided to adopt an agnostic point of view and allocated equal weights. The results are not very sensitive to the equal weights assumption, as can be seen further below in Figure 3.2.

$$E_i^{CH} = \gamma_0 + \gamma_1 \text{ HKPARENTS} + \lambda X_i + \epsilon_i. \tag{3.4}$$

$$E_i^{CH} = \gamma_0 + \gamma_1 \text{ HKPARENTS} + \gamma_2 \text{HKGPAR} + \lambda X_i + \epsilon_i. \tag{3.5}$$

Aggregating allows us also to reduce measurement error in the education variable, and it ensures that the variable behaves similarly to a continuous variable. For grandparents, we make the same distinction, and if we estimated an aggregated model for the parents we used also the aggregated measure for the grandparents (HKGPAR).

2. Drop all cohort members who are male, and estimate father's and mother's effects as implied by Equations (3.2) and (3.3).

Note also that we are leaving out a grandparent couple, since we have only detailed information on the grandparents of the cohort member. Thus any effect found could well be larger or smaller, taking into account as well the influence of partners' grandparents.

We use a large set of background controls, including grandparents' cohort effects, inter-sibling birth space indicators, survey year dummies and parent and child characteristics. We do not use age of the children, since this is directly related to the censoring mechanism. A higher age implies a much higher probability of being censored and this might bias our results. For convenience we extract a single principal component from the test scores up to age eleven of the cohort member, so we can include a single measure which we coin "ability" in all equations. Clearly this measure might suffer from endogeneity bias and might be correlated to the control function we are trying to estimate and include in the regressions. However, when we exclude ability, the coefficients on the control function are larger and more precisely estimated as well as the effects on fathers education, but our main conclusions do not change significantly. See the Appendix of this chapter for more details. We also show there the same regressions including age effects. There is also no significant change

We propose then the following steps, which will appear below in the Results section.

The aggregated human capital and mother's/father's human capital will be instrumented by the 1973 school leaving reform that affected the 1957 birth cohort in the UK, and mothers birth order. We try also a wider definition of the reform (dummies for years around the date) and the presence of older male siblings as additional instruments. To implement these instrumental variable techniques in a censored variables context, we will estimate a first stage equation of the following type:

$$\text{HKPARENTS}_i = \alpha + \rho_1 \text{ ROSLA73} + \rho_2 \text{ BIRTH ORDER} + \lambda X_i + \eta_i. \quad (3.6)$$

As explained above, HKPARENTS denotes parents' combined stock of education.

ROSLA73 is a dummy variable taking value one if the individual was subject to the raising of the school minimum leaving age. BIRTH ORDER reflects the birth order position of the mother. We construct the control functions from these first stage regressions. We compute the standardized residuals as described in the two-step methodology in the next subsection.

For grandparents, we estimate a similar first stage equation, and construct control functions analogously. For example, for grandfathers we would estimate

$$E_i^{GF} = \alpha + \psi_1 \text{ ROSLA47F} + \lambda X_i + \eta_i. \quad (3.7)$$

Here ROSLA47F is a dummy variable taking the value one if the grandfather had been subject to the raising of the minimum school leaving age in 1947 (analogously ROSLA45M and grandmother's birth-order as an additional instrument for grandmothers).

Introducing control functions for parents and grandparents into equation (3.5) allows us then to control for endogeneity of parent's and grandparent's education. This three-generational approach to education transmission is new and allows us for the first time to disentangle causal effects outside a two-generation framework.

The proposed steps of estimation are detailed in Table 6 in the Results section.

We test the validity of all instruments and first stage equations following the methodology discussed in Murray (2005). We assess the explanatory power using a Stock-Yogo test. Additionally, we can also estimate a two-generational transmission model using grandparents and parents only, and test the overidentifying restrictions implied by having one additional instrument for the grandmother[7]. We also present various robustness checks in the last section which reinforce the confidence in the stability of the obtained results.

3.4.2 A censored normal regression model with endogeneous variables using control functions

In the NCDS, child education is not fully observed, but is censored for some children as they had not yet completed their education at the time of the survey. Thus we need to develop a framework in which to analyse this censoring (which can be seen as a so-called top-coding problem), and incorporate endogeneity concerns regarding the education transmission. We cannot use the OLS Instrumental variable estimator since it would not be able to account for censoring. Another intuitive approach is to model the unobservables directly and include them as an explanatory variable in the equation. This is the method of control functions. The method of control functions has the advantage that the reasoning can be applied similarly to the semi-parametric case.

Denote the outcome of interest as y. Suppose

$$
\begin{aligned}
y^* &= \gamma_0 + \lambda X + u, \\
X &= (X^e, X), \\
X^e &= \Pi Z + v.
\end{aligned}
$$

Suppose observations are censored at a threshold c_i, different for each individual. Thus

[7] In this case grandmother's birth-order and the ROSLA-schooling reform affecting grandmothers

$$y = y^* \text{ if } y^* < c_i,$$
$$y = c_i \text{ otherwise.}$$

A convenient assumption is to give u and v a joint normal distribution. This will be used for parametric identification. The error terms (u, v) are distributed bivariate normal with mean 0 and variance-covariance matrix Σ, in which element Σ_{11} is given by σ_{uu} and element Σ_{12} is given by σ_{uv} etc. Let $\beta = (\gamma_1, \lambda)$ in the case of a two generational transmission model.

Then we can write the conditional expectation of y^* as

$$E(y^* \mid X, Z) = \gamma_0 + \beta X + E(u \mid X, Z).$$

by the independence of u and X given v, the key assumption of all control function methods.

$$E(y^* \mid X, Z) = \gamma_0 + \beta X + E(u \mid v, Z).$$

Similarly, the variance is given by

$$V(y^* \mid X, z) = V(y^* \mid v, Z).$$

Then

$$E(u \mid v, Z) = \frac{\sigma_{uv}}{\sigma_v^2} v.$$

which follows from simple OLS considerations. This can be decomposed as

$$= \frac{\sigma_{uv}}{\sigma_v \sigma_u} \frac{\sigma_u}{\sigma_v} v.$$

and we can define

$$\rho = \frac{\sigma_{uv}}{\sigma_v \sigma_u}.$$

Then

$$V(u \mid v, Z) = \sigma_u^2 (1 - \rho^2)$$

Thus $y^* \mid X, Z$ is normally distributed with mean

$$\gamma_0 + \beta X + \rho \frac{\sigma_u}{\sigma_v} v,$$

and variance as above.

3.4.3 Maximum Likelihood Estimation

We can express the conditional density of the model as a product of two densities changing the conditioning set.

$$f(Y, X \mid Z) = f(Y \mid X, Z) f(X \mid Z).$$

Take logarithms of this expression

$$\log f(Y, X \mid Z) = \log(f(Y \mid X, Z)) + \log(f(X \mid Z)).$$

$$\begin{aligned}
\text{LLik} &= \sum_i^N d_i \left[-\frac{1}{2} \log(\sigma_u^2 (1-\rho^2)) - \frac{1}{2\sigma_u^2(1-\rho^2)} \left(y_i - \gamma_0 - \beta X_i - \rho \frac{\sigma_u}{\sigma_v} v_i \right)^2 \right] \\
&+ (1-d_i) \log \left[1 - \Phi \left(\frac{y_i - \gamma_0 - \beta X_i - \rho \frac{\sigma_u}{\sigma_v} v_i}{\sigma_u \sqrt{(1-\rho^2)}} \right) \right] \\
&- \frac{1}{2} \log(\sigma_v^2) - \frac{1}{2\sigma_v^2} \left(x_i - \Pi' Z_i \right)^2 .
\end{aligned}$$

We can motivate this formula in three parts as follows. The variable d_i is the censoring indicator, $d_i = \mathbf{1}(y^* < c_i)$. The first part of the likelihood represents the contribution of the uncensored observations. They contribute with a standard normal density, augmented by a control function element v_i. The second part is the contribution of the censored observations, modeled as the probability complement of the normal distribution. This relates to the feature of our data that we observe children reaching "at least" this level of education. The third term models the error distribution, depending on the estimated covariates, or can also be seen as the usual first stage regression, where $v = X^e - \Pi' Z$

3.4.4 A two-step approach

First we estimate a model of the reduced form. This can be done via simple OLS to obtain:

- STEP 1

  $$X^e = \Pi' Z + v.$$

  and recover $\hat{\Pi}$ and $\hat{\sigma_v^2}$.

  Then calculate

  $$\hat{v} = X^e - \hat{\Pi} Z,$$

  and the standardised error, where

  $$\hat{\epsilon} = \frac{\hat{v}}{\hat{\sigma_v}}.$$

- STEP 2

  Do Tobit (or censored regression) for Y on a constant, X and $\hat{\epsilon}$. This way we obtain consistent estimates of $\hat{\gamma}_0, \hat{\beta}$ and $\hat{\gamma} = \rho\hat{\sigma}_u$

  From there we can identify and calculate

  $$\hat{\sigma^*} = \hat{\sigma}_u^2(1 - \hat{\rho}^2).$$

Now we can recover the original parameters of the model. We know that the variance we have calculated here is

$$V(u \mid v, Z) = \sigma_u^2(1 - \rho^2),$$

call it

$$\hat{\sigma}_*^2 = \sigma_u^2(1 - \rho^2).$$

Then after some manipulation and using the fact that

$$\hat{\gamma} = \hat{\rho}\hat{\sigma}_u,$$

we can recover

$$\hat{\rho}^2 = \frac{\hat{\gamma}^2}{\hat{\gamma}^2 + \hat{\sigma}_*^2},$$

then we can calculate

$$\hat{\sigma}_u^2 = \frac{\hat{\gamma}^2}{\hat{\rho}^2}.$$

3.4.5 Censored regression quantiles with endogeneous regressors

Another approach to deal with endogeneity in the presence of censored observations would be the method proposed by Blundell and Powell (2004). The authors extend ideas proposed by Chen and Khan (2001) to the case where the first-stage control functions are explicitly estimated. Other related models are the ones of Das et al. (2003), Powell and Honore (2005) and an appealing approach of Chernozhoukov and Hong (2002). These methods leave the error distribution unspecified and can thus be seen as semi-parametric or non-parametric methods.

The identification of endogeneous regressors stems from a distributional exclusion restriction on the unobservable errors. One could refer to this as a quantile variant of the control function approach. Blundell and Powell (2004) propose a two-step approach. In a first stage, the control variable is estimated non-parametrically, as well as the conditional quantiles, given regressors and control functions. In a second stage the (weighted) difference between the estimated quantiles is calculated, with weights being non-zero only if (i) both *estimated* quantiles are below the censoring threshold c_i and (ii) the difference in control variables of any pair of observations is small. Thus, intuitively, we take differences of the estimated quantiles of two individuals who have the same or at least similar unobservables. An intuitive comparison would be a classical matching estimator and here we would match on the estimated unobservables. More formally (see Blundell and Powell (2004)), assume our model is

$$y_i = \min\{x_i'\beta + \epsilon_i, c_i\}.$$

Suppose now that, as before, some regressors are endogeneous, named x_i^e and we have instruments for these, z_i. The x^e are generated by the reduced form

$$x_i^e = \pi(z_i) + v_i.$$

The function π is some possibly non-parametric function of the instruments. A condition implied by the usual independence assumption of u, v is the distributional exclusion restriction, which can be formulated as a quantile restriction. All conditional quantiles of the error distribution ϵ_i, given x_i, z_i, are functions only of the control variable v_i. This approach does not impose any distribution on the error terms, but it is also somehow restrictive since the conditions below must hold for every quantile of the distribution. Thus

$$\mathcal{Q}_\alpha [u_i \mid x_i, z_i] = \mathcal{Q}_\alpha [u_i \mid v_i] \qquad \text{w.p.1.}$$

where the $\mathcal{Q}_\alpha [u_i \mid x_i, z_i]$ is the α conditional quantile of u_i, given x_i and z_i, and under the restriction above it can be shown that

$$\begin{aligned} q_i &\equiv \mathcal{Q}_\alpha [y_i \mid x_i, z_i], \\ &= \mathcal{Q}_\alpha \left[\min\{x_i'\beta_0 + \epsilon_i, c_i\} \mid x_i, z_i\right], \\ &= \min\{x_i'\beta_0 + \mathcal{Q}_\alpha [\epsilon_i \mid x_i, z_i]\}, \\ &= \min\{x_i'\beta_0 + \lambda_\alpha(v_i)\}. \end{aligned}$$

where the term $\lambda_\alpha(v_i)$ represents the control function.

Now we estimate these conditional quantiles and construct differences in observations of individuals with identical control variables. This will difference out the control function. Suppose

$$\hat{q}_i < c_i \qquad \hat{q}_j < c_j \qquad \text{and} \qquad v_i = v_j.$$

then it follows that

$$\hat{q}_i - \hat{q}_j = (x_i - x_j)'\beta_0.$$

which indentifies β_0.

According to Blundell and Powell (2004), the control function is to be estimated by non-parametric methods, either kernel or local polynomial regression, to guarantee the desired asymptotic behaviour of the estimator. We fit a local-linear estimator by using the locfit package, developed by Loader (1997) in R (2005). Usually, the method uses a backfitting algorithm to determine a suitable first-step bandwidth. We, however, use only bandwidths falling into the theoretically prespecified bands, since (i) the estimator is sensitive to the choice of bandwidth and (ii) the generalised cross-validation criterion leads to a different bandwidth from the prespecified, theoretical one. After having estimated the error distribution we then compare the estimated quantiles for those individuals with similar control functions in the spirit of Blundell and Powell (2004): that is constructing the weighed least squares estimator

$$\hat{\beta} = \left[\sum_{i<j} K_v\left(\frac{v_i - v_j}{h_n}\right)\hat{t}_i\hat{t}_j(x_i - x_j)(x_i - x_j)'\right]^{-1}$$

$$\sum_{i<j} K_v\left(\frac{v_i - v_j}{h_n}\right)\hat{t}_i\hat{t}_j(x_i - x_j)(\hat{q}_i - \hat{q}_j)'.$$

K_v is a kernel function that ensures that our estimates of the control functions of the two individuals are close, h_n is a sequence of scalar bandwidth terms which tend to zero at an appropriate rate with the sample size and $\hat{t}_i$ is a trimming term that cuts off observations unless the estimated quantiles are positive and the v_i fall in a compact set S. Note that we can take only the sum over observations $i < j$, otherwise the resulting matrix would be singular. Note as well the asymmetric way in which (compared to a standard GLS estimator) the weights enter this estimator.

This method is computationally more demanding, but the main advantage remains that it does not assume any particular error distribution.

3.5 Results

As the data structure of three generations is very challenging, we will extend the model step by step. We outline two generational results first and present thereafter three generational results. We will then try to assess to what extent the transmission process between generations is non-linear in the observables and unobservables [8]. Finally we

[8]Additional probit results on staying on in school can be found in the Appendix

Samples used Sample 1	 Sample 2	Estimation framework used		How to treat grandparents' variables ?
Yes	Yes	Censored regression	IV[a]	no grandparent education variables
Yes	Yes	Probit	IV	
Yes	Yes	Clustered Probit	IV	
Yes	Yes	Censored regression	IV	**grandparents exogeneous**
Yes	Yes	Censored regression	IV	**grandparents endogeneous**

Table 3.10: Details of the estimated models

[a]Instrumental variables

are going to relax the bivariate joint normality assumption and we will implement for the first time a distribution-free approach to the censored regression context under endogeneity. We will argue all through the results section that the estimated effects reveal that the previously presented censored regression transmission coefficients can be seen as upper bounds. We argue as well that parametric and non-parametric approaches provide *complementary* information.

Throughout the Results section we will be using two samples. Sample 1 uses an aggregate measure of human capital of the parents and also an aggregated measure of the human capital of grandparents. Sample 2 drops all mothers who are not cohort members. This avoids any identification problem which might arise because instruments apply to the cohort members and partners seperately, which is not equivalent to the split between mothers and fathers. This sample 2 allows us to present effects for mothers and fathers separately.

We introduce control function estimates into censored regressions to control for endogeneity. The way of proceeding is illustrated in Table 3.10. We compare the IV results with the previously obtained censored regression results.

After having examined these results, we show then that the estimated control functions are likely to violate the normality assumption. Therefore we estimate a non-parametric quantile regression model to account for this.

3.5.1 First step: Estimating the control functions

Table 3.11 presents the estimated control functions. These are equivalent to the traditional first-stage estimates of an IV model. We opted for parametric first stage regression to establish the parametric model as a benchmark case.

	Dependent variables in parametric control functions					
	(1) HKPARENTS	(2) **Mother's education**	(3) **Father's education**	(4) HKPARENTS	(5) **Mother's education**	(6) **Father's education**
Mother's position birthorder	−.051 (.02)	−.073 (.02)		−.031 (.02)	−.057 (.02)	
Affected by 1973 reform	.237 (.05)		.412 (.09)	.201 (.05)		.356 (.11)
Number of siblings	−.133 (.08)	−.113 (.09)	−.150 (.14)	−.044 (.09)	−.006 (.10)	.003 (.17)
Ability (factor)	.393 (.02)	.438 (.03)	.455 (.05)	.319 (.03)	.373 (.03)	.356 (.06)
Number of siblings (parents)	.014 (.01)	.032 (.02)	.023 (.02)			
Child female dummy	−.012 (.05)	.064 (.06)	.096 (.09)	−.029 (.05)	.014 (.06)	.087 (.11)
Joint grandparents' human capital				.355 (.03)		
Grandmother's education					.248 (.03)	.218 (.05)
Grandfather's education					.129 (.03)	.203 (.05)
Constant	3.225 (.38)	3.795 (.42)	2.579 (.72)	−2.159 (.66)	−2.049 (.76)	−4.009 (1.254)
R-squared	.11	.11	.07	.16	.18	.11
N	2348	1790	1509	1836	1394	1177
F-test on IV's	12.56 (0.000)	17.19 (0.000)	6.43 (0.013)	7.22 (0.000)	5.75 (0.0166)	9.83 (0.002)

Table 3.11: First stage estimation for parents' endogeneous variables to construct the control functions

When reading this table, remember from Table 3.10 that we used aggregated human capital of the parents (HKPARENTS) and separate variables (Mothers/Fathers). The first three columns report the estimates excluding grandparents. The last three columns include grandparents as exogeneous variables.

We aggregated the parents' variables, giving equal weight to mother's education and to father's. A priori, and as outlined in the Implementation section, there is reason to believe that mother's weight might be higher, but the results are robust even if we vary the father's weight, see Figure 3.2. There are various reasons why one can argue for and against aggregating or separating parent's education variables. An argument for aggregating is that the aggregated variables provide a more continuous measure of education than the single measures. Also, they might reduce measurement error by averaging. In contrast, it seems interesting to be able to isolate mothers' and fathers' effects.

The control function estimations indicate that we have good instruments. The F-test on the instruments is given in the last line of the table. These all are well above the Stock-Yogo critical values (which in this case boils down to a simple F-test) for all sets of variables, and we can thus reject the hypothesis of weak instruments for all, Murray (2005). However, when grandparents enter the first stages, the F-values weaken slightly. We obtain reasonable values of R-squared, explaining some of the variation in the endogeneous variables. Having been subject to the raising of the minimum school leaving age raises schooling by 0.20 to 0.4 years, whereas each step up in mother's birth order lowers schooling by -0.07 to -0.10 years of education. Note that we need to double the coefficients of birth order to calculate the effect on education since the variable HKPARENTS was aggregated from mother's and father's education with a weight of $\frac{1}{2}$ respectively. The two coefficients are highly significant in all specifications, except in (4) where identification might be threatened. Note that the coefficient on ability falls, once we control for grandparents' education in these control function estimates. [9]

In Table 3.12, we present the estimated control functions for grandparents. We used the two aforementioned schooling reforms as well as grandmother's birth order. Note that this is an overidentified equation, a fact we will use later to conduct a robustness check.

As one can see from the table, the effects of the reform are statistically and economically significant, raising schooling between 0.2-0.6 years on average in the reform group. Especially for grandmothers, the effect is large, visible in the aggregate and the separate equations. Grandmothers gain on average more than half a year of schooling. This seems to be a substantial effect, whereas grandfathers seem to gain only 0.4 years. This is still above the effect for parents, the reform thus having a larger impact on the grandparent generation than on the parents' generation. We can again statistically

[9] The estimation was also done without the ability component, which induces a more precisely estimated control function part. Results can be obtained from the author

	Dependent variables of grandparents' parametric control functions		
	(1) Aggregated GP education[a] (HKGPAR)	(2) Grandfather's education (FEDUC)	(3) Grandmother's education (MEDUC)
Grandfather subject to 1947 reform	.175 (.07)	.431 (.09)	
Grandmother subject to 1947 reform	.30 (.06)		.620 (.07)
Birth order grandmother	−.049 (.008)		−.067 (.009)
Number of siblings	−.078 (.07)	−.087 (.06)	−.054 (.069)
Ability (factor)	.20 (.02)	.216 (.02)	.20 (.02)
Number of siblings (parents)	−.015 (.01)	−.038 (.01)	−.017 (.01)
Child female dummy	.018 (.03)	−.031 (.04)	.046 (.04)
Constant	15.626 (.25)	15.68 (.33)	15.557 (.29)
	(.2544)	(.33582)	(.29214)
R-squared	.12	.07	.15
N	2150	2298	2199
F-test on IV's	24.18	81.86	67.91
	(0.000)	(0.000)	(0.000)

[a]These equations include cohort and inter-siblings birth space indicators.

Table 3.12: First stage estimation for grandparents' endogeneous variables to construct the control functions

reject the hypothesis of weak instruments.

3.5.2 Results for the parametric models

In Table 3.13, we present the results of the second stage estimation for all models *as if* we did not have information on the grandparents. This amounts to saying that we treat the problem of endogeneous education transmission in this table as a two-generational transmission. This has been the standard approach in the literature, which we now try to extend.

One can notice that the assumption of an endogeneous effect of parents seems to yield negative point estimates for parent's effects in the aggregated measures. These are imprecisely estimated and not statistically significant. In the separate sample, the effect of mothers is lower than censored regression and positive, whereas father's effect is negative and marginally significant. Thus, these results relate to the twin literature, Behrman and Rosenzweig (2002), where mother's coefficients turned out negative. Here we find that fathers have a negative effect - an awkward finding and difficult to reconcile with economic theory. This casts the first shadow of doubt on identification and thus one cannot pretend to have isolated causal effects. It also points out another potential issue of this analysis: measurement error in the dependent variable. We refine our instrumental variable by having a larger effect around the reform date and also by including an additional potential instrument, in this case if the mother had any older brother. The results for this specification are given in the last two columns. The coefficients change little, although mothers coefficient drops slightly.

	Dependent variable: child's education					
	Censored reg.	IV (control function)	Censored reg.	IV (control function)	Censored reg.	IV (control function)
Parents' joint human capital	.436 (.04)	−.461 (.35)				
Mother's education			.391 (.07)	.470 (.78)	.402 (.07)	.323 (.64)
Father's education			.189 (.04)	−.559 (.31)	.188 (.04)	−.64 (.30)
Number of siblings	−.448 (.14)	−.569 (.14)	−.574 (.18)	−.707 (.21)	−.565 (.18)	−.733 (.20)
Ability(factor)	.239 (.05)	.591 (.14)	.261 (.07)	.560 (.37)	.261 (.07)	.667 (.31)
Number of siblings (parents)	−.038 (.02)	−.051 (.02)	−.034 (.03)	−.018 (.03)	−.029 (.03)	−.013 (.03)
Child female dummy	.617 (.09)	.607 (.09)	.619 (.12)	.696 (.14)	.635 (.12)	.733 (.14)
Control function (aggregated)		1.149 (.45)				
Control function mother				−.11 (1.03)		
Control function father				1.372 (.578)		
(*extended IV set*) Control function mother						.093 (.84)
(*extended IV set*) Control function father						1.5223 (.55)
Constant	16.407 (.67) (.67111)	19.347 (1.338) (1.3383)	16.533 (.91) (.91829)	18.358 (3.07) (3.0703)	16.487 (.91) (.91602)	19.126 (2.621) (2.6214)
N	2182	2182	1328	1328	1325	1325

Table 3.13: Second stage for the censored normal regression model

	Dependent variable: parents' education			
	OLS	IV	OLS	IV
Grandparents' joint education (HKGPAR)			.356 (.03)	−.40 (.25)
Grandfather's education	.142 (.02)	−.026 (.24)		
Grandmother's education	.190 (.03)	−.191 (.14)		
Ability (factor)	.412 (.03)	.513 (.06)	.324 (.03)	.480 (.06)
Cohort member female	−.108 (.06)	−.148 (.06)		−.164 (.07)
≠ Siblings parents	−.015 (.01)	−.013 (.01)	−.008 (.02)	−.008 (.02)
≠ Siblings (grandmother)	−.006 (.01)	−.039 (.02)	−.012 (.01)	−.051 (.02)
Constant	11.287 (.65)	19.905 (4.34)	−2.550 (.66)	9.392 (3.98)
R-squared	.14	.04	.11	.05
N	2142	2117	1875	1851
F-test (IV first stage)	separate measures		aggregated measures	
ROSLA Father	25.01			
ROSLA Mother, Birthorder	38.13			
ROSLA Father,Birthorder,ROSLA Mother			14.25	
Overidentifiaction				
Sargan (NR^2)	0.479 $\chi^2(1)$ p= 0.488		0.953 $\chi^2(1)$ p = 0.328	

Table 3.14: Grandparents two-generational effect on parents

In general, father's unobservables enter the equation significantly. One can only speculate that this might reflect the missing proxy for ability. For mothers, we have included the ability factor, whereas for fathers (as partners) there is not such a measure in the data. The effect on fathers is relatively independent of the weight attached to father's education in the "joint human capital" variable as the next figure illustrates. However Figure (3.2) illustrates that the unobservable effect comes exclusively from the father. As the weight on the father becomes lower, at about < 0.2 the control function is imprecisely estimated.

We repeat now the steps above for the pair grandparents-parents (two-generational). As we can use grandmother's birthorder as an additional instrument, we can test the overidentifying restrictions implied by having more instruments than endogeneous variables [10]. As can be seen, the results for the pair grandparents-parents are very similar to the pair parents-children.

[10]We correct STATA's output slightly, since STATA erroneously assumes that the aggregated variable $HKGPAR$ is only one variable. However, we need two instruments to identify the effects and thus the correct degrees of freedom for the Chi-square test is always 1 (more restrictive)

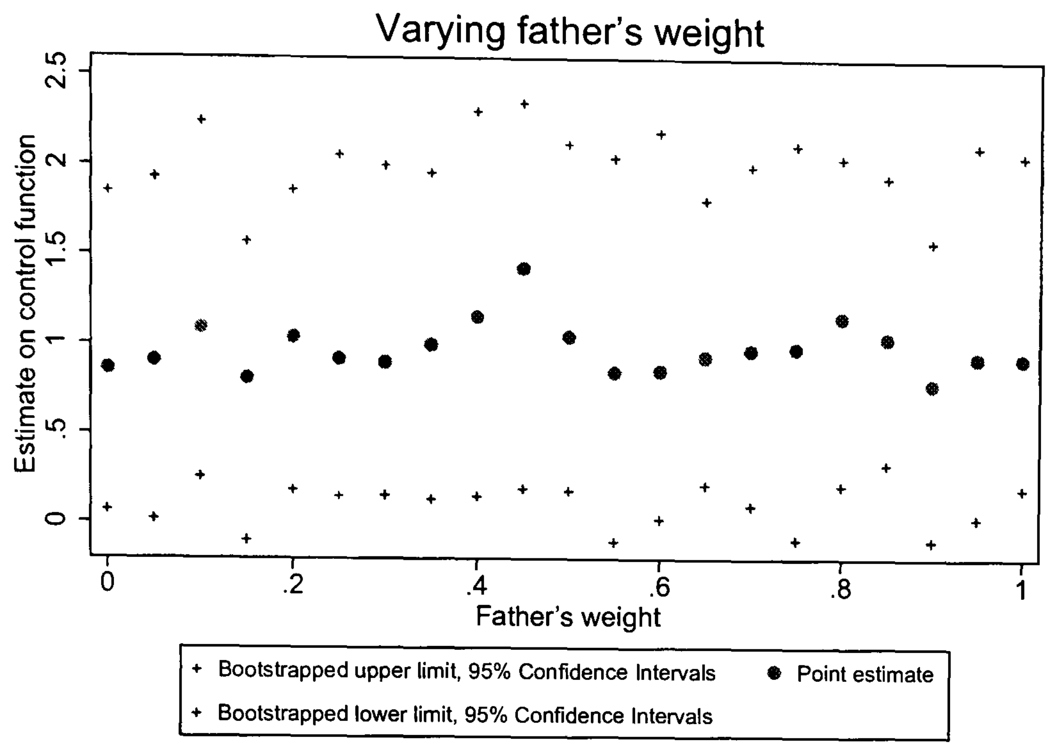


Figure 3.2: Estimates of the control function varying fathers' weight

The grandparents-parents generation equation might be less prone to measurement error, since we observe people who have already completed their education. This test is thus useful not only to assess the instruments but also to establish that the results are most likely not due to measurement error and/or censoring.

In summary, the censored regression estimates appear to be an upper bound on the true coefficients in the two cases presented. This result would be compatible with related work by Dearden et al. (1997) which assumed such a relationship. Unobserved factors as measured in the control function are significantly and positively related to outcomes. This would not be obvious in the standard IV model, since we would not estimate these control function components. However, these control functions do not tell us much about unobserved heterogeneity except for a summary measure of relatedness. We have to assume that there is a whole vector of unobserved characteristics subsumed into the univariate error. Thus, it seems that only *on average* are unobserved characteristics positively and significantly influencing outcomes. As far as the results compare to similar studies, Oreopoulos et al. (2006) came to a different conclusion, as they find that the OLS coefficients on an equation relating parental education to a child's grade progression are underestimates for the true transmission coefficients. In our data, two potential explanations exist why our effects are different. First of all we select a particular sample of children and parents, since our observations are still censored. Thus we also implicitly select only a subgroup of grandparents. Thus our estimates apply only to a subgroup of families who had children early and are relatively uneducated. Thus, we would expect our results to be different. Still, the results are in line with those of Black et al. (2003), who also find overestimated coefficients in their administrative and representative data. A second explanation of finding negative and sometimes insignificant results could be the parametric specification and in some cases the heavy censoring of the outcome variable. If we compare coefficients in the censored regressions with what other researchers have found in non-censored samples, we realise that our coefficients indicate large effects. Therefore, we should favour the results of the censored quantile regression estimator, which indicates small but positive transmission coefficients, in accordance with much of the twin and adoptees literature,Rosenzweig and Evenson (1977), Plug and Vijverberg (2005) .

3.5.3 Censored regression results introducing grandparents

As shown in Table 3.10, a first step is to construct a model into which grandparents enter exogeneously. As explained in the literature section, this has been the cornerstone of some identification work which even used grandparents as an instrument for parents, assuming that they were not related to the child's generation. The second step is then to instrument for grandparents using the outlined IV strategy.

Table 3.15 summarizes all the results obtained for this proceedure. The table might be difficult to read, but we have presented the aggregated results (grandparents' and parents' education in two variables called HKPARENTS and HKGPAR in Columns (1) and (2)) and then separated out the effects by mother/father and grandmother/grandfather in columns (3) and (4). The results document some effect of grandparents' education, as already seen in the OLS results. However, using grandparents in this estimation changes mother's coefficients drastically and makes the coefficients remain at the same level approximately as in the censored regression model without considering endogeneity. This is rather worrying, since it would cast doubt on the previous results which seemed to suggest that OLS estimates are upper bounds. However, the coefficient on mother's education is never statistically different from zero. We observe consistently that the strong correlation with ability disappears after instrumenting mother's education and this could well be picked up by the education coefficient. Otherwise, we see again fathers' coefficients negative and significant in some IV results (grandparents exogeneous), but fathers' unobservables entering significantly positively in all IV specifications

In the aggregated grandparents' control funtion, we observe that grandparents unobservables enter significantly and with a large coefficient into the transmission equation. This would tell the same story as the previous two-generational results, but now having grandparents is the important contributor in terms of unobservables. In the disaggregated results, we can see that grandfather's control function is marginally significant.

How can we reconcile these findings? It seems extremely difficult to estimate the education transmission effect with precision, since the process carries a great deal of noise. Our moderate sample sizes do not allow us to improve on this.

All results, whether three or two generational, point however to *no statistically significant effect for parents' or grandparents' education, but a large effect of unobservables.* All comparisons between censored regression and IV estimates seem to point to an upper bound argument. Consistently, we seem to find a large effect of father's control function on outcomes.

The Appendix reports estimates for repeating the above setup for two and three generations, using probit specifications. These probit specifications measure the effect of parents' characteristics on the "staying on decision" of children, by which we mean the decision to stay on in education after the minimum legal age. The effects of education are weaker in these specifications, but the unobservables turn out nevertheless again important on the fathers' side. We will now turn briefly to the full MLE model, since this is the most efficient (parametric) estimation method.

3.5.4 Results from full Maximum Likelihood estimation

This short section presents the results of implementing the full maximum likelihood, Table 3.16. The main advantage of this method is the simultaneous estimation of all equations and the more precise estimation of the control function parts in the regression.

The MLE results confirm the previously established censored OLS results, only that the unobservable components are more precisely estimated, as well as the effect of education of grandparents and parents appears to have fewer negative point estimates. In addition, we notice that the first stage regressions are less precisely estimated in this simultaneous set-up. The fact of negative point estimates here and earlier leads us to consider that the parametric approach might be partially invalid due to the restrictive distributional assumptions. It is therefore reasonable to examine also the results of the non-parametric approach detailed below.

3.5.5 Results for the censored quantile regression model

Figure 3.3 depicts why the censored quantile regression model might be a preferred model. The parametrically estimated control functions seem not to satisfy the normality assumption. We plot here the estimated density of the control function from the aggregated parental education variable (residuals from a regression of HKPARENTS on instruments and controls)

We have tried only two specifications relating to the censored regression results: using aggregate parents and grandparents, and controlling for their endogeneity. The method is computationally intensive, especially for the construction of artificially differenced datasets of size $N = \frac{n}{2}(n-1)$.

Table 3.17 presents the results assuming parents' education is endogeneous and Table 3.18 presents the results assuming parents' and grandparents' education are endogeneous.

Note that the results are sensitive to the bandwidth used. The first step bandwidth was *not* chosen by minimising a generalised criterion (see Appendix) since this led to a relatively large bandwidth (2.2), having a window width including about 23% of the sample at each point. We relied on the theoretical ranges derived by Blundell and Powell (2004). The choice of both first step and second step bandwidths influences the coefficients and this is certainly a drawback of this estimator: the results are specific to these bandwidth choices. To illustrate the sensitivity of the results, consider Table

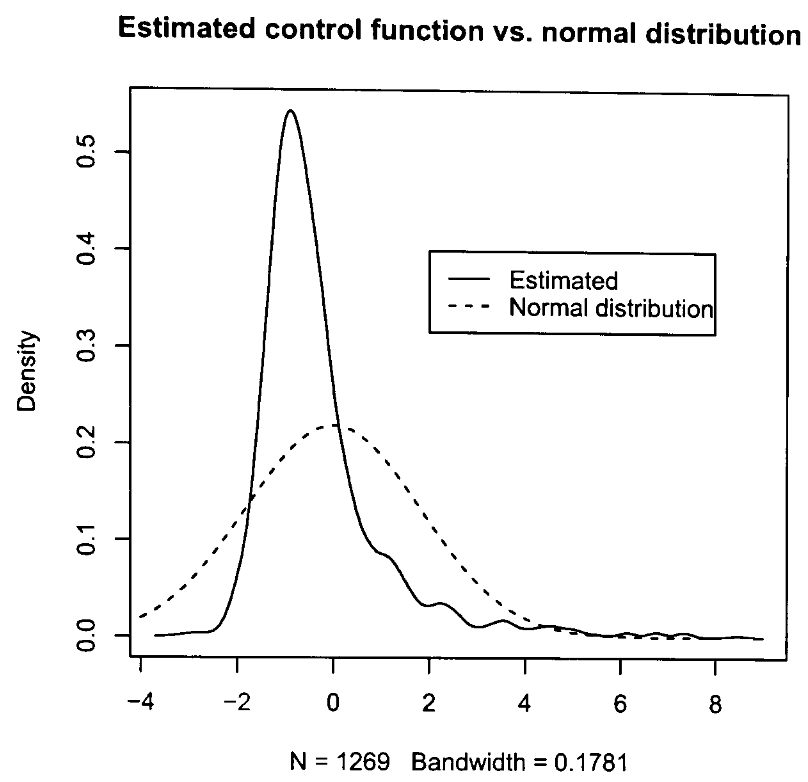


Figure 3.3: The estimated control function and the normal distribution of the same variance

3.17. Here we varied the first and second step bandwidth. One can see that coefficients are relatively stable, but standard errors are estimated less and less precisely, from column (1) to column (4), and again from (5) to (7). This suggests that precision is decreasing with decreasing bandwidths. On the other hand, it seems that precision can be artificially increased by increasing the bandwidth. A t-ratio on the female dummy of about 21 seems unacceptable in column (1), for example. Increasing the bandwidth includes more observations in the final sample and lowers thus the estimated sample variance. As a guide to which bandwidth might be appropriate (given that it does not affect the estimated coefficients as much as the first step bandwidth) we calibrate it in such way that the standard errors are about the size of the MLE estimated standard errors. This would imply $h_n = 0.001$.

The first impression is that, although imprecisely estimated, the effect of parents' human capital on the median is much lower than the effect in the censored OLS case. It is however positive and above the parametric IV results. It is also in line with estimates of the twins' literature of 0.05-0.10. As we calibrated the standard errors to match the MLE performance in column 6 for example, we can also note that the coefficients on

HKPARENTS is not more precisely estimated as under the MLE approach and that the variance ratios among the coefficients are the same as in the MLE estimation. However, as mentionned, we always obtain positive and small point estimates.

Including now the grandparents as an additional endogeneous variable lowers the parents coefficient slightly, but we still have a small positive parents' coefficient. The grandparents' coefficient on the other hand becomes slightly negative. This is consistent with a story of indirect influence through the education variable but direct influence through unobserved factors (compared to the censored regression case, where grandparents point estimate on education was negative, but their unobservables were positively related to grandchildrens' education).

The quantile estimator concludes the results section. The interested reader is referred to the Appendix, where additional probit results and some explanations of the methods used are available. In the next section we explore the robustness of the estimates, establishing various categories of potential worries, which are formulated as questions.

3.6 Alternative estimation method and sensitivity checks

Is censoring an important issue in the data ? Accounting for censoring is crucial, since Table 3.19 shows that ignoring censoring would yield much lower coefficient estimates. The results indicate that estimated coefficients are all very different in magnitude, especially the coefficient relating to human capital of parents. In the pure censored regression model, without censoring, this coefficient appears about four times smaller than in the censored case. Note that the number of siblings measure even turns from a positive effect into a negative one. The signs of the other estimated coefficients are mostly in line with the signs of the censored regression model and have, as expected, smaller magnitudes.

Do the results vary with the set of instruments used ? We do not have credible additional instruments to test the parents' first stages. Interestingly, we can test an over-identification restriction in the two-generational setting of grandparents' effects on parents. For grandparents, we have the reform dummies but we can also use grandmother's birth-order. This gives an over-identified grandmother and joint human capital equation. The acceptance of the test increases confidence in the validity of the IV results.

Do the instruments have explanatory power ? Simple F-testing of the first stage equations for one endogeneous variable reveals that we reject the null hypothesis of weak instruments with a comfortable margin. The critical values for the so-called Stock-Yogo test Murray (2005) are the simple F-test, since we have only one troublesome variable in the model with an aggregated parents measure. The test statistic is a modified F-test when we use another endogeneous variable or more of them. We are able to reject the null hypothesis for these since the values of the test are slightly lower than the F-test in the case of two endogeneous variables.

In the case grandparents are introduced as exogeneous variables, we reject easily the hypothesis that grandparents' instruments are weak, and parents' instruments pass the test with a smaller margin (simple F-test)

All first stage equations have the expected signs and also generate sufficient identifying variation. The control functions are not estimated with great precision, but the control functions of the father indicate that unobservables enter significantly into the equation. Although we are not primarily interested in these results, the t-test on the coefficient of the control function could be interpreted as a test of exogeneity, Wooldridge (2002), which we cannot reject in all cases. This confirms the view that the education transmission mechanism does suffer (at least in this data context) from bias through endogeneity.

Are the instruments valid ? The validity of the reform instruments seems hardly contestable since the reforms were truly unanticipated and implemented across the UK. These instruments are natural or quasi-natural experiments, which enables us to believe in their orthogonality. The grandparents' instruments were used by previous studies and appropriate checks to their relevance as well as validity precede this study. Nevertheless, we should perhaps examine some features of the reform at greater depth.

The instrument of grandparents selects grandparents less than or equal to 24 years of age, since these were subject to the 1947 reform. Mothers below and mothers above 24 years might be different in observed and unobserved characteristics. Therefore we will check what one could call, in reference to the treatment effect literature, "Covariate balance". We will run a simple probit model on the reform dummy to see if there are any statistically and economically significant differences for (grand-) mothers subject to the 1947 reform or those who have not been affected by the reform, (Table 3.20).

As one can see, most of the significant differences can be attributed to cohort effects. The only significant variable is grandmothers weight. But the effects are negligibly small, compared to the effect of the (grandfather's) cohort dummies on these variables. Note the increase in R-squared by almost 0.26 when including cohort dummies.

We can in addition only imperfectly control for cohort dummies by using the grandfather's (husbands) cohort dummies. Grandmothers' cohort dummies would be perfectly collinear with the reform dummy. If we could directly introduce mother's cohort dummies, any effect would most likely disappear in this probit equation. A significant variable such as grandmother's weight, which is possibly totally unrelated to the outcome of child education is certainly due to such effects. We can thus conclude that the instrument is valid once we control for cohort effects.

Do the parametric assumptions influence the results ? Finally, the reader might be worrying about the parametric assumptions we made. This concerns especially the joint normality assumption in the censored normal regression model, which identifies our model. As we have explicitly estimated a non-parametric model, these worries should be addressed. Indeed, one can see that negative coefficient estimates for the effect of parental education under the joint normal distributed models seem implausible. They disappear in the non-parametric case, which indeed suggests that the parametric assumptions influence the results to some degree. All the other coefficients, with the exceptions of the number of siblings variable, have the same sign and value as in the censored normal case. It seems that the distribution of the endogeneous variable, education, is mostly affected by the parametric assumptions. Note finally that the full MLE estimates less extreme coefficients with a higher precision, which seems to weaken the criticisms of the jointly normally distributed model.

3.7 Conclusions

This paper has analysed intergenerational educational transmission across three generations. It has reviewed and applied several econometric techniques for dealing with censoring and endogeneity in various settings which are interesting in their own right.

Grandparents' education enters in many censored regression specifications as significantly positively related to grandchildrens' outcomes, especially for grandfathers in relation to male grandchildren. In the context of exogeneity, we find some evidence for a strategic role, in the form of bequests or presence of grandparents. The censored regression results support the interpretation of a positive effect of grandparents two generations ahead.

Endogeneity emerges as an important concern in the transmission mechanism. We found a significant effect of unobservables in the transmission equations and much lower education transmission coefficients than the censored regression results would suggest. Certainly, the IV results point to an omnipresent measurement problem. Finding however similar effects in a two-generational grandparent to parent setting, where such measurement problems are much less important, seems to confirm the hypothesis of unobserved factors such as, in a hypothetical suggestion innate ability, to be of prime importance. Also, without necessarily interpreting them as causal effects, the IV results help to accumulate evidence that censored regression effects are overestimating the true effects of education on the next generation's educational attainment. The non-parametric model and the parametric model seem to confirm this interpretation. Both approaches are complementary, as the parametric approach allows us to "observe the unobservable" under restrictive assumptions, whereas the non-parametric approach differences out the unobservable, but needs no distributional assumptions.

The importance of unobserved factors suggests from a policy point of view that increasing educational attainment of a cohort might have modest effects on the education of their descendants. It would certainly be interesting to learn more about these important but unobservable influence factors and if or how policy can affect them.

These challenging questions could be a program of future research in this area. The next chapter will explore how the design of the education system might affect educational and labour market outcomes. By providing a new evaluation of the policy of comprehensive schooling, it provides a complementary analysis on how educational outcomes of children are affected by the school system additionally to parental background and education transmission within the family.

	Dependent variable: child's education in years					
	Grandparents assumed exogeneous				Grandparents assumed endogeneous	
	Aggregated education		Mothers/Fathers		Aggr.	Mothers/Fathers
	Cens. reg.	IV	Cens. reg.	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Parents' joint education (HKPARENTS)	.355 (.05)	−.315 (.46)			−.206 (.38)	
Grandparents' joint education (HKGPAR)	.172 (.07)	.419 (.18)			−.504 (.31)	
Mother's education			.361 (.07)	.413 (.81)		.383 (.83)
Father's education			.122 (.04)	−.853 (.40)		−.571 (.35)
Grandmother's education			−.008 (.08)	.188 (.23)		−.459 (.28)
Grandfather's education			.189 (.07)	.288 (.15)		.104 (.55)
Child female dummy, 1=female	.659 (.10)	.642 (.10)	.642 (.13)	.736 (.14)	.661 (.10)	.755 (.15)
# Siblings	−.191 (.15)	−.224 (.16)	−.221 (.20)	−.254 (.20)	−.326 (.16)	−.470 (.24)
Ability (factor)	.265 (.06)	.480 (.16)	.269 (.08)	.586 (.34)	.618 (.17)	.687 (.40)
# Siblings (CM)					−.054 (.03)	−.009 (.04)
CF (aggregated)		.84 (.57)				
Parents' control function					.714 (.49)	
Grandparents' control function					.634 (.28)	
CF Father				1.848 (.75)		1.254 (.64)
CF Mother				−.064 (1.03)		−.003 (1.10)
CF Grandfather						.14952 (.71)
CF Grandmother						.481 (.30)
Constant	13.388 (1.35)	11.814 (1.73)	12.998 (1.85)	9.391 (3.04)	25.905 (5.00)	23.480 (9.23)
N	1711	1711	1075	1075	1680	1015

Table 3.15: The second stage estimation of grandparents' effects assuming both exogeneity (Columns 1-4) and endogeneity (Columns 5-6) for grandparents

	Dependent variable: Child's education		
HKPARENTS	0.367 (.05)	−.222 (.40)	−.037 (.57)
HKGPAR			−.634 (.45)
Number of siblings	−.107 (.15)	−.114 (.16)	−.386 (.25)
Child gender	.645 (.10)	.688 (.11)	.764 (.17)
Ability (factor)	.300 (.06)	.525 (.16)	.694 (.25)
σ_u^2	3.414 (.16)	5.189 (2.44)	10.431 (3.85)
σ_{v1}^2		5.081 (.17)	2.541 (.12)
σ_{v2}^2			.336 (.15)
ρ_1	=0(imposed)	.587 (.26)	.228 (.25)
ρ_2			.336 (.15)
Constant	16.077 (.72)	17.895 (1.46)	27.753 (7.30)
First stages			
Parents			
ROSLA73		.232 (.05)	.231 (.07)
Birth order mother		−.059 (.02)	−.060 (.03)
Grandparents			
ROSLA45F			.192 (.10)
ROSLA45M			.359 (.09)
Grandmothers' birth order			−.046 (.01)

Table 3.16: Results for the full MLE

	Dependent variable: Estimated medians[a]						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HKPARENTS	.088 (.01)	.085 (.01)	.080 (.02)	.076 (.02)	0.061 (0.05)	.059 (.11)	.056 (.34)
# Siblings	.776 (.04)	.763 (.04)	.763 (.05)	.768 (.06)	.764 (.06)	.764 (.13)	.776 (.39)
Child female dummy	.613 (.03)	.605 (.03)	.595 (.03)	.577 (.06)	.624 (.06)	.625 (.04)	.600 (.28)
Ability (factor)	.072 (.01)	.079 (.01)	.086 (.02)	.100 (.02)	.068 (.02)	.068 (.06)	.069 (.15)
First step quantile weighing Kernel δ_n	including all observations						
First step bandwidth δ_n	0.01	0.01	0.01	0.01	0.03	0.03	0.03
Second step bandwidth h_n	0.6	0.2	0.05	0.001	0.6	0.2	0.001
Real (N)	700000	529990	341553	92960	145829	36802	768
effective (n)	1183	1030	826	432	541	271	40

[a]Controls such as cohort dummies, inter-sibling birth spacing and other variables included which also appear in the censored OLS specifications

Table 3.17: Second step estimation of the censored quantile regression model for increasing bandwidths

	Dependent variable: Estimated medians[a] grandparents aggregated		
HKPARENTS	.046 (.15)	.045 (.32)	.029 (.96)
Number of siblings	0.613 (.18)	.651 (.38)	.689 (1.25)
Child female dummy	.586 (.11)	.591 (.24)	.610 (.74)
Ability	.134 (.07)	.132 (.14)	.104 (.45)
HKGPAR	−.038 (.20)	−.045 (.44)	−.074 (1.26)
Sample size			
Real (N)	73616	7507	1166
Effective (n) (for asymptotic results)	429	137	54
Bandwidth			
First step quantile weighing Kernel δ_n	including all observations		
First step non-parametric control function δ_n	0.01	0.01	0.01
2nd step Kernel (h_n)	0.6	0.2	0.05

[a]Controls such as cohort dummies, inter-sibling birth spacing and other variables included

Table 3.18: Second step estimation of the censored quantile regression model, assumimg grandparents endogeneous

	Dep. Variable: Childs' education **OLS**	**Censored regression**
HKPARENTS	.103 (.02)	.434 (.04)
Ability (factor)	.065 (.03)	.238 (.05)
Number of siblings	.258 (.09)	−.429 (.13)
Number of siblings (parents)	−.003 (.01)	−.035 (.02)
Child female dummy, 1=female	.346 (.06)	.620 (.09)
Cohort 1910 Grandfather	.038 (.17)	−.089 (.26)
Cohort 1920 Grandfather	.101 (.18)	.137 (.28)
Cohort 1930 Grandfather	.149 (.22)	.033 (.33)
Cohort 1910 Grandmother	1.022 (.47)	1.568 (.64)
Cohort 1920 Grandmother	1.086 (.47)	1.438 (.64)
Cohort 1930 Grandmother	1.033 (.48)	1.376 (.65)
Constant	15.888 (.48)	16.403 (.48)
R-squared	.08481	0.0382
N	2200	2200

Table 3.19: Aggregated model, estimated with and without considering the censoring of the education distribution

	Dependent Variable: Grandmothers reform dummy[a]	
Grandmother has worked since child at school	.226 (.05)	.124 (.06)
Grandmother smoking prior to pregnancy	−.012 (.05)	.005 (.01)
Substantial help of parents while child at school	−.055 (.05)	−.073 (.06)
Grandmother's weight in stones, 1958	.039 (.007)	.028 (.009)
Birth order-all mother's siblings	−.023 (.01)	−.027 (.02)
Grandmother reads to her child	.015 (.02)	.011 (.04)
Cohort 1910 Grandfather		.699 (.40)
Cohort 1920 Grandfather		2.315 (.39)
Cohort 1930 Grandfather		4.319 (.44)
Constant	−.555 (.12)	−2.296 (.532)
R-squared	0.0367	0.29
N	2501	2197

[a]Regional dummie s, as well as (fathers) cohort effects are included in the second column. Significance levels are $+ \ p < 0.10$, * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 3.20: Checking for covariate balance

Chapter 4

Accounting for Unobservables in Comparing Selective and Comprehensive Schooling

Today, despite the fact that 116 Labour MP's benefited from a grammar school education, Labour remains ideologically blinkered about the contribution grammar schools make to our education system.... Under the Conservatives, grammar schools will survive and thrive.

Michael Howard MP; leader of the Conservative Party, 2004

If it's the last thing I am going to do, I'm going to destroy every [...] grammar school in England. And Wales. And Northern Ireland.

Anthony Crosland, 1965, then Education Secretary

4.1 Introduction

In this chapter, we compare the effects of selective and non-selective education on various educational and labor market outcomes, using data from the National Child Development Study (NCDS). The recent history of British secondary education has been characterized by the shift from a selective towards a non-selective or "comprehensive" system. After the Second World War, and starting with the Butler Education act (1945) the education system in the U.K. was mostly selective. Namely, children were assigned to two different types of schools depending on their results to a test score at age 11, called the *11-plus* exam. Successful children went to grammar schools, while the others attended less demanding secondary modern schools. Then, an important change was initiated in 1965 by the Crosland Circular, by which a possibility of comprehensivisation was introduced, creating schools where children of different ability levels were pooled together. As the circular did not force schools to be comprehensive, the timing of the shift towards comprehensive schools has been heterogeneous. Still in 1965, less than 5% of all public schools were comprehensive. By 1975 this proportion had reached 60%. In the period when the children of the NCDS, all born in 1958, were at school, the two systems coexisted. This makes these data a potential laboratory for comparing the two systems.

There have been several recent attempts at measuring the effect of pupils' selection on outcomes using the NCDS data. In an early study, Kerckhoff (1986) measures the effects of school type on later educational outcomes. He finds that attending a grammar or secondary modern school is associated with higher and lower educational attainment, respectively, than attending a comprehensive school. Dearden et al. (2002) define "selective schools" as either grammar or private schools. Using propensity score matching techniques to control for a large set of covariates, they find that attending either of the two types is associated with better outcomes. More recently, Galindo-Rueda and Vignoles (2005) focus on the differences between school systems: selective (either grammar or secondary modern) and comprehensive. They find that attending a selective school yields better educational outcomes. Moreover, the gain is found to be concentrated on high-ability pupils.

In an important methodological criticism to these findings, Manning and Pischke (2006) question the exogeneity of the comprehensivisation reform. They measure the effect of attending a selective school on test scores at age 11, that is before entering a secondary school. They find positive effects, of similar magnitude or larger than the effects on subsequent outcomes. They interpret this exercise as a falsification test, which suggests that attending a selective school is likely to be affected by unobservables that in turn affect later outcomes. They conclude that "we don't know very much about the effects of comprehensive schooling in Britain" (p.19).

The point raised by Manning and Pischke (2006) is the main motivation of this chapter. We start by focusing on the differences between the selective and comprehensive schooling systems. Attending a comprehensive school is the determinant, or "treatment", of interest. Given the non exogeneity of the treatment, consistent estimation of treatment effects on the mean and variance of outcomes is a difficult task. Our strategy is inspired from recent advances in the education production function literature. We build a model, where test scores depend on parental and school inputs, as well as on the child's endowment, as in Todd and Wolpin (2003, 2004). The endowment is unobserved to the econometrician, and can be multidimensional. We embed this model into Rubin's (1974) framework, where we do not simultaneously observe the two potential post-treatment outcomes. We assume that the same unobserved endowment affects both pre- and post-treatment outcomes as well as the probability of attending a selective or comprehensive school.

In the presence of selection-on-unobservables, usual matching estimators are biased. To correct the matching estimand for the bias created by the presence of the endowment, we develop a quasi-differencing approach inspired from the "within" approach to eliminate fixed effects in linear panel data models (*e.g.* Holtz-Eakin et al. (1998)). For this purpose, we make use of the availability of several test scores at age 11 (maths, reading, verbal), which are also affected by the endowment. Then, estimating the Average Treatment Effect (ATE) in our model is simple and transparent. Under the main assumption that pre-treatment outcomes are not affected by the treatment other than through the child's endowment, we obtain that the bias of the matching estimand on post-treatment outcomes is proportional to the difference in pre-treatment test scores between the two schooling systems. Moreover, the coefficient of proportionality can be estimated by using other test scores (such as the reading score at age 11) as alternative measures of the endowment.

Following this strategy one is able to consistently estimate the ATE, as well as other treatment effects (AT for the Treated, ATE and ATT on variances). It is to be noted that we do not assume anything on the correlation between the endowment and observed covariates, such as parental inputs. This fixed effects approach is especially appealing in the education production function perspective, where parents take decisions based on their child's ability. Moreover, the model implies testable restrictions on the data. This property is important in order to check if the assumed structure is correct.

Our identification strategy is closely linked to a class of models introduced by James Heckman and coauthors. Starting with Carneiro et al. (2003) and Hansen et al. (2004), these authors use factor models to restrict the correlation between measurements and achieve identification. Compared to the recent models in this literature (*e.g.* Cunha and Heckman (2006), and Cunha et al. (2006)) our framework is more restrictive as the identifying content of the model is limited. In particular, we are only able to identify effects of the treatment on the mean and the variance, as we make no distributional assumptions (such as independence) on the errors. Another restriction compared to recent advances in that literature is the linear way the endowment enters the model. One virtue of our approach is that it does not restrict the correlation between the endowment and the covariates, and that it yields easily testable implications.

Applying our methodology to the NCDS data, we find evidence of a bi-dimensional endowment. We find that, correcting for the differences in endowment at age 11, attending a comprehensive rather than a selective school has a negative effect on means of outcomes, but small and insignificant in many cases. Stronger are the results we obtain for variances, as the selective system is found to present significantly more dispersion than the comprehensive one. We then focus on the variance within the selective system. We argue that one important part of this variance comes from the differences between achievements at grammar and secondary modern schools. In a tentative attempt to measure the premium associated to the fact of passing the *11-plus* exam, we find strong and often significant effects. We interpret these results as a potential explanation of risk-averse parents being likely to have pushed towards comprehensivisation.

The effects of pupils' selection and "ability tracking" on outcomes are currently attracting interest ouside the U.K. too. In the U.S., there is already a large literature on the subject, with mixed evidence (see Figlio and Page (2002), and references therein). Comparable to the British case, one of the main features of the Swedish reform that took place in the 1950's was to abolish selection by ability into academic and non-academic streams. Meghir and Palme (2005) exploit the fact that the reform was preceded by a social experiment to estimate its effects on educational and labor market outcomes. They find positive effects, concentrated on children with lower parental background. Using country-level data and a difference-in-difference methodology, Woessmann and Hanushek (2006) find little evidence of positive effects of ability tracking on mean outcomes. However, they also find that the variance of educational attainment is larger in countries with selective education (but see also Waldinger (2006)). Maurin and McNally (2006) study the effects of a shift in the admission criteria in grammar schools in Northern Ireland. They find that the probability of attending a grammar school and of achieving higher educational qualifications both strongly increased after the reform.

The outline of the chapter is as follows. In Section 4.2 we present the NCDS data and compute some preliminary estimates of the effect of selection on various outcomes. We also present some evidence that the treatment consisting in attending a comprehensive school is not exogenous, building on Manning and Pischke (2006). We present our model in Section 4.3 and list and motivate the assumptions we make. Section 4.4 is devoted to the identification of causal parameters, starting with the ATE. In that section we also provide estimators of the parameters of interest, and derive some testable implications of the model. In Section 4.5 we apply our methodology to the NCDS data, and compare and contrast the two schooling systems. In Section 4.6 we focus on the differences that exist within the selective system, between school types. Lastly, Section 4.7 concludes.

4.2 A first look at the data

The NCDS is an ongoing longitudinal survey of a British birth cohort born between March 3 and March 9 of 1958. The initial sample consisted of 17,634 individuals which were resurveyed on six further occasions in order to monitor their changing health, education, social and economic circumstances. Attrition has led the sample size to shrink in the subsequent waves:[1] in 1965 at age 7, 1969 at age 11, 1974 at age 16, 1981 at age 23, 1991 at age 33, and 1999/2000 when the cohort had reached the age of 42. We will be looking at schooling achievements of the NCDS cohort members and will use information from all five waves. The NCDS children received secondary schooling between 1969 and 1974, and thus lived through the change of the British education system.

Sample selection. We are interested in the relative effects of selective and comprehensive schooling on outcomes. We use a NCDS variable that gives the type of the school attended by the child at age 16, when finishing secondary education. We do not use observations for which this variable is missing. We also exclude other types of schools, such as technical or *public* (*i.e.* private) schools. These schools are likely to be

[1]There has been a large amount of attrition in the NCDS. Current practice has treated attrition as exogenous. Conolly *et al.* (1992) show that attrition is somewhat stronger among children from lower parental background. Dearden et al. (2002) argue that the results are not necessarily biased, as they dispose of many indicators of parental characteristics, as we also do.

very different from the ones we consider. Moreover, they represent a small percentage of all schools (less than 10%). Note that the school type variable also indicates which kind of selective school, grammar or secondary modern, the child was attending at age 16. We will use this information in Section 4.6.

For the purpose of comparing children that went through fully selective or fully comprehensive secondary education, we keep children who stayed during five years in the same schooling system. This avoids re-organisation effects in our data, as we would expect there to be a large re-organisation effect once a school changes type, say from grammar to comprehensive, within the five year period we observe. To select the sample, we use a variable which indicates the time stayed in the same school. We select only pupils that stayed for five years in the same school. We obtain a sample of 6870 observations in the sample, 3865 (56%) of which are comprehensive. We also use other samples for the purpose of replicating samples in the literature and also to compare pupils in grammar and comprehensive schools. One sample takes only purely selective and purely comprehensive schools, that is schools in local education areas which have either fully introduced comprehensive schools and areas which are still entirely in the previous, selective system. Another sample takes only pupils in grammar and secondary modern schools in order to compare differences between these schools. See the data Appendix for details about sample construction and all the samples used.

Variables. We are interested in the effect of the schooling system on several educational outcomes. We shall consider one outcome at age 16: the maths score given by the NCDS staff. It is important to notice that, unlike the math test, the reading test was the same at age 11 and age 16. For this reason, the distribution of the reading score at 16 is concentrated at high values and we do not take it as an outcome. In addition, we will look at later outcomes: age when the child left school, vocational and academic qualifications obtained by the child, as well as wages.

We will also use indicators of educational level before secondary schooling. In particular, we will use the results of NCDS test scores at 7 and 11, before the *11-plus* examination took place. Maths and reading tests were administered at ages 7 and 11, and an additional verbal test was administered at age 11 only.

We will also use parental background, such as the social class and education levels of the parents measured when the child was 7 or 11 years old, and school characteristics, such as class size variables, at age 7, 11 and 16. The complete list of these variables is given in the Appendix. Finally, we complement the data by constructing a set of local controls. For this purpose, we merge the 1971 Census statistics to the NCDS data. We then obtain characteristics of the enumeration district in which the child lives. In addition, we construct a political variables set. As the education reform was first implemented on a voluntary basis, political control of the Local Authority, e.g. City or County Borough Council seems an important variable to explain the shift towards comprehensive schooling. We aggregated the results of the 1970 general election for each constituency. Furthermore, we used local authority level information on school resources to complement our analysis. See the Appendix of this chapter for a more precise description of all data sets involved.

Descriptive statistics. Table 4.1 shows some descriptive statistics for the two groups of children in the sample, attending a comprehensive or a selective system. Educational outcomes are higher for children attending selective schools, who score on average 1.9 points higher in mathematics and 1.3 points in reading than children at comprehensive schools, about 25% of one standard deviation in each case. Moreover, they leave school on average 0.4 years later, and obtain more academic and vocational qualifications. There is a small difference in weekly wages. The two groups are also very different in terms of intake, as children at selective schools score better at all tests at age 7 and 11. For instance, they score 3 points higher in mathematics at age 11, 30% of one standard deviation. We find similar discrepancies for the dispersion of test scores, selective schools showing higher variance of pre-treatment variables at age 11, and post-treatment outcomes except the reading score at 16.

Another way to illustrate the differences pre-secondary education is to summarize the information contained in the test scores previous to the *11-plus* examination into one scalar indicator. It is common practice to compute the first principal component of

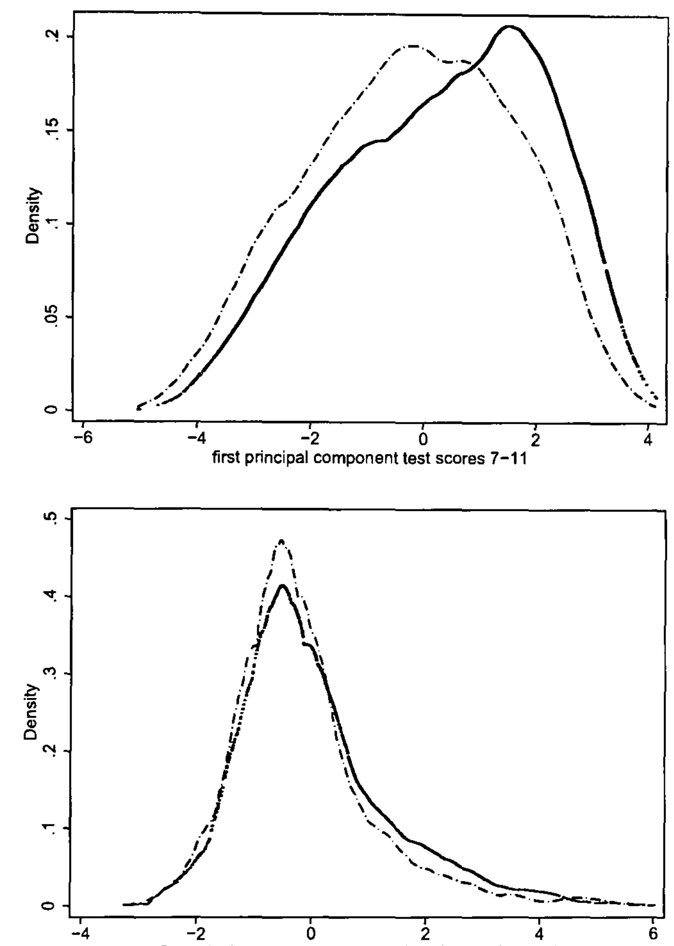


Figure 4.1: The "ability" and "parental background" principal components for pupils attending comprehensive (dash-dotted line) and selective schools (solid line).

Variable	Comprehensive Mean	Std.Dev.	N	Selective Mean	Std. Dev.	N
Maths score (Age 16)	11.921	6.294	3600	13.826	7.135	2872
Reading score (Age 16)	25.164	6.603	3614	26.471	6.382	2882
Maths score (Age 11)	15.693	9.65	3434	18.673	10.371	2636
Reading score (Age 11)	15.672	5.912	3434	17.119	6.045	2636
Verbal score (Age 11)	21.596	8.996	3436	24.246	9.076	2636
BSAG score (Age 11)	8.152	8.692	3439	7.325	8.015	2636
Maths score (Age 7)	5.091	2.389	3431	5.443	2.442	2701
Reading score (Age 7)	23.091	6.890	3437	24.385	6.359	2717
Age left full time education	17.136	2.085	2701	17.513	2.403	2165
Qualifications 2004	2.246	1.35	3168	2.494	1.361	2513
Academic qualifications (2004)	1.781	1.304	2059	2.043	1.384	2507
Vocational qualifications (2004)	1.511	1.497	3063	1.642	1.556	2437
1981 log weekly net wage	4.208	0.369	2059	4.213	0.347	1653
1991 log weekly net wage	4.409	0.704	1682	4.445	0.742	1354
2000 log weekly net wage	4.466	0.793	1928	4.513	0.78	1556
2004 log weekly net wage	4.741	0.65	1708	4.784	0.715	1364

Table 4.1: Comparing outcomes in selective and comprehensive schools; sample 1.

the 7 and 11 test scores and to label it as "ability" (*e.g.* Galindo-Rueda and Vignoles (2005)). Drawing the densities of the "ability" variable for the two groups shows strong differences, as the top panel of Figure 4.1 illustrates. On the bottom panel, we draw the density of a "parental background" variable, similarly computed as the first principal component of the many NCDS variables related to the social class of parents or their education. Figure 4.1 shows that there are less marked but still noticeable differences between the two groups, children in the selective system coming from better background on average.

Raw effects. Next, we regress the maths score at age 16 on the indicator of comprehensive schooling, with various controls. Table 4.2 presents the results. The unconditional effect reported in column (1) is negative. In columns (2)-(5), we adopt a value-added specification and use as controls lagged test scores, at age 11 and 7. Such specifications have been widely used in the education production function literature. The results show a consistently negative effect of attending a comprehensive school on the math score. The introduction of lagged test scores as controls lowers the effect measured by the coefficient of the comprehensive reform dummy, C, by a large amount. Still, the effect, although not especially large (about 6% of one standard deviation) is significant at the 5% level in all the specifications. Lastly, in column (6), we replace the lagged test scores by their first principal component (the "ability" variable). Columns (5) and (6), comparable in terms of other included covariates, are very similar, with a slightly lower R-squared in column (6).

	Outcome: Mathematics score at 16					
	(1)	(2)	(3)	(4)	(5)	(6)
C (in comprehensive school)	−1.904 (.167)	−1.540 (.155)	−1.8119 (.205)	−1.660 (.185)	−.464 (.121)	−.556 (.133)
Child female		−1.327 (.154)	−1.328 (.154)	−1.266 (.192)	−1.125 (.112)	−1.652 (.127)
Father's education		.354 (.058)	.352 (.058)	.288 (.073)	0.135 (0.043)	0.175 (0.050)
Mother's education		.619 (.072)	.622 (.072)	.602 (.091)	0.244 (.054)	0.251 (.061)
Maths score 11					.473 (.006)	
Ability (factor)						2.492 (.037)
Constant	13.825 (.124)	−11.090 (61.146)	−11.09 (61.16)	6.760 (1.328)	3.028 (.813)	11.539 (.91870)
R-squared	.019	.1626	.185	.198	.597	.559
N	6472	6472	6472	4279	5738	5226
Controls						
Individual	no	yes	yes	yes	yes	yes
Parental background	no	yes	yes	yes	yes	yes
School and local	no	no	yes	yes	yes	no
LEA dummies	no	no	no	yes	no	no
Ability	no	no	no	no	no	yes

Table 4.2: The effect of attending a comprehensive school on the math score at 16, conditional on increasing sets of covariates.

We present propensity-score matching estimates in Table 4.3. We use the inverse probability weighting of Hirano et al. (2003) to compute the effects of comprehensivisation on the various outcomes. Details on the estimation will be given in the next sections. In the first row of the table, we use as controls the parental, school and local covariates used in column (6) of Table 4.2. We find negative and significant effects of attending a comprehensive school on all outcomes. These effects remain negative when we include the "ability" factor as a covariate, but they turn mostly insignificant.

	Years educ. educ. (1)	**Academic qual.** (2)	**Vocational qual.** (3)	**Weekly wage 2004** (4)	**Average of weekly wages**[a] (5)
Controls					
(3) [b]	−.292	−.197	−.099	−.053	−.036
	(.061)	(.034)	(.042)	(.022)	(.023)
(3) + LEA dummies	−.397	−.237	−.188	−.077	−.071
	(.078)	(.044)	(.054)	(.028)	(.030)
(3) + Ability	−.131	−.048	.006	−.010	−0.003
	(.064)	(.034)	(.047)	(.023)	(.025)

Table 4.3: The effect of attending a comprehensive school on other outcomes, using the same specification (3) of Table 4.2 and the indicated additional variables

[a]Simple average over wages in 1991, 2000 and 2004
[b]Parental school and local controls used

To summarize, results conditional on school, parental and local covariates suggest that attending a comprehensive school is associated with lower educational achievement. Controlling for lagged test scores, or alternatively for their first principal component, yields also negative estimates, though insignificant except for the math score at 16. However, the "value-added" strategy, which consists in directly controlling for lagged test scores in the outcome equations, is not without problems. Todd and Wolpin (2003) show that value-added specifications impose strong restrictions on the parameters of the production function of education outcomes. In a more general model, later test scores do not depend directly on lagged ones, but on the child's endowment that is unobserved to the econometrician. In this perspective, the test scores used in the value-added specification can be seen as informal proxies for the child's endowment at age 11. If the restrictions that this specification imposes are not satisfied, it is not clear what we are estimating by controlling for these variables as we did in Tables 4.2 and 4.3.

"Falsification test". Moreover, even within the logic of value-added specifications, there are reasons not to believe the above estimates of the effect of attending a comprehensive school. To see why, we follow Manning and Pischke (2006), and estimate the effect of attending a comprehensive school between age 11 and 16 on outcomes *prior to* entering secondary education. Table 4.4 shows the results of the regressions of the test scores at age 11, controlling for an increasing number of covariates.[2] We find strong effects, indeed even stronger than on outcomes at age 16. For instance, the effect on the math score is never lower than 1.4 in absolute value, that is 14% of one standard deviation. Estimates obtained using propensity score matching give a similar picture.

	Outcome: Mathematics score at 11					
	(1)	(2)	(3)	(4)	(5)	(6)
C (in comprehensive school)	−2.979 (.258)	−2.403 (.244)	−2.387 (.250)	−2.785 (.321)	−1.801 (.225)	−1.461 (.208)
Child female		−.667 (.240)	−.630 (.240)	.437 (.300)	.202 (.216)	−1.006 (.203)
Father's education		.521 (.092)	.509 (.091)	.761 (.114)	.457 (.085)	.412 (.079)
Mother's education		.902 (.115)	.772 (.114)	.761 (.143)	0.574 (.105)	0.454 (.097)
Maths score 7					2.014 (.045)	1.324 (.047)
Reading score at 7						.552 (.017)
Constant	.194 (18.672)	9.401 (1.034)	5.487 (1.674)	6.342 (2.127)	−2.765 (1.536)	−9.215 (1.432)
R-squared	.021	.14	.174	.188	.392	.484
N	6070	6070	6070	3936	5537	5521
Controls						
Individual	no	yes	yes	yes	yes	yes
Parental background	no	yes	yes	yes	yes	no
School and local	no	no	yes	yes	yes	yes
LEA dummies	no	no	no	yes	no	no
Math7	no	no	no	no	yes	yes
Read7	no	no	no	no	no	yes

Table 4.4: The effect of attending a comprehensive school on the math score at 11, conditional on increasing sets of covariates.

This evidence casts serious doubts on the exogeneity of the school system variable in outcome equations. Of course, this "falsification test" needs very special conditions in order to be a proper test of exogeneity. For the effect on age 11 test scores to be zero, one would in particular need a stationarity condition to be satisfied, Imbens

[2]Column (1) corresponds to the case without controls. Then we add: (2) the maths score at 7; (3) the reading score at 7; (4) individual and family characteristics until 1969; and (5) regional controls.

(2004), requiring that the effect of age 7 test scores on scores at age 11 be the same as the effect of age 11 test scores on subsequent outcomes. In our data this condition is not satisfied, as is clear from the difference in R^2 between Tables 4.2 and 4.4. In any case, this results point at the presence of unobserved variables that cannot be properly controlled for by the many covariates present in the NCDS data.

Alternative strategies. In this paper, we explicitly address the concern, raised by Table 4.4, that attending a selective or comprehensive school may be endogenous in test score equations. Our approach is motivated by the failure of standard methods to provide credible inference on the parameters of interest. A natural approach would be to find an instrument for the schooling system. Galindo-Rueda and Vignoles (2005) propose to use the political colour of the LEA for this purpose. Indeed, in the period schools in Labour dominated LEA's are more predominantly comprehensive than conservative ones. Although a natural idea, using political colour as an instrument fails to remove the bias as Manning and Pischke (2006) show. This is because Labour dominated LEA's have on average children of lower ability and parental background. Using the IV strategy in the age 11 test scores equations yields a large effect of attending a selective school (see Table C.7 in Appendix). Another approach would be to derive bounds for the effect of comprehensive schooling, as the effect obtained by OLS or matching can be interpreted as an upper bound (in absolute value). To find a lower bound, we would need to find a variable that influences positively the probability of attending a selective school, yet is negatively correlated with the unobservables in the outcome equations. One example of this strategy is the use of the affirmative action legislation in Rothstein and Yoon (2006). In our case, it is not clear that such a variable exists.

Lastly, one might want to check the sensitivity of the results to departures from exogeneity. Commonly used techniques assume the existence of an unobserved "confounding" regressor, conditional on the treatment, here the type of school attended, being randomly assigned. See *e.g.* Rosenbaum and Rubin (1983), Imbens (2003), and recently Ichino et al. (2006). In this approach, one checks the sensitivity of treatment effect estimates under various assumptions on the unobserved confounder. One weakness of the approach is that it is not always clear how to calibrate the magnitude of the sensitivity check. Altonji et al. (2005) propose to use *observed* covariates as a guide to calibrate the sensitivity parameters. However, in our case, the central role played by the child's ability, that is unobserved to the analyst, makes this method difficult to apply.[3]

Our strategy is close in spirit to the sensitivity approach. We also assume that there exists an unobserved confounder. In addition, we use information on the test scores prior to secondary education to remove the bias created by the presence of the confounder on the treatment effect estimate. The structure we assume has the advantage of being in line with the recent literature on the education production function. We now formally present our approach.

4.3 A model accounting for selection on unobservables

We start with the Rubin (1974) framework. Attending a comprehensive school is the treatment the effects of which we want to study. We denote by $T = 0$ the fact of attending a selective schooling system, while the comprehensive alternative is denoted by $T = 1$. We consider an outcome that has two potential values, in the presence (Y_1) and the absence (Y_0) of treatment. To fix ideas, we think of the outcome as the math test score at age 16. We shall also use other outcomes, namely the age when left school, the educational qualifications obtained and weekly wages. The two potential outcomes are not simultaneously observed by the econometrician, only the combination $Y = TY_1 + (1 - T)Y_0$.

In addition, we consider two outcome variables, W and Z, that we think of as pre-treatment outcomes. For instance, W is the math score at age 11 and Z is another test score (*e.g.* reading or verbal) at the same age. These variables are not causally affected by the treatment and are always observed.

[3]If one does not include lagged test scores in the set of covariates, then the estimates seem very robust. But if we include them, then they become close to zero. One source of the problem could be that the child's ability is a very special regressor, unfit for the framework in Altonji et al. (2005), where observed and unobserved variables are assumed to be randomly drawn from the same pool of covariates.

Our modelling of the four outcomes variables reads as follows:

$$\begin{cases} Y_0 &= f_0(X) + \alpha_0\eta + \varepsilon_0, \\ Y_1 &= f_1(X) + \alpha_1\eta + \varepsilon_1, \\ W &= g(X) + \beta\eta + U, \\ Z &= h(X) + \gamma\eta + V. \end{cases} \quad (4.1)$$

In this model, X are covariates that include children's characteristics as well as parental and school inputs. Their effect on the outcomes is modelled by the unknown functions f_0, f_1, g and h. Then, η is an unobserved variable that represents the child's endowment at the time the treatment is received, that is at age 11. The endowment, that includes child and school unobserved characteristics, has a distinct effect on the various outcomes, through the parameters α_0, α_1, β and γ.

We make the following assumptions.

Assumption 1 *(selection on unobservables)*

$$\mathbb{E}\left(\varepsilon_0|X,T\right) = \mathbb{E}\left(\varepsilon_0|X\right); \quad \mathbb{E}\left(\varepsilon_1|X,T\right) = \mathbb{E}\left(\varepsilon_1|X\right).$$

Assumption 1 is reminiscent of the "selection-on-observables" assumption often used in policy evaluation, where the treatment is independent of the potential outcomes conditional on observed covariates:

$$(Y_0, Y_1) \perp\!\!\!\perp T| \quad X, \quad (4.2)$$

where ". $\perp\!\!\!\perp$.| ." represents conditional independence. A weaker assumption requires that treatment and outcomes be independent conditional on observed and unobserved covariates:

$$(Y_0, Y_1) \perp\!\!\!\perp T| \quad X, \eta. \quad (4.3)$$

Assumption (4.3) is typically made when checking the sensitivity of treatment effects estimates to departure from exogeneity, as in Imbens (2003). In that approach, one calibrates the contribution of η to the model and check by how much the estimates vary.

Assumption 1 is slightly weaker than (4.3). In particular, mean independence is assumed instead of full independence.

Assumption 2 *(pre-treatment outcome)*

$$\mathbb{E}(U|X,T) = \mathbb{E}(U|X).$$

Assumption 2 requires that W be not affected by the treatment. Hence differences between $\mathbb{E}(W|T=1,X)$ and $\mathbb{E}(W|T=0,X)$ simply reflect composition effects, as the composition in terms of η is different in the treated and non treated groups. We have in mind pre-treatment variables, that are often used in order to test for the selection-on-observables assumption (4.2).

Assumption 3 *(additivity)*

$$\eta \perp (\varepsilon_0, \varepsilon_1, U, V) \mid \quad X, T,$$

Assumption 3 emphasizes the additive structure of the model. In the assumption, the notation ". $\perp$.| ." stands for conditional uncorrelatedness. A sufficient condition for this assumption to hold is:

$$\mathbb{E}(\varepsilon_0|X,\eta,T) = \mathbb{E}(\varepsilon_1|X,\eta,T) = \mathbb{E}(U|X,\eta,T) = \mathbb{E}(V|X,\eta,T) = 0.$$

Note that these assumptions do not assume any structure for the correlation between η and X. η is analogous to a "fixed" effect in a linear panel data model of the form:

$$y_{it} = x_{it}'\theta + \eta_i + u_{it},$$

where $\mathrm{Cov}(\eta_i, u_{it}) = 0$ for all t, but $\mathrm{Cov}(\eta_i, x_{it})$ is not restricted.

Assumption 4 *(instrument)*

$$V \perp (\varepsilon_0, \varepsilon_1, U) \mid \quad X, T,$$

Assumption 4 requires that Z be an "instrument" in the following sense. We assume that Z is uncorrelated with the residuals of the other outcome variables, conditional on observed and unobserved variables. Of course, the presence of the unobserved η creates a correlation between the variable Z and the potential outcomes Y_0 and Y_1. Hence, Z cannot be used directly, as a proper instrumental variable.

Assumption 5 *(rank conditions)*

$$\beta \neq 0; \qquad \gamma \neq 0; \qquad \mathrm{Var}\,(\eta|X,T) \neq 0.$$

Lastly, Assumption 5 requires that η presents variation conditional on X and T and has an effect on the pre-treatment outcome and the instrument. It means that W and Z have an unobservable in common.

As presented here, the model allows for a one-dimensional endowment. It can immediately be extended in order to allow for a multivariate endowment $\eta \in \mathbb{R}^K$, and vectors of pre-treatment outcomes and instruments W and Z. The model is presented in this general version in the Appendix. We require the same assumptions, except that Assumption 5 is replaced by the condition that β, γ and $\mathrm{Var}\,(\eta|X,T)$, that are now matrices, have all rank K, the number of dimensions of the endowment. In particular, we need at least K pre-treatment outcomes and K instruments for this assumption to be satisfied. As is intuitive, data requirements increase if we want to account for a multivariate structure of unobservables.

The model is in line with recent models of the education production function literature, as in Todd and Wolpin (2004, 2003). In that literature, test scores are modelled as combinations of present and past parental and school inputs, and of the child's endowment. We allow that parental and school decisions be based on the endowment, as the correlation between η and X is left unrestricted. Moreover, the residuals of the test score equations, ε_0 and ε_1, can include unobserved parental and school inputs, as they can also be correlated with X.

In the model, the child's endowment at age 11, which may include individual and school effects, can be correlated to attending a selective system. The endowment η is the only variable that correlates the treatment with the outcomes. In particular the residuals $(U, \varepsilon_0, \varepsilon_1)$ are mean independent of the treatment. In contrast, they can be freely correlated with each other. This correlation can represent a special ability that the child posseses for maths, compared to other subjects. This special ability is imposed not to influence the treatment, the endogeneity of the latter coming only from the endowment.

As emphasized by Todd and Wolpin (2003), it is important that η be possibly correlated to the parental and school variables. Two common strategies to cancel out the child's endowment are the use of data on siblings, and the use of time series of test scores. In both cases, a "within" estimator provides consistent estimates of the model's parameters. We do not have information on siblings, and the time series we have are rather short (age 7, 11 and 16). Moreover, it is not clear how to adopt a "within" approach in the presence of a treatment that can be correlated to the endowment. Lastly, approaches such as the ones advocated by Todd and Wolpin (2003) are not applicable if the endowment is multidimensional.

4.4 Identification of policy parameters

In this section, we show how to identify and estimate several treatment effects of interest if Assumptions 1-5 hold. The discussion is conducted in the case of the model in its simplest form, where the endowment η is scalar. A treatment of the multidimensional endowment case is provided in the Appendix, as well as a proof of all statements appearing in this section. Let us start with the Average Treatment Effect (ATE) parameter: $\Delta = \mathbb{E}(Y_1 - Y_0)$. We have:

$$\Delta = \mathbb{E}\left[\mathbb{E}(Y_1|X) - \mathbb{E}(Y_0|X)\right].$$

In contrast with the case where there is no selection on unobservables (e.g.Rosenbaum and Rubin (1983)), this quantity has no direct counterpart in the data. Formally, for $j = 0, 1$ we have:

$$\mathbb{E}(Y_j|X) \neq \mathbb{E}(Y_j|X, T = j).$$

To illustrate how our method works, consider the case where $\alpha_j = \beta = 1$. Then it follows immediately from Assumptions 1 and 2 that:

$$\mathbb{E}(Y_j - W|X) = \mathbb{E}(Y_j - W|X, T = j).$$

The selection on the differenced outcomes is only driven by the covariates X. We can thus estimate the effect of the treatment on $Y - W$, as the endowment η has been differenced out. This strategy, inspired from the "within" estimation of linear panel data models with fixed effects, is used in studies where data on siblings is available Altonji and Dunn (1996) .

In our case, we have several test scores taken at different years, but the individual time series is rather short (three dates), and we see no *a priori* reason to assume that the weight of the ability is the same in every of them. We propose a generalization of the "within" approach, which is based on the following property, that is also immediately based on Assumptions 1 and 2.

$$\mathbb{E}\left(Y_j - \frac{\alpha_j}{\beta} W|X\right) = \mathbb{E}\left(Y_j - \frac{\alpha_j}{\beta} W|X, T = j\right). \tag{4.4}$$

A first interpretation of (4.4) is obtained by remarking that we have, equivalently:

$$\mathbb{E}\left(Y_j|X, T = j\right) - \mathbb{E}\left(Y_j|X\right) = \frac{\alpha_j}{\beta}\left[\mathbb{E}\left(W|X, T = j\right) - \mathbb{E}\left(W|X\right)\right]. \tag{4.5}$$

The term on the left-hand side of (4.5) can be interpreted as one of the two parts of the bias of the matching estimand, for given covariates X. If the selection-on-observables assumption holds, then this bias term is zero and the ATE parameter is equal to the matching estimand. If not, then the bias is proportional to the difference in *pre*-treatment outcomes W between the treated and non-treated. Hence, under the model's assumptions, "falsification tests" à la Manning and Pischke (2006) are useful, as they indicate if matching on X will provide a consistent estimate of the ATE parameter.

A second interpretation of (4.4) is related to the panel data literature. Indeed, our approach can be seen as a "quasi-differencing" approach, close to the method used by Holtz-Eakin et al. (1998) to estimate panel data models with interactive fixed effects of the form:

$$y_{it} = x_{it}'\theta + \alpha_t \eta_i + u_{it},$$

where α_t are time-varying parameters.

Our approach requires the identification of the ratios α_j/β, for $j = 0, 1$. The necessary information for that purpose is provided by the availability of other pre-treatment outcomes Z ("instruments"). We show in the Appendix that, if the model's assumptions hold then this ratio is given by:

$$\frac{\alpha_j}{\beta} = \frac{\mathbb{E}\left[\mathrm{Cov}\left(Z, Y|X, T = j\right)|T = j\right]}{\mathbb{E}\left[\mathrm{Cov}\left(Z, W|X, T = j\right)|T = j\right]}, \tag{4.6}$$

$$= \frac{\mathrm{Cov}\left[Z - \mathbb{E}\left(Z|X, T = j\right), Y - \mathbb{E}\left(Y|X, T = j\right)|T = j\right]}{\mathrm{Cov}\left[Z - \mathbb{E}\left(Z|X, T = j\right), W - \mathbb{E}\left(W|X, T = j\right)|T = j\right]}. \tag{4.7}$$

It follows from equation (4.6), or alternatively (4.7), that the weight of the endowment in post-treatment relative to pre-treatment outcomes can be consistently estimated from the data. It can be interpreted as the IV estimand in the quasi-differenced equation. A first implication of this result is that the ratio α_1/α_0 can also be estimated from the data. This is especially interesting in the case of selective schooling, as the literature has shown interest in the different returns to the child's ability in the two education systems.

A second implication of the identification of α_j/β concerns the estimation of policy parameters. Indeed, from (4.4) the ATE on $Y - \frac{\alpha_j}{\beta}W$ can be estimated by matching. It follows that the ATE on Y is also identified, as:

$$\Delta = \mathbb{E}\left[\mathbb{E}\left(Y - \frac{\alpha_1}{\beta}W|X, T=1\right) - \mathbb{E}\left(Y - \frac{\alpha_0}{\beta}W|X, T=0\right)\right] + \frac{\alpha_1 - \alpha_0}{\beta}\mathbb{E}(W). \tag{4.8}$$

See the Appendix for a straightforward derivation of (4.8).

Extending this approach, we can also identify the Average Treatment Effect for the Treated (ATT): $\Delta^{TT} = \mathbb{E}(Y_1 - Y_0|T=1)$. We show in the Appendix that the following identity holds:

$$\Delta^{TT} = \mathbb{E}[Y|T=1] - \mathbb{E}\left[\mathbb{E}\left(Y - \frac{\alpha_0}{\beta}W|X, T=0\right)|T=1\right] - \frac{\alpha_0}{\beta}\mathbb{E}(W|T=1). \tag{4.9}$$

In Section 4.6 we will be interested in the AT on the Non Treated (ATNT), that is obtained in a similar way.

A last parameter we are interested in is the ATE on variances:

$$\Delta^V = \mathrm{Var}(Y_1) - \mathrm{Var}(Y_0).$$

Comparing the variances of (potential) outcomes in the two schooling systems is interesting, as it reflects within-system dispersion and inequality in educational performance. In order to achieve identification we augment Assumptions 1 and 2 with the two following assumptions.

Assumption 6 *(selection on unobservables, variances)*

$$\mathrm{Var}(\varepsilon_0|T=1, X) = \mathrm{Var}(\varepsilon_0|X); \quad \mathrm{Var}(\varepsilon_1|T=1, X) = \mathrm{Var}(\varepsilon_1|X).$$

Assumption 7 *(pre-treatment outcomes, variances)*

$$\text{Var}\,(U|T=1,X) = \text{Var}\,(U|X)\,.$$

Assumptions 6-7 rule out effects of the treatment on the variances of pre- and post-treatment outcomes. Under Assumptions 1-7 we show in Appendix that

$$\text{Var}(Y|T=j,X) - \text{Var}(Y_j|X) \;=\; \left(\frac{\alpha_1}{\beta}\right)^2 \left[\text{Var}(W|T=j,X) - \text{Var}(W|X)\right]. \tag{4.10}$$

Here also, as in equation (4.5), the bias of the matching estimand (on the variance) of the post-treatment outcome is proportional to the difference in pre-treatment variable between the two groups. We thus obtain the within-X variance. The between-X variance is easily obtained using the results for the ATE case. Likewise we can also derive the expression of the ATT on variances $\Delta^V = \text{Var}(Y_1|T=1) - \text{Var}(Y_0|T=1)$. See the Appendix for details.

In this paper, we shall focus on the above parameters: ATE and ATT on means and variances. One could be interested in many other policy-relevant effects of selective schooling. Their identification would require more assumptions, such as ones of conditional independence between the residuals. Carneiro et al. (2003) show how to identify Marginal Treatment Effects using a flexible independent factor analytic structure for the unobserved variables.

Multidimensional endowment. Allowing for multivariate η does not complicate the analysis very much. To simplify the presentation, we select K linearly independent rows of β and K linearly independent rows of γ. This is possible as the two matrices have rank K. With abuse of notation, we still call the resulting vectors of pre-treatment outcomes W and instruments Z as W and Z, respectively, and similarly still denote the matrices of coefficients as β and γ. Note that these are now $K \times K$ non singular matrices.

With this notation we show in the Appendix that the ratio of the return to endowment in post-treatment and pre-treatment outcomes, respectively, is given by:

$$\alpha_1\beta^{-1} = \mathbb{E}\left[\mathrm{Cov}\left(Y, Z|X, T=j\right)\right]\left\{\mathbb{E}\left[\mathrm{Cov}\left(W, \boldsymbol{Z}|X, T=j\right)\right]\right\}^{-1}, \tag{4.11}$$

$$= \mathrm{Cov}\left[Y - \mathbb{E}\left(Y|X, T=j\right), Z - \mathbb{E}\left(Z|X, T=j\right)|T=j\right]$$
$$\times\left\{\mathrm{Cov}\left[W - \mathbb{E}\left(W|X, T=j\right), Z - \mathbb{E}\left(Z|X, T=j\right)|T=j\right]\right\}^{-1} \tag{4.12}$$

Quasi-differencing approach yields:

$$\mathbb{E}\left(Y_j - \alpha_j\beta^{-1}W|X\right) = \mathbb{E}\left(Y_j - \alpha_j\beta^{-1}W|X, T=j\right), \tag{4.13}$$

and we immediately obtain the identification of the ATE parameter, as well as the ATT and the ATE on variances. An important difference with the univariate endowment case, however, is that the relative returns to the various components of η for the treated and non treated children (α_j/β in the univariate case) are not identified if $K > 1$.

Practical issues. Given an i.i.d. sample $\{Y_i, W_i, Z_i, X_i\}$, $i = 1, ..., N$, we proceed in two steps for the estimation of the ATE, ATT and ATE on variances parameters. In the first step, we estimate the ratio α_1/β by replacing expectations by empirical means in (4.7). This is straightforward to do if there are few discrete covariates. In our model of schooling, we want to condition on many covariates, discrete and continuous, as parental and school characteristics. One option is to adopt a flexible form for the conditional expectations, using for instance series estimators. We go half way in that direction, and adopt a linear specification. In practice, this amounts to first computing residuals of Y, Z and W on covariates X, and then estimating the covariances of these residuals, for the treated or non treated individuals. Including interaction terms had little effect on the results, suggesting that the linear specification is somewhat satisfying.

Note that in the estimation we have as many overidentifying restrictions as instruments. We use these restrictions for specification testing. Let τ be the vector of two-by-two equalities of the ratios given by (4.7). Under the model's assumptions τ is zero. The statistic we consider is $\widehat{\tau}'\left(\mathrm{diag}\,\widehat{\Omega}\right)^{-1}\widehat{\tau}$, where $\widehat{\tau}$ is an analog estimator of τ and $\widehat{\Omega}$ is a consistent estimate of the variance-covariance matrix of $\widehat{\tau}$. We take the diagonal as in Altonji and Segal (1996) . We estimate $\widehat{\Omega}$ by nonparametric bootstrap. We compute critical values from the bootstrap distribution of $\widehat{\tau}'\left(\mathrm{diag}\,\widehat{\Omega}\right)^{-1}\widehat{\tau}$.

We also use the overidentifying restrictions in order to improve the efficiency of the estimator. To do so, we weight the various estimates of α_j/β by the inverse of their bootstrap variance.

Then, given first stage estimates, say $\frac{\widehat{\alpha_1}}{\beta}$ and $\frac{\widehat{\alpha_2}}{\beta}$, we estimate the ATE by matching on the propensity score. We use the inverse probability weighting of Hirano et al. (2003), and remark that:

$$\Delta = \mathbb{E}\left[\frac{T\left(Y - \frac{\alpha_1}{\beta}W\right)}{\pi(X)} - \frac{(1-T)\left(Y - \frac{\alpha_0}{\beta}W\right)}{1-\pi(X)}\right] + \frac{\alpha_1 - \alpha_0}{\beta}\mathbb{E}(W), \quad (4.14)$$

where $\pi(X) = \mathrm{P}(T = 1|X)$ is the propensity score with respect to X. Note that it is not possible to match on $\pi(X)$ directly to estimate the ATE on levels Y. For this, knowledge of the propensity score with respect to X and η would be necessary. However, matching on $\pi(X)$ is feasible to estimate the ATE on quasi-differences $Y_j - (\alpha_j/\beta)W$.

We plug our estimates $\frac{\widehat{\alpha_1}}{\beta}$, $\frac{\widehat{\alpha_2}}{\beta}$ and $\widehat{\pi}(X)$ into (4.14), and replace the (unconditional) expectations by empirical means. To estimate the propensity score we use Logit ML, which can be seen as a first approximation to the series Logit estimator proposed by Hirano et al. (2003). Again, including interaction terms had little effect on the results. Lastly, note that, for (4.14) to be valid, we need that $\pi(X)$ be strictly comprised between 0 and 1. In practice, we compute the expectation ouside of the tails of the propensity score distribution, selecting the portion between the 5^{th} and 95^{th} percentiles.[4]

We proceed similarly to estimate the ATT parameter. Estimation of the ATE on variances is slightly different. Details on the estimation, including the case where η is multidimensional, are provided in the Appendix. Lastly, standard errors are computed by 500 bootstrap iterations.

4.5 Mean and variance of educational outcomes

We present in Table 4.5 the correlation matrix of the various test scores at age 7 and 11, for children attending selective and comprehensive schools. The estimates are all significant, and remain so when we include covariates (not shown). This suggests that

[4]We also restricted the estimation to the "thick support" (percentiles 33^{rd} and 67^{th}) as an informal check of robustness of the effect to the presence of other unobservables (see Black and Smith (2004)). Doing so had some effects on the results, but did not modify the qualitative conclusions below.

	Math 11	Read 11	Verbal 11	Math 7	Read 7
Comprehensive					
Math 11	1.00				
Read 11	.69	1.00			
Verbal 11	.74	.71	1.00		
Math 7	.52	.41	.45	1.00	
Read 7	.55	.56	.62	.49	1.00
Selective					
Math 11	1.00				
Read 11	.74	1.00			
Verbal 11	.78	.74	1.00		
Math 7	.55	.46	.48	1.00	
Read 7	.60	.60	.66	.48	1.00

Table 4.5: Correlation matrices of pre-treatment outcomes, by schooling system.

Assumption 5, which requires intuitively that the pre-treatment variables have at least one observable in common, is satisfied. Moreover, correlations between test scores are higher in selective schools, with the exception of the age 7 math and reading scores. Most of these differences are significant at 5%, and also robust to the inclusion of covariates. This is consistent with children at selective schools having a greater initial endowment, which in turn affects positively all educational outcomes.

In Table 4.6 we report the ATE estimates of the effect of attending a comprehensive school on the six outcomes of interest: the math and test scores at 16, age when the child left school, the number of vocational and academic qualifications, the weekly log wage in 2004 and the log of the average weekly wage of 1991, 2000 and 2004. We present the original ATE estimate and then the corrected estimate with its standard error. The corrected estimate is obtained by the quasi-differencing method assuming a single scalar endowment η. As a pre-treatment variable W, we have chosen the age 11 math score for all outcomes. The math age 11 score offers the best basis for a pre-treatment variable, as it is highly correlated with the outcomes and is, by definition, not affected by the treatment. The variables Z used as "instruments" are all the other pre-treatment outcomes. Conditional on observed and unobserved factors, the errors of the instrument equation, V, are assumed to be independent of the remaining error terms of the structural model. Note that the error terms ϵ_0, ϵ_1 and U can be freely correlated between each other. This correlation might represent a special ability, say for example for mathematical reasoning. We require however that V be uncorrelated to ϵ_0, ϵ_1 and U. If we would believe in such a correlation between ϵ_0, ϵ_1 and U representing a special ability, we would need an instrument which does not contain this factor after conditioning on the unobservable η. For example, the reading score at age 7 might satisfy this assumption particular well, as no special mathematics ability might be measured with it. A priori, it is however an important task to choose the Z variables. Fortunately, as we will see from our Sargan test, using different instruments leads us essentially to the same parameter estimates which indicates that there is, if any, very little residual correlation between the error terms of the instrument which indicates that we may use the other test score as Z.

The estimates of α_j/β were weighted in inverse proportion to their bootstrap variance. For the uncorrected and corrected estimates, the three columns correspond to (1) no covariates, (2) parental controls, and (3) a large set of parental, school and local controls.

We find that accounting for selection on unobservables matters, as it leads to a five to ten-fold reduction of the ATE estimates. The corrected ATE is still mostly negative. The effect is marginally significant for the math score at 16 in column (1), but insignificantly different from zero for the other outcomes. The effects on qualifications are positive in column (1), but also statistically insignificant.

In Table 4.7 we report the estimates of the returns to the endowment η, in the post-treatment relative to the pre-treatment outcomes. The variables Z used as "instruments" are ordered as in Table 4.5: Read 11, Verbal 11, Math 7 and Read 7. The age 11 maths score is used as W throughout, analogously to the previous specifications. For instance, the first figure in the first column of Table 4.7 refers to the age 11 math score used as W and the the age 11 reading score used as Z. The fourth figure in the same column corresponds to W being the age 11 math, and Z being the age 7 reading test scores. In the upper part of the table, we present the estimates of α_1/β, the relative return in the comprehensive system. In the lower part we report

	ATE			ATE (corrected)		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	−1.83	.20	−1.73	−.339	−.341	−.285
	(.19)	(−1.76)	(.30)	(.12)	(.15)	(.21)
Years of education	−.667	−.784	−.490	−.172	−.499	−.138
	(.26)	(.30)	(.44)	(.26)	(.30)	(.46)
Academic qual.	−.207	−.226	−2.65	.018	−.015	−.033
	(.04)	(.04)	(.06)	(.03)	(.03)	(.04)
Vocational qual.	. − 131	−.130	−.223	.008	−.022	−.099
	(.04)	(.05)	(.07)	(.04)	(.05)	(.06)
Weekly log wage 2004	−.084	−.176	−.086	.021	−.133	−.044
	(.13)	(.15)	(.21)	(.14)	(.15)	(.20)
Average log wage	−.048	−.12	−.198	.03	−.094	−.146
	(.18)	(.24)	(.26)	(.18)	(.24)	(.285)

Table 4.6: ATE and bias corrected ATE for various specifications of covariates; one-dimensional η.

α_0/β, the relative return in the selective one. We also report the weighted means of the estimates, that we used in order to estimate the corrected ATE in Table 4.6. Lastly, the included covariates correspond to specification (3) of Table 4.6.

Several interesting insights can be obtained from Table 4.7. First, the magnitude of the effects is informative. The endowment at age 11, η, has a lower effect on later outcomes than on the age 16 test score. Finding coefficients very different from one for the other outcomes motivates our quasi-differencing approach, which allows for different returns for various test scores. Second, for the math score at 16 and the years of education, we find some evidence that the return to η is higher in the selective system. However, the differences are statistically only weakly significant: for instance, the p-value of the difference of the weighted means estimates is .18 for the math score, and .20 for the years of education measure. In any cases the differences are small, also from a point of view of orders of magnitude. This result challenges the intuition that there is a higher reward to "ability" in selective schools, if we are ready to interpret the unobservable η as ability. At the same time, this result is in line with the results of Table 4.6, which show only few differences between the two schooling systems.

Under the null that the model's assumption are satisfied for one single η, the estimands obtained using different variables as instruments should be equal. Indeed, the estimates reported in Table 4.7 are of the same order of magnitude, suggesting

	Maths 16	Years education	Academic qual.	Vocational qual.	Log wage 2004	Average log wage
Comprehensive	.52 (.01)	.094 (.006)	.089 (.003)	.038 (.005)	.018 (.002)	.016 (.002)
	.50 (.01)	.069 (.005)	.076 (.004)	.035 (.004)	.014 (.002)	.011 (.002)
	.53 (.02)	.067 (.009)	.083 (.004)	.028 (.006)	.017 (.003)	.011 (.002)
	.53 (.02)	.067 (.009)	.083 (.004)	.028 (.006)	.017 (.003)	.013 (.002)
weighted mean	.51 (.01)	.074 (.004)	.077 (.003)	.036 (.004)	.016 (.002)	.012 (.002)
Selective	.56 (.01)	.116 (.006)	.088 (.004)	.037 (.005)	.016 (.003)	.016 (.002)
	.55 (.01)	.089 (.006)	.075 (.003)	.031 (.004)	.014 (.002)	.013 (.002)
	.58 (.01)	.084 (.008)	.073 (.004)	.035 (.007)	.017 (.004)	.013 (.002)
	.57 (.02)	.093 (.007)	.080 (.004)	.040 (.006)	.011 (.004)	.011 (.003)
weighted mean	.56 (.01)	.097 (.006)	.079 (.003)	.035 (.004)	.015 (.003)	.013 (.002)

Table 4.7: Relative returns to the endowment; one-dimensional η; specification (3) of Table 4.6.

	Maths 16	Years education	Academic qual.	Vocational qual.	Log wage 2004	Average log wage
One-dimensional η						
Comprehensive	.07	.00	.00	.18	.19	.16
Selective	.34	.00	.00	.31	.28	.19
Two-dimensional η						
Comprehensive	.98	.31	.25	.47	.35	.35
Selective	.12	.92	.57	.29	.61	.36

Table 4.8: P-values of the test of overidentifying restrictions, $K = 1$ and 2; specification (3) of Table 4.6.

that a single-η structure is a reasonable assumption, less so in the case of years of education and the number qualifications obtained. To check the statistical validity of this assumption, we then report in Table 4.8 the p-values of the tests that all returns are equal. In the first two rows, equalities of the estimates α_1/β (comprehensive) and α_0/β (selective), for different variables used as "instruments", are tested. Included covariates are the same as in the previous table. At conventional levels, the tests do not reject the equality in the case of the age 16 math score, vocational qualifications and wages, but they do reject it for the two other outcomes. We interpret this rejection as evidence that, though not a bad first-order approximation (see Table 4.7), a uni-dimensional η fails to capture a statistically significant part of the endowments of children. We then investigate if a bi-dimensional η would do better. The two last rows of Table 4.8 shows that it does, with much higher p-values.

For this reason, we then report the results using the quasi-differencing approach able to deal with a bi-dimensional η. In Table 4.9, we report the corrected ATE estimates, where selection on a bivariate unobservable is accounted for. The results indicate a similar picture as in Table (4.6). For all outcomes but the wages in column (1) , attending a comprehensive school is associated with slightly lower later outcomes. The effects are very smilar in magnitude to those in the case of one single η. Only the significance of the results is marginally affected.

	ATE			**ATE (corrected)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	−1.83 (.19)	−1.76 (.20)	−1.73 (.30)	−.316 (.13)	−.324 (.14)	−.233 (.28)
Years of education	−.667 (.26)	−.784 (.30)	−.490 (.44)	−.206 (.29)	−.532 (.27)	−.175 (.44)
Academic qual.	−.207 (.04)	−.226 (.04)	−2.65 (.06)	−.003 (.03)	−.030 (.04)	−.061 (.05)
Vocational qual.	. − 131 (.04)	−.130 (.05)	−.223 (.07)	−.013 (.05)	−.019 (.05)	−.106 (.06)
Weekly log wage 2004	−.084 (.13)	−.176 (.15)	−.086 (.21)	.032 (.15)	−.137 (.14)	−.055 (.19)
Average log wage	−.048 (.18)	−.12 (.24)	−.198 (.26)	.042 (.21)	−.103 (.24)	−.163 (.30)

Table 4.9: Uncorrected ATE (equivalent to table 1) and bias corrected ATE for various specifications of covariates; bi-dimensional η.

We then report the estimates of the ATE on variances. The results we obtain in Table 4.10 are much more clear-cut than in the case of the ATE on means. We find that, even controlling for a bivariate unobserved endowment, selective schools are associated with a larger variance. This is true in the case of the math score, for which the difference in variance between the two systems is about 10 percent. This is even more true in the case of later educational outcomes, where the difference in variances is about 25 percent. The variance is also higher in the case of educational qualifications and wages, and often significantly so.

	ATE on variances			ATE on variances (corrected)		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	−10.102 (1.35)	−9.677 (1.43)	−9.528 (1.74)	−6.311 (1.87)	−5.953 (1.47)	−3.068 (2.39)
Years of education	−1.180 (.33)	−1.343 (.292)	−1.815 (.457)	−1.108 (.34)	−1.286 (.29)	−1.672 (.447)
Academic qual.	−.172 (.05)	−.222 (.05)	−.198 (.07)	−.117 (.067)	−.18 (.06)	−.11 (.07)
Vocational qual.	−.141 (.04)	−.179 (.05)	−.240 (.08)	−.132 (.05)	−.183 (.05)	−.227 (.09)
Weekly log wage 2004	−.08 (.03)	−.073 (.03)	−.083 (.04)	−.08 (.03)	−.071 (.03)	−.074 (.04)
Average log wage	−.058 (.02)	−.060 (.02)	−.054 (.03)	−.058 (.02)	−.059 (.027)	−.053 (.03)

Table 4.10: ATE and bias corrected ATE on variances for various specifications of covariates; bi-dimensional η.

The results in Table 4.9 and 4.10 suggest that comprehensive shools do differ from selective schools. Moreover, differences appear not to be so much related to the average achievement of children, but to their dispersion. As an implication, children at selective schools have slightly better educational results, but they face more inequality. We view this result as a candidate to explain the push towards comprehensivisation. In the next section, we shall investigate this issue more thoroughly.

We also performed a series of robustness checks. We estimated the effects of attending a comprehensive school, restricting the sample to purely selective and purely comprehensive LEA's, as proposed by Manning and Pischke (2006). Indeed, if for example grammar schools coexist alongside comprehensive schools, they might attract the best students (the "cream-skimming" effect). By excluding not fully selective LEA's, we might expect to lower the differences in the age 11 test scores. In that smaller sample (638 children in the comprehensive system, 729 in the selective one), we were unable to allow for the presence of a bi-dimensional unobserved endowment, as we obtained very imprecise estimates. Using one factor, the results are in line with the previous ones, though less precisely estimated. Lastly, we estimated the effect of attending a comprehensive school for a sample of children that were directly affected by the comprehensivisation reform. We found a somewhat more negative effect, suggesting that there might have been a short-run impact of the reform (caused *e.g.* by reorganization costs at the school level). These additional results are available from the authors.

4.6 The two sides of selective schooling

Selective schools can be of two types, either grammar or secondary modern. By construction, these two types of schools are very different in many respects. In this section, we modify the previous analysis in order to account for these differences.

In Table 4.11, we report the mean outcomes for children attending grammar and secondary modern schools. We see that differences within the selective system are much stronger than between the selective and comprehensive systems. Indeed, children at grammar school score on average 10.1 points more than the ones at secondary schools at mathematics at age 16. Differences previous to entry are also enormous, as children at selective schools score on average 15.2 points more at maths at age 11.

Consider the following modification of Model (4.1), where the potential outcome in the selective system is replaced by two potential outcomes:

$$\begin{cases} Y_{0G} &= f_{0G}(X) + \alpha_{0G}\eta + \varepsilon_{0G}, \\ Y_{0S} &= f_{0S}(X) + \alpha_{0S}\eta + \varepsilon_{0S}. \end{cases}$$

In the model, Y_{0G} is realized if the child belongs to the selective system ($T = 0$), and if she is assigned to a grammar school (that we denote as $G = 1$). In turn, Y_{0S} is realized if $T = 0$ and the child is assigned to a secondary modern school ($G = 0$). The assumptions of Model (4.1) are readily extended to this model.

In the context where grammar and secondary modern schools are different, the treatment effects calculated in the previous section have no simple interpretation. In-

Variable	Grammar Mean	Grammar Std.Dev.	Grammar N	Secondary Modern Mean	Secondary Modern Std. Dev.	Secondary Modern N
Female dummy	0.543	0.498	1016	0.485	0.5	1989
Mathematics score 16	20.482	5.223	985	10.351	5.295	1887
Reading score 16	31.22	2.98	988	23.994	6.281	1894
Mathematics score 11	28.538	6.361	913	13.445	8.013	1723
Reading score 11	22.138	4.412	913	14.459	5.017	1723
Verbal score 11	31.946	4.827	913	20.165	8.099	1723
BSAG score 11	4.055	5.297	915	9.064	8.648	1721
Math score 7	6.884	2.075	936	4.679	2.272	1765
Reading score 7	28.621	2.244	939	22.147	6.683	1778
Average wage	4.897	0.559	297	4.54	0.607	428
1981 log weekly net wage	4.264	0.286	603	4.184	0.375	1050
1991 log weekly net wage	4.621	0.699	520	4.336	0.746	834
2000 log weekly net wage	4.727	0.754	566	4.39	0.768	990
2004 log weekly net wage	4.972	0.716	532	4.665	0.689	832
age left full-time education	19.072	2.829	739	16.705	1.647	1426
Academic qualifications	2.947	1.234	881	1.553	1.203	1626
Vocational qualifications	1.876	1.658	855	1.516	1.483	1582
All qualifications	3.232	1.047	881	2.096	1.344	1632

Table 4.11: Comparing outcomes in grammar and secondary modern schools

deed, it is difficult to give a precise meaning to Y_0, the potential outcome in the selective system. One possibility would be to define $Y_0 = G^*Y_{0G} + (1 - G^*)Y_{0S}$, and define G^* as a latent assignment to a grammar school for children attending the comprehensive system, while $G^* = G$ for the ones in the selective system. However, taking this route does not appear straightforward given our lack of data on pupils' ability and the precise nature of the assignment process.

Instead, we report in Tables 4.12-4.13 the estimates of the treatment effects of the fact of attending a comprehensive school on the outcomes, calculated for children attending a selective school. The effects reported in Table 4.12 are formally Average Treatment effects for the Non Treated (ATNT):

$$\Delta^{TNT} = \mathbb{E}\left(Y_1 - Y | T = 0\right).$$

The effects given in Table 4.13 are ATNT on the variances:

$$\Delta^{V,TNT} = \text{Var}\left(Y_1 | T = 0\right) - \text{Var}\left(Y | T = 0\right).$$

	ATNT			**ATNT (corrected)**		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	−1.874 (.16)	−1.876 (.15)	−1.611 (.20)	−.236 (.14)	−.331 (.15)	.081 (.294)
Years of education	−.521 (.14)	−.514 (.15)	−.428 (.27)	−.105 (.18)	−.475 (.19)	.04 (.51)
Academic qual.	−.240 (.03)	−.251 (.04)	−.254 (.05)	−.03 (.03)	−.066 (.03)	−.062 (.05)
Vocational qual.	−.12 (.04)	−.128 (.04)	−.182 (.05)	−.032 (.04)	−.035 (.05)	−.106 (.07)
Weekly log wage 2004	−.056 (.05)	−.109 (.06)	.006 (.11)	.051 (.07)	−.160 (.079)	.114 (.18)
Average log wage	−.052 (.08)	−.100 (.09)	−.021 (.18)	.022 (.09)	−.152 (.13)	.117 (.27)

Table 4.12: ATNT and bias corrected ATNT for various specifications of covariates; bi-dimensional η.

The results strengthen the ones obtained in Tables 4.9 and 4.10. For the math score, we find negative effects on mean outcomes, for example in column (2). These effects are smaller than in Table 4.9 and become insignificant when controlling for parental characteristics and LEA characteristics. Overall, the table offers the same picture as Table 4.9. However, effects on variances are still large and significant, only the wage variables seem statistically weakly significant.

	ATNT on variances			ATNT on variances (corrected)		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	−10.101	−9.692	−10.385	−6.189	−5.260	−2.695
	(1.26)	(1.35)	(1.68)	(1.32)	(1.66)	(2.29)
Years of education	−1.101	−1.408	−1.564	−1.01	−1.315	−1.428
	(.35)	(.34)	(.41)	(.35)	(.37)	(.43)
Academic qual.	−.170	−.232	−.221	−.115	−.181	−.125
	(.06)	(.05)	(.08)	(.06)	(.06)	(.09)
Vocational qual.	−.134	−.186	−.246	−.134	−.188	−.246
	(.05)	(.06)	(.09)	(.05)	(.07)	(.10)
Weekly log wage 2004	−.091	−.072	−.070	−.087	−.066	−.058
	(.03)	(.04)	(.04)	(.03)	(.04)	(.04)
Average log wage	−.053	−.059	−.041	−.050	−.056	−.038
	(.03)	(.02)	(.04)	(.03)	(.02)	(.04)

Table 4.13: ATNT and corrected ATNT on variances for various specifications of covariates; bi-dimensional η.

As in the previous section, the strongest result we obtain is the fact that selective schools show more dispersion in educational outcomes than comprehensive ones. The effect on means, when significant, is of a smaller order of magnitude. In the rest of this section, we intend to better understand the sources of this greater dispersion.

One source of the higher variance shown by the selective sytem is the variance between grammar and secondary schools. To investigate if these differences are robust to the inclusion of covariates we would ideally like to estimate the ATE parameter:

$$\Delta_{GS} = \mathbb{E}\left(Y_{0G} - Y_{0S} | T = 0\right).$$

Still, this overall average is meaningless as, by construction, there are very few children in grammar schools with a low η, say in the lower quartile of the distribution. The low-η children in secondary modern schools have no counterparts in grammar schools to whom they could be compared. Note that the lack of overlap occurs because of the unobserved regressor η, for which we have no direct measure. In Figure 4.2 we draw the densities of the "ability" and "parental background" components in grammar and

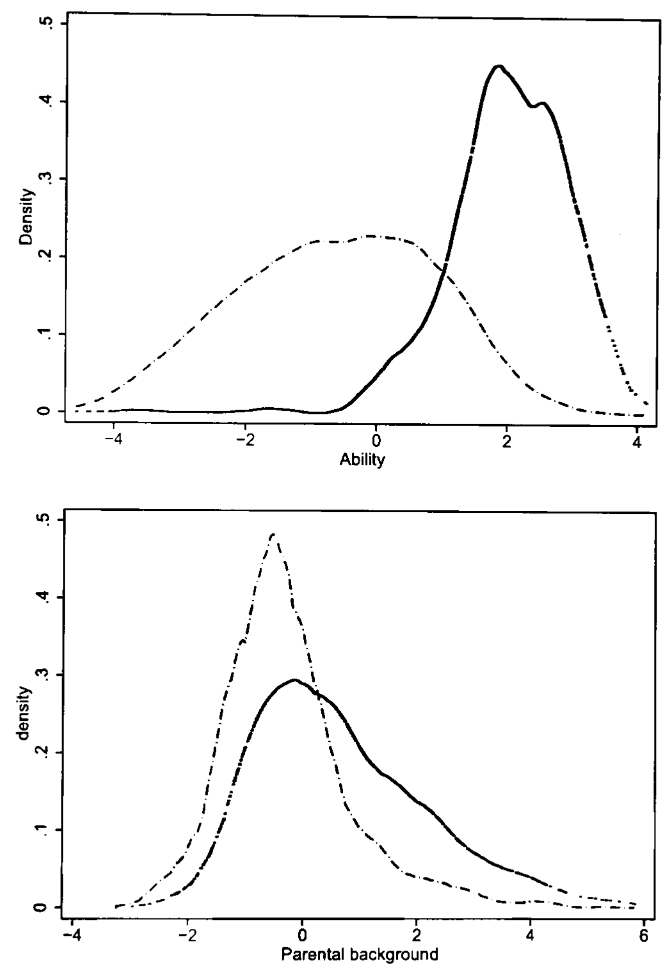


Figure 4.2: The "ability" and "parental background" principal components for pupils attending secondary modern (dash-dotted line) and grammar schools (solid line).

secondary modern schools. The upper panel, which represents "ability", illustrates what is likely to be the case of the distribution of η, showing little room for comparison.

For this reason, we choose to compute averages on children that have an intermediate level of the "ability" factor, between the percentiles .40 and .80. These children have significant probabilities of attending the two school types. We report in Table 4.14 the corrected ATE estimate, which allows for a uni-dimensional η. We find strong some strong positive effects of the fact of attending a grammar school on all outcomes. When corrected by the quasi-differencing method allowing for one scalar endowment, the effects are much reduced but still strongly significant for the math scores and the wages. The effect on the years of education and academic qualifications obtained is insignificantly different from zero. Lastly, the effect on the number of vocational qualification levels obtained is negative and insignificant.

	ATE			ATE (corrected)		
	(1)	(2)	(3)	(1)	(2)	(3)
Math 16	10.645 (.96)	9.991 (.92)	10.664 (1.03)	2.705 (.87)	2.385 (.76)	2.678 (.78)
Years of education	1.366 (1.56)	3.057 (1.285)	4.170 (1.51)	1.124 (1.24)	1.264 (1.16)	2.152 (1.19)
Academic qual.	1.210 (.19)	1.226 (.17)	1.377 (.23)	.16 (.17)	.08 (.18)	.14 (.25)
Vocational qual.	.35 (.19)	.369 (.17)	.444 (.21)	−.132 (.24)	−.127 (.21)	−.26 (.26)
Weekly log wage 2004	1.358 (.53)	1.071 (.50)	1.807 (.618)	.875 (.50)	.589 (.52)	1.109 (.49)
Average log wage	1.687 (.70)	1.620 (.69)	3.006 (.84)	1.113 (.59)	.991 (.56)	2.206 (.88)

Table 4.14: Bias corrected ATE of attending a grammar rather than a secondary modern school, for various specifications of covariates; uni-dimensional η.

These results shed some light on the differences between selective and comprehensive secondary schooling. Comparing the orders of magnitude in Tables 4.12 and 4.14 shows that there are few differences in means between the two systems, but there are strong differences between the two school types of the selective system. This has interesting implications for the understanding of the schooling reform in Britain. As any test, the *11-plus* exam involves several factors: the ability of the child, her specific preparation and, importantly, a certain amount of luck. By comparing children with similar ability who ended up at a grammar or a secondary modern school, we measure the "premium" that is associated with being sucessful at the exam. The estimates we obtain suggest that this premium is large, and explains a part of the variance in outcomes in the selective system. The conclusions of this exercise are related to the findings of Maurin and McNally (2006), who show that relaxing the admission criteria to grammar schools in Northern Ireland has led to an increase in total educational attainment. However, to our knowledge no such reform was undertaken in Britain. From the perspective of parents, the variance between school types represents a risk associated with the probability of failing the *11-plus* exam. As we find small differences in means between the two schooling systems, it may have been rational for risk-averse parents to support the push towards comprehensivisation.

4.7 Conclusion

Several recent papers in the literature insist on the fact that it is difficult to separate the effect of school type on educational outcomes from other factors, most importantly the child's ability. Possible answers to this problem are to use "natural experiments" (Meghir and Palme (2005), and Maurin and McNally (2006)), or to exploit cross-country and time variation (Woessmann and Hanushek (2006)). The absence of such data motivates our quasi-differencing method, that uses the availability of various test scores previous to secondary schooling.

It is often thought that when choosing between selective and comprehensive education one faces a trade-off between efficiency and equity. We find little evidence of such a trade-off. The raw data from the NCDS, as well as regressions using a large set of covariates, suggest that children attending selective schools have better results on average. We show that this is mainly the result of composition effects in terms of unobservables, selective schools having a better intake. When we difference the unobservables out, the effect of attending a selective school is still positive, but small and most often insignificantly different from zero. In contrast, the difference in variance remains, at least for the maths and years of education outcomes. In addition, we give some evidence on the sources of the higher variance in the selective system, showing that there is a large premium to passing the admission exam and attending a grammar school.

Conclusions

This thesis has examined the economics of child labour and child education. It has provided a new theoretical model of the commitment problem of child labour in developing countries and it has empirically assessed the intergenerational transmission of education across three generations and evaluated the impact of a schooling system reform in the United Kingdom.

The thesis focused first on developing countries. Here economic motivations for parents to chose child labour instead of child education are very strong if parents are poor and if, additionally, commitment problems prevent efficient educational investment. In a theoretical model, this thesis showed that the informational requirement of an observable history of play is crucial to equilibria proposed so far in the literature. Without this observability, additional sanctions or bequests are needed to enforce first-best intergenerational contracts. We also derived an important role for grandparents in this model, as they save over a longer horizon, can more credibly commit to bequests than parents and observe the game for a longer time-span. Bequest sanctions act in this model as a punishment, but also as a costly signal. The implications of this analysis were that families might have difficulties implementing intergenerational contracts. This gives room for public policy, which needs to address not only the poverty problem, but needs to provide a solution to the intergenerational dynamic commitment problem.

In chapter three we tried, to some extent, to test the influence of grandparents on grandchildren. In censored regressions, and without taking into account the potential endogeneity of education, we find strong effects for grandparents bequests, education and the presence of grandparents on grandchildren's education. As soon as we address the endogeneity using IV techniques, the strong effect of education disappears. We then observe, however, a significant impact of unobservable factors of parents and grandparents on grandchildren's education. This is of considerable interest to policy and agrees with other studies in this area. It is not education itself, but rather other important factors that determine the education choice. This could be ability, parenting skills or attention dedicated to the child. A closer examination of these unobservable factors is a challenging starting point for future research. Policy could try to affect these variables in order to increase educational attainment of generations.

Finally, chapter four evaluates another aspect of educational choice, a schooling reform within England and Wales. There has been a large discussion on the effects of this reform. Previous studies, who evaluated this reform, are potentially biased due to so called composition effect, i.e. the fact that comprehensive schools have a much worse intake of students than selective schools. This last chapter addresses this bias by developing an original econometric technique of quasi-differencing. We use test scores prior to pupils' enrolment in the comprehensive school to estimate one or more unknown factors, such as ability, in the data. The analysis is then conducted by matching on the difference between pre-and post reform outcomes. The bias is removed by differencing out the unobservable component from the matching estimand. Using this technique, we find that comprehensive school pupils do, on average, no worse than a child in the selective system. We find, however, a higher variance in outcomes for those pupils attending schools within the selective system. This is indicative, in our interpretation, of a higher risk associated with the selective system. Indeed, we show that once pupils pass the selection exam to enter grammar school, there is a large premium associated with attending a grammar school rather than a secondary modern school. Thus, if a pupil belongs to the middle range of a hypothetical ability scale, there is a non-ignorable probability of failing the admission exam, due to factors outside of control of the child or the parent (i.e. "having a bad day" during the exam). This might influence the schooling choices of the child in a negative way. Thus parent might have been aware of this risk and initiated a movement for the establishment of comprehensive schools, which as documented by sociologists, was a "grass-roots" initiative.

In summary, this thesis presents some original theoretical and empirical work which links with debates in the literature on the economics of child labour and child education and provides new results. The thesis aims to become a reference work in this area, but also a basis for interesting future work. This would include a more thorough investigation of the role of grandparents in association with intergenerational transfers, a closer scrutiny of unobservable factors in the data which influence educational attainment and a re-evaluation of the influence of the school system on outcomes using more recent data.

Appendix A

Chapter 2

A.1 Symmetry of equilibrium in two generational cyclic games

A useful property of cyclic games can be derived if we elaborate on the similarities between two-generation cyclic OG games and a general class of symmetric games.

The proceeding will be as follows. We first define cyclic two-generation OG games. Then we define symmetric games. We then show that a cyclic two-generation OG game has a finite symmetric normal form representation for every player t interacting with $t+1$. From this, we can invoke Theorem 2 of Nash (1951), which asserts that any finite symmetric game must have a symmetric equilibrium point. This is thus also special property of cyclic two-generation OG games, which simplifies the search for equilibria considerably.

Definition 5 *A cyclic two-generation OG game is characterised by*

(i) a (possibly) infinite set of players $t \in \mathbb{N}$,

(ii) a finite set S, composed of k pure strategies, each player t chosing $s_{tk} \in S$, S being the same for all players $t \in \mathbb{N}$,

(iii) a payoff function W which depends exclusively on the strategies played by player t and player $t+1$, thus

$$W_t(s_t, s_{t+1}).$$

(iv) an associated normal form representation as a matrix M of dimension $k \times k$, each element being a two-element row vector, denoting the payoffs to player t and $t+1$.

Example: Prisonners Dilemma Consider the famous prisonners' dilemma game, played as a cyclic two generation OG game. Player t is in the game, player $t+1$ enters the game. Player t plays against player $t+1$, and then player t exits. Player $t+1$ is then already in the game and player $t+2$ will enter. Thereafter, player $t+1$ plays against player $t+2$, and then player $t+1$ exits...et cetera. The player set can thus clearly be infinte.

Furthermore, the strategy set of player t is the same as for player $t+1$ and so forth and equal to confess, C, or not confess, NC, thus $S = \{NC, C\}$. How can we obtain a normal form representation ? The game is forward looking, i.e. we don't know what player $t+1$ will do against player $t+2$, since this will occur in the next step of the game. We can however use an equivalence: As the strategies of player $t+2$ are the same as those of player t, we can, label a player t equivalently as a player $t+2$ and construct our payoff matrix.

Payoffs to player t only given actions of $t+1$ can be expressed in the following matrix, G_t.

		Player t	
		NC	**C**
Player t+1	**NC**	1	-1
	C	2	1

Similarily, we can do exactly the same for player $t+1$ and obtain G_{t+1}:

		Player t+1	
		NC	**C**
Player t+2	**NC**	1	-1
	C	2	1

which is trivially equivalent. To reconsider the problem mentionned above, player t only has the following information about the normal form representation in period, a 2×2 matrix, M_t with vector valued entries:

		Player t	
		NC	**C**
Player t+1	**NC**	$(1, a)$	$(-1, b)$
	C	$(2, c)$	$(0, d)$

We ignore what player $t+1$ payoffs against player t will be: Player $t+1$ moves against player $t+2$, and not against t. Thus we wrote abitrary values a, b, c, d into the matrix for the moment. Choices of player $t+1$ affect player t. And we know player $t+2$ has the same strategies than both players player t and $t+1$ and furthermore the same payoff function.

In M_t player $t+1$ is a column player. In the (individual) payoff matrix for player $t+1$, he is a row player. To know what player $t+1$ chose against $t+2$ as a column player, I will have to transpose matrix G_{t+1}. Now use the fact that player t's strategies are equivalent to those of player $t+2$. Let the matrix G_t be renamed P with a typical element being p_{ij}. Denote the transpose of G_{t+1} as Q. Denote a typical element of Q as q_{ij}. Then we can form the following normal form representation, M:

		Player t	
		NC	**C**
Player t+1	**NC**	(p_{11}, q_{11})	(p_{12}, q_{12})
	C	(p_{21}, q_{21})	(p_{22}, q_{22})

Where clearly

$$Q \equiv \begin{bmatrix} (q_{11}) & (q_{12}) \\ (q_{21}) & (q_{22}) \end{bmatrix} = P'.$$

Then in the Prisonners Dilemma cyclic OG game, the normal form representation is given by

		Player t	
		NC	**C**
Player t+1	**NC**	$(1,1)$	$(-1,2)$
	C	$(2,-1)$	$(0,0)$

From this example, it is now obvious that there is some for of symmetry in cyclic games, since players have the same strategies in every period, and we assumed here that they receive the same payoff. In order to deepen our understanding of symmetry, let us state the symmetry definition of Nash (1951).

Definition 6 (Symmetry, Nash (1951))

Suppose there are n players. An automorphism or symmetry of a game will be a permutation of its pure strategies which satisfies certain conditions on the permutation of players. Let ϕ be a permutation of the pure strategies. ϕ induces a permutation ψ of the players. Let χ denote the permutation of the pure strategies induced. Each n-tuple of pure strategies is therefore permuted into another n-tuple of pure strategies. Let ξ be one such n-tuple of pure strategies. Let $W_i(\xi)$ denote the payoff to player i when such an n-tuple is employed. It is required that if

$$j = i^{\psi},$$

$$W_j(\xi^{\chi}) = W_i(\xi).$$

Example We require only the two player case, whereas Nash (1951) considered the n-player setup. Consider the general matrix representation of a two-player, two-strategy game. The strategies here are x and y. Payoffs to each player are denoted by W, subscripted by strategies played by the row player (first subscript) and by the column player (second subscript) and superscripted by the player index.

		Player t	
		x	y
Player t+1	x	(W^1_{xx}, W^2_{xx})	(W^1_{xy}, W^2_{xy})
	y	(W^1_{yx}, W^2_{yx})	(W^1_{yy}, W^2_{yy})

If players are now permuted by the rule $j = i^{\psi}$, player 2 is interchanged with player 1 and vice-versa. According to Nash (1951), it should be true that if players are permuted, and they play strategies permuted in the same manner, payoffs should be equal. More formally,

$$W^1_{yx} = W^2_{xy}.$$

Here the player index was permuted, $1^{\psi} \to 2$, $2^{\psi} \to 1$. In the simple two player case, this implies *interchange* of players, thus their strategies should also be interchanged. Similarily it must hold,

$$W^1_{xy} = W^2_{yx},$$

and of course

$$W^1_{xx} = W^2_{xx},$$

$$W^1_{yy} = W^2_{yx}.$$

These four conditions encompass all the pure strategy combinations in the two player case.

If one examines now player 1's payoffs only, collected in a matrix A, we have

$$A = \begin{bmatrix} (W^1_{xx}) & (W^1_{xy}) \\ (W^1_{yx}) & (W^1_{yy}) \end{bmatrix}.$$

Similarily, collect all the payoffs of player B in a matrix

$$B = \begin{bmatrix} (W^1_{xx}) & (W^1_{xy}) \\ (W^1_{yx}) & (W^1_{yy}) \end{bmatrix}.$$

The four conditions on pure strategies will only be true if

$$A' = B.$$

Proposition 6 *A cyclic 2-generation OG game with identical payoffs for each time period has a finite symmetric normal form representation.*

Proof *The payoffs to player $t+1$ facing player $t+2$, playing strategies s_{t+1} and s_{t+2} respectively are the same as the payoffs to player t facing player $t+1$, playing strategies s_t and s_{t+1}. This follows from the identical payoff functions for each time period, $W^{t+1}(s_{t+1}, s_{t+2}) = W^t(s_t, s_{t+1})$ whenever $s_{t+1} = s_t$ and $s_{t+1} = s_{t+2}$.*

A $k \times k$ payoff matrix for player t may be formed as follows. Playing against player $t+1$, player t will receive the following payoffs.

$$P = \begin{bmatrix} W_{11}^t & W_{12}^t & \dots & W_{kk}^t \\ W_{21}^t & W_{22}^t & \dots & W_{2k}^t \\ W_{31}^t & \ddots & \dots & W_{3k}^t \\ & \ddots & & \\ W_{k1}^t & W_{k2}^t & \dots & W_{kk}^t \end{bmatrix}$$

Player t's k strategies are the rows of this matrix, whereas player $t+1$'s k strategies are the columns of this matrix. A typical element of P is given by p_{ij}.

Player $t+1$ then plays against $t+2$. Thus player $t+1$ has a payoff matrix of the following form.

$$P = \begin{bmatrix} W_{11}^{t+1} & W_{12}^{t+1} & \dots & W_{kk}^{t+1} \\ W_{21}^{t+1} & W_{22}^{t+1} & \dots & W_{2k}^{t+1} \\ W_{31}^{t+1} & & \dots & W_{3k}^{t+1} \\ & \ddots & & \\ W_{k1}^{t+1} & W_{k2}^{t+1} & \dots & W_{kk}^{t+1} \end{bmatrix}$$

Player $t+1$'s k strategies are the rows of the same matrix P, whereas player $t+2$'s k strategies are the columns of this same matrix P. The matrices are identical since, $W^t = W^{t+1}$ whenever $s_{t+1} = s_t$ and $s_{t+1} = s_{t+2}$.

In the game of player $t+1$ against player $t+2$, the strategy of player $t+2$ is equal to the strategy of player t. The strategy set of player $t+2$ is equal to the strategy set of player t. This follows from the definition of a cyclic 2-generation OG game. Transposing *the matrix P to yield $P' = Q$, relabelling player $t+2$ player t and forming a new matrix by joining P and Q, we obtain a new $k \times k$ matrix normal form representation M with each entry consisting of a vector, $m_{ij} = (p_{ij}, q_{ij})$.*

The payoff to player t is given in the first entry of m_{ij} and the payoff to player $t+1$ in the second entry of m_{ij}. Nash's (1951) symmetry condition requires

$$W^t(i,j) = W^{t+1}(j,i),$$

and thus

$$p_{ij} = q_{ji}.$$

Consider player t playing strategy i and player $t+1$ playing strategy j. This yields $m_{ij} = (p_{ij}, q_{ij})$ and hence payoff p_{ij} to player t. Player $t+1$ playing strategy i and player t playing strategy j (interchange of strategies) yields $m_{ji} = (p_{ji}, q_{ji})$ and hence payoff q_{ji} to player $t+1$.

It follows that $p_{ij} = q_{ji}$ since q_{ji} is obtained from entry p_{ij} of the original matrix by the usual rule of transposition. Since $q_{ij} = p_{ij}$, player $t+1$ earns the same payoff as player t when interchanging strategies, which completes the proof of symmetry in the sense of Nash (1951) □

Proposition 7 *A cyclic 2-generation OG game with time invariant payoffs has a symmetric equilibrium point*

Proof *Since the cyclic 2-generation OG game with time invariant payoffs has a finite symmetric normal form representation, it follows immediately by Theorem 2 in Nash (1951) that it has a symmetric equilibrium point.* □

The above proposition simplifies the search for equilibria, since only symmetric equilibrium points can be candidates for a long run solution. However, another twist occurs in our game, which we need to account for. Every period, payoff to members of a given generation increases by a proportional factor because of wage growth g. Thus, suppose we have the payoffs received in the payoff points as $W_{i,t}$, the payoff received in payoff point ∇_1 at time t. Because of wage growth g, we have $W_{i,t} = \rho_1 W_{1,t+1}$,

$W_{2,t} = \rho_2 W_{2,t+1}$, ..., $W_{16,t} = \rho_{16} \nabla_{16,t+1}$ in our particular game. How does this affect the equilibrium ?

Proposition 8 *(i) A unique Nash equilibrium in pure strategies in a 2-player normal form game is not affected by a transformation that multiplies all the payoffs of one player by the same constant. (ii) It is also not affected by a transformation that multiplies every payoff of that player by a different constant, but keeps the original ranking of the payoffs.*

Proof *A Nash equilibrium point in pure strategies is characterised by each player chosing a pure strategy s_i, such that*

$$W_i(s^*) = \max_{s_i} W_i(s^*_{-i}, s_i).$$

If the ordering of the payoffs is unchanged for player i, the condition will still hold true. Suppose $W_i' = \rho W_i$ for all payoffs of player i. Then it will still be true that

$$W_i'(s^*) = \max_{s_i} W_i'(s^*_{-i}, s_i) = \rho \max_{s_i} W_i(s^*_{-i}, s_i).$$

(ii) Suppose the payoffs of player i are ranked in the following way using the abbreviation notation

$W_j = W_i(s^*_{-i}, s_j)$ and $W_k = W_i(s^*_{-i}, s_k)$. Suppose the payoffs of the best response are ranked before the transformation as $W_j > W_l > W_k$. Thus, $\max_{s_i} W_i(s^*_{-i}, s_i) = W_j$ in this case. Now suppose we multiply each of these best response payoffs by the constant of the corresponding index, the constants satisfying $\rho_j \geq \rho_l \geq \rho_k$, thus multiplying each payoff by a different constant, but keeping the original ordering of payoffs. Consequently W_j is still the maximum of the best response payoffs. □

Of course, the previous proposition does not hold for multiple equilibria, and not generally for a unique mixed strategy equilibrium. In a multiple equilibria setting, a player must be indifferent between at least two strategies in the support of his mixed strategy - implied by the linearity of expected payoffs in probabilities. Thus, changing one payoff by a greater factor than the other will not preserve the ordering (!) and hence only fully proportional transformations preserve the ordering in a multiple equilibria setting.

Example Prisonners dilemma

We have:

		Player t	
		NC	**C**
Player t+1	**NC**	(π_C, π_C)	(π_D, π_P)
	C	(π_P, π_D)	(π_N, π_N)

The ranking of the payoffs is as usual, $\pi_D > \pi_C > \pi_N > \pi_P$.

The equilibrium point here is (π_N, π_N), since each player prefers to play D, since $\pi_D > \pi_C$. A transformation, multiplying $\rho \geq 1$ into each of the payoffs of player 2 (column player) only, does not change his payoff ordering and thus not his strategy. Even if a transformation, which preserves the ordering, is allowed for $\rho_1\pi_D > \rho_2\pi_C > \rho_3\pi_N > \rho_4\pi_P$, the pure strategies of player 2 will not be affected.

A.2 Choice of the first-best contract

It is not without implications how to solve the joint maximisation problem. We assume first that the individuals utility function is strictly concave in the transfer α - i.e. there exists a transfer that improves total family utility. Furthermore, the child labour wage is always smaller than or equal to the adult uneducated wage - i.e. no unrealistic corner solutions with respect to l exist.

In this situation, the first best outcome is reached, if an individual first maximises his lifetime utility with respect to α^* and then parents chose the appropriate level of human capital investment l^*. This can be shown as follows. We denote the scenario when parents chose first l and then children chose α as maximisation problem 1, with l_1 as optimal solution and we denote the maximisation problem where children choose α first and then parents l as scenario 2, with solution l_2.

Given parents maximise the first order with respect to l_1, the first order condition for l_1 is given by

$$-h'(l_1) = \frac{w^{CH}}{\alpha}. \tag{A.1}$$

Given children chose first the optimal α^* and parents choose l_2, we can show, using the result

$$\alpha^* = \frac{h(l_2) - M'(g)h(l_2) - M'(g)gl_2w_t^{CH}}{M'(g)gh(l_2) + h(l_2)},$$

that this implies parents maximising the following FOC over l_2:

$$\max_{l_2} \frac{h(l_2) - M'(g)h(l_2) - M'(g)glw_t^{CH}}{M'(g)gh(l_2) + h(l_2)} gh(l_2) + l_2gw^{CH},$$

which has the solution

$$-h'(l_2) = \frac{w^{CH}M'(g)}{(1 + M'(g))}. \tag{A.2}$$

Note that for the above α^* to be positive, we need the concavity condition of the utility function in α and this implies $g > 1$.

Now, one can show that for human capital to be greater under scenario 1, the amount of child labour is lower for $\alpha \in [0, 1]$. The amount of child labour is implicitly indicated by the derivative function $h'(l)$, which is decreasing in its argument. Applying again an inverse function, $h^{'-1}$, we obtain that the solution l must be increasing in the argument. Thus we can compare the two right hand sides of Equation (A.1) and (A.2). Child labour in scenario 1 can only be lower if

$$\frac{w^{CH}}{\alpha} < \frac{w^{CH}M'(g)}{(1 + M'(g))},$$

or if $\alpha > \frac{(1+M'(g))}{(M'(g))} > 1$.

Since $\alpha \in [0, 1]$, this implies that $l_1 > l_2$.

Now, an argument from the optimal first oder conditions on α shows that any $l > l_2$ must yield a lower utility.

The optimal first order conditions on α_1 imply

$$W'(h(l_1)(1 - \alpha_1)) = gW'(h(l_1)(1 + g\alpha_1) + l_1gw^{CH}). \tag{A.3}$$

for an interior solution for α. Now suppose $h(l_1)(1-\alpha_1))$ with $l_1 > l_2$, then if $\alpha_1 = \alpha_2$, the first part of the utility function satisfies

$$W'(h(l_2)(1-\alpha_2)) < W'(h(l_1)(1-\alpha_2))$$

If $\alpha_1 \geq \alpha_2$, the above condition is still satisfied, and we obtain the result immediately: the FOC equality, Equation (A.3) or simplified $W'(.) = gW'(.)$, implies that the second part of the utility function satisfies:

$$gW'(h(l_2)(1+\alpha_2) + l_1 gw^{CH}) < gW'(h(l_1)(1+\alpha_1) + l_2 gw^{CH}).$$

It follows that both arguments of the utility function under maximisation problem 2 are also greater, by the properties of marginal utilities, namely if $W'(a) > W'(b) \rightarrow a < b$.

If $\alpha_1 < \alpha_2$, then there can be no solution, since a reduction in α_2 below α_1 entails an utility loss of $(1-g)W'()$, which is negative for $g > 1$. Thus total utility will fall further from the situation where $\alpha_2 = \alpha_1$ where equal.

Illustration Figure A.1 illustrates the above result. The total utility with transfer α is depicted under the two maximisation schemes. One can see that both functions are concave in α, but that the function for maximisation problem 1 stays strictly within the enveloppe formed by the function from maximisation problem 2.

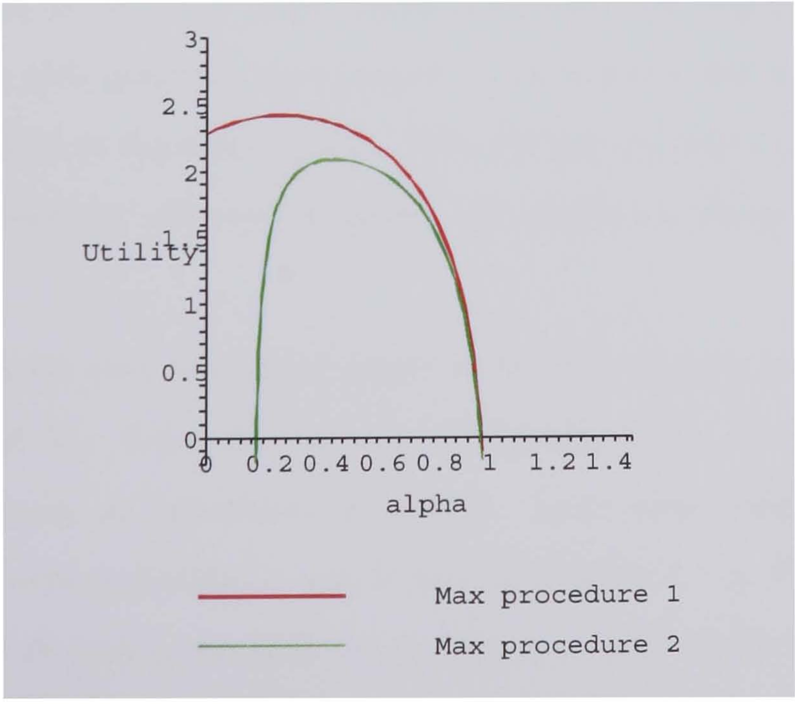


Figure A.1: Example of the two maximisation procedures

A.3 The relation between cyclic OG games, the normal form and overlapping generations games

Cyclic OG games have not been intensively studied in the literature. However, they possess interesting properties. This appendix is designed to highlight specificities and similarities of cyclic OG games. Let us start with the similarities by arguing that a cyclic OG game has the same (pure or mixed) equilibria as the normal form static game. Suppose we have the following normal form representation of a game for players 1 and 2, G_1:

		Player 1	
		NC	**C**
Player 2	**NC**	$(1,1)$	$(-1,2)$
	C	$(2,-1)$	$(0,0)$

Player 1 plays rows and player 2 columns. How can we obtain the cyclical game representation of this normal form game ? Cyclic games are dynamic games, thus they have to be written in extensive form. Also, we have to take into account that player 1 enters the game at his entry point, makes a decision and then exits the game while he is overlapping with player 2. Since player 1 has an entry and an exit point, we can always cut the game at those two points. This will leave us with a finite, acyclic graph. In the case of Prisonners' dilemma as above, we would have the graph of figure A.2 for player 1:

As in the main text, the player enters at his first decision point and exits at the exit point marked ∇_i. From this tree we can quickly derive the equilibrium solution looking, for example, at behavioural strategies. Each player randomises his choices. Player 1 plays C with probability p and D with probability $1-p$, Player 2 plays C with probability q and D with probability $1-q$. One can easily check that player 1 should play with probability $p^* = 0$ and Player 2 should play with probability $q^* = 0$, thus the only equilibrium in this game is the usual, namely $p^* = q^* = 0$, and thus (D,D) in pure

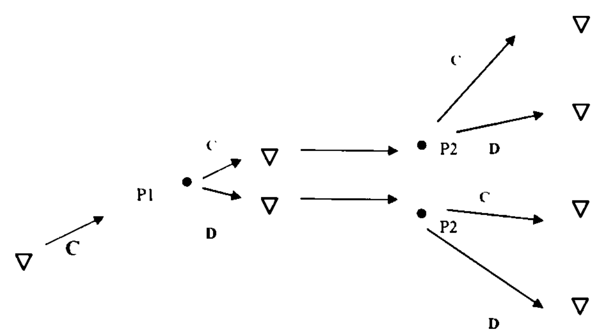


Figure A.2: A finite graph from the Prisonners dilemma game, assuming player 1 moves first.

strategies.

As the other normal form games, also cyclic games have at least one equilibrium. Consider this example which is a game that cannot be solved by applying pure strategies or for example iteratively eliminating dominated strategies.

		Player 1	
		NC	**C**
Player 2	**NC**	$(4,0)$	$(0,4)$
	C	$(1,2)$	$(2,1)$

Applying the same methodology as above, one can determine that player 1 should play $p^* = \frac{2}{5}$ player 2 should play $q^* = \frac{3}{5}$. Since cyclic overlapping generation games use as the main building blocks of finite extensive form games and these can be converted to normal form, the Nash (1950) result applies as well here: every finite normal form game has a mixed strategy Nash equilibrium. Thus also every cyclic OG game must have at least one (pure or mixed) strategy equilibrium.

Cyclic OG games behave as normal form games. However, they are dynamic games with players that enter and exit all along the game. They do not have an end and are thus infinite. If we look at approaches in the literature that come close to such a setup, we could think of comparing repeated play of a stage game in normal form with a cyclic game. The usual infinite repeated stage game has two players playing repeatedly against themselves a simultaneous move game. In this traditional setup players stay the same. Therefore, a better comparison would be to consider play of stage games by overlapping players, (Smith (1992), Bhaskar (1998), Yoon (2001), Kandori (1992)). We will illustrate the differences of these games with respect to cyclic games with the example of Prisonners' dilemma as this was almost exclusively used in the literature. The first major difference is obviously, that cyclic games have a dynamic structure where players do not move simultaneously. This is crucial for the equilibrium strategy proposed within the cyclic game.

Two types of setups will be compared. First the repeated play of Prisonners' dilemma is studied with players that overlap for one period and play the game twice. In the jargon of the literature on overlapping generation games, players will play a first time when they are young and a second time when they are old. Thus in the young generation the player plays C or D against a player of the old generation. Then the young player takes the place of the old player. The old player plays then against the next young player, chosing C or D. Which equilibria can be implemented ?

As Bhaskar (1998) states, these players can implement two types of strategies considered by the literature, they can play GRIM or RESILIENT. I will paraphrase Bhaskar (1998) in the following two sections. Both of the strategies GRIM and RESILIENT suppose knowledge of the history of play. Consider strategy GRIM: A young agent plays D while the older agent (with whom he overlaps) plays C. Thus a young agent makes a transfer to the older player of -1. Each subsequent player does so if all the history of play has been (C,D). An old player receives thus 2 always and a young player -1, thus each a lifetime payoff of $\pi = -1 + \delta 2$. As long as $\delta > \frac{1}{2}$ such an equilibrium is possible. If the young agent would play D and fail to make a transfer, all other agents would revert to play D since this is the NE of the game, and thus a best response. Knowledge of the entire history of play is necessary, and we will explain this in more detail below.

Consider strategy RESILIENT: In this set-up, only deviants will be punished: If a young person deviates from the above equilibrium, he will be punished by playing D in the next young generation, when the deviator is old. This yields the deviator a lifetime payoff of 0. The next young are identified as "police man" and can still expect a transfer when they enter the old stage. Again, knowledge of the entire history of play is necessary, since otherwise a player cannot distinguish between individuals who are "police men" or who are deviators.

Bhaskar (1998) and Yoon (2001) give the same, relevant example concerning observability of the history of play. Consider the strategy RESILIENT or GRIM: Suppose every player can only remember what the previous generation did. So a player in period t can only remember what players in peridod $t-1$ did. Player t must give a transfer to $t-1$ when he is young. If he fails to do so, then he must be punished by player $t+1$. However, when player $t+1$ initiates the punishment sequence, player $t+2$ cannot distinguish if the agent justifiably initiated a punishment or if that agent simply deviated. Player $t+2$ cannot verify what player t did because he does not observe the history of play. Thus, agent $t+1$ will also be punished in any case, since this is the best reply. Thus agent $t+1$ has an incentive to conceal player t's deviation and do as if nothing happened. Clearly then player t plays a deviation - thus there is no other equilibrium than the non-cooperative Nash solution in this context. This has been referred to as the "Anti-Folk Theorem", Bhaskar (1998). Cyclic games push this logic to an extreme, since in cyclic games, there is absolutely no memory. Some other equilibria are possible in an extension of the Prisonners' Dilemma, because cyclic games feature sequential moves.

I will show now how an equilibrium can be sustained in a cyclic game, when there is no memory but sequential moves and a conditional transfer or a bequest.

Let us stay in the prisonners dilemma example, but let us adopt the more general matrix formulation where π^C is the cooperative payoff, π^D is the payoff of playing D when the other player plays C and thus the deviation payoff and π^P is the payoff when the player plays C when the other play D. Clearly $\pi^D > \pi^C > \pi^{NE} > \pi^P$. Let all payoffs be normalised so that the Nash-Equilibrium payoff, $\pi^{NE} = 0$.

		Player 1	
		NC	**C**
Player 2	**NC**	(π^C, π^C)	(π^P, π^D)
	C	(π^D, π^P)	$(0, 0)$

Suppose as in the main text, that some part of the old generations' consumption can be earmarked as a bequest. This implies that this resource is now only available to be left to the next generation at the end of the players life. The old player cannot consume the bequest, but he can disinherit. Every old player proposes thus the following contract: respond with cooperation to my cooperative move (dynamic). I will then bequest to you b^*. What are the decisions of the young player ? Suppose the young player decides to cooperate, his payoff will be

$$\pi^C + b^* + \delta(\pi^C - b^*),$$

if he applies the contract to his an the next generations.

If he does not cooperate he can at most obtain

$$\pi^D + \delta(\pi^C - b^*),$$

trying to deviate and implement the contract with the next generation. Thus in order for him to accept the contract, it must hold that

$$\pi^C + b^* + \delta(\pi^C - b^*) > \pi^D + \delta(\pi^C - b^*),$$

and thus simply the bequest needs to be the difference in payoffs, $b^* = \pi^D - \pi^C$. By construction, it must hold that for positive bequests, $b^* < \pi^C$, implying $\pi^D < 2\pi^C$ for this equilibrium to exist.

Does an old player need to worry that a young player will deviate ? No, since the young player (by subgame perfection) will always choose to play the contract if the above condition holds. Let us impose first the fact that parents will never bequest to a deviator and always bequest to a cooperating young. Thus, given the same payoff, parents weakly prefer the cooperation outcome. This can be seen as "revenge thinking" in a more moderate setting. Once the bequest is earmarked, parents forego it willingly. Therefore their best response to deviation of the children is to disinherit. Indeed, parents utility is not affected by the destination of the resource they cannot consume. Thus their best response to deviation of children (payoff reducing pay) is not to bequest and their best response to cooperation is to bequest. The following game structure (figure 6) illustrates entering, choices and exit of players and the game tree which cuts at the entry and exit point of the child take this best response into account.

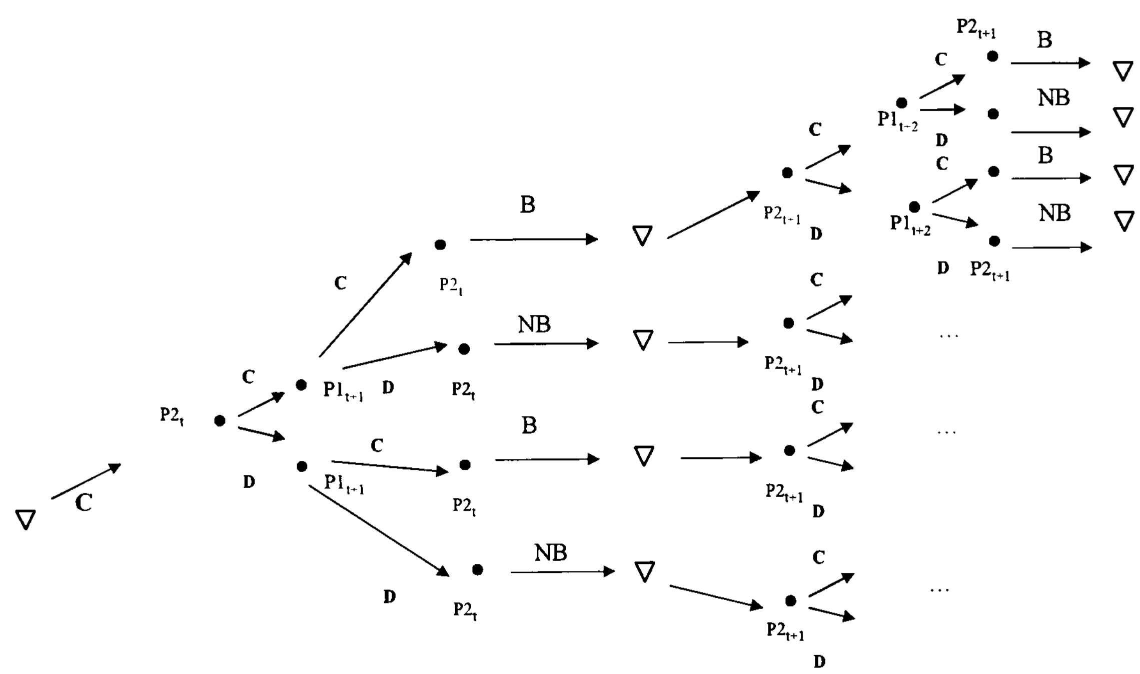


Figure A.3: The simplified overlapping Prisonners dilemma game, eliminating parents weakly dominated bequest strategies.

We depict a child (P1) in generation $t+1$ and all his possible choices given his parents t and his future children's choices $P1_{t+2}$. In order not to overcharge the picture, we only depict the first four exit points, although they are in total 16 exit points. Notice that the game is dynamic and thus *not* the usual simultaneous move game. The own choices of player $t+1$ are underlined. The payoffs for the players at the exit points are displayed in the next table.

Given parents played C			
children playing C		**children playing D**	
Move	Payoff	Move	Payoff
CCC	$\pi^C + b^* + \delta(\pi^C - b^*)$	DCC	$\pi^D + \delta(\pi^C - b^*)$
CCD	$\pi^C + b^* + \delta\pi^P$	DCD	$\pi^D + \delta\pi^P$
CDC	$\pi^C + b^* + \delta\pi^D$	DDC	$\pi^D + \delta\pi^D$
CDD	$\pi^C + b^* + \delta\pi^{NE}$	DDD	$\pi^D + \delta\pi^{NE}$
Given parents played D			
children playing C		**children playing D**	
Move	Payoff	Move	Payoff
CCC	$\pi^P + \delta(\pi^C - b^*)$	DCC	$\pi^{NE} + \delta(\pi^C - b^*)$
CCD	$\pi^P + \delta\pi^P$	DCD	$\pi^{NE} + \delta\pi^P$
CDC	$\pi^P + \delta\pi^D$	DDC	$\pi^{NE} + \delta\pi^D$
CDD	$\pi^P + \delta\pi^{NE}$	DDD	$\pi^{NE} + \delta\pi^{NE}$

As the game is dynamic, we can eliminate strategies which are not subgame perfect. Looking at the second part of the table, given parents played D, we note that children must always play D in this case. All the payoffs are indeed higher responding with D, which is of course also the Nash-best reponse in the original game. Based on this fact, we can then eliminate from the top part of the table the strategies CDC and DDC respectively, given parents played C. This is simply because D should be followed by D, and because children observe parents' move. Since by definition of our bequest $\pi^C + b^* > \pi^D$, we can see that all other payoffs are lower than CCC, thus children will always play C given that parents play C and can commit to such a contract.

The equilibrium in this game rests strongly on the ability to commit to a bequest and being able to disinherit. Thus there is a natural tension between commiting to reserve an amount of money and then perhaps not giving it. This tension is resolved by the fact that by our definition of bequest resources, these are impossible to consume. It is therefore not possible for parents to renege on their choices since they cannot gain by refusing the bequest[1].

There are two interesting and interdependent extensions of this model. Suppose a random cooperation payoff, i.e. a risky environment. Then sometimes, the parents would

[1]This simple set-up does not work in the game proposed in the main text because of the wage growth over generations: bequests also need to grow at the same rate.

not be able to commit to a bequest because of lack of income. A certain threshold needs indeed to be crossed for bequests to be applicable. Thus children will play cooperation and deviation with mixed strategies and some child labour may exist in equilibrium. A related issue to randomness is the following: Suppose income was low in the past and no family could apply the contract because no bequest resources were available. But there is now a durable increase in parents income to sustain a bequest. Then we can explain the emergence of cooperative equilibrium play. However, the counterpart of this assumption is that all generations must be convinced that future incomes will be high enough, otherwise the contract will also break down, or some fraction of non-cooperative moves may exist in a new equilibrium. Using mixed strategies, cyclic games can unfold their full potential, since they can be used to derive stationary mixed strategy equilibria. Because of constraints in space an time, we have not elaborated on this interesting property of cyclic games, but this will constitute an interesting point on a future research agenda.

Appendix B

Chapter 3

B.1 Results for the probit specification

This section explores the relationship between the NCDS grandchildren staying on at school after the minimum leaving age (16 years) or leaving school, given that we observe that they are already 16 years or older. In this set-up, there is no censoring. We can still use the control function approach in the probit case, but do not have to account for censoring. This provides also an additional robustness check on the results. We constructed a dummy variable equal to one for children staying on at school, and present some further evidence on the education transmission effects.

The results for the probit model presented in Table B.1 are different from the estimates of the censored normal model. This is also to be expected, as the decision to stay on in school after the minimum leaving age might be different from the decision to take more education. Note the relatively small effect of parents, and note that mother's education seems barely significant in the staying-on decision of children. Notice as well that the aggregate measure of human capital displays a negative effect on the staying-on decision, although not significant. This negative influence is difficult to interpret, but it is consistent with some of the results in the tables in the main text. Also the father's control function is again significant, indicating that unobserved factors enter these relationships. Notice that mother's education does play statistically and economically a marginal role in the staying-on decision, but ability measures of the mother are strongly correlated

	Dependent variable is a dummy variable: child stays on at school after age 16 Sample: oldest siblings only			
	Probit	Probit IV (control function)	Probit	Probit IV (c.f.)
Joint education (HKPARENTS)	.030 (.03)	−.284 (.30)		
Mother's education			.085 (.05)	−.146 (.63)
Father's education			.037 (.03)	−.396 (.24)
# Siblings	.022 (.11)	−.023 (.11)	.0006 (.14)	−.114 (.16)
Ability (factor)	.183 (.04)	.307 (.12)	.177 (.05)	.480 (.30)
# Siblings (CM's family)	−.033 (.01)	−.037 (.02)	−.035 (.02)	−.026 (.03)
Child female dummy, 1=female	−.434 (.08)	−.439 (.08)	−.363 (.10)	−.308 (.11)
CF aggregate		.403 (.38)		
CF Mother				.301 (.83)
CF Father				.801 (.45)
Constant	1.237 (.46)	2.256 (1.08)	1.195 (.66)	3.245 (2.46)
N	2325	2325	1434	1434

Table B.1: Second stage of the probit model using only the sample of older siblings: marginal effects

with children staying on. This seems to confirm our conjecture that it is not education itself but other factors, such as innate parental ability and perhaps parents parenting skills, that induce their children to stay longer in school.

We can compare the "oldest-sibling" results with the "all-siblings" sample, which are presented in Table B.2. We pick up very similar effects to the censored normal regression model, except again a very small coefficient on mother's education. Also, father's negative effect, as well as the positive impact of unobservables, seems only to count for one third of the original coefficient. Thus, later-born children seem to be less attached to fathers than earlier-born ones.

The results confirm that the decision to stay on at school might be different from the decision to continue to A-Levels or university. We also tried an ordered probit model of these decision levels, but the results did not change.

The IV control function approach seems to tell a consistent story when we look at the change in coefficients overall, but negative coefficients on parents raise doubts not necessarily about identification but rather about possible measurement error. Since

	Dependent variable is a dummy variable: child stays on at school after age 16 Sample: All siblings			
	Probit	Probit IV (control function)	Probit	Probit IV (c.f.)
Joint parents' education (HKPARENTS)	.054 (.03)	.039 (.15)		
Mother's education			.108 (.05)	.397 (.31)
Father's education			.042 (.02)	−.205 (.11)
# Siblings	−.165 (.04)	−.167 (.04)	−.188 (.05)	−.163 .07
Ability (factor)	.145 (.03)	.151 (.07)	.128 (.04)	.121 .14
# Siblings (parents)	−.028 (.01)	−.028 (.01)	−.024 (.01)	−.017 .01
Child female dummy, 1=female	−.389 (.06)	−.389 (.06)	−.396 (.08)	−.393 (.08)
CF aggregate		.018 (.19)		
CF Mother				−.388 (.41)
CF Father				.480 (.21)
Constant	1.086 (.37)	1.132 (.60)	1.68 (.54)	1.23 1.24
N	3634	3634	2311	2311

Table B.2: Second stage probit using all siblings: marginal effects

the coefficients change in the right direction, we might have an additional measurement error problem.

B.2 Details relating to the estimation of the censored quantile regression model with endogeneous regressors

The censored quantile regression model uses first stage estimators of the conditional quantile of the following format.

$$(\hat{\gamma}, \hat{q}) = \arg\min \sum_{l=1}^{n} K_w \left(\frac{w_i - w_l}{\delta_n} \right) \rho\alpha \left(y_l - q - g(w_i^c - w_l^c, p, \gamma) \right).$$

Here w_i represents a vector of covariates x and the instruments z of dimension r. In particular w^c are the continuous covariates of w. K_w is an r-dimensional product Kernel on the uniform distribution from $\mathcal{U} \in (-0.5, 0.5)$, and thus determines if any of the bandwidth-adjusted differences in all covariates fall outside this interval, and assigns them value zero; δ_n is a bandwidth chosen to satisfy $\delta_n = cn^{-\kappa}$, for some $c > 0$ and $\kappa \in \left(\frac{1}{2p+s}, \frac{1}{4s} \right)$ where $s = \dim(w_i^c)$, and p is the order of the polynomial with $p > \frac{3s}{2}$.

The function ρ_α is the quantile regression check function, and $g()$ represents a polynomial of order p in the differences of the continuous variables in w.

We estimate this equation on all the pairs of differences in the control variables. The estimated intercept of this equation $\hat{q}$, is then used to construct the second stage estimates, outlined in the text.

Some remarks seem important to the practicality of this estimator. We use only a multivariate polynomial of order four in the continuous variables of w although Blundell and Powell (2004) suggest a much higher order. This is however unfeasible, given that we have only around 1700 observations to fit such a model. We also choose a bandwidth that does not trim off many observations in the first stage quantile regression. The trimming of observations is problematic in the presence of dummy variables. Suppose a dummy is only informed for 10% of the observations. If we trim away 50% of all observations, the dummy might only consist of zeros. This will render the design matrix singular and stop the algorithm.

We use nonparametric first stage regressions to estimate the first stage control functions. We use a local-linear fit. The bandwidth chosen for the local-linear regression follows Blundell and Powell (2004) in setting it at $h_n = c_n n^{-\gamma}$, where $c_0 < c_n < c_0^{-1}$ for some $c_0 > 0$ and $\gamma \in \left(\frac{1}{2M}, \frac{1-2\kappa s}{4L+8}\right)$, where additionally M is the order of the second step kernel, κ is again the first step bandwidth and $L = \dim(v_i)$. While staying within the range proposed by Blundell and Powell (2004), we conduct a cross-validation check to assess the validity of the bandwidth choice. The locfit package of Loader (1997) suggests a bandwith of 2.2[1], smoothing in each window over about 23% of the sample. This is a fairly large value, and in contradiction with the theoretical values. Since these two values differ and since the bandwidth does influence the results, we prefer to apply the theoretical bandwidths.

In doing so, we implicitly make the assumption that our nonparametric control function estimate

$$\hat{v}_i = x_i - \hat{\Pi} z_i,$$

satisfies

[1] On the graph are shown the degrees of freedom, not the bandwith

$$\max \|\hat{v}_i - v_i\| = o_p(n^{-\frac{3}{8}}).$$

by virtue of including higher order terms in the first stage estimation. The choice of bandwidth is therefore determined by this consideration rather than by choosing the lowest variance bandwidth.

Finally, as the asymptotic results for the estimated variance are given for n, we need to re-scale our variances not by the full sample size (considering degrees of freedom) but by a smaller sample size. In fact, by using the approach of Blundell and Powell (2004), we articfially construct a much larger sample of size $N = \frac{n}{2}(n-1)$, which amounts in total to aproximately 1.444.150 observations in our case. In fact, we need to count how many original observations are present in the construction of our final Kernel weighted estimator and re-scale the variances accordingly. Notice as well the special way the weights enter into this estimator.

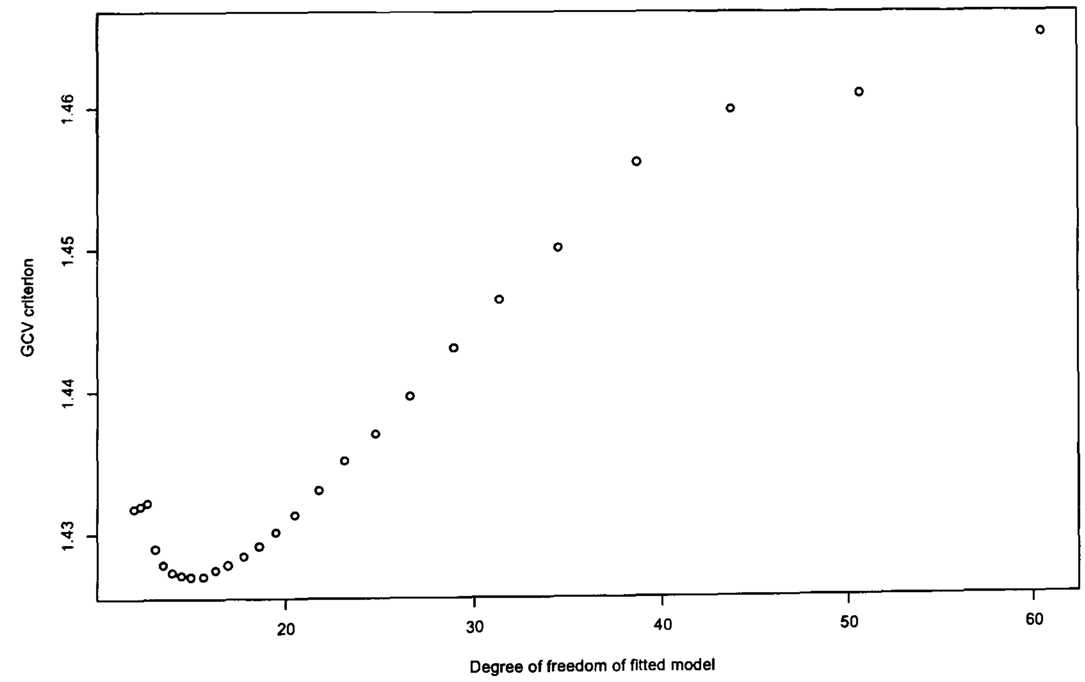


Figure B.1: Graphic illustration of the GCV-criterion for various choices of bandwidths in the *locfit* package

B.3 Additional results and summary statistics

As outlined in the text, we estimated principal components for ability and social class. For ability, we used all the available test scores of the cohort member up to age 11 (if we had used later test scores, this would be inconsistent with the identification which relies on a schooling reform implemented before age 16) and extracted one common factor, labelled ability.

For social class, we decided to also use a single factor, although with the methodology proposed by Tipping and Bishop (1999) which allows us to impute missing values. This is done via the iteration of an EM algorithm and can be executed with the STATA command *pcmav* for example. The social class variables are affected by missing values. With this method, we are able to recover most of the missing values through imputation. We are not interested in the coefficients of the social class variables: we think that these might be endogeneous as well. But we want to control for social class in the censored regressions and thus such an imputation is reasonable. Below we illustrate the correlation of the factors with actual outcomes. Note that we never used the social class variables in the estimation of the IV estimates.

The abbreviations in Table B.3 stand for mathematics and reading tests at different ages (7,11,16) and the British Social Adjustment grade (BSAG) at age 11. The latter is measured on an inverse scale to the other test results, thus the negative correlation is consistent.

Table B.4 shows a similar correlation matrix with the social class measures. Note that the imputation gains are large, since we recover nearly the whole sample of families (compared with the ability measure which is observed for the whole sample). The abbreviations in the table are *cm* for cohort member and *p* for partner. The numbers indicate the year in which social class was recorded. Note that a higher social class factor implies a **lower** social class - the factor being inversely ranked because lower social classes have higher values in the data.

	Correlation of ability factor and test scores						
	Ability	Math16	Read16	Math11	Read11	Math7	Read7
Ability	1.0000						
N	4243						
Math16	.694 .00						
N	4243						
Read16	.643 .00	.824 .00					
N	4243	4243					
Math11	.802 .00	.404 .00	.294 .00				
N	4243	4243	4243				
Read11	.744 .00	.310 .00	.295 .00	.796 .00			
N	4243	4243	4243	4243			
Math 7	.688 .00	.229 .00	.218 .00	.412 .00	.393 .00		
N	4243	4243	4243	4243	4243		
Read 7	.659 .00	.229 .00	.178 .00	.411 .00	.333 .00	.702 .00	
N	4243	4243	4243	4243	4243	4243	
Bsag 11	−.188 .00	−.176 .00	−.143 .00	−.091 .00	.056 .00	−.126 .00	−.078 .00
N	4243	4243	4243	4243	4243	4243	4243

Table B.3: The estimated principal component "Ability" and its correlation with test scores, p-values in small font below estimates

B.3.1 Robustness check: Modifying the specification

In our specification, we use as control variables the sex of the child, the inter-sibling birth-spacing intervals, the ability measure of the cohort member, the number of siblings of the mother, the number of siblings of the grandmother and cohort dummies for both grandparents. Ability is a variable that could suffer from endogeneity, as schooling and unobserved components could be correlated with ability. In order to minimise the effect of schooling, we only take ability measures up to age eleven of the cohort member, that is well in advance he or she entered post-compulsory education. Nevertheless, ability could still be correlated with the unobservables in our transmission equation and thus we should test if the results change significantly if we exclude ability. Also, we have not included measures of age of the children, as we feared that these might nearly perfectly explain the censoring mechanism and therefore bias the results. In Table B.5, we present the estimation of the two-generational model including and excluding the ability and child age measure (age and age squared of the child).

	Correlation of social class factor and social class measures							
	Social class	cm1981	cm1970	cm1980	p1970	cm1991	p1981	cm2000
Social class (estimated factor)	1.0000							
	4184							
cm1981	.4948							
	.00							
N	3529							
cm1970	.680	.232						
	.00	.00						
N	2952	2952						
cm1980	.711	.312	.966					
	.00	.00	.00					
N	2997	2997	2952					
p1970	.942	.094	.213	.164				
	.00	.00	.00	.00				
N	2231	2209	1868	1886				
cm1991	.466	.107	.293	.207	.142			
	.00	.00	.00	.00	.00			
N	3425	2951	2481	2510	1892			
p1981	.961	.071	.151	.150	.952	.124		
	.00	.00	.00	.00	.00	.00		
N	2278	2256	1904	1925	2231	1926		
cm2000	.247	.177	.333	.258	.152	.359	.105	
	.00	.00	.00	.00	.00	.00	.00	
N	3495	2942	2455	2493	1892	2916	1933	
p2000	.141	.145	.188	.155	.377	.087	.285	.155
	.00	.00	.00	.00	.00	.00	.00	.00
N	2186	1877	1556	1577	1285	1851	1307	1965

Table B.4: The estimated principal social class component and it's correlation with social class measures, p-values in small font below estimates

B.3.2 Robustness check: implementing a RD design

The attempts to identify causal effects in this paper rely extensively on schooling reforms, implemented in the UK in 1947 (grandparent) and 1973 (partners). There has been a growing trend in the education achievement of the partners and parents of cohort members over time. A split in pre- and post reform cohorts is likely to confuse the reform and the age effect. While we acknowledge that this might be of some concern for the partners of the cohort members, we argue that our results are, overall, not substantially biased. The main conclusions still apply, implying that our data shows positively biased coefficients for education when not taking endogeneity into account. First we illustrate in Figure B.2 the impact of the reform on the education achievement of grandmothers and partners of cohort members, using a RD-type design by fitting quartic polynomial in the age of partners and grandmothers respectively. One can see that while the effect is substantial in the case of grandmothers, the effect for partners is smaller although still statistically significant (estimate at the threshold: .38 with standard error .15).

Secondly we re-test the over-identified, two-generational transmission between

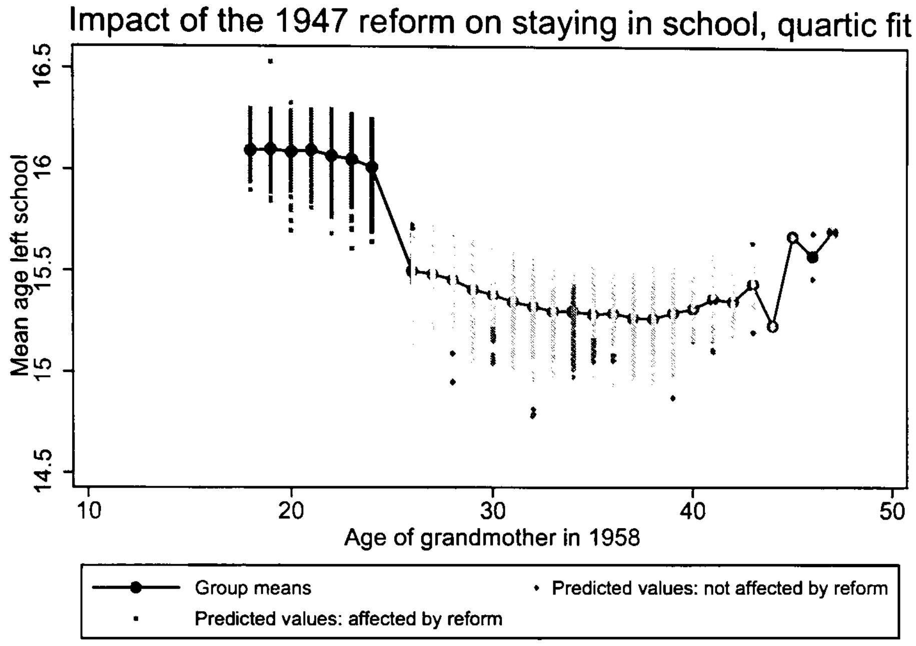
	Dependent variable: childs' education			
	Child age and ability included			
	Censored reg.	IV	Censored reg.	IV
HKPARENTS	.382 (.05)	−.684 (.45)		
Mothers' education			.434 (.08)	.354 (.59)
Fathers' education			.126 (.05)	−.791 (.40)
Control function mother				.22 (.75)
Control function father				1.720 (.74)
Control function (aggregated)		1.360 (.57)		
	No ability measure but age included			
HKPARENTS	.438 (.04)	−.325 (.41)		
Mothers' education			.524 (.08)	.542 (.42)
Fathers' education			.127 (.04)	−1.008 (.43)
Control function mother				−.038 (.57)
Control function father				2.147 (.82)
cfa		1.009 (.54)		
	No ability measure but age included			
	+ additional instrument: presence of an older brother			
HKPARENTS	.438 (.04)	−.323 (.41)		
Mothers' education			.520 (.08)	.398 (.39)
Fathers' education			.128 (.04)	−.983 (.43)
Control function mother				.154 (.52)
Control function father				2.102 (.82)
Control function (aggregated)		1.006 (.54)		

Table B.5: A robustness check: Including/Excluding age and ability measures

parents and grandparents, replacing cohort dummies by a cubic fit in age of grandparents (and all cross-products). We reproduce the main results below, and thus it appears that education coefficients are over-estimated even when modelling the trend in age with a cubic fit rather than with cohort dummies. This increases the confidence that our two-generational model is capturing at least the appropriate direction of the effect, that is smaller coefficients on education transmission in the IV models.

	Dependent variable: parents' education			
	OLS	IV	OLS	IV
FEDUC	.173 (.02)	−.419 (.50)		
MEDUC	.264 (.03)	−.066 (.25)		
HKGPAR			.434 (.03)	−.488 (.39)
Cons	9.913 (2.90)	30.041 (14.87)	−4.143 (2.49)	14.701 (8.74)
R-squared	.091238	.04	.11354	.05
N	2123	2123	1825	1825
F-test (IV first stage)	separate measures		aggregated measures	
ROSLA Father	16.51			
ROSLA Mother, Birthorder	17.27			
ROSLA Father,Birthorder,ROSLA Mother			4.85	
Overidentifiaction				
Sargan (NR^2)	1.11 $\chi^2(1)$ p= 0.29		0.324 $\chi^2(1)$ p = 0.56	

Table B.6: Grandparents two-generational effect on parents



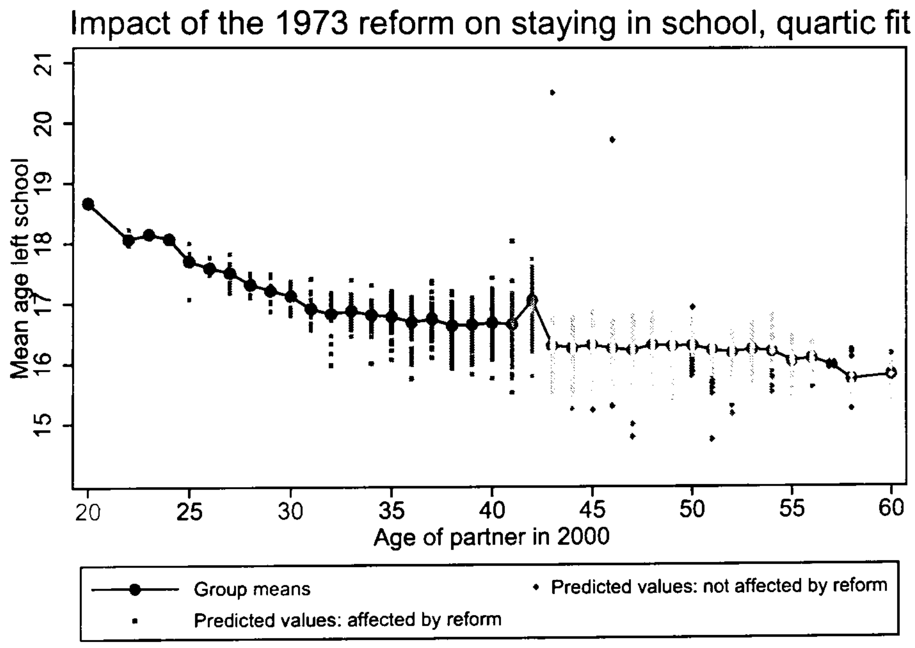


Figure B.2: The reform impact on the educational achievement (in years) of mothers (top) and partners (bottom) of the cohort member

Appendix C

Chapter 4

C.1 Model identification

In its most general form, the model reads as follows:

$$\begin{cases} Y_0 & = & f_0(X) + \boldsymbol{\alpha}_0'\boldsymbol{\eta} + \varepsilon_0, \\ Y_1 & = & f_1(X) + \boldsymbol{\alpha}_1'\boldsymbol{\eta} + \varepsilon_1, \\ \boldsymbol{W} & = & \boldsymbol{g}(X) + \boldsymbol{\beta}\boldsymbol{\eta} + \boldsymbol{U}, \\ \boldsymbol{Z} & = & \boldsymbol{h}(X) + \boldsymbol{\gamma}\boldsymbol{\eta} + \boldsymbol{V}, \end{cases}$$

where $\boldsymbol{\eta}$ is a $K \times 1$ vector of endowments, $\boldsymbol{W}$ is a $L_1 \times 1$ vector of pre-treatment outcomes, $\boldsymbol{Z}$ is a $L_2 \times 1$ vector of instruments, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $L_1 \times K$ and $L_2 \times K$ matrices of parameters.

We require the following set of assumptions:

$$\mathbb{E}\left(\varepsilon_0|X,T\right) = \mathbb{E}\left(\varepsilon_0|X\right); \quad \mathbb{E}\left(\varepsilon_1|X,T\right) = \mathbb{E}\left(\varepsilon_1|X\right), \tag{C.1}$$

$$\mathbb{E}\left(\boldsymbol{U}|X,T\right) = \mathbb{E}\left(\boldsymbol{U}|X\right), \tag{C.2}$$

$$\boldsymbol{\eta} \perp \left(\varepsilon_0, \varepsilon_1, \boldsymbol{U}, \boldsymbol{V}\right) | \quad X, T, \tag{C.3}$$

$$\boldsymbol{V} \perp \left(\varepsilon_0, \varepsilon_1, \boldsymbol{U}\right) | \quad X, T, \tag{C.4}$$

$$\text{Rank}\left(\boldsymbol{\beta}\right) = \text{Rank}\left(\boldsymbol{\gamma}\right) = \text{Rank}\left(\text{Var}(\boldsymbol{\eta}|X,T)\right) = K, \tag{C.5}$$

$$\text{Var}\left(\varepsilon_0|X,T\right) = \text{Var}\left(\varepsilon_0|X\right); \quad \text{Var}\left(\varepsilon_1|X,T\right) = \text{Var}\left(\varepsilon_1|X\right), \tag{C.6}$$

$$\text{Var}\left(\boldsymbol{U}|X,T\right) = \text{Var}\left(\boldsymbol{U}|X\right). \tag{C.7}$$

We start by selecting K linearly independent rows of $\boldsymbol{\beta}$, and K linearly independent rows of $\boldsymbol{\gamma}$. For notational simplicity we still denote the subsets of K pre-treatment outcomes as $\boldsymbol{W}$, the matrix of factor loadings as $\boldsymbol{\beta}$, and similarly for the instruments. In the submodel, Assumptions (C.1)-(C.7) are satisfied, $L_1 = L_2 = K$, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are invertible.

We have, using (C.3) and (C.4):

$$\begin{aligned}
\mathrm{Cov}\,(Y, \boldsymbol{Z}|X, T=j) &= \boldsymbol{\alpha}_j' \,\mathrm{Var}\,(\boldsymbol{\eta}|X, T=j)\,\boldsymbol{\gamma}', \\
\mathrm{Cov}\,(\boldsymbol{W}, \boldsymbol{Z}|X, T=j) &= \boldsymbol{\beta}\,\mathrm{Var}\,(\boldsymbol{\eta}|X, T=j)\,\boldsymbol{\gamma}'.
\end{aligned}$$

Hence, making use of (C.5) we obtain:

$$\begin{aligned}
\boldsymbol{\alpha}_j'\boldsymbol{\beta}^{-1} &= \boldsymbol{\alpha}_j'\mathbb{E}\left[\mathrm{Var}\,(\boldsymbol{\eta}|X, T=j)\,|T=j\right]\boldsymbol{\gamma}'\left\{\boldsymbol{\gamma}'\right\}^{-1}\left\{\mathbb{E}\left[\mathrm{Var}\,(\boldsymbol{\eta}|X, T=j)\,|T=j\right]\right\}^{-1}\boldsymbol{\beta}^{-1}, \\
&= \boldsymbol{\alpha}_j'\mathbb{E}\left[\mathrm{Var}\,(\boldsymbol{\eta}|X, T=j)\,|T=j\right]\boldsymbol{\gamma}'\left\{\boldsymbol{\beta}\mathbb{E}\left[\mathrm{Var}\,(\boldsymbol{\eta}|X, T=j)\,|T=j\right]\boldsymbol{\gamma}'\right\}^{-1} \\
&= \mathbb{E}\left[\mathrm{Cov}\,(Y, \boldsymbol{Z}|X, T=j)\,|T=j\right]\left\{\mathbb{E}\left[\mathrm{Cov}\,(\boldsymbol{W}, \boldsymbol{Z}|X, T=j)\,|T=j\right]\right\}^{-1}. \qquad \text{(C.8)}
\end{aligned}$$

Then we also have, using (C.1) and (C.2):

$$\mathbb{E}\left(Y_j - \boldsymbol{\alpha}_j'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X\right) = \mathbb{E}\left(Y_j - \boldsymbol{\alpha}_j'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X, T=j\right), \qquad \text{(C.9)}$$

from which it follows that:

$$\begin{aligned}
\Delta &= \mathbb{E}\,(Y_1 - Y_0) \\
&= \mathbb{E}\left(Y_1 - \boldsymbol{\alpha}_1'\boldsymbol{\beta}^{-1}\boldsymbol{W} - Y_0 + \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}\right) + (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0)'\,\boldsymbol{\beta}^{-1}\mathbb{E}\,(\boldsymbol{W}) \\
&= \mathbb{E}\left(\mathbb{E}\left(Y_1 - \boldsymbol{\alpha}_1'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X\right) - \mathbb{E}\left(Y_0 - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X\right)\right) + (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0)'\,\boldsymbol{\beta}^{-1}\mathbb{E}\,(\boldsymbol{W}) \\
&= \mathbb{E}\left(\mathbb{E}\left(Y_1 - \boldsymbol{\alpha}_1'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X, T=1\right) - \mathbb{E}\left(Y_0 - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X, T=0\right)\right) + (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0)'\,\boldsymbol{\beta}^{-1}\mathbb{E}\,(\boldsymbol{W}) \\
&= \mathbb{E}\left(\mathbb{E}\left(Y - \boldsymbol{\alpha}_1'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X, T=1\right) - \mathbb{E}\left(Y - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X, T=0\right)\right) + (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0)'\,\boldsymbol{\beta}^{-1}\mathbb{E}\,(\boldsymbol{W}),
\end{aligned}$$
$$\text{(C.10)}$$

where we have used (C.9) to show the fourth equality.

Similarly we have:

$$\begin{aligned}
\Delta^{TT} &= \mathbb{E}\left(Y_1 - Y_0|T=1\right) \\
&= \mathbb{E}\left(Y_1|T=1\right) - \mathbb{E}\left(Y_0 - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}|T=1\right) - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\mathbb{E}\left(\boldsymbol{W}|T=1\right) \\
&= \mathbb{E}\left(Y_1|T=1\right) - \mathbb{E}\left(\mathbb{E}\left(Y_0 - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X, T=1\right)|T=1\right) - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\mathbb{E}\left(\boldsymbol{W}|T=1\right) \\
&= \mathbb{E}\left(Y_1|T=1\right) - \mathbb{E}\left(\mathbb{E}\left(Y_0 - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X, T=0\right)|T=1\right) - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\mathbb{E}\left(\boldsymbol{W}|T=1\right) \\
&= \mathbb{E}\left(Y|T=1\right) - \mathbb{E}\left(\mathbb{E}\left(Y - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}|X, T=0\right)|T=1\right) - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\mathbb{E}\left(\boldsymbol{W}|T=1\right).
\end{aligned} \tag{C.11}$$

The formula for Δ^{TNT} is obtained by interverting $T = 1$ and $T = 0$ in this expression.

Then, the ATE on variances is obtained as follows. Let

$$\Delta^V = \mathrm{Var}(Y_1) - \mathrm{Var}(Y_0).$$

Let the index j take two possible values, $j = 1, 2$. We have:

$$\mathrm{Var}(Y_j) = \mathrm{Var}(\mathbb{E}\left(Y_j|X\right)) + \mathbb{E}\left(\mathrm{Var}(Y_j|X)\right).$$

We have, using (C.3):

$$\mathrm{Var}(\boldsymbol{W}|X) = \boldsymbol{\beta}\,\mathrm{Var}\left(\boldsymbol{\eta}|X\right)\boldsymbol{\beta}' + \mathrm{Var}\left(\boldsymbol{U}|X\right),$$

and, using (C.3) and (C.2):

$$\mathrm{Var}(\boldsymbol{W}|X, T=j) = \boldsymbol{\beta}\,\mathrm{Var}\left(\boldsymbol{\eta}|X, T=j\right)\boldsymbol{\beta}' + \mathrm{Var}\left(\boldsymbol{U}|X\right).$$

We obtain:

$$\mathrm{Var}(\boldsymbol{W}|X, T=j) - \mathrm{Var}(\boldsymbol{W}|X) = \boldsymbol{\beta}\left[\mathrm{Var}\left(\boldsymbol{\eta}|X, T=j\right) - \mathrm{Var}\left(\boldsymbol{\eta}|X\right)\right]\boldsymbol{\beta}'.$$

Likewise, making use of (C.1) we have:

$$\mathrm{Var}(Y_j|X, T=j) - \mathrm{Var}(Y_j|X) = \boldsymbol{\alpha}_j\left(\mathrm{Var}\left[\boldsymbol{\eta}|X, T=j\right) - \mathrm{Var}\left(\boldsymbol{\eta}|X\right)\right]\boldsymbol{\alpha}_j'.$$

It follows that:

$$\mathrm{Var}(Y_j|X, T=j) - \mathrm{Var}(Y_j|X) = \boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\left[\mathrm{Var}(\boldsymbol{W}|X, T=j) - \mathrm{Var}(\boldsymbol{W}|X)\right]\left\{\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\right\}'.$$

Hence the within-X component of the variance:

$$\mathbb{E}\left(\mathrm{Var}(Y_j|X)\right) = \mathbb{E}\left\{\mathrm{Var}(Y|T=j,X) - \boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\left[\mathrm{Var}(\boldsymbol{W}|X,T=j) - \mathrm{Var}(\boldsymbol{W}|X)\right]\left\{\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\right\}'\right\}. \tag{C.12}$$

Lastly, using (C.9) we obtain the between-X part, as:

$$\mathrm{Var}\left(\mathbb{E}\left(Y_j|X\right)\right) = \mathrm{Var}\left\{\mathbb{E}\left(Y|X,T=j\right) - \boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\left[\mathbb{E}\left(\boldsymbol{W}|X,T=j\right) - \mathbb{E}\left(\boldsymbol{W}|X\right)\right]\right\}. \tag{C.13}$$

The within-X and between-X components of $\mathrm{Var}\left(Y_0|T=1\right)$, that intervenes in $\Delta^{V,TT}$, are obtained similarly as:

$$\begin{aligned}\mathbb{E}\left(\mathrm{Var}(Y_0|T=1,X)|T=1\right) = \mathbb{E}\Big\{&\mathrm{Var}(Y|T=0,X) \\ &-\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\left[\mathrm{Var}(\boldsymbol{W}|X,T=0) - \mathrm{Var}(\boldsymbol{W}|X,T=1)\right]\left\{\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\right\}'\Big|T=1\Big\},\end{aligned} \tag{C.14}$$

and

$$\begin{aligned}\mathrm{Var}\left(\mathbb{E}\left(Y_0|X,T=1\right)|T=1\right) = \mathrm{Var}\Big\{&\mathbb{E}\left(Y|X,T=0\right) \\ &-\boldsymbol{\alpha}_j\boldsymbol{\beta}^{-1}\left[\mathbb{E}\left(\boldsymbol{W}|X,T=0\right) - \mathbb{E}\left(\boldsymbol{W}|X,T=1\right)\right]\Big|T=1\Big\}.\end{aligned} \tag{C.15}$$

C.2 Parameter estimation

We rewrite (C.8) as:

$$\begin{aligned}\boldsymbol{\alpha}_j'\boldsymbol{\beta}^{-1} = &\mathrm{Cov}\left[Y - \mathbb{E}\left(Y|X,T=j\right), \boldsymbol{Z} - \mathbb{E}\left(\boldsymbol{Z}|X,T=j\right)|T=j\right] \times \\ &\left\{\mathrm{Cov}\left[\boldsymbol{W} - \mathbb{E}\left(\boldsymbol{W}|X,T=j\right), \boldsymbol{Z} - \mathbb{E}\left(\boldsymbol{Z}|X,T=j\right)|T=j\right]\right\}^{-1}.\end{aligned} \tag{C.16}$$

Then, we write the ATE and ATT estimands as:

$$\Delta = \mathbb{E}\left(\frac{T\left(Y - \boldsymbol{\alpha}_1'\boldsymbol{\beta}^{-1}\boldsymbol{W}\right)}{\pi(X)} - \frac{(1-T)\left(Y - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}\right)}{1-\pi(X)}\right) + \left(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0\right)'\boldsymbol{\beta}^{-1}\mathbb{E}\left(\boldsymbol{W}\right), \tag{C.17}$$

and:

$$\Delta^{TT} = \mathbb{E}\left(Y|T=1\right) - \mathbb{E}\left(\frac{\pi(X)}{\mathrm{P}(T=1)}\frac{(1-T)\left(Y-\boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\boldsymbol{W}\right)}{1-\pi(X)}\right) - \boldsymbol{\alpha}_0'\boldsymbol{\beta}^{-1}\mathbb{E}\left(\boldsymbol{W}|T=1\right). \tag{C.18}$$

To derive these expressions, see Hirano et al. (2003).

Lastly, we estimate the ATE on variances on the basis of (C.12) and (C.13), replacing conditional expectations by linear projections. We proceed similarly for the ATT on variances.

	Frequency	Percentage	Cumulative
Non Lea school	6,287	38.78	38.78
Comprehensive	5,895	36.36	75.15
Grammar	1,294	7.98	83.13
Secondary modern	2,582	15.93	99.06
Technical	65	0.40	99.46
All-age school	1	0.01	99.46
Day special school	1	0.01	99.47
Other	85	0.52	99.99
Res./Esn. school	1	0.01	100.00
Total	16,211	100.00	

Table C.1: The raw distribution of schools at age 16

C.3 Data Appendix

C.3.1 Construction of samples

Samples We construct various samples from the NCDS to estimate our model. Here we justify the samples selected and also explain how they relate to other samples used in the literature. It is useful to examine the raw data (see Table C.1).

Note that all authors only keep either comprehensive, grammar or secondary modern students. All others (notably private schools, here denoted "non-LEA") are dropped. Notice as well that the "non-lea school" reflect also missing values, since there is no other category to code missing values in this question.

This gives the distribution of comprehensive versus selective pupils at age 16 in 1974.

The information in 1969 is given by three variables. First of all, one variable contains the response of the teacher to the question: "What type of school is the NCDS child likely to attend in the next year ?" The answer can be 1) non-LEA (Local Education Athority) school, 2) selective school, 3) maintained (comprehensive) school.

Another variable in 1969 indicates the year a school went comprehensive. Yet another variable indicates the number of years a child spent in a given school, recorded in 1974. This last variable allows us to form three groups of pupils. Those who stayed five years or more in a comprehensive or a selective school system. We call this sample of pupils sample 0, our basic sample. Sample 0 gives a close approximation to what

Kerckhoff (1986) and Galindo-Rueda and Vignoles (2005) used.

We also construct a subsample keeping only schools in either purely selective LEA's or purely comprehensive LEA's. We use the list in Manning and Pischke (2006) and select only those LEA's from the NCDS. This suggestion was made to avoid the cream-skimming effect, i.e. able pupils still attending a grammar school if in there are still grammar schools present within their catchment area (i.e. grammar and comprehensive schools coexist in one LEA).

In addition, we keep in a sample we call "sample0GS" only those pupils who have been at either grammar or secondary modern school for five years before 1974 to evaulate the effect of attending a grammar vs. a secondary modern school.

Finally, we construct a sample in order to control for LEA fixed effects. Some LEA's have only a observations, and there might be for example, only grammar or only secondary modern students. This presents a problem in estimation, since the fixed effect cannot be estimated in these cases. We decided to drop these problematic LEA's, and we retain a sample of LEA's where we at least observe people in the three systems, which we call "sample 0 mixed".

Political variables In order to match local election results, we need to map constituency information into Local Education Authorities (LEA). We only have information on constituency level election results for the period. There can be an overlap between sets of constituencies and LEA's in the following sense. LEA's are larger administrative units than constituencies, since they are bodies of the local authorities and these may regroup one or more constituencies. LEA's are defined as a collection of electoral wards and these electoral wards may or may not coincide with the boundaries of a set of constituencies. Furthermore constituency boundaries are subject to change in the period we are considering. Since we have no information on which electoral wards made up an LEA exactly, and we do not have access to local election results, we cannot pin down precisely what happened at the local level. Instead, we tried our best to match constituencies to LEA's, that is we checked with various sources, such as the Constituency boundary commission and local authority websites to establish which constituencies belonged to which LEA's in the end of the 1970's. A majority of constituencies can be aggregated into an LEA. Some constituencies have some electoral wards in one LEA and other wards in the other LEA because of redrafting of boundaries. In this case, we aggregate the results using the "redrafted constituency" results for both LEA's. The NCDS has only confidential information on constituencies, so a first step was to assign each constituency an LEA identifier. Several constituencies can fall within one LEA, sometimes just a single constituency. Then we aggregated the results taking simple means to the LEA level.

A more satisfactory solution to this problem was provided by the matched LEA-variable data, described below. In this data a variable indicated the proportion of Labour control in the period 1957-1970. We use this variable in the IV regressions, the results did not change, no matter which variable we used.

LEA variables We matched a set of LEA-wide variables, collected in a special dataset provide by the Data Archive[1] to the NCDS data. These variables contained useful information on the proportions of pupils enrolled in different school types within the LEA in two time points, 1967 and 1972. It provides furthermore some political variables, namely the proportion of years the LEA was labour-controlled, and it provides measures of school quality, such as teachers salaries, total cost per pupil and indebtedness at the LEA level.

Missing values Missing values are of some concern, especially in longitudinal studies, where there is attrition. The NCDS is no exception in this respect, and if we did not address the missing values problem, we would loose a part of our data.

First we checked roughly if the missing values are missing at random. For this, we replace missing values with the sample means of the variables and run some of our regression equations. If the coefficients do not change drastically from the specification before, we conclude that the missing values are at random. We find very little change in our estimates when we do this. In a next step, we replace the missing variables by dummy variables, taking the value 0 if the variable is not missing and 1 if it is missing. Using this method, we preserve most of our sample intact.

[1]Effect of LEA resources and policies on educational attainment, 1972-1974, SN199

Additional education measures used The NCDS has a "years of education" measure as a proxy for educational achievement but a wealth of individual qualifications obtained. We created a 6-way classification of academic and vocational achievement (see Jenkins and Sabates (2007)). For this purpose we scanned the NCDS qualification variables in the 1991, 2000 and 2004 sweep for the highest qualification a cohort member has achieved. For each sweep, we have academic, vocational and the highest qualification (the maximum of academic or vocational). We have thus 9 variables, these relate to total qualification level up to the year of the sweep. The education levels correspond to the following classification, which in turn corresponds also to the ranking in the National qualifications framework.

Level	Vocational	Academic
0	**no qualifications**	
1	National Vocational Qualification Level 1	no formal education
	Royal Society of Arts (RSA) Level 1	CSE/O-Levels
	GNVQ (Foundation)/Other GNVQ	at grades 2 to 5 (or D-G)
	Pitmans Level 1	
	HGV license	
	Other vocational qualifications not specified	
2	National Vocational Qualifications (NVQ) level 2	GCSE grade A*-C
	Pitmans level 2	O-Levels grade A-C
	RSA Level first diploma (2)	O-levels D-E or CSE grade 1
	Apprenticeships	
	City and Guilds qualification 1 and 2	
	GNVQ (Intermediate)	
	Diploma of TEC/BEC/SCOTEC/SCOTBEC	
3	National Vocational Qualifications (NVQ) level 3	
	City and Guils qualifications Level 3 (final)	A-level
	RSA Advanced Diploma	AS-Level
4	National Vocational Qualifications (NVQ) level 4	
	Higher National Certificate (HNC,HND)	Degree, Higher education Diploma
	also HND TEC/BEC/SCOTEC/SCOTBEC	
	Nursing Paramedic	
	RSA Higher Diploma	
	Other teacher training qualification	
	City and Guilds Part 4 Career Ext/Full Tech	
5	National Vocational Qualifications (NVQ) level 5	
	Post graduate certificate of education	Higher degree
	Professional degree qualifications	

Table C.2: Table of equivalence between vocational, academic and the 6-Level Classification

In our analysis, we used only the variables relating to the highest qualifications obtained until 2004, thus total qualifications obtained up to 2004.

Different specifications used The specifications labelled (1), (2) and (3) used in the ATE/ATE variance tables are as follows:

- (1) is a specification using parental and social class characteristics and a child female dummy. The variables included are: Fathers and mothers education, number of older siblings, Social class categorisation defined from all the social class measures available in the 1958, 1965 and 1969 survey rounds. Measures of fathers and mothers income (only banded/categorized income measures are available in NCDS) are included as well as missing variable indicators as mentioned in the Missing variables section.

- (2) is specification (1) augmented by: if parents want child to stay at school after minimum age (1965), if mother is working in 1965, number of children in child's lass (1965), number of schools attended (1969), pupil teacher ratio (1969), if child in junior school in 1969, sex of the headteacher (1969), Age of the main school buildings (1969), dummies for an ability streamed class in 1969, The proportion of unemployed and sick in the census ED, the proportion of mining workers in the census ED, the proportion of working mothers (census ED), the proportion of skilled, semi-skilled, managerial and unskilled workers in census ED, the proportion of households with indoor WC in census ED, the proportion of immigrants in census ED, and the proportion of labour voters in the 1970's general election, percentage of Labour votes in the local area, regional dummies for the Government Office Regions in England and Wales.

- (3) is (2) with LEA dummies, but without regional dummies and the precentage of Labour votes in the local area variable. Note that specification (3) is used only in conjunction with the sample of mixed LEA's, as described in the section on Samples in this appendix.

Notice that we only include variables which intervene before or at the date of treatment. We fear that any variables in 1974 might suffer from endogeneity bias.

C.3.2 Variable statistics

The following statistics are given for sample 0.

Table C.3: Family and child characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
Female dummy	0.494	0.5	0	1	6870
Mathematics score at 16	12.766	6.747	0	31	6472
Reading score at 16	25.744	6.538	1	35	6496
Mathematics score at 11	16.987	10.077	0	40	6070
Reading score at 11	16.3	6.012	0	35	6070
Verbal score at 11	22.747	9.125	0	40	6072
BSAG score at 11	7.793	8.414	0	57	6075
Mathematics score at 7	5.246	2.418	0	10	6132
Reading score at 7	23.662	6.691	0	30	6154
Average weekly net income	4.657	0.585	2.187	6.919	1644
1981 log weekly net wage in 1981 prices	4.211	0.359	1.609	5.106	3712
1991 log weekly net wage in 1981 prices	4.425	0.721	1.329	7.351	3036
2000 log weekly net wage in 1981 prices	4.487	0.787	-0.863	7.469	3484
2004 log weekly net wage in 1981 prices	4.76	0.68	1.678	7.98	3072
Age when left full-time continuous education	17.303	2.24	14	30	4866
Academic qualification level (up to 2004)	1.896	1.346	0	5	5675
Vocational qualification level (up to 2004)	1.569	1.525	0	5	5500
Total qualification level (up to 2004)	2.355	1.36	0	5	5694

Table C.4: Family and child characteristics - continued

Variable	Mean	Std. Dev.	Min.	Max.	N
Fathers completed education, years	3.086	2.226	0	11	6870
Mothers completed education, years	3.104	1.98	0	11	6870
Total number of older siblings (girls/boys) of NCDS child	0.366	0.616	0	8	6870
Do you wish you had left school earlier, NCDS child answer	0.234	0.424	0	1	6870
Age you like to leave school, NCDS child answer	1.346	0.940	0	3	6870
Will child benefit staying on ? Teacher answer	0.42	0.494	0	1	6870
How long (in years) has child been at this school	5	0	5	5	6870
Total school attendance	83.355	24.933	0	100	6870
Child has a room to study	0.835	0.371	0	1	6870
Natural father takes care of child in 1974	0.712	0.453	0	1	6870
Natural mother takes care of child in 1974	0.771	0.42	0	1	6870
Father born in British Isles	0.903	0.295	0	1	5828
Mother born overseas	0.03	0.169	0	1	5828
Mother born in British Isles	0.918	0.274	0	1	5828
Child has no father figure, 1965	0.025	0.156	0	1	6085
Child has no mother figure, 1965	0.003	0.053	0	1	6085
Natural mother in household, 1965	0.002	0.048	0	1	6085
Natural mother in household at age 7	0.013	0.112	0	1	6085
Persons per room in 1974	1.091	0.782	0	4	6870
Accomodation type in 1974	1.632	0.894	1	6	5543
Social Class 1958 - I	0.107	0.309	0	1	6870
Social Class 1958 - II	0.091	0.288	0	1	6870
Social Class 1958 - III	0.478	0.5	0	1	6870
Social Class 1958 - IV	0.115	0.319	0	1	6870
Social Class 1965 - I	0.038	0.192	0	1	6870
Social Class 1965 - II	0.115	0.319	0	1	6870
Social Class 1965 - III	0.088	0.284	0	1	6870
Social Class 1965 - IV	0.408	0.492	0	1	6870
Social Class 1965 - V	0.014	0.119	0	1	6870
Social Class 1965 - VI	0.141	0.348	0	1	6870
Combined social class measure					
Proportion in Class I	0.045	0.206	0	1	6870
Proportion in Class II	0.17	0.375	0	1	6870
Proportion in Class III	0.526	0.499	0	1	6870
Proportion in Class IV	0.181	0.385	0	1	6870
Proportion in Class V	0.052	0.222	0	1	6870
Fathers and mothers income categories					
Father					
Category I	0.08	0.272	0	1	6870
Category II	0.248	0.432	0	1	6870
Category III	0.162	0.368	0	1	6870
Category IV	0.101	0.301	0	1	6870
Mother					
Category I	0.444	0.497	0	1	6870
Category II	0.044	0.205	0	1	6870

Table C.5: School characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
Enrolled in a grammar school 1974	0.148	0.355	0	1	6870
Enrolled in a secondary modern school in 1974	0.29	0.454	0	1	6870
Years spent in comprehensive between 1969 and 1974	0.667	0.471	0	1	3865
Proportion of LEA comprehensive in 1974	59.02	28.064	0	99.900	6870
Comprehensive school formed out of a Grammar school	0.142	0.349	0	1	3825
Comprehensive school formed out of secondary modern	0.303	0.46	0	1	3825
Comprehensive school purpose built	0.194	0.396	0	1	3825
Comprehensive school formed by amalgamtion	0.329	0.47	0	1	3825
Child is not in an ability streamed class, 1969	0.561	0.496	0	1	6870
Child is in low ability streamed class, 1969	0.085	0.279	0	1	6870
Child is in average ability streamed class, 1969	0.093	0.291	0	1	6870
Child is in high ability streamed class, 1969	0.133	0.34	0	1	6870
Number of schools attended since age of 5, 1965	1.084	0.681	0	11	6870
In infant school in 1965	0.49	0.5	0	1	6870
In junior and infant school in 1965	0.373	0.484	0	1	6870
In other school, 1969	0.012	0.108	0	1	6870
Free school meals at school, 1969	0.076	0.264	0	1	6870
Sex of headteacher, 1969	0.15	0.357	0	1	6870
Age of the main school buildings, 1969	40.036	37.984	0	193	6870

Table C.6: Local area and political variables

Variable	Mean	Std. Dev.	Min.	Max.	N
RE161	0.145	0.352	0	1	687
RE162	0.084	0.278	0	1	687
RE163	0.081	0.273	0	1	687
RE164	0.087	0.282	0	1	687
RE165	0.098	0.297	0	1	687
RE166	0.188	0.39	0	1	687
RE167	0.068	0.251	0	1	687
RE168	0.065	0.247	0	1	687
RE169	0.121	0.326	0	1	687
Matched census statistics					
% Unemployed and sick, 1971 Census ED	2.635	4.996	0	60	687
% Married women working, 1971 Census ED	21.102	23.828	0	80	687
% Mining & Manuf. Employment, 1971 Census ED	20.869	23.949	0	100	687
% Managment Employment, 1971 Census ED	6.514	10.902	0	78.570	687
% Skilled manual Employment, 1971 Census ED	15.622	17.615	0	80	687
% Semi-skilled manual Employment, 1971 Census ED	10.272	12.597	0	71.42	687
% owner occupied, 1971 Census ED	45.364	34.409	0	111.11	364
% council tenants, 1971 Census ED	38.857	39.255	0	103.33	364
Persons per room, 1971 Census ED	0.327	0.319	0	1.1	687
Households lacking inside WC, 1971 Census ED	5.38	13.106	0	97.72	687
% New commonwealth immigrants, 1971 Census ED	0.879	4.008	0	57.28	687
Matched political variables to local authority level					
Labour win in Local Authority	0.29	0.101	0	0.509	687
Conservative win in Local Authority	0.352	0.067	0.077	0.448	654
Liberal win in Local Authority	0.06	0.049	0	0.316	654
Matched data on LEA resources, 1967-1972					
% 13 year olds in comprehensives, 1967	12.489	19.067	0	99.8	67
% 13 year olds in comprehensives, 1972	29.252	26.819	0	99.900	67
Percentage of years 1958-1969 that LA was labour controlled	31.156	34.8	0	100	67

C.4 Complementary tables and figures

	First stage	IV Dep. variable Maths score at 16	First stage	IV Dep. variable Maths score at 11
Comprehensive LEA	Dep. var	−1.12 (.54)	Dep. var	−3.074 (1.01)
Labour control	.009 (.001)		.009 (.001)	
Ability at 11 (factor)	−.012 (.006)	2.423 (.12)		
Ability at 7 (factor)			−.004 (.006)	4.917 (.23)
Female dummy	−.019 (.014)	−1.85 (.30)	−.020 (.013)	−.69 (.55)
Father's education	−.008 (.004)	.33 (.24)	−.007 (.004)	.51 (.14)
Mother's education	.009 (.008)	.178 (.16)	.006 (.008)	.24 (.40)
Constant	.280 (.25)	12.67 (2.08)	.27 (.217)	8.80 (4.18)
R-squared	.689	.577	.69	.52
N	1168	1094	1257	1168

Table C.7: Instrumental variable regression using political colour of the LEA as instrument for the indicator of being in a comprehensive LEA

	Maths 16	Years education	Academic qual.	Vocational qual.	Log wage 2004	Average log wage
One-dimensional η						
Grammar	.23	.01	.01	.68	.13	
Secondary Modern	.72	.01	.04	.08	.68	
Two-dimensional η						
Grammar	.49	.87	.92	.24	.9	.93
Secondary Modern	.66	.98	.96	.82	.74	.38

Table C.8: P-values of the test of overidentifying restrictions, $K = 1$ and 2. Specification (3). for Grammar vs. Secondary Modern Schools

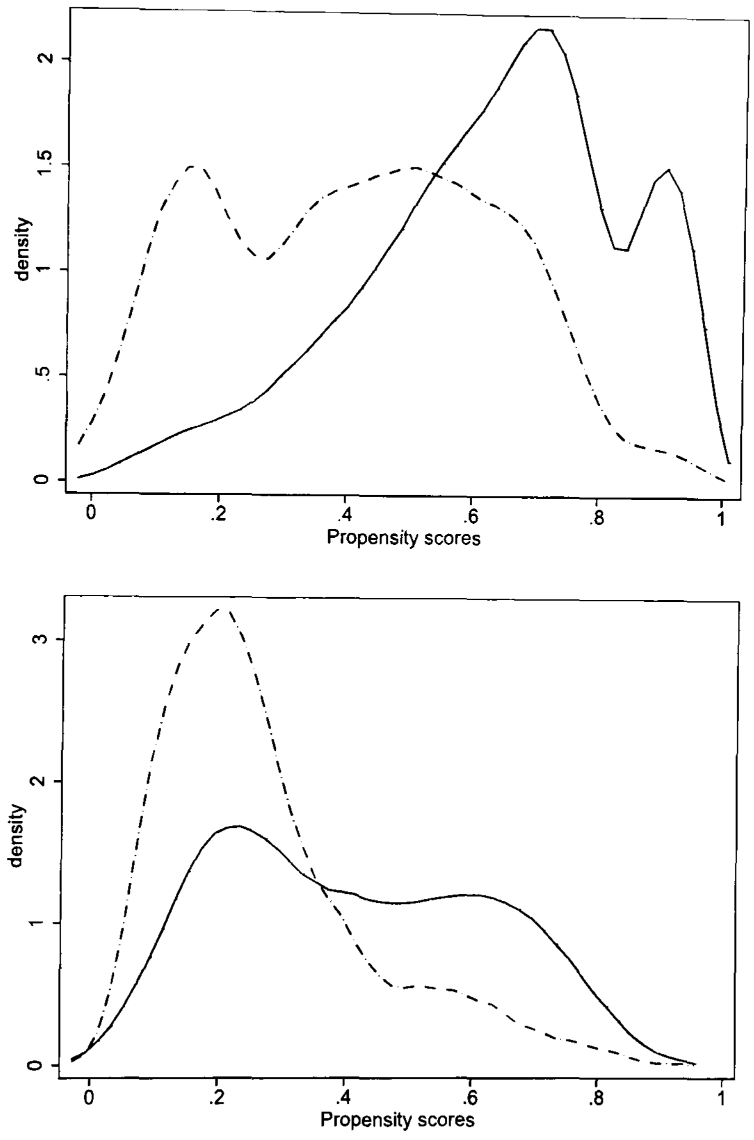


Figure C.1: The estimated propensity scores of attending a comprehensive vs a selective school (up), and a grammar vs a secondary modern school (bottom). In both panels the density for the treated group is given in solid line.

C.5 The comprehensive policy in the UK and mobility between LEA's

This part of the appendix is designed to retrace the reform of the British education system: the implementation of comprehensive schools in the period 1945-1975. Our dataset covers the period 1958-1974. It is important to be aware of the details of the comprehensive reform, since this leads to the modelling choices we made. This section is divided into two parts. We describe in a first part the comprehensive reform in greater detail. In a second part, we address a potential misspecification issue of our model, related to this reform: the mobility of families. Was there moving to selective areas after the reform was implemented? Mobility is important if children and parents move as a consequence of the reform. This would yield very different intakes for comprehensive and selective areas and it would be difficult to compare them without taking the possible endogeneity of moving into account. Thus, we illustrate in this second part the decision of parents to move and will argue that it was not linked with moving to a less comprehensive area.

C.5.1 The comprehensive reform

This subsection will draw extensively from a volume of the Oxford Review of Education, published in September 2002, "A century of local education authorities".

The postwar English school system was either bi-partite or tri-partite. It consisted mainly of grammar and secondary modern schools in the bipartite version, and to a smaller extend of grammar, secondary modern and so-called technical schools in the tripartite arrangement. Both variants were part of a 'national system, locally administered', Chitty (2002). The linkage between government, local authorities and schools should provide "checks and balances" and ensure no single element would acquire excessive power. The 1944 Education Act left indeed a large margin of autonomy to LEA's. It was for them to decide on curricula and the implementation of secondary education, Crook (2002). Often, the central government was not aware of the exact curriculum choices, this being implemented by individual schools and headmasters, Chitty (2002).

Already in 1947 some 54 LEA's were making plans for establishing non-selective schools in their area, Simon (1991). While a majority of LEA's maintained their existing grammar schools, some decided to provide some new form of secondary education, Peterson (1965). The term comprehensive schools was used instead of the term "multilateral" schools. Haydn (2004) argues that a movement for multilateral schools, composed of socialist politicians and union officials, is identifiable from the 1920's onwards. Some multiateral schools existed already before the great reform, purpose built for children aged 11-18, Crook (2002), in the aftermath of the Second World War. Kerckhoff (1986) states that comprehensive schools were, in the beginning of their existence, often thought as being viable only if they were big enough, needing to accommodate at least some 1500 pupils. This conception, so Crook (2002) says, was "to be substantially revised over the next 30 years".

Crook (2002) argues that "the post-Second World War drive for comprehensive education was a grass roots initiative". This is confirmed by Kerckhoff (1986).

> The fact is that there has been a growing local movement against the eleven-plus examination and all that it implies. The movement has not been politically inspired or imposed by the Centre. It has been a spontaneous growth at the grassroots of education, leading to the widespread conviction that separation is an offence against the child as well as a brake on social and economic progress. (p.28)

Crook (2002) and Chitty (2002) describe some LEA's experimenting with non-selective schools well before the Crosland Circular implemented this option officially. In the 1950's the West-Riding and Anglesey went fully comprehensive, Crook (2002). In urban areas the authorities saw comprehensivisation as adequate in areas where wartime bomb damage was large. Peterson (1965) describes fears by local authorities, that in a period of housing shortage and insufficient construction, additional resources should be devoted to purpose built comprehensives instead of using the existing stock of buildings. Thus Bristol, Coventry and North London, Crook (2002), experimented first with comprehensive schools, due to the increased destruction they had faced during the War.

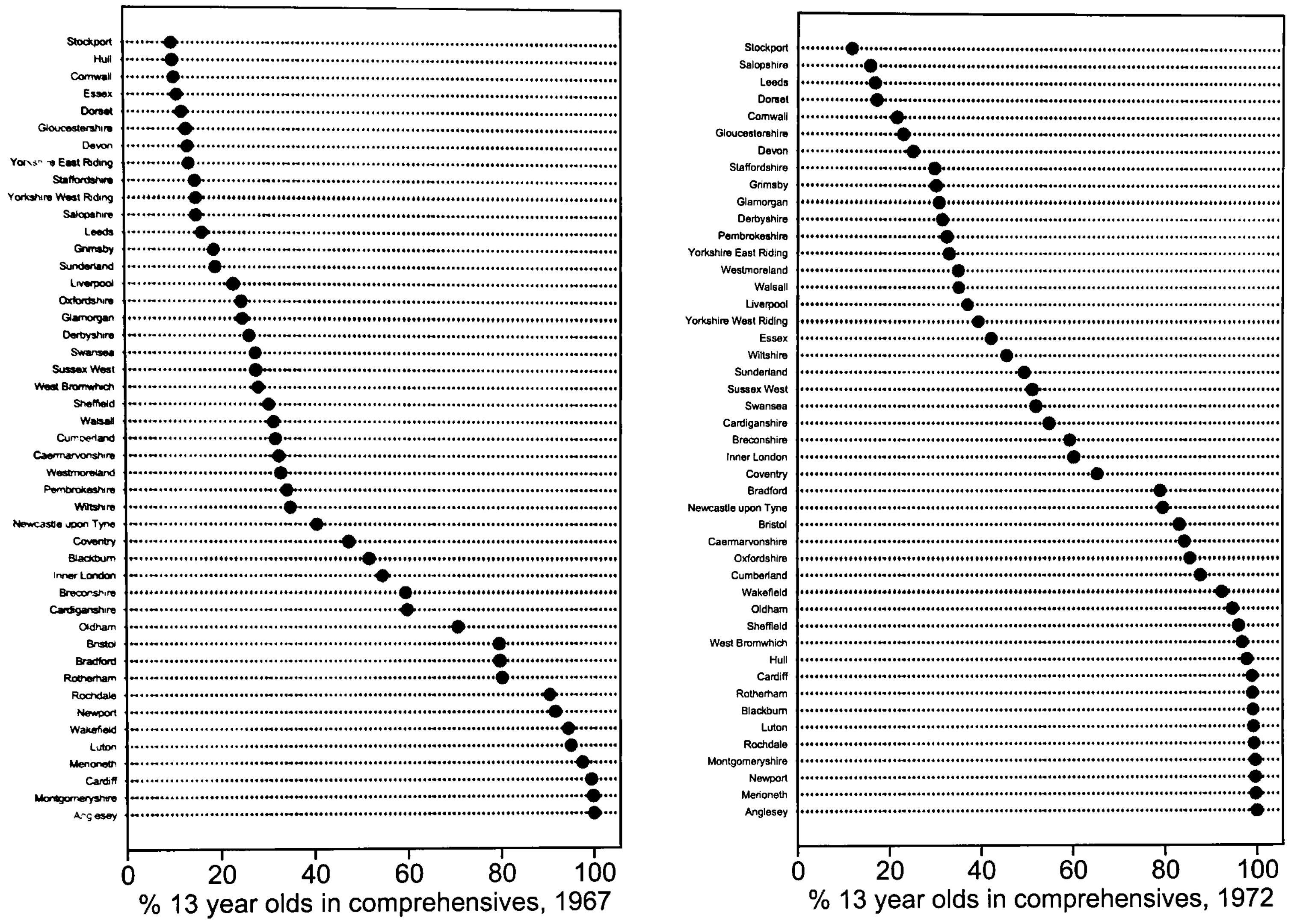


Figure C.2: Comprehensivisation at different speeds: proportion of thirteen year olds in comprehensive schools by LEA, plotting only those LEA's which had a positive proportion of comprehensive pupils in 1967

In London, comprehensive schools were established in former senior and central school buildings, Crook (2002), but these did not tempt parents to opt out of the grammar school system. When the London LEA wanted to abolish selective schools in this area, they were stopped by the newly elected conservative majority in 1951, Peterson (1965). Even though some former grammar schools became comprehensives, the 11-plus exam still continued to operate in London and denied these schools a comprehensive intake. Crook (2002) cites Bristol LEA, which established 14 non-selective secondary schools between 1954 and 1963, serving working class housing estates. Bristol LEA also continued to allow the LEA to purchase direct grant places for pupils up until 1965. Also Coventry, seems according to Crook (2002), to have kept open all of their grammar schools along the re-organisation. This was obviously a drawback of the policy, since parents could still apply to the LEA to have their children sent to a grammar school within the local catchment area.

An increased awareness that psychometric testing was flawed, Yates (1971), and that success on the eleven plus exam was influenced by social background and the inputs of parents, the pressure on LEA's increased for comprehensive reorganisation. Some LEA's, starting with the West Riding of Yorkshire, introduced what was known as the "Thorne scheme". This scheme consisted of abolishing the eleven-plus exam and allocating to each primary school a fixed number of grammar school places, which eliminated tests. The ablest pupils were then to go to Grammar school based on teachers recommendations, whereas tests were only used for borderline cases, Peterson (1965). By 1964, two thirds of the West Riding adopted this scheme, and some other LEA's were to follow, Crook (2002).

The election victories of Labour in 1963 (local elections) and 1964 (general election) became one of the motors of comprehensive re-organisation. Even though Macmillan's educational minister, Edward Boyle, was sympathetic to a comprehensive reform without closing Grammar schools, the real impetus to comprehensive organisation was provided by the Labour governments 1964 general election victory under Harold Wilson. The outline of the Crosland Circular was produced relatively quickly, but the cabinet did not enact legislation at this point, Crook (2002). Instead, a circular was sent to all LEA's requesting, Crook (2002), them to submit plans for comprehensive re-organisation. There seems to have been an extensive debate in the Department of Education and Science "whether the proposed Circular should *require* or *request* the local authorities to prepare plans for comprehensive reorganisation", Chitty (2002). Anthony Crosland seemed to have opted for the wording *request*, following the established system of local autonomy in implementation of reforms. Six methods for implementing the reform were announced in the Circular, Yates (1971). These were, citing from Yates (1971):

1. All through comprehensives, age range 11-18

2. A two-tier system, all pupils attending a junior comprehensive from 11-13 years,

some moving on at age 13-14 to a senior school, the remainder staying behind

3. A two-tier system, all pupils attending a junior comprehensive and some moving on at age 13-14 to a senior school, the remainder staying behind

4. A two-tier system, in which all pupils attend a junior comprehensive school and then, at the age of 13 or 14, all have the choice between a senior school catering for those who propose to leave when they reach the statutory school leaving age and one which provides for those who wish to continue education at this stage

5. Comprehensive schools with an age range from 11-16 combined with sixth form colleges for pupils over 16

6. Middle schools with an age range from 8-9 to 12-13, followed by comprehensive schools with an age range from 12/13-18.

According to Crook (2002) most of the large, Labour controlled LEA's adopted all-through comprehensives, often however funding and long term planning security were lacking to implement these plans quickly. "Compromises inevitably took the form of interim arrangements, the phased abolition of the 11+ and school amalgamation that created comprehensive schools on two or more sites", Crook (2002) p.252.

Leicestershire drafted a plan for a two-tier model, delaying selection at ages 13-14 and involved guided parental choice at transfer age to make sure Grammar schools would not be populated by the socially advantaged and avoiding the relative disadvantage experienced by girls. This was essentially done to use existing building resources appropriately, Peterson (1965).

Some rural LEA'S defied openly the plans for reorganisation, Crook (2002), and were leaning towards the Leicestershire model. Richmond-upon-Thames merely renovated, enlarged and then renamed its secondary modern premises as "comprehensives".

The Plowden report (1967) advocated middle schools and transfers at the ages of 8 and 12, which was particularily attractive to Conservative controlled authorities, "seeking to abolish the 11+ without creating 'monster' comprehensives", Crook (2002)

Despite conservative resistance by the secretary of State for Education, Margaret

Thatcher, more than 1400 comprehensive schools (including middle schools) were established between 1970 and 1974.

Support for comprehensive re-organisation came mainly from big cities and Labour dominated areas. But Crook (2002) mentions that support was not unanimous, especially because many of the Socialist politicans benefited themselves from a Grammar school education. Peterson (1965) identified similar resistance to the re-organisation of secondary education after the War, as "leading citizens were former pupils of grammar schools, regarding them as elite schools, and fought bitterly against any plan to absorb them in a new type of common school" p.161.

The parents supporting the comprehensive movement appear not to be very different from the parents in favour of the selective education system. In general, re-organisation was a bottom-up process, according to Crook(2002), where parents and LEA's where the main stakeholders. Also, resistance to comprehensive reorgansation was often focused on the practicality of the reform at the local level. Crook(2002) gives the example of defining new catchment areas or the need to bus children to larger comprehensive schools.

C.5.2 Mobility and comprehensive re-organisation

Mobility of parents caused by the comprehensive reorganisation is a serious concern in our analysis. If parents could chose to move to an area of their choice, richer parents would be able to re-locate to better areas which still maintain selection. Yates (1971) analyses the education system as follows: he argues that richer parents chose, in general, the more prestigious private schools for their children. Thus, his argument, "comprehensive re-organisation for the remainder could considerbaly lead to grosser inequalities than at present", Yates (1971, p.58). He asserts that richer parents will anyways chose private schools if they can afford them, and that grammar schools serve the able, from middle to low class background. Thus, we need at least to establish two results for our analysis to be unbiased by factors of moving. First we need to analyse which were the determinants of moving and if they were related to comprehensive re-organisation within the local area. Second, we need to be sure that there was no additional effect on private schools, for example was the comprehensive re-organisation pushing parents to enrol their children in private schools ?

In order to understand the mobility issue better, recall that the English and Welsh education system is based on catchment areas. As Haydn (2004) states, LEA's had to ensure a balanced social intake for each school, so that schools would comprise all different ability groups and social classes. In general pupils would be allocated to a specific secondary school by the LEA. There are the above described exceptions, the coexistence of comprehensives and grammar schools and some limited influence of parents, Crook (2002), but it is to be believed that in the great majority of the cases, the LEA would chose. Haydn (2004) describes that each LEA would select pupils, using data from primary schools and by drawing up catchment areas, which were pie-segments, from the centre to the peripheric areas of cities. Thus the LEA's tried to mix poorer inner city areas with the more affluent suburbs. "In cities such as Manchester, which was 'long and thin' in terms of its population distribution, and which made the application of the 'pie' model problematic, pupils were 'bussed' accross the city in order to maintain a balanced social mix of pupils", Haydn (2004, p.419).

Thus, mobility could occur in moving from a comprehensive area to a selective area. We can check this by using a variable that indicates if the family moved LEA between 1969 and 1974 in our data. We can then estimate a standard probit model of the probability of moving. In order to control for the degree of comprehensiveness in the origin area and the expected degree of comprehensiveness in the destination area, we use the matched LEA level data in our dataset.

In our matched administrative data, we do not have an exact date match withe the NCDS survey. We have records for 1967 and 1972, whereas the NCDS has records for 1969-1974. We would nevertheless like to use the 1969 and 1972 measures as controls for the share of comprehensives in the origin ($-$) area in 1967, s^{-}_{c1967}, the share of comprehensives in the destination ($+$) area in 1967 s^{+}_{c1967} and the respective shares of origin and destination in 1972 s^{-}_{c1972} and s^{+}_{c1972} in order to examine their influence on mobility. Furthermore, we include measures of parental background (factor parback) and ability of the child (factor ability) and interactions with the share of comprehensives as important controls.

Table (C.10) illustrates that the moving decision overwhelmingly depended on

	Dependent variable: 0/1 variable for moving		Dep. var.: % C in destination 1972 only movers	
	(1)	(2)	(3)	(4)
% C in origin 1972 s^{-}_{c1972}	−.0061 (.001)			−.09458 (.05)
% C in destination 1972 s^{+}_{c1972}	.003 (.001)			
ability	.029 (.03)	.027 (.02)	−.350 (1.104)	−.394 (1.363)
parback	.156 (.03)	.152 (.02)	−2.168 (1.16)	−3.107 (1.458)
ability × % C origin	−.004 (.001)			.017 (.04)
ability × % C destination	.004 (.0009)			
parback × % C origin	.003 (.001)			.042 (.04)
parback × % C destination	−.004 (.001)			
Δ		.004 (.001)		
ability ×Δ× %C destination		.004 (.0009)		
parback ×Δ× %C destination		.004 (.001)		
% C in origin 1969 (s^{-}_{c1967})			−.082 (.06)	
ability × % C in origin 1967 (s^{-}_{c1967})			.056 (.06)	
parback × % C in origin 1967 (s^{-}_{c1967})			.012 (.06)	
Constant	−1.377 (.04)	−1.459 (.02)	30.990 (1.700)	32.742 (2.06)
R-squared	.03	.04	.003572	.008554
N	6164	6265	478	478

Table C.9: Moving from one LEA to another, between 1969 and 1974, conditional on ability, parental background and LEA characteristics parental background. Specification (1) indicates that if comprehensivsiation in origin in 1970 was large, and children relatively more able, families were less likely to move, and equivalently if comprehensivisation was high in the destination and children more able (given parental background), then there were more likely to move. Specification (2) asseses the reaction of parents to a change in the speed of comprehensivisation. If the difference in comprehensivisation between destination and origin ($\Delta = s^{+}_{1972} - s^{-}_{1972}$) was positive, parents were more likely to move. In columns (3) and (4) we regress the share of comprehensive schools in each LEA on parental characteristics and interactions with the shares of comprehensives in 1967 (column (3)) and 1972 (column (4)) *for those parents that moved*. It seems that parents with high parental background were more likely to be in an area with lower comprehensive shares. We see no effect for ability of the child affecting the share of comprehensives in the area, which is reassuring in specification (4). Also, we see no effect of comprehensivisation in 1967 in the origin area to affect the share of comprehensvisation in 1972 in the destination, which is equivalent to saying that there was no systematic relationship for people to move from areas with higher proportions of comprehensive schools to areas with lower proportions of comprehensive schools.

With respect to assessing the private school system, we use the same set of controls as above and determine the likelihood of entering private education, but in addition we compare movers' and stayers' private school determinants.

Dependent variable: 0/1 variable for private school						
	Probit estimation					
	(1)	(2)	(3)	(4)	(5)	(6)
	$s^-_{c1972} > 50\%$	$s^-_{c1972} \leq 50\%$	$s^-_{c1967} > 50\%$	$s^-_{c1967} \leq 50\%$		
parback	.136 (.02)	.096 (.01)	.07 (.03)	.112 (.01)	.095 (.01)	.107 (.01)
ability	.016 (.02)	.001 (.01)	.051 (.03)	.003 (.01)	.0006 (.01)	.004 (.01)
% of origin LEA comp. in 1970					−.0024 (.0017)	
% of destination LEA comp. in 1970					−.00001 (.0019)	
% of origin LEA comp. in 1967					.0056 (.0028)	
% of destination LEA comp. in 1967					−.0065 (.003)	
Teacher salary (1970)					.00007 (.0003)	
Per-pupil cost					−.0028 (.001)	
Industrialisation index					−.0088 (.002)	
Population density					.0152 (.003)	
Teaching amenities					.0126 (.003)	
Teacher salary 1969					.00976 (.004)	
Overcrowding index 1969					.0249 (.007)	
Mobility					.383 (.05)	
$\Delta\times$ comp						.002 (.001)
abi $\times\Delta\times$ %C dest.						−.001 (.0009)
par $\times\Delta\times$ %C dest.						−.001 (.0009)
Constant	−.940 (.03)	−.881 (.02)	−.959 (.06)	−.892 (.02)	−3.949 (.84)	−.905 (.02)
N	1736	6212	708	7121	7440	7691

Table C.10: Selecting a private school between ages 11-16, conditional on ability, parental background and LEA characteristics

As one can see, parental background is the dominant factor when chosing a private school. None of the specifications suggests that more able pupils attend private schools, and thus it is unlikely that there be a cream-skimming effect. We observe that there is some indication that comprehensivisation in the origin makes one less likely to attend private schools. This is perhaps consistent with the fact that more defavourised areas where comprehensivised first, leading to there being less pupils attending private schools by definition. We also see that various proxies for a defavorised area (industrialisation index, teaching amenities and overcrowding measures) affect the decision for a private school in the same manner. People who moved were also more likely to send their children to private schools, as they were better off financially. Finally, column (6) illustrates that there is no effect of the comprehensivisation differential on the decision to go private.

Finally, we compare ability in private schools and selective schools for fully selective areas with ability in private schools and comprehensive schools in fully comprehensive areas. This graph is given in the next figure.

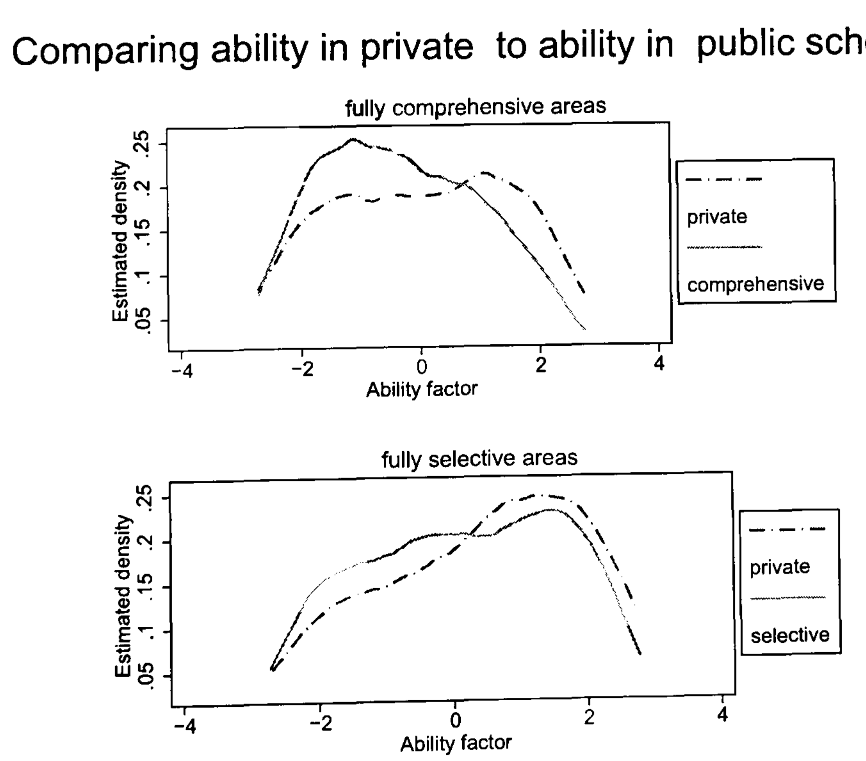


Figure C.3: Ability in fully comprehensive vs fully selective areas, graphs by individuals in private or selective/comprehensive schools

There seems to be not more cream-skimming for the private sector when one considers fully comprehensive versus fully selective areas. Actually, more individuals of higher ability are in private schools in selective areas than in comprehensive areas. Again, there is not much evidence of private schools having expanded more in the comprehensive areas, and that high ability students in comprehensive areas attended private schools in higher proportions than in fully selective areas - the opposite seems rather the case.

It seems that we can detect some mobility in the data, but this mobility seems unrelated to the comprehensive reform but rather due to parental and local characteristics. Also, we find little evidence that in comprehensive areas, more people moved their children to private schools in response to increased comprehensivisation in that area.

Bibliography

Abel, A. B.: 1985, Precautionary saving and accidental bequests, *American Economic Review* **75**(4), 777–91. available at http://ideas.repec.org/a/aea/aecrev/v75y1985i4p777-91.html.

Altonji, J. G. and Dunn, T.: 1996, Using siblings to estimate the effects of school quality on wages, *Review of Economics and Statistics* **78**, 665–671.

Altonji, J. G., Elder, T. E. and Taber, C. R.: 2005, Selection on observed and unobserved variables: Assesing the effectiveness of catholic schools, *Journal of Political Economy* **113**, 151–184.

Altonji, J. G. and Segal, L. M.: 1996, Small sample bias in GMM estimation of covariance structures, *Journal of Business and Economic Statistics* **14**, 353–366.

Anderberg, D. and Balestrino, A.: 2003, Self-enforcing intergenerational transfers and the provision of education, *Economica* **70**, 55–71.

Anker, R., Ashraf, S., Barge, S. and Levison, D.: 1996, Is child labour really necessary in India's carpet industry?, *ILO Labour Market Paper* **15**, 1–36.

Arrondel, L. and Masson, A.: 2002, Altruism, exchange or indirect reciprocity: What do the data on family transfers show, *Working Paper DELTA-ENS Paris* .

Baland, J.-M. and Robinson, J. A.: 2000, Is child labour inefficient ?, *Journal of Political Economy* **108**, 663–678.

Barro, R. J.: 1974, Are government bonds net wealth ?, *Journal of Political Economy* **82**, 1095–1117.

Bartolucci, F. and Scaccia, L.: 2004, Testing for positive association in contingency tables with fixed margins, *Computational Statistics and Data Analysis* **47**, 195–210.

Basu, A. K., Chau, N. H. and Grote, U.: 2006, Guaranteed manufactured without child labor: the economics of consumer boycotts, social labelling and trade sanctions, *Review of Development Economics* **10**, 466–491.

Basu, K.: 1999, Child labour: Cause, consequences and cure with remarks on international labour standards, *Journal of Economic Litterature* **37**, 1083–1119.

Basu, K. and Tzannatos, Z.: 2003, The global child labor problem: What do we know and what can we do ?, *The World Bank Economic Review* **17/2**, 147–173.

Basu, K. and Van, P. H.: 1998, The economics of child labour, *American Economic Review* **88**, 412–427.

Becci, E. and Julia, D.: 1998, *Histoire de l'enfance en Occident - du XVIIIe siècle à nos jours*, Seuil.

Becker, G. S. and Tomes, N.: 1979, An equilibrium theory of the distribution of income and intergenerational mobility, *Journal of Political Economy* **87**(6), 1153–1189.

Becker, G. S. and Tomes, N.: 1986, Human capital and the rise and fall of families, *Journal of Labor Economics* **4**(3), S1–39. available at http://ideas.repec.org/a/ucp/jlabec/v4y1986i3ps1-39.html.

Behrman, J. R. and Rosenzweig, M. R.: 2002, Does increasing womens schooling increase the schooling of the next generation ?, *American Economic Review* **92**(2), 323–335.

Ben-Porath, Y.: 1967, The production of human capital and the lifecycle of earnings, *Journal of Political Economy* **75**, 352–365.

Bernheim, B. D., Shleifer, A. and Summers, L. H.: 1985, The strategic bequest motive, *Journal of Political Economy* **93**(6), 1045–76. available at http://ideas.repec.org/a/ucp/jpolec/v93y1985i6p1045-76.html.

Bernheim, B. and Severinoro, S.: 2000, Bequests as signals, *Working Paper 7791*, NBER.

Bhalotra, S.: 2004, Parent altruism, cash transfers and child poverty, *Technical report*, Mimeo University of Bristol.

Bhalotra, S.: 2007, Is child work necessary, *Oxford Bulletin of Economics and Statistics* **69**, 29–56.

Bhalotra, S. and Heady, C.: 2003, Child farm labour: The wealth paradox, *The World Bank Economic Review* **17**, 197–227.

Bhalotra, S. and Tzannatos, Z.: 2003, Child labour: What have we learnt ?, *Social protection discussion paper, No.0317,The World Bank* .

Bhaskar, V.: 1998, Informational constraints and the overlapping generations model: Folk and anti-folk theorems, *Review of Economic Studies* **65**, 135–149.

Black, D. A. and Smith, J. A.: 2004, How robust is the evidence on the effects of college quality: Evidence from matching, *Journal of Econometrics* **121**, 99–124.

Black, S., Devereux, P. and Salvanes, K.: 2003, Why the apple doesn't fall far - understanding intergenerational transmission of human capital, *NBER Working Paper 10066* .

Black, S. E., Devereux, P. J. and Salvanes, K. G.: 2005, The more the merrier ? the effect of family size and birth order on children's education, *Quarterly Journal of Economics* **120**, 669–700.

Blow, L., Goodman, A., Kaplan, G., Walker, I. and Windmeijer, F.: 2005, How important is income in determining children's outcomes ? - a methodology review of econometric approaches. Mimeo University of Warwick.

Blow, L., Walker, I. and Zhu, Y.: 2004, Who benefits from child benefit, *Technical report*, Mimeo, University of Warwick and IFS.

Blundell, R. and Powell, J. L.: 2004, Censored regression quantiles with endogeneous regressors. UCL and University of California.

Blundell, R., Preston, I. and Walker, I.: 1994, *The Measurement of Household Welfare*, Cambridge University Press.

Bourguignon, F., Ferreira, F. H. and Leite, P. G.: 2003, Conditional cash transfers, schooling and child labour: Micro-simulating Brazil's bolsa escola program, *World Bank Economic Review* **17**, 229–254.

Boutros-Ghali, B.: 1992, Empowering the United Nations, *Foreign Affairs* .

Bowning, M.: 1992, Children and household economic behaviour, *Journal of Economic Litterature* **30**, 1434–1475.

Carneiro, P.: 2005, Intergenerational effects of mother's schooling on children's outcomes: Causal links and transmission channels. UCL and IFS Working Paper.

Carneiro, P., Hansen, K. and Heckman, J.: 2003, Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college, *International Economic Review* **44**(2), 361–422.

Chaudhuri, S. and Gupta, M. R.: 2004, Child labour and trade liberalisation, *The Japanese Economic Review* **55**, 201–211.

Chen, S. and Khan, S.: 2001, Semiparametric estimation of a partially linear censored regression model, *Econometric Theory* **17**(3), 567–590.

Chernozhoukov, V. and Hong, H.: 2002, Three-step censored quantile regression and extramarital affairs, *Journal of the American Statistical Association* **97**(459), 872–882.

Chevalier, A.: 2004, Parental education and child's education: A natural experiment, *Centre for the Economics of Education Working Paper* .

Chevalier, A., Harmon, C., O'Sullivan, V. and Walker, I.: 2005, The impact of parental income and education on the schooling of their children. IFS Working Paper WP05/05.

Chevallier, A., Harmon, C., Zhu, Y. and Walker, I.: 2004, Does education raise productivity or just reflect it ?, *The Economic Journal* **114**.

Chitty, C.: 2002, The role and status of LEA's: Post-war pride and fin de siècle uncertainty, *Oxford Review of Education* **28**(2-3), 261–273.

Cigno, A.: 1993, Intergenerational transfers without altruism: Family, market and state, *Journal of Political Economy* **9**, 505–18.

Cigno, A., Rosati, F. and Guarcello, L.: 2002, Does globalisation increase child labour, *Technical report*, IZA Working Paper No. 470.

Cockburn, J.: 2001, Income contributions of child work in rural Ethiopia, *CSAE Working Paper* .

Cox, D.: 1987, Motives for private income transfers, *Journal of Political Economy* **95**, 508–546.

Crook, D.: 2002, Local authorities and comprehensivisation in England and Wales, *Oxford Review of Education* **28**(2), 247–260.

Cunha, F. and Heckman, J.: 2006, Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation, *Mimeo* .

Cunha, F., Heckman, J. and Schennach, S.: 2006, Estimating the technology of cognitive and non-cognitive skill formation, *Mimeo* .

Cunningham, H. and Viazzo, P. P.: 1996, *Child Labour in Historical Perspective 1800-1985, Case Studies from Europe, Japan and Colombia*, UNICEF International Child Development Centre.

Das, M., Newey, W. and Vella, F.: 2003, Non-parametric estimation of sample selection models, *Review of Economic Studies* **70**, 33–58.

Davies, R. B.: 2005, Abstinence from child labor and profit seeking, *Journal of Development Economics* **76**, 251–263.

Dayiolu, M.: 2005, Patterns of change in child labour and schooling in Turkey: The impact of compulsory schooling, *Oxford Development Studies* **33**, 195–210.

Dearden, L., Ferri, J. and Meghir, C.: 2002, The effect of school quality on educational attainment and wages, *The Review of Economics and Statistics* **84**(1), 1–20.

Dearden, L., Machin, S. and Reed, H.: 1997, Intergenerational mobility in Britain, *Economic Journal* **107**(440), 47–66.

Dehejia, R. and Gatti, R.: 2002, Child labour: The role of income variability and access to credit across countries, *NBER Working Papers* **W9018**.

Dessy, S. and Pallage, S.: 2005, A theory of the worst forms of child labour, *The Economic Journal* **115**, 68–87.

Doepke, M. and Zilibotti, F.: 2003, Voting with your children: a positive analysis of child labour laws, *Working Paper UCLA* **828**, 1–57.

Duflo, E.: 2000, Child health and household resources in South Africa: Evidence from the old age pension program, *American Economic Review* **90**, 392–398.

Duncan, T., Cortreras, D. and Elizabeth, F.: 2002, Distribution of power within the household and child health, *Technical report*, Mimeo UCLA.

Edlund, L. and Rahman, A.: 2005, Household structure and child outcomes: Nuclear vs. extended families – evidence from Bangladesh. UCL and Columbia University.

Edmonds, E.: 2006a, Child labour and schooling responses to anticipated income in South Africa, *Journal of Development Economics* **81**(2), 386–414.

Edmonds, E. and Pavcnik, N.: 2004, International trade and child labour: Cross-country evidence, *NBER Working Paper* **10317**, 1–36.

Edmonds, E. V.: 2006b, Understanding sibling differences in child labour, *Journal of Population Economics* **19**, 795–821.

Edmonds, E. V.: 2007, Child labour, *NBER Working Paper* **12926**.

Figlio, D. and Page, M. E.: 2002, School choice and the distributional effects of ability tracking: Does separation increase inequality, *Journal of Urban Economics* **51**(3), 497–514.

Galindo-Rueda, F. and Vignoles, A.: 2005, The heterogeneous effect of selection in secondary schools: Understanding the changing role of ability, *Centre for the Economics of Education Discussion Paper* .

Gary-Bobo, R. J., Prieto, A. and Picard, N.: 2006, Birth order and sibship sex composition as instruments in the study of education and earnings. CEPR Working Paper 5514.

Goldin, C. and Parsons, D. O.: 1989, Parental altruism and self-interest: Child labour among late nineteenth-century American families, *Economic Enquiry* **27**, 637–659.

Grootaert, C. and Kanbur, R.: 1995, Child labour, an economic perspective, *International Labour Review* **134**, 187–203.

Grootaert, C. and Patrinos, H.: 1998, *The Policy Analysis of Child Labour: A Comparative Study - Child Labour in Côte d'Ivoire: Incidents and Determinants*, The World Bank, Washington.

Grossmann, H. and Michaelis, J.: 2007, Trade sanctions and the incidence of child labour, *Review of Development Economics* **11**, 49–62.

Groves, L.: 2004, Implementing ILO child labour convention 182: Lessons from Honduras, *Development in Practice* **14**, 171–182.

Gupta, M. R.: 2000, Wage determination of a child worker: A theoretical analysis, *Review of Development Economics* **4**, 219–228.

Hagemann, F., Diallo, Y., Etienne, A. and Mehran, F.: 2006, *Global Child Labour Trends*, ILO Geneva.

Hansen, K., Heckman, J. and Mullen, K.: 2004, The effect of schooling and ability on achievement test scores, *Journal of Econometrics* **121**(1-2), 39–98.

Harmon, C. and Walker, I.: 2000, The returns to the quantity and quality of education: Evidence for men in England and Wales, *Economica* **67**(265), 19–35.

Hart, O. and Moore, J.: 1999, Foundations of incomplete contracts, *Review of Economic Studies* **66**, 115–138.

Haydn, T.: 2004, The strange death of the comprehensive school in England and Wales, 1965-2002, *Research Papers in Education* **19**(4), 415–432.

Hazan, M. and Berdugo, B.: 2002, Child labour, fertility and economic growth, *The Economic Journal* **112**, 810–828.

Heady, C. and Bhalotra, S.: 2000, Determinants of child farm labour in Ghana and Pakistan. University of Bath and University of Bristol.

Hirano, K., Imbens, G. and Ridder, G.: 2003, Efficient estimation of average treatment effects using the estimated propensity score, *Econometrica* **71**(4), 1161–1189.

Holtz-Eakin, D., Newey, W. and Rosen, H.: 1998, Estimating vector autoregressions with panel data, *Econometrica* **56**(6), 1371–1395.

Horrel, S. and Humphries, J.: 1995, Child labour and the family economy in the Industrial Revolution, *Explorations in Economic History* **32**, 585–516.

Hussain, M. and Maskus, K. E.: 2003, Child labour use and economic growth: an econometric analysis, *World Economy* **26**, 993–1017.

Ichino, A., Mealli, F. and Nannicini, T.: 2006, From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity, *CEPR Discussion Papers* .

ILO: 2002, *Every Child Counts: New Global Estimates on Child Labour*, ILO Geneva.

Imbens, G.: 2003, Sensitivity to exogeneity assumptions in program evaluation, *American Economic Review* **93**(2), 126–132.

Imbens, G.: 2004, Nonparametric estimation of average treatment effects under exogeneity: A review, *The Review of Economics and Statistics* **86**(1), 4–29.

Jacoby, H. G.: 1993, Shadow wages and peasant family labour supply: An econometric application to the Peruvian sierra, *Review of Economic Studies* **60**, 903–921.

Jafarey, S. and Lahiri, S.: 2002, Will trade sanctions reduce child labour ?, *Journal of Development Economics* **68**, 137–156.

Jafarey, S. and Lahiri, S.: 2005, Food for education versus school quality: a comparison of policy options to reduce child labour, *Canadian Journal of Economics* **38**, 394–419.

Jenkins, A. and Sabates, R.: 2007, The classification of qualifications in social surveys, *Technical report*, CLS Working Paper 2007-2.

Kandori, M.: 1992, Repeated games played by overlapping generations of players, *Review of Economic Studies* **59**, 81–91.

Kerckhoff, A. C.: 1986, The effects of ability grouping in British secondary schools, *American Sociological Review* **51**, 842–858.

Kirby, P.: 2005, A brief statistical sketch of the child labour market in mid-nineteenth century London, *Continuity and Change* **2**, 229–245.

Kis-Katos, K. and Schulze, G. G.: 2005, Regulation of child labour, *Economic Affairs* **25**, 24–30.

Laitner, L. and Ohlsson, H.: 2001, Bequest motives: a comparison of Sweden and the US, *Journal of Public Economics* **79**(1), 205–236.

Levy, V.: 1985, Cropping patterns, mechanisation, child labour and fertility behaviour in a farming economy, *Economic Development and Cultural Change* **33**, 777–791.

Loader, C.: 1997, Locfit: an introduction, *Statistical Computing and Graphics Newsletter* pp. 1–10.

Lopez-Calva, L. F.: 2001, Child labour: Myths, theories and facts, *Journal of International Affairs* **55**, 59–73.

Lopez-Calva, L. F.: 2003, Filial obligations and child labor, *Working Paper Colegio de Mexico* .

Maffei, A., Raabe, N. and Ursprung, H. W.: 2006, Political repression and child labour: Theory and empirical evidence, *The World Economy* **2**, 211–239.

Manacorda, M.: 2006, Child labor and the labor supply of other household members: Evidence from 1920 America, *American Economic Review* **96**(5), 1788–1800.

Manning, A. and Pischke, J.-S.: 2006, Comprehensive versus selective schooling in England and Wales: What do we know ?, *NBER Working Paper* (12176).

Marx, K.: 1867, *Capital: A Critique of Political Economy*, Progress Publishers.

Maskin, E. and Tirole, J.: 1999, Unforeseen contingencies, property rights and incomplete contracts, *Review of Economic Studies* **66**, 83–114.

Maurin, E.: 2002, The impact of parental income on early schooling transitions. a re-examination using data on three generations, *Journal of Public Economics* .

Maurin, E. and McNally, S.: 2006, Selective schooling, *Mimeo* .

McElroy, M. B.: 1990, The empirical content of nash-bargained household behaviour, *The Journal of Human Resources* **25**, 559–583.

McElroy, M. B. and Horney, M. J.: 1981, Nash bargained household decisions: toward a generalisation of the theory of demand, *International Economic Review* **22**, 333–49.

McGarry, K.: 1997, Inter-vivos transfers and intended bequests, *Working Paper W6345*, NBER.

Meghir, C. and Palme, M.: 2005, Educational reform, ability and family background, *American Economic Review* **95**(1), 414–424.

Merkeerk, E. v. N.: 2004, Peter Kirby: Child labour in Britain, 1750-1870, book review, *International Review of Social History* **3**, 517–518.

Michael, R. T.: 2005, Children's reading and math skills: The family's two constraints. Harris School Working Paper, 0511.

Moehling, C. M.: 1995, The intrahousehold allocation of resources and the participation of children in household decision making: Evidence from early 20th century Northamerica. Mimeo, Northwestern University.

Moehling, C. M.: 1999, State child labour laws and the decline of child labour, *Explorations in Economic History* **36**, 72–106.

Moehling, C. M.: 2003, The incentives to work: Working children and household decision making. Yale University, Working Paper.

Murray, M. P.: 2005, The bad, the weak and the ugly: Avoiding the perils of instrumental variable estimation. Bates College Working Paper.

Nardinelli, C.: 1980, Child labor and the factory acts, *The Journal of Economic History* **40**, 739–755.

Nash, J.: 1951, Non-cooperative games, *The Annals of Mathematics* **54**(2), 286–295.

Neumayer, E. and Soysa, I. d.: 2005, Trade openness, foreign direct investmet and child labour, *World Development* **33**(1), 43–63.

Nieuwenhuys, O.: 2007, Embedding the global womb: Global child labour and the new policy agenda, *Children's Geographies* **5**(1-2), 149–163.

Nkamleu, G. B. and Kielland, A.: 2006, Modelling farmers' decisions on child labour and schooling: a multinomial logit analysis in Côte d'Ivoire, *Agricultural Economics* **35**, 319–333.

OECD: 2003, *Combating Child Labour: a Review of Policies*, OECD Paris.

Oreopoulos, P., Page, M. E. and Stevens, A. H.: 2006, The intergenerational effect of compulsory schooling, *Journal of Labour Economics* **24**(4), 729–760.

Page, B. R.: 2003, Bequest taxes, inter-vivos gifts and the bequest motive, *Journal of Public Economics* **87**(6).

Perozek, M. G.: 1998, A reexamination of the strategic bequest motive, *Journal of Political Economy* **106**(2), 423–445. available at http://ideas.repec.org/a/ucp/jpolec/v106y1998i2p423-445.html.

Peterson, A. D. C.: 1965, Secondary reorganisation in England and Wales, *Comparative Education* **1**(3), 161–169.

Pettit, B.: 1998, *Children and Work in the UK: Reassesing the Issues*, Child Poverty Action Group.

Plug, E.: 2004, Effect of mothers' schooling on schooling of adoptees, *American Economic Review* **94**(1), 358–368.

Plug, E. and Vijverberg, W.: 2005, Does family income matter for schooling outcomes ? using adoptees as a natural experiment, *Economic Journal* **115**(506), p.879–906.

Powell, J. and Honore, B.: 2005, *Identification and Inference in Economic Models*, Cambridge University Press.

Psacharopoulos, G.: 1997, Child labour versus educational attainment: some evidence from Latin America, *Journal of Population Economics* **10**, 377–386.

Psacharopoulos, G. and Patrinos, H.: 2002, Returns to investment in education: A further update, *Technical Report 2881*, World Bank.

R, D. C. T.: 2005, *R, A language and Environment for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0.

Rammohan, A.: 2001, Development of financial capital markets and the role of children as economic assets, *Journal of International Development* **13**, 45–58.

Rangel, A.: 2003, Forward and backward intergenerational goods: Why is social security good for the environment, *American Economic Review* **93**(3), 813–834.

Ranjan, P.: 2001, Credit constraints and the phenomenon of child labour, *Journal of Development Economics* **64**, 81–102.

Ray, R.: 2002, The determinants of child schooling and child labour in Ghana, *Journal of African Economies* **11**, 561–590.

Robson, A.: 2003, A 'bioeconomic' view of the neolithic and recent demographic transition, *Working Paper* .

Rosenbaum, P. and Rubin, D.: 1983, The central role of the propensity score in observational studies for causal effects, *Biometrika* **70**(1), 41–55.

Rosenzweig, M. and Evenson, R.: 1977, Fertility, schooling and the economic contribution of children in rural India: An econometric analysis, *Econometrica* **45**, 1065–1079.

Rosenzweig, M. R.: 1995, Why are there returns to schooling ?, *American Economic Review* **85**, 153–158.

Rothstein, J. and Yoon, A.: 2006, Mismatch in law school, *Northwestern Law and Economics Research Paper No.881110* .

Rubin, D.: 1974, Estimating causal effects to treatments in randomised and non-randomised studies, *Journal of Educational Psychology* **66**, 688–701.

Sacerdote, B.: 2005, Slavery and the intergenerational transmission of human capital, *The Review of Economics and Statistics* **87**(2), p.217–234.

Satz, D.: 2003, Child labour: A normative perspective, *World Bank Economic Review* **17**, 311–324.

Schultz, P. T.: 2004, School subsidies for the poor: Evaluating the Mexican Progresa program, *Journal of Development Economics* **74**, 199–250.

Selten, R. and Wooders, M. H.: 2001, Cyclic games: An introduction and some examples, *Games and Economic Behaviour* **34**, 138–152.

Simon, B.: 1991, *Education and the Social Order, 1950-1999*, Lawrence and Wishart, London.

Skoufias, E.: 1994, Using shadow wages to estimate labour supply of agricultural households, *Agricultural Economics* **76**, 215–227.

Skyt-Nielsen, H.: 2001, How sensitive is the demand for primary education to changes in economic factors?, *Journal of African Economies* **10**, 191–218.

Skyt-Nielsen, H. and Dubey, A.: 2002, Child labour in rural India: A microeconomic perspective, *The Indian Journal of Labour Economics* **45**, 479–497.

Smith, L.: 1992, Folk theorems in overlapping generations games, *Games and Economic Behaviour* **4**(3), 426–449. available at http://ideas.repec.org/a/ecm/emetrp/v62y1994i4p939-48.html.

Solon, G.: 1992, Intergenerational income mobility in the United States, *American Economic Review* **82**(3), 393–408. available at http://ideas.repec.org/a/aea/aecrev/v82y1992i3p393-408.html.

Souza, A. P. and Emerson, P. M.: 2003, Is there a child labour trap ? intergenerational persistence of child labour in Brazil, *Economic Development and Cultural Change* **51**, 375–399.

Stark, O.: 1995, *Altruism and Beyond*, Cambridge University Press.

Strulik, H.: 2004, Child mortality, child labour and economic development, *The Economic Journal* **114**, 547–568.

Swinnerton, K. and Rogers, C.: 1999, The economics of child labour: Comment, *American Economic Review* **89**, 1382–85.

Swinnerton, K. and Rogers, C.: 2004, Does child labor decrease when parental income rise ?, *Journal of political economy* **112**, 939–47.

Tanaka, R.: 2003, Inequality as a determinant of child labour, *Economics Letters* **80**, 93–97.

Tipping, M. E. and Bishop, C. M.: 1999, Probabilistic principal component analysis, *Journal of the Royal Statistical Society - Series B (Statistical Methodology)* **61**(3), 611–622.

Todd, P. E. and Wolpin, K. A.: 2003, On the specification and estimation of the production function for cognitive achievement, *Economic Journal* **113**(485), 3–33.

Todd, P. and Wolpin, K.: 2004, The production of cognitive achievement in children: Home, school and racial test score gaps, *University of Pennsylvania Workig Paper* .

Todd, P. and Wolpin, K. I.: 2002, Using a social experiment to validate a dynamic behavioural model of child schooling and fertility, *Pier Working Paper* .

Udry, C.: 1996, Gender, agricultural production and the theory of the household, *The Journal of Political Economy* **104**, 1010–1046.

Udry, C. and Pranab, B.: 1999, *Development Microeconomics*, Oxford University Press.

Wagener, A.: 2002, Intergenerational transfer schemes as incomplete social contracts, *Journal of Political Economy* **13**, 337–359.

Waldinger, F.: 2006, Does tracking affect the importance of family background on student's test scores, *Mimeo CEP-LSE* .

Woessmann, L. and Hanushek, E.: 2006, Does education tracking affect performance and ineqality ? differences-in-differences evidence across countries, *The Economic Journal* **116**(3), C63–C76.

Wooldridge, J. M.: 2002, *Econometric Analysis of Cross-section and Panel Data*, MIT Press.

Yates, A.: 1971, The development of comprehensive education in England, *International Review of Education* **17**(1), 58–65.

Yoon, K.: 2001, An anti-folk theorem in overlapping generations games with limited observability, *Review of Economic Dynamics* **4**, 736–745.