Essays in Empirical Development Economics

by

Esther Duflo

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1999
[June, 1999]

Author ....
Department of Economics
May 15, 1999

Certified by ....
Abhijit V. Banerjee
Professor of Economics
Thesis Supervisor

Certified by ....
Joshua Angrist
Professor of Economics
Thesis Supervisor

Accepted by ....
Peter Temin
Chairman, Departmental Committee on Graduate Studies

Essays in Empirical Development Economics

by

Esther Duflo

Submitted to the Department of Economics
on May 15, 1999, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis is a collection of three essays in empirical development economics.

The first chapter evaluates the effects on education and wages of a large school construction program undertaken by the Indonesian government between 1973 and 1978. I evaluate the effect of this program on education and wages by combining differences across regions in the number of schools constructed with differences across cohorts induced by the timing of the program. The estimates suggest that the construction of primary schools led to a substantial increase in education and earnings. These estimates implie returns to education ranging from 6.4% to 9.1%.

This second chapter studies the impact of household resources on child nutrition in South Africa. In the early 1990s, the benefits and coverage of the South African social pension program were dramatically expanded for the black population. About a third of black South African children under age 5 live with an elderly person. This chapter examines whether this large positive income shock was followed by an improvement of anthropometric status. Estimates suggest that pensions received by women had a large and significant impact on the anthropometric status of girls and a smaller and insignificant effect on that of boys. I found no effect of the pension on child nutrition when it is received by men.

The third chapter examines the role that reputation plays in determining contractual outcomes, using a data set containing detailed information about 230 projects carried out by 125 software firms that I have collected for this purpose. Ex ante contracts as well as the outcome after ex post renegotiation vary with firms' characteristics plausibly associated with reputation. I propose a model of the industry where reputation determines contractual outcomes, whose predictions are consistent with several facts observed in the data.

Thesis Supervisor: Abhijit V. Banerjee
Title: Professor of Economics

Thesis Supervisor: Joshua Angrist
Title: Professor of Economics

Acknowledgments

I thank my advisors, Abhijit V. Banerjee and Joshua Angrist. I owe much more to them than what is reflected in this thesis.

I also thank Michael Kremer, Jonathan Morduch, and Kaivan Munshi for many long and important conversations, and MIT professors Daron Acemoglu, Jonathan Gruber, Paul Joskow, Steve Pishke, Michael Piore, and Jim Poterba, as well as my fellow students Alberto Abadie, Aymee Chin, John Johnson, Stuti Khemani and Guido Lorenzoni, for their support and their comments. I am grateful to Anne Case and Angus Deaton for their hospitality in Princeton and their help to my work on South Africa, and to Daniel Cohen and Thomas Piketty for convincing me that economics was worth studying.

I acknowledge financial support from the Ecole Normale Supérieure, the American Womens' Group in Paris, the Fondation Thiers pour la Recherche en Sciences Sociales, the Sloan doctoral Dissertation Fellowship, the World Bank, and the George and O'Bie Shultz fund.

I particularly grateful to Emmanuel Saez for his help, his insights, and his tireless support.

This thesis is dedicated to the memory of Geneviève Duflo.

Contents

Introduction

This dissertation is a collection of three independent essays in empirical development economics. In the first two chapters, I evaluate the causal effects of public policies. The first chapter evaluates the impact of a large school construction program on education and labor market outcomes in Indonesia. The second chapter exploit the variation introduced by important changes in the old age pension program to estimate the effect of household resources on child nutrition in South Africa. The third chapter examines reputation effects in the Indian software industry, using a data set I have collected for this purpose.

The first chapter is entitled "Schooling and Labor Market Consequences of School construction in Indonesia: Evidence from an Unusual Policy Experiment". This essay evaluates the effects on education and wages of a large school construction program that took place in Indonesia in the mid-1970s. Between 1973 and 1978, the Indonesian government constructed over 61,000 primary schools throughout the country (this is one of the largest school construction programs on record). The allocation rule specified that more schools had to be built in regions where enrollment rates in 1972 were low. This rule introduced substantial variation in the intensity of the program across regions. In addition, the program was an important change in policy: before 1973, capital expenditures in the education sector were very limited. Because Indonesian children attend primary schools only between ages 7 and 12, there is a variation in exposure to the program across cohorts of birth: children who turned 13 before 1973 did not benefit from the program while younger children did. I evaluate the effect of this program on education and wages by combining these two sources of variation. I use a large survey of Indonesian households (the 1995 Intercensal Survey of Indonesia), combined with administrative data on the allocation of schools in each region. The household data set contains information on education, wages, and year and region of birth. The estimates suggest that the construction of primary schools led to a substantial increase in education (especially at the primary level) and earnings. Children aged 2 to 6 in 1974 received 0.12 to 0.18 more years of education for each school constructed per 1000 children in their region of birth. Using the variation in schooling generated by this policy as an instrumental variable for the impact of education on wages generates estimates of economic returns to education ranging from 6.4% to 9.1%. A cost-benefit analysis of the program based on these estimates suggests that the internal rate of return of this program was at least 7%.

The second chapter is entitled "Child Health and Household Resources in South Africa: Evidence from the Old Age Pension Program". This essay studies the impact of household resources on child nutrition in South Africa. In the early 1990s, the benefits and coverage of the South African social pension program were dramatically expanded for the black population. In 1993, nearly 85% of black South Africans eligible on the basis of age were receiving an old age pension. The benefits level was twice the median income per capita in rural areas. Due to traditional living arrangements, about a fourth of black South African children under age 5 live with an elderly person. This reform provides a unique opportunity to estimate the impact of an exogenous increase in income on child health. Using a nationally representative survey of South African households conducted in 1993, this paper examines whether this large positive income shock was followed by an improvement in the anthropometric status of children living with a pension recipient. Simple correlations between the amount of pension income received by the household and child height (a long run measure of nutritional status) tend to be negative in the complete sample, which simply reflects the fact that pensioners live in poorer households. However, height reflects *cumulated* effects of nutrition and illnesses. Therefore, younger children, who have been exposed to the program for all their lives, do relatively better in households where a member is eligible for pension. Older children, on the other hand, do worse, which reflects the fact that they are coming from a more disadvantaged background and were exposed to the program only during a fraction of their lives. Estimates based on this observation suggest that pensions received by women had a large and significant impact on the anthropometric status of girls (it increased their standardized height by 0.80 standard deviation, bridging more than half the gap with American girls the same age), and a smaller and insignificant effect on that of boys. In contrast, I found no effect of the pension on child nutrition when it is received by men.

The third chapter is entitled "Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry". It is co-authored with Abhijit V. Banerjee. This paper examines the role that reputation plays in determining contractual outcomes. The question we address is the following. In an incomplete contracting environment, does the reputation of a firm affect the way its contracts are written? How important are these effects?

To address these questions, we conduct an empirical analysis of the Indian customized software industry. Customized software is an obvious place to study such effects since the desired end-product tends to be extremely complex and difficult to describe ahead of time in a way that a third party (such as a court) would understand.

We interviewed the CEOs of 125 firms in India in the Winter 1997/1998, and we have collected a data set containing detailed information about the firms and 230 projects carried out by these firms, which we analyze in this paper. We have obtained information on the initial contracts, the type of project, the cost overruns and their sources, and outcomes of ex post renegotiation. The evidence supports the view that reputation matters. Ex ante contracts as well as the outcome after ex post renegotiation vary with firm characteristics plausibly associated with reputation (age, certification, previous relationship with the client). We argue that this pattern is not consistent with optimal risk sharing and propose a model of the industry where reputation determines contractual outcomes, whose predictions are consistent with several facts observed in the data. We argue that there is no obvious alternative explanation to the patterns present in the data.

Chapter 1

Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment

1.1 Introduction

The question of whether investment in infrastructure increases human capital and reduces poverty has long been a concern to development economists and policy makers. For example, availability of schooling infrastructure has been shown to be positively correlated with completed schooling or enrollment by Birdsall (1985) in urban Brazil, DeTray (1988) and Lillard and Willis (1993) in Malaysia, Lavy (1996) in Ghana, and Case and Deaton (1996) in South Africa. The principal methodological problem with these studies is that schools are not randomly allocated across communities. In education systems relying on local financing, more affluent communities can afford to build more schools. Children in these communities are likely to be more educated (and earn more in adulthood). On the other hand, in centralized education systems, government resources may be allocated to regions that are lagging behind (as it was the case with school building in Indonesia in the 1970's). As a result, education and wages may be lower in the regions that have more government schools. The ideal experiment to estimate the effects of building schools would be to allocate schools randomly to some communities and not to others, and then to compare education and earnings across communities. In the absence of evidence from such an experiment, it is necessary to rely on exogenous natural variation in combination with statistical modeling strategy.

This paper exploits a dramatic change in policy to evaluate the effect of building schools on education and earnings in Indonesia. In 1973, the Indonesian government launched a major school construction program, the Sekolah Dasar INPRES program (or SD INPRES program). Between 1973/74 and 1978/79 (the duration of Indonesia's second five-year plan), more than 61,000 primary school buildings were built (an average of two schools per 1000 children aged 5 to 14 in 1971). The government's goal was to increase enrollment rates among children aged 7 to 12 from 69 percent in 1973 to 85 percent by 1978. In 1978, the enrollment rate reached 84 percent for males and 82 for females (World Bank (1990)). Indonesia's primary schooling expansion is quoted by the World Bank as "one of the most successful cases of large-scale school expansion on record" (World Bank(1990)). This program represented a drastic change in policy. Prior to 1973, capital expenditures in the education sector were very low (Davoesan (1971)) and it appears that enrollment rates were actually declining in the early 1970s (Davoesan (1971), Heneveld (1978)).

The identification strategy in this paper uses the fact that exposure to the school construction program varied by region of birth and date of birth. There was substantial variation in program intensity across regions due to the government effort to allocate more schools in regions where initial enrollment was low. Therefore education of men who were young enough to be in school when the program was launched should be higher than education of older men in all regions, but the difference should be larger in the regions which received more schools. A difference in differences estimator controls for (additive) systematic variation of education both across regions and across cohorts. Only the combination of the two variations is treated as exogenous. Similar strategies are often used in the public finance literature, to evaluate the effects of public policies. Rosenzweig and Wolpin (1988) first proposed to use fixed effects procedures for the evaluation of the usefulness of public infrastructure in developing countries and they were applied in the Indonesian context by Pitt, Rosenzweig and Gibbons (1993) and Gertler and Molyneaux (1994). This schooling reform is particularly suited for this method because the variation in inputs comes, unlike in most other studies, from a well defined reform, so that the source of variation is better understood. Furthermore, an implication of this identification assumption can be tested: I show that among early cohorts, who did not benefit from the program because they were too old to attend primary school when it started (individuals 12 or older in 1974), the increase in educational attainment from one cohort to another is not correlated with the number of INPRES schools per capita built from 1973 to 1978. This identification strategy allows me to estimate the effect of the INPRES program on education. The same strategy is used to estimate the impact of this program on 1995 wages. I then use this exogenous source of variation in education to estimate the impact of years of schooling on wages.

The question of whether an increase in educational attainment would cause an increase in income levels is a basic concern to development economists. There is a large literature on returns to education in developing countries (See Psacharopoulos (1973,1981,1985, 1994) for surveys). Estimated returns are in general at least as large, and often larger, in developing countries compared to industrialized countries. The relationship between education and economic growth has also been extensively studied by macroeconomists. Surprisingly, however, there has been very little effort to estimate returns to education using only exogenous variation in schooling. The bias in estimates that treat an individual's education level as exogenous is likely to be important in developing countries: in particular, liquidity constraints and family or community background are likely to influence both education and earnings. Behrman's (1990) assessment of the existing literature is that most standard estimates of return to education in developing countries are likely to overstate the returns to education. Strauss and Thomas (1995) survey is less negative, and concludes that the evidence is inconclusive, and deserves further study. However, finding sources of exogenous variation in education is difficult. Most factors influencing education are likely to also have other indirect effects on income. This is clearly the case of family background variables (assets and parental education) which are often used as instruments, or included in the set of instruments on the ground that they are good predictors of education. If the concern is that unobserved family and community background characteristics are sources of bias in OLS estimates of returns to education, observed family and community variables should be entered as covariates in the wage equation, and are not likely to be valid instruments. This is often also true of other potential instruments. For example, birth order has been shown to affect education. But it also affects health, which in turn affects income.[1] Proximity of parent's residence to educational facilities has been used as an instrument for college education in the US (Card (1993), Kane and Rouse(1995)), and years of secondary education education in the Philippines (Maluccio (1998)).[2] These studies suffer from the problem outlined above: schools are not randomly allocated, and depending on which mechanisms generate the allocation of schools, schooling and wages might be lower or higher in households leaving near or far away from a school, even if there are no causal effect of the proximity of a school on education. This paper exploits the exogenous variation in education created by the INPRES program to construct instrumental variables estimates of the effect of education on wages.

Using a large cross section of men born between 1950 and 1972 from the 1995 intercensal survey of Indonesia (SUPAS), I was able to link an adult's education and wages with district level data on the number of new schools built between 1973/74

[1] Similarly, Case and Butcher (1994) use sibling composition as an instrument for women's education in the US. Indeed, Garg and Morduch (1996) find that sibling composition affects education in poor households in Ghana, but they also find evidence of an impact of sibling composition on health, which is likely to affect income directly.

[2] Malucio uses a 1994 sample of 250 wages earners in the Philippines, and uses distance to school reported by the parents in 1978 as an instrument for education of the individuals interviewed in 1994.

and 1978/79 in his region of birth. The exogenous variables (and the instruments in the wage equation) are interactions between dummy variables indicating the age of the individual in 1974 and the intensity of the program in his region of birth between 1973 and 1978. Similar strategies have been used to estimate the effect of school quality on returns to education (Card and Krueger (1992)), the effect of teen fertility on educational and labor market outcomes (Angrist and Evans (1996)) and the returns to college education (Card and Lemieux (1998)).

The remainder of this paper is organized as follows: in section 2, I describe the SD Inpres program and the data I use. In section 3, I present the identification strategy and discuss the identification assumption using a simple model of endogenous schooling. In section 4, I present the results of the estimation of impact of the program on education. Section 5 is devoted to the estimation of the effect of the program on wages. In section 6, I estimate economic returns to education. In section 7, I combine estimates of the program effect on wages and detailed data on the costs of education in Indonesia to present a tentative cost-benefit analysis of the program. Section 8 concludes.

1.2 The program

1.2.1 Data

The 1995 intercensal survey of Indonesia (SUPAS), is a sample of over 200,000 households. The SUPAS is conducted every 10 years by the Central Bureau of Statistics of Indonesia. Basic data is collected on each individual in the household. In this study, I focus on men born between 1950 and 1972 (which insures that the individuals in the sample have completed their education). Summary statistics for this sample are presented in table 1. There are 152,989 individuals in the sample, with an average level of 7.8 years of completed education (6 years of education correspond to graduation from primary school). Over 91 percent of the sample are individuals who are working, but only 45 percent of the sample are working for a wage (the others are self employed). The SUPAS collects data on last month's wage for people who are working for pay. From this, I calculate the hourly wage data by dividing the monthly wage by the number of hours worked during the month. I estimate the effects of the program on education using the complete sample, but the wage equation is estimated using the reduced sample, which has only 60,633 individuals (sample selection issues are examined below).

The SUPAS asks in which district the individual was born. I used this information to match the individual survey with district level data (pertaining to the situation in the district in the 1970s) collected from various sources: Ministry of Education and Culture (MOEC), presidential instructions published by the Bappenas (the Planning Agency), and published results from the 1971 census.[3] Descriptive statistics are presented in table 1.

1.2.2 The Sekolah Dasar INPRES program

Since 1973, the "Presidential instructions" (INPRES) are the main centrally controlled fiscal mechanism determining spatial redistribution of the aggregate gains to Indonesia from the oil boom (Ravallion (1988)). Over the years, the scheme has evolved into a complex system of grants for various purposes, such as building of schools, health clinics and roads, as well as more routine government spending.

The Sekolah Dasar Inpres program was one of the first INPRES programs, and by far the largest at the time it was launched (in 1973/74). During the first five year plan (Repelita I), the emphasis was on basic infrastructure and sectoral development. Agriculture, industry/mining and communication were absorbing 70% of the development budget (H. Hill (1996)). At the outset of the second five year plan (Repelita II), which emphasized the need for equity, changing priorities were in evidence. Regional development became an important item in the budget (absorbing 15% of it). The Sekolah Dasar Inpres program represented in turn 12% of the regional development budget in 1973, and 28% in 1979 (for comparison, health expenditures represented

[3]The matching was complicated by the fact that some districts changed boundaries or name. I used maps of Indonesia to solve these ambiguities.

only 3.4% of budget for regional development in 1973 (and 5.5% in 1979)).[4] The budget itself, thanks to the oil boom, expanded very rapidly during this period (real expenditures more than doubled between 73 and 80. The share of oil in government revenues grew from 25% in 1971 to 48% in 1974 and a pic of 62% in 1981). Due to the combination of these two factors (change in priority and increase in revenues), the Sekolah Dasar Inpres program became extremely important.

Basic data about the program are presented in table 1: Between 1973/74 and 1978/79, more than 61,000 new buildings were constructed.[5] This represented on average 220 new schools (and 660 teachers) per district or about one school per 500 children aged 5-14 in 1971. This amounted to double the number of existing school buildings; since the INPRES schools were smaller than most existing schools (3 teachers), the increase in the number of teachers was only 43%. [6] Prior to 1973, in contrast, very few new schools were constructed. There was a complete ban on the recruitment of new civil servants, and some newly trained teachers could not find employment (Davoesan, 1971).

Once an INPRES school was established, the central government recruited the teachers and paid their salaries (each school was planned for three teachers and 120 pupils). The minimum requirement to be a primary school teacher was an upper secondary school degree, generally obtained in a special training school. In 1971, 71% of the primary school teachers met this qualification, while 29% were under-qualified. The program went in parallel with an effort to train more teachers (World Bank (1989)).

The program was explicitly designed to target children who had not previously been enrolled in school, and a separate budget was designed for the rehabilitation

[4]This share does not include teachers' salaries, paid out of the routine component of the budget.

[5]The program did not stop at that date. I chose to consider the construction between 1973/74 and 1978/79 date for the following reasons: it corresponds to the end of the second five year plan, a very high primary enrollment rate was achieved in 1978 and people born in 1972 (the youngest cohort in my sample), turned 7 in 1979: they were therefore fully exposed to the program. The program slowed down considerably after 1978/1979.

[6]I chose 1971 as a base year for the population of children because 1971 was a census year; in census publications, the district population is broken down by wide age categories, so I have to use children born 5-14 as the reference group -instead of 7-12 children, which would correspond to the primary school aged children.

of existing run-down buildings (Heneveld(1978) and Bappenas(1973 to 1979)). The distribution of funds is described in detailed governmental instructions (the "Presidential instructions"). All schools were constructed identically. The instructions were also explicit about the allocation rule: in 1973/74 and 1974/75, the number of schools to be constructed in each district was proportional to the number of children of primary school age *not enrolled in school* in 1972. From 1975/76 on, the rule was spelt out slightly differently, but had similar implications: the number of schools to be constructed was proportional to the number of new pupils to be accommodated between 1972 and 1978 in this region to satisfy the target of an enrollment rate of 85% in 1978. More schools were allocated to the transmigration regions.[7] The final allocation was decided by planners in the Department of Education of culture, with the approval of the Department of Finance and the Bappenas, the administration responsible for the final implementation of the program. Funds were then sent through the Governor's office to the local administrations (bupati), who supervised the actual construction. The instructions listed the exact number of schools to be constructed in each of the 281 districts (kabupaten/kotamadya).

I use this number in my analysis, rather than the actual number of schools constructed (which is not available). In 1983, the Department of Education and Culture conducted a survey of the implementation of the program from 1973 to 1983. According to this study the actual number of schools constructed matched the plans until 1980. Some discrepancy occurs thereafter. The Department of Education and Culture has also published data on the number of schools in 1973/74 and 1978/79. This data suggests that the actual growth in the number of functioning schools was lower than the number of schools constructed under the SD INPRES program (even accounting for the time lag in construction). One reason is that prior to 1973, several schools were frequently operating in the same building (as soon as a school had more than one class per grade, it became two schools, with separate head-teachers and administrative status). School buildings in urban areas could operate in as many as

[7]The transmigration regions are the areas where the government of Indonesia encouraged new settlement as a solution to the overcrowding of Java.

four shifts a day (Davoesan (1971)). It is quite possible that some new buildings were used to reduce overcrowding in the old ones. The average increase in the number of teachers implied by the allocation of INPRES schools is very close to the increase in the actual number of teachers recorded by the Ministry of Education.

Using this data, I first checked whether the allocation decided by the Ministry corresponds to the general allocation rule. The 1971 census publications do not give enrollment rates among children aged 7 to 12, but only the fraction of the overall population attending school. It is therefore not possible to run the actual formula using this data. But the rule also implies that the logarithm of the number of schools had to be correlated with the logarithm number of children and (negatively) with enrollment rate in the population in 1971. The actual rule would predict that each of this coefficient are close to unity (in absolute value). Table 2 presents the results of a simple regression testing this implication. The number of schools built in each region between 1973/74 and 1978/79 is positively correlated with the number of children, and negatively correlated with the enrollment rate. Even after accounting for measurement errors[8], the coefficients are not 1 and -1, however, suggesting an imperfect application of the rule. A substantial part of the variation between regions is explained by these two factors alone: the R squared is 0.76. Remaining noise might be explained by the fact that I use the population enrollment rate (instrumented with average education in the pre-program period) instead of enrollment among children of primary school age, and that the actual formula implied non-linearity (regions that had an enrollment rate of 85% among children in 1972 where not supposed to get any school). Finally, the implementation of such a massive program in a developing country was bound to involve some deviation from the general rule.

[8]To correct for measurement error in enrollment rate, I use an uncorrelated measure of regional education as instrument for enrollment rate in 1971: the average education in the cohorts non affected by the program (calculated from my data).

1.3 Identification strategy

1.3.1 Sources of variation

The date of birth and the region of birth jointly determine the exposure of an individual to the program. First, a child born in 1962 or before was 12 or older in 1974, when the first INPRES schools were constructed. Indonesian children attend primary school between age 7 and age 12. So a child aged 12 or older in 1974 did not benefit from the program, since he left primary school before the school year 1974/75, when the first INPRES schools (built in 1973/74) were opened. This discussion assumes that all children effectively start school at age 7 and don't repeat. In fact, there is evidence that age at school entry varies in developing countries (Glewwe and Jacoby (1995) and that grade repetition are important in Indonesia (Berhman and Deolalikar (1991)). I reproduce in panel C of table 1 information from the IFLS data set, a smaller but richer data set fielded in 1993. The survey asks about grade repetition. 20% of children have repeated at least one grade. The proportion drops to 6% among children who complete more than primary school. The survey ask in which year an individual left school. I combine this information with the information on grade repetition (for people completing more than primary) to calculate the number of people still attending primary school after age 12 and 13. 16% of students were still in primary school at age 13. Only 7% were still in primary school by age 14. Even taking into account grade repetition, the exposure of children 12 or older in 1974 is very limited. Less than 3% of them were still in primary school after 1974 (the very first year of the program). A child born in 1968 was 6 in 1974 and 11 in 1979. He was exposed to the first wave of constructions while he was of primary school age (24% of these children left schools after 1974), but was exposed only partly to the next waves. A child born in 1972 was fully exposed. In summary, children 12 or older in 1974 were not exposed to the program, and for younger children the exposure is increasing with the age in 1974. Likewise, the effect of the program should be zero for children 12 or older in 1974, and increasing for younger children.

Second, the intensity of the program varied across regions. A simple way to separate the regions into two groups is to consider that the level of program in a region is "high" if the number of schools built per child (in 1971) is higher than average, and "low" if it is lower than average. I compare the education of individuals born in high and low level regions. The region of birth is highly correlated with the region of education. In the IFLS sample, 8.5% of individuals were living at age 12 in another district than the one they were born in (cf. panel C, table 1). This migration introduces measurement error , and can only lead to downward bias in the estimation of the program effect. Moreover, it is known that endogenous migration can bias estimates of program effects (Rosenzweig and Wolpin(1988)). Some people might have move between the birth of the child and his education period to benefit from the program. However, region of birth can not be endogeous with respect to the *program*: all individuals in the sample are born *before* the program was started. Therefore, the parents could not have moved to the high program regions before the birth of the child to benefit from the program, since it was not yet implemented.[9]

Thus, the basic idea behind the identification strategy can be illustrated using simple 2 by 2 tables. In table 3, I present results which illustrate the identification strategy and a test of an implication of the identifying assumption. These results are imprecise, due to the fact that a small part of the available information is used. They mean to be illustrative rather than definitive. This table shows means of education and wages for different cohorts and program levels.[10] In panel A (left panel), I indicated the program intensity in both types of regions. In high program regions an average of 2.53 schools per 1000 children was built; in low program regions an average of 1.51 schools per 1000 children was built: the difference was 1.02 school per 1000 children. In panel B, I present the main experiment. I compare the educational attainment and the wage of individuals who were not exposed to the program (they were 12 to 17 in 1974) to that of individuals who were exposed during all the time they spent in primary school (they were 2 to 6 in 1974), in both types of regions. The

[9]This is not saying, of course, that region of birth is not endogenous with respect to education. This introduced region specific effects, for which the identification strategy will attempt to control.

[10]To make Wald estimates meaningful, estimates in this tables are presented for the sample with valid wage data. Results are similar for education in the complete sample.

program provision that more schools get built in lower enrollment regions is reflected in the differences between the education in low and high level regions. In both cohorts the average educational attainment in regions which received *fewer* schools is *higher* than in regions which received more schools. The same is true for wages. In both types of regions, education level increased over time. However it increased more in regions which received more schools. The difference in these differences can be interpreted as the causal effect of the program, under the assumption (which will be discussed below) that, in the absence of the program, the increase in educational attainment would not have been systematically different in low and high program regions[11]. An individual young enough, and born in a high program region, received on average 0.12 more years of education and the logarithm of his wage in 1995 was 0.021 higher. The differences in differences are not significant from 0. Controlling for region of birth and year of birth does not change the estimate, but reduces the standard error, and the estimate of the program effect on education becomes significant. This simple estimator suggests that one school per 1000 children contributed to raise education by 0.12 years (0.12 divided by 1.02) and wages by 0.021 for children aged 2 to 6 when the program was initiated. The ratio of these two estimate can be computed. This is the Wald estimate of returns to education, which under conditions which will be spelt out below is an estimate of average returns to education for people affected by the instruments.

This difference in differences estimator is comparable to the fixed effect procedure proposed for the evaluation of social programs in developing countries[12] As Strauss and Thomas (1995) point out in their assessment of the approach adopted by these papers, the identification assumption should not be taken for granted: the pattern of increase in education could vary systematically across regions. Moreover, the simple differences (the differences in education across cohorts and between regions) are large. This makes the difference in differences sensitive to assumptions about functional

[11]In regions were enrollment at the primary level were close to 100%, the increase of education should have come from increased enrollment at the junior high school level.

[12]Rosenzweig and Wolpin (1988), by Pitt, Rosenweig and Gibbons (1993), and Gertler and Molyneaux(1994).

form (Heckman (1996)). In particular, if the increase in education were negatively correlated with initial levels this pattern would be observed in the data even if the program had no effect.

Therefore, an interesting aspect of this experiment is that an implication of the identification assumption can be tested. Individuals aged 12 or older in 1974 were not exposed to the program. Therefore, in this age group, the increase in education between cohorts should not differ systematically across regions. This control experiment exploits the presence of multiple control groups (the successive cohorts not exposed to the program) (cf. Heckman and Hotz (1989) and Rosenbaum (1987) for similar ideas). Lack of pre-program information prevent most studies using fixed effect procedures to present similar evidence in favor of their identifying assumption.

In panel C, I present this control experiment. I consider a cohort aged 18 to 24 in 1974 and a cohort aged 12 to 17 in 1974. The estimated difference in differences is very close to zero for education and for wages, and not significantly different from zero. The assumption that the dynamics would not have been systematically different across regions in the absence of the program is not rejected by this pre-program test. These results suggest that the differences in differences are not driven by inappropriate identification assumptions, but this is far from not definitive evidence, in particular because differences in differences are imprecisely estimated. In panel C for example, the differences in differences are statistically insignificant from 0, but from the differences in differences in panel B as well. The remainder of this paper will elaborate this strategy to lead to more convincing results.

1.3.2 Conceptual framework

In this subsection, I use a simple version of the model of endogenous schooling developed in Card (1995,1999,2000), who draws from Becker (1967). I extend it to take into account the general equilibrium implications of the program. The goal of this exercise is to spell out explicitely under what conditions an identification strategy based on interactions of cohort and region contrasts is appropriate, and what evidence can be used to give credit to it.[13]

Individual's utility is written

$$U(w, S) = \ln w(S) - h(S),$$

where $h(S)$ is the cost of schooling function, and w(s) it the income of an individual with schooling S. I will assume that returns to schooling are linear, and person specific[14], i.e. income is written:

$$y_{ijk} = lnw_{ijk} = a_{ijk} + b_{ijk}S, \tag{1.1}$$

where y_{ijk} is the logarithm of the wage of an individual i born in region j in cohort k.

Following Card, the cost of schooling is written: $h(S) = r_{ijk} + \phi S$.

Optimal choice of schooling implies:

$$S_{ijk} = \frac{E_k b_{ijk} - r_{ijk}}{\phi}, \tag{1.2}$$

where $E_k b_{ijk}$ denotes the expectation of future returns to education at a time the individual makes his schooling decision.

Heterogeneity is modeled additively as follows:

$$b_{ijk} = b_{jk} + \nu_i$$
$$r_{ijk} = r_{jk} + \omega_i,$$

where b_{jk} (resp. r_{jk}) is the average returns to (resp. cost of) education for cohort k in region j and ν_i (resp. ω_i) is the individual deviation from the region average.

Returns to education for a cohort in a region are a function of regional and national

[13]This model was used by Card to interpret IV estimates of returns to schooling, and by Heckman and Vytlacil (1999) in their application of an IV estimator for the correlated random coefficient model.

[14]Card assumes concave returns to schooling. I abstract from this to focus on the most important assumptions underlying the identification strategy in this context.

economic conditions (Foster and Rosenzweig (1996)) average education in the region and in the country(due to general equilibrium effects (cf. Angrist (1996)) or positive externalities (cf. Foster and Rosenweig (1995,1996)) and the quality of education (cf. Berhrman and Birdsal (1983), Card and Krueger (1992)). I assume a linear form:

$$b_{jk} = 2\beta_1 S_j + 2\beta_2 \overline{S} + \beta_3 \lambda_j + \beta_4 q_{jk} + \beta_0, \tag{1.3}$$

where S_j is the average education in the region, S is the average education in the country (in 1995), λ_j are the economic conditions in the region (or any other region specific factor), and q_{jk} denotes the quality of education in region j for cohort k.

Cost of education in a cohort and a region depend on the number of schools per capita (the variable affected by INPRES), and other variables which will differ accross regions and cohorts (general infrastructure -such as roads- average income and wealth distribution in the region, etc...). I capture this by writing:

$$r_{jk} = \alpha_1 Z_{jk} + \alpha_2 \mu_{jk} + C, \tag{1.4}$$

where Z_{jk} denotes the number of schools per capita in region j at the time cohort k got an education, and μ_j denotes all the other region-specific factors.

Effect of the program on education

Consider a "young" cohort, fully exposed to the program (children aged 2 to 7 in 1974) and an "old" cohort, who was not exposed to the program (children aged 12 to 17 in 1974). Note $k = Y$ for the young cohort, and $k = O$ for the old cohort. To simplify notation, assume that the cohorts have the same size. The rational expectation equilibrium, assuming that the old generation did not anticipate the program (i.e. thought that cost and quality of education would not change between the time they receive and education and the time the young receives their education) implies the following expression:

$$
\begin{aligned}
(\phi - \beta_1)(S_{jY} - S_{jO}) &= \beta_2((E_Y\overline{S_Y} - E_Y\overline{S_Y} - \beta_1(E_O\overline{S_Y} - E_O\overline{S_O})) \\
&\quad + \beta_3\left(E_Y\lambda_{jY} - E_O\lambda_{jO} - \beta_1(E_O\lambda_{jY} - E_O\lambda_{jO})\right) + \beta_4(q_{jY} - q_{jO}) \\
&\quad - \alpha_1(Z_{jY} - Z_{jO}) - \alpha_2(\mu_{jY} - \mu_{jO}) + (\nu_{jY} - \nu_{jO}) - (\omega_{jY} - \omega_{jO}),
\end{aligned} \tag{1.5}
$$

where ν_{jk} is the average of v_i in region j, cohort k and ω_{jk} is defined in a similar fashion.

Equation 1.5 can be rewritten:

$$
S_{jY} - S_{jO} = \pi_0 + \pi_1(Z_{jY} - Z_{jO}) + \xi_j = \pi_0 + \pi_1 P_j + \xi_j \tag{1.6}
$$

where $\pi_1 = \frac{\alpha_1}{\phi - \beta_1}$, P_j denotes the INPRES allocation, and the expressions for the other terms follow directly from equation 1.5.

What is identified here is, at best, the reduced form parameter π_1. The parameters α_1, β_1 and ϕ are not separately identified. This shows that when we take explicitely into account general equilibrium effects, we can only estimate a mixture of a behavioral parameter (the elasticity of education with respect to school infrastructure) and the effect of the program on average education.[15]

We can now discuss under what conditions OLS estimation of equation 1.6 will produce consistent estimation of π_1.

• **Linearity**

One of the consequence of the allocation rule was that more schools were allocated in places where initial number of school per capita was low (this appears in table 3, panel 1). This implies that there might be a positive correlation between program effect and the allocation if the effect of school per capita is concave, in which case the OLS coefficient is an upward biased estimate of the average program effect. This can be addressed by estimating a non parametric version of equation 1.6. Kernel estimators will be presented. More generally, I assumed that the parameters of the

[15]This last term would be identified if the constant π_0 did not incorporate the difference between nationwise education in the two cohorts, so if Indonesian economy were fully integrated.

$r(.)$ and $b(.)$ functions do not vary accross regions. If they vary, the variation must be uncorrelated with the intensity of the program.

- **ξ_j and P_J uncorrelated**

First, $\nu_{jY} - \nu_{jO}$ and $\omega_{jY} - \omega_{jO}$ must be uncorrelated with P_j. This is a reasonable assumption, which will be satisfied if the distribution of individual heterogeneity did not change over time accross regions in ways correlated with the program (once we account for all regional effects). Next, the expected market conditions $(E_k\lambda_{jk})$, and the regions specific determinant of costs of education (μ_{jk}) must not have changed in ways correlated with the program. These assumptions are clearly strong, given the allocation rule. Recall that P_j is determined to a large extent by enrollment rate in 1972, and therefore of S_{jO}. Therefore these assumptions require that changes in expected and actual regional conditions be uncorrelated with initial levels.[16] Moreover, it will be violated if the allocation of other govermental programs initiated as a result of the oil boom and potentially affecting education, either directly (such as health programs or road construction) or indirectly through improvement of income levels[17] or expected returns to education were correlated with the allocation or INPRES schools.

- **Specification checks**

A first specification check is to run a regression of the difference in average education between two cohorts who were not affected by the program (the "young" cohort is the cohort aged 10 to 17 in 1974, and the "old" cohort is the cohort aged 18 to 24 in 1974) on the allocation of the INPRES program. The program was not yet in place, so the true coefficient should be zero. However, convergence or divergence in education levels due to unobserved variation in determinants of schooling costs and market returns would lead to a spurious coefficient. If the coefficient is indeed 0, it still leaves open

[16]Reverse causality *per se* (improvement in performance *causing* improvement in inputs, or dynamic feedback in the terminology of Gertler and Molyneaux) is not a potential problem in this case (unlike in Gertler and Molyneaux (1994) and Pitt, Rosenzweig and Gibbons (1993)) because the allocation rule was based upon enrollment in 1972, and was not updated over the period

[17]In practive, growth rates accross Indonesian provinces have been relatively uniform since 1966 and there is no evidence of convergence (Hill (1996)).

the possibility that there was indeed no convergence in education levels before the program but that there was convergence after the program was introduced.

A second specification check is therefore to use more specifically the timing of the program. An improvement in income levels or an improvement in roads or sanitation conditions should lead to an improvement in the education of *all* children in the affected regions, not only for children aged 12 or less in 1974. If we run successively OLS on equation 1.6 using in turn each year of birth cohort as the "young" cohort and the cohort aged 24 in 1974 as the "old" cohort, the identification assumption implies that the coefficients should be positive only for the cohorts aged 12 or less in 1974.

A third specification check is to consider how the probability of completing any given grade of education is affected by INPRES. The program changed the cost of education only at the primary level. There was a moderate increase in the provision of junior high school during the period, but not targeted to low enrollment regions in particular (Heneveld (1978)). There could be some spillover effect of the decrease in primary schooling cost on junior high school attendance, but we expect them to be smaller than the direct effects on completion of additional primary school grades. However, any omitted changes in regional conditions in equation 1.6 would be likely to influence the attendance of junior high school at least as much as attendance of primary school.

A final strategy is to explicitely control for education levels before the program[18] in regression 1.6, as well as for other regional variables that might influence education. In particular I control for the allocation of the water an sanitation program, which was the second largest INPRES program centrally administered at the time.

- **Changes in quality**

Lastly, π_1 will not be equal to $\frac{\alpha_1}{\phi-\beta_1}$ if the program has affected the quality as well as the quantity of education, which is conceivable with such a large program. The

[18]I present regressions controlling enrollment rate in 1971, which is the closest I have to the measure used by the Bappenas. I ran the same regressions using both enrollment in 1971 and average education in the cohort 1950-1960, which corresponds more closely to the specification here, and I find the very similar results

direction of the change in quality is, however, not obvious *a priori*. Several factors may play a role. First, a rapid increase in the quantity of education would probably cause a decrease in the quality, if new teachers could not be hired fast enough to accommodate the new pupils, or if their training was not adequate. If this effect were dominant, quality would have decreased more in regions which received more schools. Second, however, the quality of education in Indonesia was low before the program. In particular, schools were overcrowded, and some of the new schools were probably used to reduce overcrowding in the existing schools. Moreover, in 1975/76, INPRES instructions started to allocate a separate budget for the refection of existing schools. This would tend to increase the quality of education more in high program regions. I will present some evidence below on any modification in schooling quality. But note that if the quality of education changed as a result of the program, the parameter π still estimates a relevant parameter, namely the increase in education due to the program *after taking into account* any change in quality perceived by the individual when choosing its education level.[19]

Effect of the program on wages

We can now use equation 1.1 to express the average of the log of the 1995 earnings for a given region and cohort.

$$\begin{aligned} y_{jk} &= a_{jk} + \frac{1}{N_{jk}} \sum_{i=1}^{N_{jk}} b_{ijk} S_{ijk} \\ &= a_{jk} + b_{jk} S_{jk} + \frac{1}{N_{jk}} \sum_{i=1}^{N_{jk}} \nu_i \frac{\nu_i - \omega_i}{\phi} \\ &= a_{jk} + b_{jk} S_{jk} + \epsilon_{jk} \end{aligned}$$

where N_{jk} is the number of individuals in region j in cohort k and $\epsilon_{jk} = \frac{1}{N_{jk}} \sum_{i=1}^{N_{jk}} \nu_i \frac{\nu_i - \omega_i}{\phi}$.

[19]This is obvious if we write the change of quality as a linear function of the program plus an error term.

Using the definition of b_{jk} (equation 1.3) and replacing S_{jk} (equation 1.6), we get:

$$y_{jY} - y_{jO} = a_{jY} - a_{jO} + \bar{b}\pi_1 P_j + \beta_4(q_{jY} - q_{jO}) + \epsilon_{jY} - \epsilon_{jO} + b_{jO}(\pi_0 + \xi_j) + (b_{jO} - \bar{b})\pi_1 P_j, \tag{1.7}$$

where $\bar{b}$ is the average return to education in the old cohort.

I rewrite equation 1.7 in a similar form as equation 1.6:

$$y_{jY} - y_{jO} = \pi_2 + \pi_3 P_j + \eta_j, \tag{1.8}$$

where $\pi_3 = \pi_1 \bar{b}$.

OLS estimation will give a consistent estimate of π_3 only under all the conditions spelt out above, and an array of additional assumptions.

• **Correlation of $a_{jY} - a_{jO}$ and P_j**

The changes in the intercept between must not be correlated with the program. Three factors might play a role. First, other programs correlated with the INPRES program might have improved the earnings capacity of the young cohort (for example, the health of the young might have improved more in regions that received more schools). Second, experience profiles might differ systematically across regions, in a way correlated with the program. Third, there could be cohort effects in wage determination (the condition at the time a cohort entered the job market might continue to determine their wages in the future). In this case, convergence between region might be reflected though wages are observed in the same year for the whole sample. [20]

• **Correlation of region-specific returns to education and the program**

Returns to education might vary across region in a way correlated with the program. The program was a function of initial education. Since the relationship between returns to education and average education is an equilibrium relationship, it is not clear what the correlation between initial education (and therefore the program) and

[20] The difference $\epsilon_{jY} - \epsilon_{jO}$ must also be uncorrelated with the program (a special case of the assumptions in Wooldridge (1997), and Heckman and Vytlacil (1999)). It is not a strong assumption in this case. It will be satisfied if individual characteristics of young and old are drawn from the same distribution,after taking out region and cohort effects

the returns should be, empirically. Moreover, the more rapid increase in the number of graduates in these regions due to the program should have depressed the returns further, if negative general equilibrium effects are present. Both effects can potentially lead to a bias in the estimates in the effect of the program. I will show estimate of returns to education in low and high program regions to investigate whether there is a difference in returns betwen low and high program regions Moreover, specification checks can be used to examine whether these biases are important empirically.

- **Specification checks**

These concerns can be addressed by using the same "control" experiment and the same control variables as for education. We should not see the program have an effect on any pre-program cohort (whereas a spurious effect will be obtained if these factors are at play), and we can control explicitly for determinants of the program (education level in 1971) or other programs (water and sanitation program). Controlling for education will in particular remove any correlation between the program and the returns to education due to the program targetting rule. However, conditionally on initial education, returns will tend to be lower in the high program regions, due to the decline in returns to education following the increase in educational attainment in the regions. This should lead to a downward estimate of the program effect, if these general equilibrium effects are important. But they should be lower for older generation as well. So if this effect is important, this would lead to a *negative* spurious program effect in the control experiment.

- **Changes in quality**

To write $\pi_3 = \pi_1 \bar{b}$, we need to assume the quality of education did not change as a result of the program. Even if quality changed as a result of the program, total effect of the program on education (and on wage) can still be identified. But the coefficient π_3 will not reflect both changes in average education and changes in returns to education due to the change in quality. I will show below direct and indirect evidence on how quality was affected.

If these assumptions are satisfied, the average returns to education can be then calculated by dividing π_3 by π_1 This is the indirect least squares estimate, a simple instrumental variable estimate of returns to education. If the last assumptions is not satisfied but all the other are, then the reduced form effect of the program can still be estimated, but dividing π_3 by π_1 will not result in an unbiased estimate of $\bar{b}$. The Wald estimate presented above was in effect a simpe ILS estimate, with program intensity set to 1 or 0.

To conclude, recall that this formulation is valid under the assumption of linear individual heteroneity and returns to education. Under these assumptions, changes in education caused by the program are not correlated with individual ability. If individuals differed in ϕ, the *marginal* cost of schooling instead, individual response to the program might depend on returns to education. This effect is not captured in this simple model but is clearly a possibility in reality. As shown in Imbens and Angrist (1994) in the case of a binary instrument, if returns to education vary by individuals and the assumptions of independence of the instruments and monotonicity are satisfied, the instrumental variable estimator (e.g. the Wald estimate presented in table 3 or the ILS estimator here) measures averages returns to education for the individuals who, as a result of the program, change their level of education. If the individuals affected by the program tend to be individual with higher (or lower) returns in the poorer regions, the IV estimator is therefore not an estimate of average returns in the population. It is still, however, a causal parameter of interest for policy evaluation, since it measures the returns for people affected by this policy.

1.4 Effect on education

1.4.1 Basic results

I start by estimating equation 1.6. In practice, I run:

$$S_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + (P_j * T_i)\gamma_1 + (C_j * T_i)\delta_1 + \epsilon_{ijk} \ , \tag{1.9}$$

where T_i is a "treatment dummy" indicating whether the individual belong to the "young" (or treated) cohort in the sub-sample, P_j is the intensity of the program in the region of birth. C_j are control variables (following the discussion in the preceding section, I introduce successively the enrollment rate in 1971 and the allocation of the water and sanitation program).[21]

In table 4 (columns (1) to (3), I present estimates of equation 1.9 for two subsamples. In panel A, I compare children aged 2 to 6 in 1974 with children aged 12 to 17 in 1974. The suggested effect is that one school built per 1000 children increased the education of the children aged 2 to 6 in 1974 by 0.12 to 0.18 years for the complete sample, and 0.19 to 0.22 for the sample of wage earners. Controlling for enrollment rate and water and sanitation program makes the estimates higher, suggesting that the estimates are not biased upwards by mean reversion. In panel B, I show the results of the control experiment (comparing the cohort aged 12 to 17 to the cohort aged 18 to 24 in 1974). The impact of the "program" is very small and never significant. The coefficients are statistically different from the corresponding coefficients in panel A.

Figure 1 plots the difference in education between the young and the old cohort against the program intensity in each region. The regression line corresponds to the GLS estimation of equation 1.6 (the coefficients are presented in table 4, column (1)). I plotted in addition the kernel estimator, which shows that the effect of the program in education is approximately linear. In the control experiment (panel B), the regression line as well as the non-parametric regression are flat.

1.4.2 Reduced form evidence

The identification strategy discussed in the previous section can be generalized to an interaction terms analysis.

Consider the following relationship between the education (S_{ijk}) of an individual i born in region j in year k and his exposure to the program:

$$S_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + \sum_{l=2}^{23}(P_j * d_{il})\gamma_{1l} + \sum_{l=2}^{23}(C_j * d_{il})\delta_{1l} + \epsilon_{ijk}, \tag{1.10}$$

[21] This is equivalent to a GLS estimation of equation of 1.6.

where c_1 is a constant, α_{1j} is a region of birth effect, β_{1k} is a year of birth effect, d_{il} is a dummy that indicates whether individual i is of age l in 1974 (a year of birth dummy), P_j is a measure of the intensity of the program in region j. and C_j is a vector of region specific variables (population in 1971, enrollment rate in 1971 and water and sanitation program). In these unrestricted estimates, I measure the time dimension of exposure to the program with 22 age dummies (for being 2 to 23 in 1974). The omitted dummy is the dummy for being 24 in 1974 (individuals aged 24 in 1974 form the control group). Each coefficient γ_{1l} can be interpreted as the estimated impact of the program on a given cohort. This is just a generalization of equation 1.6 to estimate cohort-by cohort contrasts. Because children aged 13 older in 1974 did not benefit from the program, the coefficients γ_{1l} should be zero for l greater than 12, and start increasing for l less than some threshold (the oldest age at which an individual can have been exposed to the program and still benefit from it). The only *a priori* restriction about this threshold is that it is smaller than 12.

Table A1 (appendix) presents unrestricted reduced form estimates of these three specifications. These reduced form estimates allow to check whether the γ_{1l} in equation 1.10 follow the expected pattern. In figure 2, I have plotted the γ_{1l}.[22] Each dot on the solid line is the coefficient of the interaction between a dummy for being of a given age in 1974 and the number of schools constructed per 1000 children in the region of birth (a 95% confidence interval is plotted in broken lines). Each dot summarizes the effect of the between-regions variation in program intensity on a given cohort. For example, a child aged 6 in 1974 received 0.2 additional years of education if he was born in a region which received one more INPRES schools per 1000 children. These coefficients bounce around zero until age 12, and start increasing after age 12. As expected, the program had no effect on the education of cohorts not exposed to it and it had a positive effect on the education of younger cohorts. All coefficients are significantly different from 0 after age 8. To show more clearly the rupture in the trend, I plotted in figure 3b a smoothed version of the same data: for each l, I plotted

[22]I have plotted the coefficients corresponding to the specification in column 2, table A1.

the average of γ_{1l}, γ_{l-1} and γ_{l+1}. The coefficients are close to 0 until age 11 and then they increase sharply.

This pattern is similar across specifications. As it was discussed in the previous section, any omitted time-varying and region specific factor affecting education would probably imply that coefficients γ_l would start to increase for some cohort l greater than 12, even if these factors had changed at the same time of the program. From these graphs, it therefore appears that the identification strategy is reasonable, and that the program had an effect on education

1.4.3 Restricted estimation

Instead of testing whether the γ_{1l} are equal to zero for $l \geq 13$, one can impose these restrictions. The equation to be estimated is then:

$$S_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + \sum_{l=2}^{12}(P_j * d_{il})\gamma_{1l} + \sum_{l=2}^{12}(C_j * d_{il})\delta_{1l} + \epsilon_{ijk}. \tag{1.11}$$

The omitted group (the control group) is now formed of individuals aged 13 to 24 in 1974. This is a more efficient way to estimate the program effect and leads to more precise estimates.

Columns 1 to 3 in table 5 show the coefficients of the interactions between age in 1974 and the intensity of the program in the region of birth in three specifications for the complete sample (columns 4 to 6 show the same results for the sample of wage earners). In all columns, the estimated effect is positive after age 10. All coefficients are significantly greater than 0 after age 8. All sets of interactions are statistically different from 0 (the F-statistic for the null hypothesis is presented at the bottom of the table). The coefficients generally increase with date of birth (decreasing with age), except for a high value at age 9 and a decline between age 6 and age 5. They increase faster between age 12 and age 9 than subsequently: this fact indicates that once the education level in the population reaches a certain level, increasing it by building primary schools becomes less effective.

The estimates in column 1 (without controls) suggest that one school per 1000 children increases the education of the youngest children by 0.16 years. On average, 1.98 schools were built per 1000 children. This implies that at its mean value, the program caused an increase in education of 0.32 years for these children (the average education in the sample is 7.8 years). As before, controlling for enrollment rate in 1971 (column 2) and the water and sanitation program (column 3) make the estimate slightly higher. In column 4 to 6, I present the same estimates for the subsample of wage earners. The program effect is higher for wage earner than in the complete sample.

More insight about why and how this program was effective is given by examining its impact in different types of regions. In table 7 (panel A), I present results equivalent to the specification in table 4 (equation 1.9) for various sub-samples or regions of birth (estimates of equation 1.11 for these sub-samples are presented in table A2, in the appendix). In column 1, I recall the result for the whole sample. In column 2 and 3 I present the program effect in sparsely and densely populated regions. In sparsely populated regions, each school constructed per child is likely to reduce the distance to school significantly (if the schools are placed relatively evenly in space). In densely populated regions, the main effect will be not to reduce the distance to school, but to increase slots availability or to reduce the overcrowding of old schools. Therefore the difference between the program effects in these two types of regions will give some information on whether distance to school or overcrowding of schools was the most important cause of the effect of the program. The results (in columns 2 and 3) indicate that the program effect is 0 in densely populated regions, while it is 0.22 sparsely populated regions. This suggests that reducing the distance to school was the most important effect of the program.[23] In column 4 and 5, I present the results in provinces where the incidence of poverty in 1976 was higher and lower than the Indonesian average. I find a larger effect in poor provinces. In column 6 and 7, I divide the sample into regions where the education of the cohort not exposed to the program (men born between 1950 and 1962) was lower or higher than the median

[23]Although this should be taken with caution, since this difference may come from other characteristics correlated with density, and not taken into account here.

(3.08 years of education). Results are similar for both sets of regions.

In summary, it appears that the school construction program had a significant impact on education. The causal interpretation of these estimates is supported by pre-program tests. It should be recalled that this program was accompanied by a general effort by the Indonesian Government in favor of education, a priority of the second five-year plan. As part of this effort, primary school fees were suppressed in 1978 (World Bank (1989, 1990)). These results can therefore not be generalized to less favorable contexts without applying caution.

1.4.4 At what level of education was the program effective ?

The consequences of the program on welfare depend on whether it affected mostly children with a low or a high level of education. It is therefore important to examine at what level of education the program was effective. The simplest way to investigate this question is to use a difference in differences estimator.[24]

I group the regions into high and low program regions and consider a cohort aged 2 to 6 in 1974 and a cohort aged 12 to 17 in 1974. Instead of considering only differences in group means, I consider differences in the cumulative distribution functions of education (the probabilities to complete any given level of education or less). Figures 4 and 5 show that in both regions, the CDF of education in the younger cohort stochastically dominates the CDF of education in the older cohort. Moreover, the CDF of education in the low program regions stochastically dominates the CDF of education in the high program regions for both cohorts, a consequence of the higher level of the program in regions with lower initial enrollment. Figure 6 plots the differences in CDF between the two cohorts for both type of regions. The between-cohort differences in CDF is larger in the low program region for the first five years of education, but lower after the sixth year (the last year of primary school).

[24]The government may care more about those children who, *in the absence of the program*, would have had the least education. However I cannot really answer this question by examining how the distribution of education is affected by the program (because the difference in quantiles is not the quantile of the difference).

Figure 7 shows the difference in differences in CDF (The 95% confidence interval is plotted in broken lines). The dot for the fifht year of education, for example, indicates that the program induced 6 percent of the sample to complete 6 years of education of more (i.e. graduate from primary school) instead of five or less. The program shows a positive effect of the program at all primary school levels. It had no effect at the junior high school level. A significantly negative effect of the interaction is shown for levels of education of 9 years and above.

We can get more precise estimates by including a full set of region of birth and year of birth dummies. To estimate an equivalent of this difference in differences controlling for these variables, I estimate a linear probability model for the probability of completing m years of education or less, for $m = 0$ to 19. For S_{ijkm} a dummy which indicates whether the individual i born in region j in year k completed m years of education or less, and for P_j a dummy indicating whether the child is born in a high program region, I estimate the following equation:

$$S_{ijkm} = c + \alpha_j + \beta_k + (P_j * T_i)\kappa_m + \epsilon_{ijk} \tag{1.12}$$

The κ_m, for $m = 0$ to 19 are the values of the estimated impact of the program at each level of education. They are plotted in figure 8 (the 95 % confidence interval is plotted in broken lines).

The shape of this function and the shape of the function estimated from the difference in differences in the CDF are similar. Both are rising until the fifth year of education, decreasing until the 12th and mildly increasing thereafter. A maximum of about 6 percent of the sample was induced by the program to complete at least primary school. This also shows some impact of the program on the probability of completing lower secondary school (1.5% of the sample is estimated to have been induced by the program to complete 7th and 8th and 9th grade or more).[25] This is reasonable: by increasing the probability of completing primary school, the program makes it more likely that somebody will actually enter lower secondary school, since

[25]The shape of this function is not affected by controlling for interactions of enrollment rate (or water and sanitation program) and year of birth dummies.

the satisfactory completion of primary school is a prerequisite for entering lower secondary school.[26] This spillover is nevertheless limited. The primary shool construction program did increase primary schooling essentially. This is a interesting result: the program was effective in increasing primary school enrollment only. This provides additional evidence that the assumption underlying the identification strategy is reasonable. The estimated effect of the program for the levels of education which it did not target is small or nul. For people who would have completed primary school in any case and were considering whether to go on to junior high school, the marginal cost of going to junior high school was not affected by the program, so they should not be affected. But the program could have induced more marginal people to complete primary school and move on to college. The fact that the program had little effect beyond primary school can be understood, even if parents are optimizing dynamically, if we recall that the direct and indirect costs of junior high school are much higher than the costs of primary education, and were not equalized across regions at the time. People who were induced by the program to complete primary are the ones who were facing high cost of primary schooling before the program. Therefore they must be facing high cost of junior high school after the program, since these were not affected by governmental intervention. This can explain why we do not observe large spillovers. The test of the human capital versus sorting models of returns to education proposed by Lang and Kropp (1986) provides another interesting light for this result. Lang and Kropp show that the sorting model implies that compulsory attendance laws, which affect the education of the low skilled workers, should also affect the education of the high skill workers (who have to get more education to show that they are different). Under the human capital model, compulsory attendance laws should not affect the education of people who are not directly constrained by them. The INPRES program directly affected primary education only, but under the sorting model, it could have lead some people who would have completed more than primary school to complete more years of education. The results in this paper seem to indicate that the human

[26]Similarly, Angrist and Imbens (1995) find that compulsory attendance law induced a fraction of the sample to complete some college as a consequence of constraining them to complete high school.

capital model of education seem to describe Indonesia better than the sorting model.

The program was effective in increasing education, in particular at the primary school level. Did it increase human capital? One way to answer the question is to look at the effect of the program on wages.

1.5 Effect on wages

1.5.1 Basic results

I start by estimating a specification equivalent to equation 1.8 for the experiment of interest and the control experiment. As for education (equation 1.9), I estimate in practice:

$$y_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + (P_j * T_i)\gamma_1 + (C_j * T_i)\delta_1 + \epsilon_{ijk}. \tag{1.13}$$

Results are presented in table 4 (columns 4 to 6) and figure 1. In table 4, panel A, I set T_i equal to 1 for children aged 2 to 6 in 1974, and I use children aged 12 to 17 as the comparison group. In figure 1 (panel A), I plotted the increase in wages between the same two cohorts against the program intensity, as well as the GLS regression line (the coefficient of which are given in column (4) in table 4) and the kernel regression line. In table 4, panel B, I set T_i equal to 1 for children aged 12 to 17 in 1967, and used children aged 18 to 24 as the comparison group. Corresponding evidence is plotted in figure 1 (panel B).

In panel A, the estimates range from 1.5 % to 2.2%. These numbers measure the average increase in wages caused by 1 school built per 1000 children in their region of birth, for children aged 2 to 6 in 1974. As for education, the estimate increase when I control for enrollment rates in 1971 and for the allocation of the water and sanitation program, although none of these estimates are significantly different from each other. In panel B, in all specifications, the interaction coefficient is small and not significantly different from zero. However, these estimates are imprecise and I cannot reject equality of the coefficients in panels (although the point estimate are much smaller in panel B). Qualitatively, the results of estimating this reduced form expression nevertheless lead to similar conclusions than for education. The program seemed to have an effect on average wages, the estimate are not smaller when control variables are introduced, and the point estimate of the program effect in an untreated sample is smaller and close to zero.

1.5.2 Reduced form evidence

As for education, we can now write an unrestricted reduced form relationship between the exposure to the program and the logarithm of the wage of an individual (y_{ijk}).

$$y_{ijk} = c_2 + \alpha_{2j} + \beta_{2k} + \sum_{l=2}^{23}(P_j * d_{il})\gamma_{2l} + \sum_{l=2}^{23}(C_j * d_{il})\delta_{2l} + v_{ijk}. \tag{1.14}$$

where y_{ijk} is the logarithm of the hourly wage of an individual i born in the year k in region j, α_{2j} is a region of birth effect and β_{2k} is a cohort of birth effect. P_j , $C_j$and d_{il} are defined as in the education equation: P_j is the intensity of the program in the region of birth, C_j is the vector of control variables and d_{il} is a dummy indicating whether individual i was of age l in 1974.

The γ_{2l} should be zero for l greater then 12 and start increasing after some threshold. Moreover, if the program affected wages only through its effect on education, the coefficients γ_{2l} should track the γ_{1l} (in the education equation). In particular the threshold after which the coefficients γ_{2l} start to increase should be the same as the threshold after which the coefficients γ_{1l} start to increase. The γ_{2l} should also track the up and downs of the γ_{1l}.

Table A1 (appendix) presents the results for the three specifications for which I estimated the education equation. Again, graphs help to interpret the reduced form coefficients. In figure 3a, the γ_{2l} are on plotted in the dotted line. They are oscillating until age 10 and start increasing after age 11. The coefficients of the interactions for education and wages track each other. Figure 3b presents the same data, but shows more clearly the impact of the program. The values on this graph are smoothed and the scales are adjusted. In this figure, the change in trend after age 11 is very apparent. The program start having a positive effect on wages and education at that point.

1.5.3 Restricted estimates

In column 7 to 9 in table 5, I present estimates of the equation:

$$y_{ijk} = c_2 + \alpha_{2j} + \beta_{2k} + \sum_{l=2}^{12}(P_j * d_{il})\gamma_{2l} + \sum_{l=2}^{12}(C_j * d_{il})\delta_{2l} + v_{ijk}. \tag{1.15}$$

It is more difficult to precisely estimate the effect of the program on wages than on education, because wages fluctuate more and the sample is smaller (since wages are not collected for self-employed people). I find therefore that few coefficients are individually significant, and that the F. statistics for the significance of the joint set of instruments are small. Nevertheless, the results mirror the estimates of the effects of the program on education. No effect is found for children 10 or older in 1974, and then the coefficients become positive (except at age 7). The interaction coefficients are generally decreasing with age. The estimates are higher when I control for enrollment rate and the water and sanitation program. The last line in this table indicates that constructing one school for 1000 children increased the 1995 wages of individuals aged 2 in 1974 by 1.6% to 3.7 % . The average number of schools constructed per 1000 children is 1.89 in the sample with valid wage data. Therefore, on average, the program caused a 3% to 7% increase in the wages of this cohort.

In table 7 (panel B), I present these estimate of the specification 1.13 for different sub-samples (estimates of equation 1.15 are presented in table A2 in appendix). The variations of the program effect across sub-samples parallel the variations of the program effect on education. In particular, we see no effect on wages in regions where there is no effect on years of education. This suggests that the program effect on wages was probably caused by the changes in years of education. In the next section, we use this to construct instrumental variables estimates of the effect of education on wages.

1.6 Estimating returns to education

The identification assumption that the evolution of wages and education across cohorts would not have varied systematically from one region to another, in the absence of the program, is sufficient to estimate the impact of the program. Since the intervention was to build primary schools, the program effect on wages was most probably caused by changes in education. At the cost of the additional assumption that the increase in the quantity of education was the only channel through which the program affected wages, I can use this program to construct instrumental variables estimates of the impact of additional years of education on wages. The most serious concern, for this interpretation, is that the program might have affected both the quality and the quantity of education, and that changes in wages could reflect both effects. Below, I examine whether there is evidence that this is a serious problem.

1.6.1 Indirect least squares estimates

Following the discussion in section 3, we can calculate indirect least squares estimates of returns to education, simply by dividing the estimate of the program effect on wage and on education. For example, dividing the estimate of panel A, column 4 by the corresponding estimate in column 1 for wage earners, we obtain an estimate of average returns to education of 7.7%. Adding control, we find respectively 8.3% and 9.73 %. This strategy can be extended to use all the information available, using a 2SLS strategy instead of the ILS approach.

1.6.2 Two stage least squares estimates of the returns to education

Estimates of equations 1.11 are of intrinsic interest because they provide an assessment of the impact of the program on education. But they also represent the first stage of a two-stages least squares estimation of the impact of education on wages. Recall equation 1.1, used used to characterize the causal effect of education on wages:

$$y_{ijk} = a_{ijk} + b_{ijk} S_{ijk}$$

Rewrite this expression as:

$$y_{ijk} = d + \alpha_j + \beta_k + S_{ijk}\bar{b} + \eta_{ijk} \ , \tag{1.16}$$

where α_j and β_k design region of birth and cohort of birth specific effects. The region specific error terms, the cohort specific error terms, and the individual error incorporates individual and regional differences in returns to education and in the specific intercepts.

Under the assumptions discussed in section 3, the interactions between the age in 1974 and the program intensity in the region of birth are available as instruments for equation 1.16. By limiting the set of instruments to the interactions in equation 1.11 (the age dummies for $l \leq 12$), I avoid potential small sample bias caused by the use of many weakly correlated instruments. Moreover, the instruments have been shown to have high explanatory power in the first stage, which indicates that the 2SLS estimates should not be affected by this problem.[27] I also estimated the same equation using a single instrument, the interaction of being in the "young" cohort and the program intensity in region of birth. Equation 1.16 can also be modified to incorporate control variables:

$$y_{ijk} = d + a_j + b_k + S_{ijk}\bar{b} + \sum_{l=2}^{12} (C_j * d_{il}) \ \pi_l + \eta_{ijk}. \tag{1.17}$$

The results are presented in panel A1 of table 6 (panel A2 presents results with the logarithm of monthly earnings as the dependent variable). The first line shows the OLS estimate. The estimated returns to education is 7.8% and is not affected by introducing control variables. This is lower than OLS estimates in Indonesia in older sample, but consistent with estimates in other samples in the 1990, and with

[27]I have also estimated this equation using LIML, which is median unbiased and does not suffer from this problem, and I find very similar result, another evidence that there is no correlated instruments problem.

the decline in estimated returns to education over time: World Bank (1990) reports estimates decreasing from 19% in 1982 to 10% in 1986.

The second line present two-stages least squares estimates of equation 1.17 (the number in square bracket is the test statistic for the overindentifying restriction). In column 1 to 3, I present the 2SLS estimates for the three specifications used throughout the paper. In column 1, I present the IV estimates without any control variables: I cannot reject that the estimate is equal to the OLS estimate, but the point estimate is slightly lower. In column 2, I introduce interactions between the enrollment rate in 1971 and year of birth dummies. The point estimate is higher than without control (7.9%), and very close to the OLS estimate. When I introduce a control for the water and sanitation program, the estimate is again slightly higher (9.1%). In the third line, I present the IV estimate using only one instrument. They are very similar to the IV estimate using more instruments (but slightly more imprecise, since they are using less variation).

These IV estimates are not very different from the OLS estimate. Behrman and Deolalikar (1993) introduce household fixed effects in an earnings function in Indonesia and report estimates that are much lower than corresponding OLS estimate.[28] The expectation in the development literature is in general that OLS estimates are likely to be biased upwards due omitted family and community background variables, which does not seem to be confirmed in this case. (Behrman (1990), (Strauss and Thomas (1995)). On the other hand, most studies in industrialized countries find IV estimates that are higher than OLS estimates, wich is also not what I find here. (Card (1995,1999, 2000), Ashenfelter and Harmon (1999)). Card (1999) discuss several reasons that might explain why IV estimates tend to be higher than OLS. I consider them briefly to examine whether they apply in this context. The first explanation, proposed by Grilliches (1977) is that ability biases in the OLS estimate of return to schooling are relatively small, and that the gaps between the IV and OLS estimates

[28]But introducing fixed effects into a wage equation reduces the sample, introduces additional selection problems, and exacerbates the attenuation bias due to measurement errors. This might explain why their fixed effect estimates are lower than the OLS estimates.

in most studies reflect the downward bias in OLS attributable to measurement error. In this light, my results would be consistent with the idea that the ability bias and the measurement error bias more or less cancel in the Indonesian context. A second explanation is that the IV estimate are even further upward biased than the OLS estimate by unobserved differences in earning ability between the "treatment" and "control" group. These differences in earnings are then "blown up" when they are divided by small differences in schooling. Card note that this problem is likely to be less important when the IV estimates exploit large variations in education and when the instrument exploits interaction between two sources of variations. My study fulfils both criteria, which may make it less biased than other IV estimates in the literature. A third explanation, proposed by Ashenfelter and Harmon (1999) is that researchers prefer to estimates with high t. statistic. With a relatively large sample and an important variations in schooling, the standard error of the IV estimates are small enough than even estimates smaller than OLS have still t. statistics above 2. Finally, Card's own explanation is that the 2SLS estimate are not estimates of average returns to schooling, due to the possibility I described at the end of section 3: people affected by the instruments might be people who have higher marginal returns to schooling. In particular if returns to education are concave and people with low levels of schooling are more affected than other people, this will be true. Only people who would have completed less than primary school were affected by the program, as described in section 4. However, there is no evidence that returns to education are concave in Indonesia. Estimating non-parametrically the shape of the true causal response function would be difficult, since the source of exogenous variation I use in this study affects only primary education. But some indication that the returns are not concave is given by OLS estimation using a dummy for each year of education. These coefficients are plotted in figure 10. Estimated marginal returns vary little until 9 years of education but are apparently high for the twelfth year of education (the last year of senior high school), and the thirteenth year of education. Similarly Strauss and Thomas (1997) found evidence of apparent convex returns to education in urban

Brazil.[29] If returns are in reality linear or even convex in developing countries, the phenomenon of "discount rate bias" (Lang (1993)) emphasized by Card has no reason to be present, which would again explain why OLS and 2SLS are similar in my study.

In table 7 (panel C), I examine whether returns to education vary across regions. The first column recalls the results for the complete sample (table 6, column 3). The next columns present results for the various sub-samples for which I have estimated the program effect on education (I have not presented the 2SLS estimate when the F. statistic for the joint significance of the instruments in the first stage was below 2, because they would not be interpretable). Returns to education do not vary substantially across regions. The noticeable results is that they are higher in sparsely populated regions and in regions where the average education level of cohorts not exposed to the program is low (they reach respectively 10% and 11%). This last result is consistent with the idea that the general equilibrium effect of an increase in education is to depress the returns, but indicate that returns before the program were not lower in high program regions

I now turn to two potential sources of bias: the assumption that the program had no impact on wages other than through the increase in the quantity of education, and problems arising from sample selection.

1.6.3 Could change in quality bias the 2SLS estimates ?

As I discussed in section 3, estimates of returns to education are biased if the program affect both the quality and the quantity of education. I present two pieces of evidence suggesting that the program did not substantially affect the quality of education.

First, in panel A of table 3 (right panel), I present difference in differences in average pupils/teachers ratio across districts.[30] In both regions, the average pupils/teachers ratio increased slightly over the period. In both years, pupils/teachers

[29]Note that the OLS coefficients indicate a correlation, and are not causal estimates: these high coefficients could be an artifact of selection into secondary education.

[30]The means are calculated from the number of pupils and the number of teachers collected in 1973/74 and 1978/79 at the district level by the Ministry of education

ratio are higher in the high program regions. The difference in differences is very close to 0. No systematic difference in quality, as measured by this indicator, is apparent. I also ran a regression of the change in pupils/teachers ratio on the number of schools per children built in the program. The coefficient is negative, very small, and not significantly different from zero.

Second, I use the fact that the program did not increase the education of people completing 9 years of education or more (as shown in section 3). For these people, the quantity of education was not affected by the program. Therefore, wages should not be affected either. In figure 9, I show the coefficients of the interactions between the program intensity and age dummies in the wage and education equations, in the sample of people whose level of education is larger than 9.[31] In contrast to figures 1 to 3, no specific pattern emerges in either equation: the interaction coefficients are fluctuating (they become negative for the youngest individuals in the education equation), and there is no rupture in trend after age 12. The evidence in table 7 (and table A2 in appendix) can be interpreted along the same line. In the densely populated regions (column 4), there is no effect of the program on years of education. If the quality of education had changed and this had affected wages, then we should see an effect of the program on wages even in this region. But there is no effect on wages either.[32]

These two separate pieces of evidence lend some support to the additional assumption that the increase in wages was due mainly to the increase in the quantity of education. There is no clear evidence against the assumption that the program affected only the quantity of education or that it affected the quality among dimensions which did not really matter for subsequent wages.

[31]Note that I partition the sample according to education, which is an endogenous variable. It not ideal, but it is not likely to be a real problem in this case, because the decision to go to senior high school is quite different from the decision to complete one more year of primary school (which is the endogenous decision I consider in this paper) -and the population who complete senior high school is also different.

[32]Even though the data shows that the program reduced pupils/ teachers ratio in densely populated regions, which is expected, given that it increased the number of teachers but not the number of pupils in these regions. This suggest that pupils/ teachers ratio did not have a big impact on effective school quality.

1.6.4 Correction for sample selection

These estimates are performed on a selected sample; only 45% of the individuals in the sample are working for a wage. Most remaining individuals are self-employed.

The probability of working for a wage is potentially affected by education. To examine this, I use two-stage least squares to estimate:

$$w_{ijk} = d + a_j + b_k + e_{ijk}\lambda + \sum_{l=2}^{12}(C_j * d_{il})\ \pi_l + \eta_{ijk}, \tag{1.18}$$

where w_{ijk} is a dummy variable indicating whether an individual reports a positive wage. Estimates of this equation are presented in table 6, panel B1. The OLS estimate is much smaller than the IV estimates (the estimate of the impact of an additional year of education on the probability of working for a wage changes from 3.3% to 10.4%). The probability of working for a wage is affected by education.

This is an interesting result in itself, because this outcome is not affected by sample selection. However this casts some doubts on the validity of the 2SLS estimates of returns to education. Because the probability of working for a wage is also affected by schooling, conditioning on the fact that wages are observed is likely to induce a correlation between the instruments and the error in equation 1.17 (the conditional expectation of the error given the instruments and the fact that an individuals reports a positive wage may not be zero). The ideal solution to this selection problem would be to find an instrument that is randomly assigned in the population with positive wages. Failing that, several econometric solutions have been proposed. Most rest on exclusion assumptions which would be strong in this context.

I implement two alternative procedures to investigate whether sample selection is likely to be an important problem in this case. First, I implement a sample correction procedure suggested by Heckman and Robb (1986), which is to introduce the probability of selection given the instruments in the estimation of the second stage. Second, I use the special income and expenditure module of another Indonesian survey (the SUSENAS, 1993) to impute and income for the self-employed individuals in my sample. Results do not seem to be sensitive to either modification.

First, I follow a suggestion Heckman (1980, 1986), then discussed by Ahn and Powell (1993), and Angrist (1995), which is to condition in the second stage on the probability of selection given the instruments.

In practice, I estimate the equation:

$$w_{ijk} = c_3 + \alpha_{3j} + \beta_{3k} + \sum_{l=2}^{12}(P_j * d_{il})\gamma_{3l} + \sum_{l=2}^{12}(C_j * d_{il})\delta_{3l} + \epsilon_{ijk}, \tag{1.19}$$

and I use the predicted value for w_{ijk} (the probability of being selected given the instruments) and the squared of the predicted value as additional regressors in equation 1.17:

$$y_{ijk} = d + a_j + b_k + S_{ijk}\bar{b} + \sum_{l=2}^{12}(C_j * d_{il})\ \pi_l + \widehat{w_{ijk}}\mu_1 + \widehat{w_{ijk}}^2\mu_2 + \eta_{ijk}, \tag{1.20}$$

where $\widehat{w_{ikj}}$ is the predicted value from equation 1.19. The instruments are interactions between year of birth dummies (for people 12 or younger in 1974) and program intensity in the region of birth.

The result of the introduction of the correction for sample selection is presented in column 5 in table 6 (panel A1) . The change in the coefficient is small: it changes from 7.9% (in column 3) to 8.9%. To check the sensitivity of this result to functional form assumptions, I have run similar specifications controlling for higher order terms of the selection probability (cubic and fourth power). The estimates are remarkably stable.[33]

I applied the same sample correction procedure to the other specifications. Conventional estimates and the selection-correction strategy generate similar results. This suggests that selection bias does probably not have a big impact in the estimation of the coefficient of education.

A problem with this procedure is that I use functions of the year and region

[33]The coefficient of education when a cubic term of the selection probability is introduced in the equation is 0.088(0.042), with a fourth power it is 0.90(0.042). When I introduce the propensity score alone, the cceefficient of education is 0.075(0.041).

of birth interactions both as regressors and as controls. The second stage is still over-identified (there are now 12 instruments for 3 parameters to estimate). But the identification is fragile (it rests on the fact that I use several instruments to measure the program intensity). This leads, in particular, to larger standard errors. An alternative approach is to impute an income to self employed individuals, and examine whether the results change when the estimation is performed in this "completed sample".

To this end, I use the income and expenditure module of the 1993 SUSENAS survey. Over 50,000 individuals are included in this module.[34] Households report the occupation and the sector of activity from which they derive their main source of income. In addition, the survey collects information on wages received by each member of the household, income derived from the sale of products and services of the household business (or farm), and operating expenses related to it. I define the income accruing to the household from each household business as the difference between sales and expenses for this business. I calculated the average income derived from the main activity of the household for cells defined by sector (9 industrials sectors and services and 4 types of agricultural activities) status and urban/rural residence. To check the consistency between the two sources, I report in table 1 the average monthly income of wage earners and their average income imputed using this procedure. The two figures are quite close. The difference is explained by the fact that the SUSENAS was fielded in 1993, while the SUPAS was fielded in 1995. The average monthly income for self-employed individuals is smaller.

The goal of this exercise is to examine whether the results are sensitive to the inclusion of self employed individuals in the estimation. Therefore I "complete" the sample by defining the dependent variable as the log of monthly earnings if they are recorded in the SUPAS data (for individuals working for pay) and the log of the average income from the SUSENAS in the individual's category for self employed individuals (multiplied by the wage inflation factor defined as the ratio of the average

[34]The SUSENAS does not report the place of birth of the individuals, and could therefore not be used as the main sample for this analysis

wage from the SUPAS and the average imputed income of wage earners).[35]

The results are presented in panel B2 of table 6. They must be compared to the results in panel A2, where the dependent variable is the logarithm of monthly earnings of wage earners. In all cases, the estimates using the completed sample are smaller than the estimation on the sample of wage earners. They are in general quite close (within 1 percentage point), except in the specification controlling for the water and sanitation program (column 4), where it drops from 7.6% to 4.9%. This particular result is surprising. The general fact that the returns for the complete sample are smaller than the returns estimated in the sample of wage earners is, however, expected (returns to education are in general found to be higher in the wage sector).

This additional evidence tend to support the idea that sample selection is probably not a very important problem in this application. A statistical sample selection correction procedure does not change the estimates significantly. Using the complete sample by imputing an income to self employed individuals also produces similar estimates in most specifications.

1.7 Comparing costs and benefits

The estimates of the program effect on wages can be used to compare the costs of building and operating the new schools to the wealth they contributed to generate, under the assumption that the increase in wages represents an increase in human capital. Note that in this case, the increase in wages is a lower bound of the total benefit generated by the program: the increase in education is likely to affect other outcomes (fertility, child morbidity and mortality, etc...).[36] These calculations require additional assumptions and should be taken with considerable caution. Nevertheless, it is useful to evaluate the magnitude of the consequences of such a large-scale program: the discounted sum of the cost of construction alone from 1973 to 1979 represented

[35]Individuals who did not work at least an hour in the previous week do not report a branch of activity; they are therefore still excluded from this sample.

[36]Strictly speaking, for it to be a lower bound, I should take into account the costs for kids to go to schools. This is however difficult to implement in practice.

more than 2 percent of the 1973's GDP.

The presidential instructions indicated each year the costs of the construction of the new schools in each region. The total cost of building over 61,000 schools reached a little over 5 billion 1990 US \$. I assume a constant real interest rate of 5% on government debt (this is slightly above the average of real discount rates between 1973 and 1995): 5% of the total capital expenditures must be paid every year. The number of new teachers is also given by the instructions: each school was designed for 120 pupils and three teachers (over 185,000 new teachers were required to operate these schools).

Detailed information on the cost of education in Indonesia has been collected in 1971 by R. Davoesan (Davoesan (1971)). She used a survey of schools she conducted and various administrative sources. A teacher's annual salary was 66,000 Rupiah (about \$ 360) in 1971. She also calculates that in 1971, the total recurrent costs (excluding repairs) were about 1.25 times the wage bill. I will assume that this holds in the entire period. I assume that all new teachers need to be trained (the cost of training a teacher in 1971 was about a third of a teacher's annual salary). Using my data, I estimated the average wage of a primary school teachers in 1995 (\$2,500), and I interpolated linearly the wage between 1971 and 1995 (this represents a annual growth slightly higher than the growth of Indonesia's GDP over the period).

Davoesan considers that a school building can remain in activity for 20 years. This estimate seems reasonable: the INPRES schools had a light structure, and could probably not last more than 20 years. In 1997, most INPRES schools constructed in the mid-1970s were either closed or crumbling. I assume that the schools operate for 20 years and have to be closed thereafter.

In summary, yearly costs are calculated using the following formula:

$$C(t) = r * K + r * TC + W(t) * 1.25,$$

where K is the total capital cost, TC is the total training cost, $W(t)$ represents the sum of teachers' wage at date t, and r is the real interest rate.

Finally, I present the cost-benefits analysis for two different assumptions about the deadweight burden of taxation (0.2 and 0.6).

Further assumptions are needed to compute the yearly benefits of the program. First, an important assumption is that the increase in wages attributed to the program represents an increase in the productivity of labor. Second, I estimated the effect of the program on men who work for a wage. I will assume that the effect is the same on (working) women and on self-employed people. I also assume that the share of total labor income going to people of any given age is constant across years, and is equal to the share of total wages going to this cohort in 1995 (which I can calculate from my data). Thus, I estimate the benefit of the program at date t for a cohort c using the following formula:

$$B(c, t) = \alpha * GDP(t) * S(c, t) * E(c),$$

where α is the share of labor in GDP (I assume that it is 0.7), $S(c, t)$ is the fraction of total wages earned by cohort c in year t, and $E(c)$ is the estimated average effect of the program on cohort c. It is 0 for people 13 or older in 1974 (they were not exposed to the program). For people aged 12 to 2 in 1974, the effect is given by the coefficients in table 4, multiplied by the average intensity of the program. For the following cohorts, I assume that the effect of the program decreases at the rate of population growth.[37] To obtain the total benefits for each year, I sum these benefits over all cohorts.

In figure 11, I show the annual cost and benefit of the program from 1973 to 2060, evaluated in millions US $ (In this picture, the GDP is assumed to grow at 2% after 1996, and the program effect is estimated in column 5 in table 4. I assume a deadweight burden of 0.2). Benefits are very low from 1973 to 1987, because the generations exposed to the program have not yet entered the job market. After 1989, benefits increase rapidly: each year, a new cohort exposed to the program enters the

[37]So for a cohort born y year after 1972, the effect would be calculated as $E(c) = \frac{E(72)}{(1+n)^y}$ if the population growth rate were constant.

job market. After 2015, when the generation exposed to the program start leaving the job market, annual benefits decrease, until the reach 0 by 2050 (when the last cohorts educated in these schools leave the job market). Costs increase rapidly from 1973 to 1979, as more and more schools are built and more teachers need to be paid. Once the total stock is built, costs increase only as teachers' salaries grow. In 1995, when the schools are closed, the costs fall to a very low level corresponding to the annuity payment on initial capital expenditures. In 1996, benefits exceed costs.

For policy purpose, the relevant variable is the discounted sum of net benefits at infinity (defined as the difference between yearly benefits and yearly costs). Figure 12 plots this series (discounted from 1973), as a fraction of 1973 GDP. From 2002 on, the discounted sum of net benefits is positive, and increases rapidly. Program's compounded benefits in 2050 reach 18 percent of 1973 GDP.[38] The program requires more almost 30 years to yield positive returns, but these returns are high.

In table 8, I present an evaluation of the program's returns for the first two specifications estimated in table 4 and three different assumptions about the growth rate of the GDP from 1996 to 2050. I present the discounted sum of program benefits in million 1990 US dollars, as a fraction of 1973 GDP, and as fraction of costs (calculated by dividing the discounted sum of program by the cost of the construction of the new buildings and training of the new teachers). To evaluate the contribution of economic growth to the benefits of the program, I also present these results with the assumption that the Indonesia's GDP grew at a rate of 2% annually from 1973 to 2050.

The costs-benefits analysis is sensitive to the specification chosen for the estimation of the program effect and to the assumptions about future growth rates in Indonesia. Nevertheless, three main points emerge from this analysis. First, a school construction program takes a very long time to generate positive returns (because the costs are incurred early on, while the benefits are spread over a generation's lifetime). Second, the returns generated are large: by 2050, in the smallest estimate, the

[38]After 2050, the benefits are stable, and I consider therefore 2050 as the infinity point.

program will have generated 9 time as much revenues as its initially cost. Third, the important benefits are to a large extent driven by the robust growth of the GDP in Indonesia from 1973 to 1997 (which is the consequences of the fact that each year's benefits are a fraction of this year GDP). If the growth rate had been very low from 1973 until today, the net present value of the program would actually have been slightly negative, according to all specifications but one. Therefore this program is justified ex-post. Investing in education is much more valuable, from a government point of view, if it expects a fast subsequent growth. A related point is made by Bills and Klenow (1998), which argue that the correlation between growth and education in cross-countries growth regression is likely to be driven by the fact that high expected growth make investment in education more profitable.

The last three lines in panel A of table 7 indicate the internal rate of return of the program (the interest rate such that the net present value is 0). Using the actual GDP growth between 1973 and 1997, they range between 8.8% and 12% depending on the specification and hypothesis chosen. These numbers are high, but reasonable. The evidence therefore suggests that the program was a profitable investment, with an internal rate of returns substantially higher than the average interest rate on government debt in Indonesia over the period. The profitability of this investment would have been much less obvious if the Indonesian growth had been slower.

1.8 Conclusion

The INPRES program of primary school construction led to an increase in educational attainment in Indonesia. The estimates of the effect of the program on education of children aged 2 to 6 in 1974 range from 0.12 to 0.18 years for each new school built per 1000 children. In particular, it has encouraged a significant proportion of the population to complete more years of primary education. This increase has translated into an increase in wages, of up to 3.8 percent for each additional school built for 1000 children. Estimates of economic returns to education using this exogenous variation in schooling (assuming that the program had no other effect than to increase the quantity of education), range from 6.4% to 9.1%. These numbers can be interpreted as weighted averages of returns to basic education for people who are affected by the instruments, a group which is likely to include the poorest segment of the population.

The findings reported here are important, because they show that, in Indonesia, a unusually large government administered intervention has been effective in increasing both education and wages. This intervention was meant to increase the *quantity* of education (measured, in the INPRES instructions, by enrollment rates). It is sometimes feared that the deterioration in the quality of education that could result from this type of programs could offset any gain in quantity. However, the program was effective in increasing not only education levels, but also wages. This suggests that the combined effect of quality and quantity changes in education was to increase human capital. I presented some evidence that the quality of education does not seem to have deteriorated significantly because of the program. But even if it has, the effect on wages shows that this decline was not sufficient to offset the impact of the increase in quantity.

I presented evidence in favor of the internal validity of these results: I have shown that changes between cohorts were not systematically different in low and high program regions before the program started, and I have tried to control for the two variables (the water and sanitation program and the enrollment rate in 1971) whose omission was most likely to bias the estimates. It remains possible that these results cannot be generalized to different situations. In particular, the emphasis given on education by the Indonesian government at the time of the program created a context particularly favorable to its success.

In future work, I plan to use the instruments generated by this INPRES program to evaluate the impact of education on fertility and child survival. The outlined methodology can also be used in other settings, characterized by intense and highly localized policy innovations. The shape of the returns to education function deserves further exploration as well. The results in this study tend to suggest that there is no evidence that returns are concave. Combining several reforms affecting education at various levels to try to recover more information about the shape of the function linking education and earnings would potentially shed light on a very important dimension of the human capital accumulation problem.

Bibliography

[1] Ahn, H. and J.L. Powell (1993) "Semi-parametric estimation of censored selection model with a nonparametric selection mechanism" *Journal of econometrics* 58:3-29

[2] Angrist, J. (1995) *Conditioning on the probability of selection to control selection bias* NBER technical working paper, n. 181

[3] Angrist, J.(1995) "The Economic Returns to Schooling in the West Bank and Gaza Strip" *American Economic review*, 85.5:1065-1087.

[4] Angrist, J. and W. Evans (1996) *Schooling and Labor Market consequences of the 1970 state abortion reforms* NBER working paper 5406.

[5] Angrist, J. and Guido Imbens(1995) "Two-stage least squares estimation of Average causal effects in Models with variable treatment intensity" *Journal of the American Statistical Association* 90.430:431-442.

[6] Becker, Gary (1967) *Human Capital and the Personal Distribution of Income* University of Michigan Press, Ann Arbor, Michigan.

[7] Bappenas (1973, 1974, 1975, 1976, 1977, 978) *Pedoman Pelaksanaan Bantuan Pembagunan Sekolah Dasar* Menurut Instruki Presiden Republik Indonesia.

[8] Behrman J.R. and Birdsall (1983) "Returns to Education, quantity alone is misleading".

[9] Behrman J.R. and Deolalikar (1991) "School Repetition, Dropout and the Returns to Schooling: the Case of Indonesia" *Oxford Bulletin of Economics and Statistics* 53.4 467-480.

[10] Behrman J.R. and Deolalikar (1993) "Unobserved household and community heterogeneity and the labor market impact of schcoling: A case study of Indonesia" *Economic Development and Cultural Change* 41:3461-488.

[11] Birdsall N. (1985) "Public inputs and child schooling in Brazil" *Journal of Development Economics* 18.1:67-86.

[12] Biro Pusat Statistic (1974) *Indonesia Sensus Penduduk, 1971, Seri S, tabel 4*

[13] Butcher K. and A. Case (1994) "The effects of sibling sex composition on women's education and earnings" *Quarterly Journal of Economics* 61.3:531-563.

[14] Card David (1993) *Using Geographic variation in college proximity to estimate the returns to schooling* NBER working paper 4483.

[15] Card David (1995) *Earnings, Schooling, and ability revisited Research in Labor Economics* col 14. Eds S. Polachek, Greewich, Coonn. JAI Press.

[16] Card David (2000) *The causal effect of education on earnings* Orley Ashenfelter and David Card (ed.s)*Handbook of labor Economics*, forthcoming.

[17] Card D. and A. Krueger (1992) "Does school quality matter? Returns to education an the characteristics of public schools in the United States" *Journal of Political Economy* 100.1:1-40.

[18] Card D, and T. Lemieux (1998) *Earnings, education and the Canadian GI Bill*, MIMEO.

[19] Case, Anne and Angus Deaton (1996) "School quality and educational outcomes in South Africa" MIMEO, Research program in Development Studies, Princeton University.

[20] Davoesan R. 1971 "Finance of education" part 1 and 2 *Bulletin of Indonesian economic studies*, 7.3:61-95.

[21] Deaton, Angus (1997) *The analysis of Household surveys, A Microeconometric approach to development policy* John Hopkins, Baltimore.

[22] Departement Dalam Negeri (1983) *Sepuluh Tahun INPRES Sekolah Dasar 1973/74-1983/84*

[23] Departemen Pendidikan dan kebudayaan, pusat informatika, (1996) *Perkembangan Jumlah Sekolah, mirid and guru tiap kabupaten/kotamadya SD, SLTP dan SM Tahum 1973/74-1995/96*

[24] DeTray D. (1988) "Government Policy, household behavior and the distribution of schooling: A case study of Malaysia" in T.P. Schultz (ed.) *Research in Population Economics*, vol.6 Greenwich CT: JAI Press.

[25] Foster, Andrew and Mark Rosenzweig (1995) "Learning by Doing and Learning from Others: Human Capital and Tecnical change in Agriculture" *Journal of Political Economy* 103.6.

[26] Foster, Andrew and Mark Rosenzweig (1996) "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution". *American Economic Review*,86.4:931-953.

[27] Garg, Ashish and Jonathan Morduch (1998)"Sibling rivalry" MIMEO, Harvard University.

[28] Gertler, P. and J. Molyneaux (1994) "How economic development and family planning programs combined to reduce Indonesian fertility" *Demography* 31-1:33-63.

[29] Glewwe P. and Hanan Jacoby: "An econometric Analysis of Delayed Primary School enrollment and Childhood Malnutrition in a Low Income Country" *Review of Economics and Statistics* 77:1 156-169.

[30] Heckman J. (1996) "Comments" to Nada Eissa in *Empirical Foundation of Household Taxation* eds. M. Feldstein and J. Poterba, University of Chicago.

[31] Heckman J. and V. Joseph Hotz (1989) "Choosing among alternative non experimental methods for estimating the impact of social programs: the case of manpower training" *Journal of the American Statistical Association*, 84.408:862-874

[32] Heckman J. Lance Lochner and Christopher Taber "General Equilibrium Treatment Effects: A Study of Tuition Policy" Wp n. 6426.

[33] Heckman J. and Richard Robb (1986) "Alternative methods for solving the problem of selection bias in evaluating the impact of treatment on outcomes" in Wainer (ed.) *Drawing Inferences from Self-selected Samples*

[34] Heneveld W (1978) "The distribution of development funds: new school building in East Java" *Bulletin of Indonesian Economic Studies*

[35] Hill H. (1996) *The Indonesian Economy since 1966: Southeast Asia's Emerging Giant* Cambridge U.P.

[36] Grilliches, Zvi (1977) "Estimating the returns to Schooling: Some Econometric Problems" *Econometrica* 45:1-22.

[37] Imbens, Guido and Joshua Angrist (1994) "Identification and estimation of local average treatment effects" *Econometrica* 62.2:467-475.

[38] Kane T, and C. Rouse (1995) "Labor market returns to two- and four-year college" *American Economic Review*85.3:600-614.

[39] Lang K. (1993) "Ability bias, discount rate bias, and the returns to Education" Unpublished discussion paper , Boston University, Department of Economics.

[40] Lang K, and D. Kropp (1986) "Human capital versus sorting: the effects of compulsory attendance laws" *Quarterly Journal of Economics* August 1986:609-624.

[41] Lavy, Victor (1996) "School Supply constraints and children's educational outcomes in Rural Ghana" *Journal of development Economics* 51.2291:314.

[42] Lillard L. and Willis R.(1994) "Intergenerational educational mobility: Effects of family and State in Malaysia" *Journal of Human Resources* 29.4:1126-1166.

[43] Maluccio, John (1998) *Endogeneity of schooling in the wage function: Evidence from rural Philippines* MIMEO , International Food Policy Research Institute.

[44] Pitt M, M. Rosenzweig and D. Gibbons (1993) "The determinants and consequences of the placement of government programs in Indonesia" *World Bank Economic Review* 7:319-348.

[45] Psacharopoulos G. (1973) *Returns to education: an international comparison* San Franciso: Jossy Bass-Elsevier

[46] Psacharopoulos G. (1981) "Returns to education: an updated international comparison" *Comparative Education* 17:321-341.

[47] Psacharopoulos G. (1985) "Returns to education: a further international update and implications" *Journal of Human Resources* 20.4:583-604.

[48] Psacharopoulos G. (1994) "Returns to investments in education: a global update" *World Development* 22.9:1325-1343.

[49] Ravallion Martin (1988) "INPRES and inequality: a distributional perspective on the center's regional disbursements" *Bulletin of Indonesian Economic Studies* 24.3:53-70.

[50] Rosenbaum (1987) "The role of a second control group in a observational study" *Statistical Science* 2:292-305.

[51] Rosenzweig Mark and Kenneth Wolpin (1988) "Evaluating the effects of optimally distributed programs: Child health and family planning interventions" *American Economic Review* 76.3:470-482.

[52] Rosenzweig Mark and Kenneth Wolpin (1988) "Migration selectivity and the effects of public programs" *Journal of Public Economics*37:265-289.

[53] Strauss, John and Duncan Thomas (1995) "Human resources: empirical modeling of household and family decisions" chapter 33, pp. 1883-2023 in *Handbook of Development economics*, edited by J. Behrman and T.N Srinivasan, Elsevier Science.

[54] Strauss , J. and Duncan Thomas(1997) "Wages, schooling and background: Investment in men and women in urban Brazil" in N. Birdsall, Burns and R. Sabot *Opportunity foregone: Education, growth and Inequality in Brazil* World Bank, 1997.

[55] World Bank (1989) *Indonesia: Basic education study* World Bank report n. 7841-IND.

[56] World Bank (1990) *Indonesia: Poverty assessment and Strategy report.*

[57] World Bank (1990) *Indonesia: Strategy for a sustained reduction in poverty* The World Bank, Washington DC.

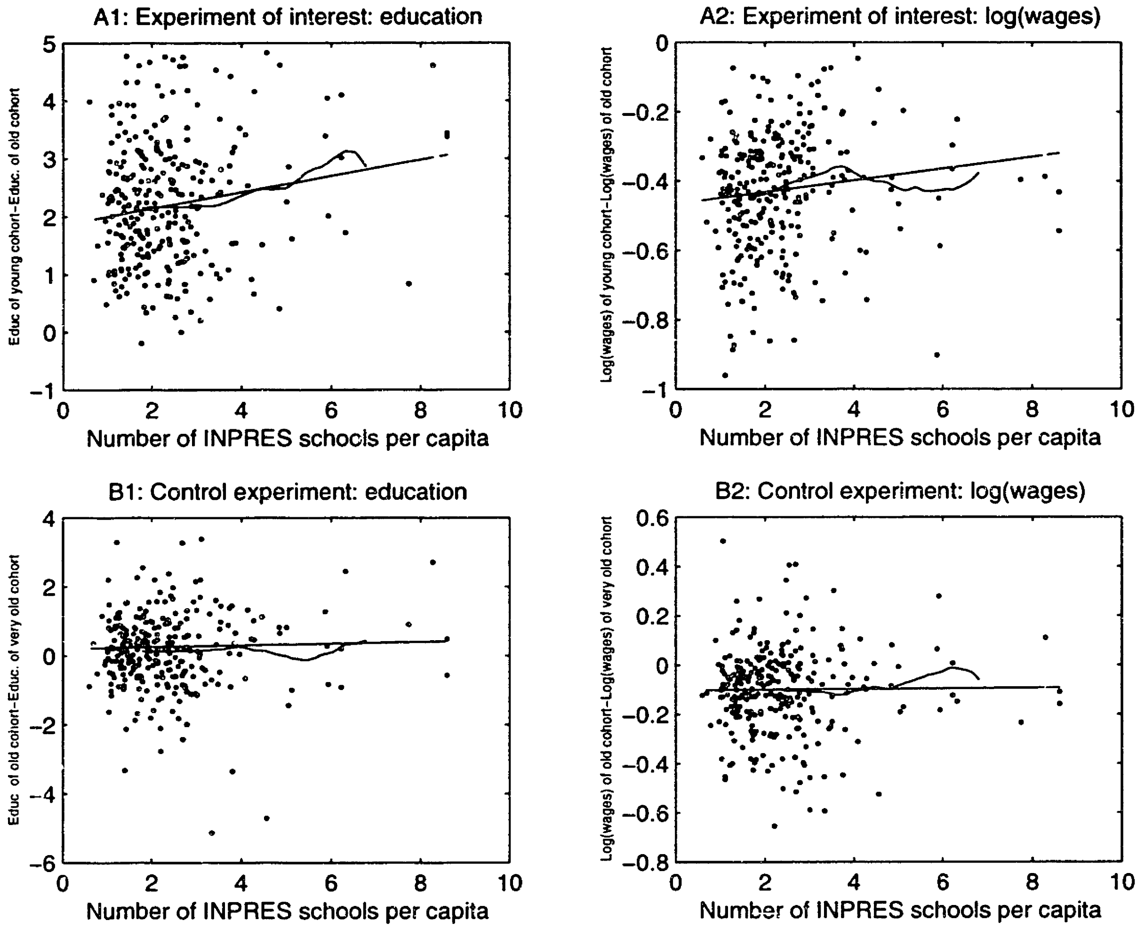


FIGURE 1: Regional Growth in education and log wages accross cohort and program intensity

Figure 2: Coefficient of the interaction region of birth Cohort of birth in the education equation

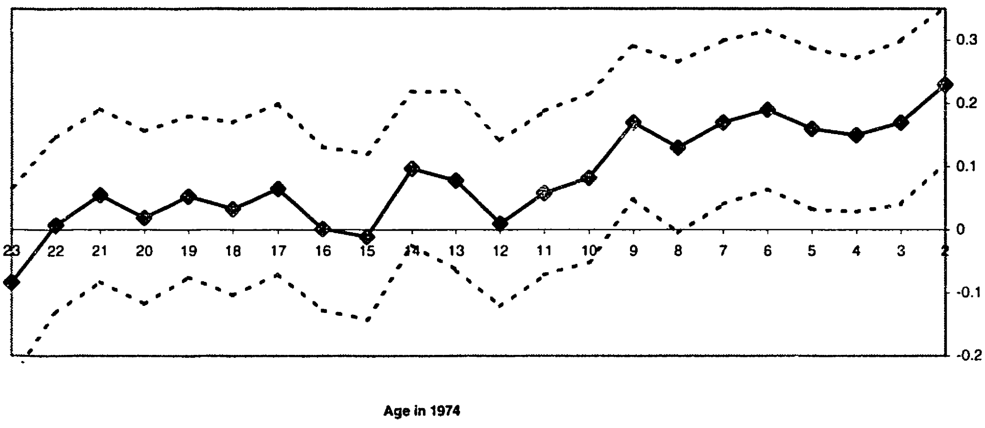


Figure 3a: Coefficients of the interaction age in 1974* program intensity in the region of birth in the wage and education equations

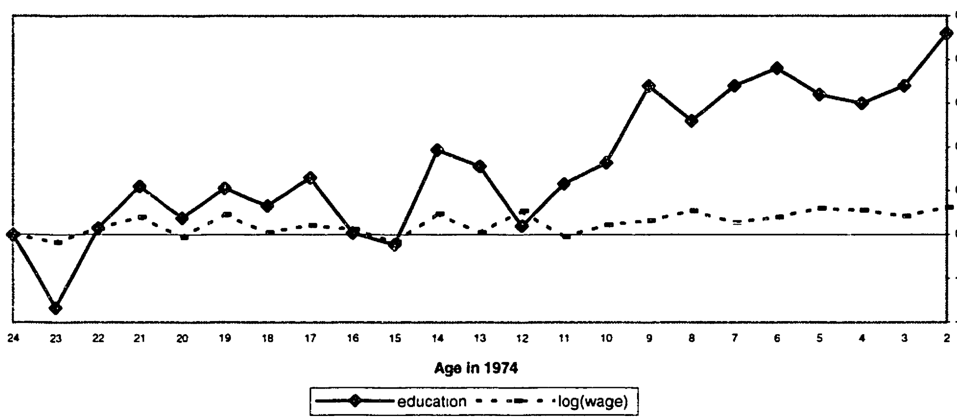


Figure 3b: Coefficients of the interaction age in 1974* program intensity in the region of birth in the wage and education equations (smoothed)

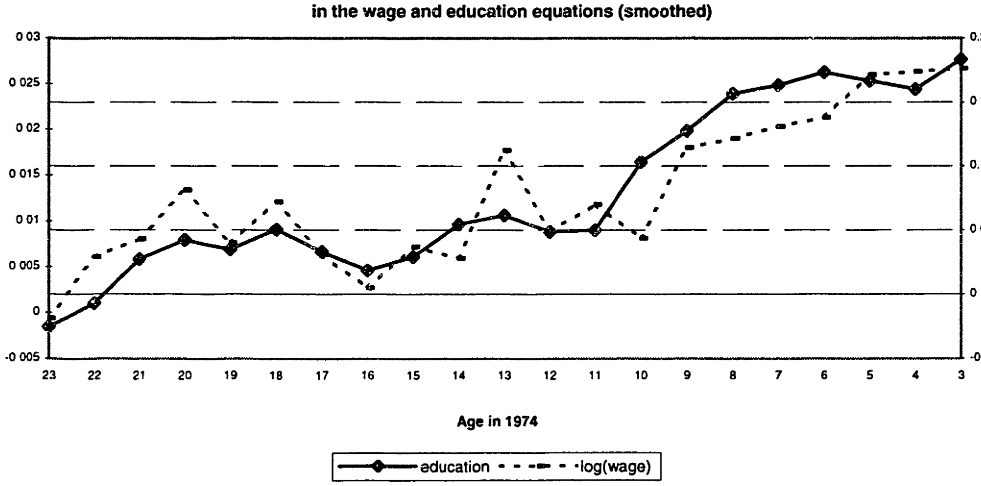


Figure 4: Cdf of education, high program region

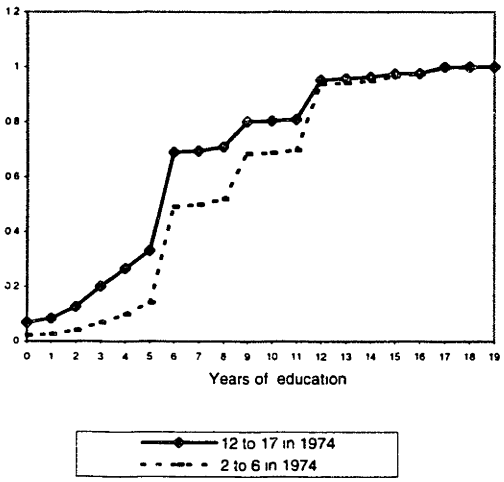


Figure 5: Cdf of education, low program region

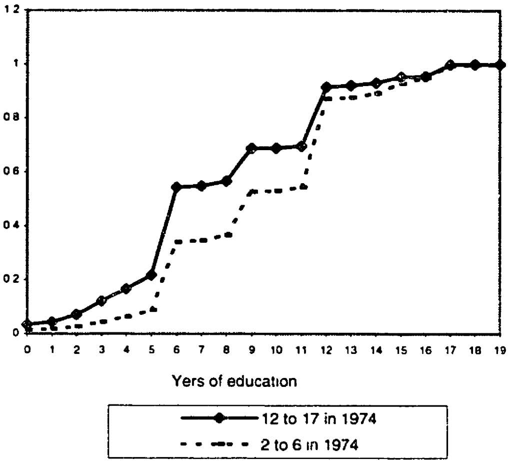


Figure 6: Between cohort differences in cdf of education.

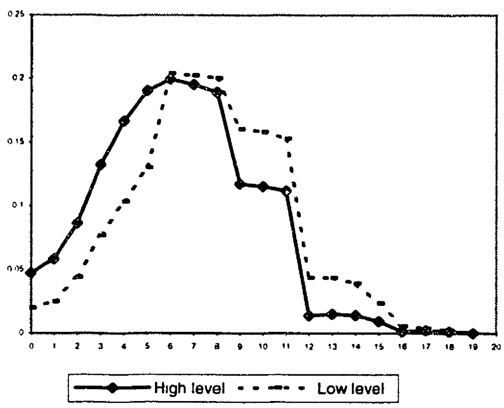


Figure 7: Difference in differences in CDF (with 95% confidence interval)

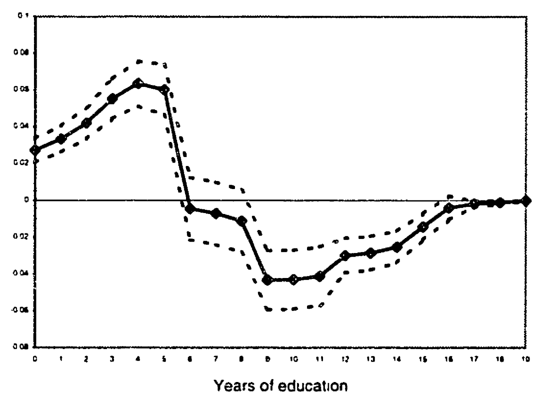


Figure 8: Difference in difference in cdf (estimated from linear probability model) with 95% confidence interval

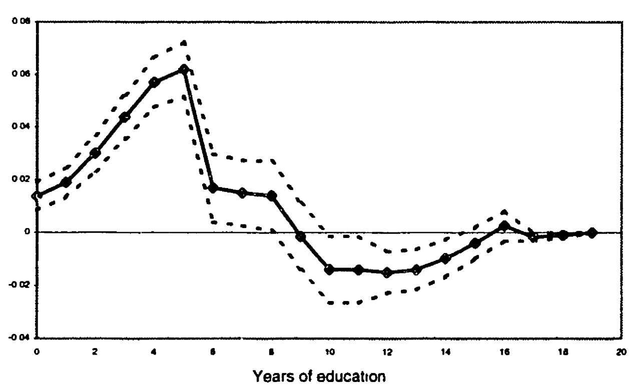


Figure 9:Coefficients of the interactions age in 1974* program intensity in the region of birth in the wage and education equation (sample: individuals who completed more than 9 years of education)

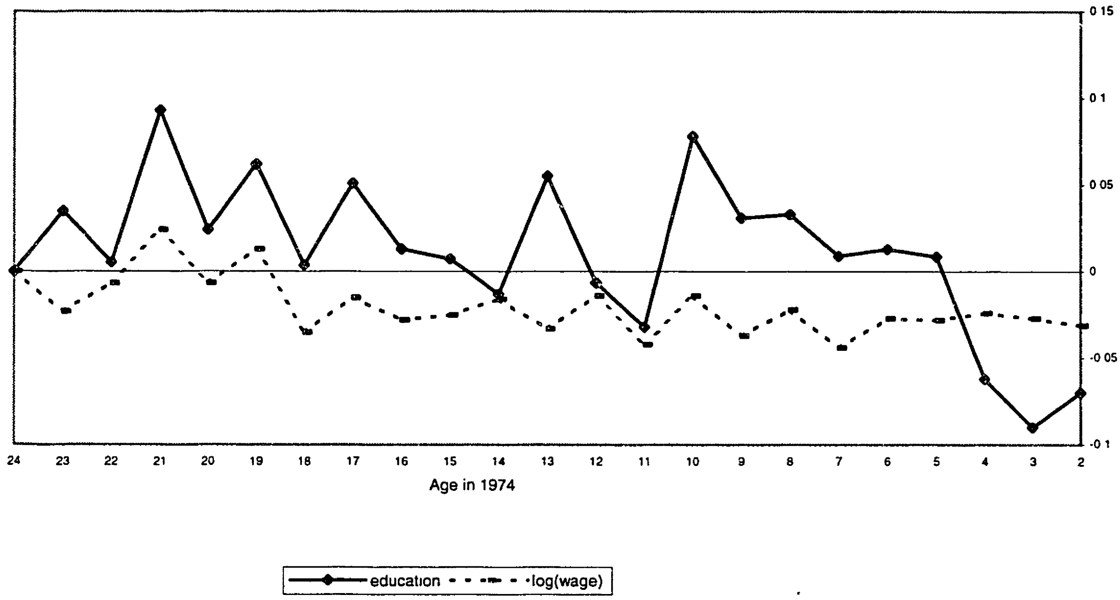


Figure 10: Returns to each year of education (OLS estimate)

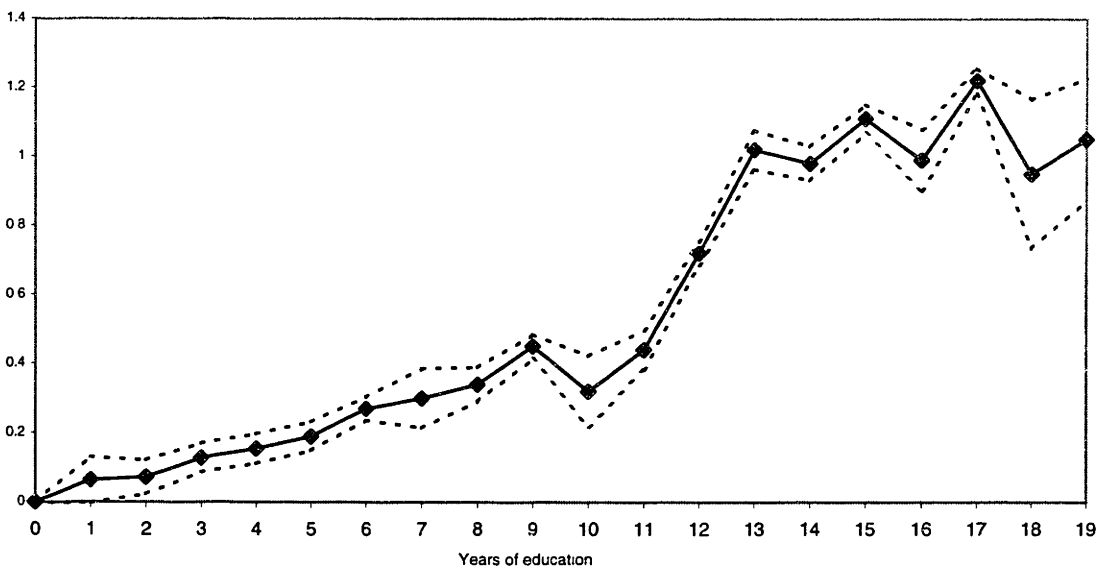


Figure 11: Annual cost and benefit of the program (Million 1990 US dollars)

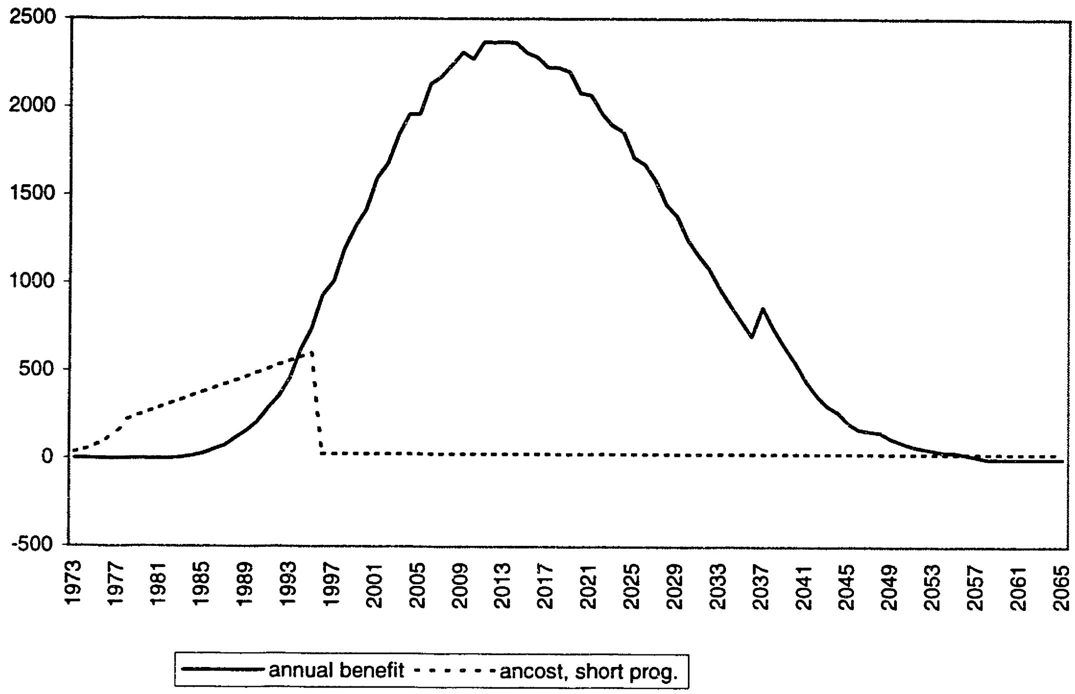
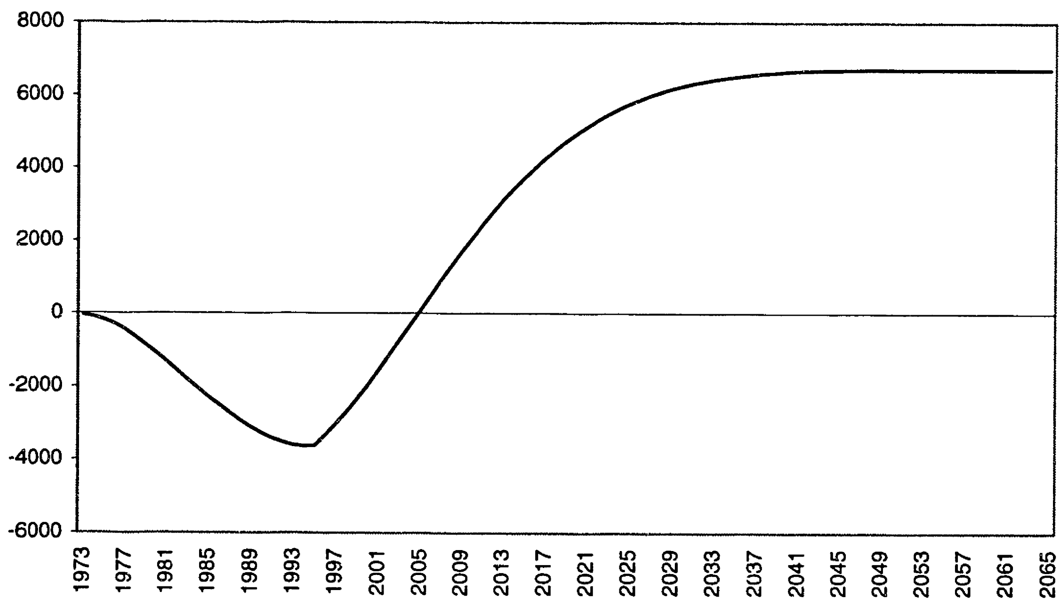


Figure 12: Discounted sum of program's net benefit



Note: Both figures assume that Indonesian GDP will grow at 2 % a year after 1998 and a DWB of 0.2
Underlying estimates are from table 4, column 6.

Table 1: Descriptive statistics

	Mean
A. Individual level means	
Education (whole sample N=153,061)	7.8
Education (sample with valid wage data N=60,714)	9.01
INPRES schools built for 1000 children	1.98
INPRES schools built for 1000 children (sample with valid wage data)	1.89
Log(hourly wage)	6.87
Monthly earnings (SUPAS 1995),thousands Rupiah	227.9
Monthly earnings (SUSENAS 1993) of wage earners, thousands Rupiah	223.8
Monthly earnings (SUSENAS 1993) of self employed individuals, thousands Rupiah	145.7
B. District level means (N=281)	
INPRES schools constructed (1973/74-1978/79)	220
INPRES schools constructed for 1000 children (1973/74-1978/79)	2.09
INPRES schools constructed (1973/74-1977/78)	167
Fraction of the population attending schools in 1971 (census)	0.174
Number of teachers in 1973/74	1530
Number of teachers in 1978/79	2082
Number of schools in 1973/74	219
C. Indonesian Family Life Survey, individuals born between 1950 and 1972	
Proportion of individuals having migrated between birth and age 12	8.5%
Proportion of people having repeated at least one grade in primary school	20.0%
Proportion of people completin more than primary having repeated at least one grade in primary school	6.0%
Proportion of individuals having attended primary school after age 12 (estimated)	15.8%
Proportion of individuals having attended primary school after age 13 (estimated)	6.8%
Proportion of individuals born 1950-1961 , completing primary or less, who left school after 1974	2.8%
Proportion of individuals born 1962-1966 , completing primary or less, who left school after 1974	24.5%

Sources: IFLS, SUPAS, INPRES instruction, census data, Department of education and Culture

Table 2: The allocation of schools.
Dependent variable :log of number of program shools built in each region

Log of number of children aged 5-14 in the region	0.69
	(0.030)
Log of fraction of population attending school in 1971	-0.46
	(0.11)

Notes:The R squared of the regression is 0.75. 281 observations are used. Average education of the individuals born 1950-1960 is used as instrument for log (attendance)

Table 3: Means of Education and log(wage) by cohort and level of program cells,
Differences in differences
and Wald estimate of returns to education.

PANEL A: Changes in number of schools per 1000 children and teacher pupils ratio

	Number of schools per 1000 children ages 5-14 in 1971				Pupil-teacher ratio			
		Level of program in Region of birth				Level of program in Region of birth		
		High	Low	Difference		High	Low	Difference
	1973/74	4.22	5.21	-0.99	1973/74	29.67	28.43	1.24
year	(n. teachers/3)					(6.11)	(6.27)	(8.75)
	1978/79	6.75	6.72	0.03	1978/79	31.51	30.14	1.37
	(imputed)					(6.72)	(7.05)	(9.74)
	Difference	2.53	1.51	1.02	Difference	1.84	1.71	0.13
	(INPRES program)					(6.38)	(6.50)	(8.76)

PANEL B: Experiment of interest

	Education				Log(wage)			
		Level of program in Region of birth				Level of program in Region of birth		
		High	Low	Difference		High	Low	Difference
	2 to 6	8.48	9.76	-1.29	2 to 6	6.60	6.73	-0.13
age in		(0.044)	(0.036)	(0.057)		(0.0078)	(0.0064)	(0.010)
1974	12 to 17	8.00	9.41	-1.41	12 to 17	6.87	7.02	-0.15
		(0.054)	(0.042)	(0.067)		(0.0085)	(0.0069)	(0.011)
	Difference	0.48	0.36	0.12	Difference	-0.26	-0.29	0.021
		(0.070)	(0.038)	(0.089)		(0.011)	(0.0096)	(0.015)
	Differences in differences (controll			0.11	Differences in differences (controlli			0.027
	for region and year of birth dummi			(0.052)	for region and year of birth)			(0.012)

Wald estimate of returns to education: 0.021/0.12=0.17

PANEL C: Control experiment

	Education				Log(wage)			
		Level of program in Region of birth				Level of program in Region of birth		
		High	Low	Difference		High	Low	Difference
	12 to 17	8.00	9.41	-1.41	12 to 17	6.87	7.02	-0.15
age in		(0.054)	(0.042)	(0.078)		(0.0085)	(0.0069)	(0.011)
1974	18 to 24	7.70	9.12	-1.42	12 to 17	6.92	7.08	-0.16
		(0.059)	(0.044)	(0.072)		(0.0097)	(0.0076)	(0.012)
	Difference	0.30	0.29	0.013	Difference	0.056	0.063	0.0070
		(0.080)	(0.061)	(0.098)		(0.013)	(0.010)	(0.016)
	Differences in differences (controll			0.037	Differences in differences (controlli			0.004
	for region and year of birth)			(0.083)	for region and year of birth)			(0.014)

Note: In panel A, the differences in schol per capita correspond to the level of the INPRES program in each t
region. For consistency with INPRES, the number of schools in 1973 is calculated as number of teachers in 1
The number of schools per child in 1978/1979 is imputed as the sum of the number of schools in 1973/1974
and the number of schools per child built under the program.

Table 4 : Effect of the program on education and wages (continued):
Coefficients of the interation between number of school constructed per child in the region of birth and cohort dummies

	Dependent variable							
	Education				Log(hourly wage)			
	N. obs	(1)	(2)	(3)	N.obs	(4)	(5)	(6)
PANEL A: Experiment of interest: individuals ages 2 to 6 or 12 to 17 in 1974.								
(Youngest cohort: individuals ages 2 to 6 in 1974)								
Whole sample	76,513	0.122	0.148	0.182				
		(0.0252)	(0.0260)	(0.0321)				
Sample of wage earners	30,065	0.188	0.198	0.223	30,065	0.0145	0.0164	0.0217
		(0.0400)	(0.0403)	(0.0500)		(0.00729)	(0.00421)	(0.00913)
PANEL B: Control experiment : individuals ages 12 to 24 in 1974								
(Youngest cohort: individuals ages 12 to 17 in 1974)								
Whole sample	76,496	0.0176	0.0238	0.0137				
		(0.0262)	(0.0271)	(0.0329)				
Sample of wage earners	30,720	0.009	0.017	0.0117	30,720	0.00385	0.00421	0.00931
		(0.0423)	(0.0429)	(0.0524)		(0.00791)	(0.00809)	(0.00994)
Control variables:								
Y.o.b.*enrollment rate in 1971		No	Yes	Yes		No	Yes	Yes
Y.o.b* water and sanitation program		No	No	Yes		No	No	Yes

Notes: All specifications include region of birth, year of birth dummies and interactions between the treatment and the number of children in the region of birth (in 1971). The numbers of observations refer to the specification in col (1) and (5), respectively.
Standard errors in parenthesis.

Table 5 :Effect of the program on education and wages:
Coefficients of the interaction between age in 1974 and number of schools per 1000 children constructed in region of birth

	Dependent variable: years of education						Dependent variable: log(hourly wage)		
	Whole sample			Sample of wage earners					
age in 1974	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
12	-0.032	-0.021	-0.026	-0.030	-0.004	-0.016	0.0169	0.0195	0.0214
	(0.047)	(0.048)	(0.048)	(0.077)	(0.078)	(0.098)	(0.0132)	(0.0133)	(0.0167)
11	0.014	0.027	0.002	0.013	0.016	-0.016	-0.0107	-0.0104	-0.0064
	(0.046)	(0.047)	(0.046)	(0.073)	(0.073)	(0.089)	(0.0125)	(0.0126)	(0.0152)
10	0.057	0.051	0.051	0.099	0.090	0.127	0.0026	0.0034	-0.0011
	(0.048)	(0.049)	(0.048)	(0.075)	(0.076)	(0.096)	(0.0128)	(0.0130)	(0.0163)
9	0.137	0.139	0.122	0.066	0.063	0.165	0.0081	0.0085	0.0208
	(0.040)	(0.041)	(0.040)	(0.065)	(0.066)	(0.083)	(0.0111)	(0.0112)	(0.0140)
8	0.080	0.100	0.082	0.178	0.191	0.257	0.0186	0.0202	0.0212
	(0.050)	(0.050)	(0.050)	(0.078)	(0.079)	(0.099)	(0.0134)	(0.0134)	(0.0167)
7	0.109	0.135	0.097	0.113	0.134	0.182	-0.0109	-0.0069	-0.0015
	(0.045)	(0.046)	(0.045)	(0.072)	(0.073)	(0.090)	(0.0123)	(0.0124)	(0.0153)
6	0.128	0.163	0.196	0.226	0.230	0.330	0.0107	0.0127	0.0153
	(0.043)	(0.044)	(0.043)	(0.070)	(0.070)	(0.088)	(0.0119)	(0.0120)	(0.0149)
5	0.100	0.131	0.105	0.129	0.148	0.255	0.0202	0.0218	0.0506
	(0.044)	(0.045)	(0.044)	(0.075)	(0.075)	(0.095)	(0.0127)	(0.0128)	(0.0161)
4	0.111	0.123	0.189	0.192	0.190	0.182	0.0185	0.0200	0.0345
	(0.040)	(0.041)	(0.039)	(0.069)	(0.069)	(0.088)	(0.0117)	(0.0118)	(0.0149)
3	0.115	0.146	0.202	0.149	0.169	0.331	0.0090	0.0135	0.0268
	(0.045)	(0.046)	(0.045)	(0.079)	(0.080)	(0.102)	(0.0134)	(0.0136)	(0.0174)
2	0.158	0.200	0.231	0.206	0.220	0.297	0.0161	0.0232	0.0369
	(0.041)	(0.043)	(0.041)	(0.073)	(0.074)	(0.093)	(0.0125)	(0.0127)	(0.0158)
Control variables:									
Y.o.b.*enr. rate in 1971	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Y.o.b* water and sanitation program	No	No	Yes	No	No	Yes	No	No	Yes
F.stat.	3.92	5.16	4.57	2.61	2.74	4.03	1.06	1.24	2.05
R squared	0.19	0.19	0.16	0.14	0.14	0.13	0.14	0.14	0.13
N. obs.	149,068	148,573	139,253	59,938	59,741	54,442	59,938	59,741	54,442

Notes: All specifications include region of birth and year of birth dummies.
Standard errors in parenthesis. The F-stat tests the hypothesis that the interaction coefficients are jointly zero.
The control group is the group of people who were 13 to 24 in 1974.

Table 6: Estimates of the coefficient of education in labor market outcomes equations.

Method	Instrument	(1)	(2)	(3)	(4)
PANEL A: SAMPLE OF INDIVIDUALS WHO WORK FOR A WAGE					
PANEL A1 Dependent variable: log(hourly wage)					
OLS		0.0776	0.0776	0.0776	
		(0.0006)	(0.0006)	(0.0006)	
2sls	Y.o.b. Dummies* program intensity in region of birth	0.0640	0.0791	0.0911	0.0890
		(0.0245)	(0.0276)	(0.023)	(0.0420)
		[0.96]	[0.9]	[0.93]	[0.9]
2sls	(Aged 2-6 in 1974)*program intensity in region of birth	0.0770	0.0821	0.0942	
		(0.0340)	(0.0336)	(0.0286)	
PANEL A2 Dependent variable: log(monthly earnings)					
OLS		0.0697	0.0697	0.0767	
		(0.0006)	(0.0006)	(0.0006)	
2sls	Y.o.b. Dummies* program intensity in region of birth	0.0721	0.0902	0.0763	0.131
		(0.0284)	(0.0281)	(0.0224)	(0.0475)
		[0.73]	[0.63]	[0.58]	[0.7]
PANEL B: COMPLETE SAMPLE					
PANEL B1 Dependent variable: participation in the wage sector					
OLS		0.0328	0.0327	0.0338	
		(0.003)	(0.0003)	(0.003)	
2sls	Y.o.b. Dummies* program intensity in region of birth	0.104	0.120	0.0760	
		(0.0216)	(0.0200)	(0.0182)	
		[0.66]	[0.93]	[1.12]	
PANEL B2 Dependent variable: log(monthly earnings), imputed for self-employed individuals					
OLS		0.0566	0.0566	0.0566	
		(0.00034)	(0.00034)	(0.00035)	
2sls	Y.o.b. Dummies* program intensity in region of birth	0.0635	0.0848	0.0489	
		(0.0167)	(0.0141)	(0.0131)	
		[0.68]	[0.58]	[1.16]	
Control variables:					
Y.o.b.*enr. rate in 1971		No	Yes	Yes	Yes
Y.o.b* water and sanitation program		No	No	Yes	No
Propensity score, propensity score squared		No	No	No	Yes

Notes: Year of birth dummies, region of birth dummies and the interactions betweeen year of birth dummies and number of children in the region of birth are included in the regressions.
The instruments are a set of year of birth dummies (for people aged 2 to 12 in 1974) interacted with the number of schools built in the region of birth.
Standard errors are in parenthesis. F statistic of the test of overidentification restriction in squared brackets.

Table 7:
Program effect and returns to education by categories of region of birth

	Whole sample	Density in R.O.B.		1976 Poverty in R.O.B		Pre-program education in R.O.B	
		<median	>median	High	Low	<median	>median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A:Effect of the program on education. Dependent variable; Years of education.Sample: individuals ages 2 to 6							
Regressors:							
Interaction	0.15	0.22	-0.017	0.13	0.08	0.13	0.12
(2-6 in 1974)*prog. Intensity in region of birth	(0.0260)	(0.032)	(0.048)	(0.056)	(0.035)	(0.040)	(0.036)
PANEL B: Effect of the program on wages. Dependent variable: log(hourly wage). Sample: individuals ages 2 to 6 or 12							
Interaction	0.016	0.029	-0.0015	0.049	-0.0014	0.016	0.0069
(2-6 in 1974)*prog. Intensity in region of birth	(0.0042)	(0.0097)	(0.012)	(0.0164)	(0.0093)	(0.0122)	(0.0095)
PANEL C: Returns to education. Dependent variable: log (hourly wage). Complete sample of wage earners							
Years of education	0.0791	0.0998	No	0.0734	No	0.107	No
	(0.0276)	(0.0241)	First stage	(0.032)	First stage	(0.0281)	First stage
	[0.9]	[0.86]		[0.88]		[0.72]	

Notes: R.O.B stands for region of birth. Region of birth dummies, year of birth dummies, and interaction of treatment dummies with number of children and enrollme
the region in 1971 are included in the regressions.

The median density (I.e. the density for the region of birth for the median person in the weighted sample) is 308 habitants per square kilometer.

The high poverty provinces are the provinces where the proportion of people consuming less than 1,500 Rp per capita is larger than the national average (11.3%) for r
(in the 1976 SUSENAS). I define as "high poverty" districts the rural districts in these provinces, which are: Lampung, Central Java, Jogyakarta, East Java, East Nus
Central Sulawesi, South Sulawesi, Southeast Sulawesi, Maluku, Irian Jaya. Source: World Bank (1979)

The pre-program education is the average education in the region of birth for people born in 1962 or before. The median is 3.18 years.

Standard errors in parenthesis, F. statistic of the overidentificaiton restriction test in squared brackets.

Table 8:
Evaluation of the program's net return

PANEL A: Results

	Deadweight loss coefficient			
	0.2		0.6	
	(1)	(2)	(3)	(4)
Control for Yob*enrollment rate	No	Yes	No	Yes
First year where benefit>costs (discount rate=5%)				
In annual value:	1996	1996	1997	1997
In discounted sum	2005	2002	2009	2005
Discounted sum of net benefits in 2050 (g after 1997=5%, discount rate 5%)				
In million 1990 US$	13,025	13,096	11,340	18,807
As a fraction of Indonesia's GDP in 1973	0.3	0.36	0.31	0.52
Divided by initial costs	24.1	24.2	21	35
Discounted sum of net benefits in 2050 (g after 1997=2%, discount rate 5%)				
In milllion 1990 US$	6,691	11,589	5,008	9,905
As a fraction of Indonesia's GDP in 1973	0.18	0.32	0.14	0.27
Divided by initial costs	12.4	21.4	9.26	18.3
Discounted sum of net benefits in 2050 (g from 1973= 2%, discount rate 5 %)				
In million 1990 US$	-631.6	1200	-2315	-483
As a fraction of Indonesia's GDP in 1973	-0.017	0.033	-0.063	-0.013
Divided by initial costs	-1.16	2.22	-4.28	-0.89
Internal rate of return				
g after 1997=5%	0.102	0.118	0.0895	0.105
g after 1997=2%	0.088	0.106	0.0750	0.0915
g from 1973=2%	0.0443	0.059	0.0326	0.0467

PANEL B: Assumptions and parameters

Population growth rate after 1997	0.015
Yearly teacher salary in 1973 (1990 US $)	363
Yearly teachers salary in 1995 (1990 US $)	2,467
Total recurrent costs/teacher salary	1.25
Total cost of construction (million 1990 US $)	522
Number of school constructed	61,800
Life time of the schools (years)	20
Share of labor income in GDP	0.7

Note: The estimates underlying these calculations are taken from tabl4 4 (col. 5 and 6)

Program effect has been set to 0 for children aged 7 or older in 1974

The internal rate of return is the interest rate such that the net present value of the project at infinity is 0.

Table A1 : Unrestricted estimates of program effect:
Coefficients of the interaction between age in 1974 and number of school per child built in the region of birth.

	education		log(wage)		education		log(wage)		education		log(wage)	
	(1)		(2)		(3)		(4)		(5)		(6)	
age in 1974	coeff	S.E.	coeff	S.E.	coeff	S.E.	coeff	S.E.	coeff	S.E.	coeff	S.E.
23	-0.073	0.074	-0.0028	0.023	-0.0835	0.077	-0.0088	0.022	-0.19	0.096	-0.0023	0.029
22	0.023	0.068	0.0062	0.022	0.0075	0.071	0.007	0.021	-0.032	0.088	0.031	0.027
21	0.06	0.066	0.017	0.021	0.055	0.07	0.02	0.02	-0.063	0.085	0.026	0.026
20	0.045	0.066	0.011	0.021	0.019	0.07	-0.0029	0.02	-0.089	0.085	-0.13	0.026
19	0.052	0.063	0.019	0.021	0.053	0.065	0.023	0.019	-0.018	0.08	0.039	0.026
18	0.013	0.061	-0.0035	0.021	0.033	0.07	0.0026	0.02	-0.13	0.086	0.014	0.026
17	0.081	0.067	0.0042	0.020	0.065	0.069	0.0106	0.019	-0.053	0.084	0.018	0.026
16	-0.007	0.065	0.0026	0.020	0.0017	0.066	0.0059	0.019	-0.11	0.082	0.013	0.025
15	-0.008	0.063	-0.0103	0.019	-0.011	0.067	-0.0084	0.018	-0.11	0.083	-0.0042	0.024
14	0.094	0.059	0.0196	0.019	0.097	0.062	0.024	0.018	0.046	0.077	0.053	0.027
13	0.073	0.069	-0.003	0.022	0.078	0.072	0.0022	0.02	-0.021	0.088	0.015	0.025
12	0.0027	0.064	0.021	0.020	0.01	0.067	0.027	0.029	-0.083	0.083	0.040	0.024
11	0.048	0.063	-0.0069	0.019	0.058	0.066	-0.0026	0.018	-0.054	0.081	0.012	0.025
10	0.093	0.064	0.0066	0.020	0.082	0.068	0.011	0.018	-0.0063	0.084	0.018	0.024
9	0.17	0.058	0.012	0.019	0.17	0.062	0.016	0.018	0.064	0.076	0.040	0.024
8	0.11	0.066	0.022	0.020	0.13	0.069	0.027	0.017	0.024	0.084	0.040	0.025
7	0.14	0.062	0.0072	0.019	0.17	0.066	0.014	0.019	0.04	0.081	0.018	0.024
6	0.16	0.060	0.0145	0.020	0.19	0.064	0.020	0.018	0.014	0.079	0.034	0.026
5	0.13	0.061	0.024	0.019	0.16	0.065	0.030	0.019	0.048	0.08	0.069	0.025
4	0.15	0.058	0.022	0.020	0.15	0.062	0.028	0.020	0.132	0.077	0.053	0.025
3	0.15	0.062	0.013	0.020	0.17	0.066	0.021	0.020	0.145	0.082	0.045	0.024
2	0.19	0.059	0.019	0.019	0.23	0.063	0.031	0.019	0.173	0.079	0.056	0.025
Control variable:												
Y.o.b.*enr. rate in 1971	no		no		yes		yes		yes		yes	
Y o b* water and sanitation program	no		no		no		no		yes		yes	
R squared	0.19		0.14		0.19		0.14		0.16		0.13	
N. obs.	149,068		59,938		148,573		59,741		139,253		54,442	

Note: OLS estimates All regressions include year of birth dummies and region of birth dummies and year of birth* children in 1971 interactions
The omitted dummy (control group) is the dummy for being 24 in 1974.

Table A2 :Effect of the program on education and wages in various subsamples:
Coefficients of the interaction between age in 1974 and number of schools per 1000 children constructed in region of birth

	Dependent variable: years of education							Dependent variable: log(wages)					
	Low migration	Density		1976 Poverty		Pre-program education		Low migration	Density		1976 Poverty		Pre-progra
		<median	>median	High	Low	<median	>median		<median	>median	High	Low	<median
age in 1974	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
12	-0.017	-0.009	-0.132	-0.035	-0.104	-0.046	-0.006	0.0021	0.0238	0.0069	-0.0034	0.0255	0.0173
	(0.093)	(0.063)	(0.088)	(0.11)	(0.066)	(0.077)	(0.067)	(0.029)	(0.018)	(0.022)	(0.030)	(0.018)	(0.024)
11	0.029	-0.001	0.003	-0.110	0.048	0.030	0.047	-0.0122	-0.0060	-0.0027	-0.0003	-0.0209	0.0081
	(0.085)	(0.061)	(0.087)	(0.010)	(0.065)	(0.074)	(0.067)	(0.025)	(0.017)	(0.021)	(0.029)	(0.016)	(0.022)
10	0.175	0.112	-0.056	0.039	0.035	0.151	-0.026	0.0303	0.0289	-0.0240	0.0211	0.0006	0.0234
	(0.090)	(0.064)	(0.088)	(0.094)	(0.067)	(0.077)	(0.069)	(0.026)	(0.015)	(0.022)	(0.030)	(0.017)	(0.021)
9	0.259	0.139	0.124	0.289	0.100	0.183	0.050	0.0233	0.0091	0.0026	0.0565	0.0005	0.0369
	(0.073)	(0.052)	(0.073)	(0.11)	(0.055)	(0.064)	(0.056)	(0.022)	(0.018)	(0.018)	(0.027)	(0.016)	(0.018)
8	0.236	0.180	-0.161	-0.086	0.049	0.198	0.184	0.0614	0.0359	-0.0150	0.055	-0.0119	0.0406
	(0.095)	(0.066)	(0.090)	(0.11)	(0.068)	(0.082)	(0.066)	(0.029)	(0.017)	(0.018)	(0.031)	(0.015)	(0.023)
7	0.246	0.149	0.011	0.155	0.062	0.199	0.030	0.0059	0.0069	-0.0353	0.0473	-0.0277	-0.00884
	(0.089)	(0.061)	(0.084)	(0.099)	(0.062)	(0.075)	(0.063)	(0.027)	(0.016)	(0.022)	(0.030)	(0.016)	(0.021)
6	0.438	0.184	0.076	0.145	0.096	0.175	0.129	0.0210	0.0266	-0.0229	0.0293	-0.0187	0.0193
	(0.084)	(0.056)	(0.082)	(0.10)	(0.060)	(0.068)	(0.061)	(0.024)	(0.016)	(0.022)	(0.030)	(0.015)	(0.019)
5	0.230	0.221	-0.128	0.095	0.054	0.038	0.170	0.0443	0.0608	0.0214	-0.0069	0.0140	0.0238
	(0.085)	(0.057)	(0.083)	(0.098)	(0.060)	(0.069)	(0.063)	(0.027)	(0.018)	(0.021)	(0.031)	(0.016)	(0.022)
4	0.304	0.224	-0.127	0.156	0.005	0.140	0.045	0.0573	0.0320	0.0008	0.0659	-0.0001	0.0282
	(0.077)	(0.052)	(0.077)	(0.110)	(0.054)	(0.063)	(0.057)	(0.025)	(0.016)	(0.019)	(0.029)	(0.015)	(0.020)
3	0.317	0.205	-0.022	0.124	0.063	0.092	0.170	0.0410	0.0266	-0.0120	0.0669	-0.0021	0.0181
	(0.090)	(0.060)	(0.084)	(0.109)	(0.061)	(0.071)	(0.066)	(0.029)	(0.019)	(0.021)	(0.033)	(0.017)	(0.022)
2	0.357	0.213	0.071	0.357	0.055	0.198	0.121	0.0471	0.0239	0.0190	0.0672	-0.0043	0.0301
	(0.081)	(0.056)	(0.079)	(0.106)	(0.057)	(0.068)	(0.060)	(0.027)	(0.017)	(0.020)	(0.032)	(0.016)	(0.021)
F.stat.	6.6	5.38	1.52	2.56	1.05	2.87	2.76						
F.stat (wage earners)	1.88	3.72	0.44	3.72	0.37	2.7	0.91	1.43	1.95	0.83	1.74	0.95	0.91
R squared	0.096	0.12	0.24	0.11	0.20	0.15	0.14	0.14	0.15	0.14	0.12	0.13	0.15
N. obs.	54,648	77,018	75,994	61,408	87,744	76,499	76,563	19,033	23,672	36,069	23,152	36,859	26,672

Notes:
Standard errors in parenthesis. The F-stat tests the hypothesis that the interaction coefficients are jointly zero.
The control group is the group of people who were 13 to 24 in 1974.
Notes: Region of birth dummies, year of birth dummies, and interaction of treatment dummies with number of children and enrollment rate in the region in 1971 are included in the regressions.
The median density (I.e. the density for the region of birth for the median person in the weighted sample) is 308 habitants per square kilometer.
The high poverty provinces are the provinces where the proportion of people consuming less than 1,500 Rp per capita is larger than the national average (11.3%) for rural regions. (in the 1976 SUSENAS). I define as "high poverty" districts the rural districts in these provinces, which are: Lampung, Central Java, Jogyakarta, East Java, East Nusa Tenggara, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Maluku, Irian Jaya. Source: World Bank (1979)
The pre-program education is the average education in the region of birth for people born in 1962 or before. The median is 3.18 years.

Chapter 2

Child Health and Household Resources in South Africa: Evidence from the Old Age Pension Program

2.1 Introduction

Why are poor children in developing countries undernourished? Is it only because they live in an unhealthy environment and their parents are not educated or not healthy themselves? Does household income *per se* affect child nutrition? This would be the case if child health were in part a normal good consumed by the household or (even in child health were a pure investment good) if households were credit constrained.

More specifically, to what extent would an increase in the income of poor households in developing countries improve child health and nutrition? Would a temporary income transfer (a child grant for some years for example) have the same effect as a permanent increase in income? Would the gender of the recipient or her relationship with the child matter?

The answers to these questions have very important policy implications. Inadequate nutrition in childhood affects long run physical development, as well as the development of skills, and thus has an impact on health and productivity later in life. Low levels of investment in child health therefore have far-reaching consequences on economic growth, distribution, and welfare. Moreover, these questions are closely linked to the economics of the family. Following the tradition of the Becker (1965) model of household behavior, child health or nutritional status might be thought of as a commodity produced by the household. Calories, parental attention, community amenities, etc... are inputs into a child health production function which are chosen at the level of the household. In this context, several papers[1] have investigated the theoretical and empirical consequences of the fact that the household is a collection of individuals with different preferences. In particular, this implies that the distribution of income within the household, not only total household income, influences the choices that affect child health. Understanding how child health responds to an increase in income, and whether the source of the extra income affects this response, can therefore enrich our understanding of the economics of the family.

An important literature has examined the link between income and child height given age and sex, a long run measure of nutritional status (WHO (1986)). There is some evidence that height is positively affected by household long run resources (e.g Thomas, Strauss and Henriques (1990,1991), Thomas, Lavy and Strauss (1992), Sahn (1990)) but, according to the assessment in Strauss and Thomas (1995), much remains to be learnt. A common problem in this literature has been to identify exogenous sources of variation in income. It is clear from the household production model that income and child heath are jointly determined, and these papers do not, in general, treat total income or expenditure as exogenous. In a static household production model, non labor income or assets might be treated as exogenous; the literature often makes this assumption. However, in the more realistic dynamic model, assets and income are endogenous. Moreover they are difficult to measure accurately, and, even if we accept the static model, they are in general not randomly assigned. In

[1]Chiappori, (1988,1992) Bourguignon, Browning, Chiappori and Lechene (1994), Browning and Chiappori (1998)

most surveys, a large majority of households report having none. Households which have assets and non-labor income tend to live in less crowded areas with better hygiene, cleaner water and better health care services. For example, Thomas, Strauss and Henriques (1990b) find that controlling for observed community characteristics reduces the estimated effect of household per capita income (instrumented using non-labor income) by half, and makes it insignificant. Most estimates of the effect of family resources on child health are therefore likely to be biased by the omission of unobserved family or community characteristics.

A smaller literature tests whether income in the hands of the women of a household has a different impact on intra-household allocations than income in the hands of the men. The evidence suggests that, compared to income or assets in the hand of men, income or assets in the hands of women are associated with larger fertility decline (Schultz (1990)), larger improvement in child health (Thomas (1990)), and larger expenditure shares on household nutrients, health, housing and education (Thomas (1994b)).[2] However, these studies face the same difficulty as the studies described above. Total income and expenditures are not exogenous, and non-labor income (used in Thomas (1990,1993)), will not be a valid instrument in general. In addition, if part of a woman's assets were given by her family before or at the time of her marriage, the marriage market will in general insure that her assets and non-labor income will be correlated with (potentially unobserved) characteristics of her husband. This will invalidate any comparison, since the coefficient on the wife's non labor income will pick up the effects of unobserved husband's characteristics.

This paper seeks to estimate the effects on child health of exogenous variations in household income. The ideal experiment would be to allocate a grant randomly to some households and not to others (possibly after selecting both treatment and control groups in some population of interest) and to compare child health in both types of households. In the absence of evidence from such an experiment, it is necessary to rely on natural variations in combination with statistical modeling strategies. This

[2]Bourguignon, Browning, Chiappori and Lechene (1993,1994) and Browning and Chiappori (1998) test further implications of the collective model.

is true of the measurement of the effects of parental income on any dimension of human capital accumulation. In the US context, a handful of studies have tried to identify plausibly exogenous sources of variation in parental income and to measure their impact on child outcomes (some are reviewed in Shea (1997)). Mayer (1997) considers several approaches. Of the various pieces of evidence in her book, the fact the gap between the outcomes of children of single and married parents is not smaller in states that pay higher AFDC benefits is the most convincing. She interprets this as saying that benefits do not contribute to improve child outcomes. Shea (1997) studies whether children outcomes (education and subsequent labor earnings) are correlated with variations in father's income that are arguably due to "luck" (union status, job loss, industry). He finds no effect of these variables on child outcomes, except among poorest families. However, it is clear that the effects of parental income on child outcomes are likely to be of a larger magnitude among poor households in developing countries.

This paper exploits the dramatic increase in the coverage and the benefits of the Old Age Pension program in South Africa. At the end of the apartheid era, a pension program originally introduced as a safety net to provide for Whites reaching retirement without adequate provisions, and which remained for a long time in practice largely restricted to them, was extended to Africans.[3] The system is universal and non-contributory. All women above 60 and men above 65 are entitled to benefits, subject to a means test. In 1993, 80% of African women above 60 and 77% of African men above 65 were receiving the pension (Case and Deaton (1995)). In 1993, most pension recipients were receiving the maximum benefits of 370 Rands per months, which is roughly equivalent to twice the median income per capita in rural areas. The extension of the program to Africans caused therefore a large and unanticipated increase in the incomes of elderly Africans. But living arrangements in South Africa are such that pension recipients often live in extended households, with their children

[3]In what follows, I use in general the official denominations of racial groups in South Africa (Africans, White, Coloured and Indians). I sometimes use "Blacks" instead of "Africans" to refer to black South Africans.

and grand-children. As a result, 32% of African children under the age of 16 live with a pension recipient. The object of this paper is to evaluate whether access to the pension led to an improvement of the nutritional status of these children, and whether this effect differ by child and recipient gender.

The data used in this paper comes from a single cross section of South African households. At the end of 1993, the World Bank and the South African Living Research Unit carried out jointly a survey of a 9,000 households representative of the whole population. Measurements of height and weight of all children 72 months or younger were taken. However, a simple comparison between the average height of children living with a pension recipient and that of other children would not answer the question posed in this paper. As Case and Deaton (1996) point out, and as the descriptive statistics from my sample clearly show (table 1, columns 1 and 2), children who live with a pension recipient come on average from relatively disadvantaged backgrounds. This is reflected in their height. Children who live with a pension recipient are on average smaller for their age than other children.

The identification strategy used in this paper therefore exploits the fact that height reflects cumulated investments in child nutrition. The larger the proportion of her life during which a child was well nourished, the higher will be her standardized height for age. Due to the expansion in the program, those of qualified age became more likely to receive a pension after 1992 than they were before this date. The benefits became also substantially larger. Therefore, if part of the pension benefits was spent on improving child nutrition, the relative advantage in height for age of young children relative to older children should be larger if they live with an individual of qualified age than otherwise.

The basic idea of identification strategy is therefore to compare the differences in height for age between young and old children in households where an age-qualified person lives and in other households. This strategy allows height for age to vary systematically with age and with eligibility status of the household. The identifying assumption is that any difference between children in eligible and non-eligible households would have been the same in all age groups in the absence of the program.

By focusing on eligibility for pension instead of receipt of a pension, this strategy avoids any potential difficulty with the endogeneity of take up of pension. Eligibility for pension can then be used as an instrument for the actual receipt of a pension. There remains a potential issue of endogenous household formation, if the program caused modifications in living arrangements. This problem is addressed by using an alternative instrument. Instead of comparing children according to whether or not they live with somebody who is eligible for a pension, it is possible to determine whether the child has a grand-parent alive and likely to be eligible for a pension, even if they don't live in the same household. Children who are in this case might still be systematically different from other children (in particular, they have older parents), but this characteristics are fixed and cannot change endogenously as a response to the program.

Finally, implications of the identifying assumption can be directly tested. Weight for height is a short run measure of well being, which should be affected by the increase in income for *all* children (not only for the younger ones). Therefore there should be no difference in the differences in weight for height between young and old children between eligible and non eligible households. Likewise, no effects should be found by applying the same strategy on Whites, Coloured or Indians, because neither the pension levels nor the coverage of the pension program did improve in these racial groups during the period under consideration.

It is straightforward to extend this strategy to estimate separately the effects of pensions received by women and by men. The South African experience is therefore a unique opportunity to determine whether the gender of the recipient matters for the effect of extra income. This non-labor income is *not* tied to past labor income and savings and it was totally unexpected at the time marriage decisions were made. Moreover, even if households with a women recipient differ systematically from household with a male recipient, this will be taken into account by the identification strategy, as long as these differences are additive with respect to the age of the child.

The remainder of this paper is organized as follows. The next section presents a brief history of the South African Old Age Pension program. In section 3, I present the identification strategy and the main results. In section 4, I discuss potential problems of the identification strategy, and I present additional evidence to address those. In section 5, I present and interpret further results: a comparison of the effects of a Rand of pension income and a Rand of non-pension income, and a comparison of a Rand of the impact of pensions received by men and that of pension received by women.

2.2 Description of the program

2.2.1 The South African Old Age Pension program

This section present a brief history and an overview of the functioning of the South African Old Age Pension program. It draws extensively from Van der Berg (1994), Lunds (1993), and Case and Deaton (1996).

Social pensions were first introduced in the 1920's for Whites. They were intended mainly as a social safety net for a minority of white workers who were not covered by occupational pensions. They were gradually extended, but with very dissimilar benefits levels, to other race groups. During the apartheid era, the system was racially discriminatory in many respects. First, different means tests were applied to each race group. In 1984 for example, benefits were withdrawn for incomes larger of R 700 per annum (R 500 in Kwa Zulu) for Blacks, and for incomes larger than R 2250 per annum for Whites. Second, benefits levels were different. In the 1980s, benefits for Whites were 10 times as high as those for Blacks. Third, the delivery systems were different. Pensions of Whites were distributed through the postal offices, while they were distributed to Africans through mobile pay points that did not go very far out into rural areas. Officials seem also to have employed rather restrictive definitions of eligibility and age in order to save on the cost of pensions Lunds (1993). It must also be noted that fewer Africans survived to the age when they became eligible to the pension. In summary, before the late 1980's, even though Blacks were not formally excluded from the pension system, they did not benefit much from it.

Pressures for equity in the treatment of racial groups became strong towards the end of the Apartheid. In 1989, the government committed to achieving racial parity in pension treatment (Van der Berg (1994)). The extension of the social pensions to the whole population took several years and was operating fully in all areas (including rural areas) only by the beginning of 1993. Benefits improved gradually from the early 1980's to the early 1990, but much faster in the 1990. Benefits nearly doubled between 1990 and 1993. Parity between benefits for Blacks and Whites was achieved in 1993. In 1992, the means tests were also modified, and unified across races. The current system is universal and non-contributory. Payments are made to women older than 60 and to men older than 65, subject to a means test. The maximum amount (370 Rands per month, or $3 per day) is paid to eligible individuals whose income (including income imputed from assets) is less than 90 Rands per month. For incomes higher than 90 Rands per months, the pension is reduced by one to one until income reaches 370 Rands, at which point the pension is suppressed. For couples, household resources are roughly divided by two. Importantly, the income of other members of the household is not taken into account when implementing the means test. There are therefore no direct incentives to partition the household or to stop working for other household members. In practice, the means test does not seem to be applied very finely. It is mainly effective in excluding most Whites as well as some wealthy Africans.

In 1993, 80% of the Africans eligible on the basis of their age were receiving a pension. Of those, most were getting the maximum amount. There is no very good estimate of the coverage earlier on. First, social pensions were administered by 17 different chambers, which made any evaluation difficult. Second, surveys (including the 1991 census) excluded the "independent homelands", where many Africans live.[4] The coverage increased substantially in the 1990's, due to a new attitude of the administration, a modification of the means test, and substantial improvements in the delivery system. The delivery is now made once a month or once every two months

[4]Van der Berg (1994) reports that about half of the age-qualified Africans were receiving some benefit in 1980 (rarely the maximum amount).

by mobile pension teams equipped with ATMs with a fingerprints recognition system. This recent technology has considerably improved the day to day administration of the pension. In sum, the social pensions program, which was not generous to Blacks with respect to coverage and benefits throughout the apartheid era, became very important in the 1990s, and especially after 1992, when means tests were unified, the new delivery system was introduced, and racial parity in benefits levels was achieved.

The benefits are large: 370 Rands is about half the median household income, and more than twice the median household income per capita among Africans. Given the high level of unemployment, it is frequent that the pension recipient is the main income earner in the household. Households with a member eligible for the pension experienced therefore a substantial improvement in income in the early 1990s, and a more dramatic increase since 1992.

2.2.2 Pension benefits and pension recipients

The data for this paper come from the national survey of South Africa carried out jointly by the World Bank and the South African Development Research Unit at the University of Cape Town (SALDRU). During the last five months of 1993, randomly selected households from all races and areas, including the "independent homelands", were interviewed. This is a multipurpose household survey similar to most World Bank Leaving Standards Measurement surveys. As part of the survey, measurements of the height and weight of all children aged less than seven years were taken. I follow the norm recommended by the World Health Organization and applied by most researchers: for each age in months, I calculate the height for age Z-score by subtracting the median and dividing by the standard error in this age and sex group in the NCHS reference population (a group of well-nourished American children). The World Health Organization recommends to limit the analysis of Height-for-age measures to children 0 to 5 year old(WHO (1986)). There appears to have been difficulties in the measurement of the oldest and the youngest children.[5] I therefore

[5]Some six year old children were recorded by the interviewers to be seven, and thus were not measured. A large fraction of the Z-scores for children aged 6 months or younger seem to be obvious

restrict the sample to children born between June 88 and April 1993 (their ages range between 6 months and 5 years). Descriptive statistics of the sample of black children are presented in table 1. Column 1 and 2 show the means of the variables used in the analysis, by pension status. Households where there is a pension recipient tend to be poorer than the median (their pre-pension monthly income is 661 Rands, compared to 930 Rands in other households). Income after pension is higher in households that get a pension, but income per capita remains slightly smaller. They are, not surprisingly, often characterized by the presence of a grand parent (93%), and the absence of the child's father (64%), or mother (17%). Parental education in households where the parents are present does not differ by pension status. However, the education of the household's head is lower in pension households (this reflects the fact that the head is more likely to be one of the child's grand-parents). They are also more likely to live in a rural area.

It is therefore not surprising that children who live with a pension recipient tend to be smaller than other children. Even if child nutrition has improved as a result of the extension the coverage and benefits of the Old Age Pension program, height for age still reflects past deprivations or illnesses, especially among the oldest children. Descriptive statistics of height for age and weight for age are shown in table 2. Taking all children together, average height for age of children living with a pension recipient is -1.41, compared with -1.21 for other children. For older boys, these numbers are respectively -1.59 and -1.34 (and -1.48 and -1.26 for girls). Interestingly, weight for height, a measure of current nutritional status, are similar for children in both types of households (0.17 and 0.16, respectively). But it is likely that family background affect short run nutrition in ways that cannot directly be captured by income: the simple comparison of weight for height (or expenditures on child health and nutrition) would reflect these differences, which are systematically correlated with pension status, and would not be an unbiased indicator of the effect of pension income on child health. Systematic differences in the family backgrounds of children living with a pension recipient and that of other children seem therefore to preclude a simple analysis of the

---

outliers (larger than 10 in absolute value).

effects of the pension on their health. This was first noted by Case and Deaton (1995), who reported that height for age are significantly smaller in pension households than in non-pension households. Using the same data, Peter Le Roux (1995) found that this negative correlation disappears when controlling for family background. In the next section, I propose a strategy to identify this impact using the same data.

2.3 Estimating the effect of Social pension on Child health

2.3.1 Identification strategy

In practice, in developing countries, growth deficits are caused by two preventable factors, inadequate food and infections. Genetic factors matter for child height, but they become more critical in adolescence. In childhood, height for age and weight for height Z-scores are widely considered to be "the most useful tool for assessing the nutritional status of children" (WHO (1986)).

Health is a capital stock, which depends on lagged as well as current inputs. Height for age of young children depends on cumulated investments over the life of the child (Martorell and Habicht (n.d.)). As in most studies of the determinants of nutritional status, consider a household production model in the tradition of Becker (1965). Household members exercise choices over consumption (including leisure), and the number and the quality of surviving children. The resulting maximization problem generates reduced form expressions of height for age and weight for height as functions of individual, family and community characteristics (unobserved or observed) and exogenous income sources of family members. An unanticipated exogenous change in income should affect investments, but any change in the stock (height) will take time to be perceptible (especially if children have accumulated growth deficits). This is the basis of the identification strategy used in this paper.

Coverage and benefits increased rapidly in the early 1990s. Age-qualified individuals became much more likely than before to receive the pension, and conditional on receiving one, they also received higher benefits. The decision to apply for a pension is a decision of the household, and should be considered as endogenous.[6] However, if we assume that there was no changes in the household composition as a result of the program (this issue will be examined in details below) eligibility is a fixed household characteristic, not a choice variable. Since over 80% of eligible individuals were receiving the pension in 1993, eligibility for pension is clearly a good instrument for pension receipt. We can therefore focus on the question of whether children living with an individual eligible for the pension benefited from the extension of the program.[7]

For simplicity of exposition, assume (pending discussion) that the extension of the Old Age Pension program was not anticipated and was suddenly introduced in January 1992. The household decision problem changed in January 1992. All children were measured at the end of 1993. Children born after January 1992 have been exposed to the pension all their lives. Children born before this date have been exposed only during a fraction of their lives. Households where both children and elderly live tend to be poorer than other households. Therefore, the oldest children, who have been exposed only during a fraction of their lives and grew up in otherwise less favorable environments, should be smaller in "eligible" households (households where an eligible recipient lives) than in non-eligible households. However, if the pension had an effect, the difference between eligible and non eligible children should be smaller for younger children, or can even be reversed.

The basic idea of the identification strategy is thus to compare the differences between the height of children in eligible and in non eligible households for children exposed to the program for a fraction of their lives and children exposed all their lives. To illustrate this strategy, I present in table 2 descriptive statistics of height for age in different sub-samples. Columns 3 and 4 show the means of height for age in households where there is an eligible woman and in household where there is an

[6]Even if all eligible individuals will receive the pension eventually, the date at which they became pension recipients is potentially related to unobserved household characteristics. It is quite possible for example that poorer eligible individuals got access to the pension later than other people (as difficulties and costs faced by applicants got progressively removed, cf. Lunds (1993)).

[7]This is the reduced form that follows from the household maximizaton problem, considering that the decistion to take up the pension is endogenous.

eligible man, respectively. Column 5 shows these means in households where there is no eligible member. Among children born before January 1992, both boys and girls are smaller in households where there is an eligible member (woman or man) than in other households. However, girls born after January 1992 are taller if they live with an eligible member (especially with a woman). This is not true for boys. This suggests that the pension seemed to have an effect on the nutrition of girls, but not on that of boys.

I present a non-parametric version of this comparison in the next subsection and a differences in differences formulation in the following subsection.

2.3.2 Non-parametric approach

The least restrictive implementation of the identification strategy is to plot height for age as a function of date of birth in eligible and non eligible households, and to examine the relative positions of these two curves.

This non-parametric approach is compelling for two reasons. First, the reform was not suddenly introduced at a well identified date, but it took some years (from 1989 to 1993) to achieve universal coverage and parity in the benefits. Choosing January 1992 as the date of the reform is therefore somewhat arbitrary. Second, even if the program had been introduced once and for all in 1992, children born before 1992 would nevertheless have been exposed part of their lives. The actual dimensions in which the treatment varies are the fraction of her live during which a child has been exposed to the program, and her age at first exposure. However, we have little knowledge about the functional form of the child health production function. The non-parametric formulation imposes neither a definite reform date nor a functional form. To answer qualitatively the question of whether the pension was effective or not, it is enough to know that the coverage and the benefits were increasing over the period, so that young children were more exposed to the program than older children.

Figure 1 shows non parametric (kernel) regressions of height for age Z-score as a function of date of birth, in eligible (straight line) and non eligible (broken line) households. These curves have the shapes traditionally found in developing countries:

height for age Z-scores decline fast in the first two years of life, and then stabilize. They are negative: black South African children are smaller than young Americans their age.[8] The interesting point, for our purpose, is that the curves have the relative positions predicted in the preceding discussion. They cross in January 1992. Children born before January 1992 are smaller in eligible households (the gap starts to close for children born after July 1991). Children born after January 1992 are taller in eligible household. This pattern is even more striking when looking at girls alone (figure 3). The advantage of girls living in eligible households over those who are not, among young girls, is as large as their handicap among older girls. For boys (figure 2), however, there does not seem to be a similar effect. Younger boys in eligible and non eligible households are about the same height as other boys (while they were smaller before).

This evidence suggests that the increase in pension contributed to improve child nutrition, especially that of girls, and resulted in faster growth for the youngest children. The pensions did not only help the girls living in an eligible household to bridge the gap with the other girls, they seemed to help them to do *better* than those. In the next subsection, I propose a simple parameterization of these effects.

2.3.3 Differences in differences formulation: statistical framework

In this subsection, I present estimates based on a more restrictive difference in differences formulation. The advantage is that single parameters can be estimated with more precision, and I can present confidence intervals, introduce control variables and compare different estimates. I also use this approach to present instrumental variables estimates the effects of pension receipt. The drawback is that estimates obtained from this formulation should be interpreted with caution, because they are based on additional assumptions.

[8]A Z-score of 0 indicates that the child has the same size has an American child of the same age and sex. A Z-score of -2 is generally considered as an indicator of *stunting*.

Reduced form: effects of eligibility

The discussion in the preceding subsection suggests the following formulation:

$$h_{ifk} = \alpha 1_{(k=1)} * T_f + \beta T_f + \sum_{j=1}^{3} \gamma_j 1_{(k=j)} + X_{ifk}\delta + \sum_{j=1}^{3} 1_{(k=j)} * X_{ifk}\lambda_j + \epsilon_{ifk}, \quad (2.1)$$

where h_{ifk} is the height for age Z-score of a child born in cohort k in family f, $1_{(k=j)}$ denotes an indicator variable equal to 1 if k is equal to j, and 0 otherwise. I consider four cohorts: children born after January 1992 ($k = 1$), children born between January and December 1991 ($k = 2$) children born between January and December 1990 ($k = 3$), and children born between June 1988 and December 1989 ($k = 4$). Children born after January 1992 form the "exposed" generation. Children born before January 1992 are considered as "non-exposed". T_f is an indicator variable equal to 1 in families where there is an eligible member, and to 0 otherwise. The last two terms (X_{ifk} and $\sum_{j=1}^{3} 1_{(k=j)} * X_{ifk}$) are family background variables and family background variables interacted with cohort dummies. I estimated three forms of equation 2.1: without control variables (except for child sex), with non-interacted control variables, and with interacted control variables. I included the following family background variables: mother's and father's education, rural, urban or metropolitan residence, mother's and father's age.[9] This equation is estimated by OLS, and standard errors are adjusted to take into account correlation of errors terms between children in the same household as well as heteroscedasticity. The coefficient of interest is α, the coefficient of the interaction between the eligibility status and the dummy for belonging in the youngest cohort.

[9] I have replaced these variables by sample means in cases where the father or the mother of the child were absent, to avoid selecting the sample using this criterion. I have also estimated specifications where I control for the presence of the child parents. I don't report these estimates because parental presence is potentially endogenous to the outcomes I consider, but they are very similar to the estimates reported in this paper.

Effects of the pension

In 1993, 82% of the children living with an eligible woman and 79% of those living with an eligible man lived with a pension recipient. 6.5% of the households where nobody was eligible received a pension.[10] Pension receipt is a choice variable and must therefore be considered as endogenous. However, eligibility status can be used as an instrument for receipt of a pension after the extension of the program.

I practice, I estimate using 2SLS the following equation:

$$h_{ifk} = \alpha 1_{(k=1)} * P_f + \beta T_f + \sum_{j=1}^{3} \gamma_j 1_{(k=j)} + X_{ifk}\delta + \sum_{j=1}^{3} 1_{(k=j)} * X_{ifk}\lambda_j + \epsilon_{ifk}, \quad (2.2)$$

where P_f is a dummy equal to 1 if the household receives a pension (in 1993). The excluded instrument is the interaction $1_{(k=1)} * T_f$.[11]

The corresponding first stage is therefore:

$$1_{(k=1)} * P_f = a 1_{(k=1)} * T_f + bT_f + \sum_{j=1}^{3} c_j 1_{(k=j)} + X_{ifk}d + \sum_{j=1}^{3} 1_{(k=j)} * X_{ifk}l_j + v_{ifk}, \quad (2.3)$$

Note that some children who have been living with an eligible member part of their life, or even all their life (since some households were already receiving a pension – albeit smaller– in 1992) are included in the control group. Therefore, if the equation is otherwise correctly specified, estimates of α in equation 2.2 should be a lower bound of the true effect of the pension on the youngest children.

2.3.4 Differences in differences formulation: Results

- **Reduced form estimates**

[10]This comes in part from the fact that I consider as eligible people who were eligible in 1992 (i.e women above 62 and men above 67) and in part from the fact that individuals receive a pension despite not being eligible. This is especially frequent among men, because in some regions, officials applied the 60 years eligibility threshold to men as well as to women.

[11]The standard errors are adjusted to take into account correlation of outcomes between children of the same household.

Estimates of equation 2.1 are presented in column 1, 2 and 3 in table 3. These results are consistent with the patterns displayed by the graphs. In the complete sample, there is a positive, marginally significant difference in differences. The coefficient of the indicator for eligibility (the difference between the height for age of eligible children born before January 1992 in eligible and non eligible families) is negative and significant. Results are not affected by the introduction of the family background variables.

It turns out that the results for the complete sample confound small and insignificant differences in differences for boys, and much larger and significant estimates for girls. For girls, the lowest point estimate is 0.47, (for an aveiage height for age of -1.23 among eligible girls) which is more than twice as large as the coefficient of the indicator for eligibility. This means that young girls are taller if they live with an eligible individual than otherwise, while the reverse is true for old girls. Interestingly, the main effects are similar for boys and girls (-0.28 and -0.21, respectively).

This suggests that the fact of living with an eligible member had a large positive effect on the nutrition of girls born after the extension of the Old Age Pensions program, while it had little or no effect on the nutrition of boys.

- **2SLS estimates of Effect of the pension**

Estimates of the first stages (equation 2.3) are presented in column 1, table 4.[12] There is a strong association between pension receipt and eligibility for pension and , not surprisingly, the first stages are highly significant in all sub-samples. The coefficient of the interaction between eligibility and belonging to the youngest cohort is 0.74, with a t. statistic above 20.

2SLS estimates of equation 2.2 are presented in table 5 (column 1 to 3). For girls, the IV estimates are positive, large (0.66) and significant at the 10 percent level. They are small and insignificant for boys. The effect of the pension on girls is large enough to bridge half of the gap between African girls ages 0 to 5 and American girls

[12]First stages are virtually identical across specifications. I present only the specification with the complete set of control variables and interactions. The standard errors take into account the fact that observations are the same for all children belonging to the same household.

their age. These effects are large, but note that the pension represents a large income transfer (twice the median household income per capita in rural area).

For comparison, the results of estimating a specification similar to equation 2.1, but where an indicator for receiving a pension is used instead of an indicator for the presence of an eligible individual in the household, are presented in table 3 (columns 4 to 6). The differences in differences estimated by OLS are small and positive but not significant for girls, and very close to 0 for boys. The straight differences in differences seem therefore to underestimate the effects of the pension. This suggests that the date of the pension take-over was systematically correlated with characteristics directly affecting child health.

The IV estimates reinforce the conclusions from the previous subsection. There is a large effect of the pension on girls, and little or no effects for boys.

2.4 Discussions of the identifying assumptions and additional evidence

2.4.1 Endogenous household recomposition

Endogenous household recomposition as a source of bias

Until now, I have assumed that household eligibility status was not affected by the program. However, changes is family composition might have occurred as a response to the program. Living arrangements under which children are living with their grand-parents (with or without their parents) are traditional in South Africa. But some living arrangements might have been modified as a result of the pension, potentially in ways that differ among households. This could affect the validity of the identification strategy proposed in this paper. Assume for example that parents who care the most about the health of their children have sent them to live with their grand-mother after she started receiving the pension. This will imply that children who live with their grand-mother come from families who care more about their health, and therefore they might be taller than their peers for reasons other the increase in income per se. If this effect is additive (for example if "good" families send *all* their children to live with their grand-mother after she starts receiving the pension), this does not invalidate the spirit of the strategy, which is to compare the differences between old and young children in eligible and non-eligible households. But if health-conscious parents send only their youngest children to live with their grand parents, younger children might be taller than their peers in eligible families for reasons other than the pension. Conversely, if health-conscious parents send only the oldest children to live with their grand-mother after she starts receiving the pension, this will result in downward biased estimates of the effects of the program.

Alternative identification strategy

To address this problem, I use an alternative instrument, which is designed to be correlated with the presence of an eligible member in the household, but not affected by household decisions. This instrument is a variable that indicates whether the child has at least one grand parent who is alive and eligible, or likely to be eligible. The survey instrument asks household members whether their parents are alive. It does not ask any other question about them if they are not in the household. This allows me to construct a variable indicating whether the child has at least one grand-parent alive, but I cannot establish whether the grand-parent is eligible if he or she does not live in the household. However, if the mother and father are old enough their parents (if alive) are likely to be eligible. In practice the instrument is equal to 1 if there is an eligible person in the household or if one of the following is true: the mother (resp. the father) of the child is older than 34 and her (resp. his) mother is alive or the mother (resp. the father) of the child is older than 32 and her (resp. his) father is alive.[13] 46% of children who have an old grand parent alive (and 7.8% of those who don't) live with a pension recipient. Therefore this instrument is still strongly correlated

[13] I determined the cutoffs of 32 and 34 years by using the information on extended families in my sample. Women whose observed child is above 34 and men whose observed child is above 32 have a probability of 60% to be eligible for the pension. Results are not sensible to the choice of the cutoff. If a parent is not in the household, the survey does not indicate nor his age, nor whether his parents are alive. So some children may have an old grand parent alive not identified in my data, but chances are that they have little conexions with him or her.

with pension receipt. We can use it as an alternative instrument for pension receipt, and check whether results are consistent with those obtained using eligibility.

The two sub-samples (where this instrument is 0 or 1) are more similar to each other than eligible and non eligible households are. This is apparent in the descriptive statistics presented in table 1 (columns 6 and 7). Non pension income, head's education, fraction of households where the father is absent, etc... are now closer to each other. Two sets of characteristics are (not surprisingly) different: the child's parents are on average older and family size is on average larger when the child has an old grand parent alive.

Table 2 (columns 6 and 7) shows means of height for age. The fact that the two sub-samples are more similar is reflected in these statistics as well. The difference in height for age between children living in households who have an old grand-parent alive and the others among children before January 1992 is -0.17 for boys (against -0.30 between eligible and non eligible), and -0.04 for girls (against -0.20 between eligible and non eligible). These descriptive statistics have the expected pattern. Among girls born after January 1992, girls who have an old grand-parent alive are taller than other girls. The opposite is true for girls born before January 1992. For boys, we see a closing of the gap, but boys born after January 1992 are still smaller if they have an old grand-parent alive than otherwise. This suggests that using this instrument instead of eligibility directly should leave conclusions unchanged.

Figure 4 shows non parametric regressions of height-for-age as a function of date of birth for children who have an old grand parent alive (straight line) and other children (broken line). Not surprisingly, the differences before the program are less marked in this case. Children born before January 1992 and who have an old grand parent alive are sometimes taller, sometimes smaller than the others. This reflects the fact that they come from more similar backgrounds. However those born after January 1992 are definitely taller if they have a grand-parent alive. This graphs suggests therefore that having a grand-parent alive resulted in improved nutrition for the youngest children.

Estimates of a specification similar to equation 2.1, but where I use the indicator for whether the child as an old grand parent alive instead of eligibility status are presented in table 3 (columns 7 to 9). For girls, I find a positive (but insignificant) effect of having a grand parent alive, smaller than the estimated effect of eligibility. This is what we expected, since the probability of getting the pension is larger conditional on living with an eligible member than conditional on having an old grand parent alive. For boys, the differences in differences are larger than the estimated effect of eligibility, but they remain insignificant.

We can then compute 2SLS estimates of the effect of pension receipt (equation 2.2), using the interaction between the indicator for having an old grand-parent alive and being born after January 1992 as an instrument. The first stages are shown in table 4 (columns 2 and 3). There is still a strong relationship between the receipt of the pension and this instrument. The coefficient is 0.39 without controlling for family background variables, and 0.54 with control for these variables (interacted with cohort dummies), with t. statistics above 14.[14]

The 2SLS estimates are shown in table 5 (columns 4 to 6). For girls, the point estimates of the effect of pension using this instruments are slightly higher than those using eligibility as an instrument, but very close (especially in the specifications that uses the full set of control variable). These estimates are also less precise. This result indicates that endogenous family composition does not bias the estimates of the effects of eligibility for girls. For boys, I find large but imprecise point estimates.[15] As a consequence, point estimates in the whole sample are larger using this instrument (although this difference is not significant). It could be that using eligibility as an instrument leads to a downward bias in the estimates for boys (but not for girls). Alternatively, transfers from the grand-parent to the household could explain part of this difference (this would suggest that having a grand-parent alive and old is an instrument for living with a pension recipient *or* receiving a transfer from a pension

[14]The difference between the value of the first stage with and without controls shows that, within children who have an old grand parent alive, only those from certain backgrounds live in the same household as he or she does.

[15]It turns out (cf. section 5) that this is due to some positive effects of having a grand-father alive for boys.

recipient). There are not many instances of transfers from parents to the household recorded in this data, but transfers are difficult to measure accurately and they might be more prevalent than it appears. It is quite possible that transfers benefit boys more than girls. In any case, these alternative estimates do not suggest that using eligibility as an instrument leads to upward biased estimate of the effects.

2.4.2 Functional form

Another potential problem for the interpretation of differences in differences is that some unobserved factors correlated with the pension might have different effects at different ages. For example, unobserved quality of the family might be a stronger determinant of height for age for older children than for younger children. This is in fact likely to be the case. In family who care less about their children, older children have been exposed for a longer time to inadequate nutrition than younger children. To illustrate this possibility, I ignore for now the issue of endogeneity, and I show in figure 5, 6 and 7 that the effect of poverty (per capita income below median) seems higher for older boys than younger boys, but that it is not the case for girls. I graph height for age as a function of date of birth for children living in households whose income per capita is respectively above and below the median. Boys in households whose income is above median are taller than poorer boys at all ages, but the difference between the two groups is larger for the older boys than for the younger. Unoberved household characteristics could likewise have different effect at different ages.

The ideal strategy would be to use at least two cross-sections and to distinguish age and cohort effects. However there is to date only one representative survey where South African children were measured. However, this problem does not seem to affect my estimates of the effects of the program on girls. First, the effect of household per capita income seems to be important for girls at all ages. Second, the kernel regressions of height for age as a function of age in eligible and non eligible households actually *cross*. Younger girls are taller in eligible households and the reverse is true for older girls (cf. figure 3 and table 2). This pattern could be explained only by an unobserved factor affecting nutrition negatively at some ages and positively at other ages. However such a factor would also be reflected in weight for height, which is a short run measure of nutrition. In table 2, I show that weight for height is larger for young *and old* girls leaving with an eligible woman. This is illustrated in figure 12, which shows weight for height as a function of date of birth for girls living in eligible and non eligible households. The modest positive differences in differences found for boys might, however, be spurious. This tends to reinforce earlier conclusions: there seems to be a large effect of the program for girls, and little or none for boys.

Another possibility along the same line is that the presence of a grand-parent in the household has a direct impact on the health of the child, which is different depending on the age of the child. This is easily checked directly. First, controlling for the presence of a grand parent (or the fact that a grand-parent is alive), interacted with cohort dummies, does not change the estimates of equation 2.1. Second, the estimates are similar, and if anything slightly higher in a sub-sample of children who live with at least one grand-parent (eligible or not for the pension). Third, I estimated a difference in differences specification of the effect of the presence of a grand-parent in the household in the sample of non-eligible households (these are therefore grand-mothers younger than 60 and grand-fathers younger than 65). These differences in differences are actually negative (but not significant). So the differences in differences for eligible members do not come from the fact that living with a grand-mother is more beneficial in early childhood than afterwards.

2.4.3 Control experiments: Weight for height and evidence from other groups

Two outcomes can be used to test some implications of the identification assumptions. These outcomes should not be affected by the program, but the sources of misspecification mentioned earlier would cause the differences in differences to be positive for these outcomes. This provides useful "control experiments".

Weight for height

Weight for height is a source a precious additional evidence. Unlike height for age, weight for height is, in childhood, an short-run measure of nutritional status. Whereas height for age reflects nutrition since birth, weight for height reflects nutrition and illnesses within the past few weeks. Therefore, if the differences in differences in height for age are due to improved nutrition and not to differential effects of unobserved family characteristics (or to endogenous family recomposition affecting only the youngest children), there should be no differences in differences in weight for height. The nutrition of all children should have improved as a result of the pension. This will not necessarily imply that children in eligible households are heavier than other children, because these households are different along other dimensions, but the difference between the weight for height of children in eligible and non eligible families should not vary with age. Weight for height is therefore an interesting control experiment.

Panel B in table 2 shows means of weight for height in various sub-samples. Boys and girls of all ages are heavier if they live with an eligible woman than if they live with no eligible member. The difference is larger for girls than for boys (0.20 compared to 0.07), which is consistent with previous results (nutrition has improved for girls but not much boys). Boys living with a pension recipient are thinner than other boys whatever their age, whereas the opposite is true for girls. Weight for height is very similar among children having a grand parent alive and other children.

Non-parametric regressions of weight for height Z-scores on date of birth are shown in figure 12 for all children, and in figures 13 and 14 for boys and girls separately. Most children (younf *and* old) are slightly heavier in eligible families. The difference between the weights of girls in eligible families and that of other girls is greater than the corresponding difference for boys, except among infants. To anticipate somewhat on future results, we note in figure 15 that children are doing particularly well in families where there is an eligible grand-mother. Children in these families are almost all heavier for their height than other children.

Table 6 confirms the impression given by these graphs. In this table, I present estimates of differences in differences specifications similar to equation 2.1, but where weight for height is the dependent variable. In contrast to previous results, there is no significant difference in differences in this table. Estimates for girls are much smaller than what was found for height for age (and sometimes negative). For boys, as before, I find small and insignificant estimates. This is true across all definitions of the treatment: receipt of a pension, presence of an eligible of the household, or existence of a potentially eligible grand-parent. This gives some reassurance that the pattern found before (little or no effect of the program on the nutrition of boys, and a strong effect on that of girls) was not an artifact of misspecification.

Height for age in non-affected groups

Another control experiment is given by height for age in groups where the pension benefits did not increase in the 1990s. This is the case for all racial groups but Africans. Pension benefits for Whites actually declined somewhat in the 1990s. The coverage and the benefits levels for Indians and Coloured increased substantially in the early 1980s as part of the effort to make the three chambers parliament viable (Van der Berg (1994)), but they did not increase further in the 1990s. Therefore, there should be no differences in differences in the height for age in the set up I analyze here. There are not enough non-black children to perform a non-parametric analysis, or even to look at the results separately by race and gender. I present in table 7 the estimates of equation 2.1 in the sample of all non-African children, where T_f is respectively receipt of a pension, eligibility status, and existence of an old grand parent. Case and Deaton (1996) document that pension recipients are much poorer than non pension recipients in this group (there are more of them among Coloured than among Indians and Whites, and only the poorest Whites do not have a private pension). It is therefore not surprising that children living with a pension recipient are much smaller than other children. A spurious positive difference in differences would then be obtained if, as discussed in the previous subsection, the effect of an unobserved characteristic correlated with poverty was less important for younger children than other children. The difference in differences is positive only when I use pension status as the treatment variable, but it is small and not significant. Interestingly, it becomes much smaller when I control for family background. Using, as in the rest of the paper, eligibility status or the presence of an old grand-parent as treatment status, the differences in differences are actually negative.

In summary, I do not find in the positive difference in differences that are found among Africans in other racial groups. This result, in addition to the weight for height results, gives some reassurance that the results found for Blacks are not spurious.

2.5 Further results and interpretation

This section is devoted to a more detailed analysis of further results and of their implications. What can these estimates teach us about income effects and the economics of resources allocation within the family? Two elements are particularly striking: the magnitude of these effects on girls and the differences by recipient gender.

2.5.1 Pension versus non-pension income

In section 3, we concluded that the effects of the pension increase on boys' nutrition were at best small, but that there were large effects on the nutrition of girls. The pension led to an increase in the Z-scores of the youngest girls of more than half a standard deviation (of the size distribution of American children). The sample average is -1.31 standard deviations among eligible girls. This means that the pension helped them to bridge more than a third of the gap with American girls. How would this increase compare to the effect of another income shock?

Conceptual framework

Two main models can explain why an increase in income can be expected to affect child height. They have different implications for the comparisons of the effects of and Rand of pension income and a Rand of non pension income.

First, in a pure investment model, parents do not derive any direct utility of the child health, but healthy individuals are more productive, and this productivity will be rewarded in self employment (Strauss (199?)), or by the market (e.g. Rosenzweig (1988), Schultz (1996); cf. also Rosenzweig (1999) for returns to human capital in the marriage market). Parents invest in child health in anticipation of these returns. In this model, a change in income will change the level of investment into child human capital only if households are credit constrained. Garg and Morduch (1996) present indirect evidence that poor households in Ghana are credit constrained: keeping fixed the total number of siblings, children of both sexes are on average taller if the proportion of sisters among their siblings is larger. If households are credit constrained, this can be explained by the fact that returns to the human capital of woman are lower (and therefore each girl requires lower investment). As this example shows, any relaxation of the credit constraint (even if this does not correspond to an improvement in permanent income) will then result in better child nutrition.

Second, some models treat child health (and, in general, child quality) as a consumption good. Household members derive utility from healthy children. If child quality is a normal good, a change in permanent income will affect child health even if households are not credit constrained. A transitory change in income, by contrast, should not affect investments into child health, unless the household is credit constrained.

Third, both models can be extended to take into account the fact that the household is a multi-person decision unit. If household members have different preferences (for example, different effective discount rates, or different valuation of child health), the way in which the preferences are taken into account in the household decision process will matter for final outcomes. This point will be extended below.

Pension income has several characteristics which might lead its effect on child nutrition to be smaller or larger than the effect of another income shock. First it is a regular income. Few Africans have a stable labor force attachment or a regular source of income, especially in rural areas. Therefore, in households where there is a pension recipient, pension benefits is probably one of the most regular sources of income. The propensity to spend out of pension income might therefore be greater than the propensity to spend out of non-pension income, if non-pension income is more akin to transitory income. When Paxson (1992) shows that in Thailand, the propensity to save out of transitory income is higher than the propensity to save out of permanent income, she uses a very short term horizon when defining permanent income. Permanent income is defined in her work as expected income *for a year* conditional on the resources and the information of the household at the beginning of the period (transitory income is the difference between realized and expected income). According to this definition, in most households, pension income is a permanent income, while other sources of income will be, part permanent, part transitory. We could therefore expect that the propensity to spend out of pension income is larger than out of non pension income. Moreover, anecdotal evidence (Lunds (1993)) suggests that pension recipients can borrow against future pension income. She reports that in some village, only households where a pension recipient live that can borrow. If the pension actually relaxes the household credit constraint, this will imply that the propensity to spend on child health out of pension income should be especially high.

Second, however, if we consider a larger horizon of permanent income, a Rand of pension income today represents less than a Rand of permanent income, since it is tied to an elderly person and will stop when the elderly person dies. If households are not credit constrained and child quality is a normal good, we should therefore the propensity to spend on child health should be smaller out of pension income than out of non-pension income. In this respect, there is also a difference between a pension received by a man and a pension received by a woman. Men receive on average the pension for a much shorter time than women (both because they get it later and they die earlier). Therefore a rand of pension income received by a man represents less, in term of permanent income, than a Rand of pension income received by a woman.

Third, pension income is received by an elderly (and more often by a woman than by a man). This improves her bargaining position, and might matter for the allocation of this income. The next subsection will be devoted to this issue.

Estimation methods

The strategy used to estimate the effect of the pension can be used to estimate the effect of each Rand of pension. The following equation can be estimated by 2SLS.

$$h_{ifk} = \alpha 1_{(k=1)} * y_f + \beta T_f + \sum_{j=1}^{3} \gamma_j 1_{(k=j)} + X_{ifk}\delta + \sum_{j=1}^{3} 1_{(k=j)} * X_{ifk}\lambda_j + \epsilon_{ifk}, \tag{2.4}$$

where y_f if the logarithm of 1 plus the pension amount (expressed in hundreds of Rand). The instrument is again the interaction $1_{(k=1)} * T_f$. The first stages are similar to equation 2.3, with y_f replacing P_f.

Unfortunately, I cannot estimate consistently the effect of a Rand of non-pension income on height for age. As mentioned earlier, non-pension income is not an exogenous variable, and OLS regressions will lead to inconsistent estimates. Nevertheless, even a simple comparison of IV estimates of the effect of pension income with OLS estimates of the effect of non-pension income is instructive. The presumption is that endogeneity and non-randomness of labor income should lead to upward biased estimate of the OLS coefficient. Therefore, if we find that the effect of a Rand of pension income is larger than the effect of a Rand of non-pension income, this result should be taken as robust.

I estimate the following specification (which allows the effect of income to differ by age):

$$h_{ifk} = \alpha_1 1_{(k=1)}) * z_f + \alpha_2 (1 - 1_{(k=1)}) * z_f \sum_{j=1}^{3} \gamma_j 1_{(k=j)} + X_{ifk}\delta + \sum_{j=1}^{3} 1_{(k=j)} * X_{ifk}\lambda_j + \epsilon_{ifk}, \tag{2.5}$$

where z_f denotes household's non-pension income and the notation is otherwise unchanged.

Measurement errors in income could lead to downward biased estimate of the coefficients of non-pension income in this equation. Therefore, I have estimate equation 2.5 both by OLS and by 2SLS, using a series of indicator variables as instruments for income excluding pension.[16]

Results

Results of equation 2.4 and 2.5 are presented in table 8. Looking at all children together, we find that the effect of a Rand of pension income is very close to the effect of a Rand of labor income on the youngest children (once I correct of measurement errors in the 2SLS specification). The same result was found for the disposition of income by Case and Deaton (1995).

For boys, we find a larger effect of non pension income, although neither a Rand of pension income nor a Rand of non-pension income has a significant effect on the height of young boys. There is a positive association between non-pension income and the height of older boys. For girls we find that both non-pension income and pension income have an effect, but that the effect of a Rand of pension income is twice as large as the effect of a Rand of non-pension income (0.41 vs 0.20).

These results, even if they should be taken cautiously, are quite interesting. At least two interpretations are possible. First, if households are credit constrained, this can lead them to focus their investment on improving the health of boys, presumably because returns to the health of men (or boys) are higher (cf. Garg and Morduch (1996)). If the pension relaxes the credit constraint, girls will then benefit from this while boys will not, if investments in the health of boys were already optimal. An increase in permanent income can still lead to an improvement in the nutrition of boys to the extent that child health is a normal good. Non-pension income is auto-correlated, therefore, a Rand of non-pension income is more closely correlated to a Rand of permanent income than a Rand of pension income. This could be an explanation for why a Rand of pension income has less impact on the nutrition of boys than a Rand of non-pension income, while the opposite is true for girls.

Another interpretation is that the tastes of the elderly are different from the tastes

[16]Head of the household is employed, head is self-employed head holds a regular wage job, a casual wage job, a job in agriculture, sector of head's job, head works for the Government or an NGO, head works for a private firm, head is paid monthly, fortnightly, weekly.

of prime-age household members, and that elderly have a preference for their grandchildren health. This will be important if the household does not function as a single decision unit, but it is appropriate to consider it as a "collective" entity. In addition to being a additional Rand of income, the pension improves the bargaining position of the elderly and therefore leads to higher investment in child health. In the next subsection, I look into this issue, by examining whether the effect of the pension varies with the gender of the recipient.

2.5.2 Importance of recipient's gender

Three out of four pension recipients are women (they live longer and they are eligible earlier). The pension is therefore a transfer program biased in favor of women. It has been argued (Lunds (1993)) that it is a desirable feature, since income in the hands of women tend to be more strongly associated to "good" outcomes (heath, education, etc...) than income in the hand of men. However, this is based on evidence from elsewhere, which is moreover subject to the caveats mentioned in the introduction to this paper. The reform in the pension program is a unique occasion to examine whether pension income has a different impact in the hands of women than in the hand of men. The comparison is not affected by the two problems (endogeneity of income and functioning of the marriage market) that plague previous effort to establish similar results in the literature. Does it matter that pension income is received by a woman or a man? Or, in the reduced form, does the effect of eligibility change with the gender of the recipient?

Descriptive statistics and Non-parametric results

The descriptive statistics in table 2 indicate that this is the case. Children of both sexes are on average smaller if they live with an eligible man than with an eligible woman. Moreover, boys born before January 1992 have the same average height in both cases, whereas younger boys living with an eligible man are much smaller than other boys. For girls, the pattern is even more striking. Older girls do better when they live with an eligible man than if they are living with an eligible woman, while the opposite is true for younger girls.

Non-parametric regressions illustrate these differences. In figure 8, I show non-parametric regressions of height for age as a function of date of birth in households where there is an eligible woman, households where there is an eligible man, and households where there is no eligible member. [17] The relative positions of the curves for households where a woman is eligible and households where nobody is eligible are the same as in figure 1. Young children are doing better if they live with an eligible woman, whereas older children are doing worst. The relative positions of the curves for households where a man is eligible and households where nobody is eligible is totally different. At all ages (except for a pick for the children born in 1993, which is due to very few high values), children are doing better in households where nobody is eligible than in households where a man is eligible. Young children are substantially taller if they live with an eligible woman than if the live with an eligible man. These graphs suggest strongly that the pension is effective only in the hands of women.

In figure 9, I show a similar graph using the alternative instrument (old grand-parent alive). Older children do better if they have a old grand father alive than if they have an old grand-mother alive. This might reflect the fact that children who have an elderly male relative alive must come healthier families. However the gap closes among children born after January 1992, and the two curves cross for children born after July 1992. The conclusion is therefore again that having a grand mother alive lead the extension of the pension program to be beneficial for children, while having a grand-father alive did not.

Figure 13 shows the same graph, this time for weight for height. It is consistent with previous evidence: almost all children living with in an eligible woman are heavier than those living in an household where nobody is eligible while the opposite is true for children who live with an eligible man. This is not surprising, since weight for

[17] In some households where there is an eligible man, there is also an eligible woman. In this version of the graphs these households are included in both samples. This tends to attenuate any difference between the two regressions.

height reflects flux in nutrition. For all children, the presence of an eligible woman in the household ensures a better nutrition. This appears even though these households are more disadvantaged.

Differences in differences and 2SLS estimates

- **Statistical framework**

I extend the difference in differences formulation to take into account the gender of the eligible individual. I similarly extend the formulation using the alternative strategy (using the indicator for whether the child has a grand-parent alive) to take into account the gender of the grand-parent. This equation is similar in spirit to the formulation in Thomas (1990).

$$h_{ifk} = \alpha_w 1_{(k=1)} * TW_f + \alpha_m 1_{(k=1)} * TM_f + \beta_w TW_f + \beta_m TM_f + \sum_{j=1}^{3} \gamma_j 1_{(j=k)} + \epsilon_{ifk}, \quad (2.6)$$

where TW_f is equal to 1 if there is an eligible woman in the household (resp. if the child has an old grand-mother alive) and 0 otherwise and TM_f is equal to 1 if there is an eligible man in the household (resp. if the child has an old grand-father alive) and 0 otherwise.[18] A similar formulation is estimated with TW_f being equal to 1 if the child as a old grand mother alive and 0 otherwise and TM_f being equal to 1 if the child has a old grand-father alive (this instrument is defined exactly as before).

Likewise, the following equation is estimated using 2SLS.

$$h_{ifk} = \alpha_w 1_{(k=1)} * PW_f + \alpha_m 1_{(k=1)} * PM_f + \beta_w TW_f + \beta_m TM_f + \sum_{j=1}^{3} \gamma_j 1_{(j=k)} + \epsilon_{ifk}, \quad (2.7)$$

where PM_f is equal to 1 if a woman receives a pension in the household, and PW_f is equal to 1 if a man receives a pension in the household. The instruments for

[18] I have omitted family background variables in this notation, but I introduce them when I estimate the equation.

$1_{(k=1)} * PW_f$ and $1_{(k=1)} * PM_f$ are as before $1_{(k=1)} * TW_f$ and $1_{(k=1)} * TM_f$.

• **Results**

I present the estimates of equation 2.6 in table 9. For boys as well as girls, the difference in differences are positive for woman's eligibility or but negative (and not significant) for men's eligibility. The positive effects of woman's eligibility is larger on the height of girls than on that of boys, and it is significant only for girls. Using the indicator of whether the child has a grand-mother or a grand-father alive and old leads to the same conclusion. The effect of having a old grand mother alive is positive for both boys and girls, larger for girls than for boys, and significant only for girls. There is a small and insignificant positive effect on boys of having a male grand-parent alive, and a small and insignificant negative effect on girls.

This establishes clearly the result that a pension received by a woman is effective (especially on girls), while a pension received by a man is not. The 2SLS estimates are presented in table 5. The point estimates suggest that pension received by women lead to an increase in the height-for-age of boys by 0.37 (this is not significant) and that to that of girls by 0.83, or more than half of the mean value of the outcome for girls. In contrast, the estimates of the effect of a pension received by a man is negative and insignificant.

Interpretation

These results constitute an example of the differential impact of women's and men's income on child health which is not subject to the traditional caveats in this literature (measurement error, endogeneity of income and correlation between woman non-labor income and unobserved man's characteristics due to marriage).[19] However, there are still two interpretations possible. The first interpretation is that the household is

[19]Note that these results also indicates that the estimates for women are not likely to be driven by omitted reforms targeting the same group. I find no effect of pension when it is received by a men. Yet, the observed characteristics of households with eligible men are similar to those of households with eligible women (table 1, column 3 and 4). If anything, households where eligible men live have a lower pre-pension income, and are more likely to be rural. Therefore, if differences in differences reflected the effect of other programs targeting the same type of households, there would be no differences by gender of the potential recipient.

a collective entity, and that the same resources are spent differently when they are received by a woman and when they are received by a man. Another interpretation, however, could be that, in term of permanent income, a rand of pension income received by a man represents much less than a rand of pension income received by a woman, because men receive the pension for a shorter time. If household are credit constrained or have a very high discount rate, this should not lead to different effects of man's and woman's pensions. But in the opposite case, this effect could drive the difference that we observe here.

To help discriminate between these two interpretations, it is useful to look at the disposition of men's an women's pension income. If the household is a unitary entity, and if a man's pension income is not spent on child health because it is akin to transitory income, then we should see that the propensity to save out of men's income is much larger than the propensity to spend out of women's income (and non-pension income). I therefore estimate the following equation:

$$S_f = \alpha_w y_{fw} + \alpha_m y_{fm} + \alpha z_f + X_f \beta + \epsilon_f, \tag{2.8}$$

where S_f stands for the total savings of households (defined as total income minus expenditures), y_{fw} is pension income received by women, y_{mw} is pension income received by men, z_f is non pension income, and X_f is a set of control variables. This specification extends Case and Deaton (1996) formulation to take into account differences in the disposition of income received by men and women. The emphasis here is on the comparison between α_w and α_m. This equation is estimated by OLS, and 2SLS. The instruments in the 2SLS equations (y_{fm}, y_{fw} and z_f are instrumented) are the indicators for the presence of an eligible man and an eligible woman, and the instruments used to correct for measurement errors in non-pension income.[20]

Results are presented in table 10 (column 1 and 2). The point estimates suggest that propensity to save out of a man's pension income is actually *lower* than the propensity to save out of a woman's non-pension income (although this difference is

[20]These instruments are described above, as well as, at the end of the table.

not significant). This result indicates that the differences in the effects of woman's and man's pension income on child height is not likely to be due to their different life cycle properties. In combination with the results in the previous subsection, this result therefore suggests that the disposition of income is influenced by the gender of the recipient.

Some additional results are presented in table 10. I estimate specifications similar to equation 2.8, but using as dependent variables respectively expenditures on food, expenditure on alcohol, tobacco, entertainment and press, share of total expenditure on food, and share of total expenditures on alcohol, tobacco, etc... The propensity to spend on food with respect to man's and woman's pension income are similar. The propensity to spend on alcohol and other personal goods out of man's pension income is however larger than the propensity to spend on the same goods out non-pension income. Together, the results therefore confirms that the disposition of resources is influenced by the member of the household who receives these resources.

Among the two interpretations proposed earlier in this section, it seems therefore that the first is the correct one: pension income affects child health more if it is received by a woman than if it is received by a man because who in the household helds the resources matters for their disposition. Woman's preferences lead them to want to invest more in their children than man do, and these preferences are taken into account if they are the income earner.

2.6 Conclusion

The extension of the Old Age Pension program in South Africa has led to an improvement in the health and nutrition of children, especially girls. This is reflected in the height for age of the youngest children. Non-parametric analysis and more restrictive difference in differences estimates lead to the same conclusion. This effect is entirely due to pension received by women. I estimate that a pension received by women improved the height for age Z-scores of girls by 0.81 of a standard deviation and that of boys by 0.31 of a standard deviation. South African children are on average -1.28

standard deviations smaller than American children, so this is a large increase. This is due to the fact that the pension benefits are generous, but also that the effect of each Rand of pension is at least as large as the effect of a Rand of non-pension income. In contrast, pension received by men have no effect on the height of children.

The findings reported here are important because they show the effect of an exogenous increase in income on child health in developing countries. This question is usually difficult to address convincingly. Furthermore, this study also provides a clear example of the difference in the effects of income in the hands of men and in the hands of women, which, unlike in other studies, cannot easily be explained by endogeneity or omitted variable biases. This indicates that it is important to model expenditures allocation within the household in a collective setting. Child health alone does not allow us to distinguish between various alternative models of the household, but in future work, it would be extremely interesting to implement the tests proposed in the literature on collective model of the households.

Moreover, these findings have immediate policy implications, for South Africa as well as for other developing countries. The first implication is that direct income transfers to poor households can contribute to increase human capital. Deciding whether such transfers should be implemented require of course to compare their cost effectiveness with that of other potential measures. But these results show that even temporary grants (a child grant for some years for example) can have important effects on child health. The second implication is that the identity of the transfer recipient matters. In South Africa, the program is naturally biased toward women, both because men can in principle claim the pension only after 65 and because women tend to live longer. Without this feature, the program would not benefit children as much. The distinction between men and women is not in accordance with the South African constitution, and there is some pressure to remove it. The effectiveness of the pension program as a tool to transfer resources to children would suggest moving in the opposite direction. Here again, the pension is an instrument to achieve several objectives (including the political objective of compensating elderly South African for the deprivations of their lives under Apartheid) which must be pondered when deciding what to do. In South Africa, the pension program and a child grant program are often thought of as partial substitute. The results in this paper suggest that the pension achieves some of the objectives of a child grant. Means testing is usually impractical in developing countries. In South Africa, targeting elderly people is a simple way to reach poor children disproportionately. In this perspective, the scheme has however an important drawback, which is that the pension is tied to the elderly, not to the needs of the child. The grand-mothers of some very poor children might be dead, which prevent them to benefit from the program. Knowing that the pension is spent partly toward children therefore answers only half the question. We need to compare the effectiveness of the pension and the effectiveness of a child grant in reaching poor children and, once they are reach, in improving their health. The finding that the identity of the recipient matters imply that the answer to these questions for the child grant are far from obvious. Future work needs to establish these results. The South African government has launched in 1995 a child grant program. Comparing the effects of this program with the effects of the pension reform should be extremely interesting in this perspective.

Bibliography

Becker, Garry (1965) 'A theory of the allocation of time.' *Economic Journal* 75, 493–517

Bourguignon, Francois, Martin Browning, Pierre Andre Chiappori, and Valerie Lechene (1993) 'Efficient intra-household allocations: A general characterization and empirical tests.' *Annales d'Economie et Statistiques* pp. 137–156

—— (1994) 'Income and outcomes: A structural model of intrahousehold allocations.' *Journal of Political Economy* 102, 1067–1096

Browning, Martin, and Pierre Andre Chiappori (1998) 'Efficient intra-household allocations: A general characterization and empirical tests.' *Econometrica* 66(6), 1241–1278

Case, Anne, and Angus Deaton (1996) 'Large cash transfers to the elderly in South Africa.' Working Paper 5572, National Bureau of Economic Research

Chiappori, Pierre Andre (1988) 'Rational household labor supply.' *Econometrica* 56(1), 63–90

—— (1992) 'Collective labor supply and welfare.' *Journal of Political Economy* 100(3), 437–467

Garg, Ashish, and Jonathan Morduch (1996) 'Sibling rivalry and the theory of the household.' Mimeo

Le Roux, Pieter (1995) 'Poverty, social policies and the reconstruction and development programme.' Working Paper, Institute for Theological and Interdisciplinary Research

Lunds, Frances (1992) 'The way welfare works: Structure, spending, staffing and social work in the south african welfare bureaucracies.' Working Paper, Cooperative HSRC Programme

____ (1993) 'State social benefits in South Africa.' *International Social Security Review* 46(1), 5–25

Martorell, R., and J.P. Habicht 'Growth in early childhood in developing countries.' In *Human Growth: A Comprehensive Treatise,* ed. F. Falkner and J.M. Tanner, 2 ed., vol. 3 (New York: Plenum Press)

Mayer, Susan E. (1997) *What Money Can't buy: Family Income and Children's Life Chances* (Cambridge: Harvard University Press)

Paxson, Christina H. (1992) 'Using weather variability to estimatet the response of savings to transitory income in thailand.' *American Economic Review* 82(1), 15–33

Rosenzweig (1988) 'Labor markets in low income countries.' In *Handbook of Development Economics,* ed. H. Chenery and T.N. Srinivasan, vol. 3 (Amsterdam: North Holland)

Rosenzweig, Mark (1999) 'Missing woman, the marriage market, and economic growth.' Mimeo, University of Pennsylvania

Sahn, D. (1990) 'Malnutrition in Cote d'Ivoire.' Social Dimensions of Adjustment Working Paper 4, The World Bank

Schultz, Theodore Paul (1990) 'Testing the neoclassical model of fanily labor supply and fertility.' *Journal of Human Resources* 25(4), 599–634

—— (1996) 'Wage rentals for reproducible human capital: Evidence from two West African countries.' Mimeo, Yale University

Shea, John (1997) 'Does parents' money matter?' Working Paper 6026, National Bureau of Economic Research

Strauss, John (1986) 'Does better nutrition raise farm productivity.' *Journal of Political Economy* 94(2), 297–320

Strauss, John, and Duncan Thomas (1995) 'Human resources: Empirical modeling of household and family decisions.' In *Handbook of Development Economics*, ed. Jere Behrman and T.N. Srinivasan, vol. 3 (Amsterdam: North Holland) chapter 34, pp. 1885–2023

Thomas, Duncan (1990) 'Intra-household resources allocation: an inferential approach.' *Journal of Human Resources* 25(4), 635–664

—— (1994a) 'Like father, like son, like mother, like daughter: Parental education and child health.' *Journal of Human Resources* 29(4), 950–988

—— (1994b) 'The distribution of of income and expenditures within the household.' *Annales d'Economie et Statistiques* 29, 109–136

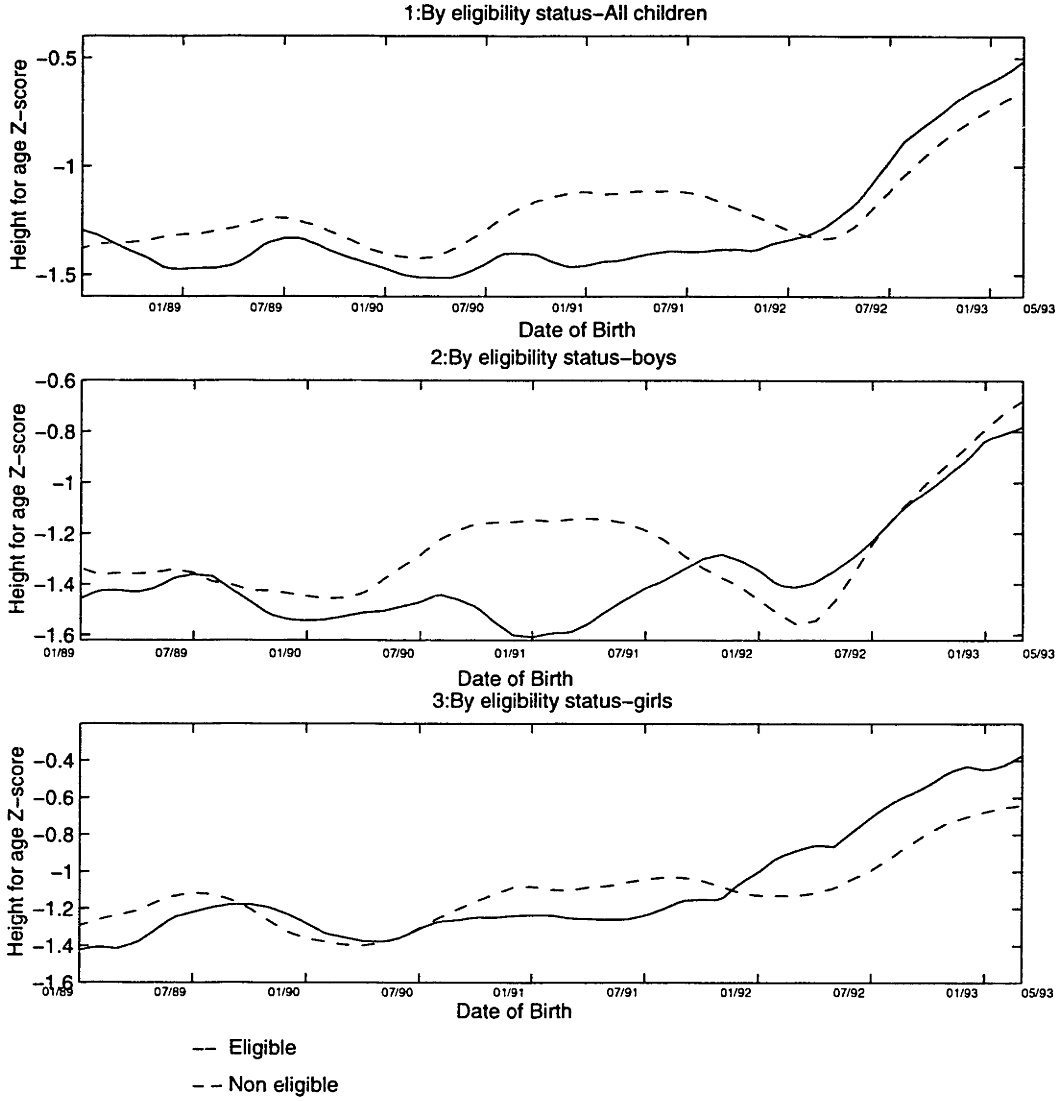
Thomas, Duncan, John Strauss, and M.-H. Henriques (1990a) 'Child survival, height for age and household characteristics in brazil.' *Journal of Development Economics* 33(2), 197–234

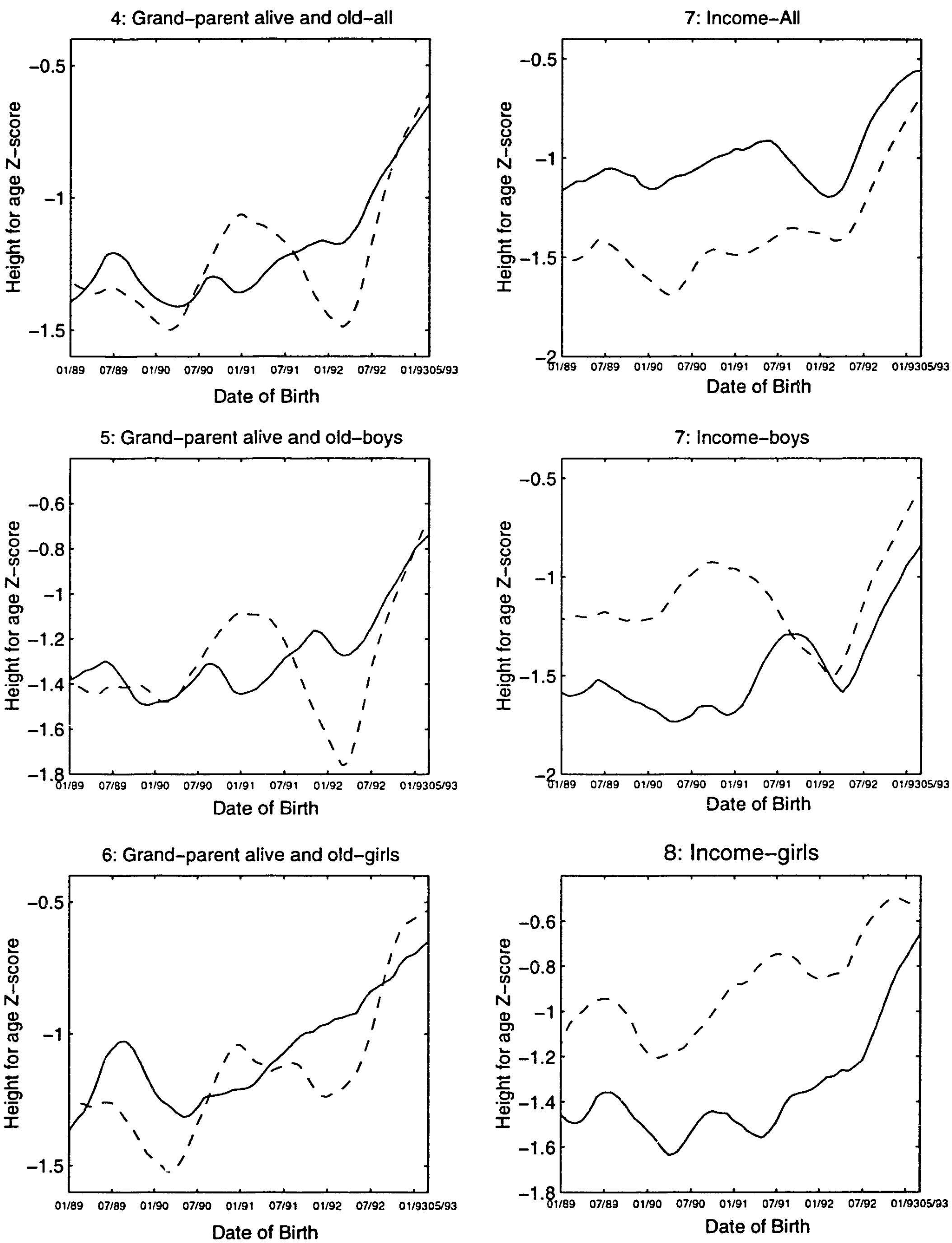
—— (1990b) 'How does mother's nutrition affects child health.' *Journal of Human Resources* 26(2), 183–211

Thomas, Duncan, Victor Lavy, and John Strauss (1992) 'Public policy and anthropometric outcomes in the Cote d'Ivoire.' Living Standard Measurement Surveys Working Paper 89, The World Bank

Van der Berg, Servaas (1994) 'Issues in South African social security.' World Bank-IFC-MIGA Office Memorandum

WHO, Working Group (1986) 'Use and interpretation of anthropometric indicators of nutritional status.' *Bulletin of the World Health Organization* 64(6), 929–941


1:By eligibility status-All children
Height for age Z-score
-0.5
-1
-1.5
01/89
07/89
01/90
07/90
01/91
07/91
01/92
07/92
01/93
05/93
Date of Birth
2:By eligibility status-boys
Height for age Z-score
-0.6
-0.8
-1
-1.2
-1.4
-1.6
01/89
07/89
01/90
07/90
01/91
07/91
01/92
07/92
01/93
05/93
Date of Birth
3:By eligibility status-girls
Height for age Z-score
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
01/89
07/89
01/90
07/90
01/91
07/91
01/92
07/92
01/93
05/93
Date of Birth
— Eligible
- - Non eligible

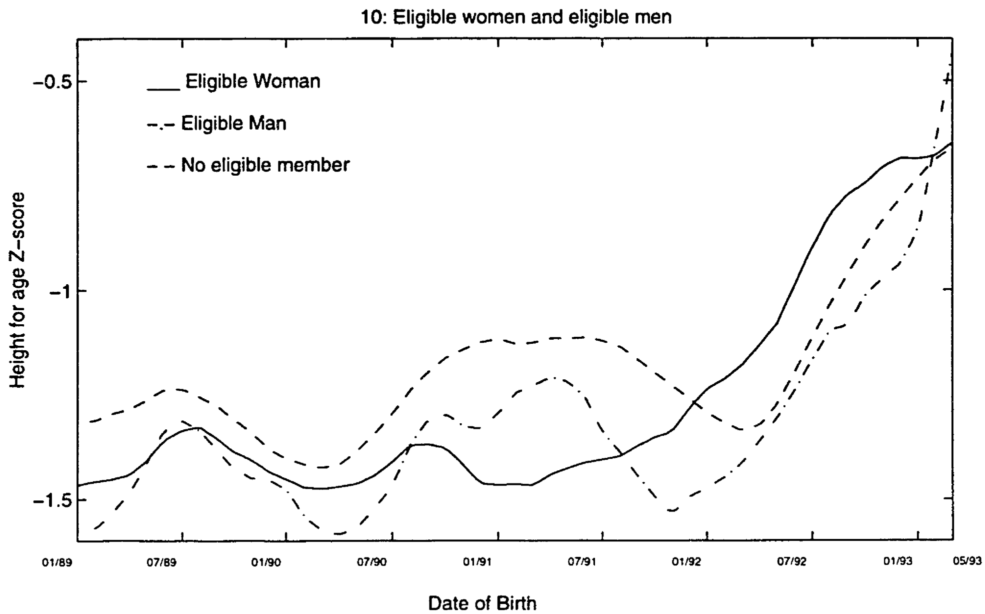


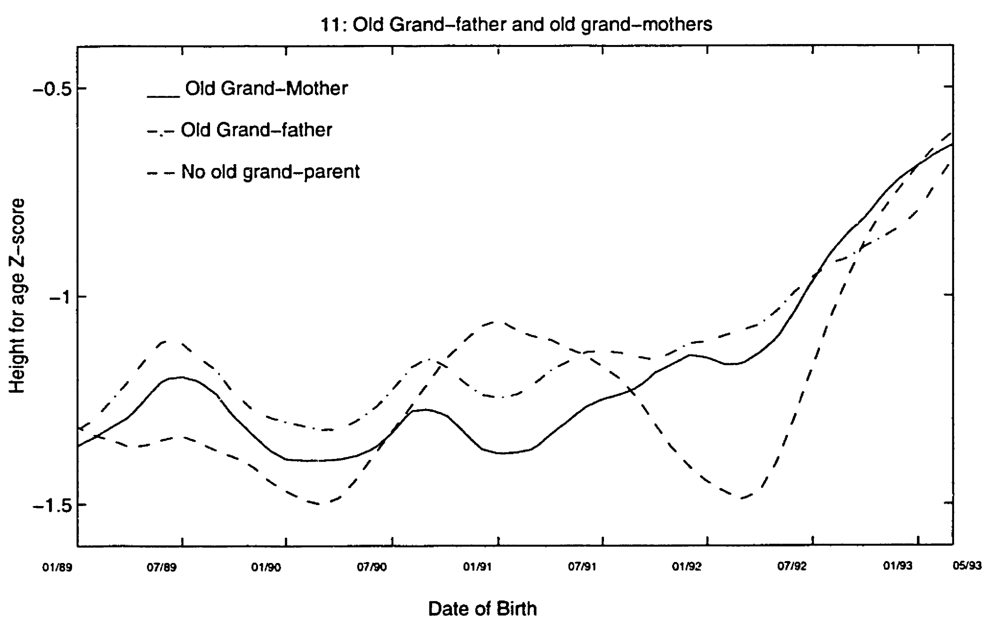
-- Old grand-parent

- - No old grand-parent

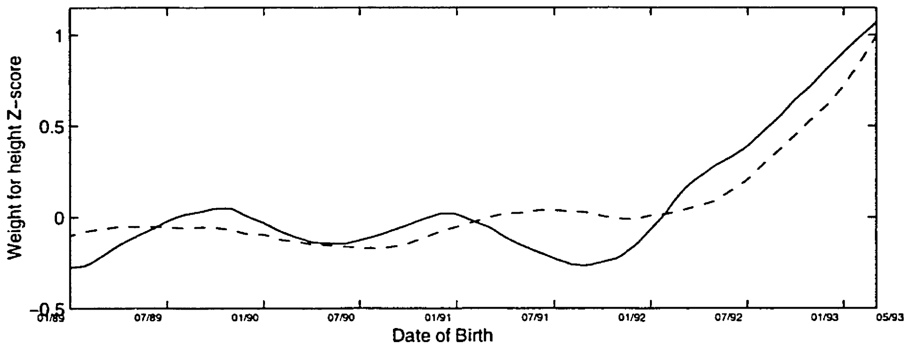
-- Income<median

- - Income>median

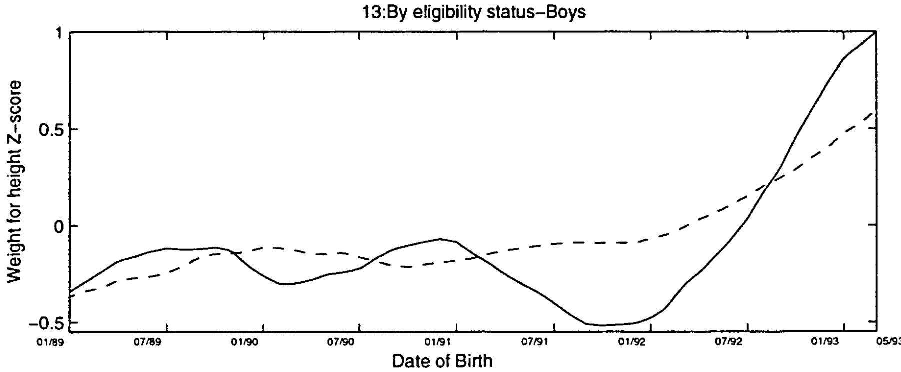

10: Eligible women and eligible men
Eligible Woman
-.- Eligible Man
- - No eligible member
Height for age Z-score
-0.5
-1
-1.5
01/89
07/89
01/90
07/90
01/91
07/91
01/92
07/92
01/93
05/93
Date of Birth


11: Old Grand-father and old grand-mothers
Old Grand-Mother
-.- Old Grand-father
- - No old grand-parent
Height for age Z-score
-0.5
-1
-1.5
01/89
07/89
01/90
07/90
01/91
07/91
01/92
07/92
01/93
05/93
Date of Birth

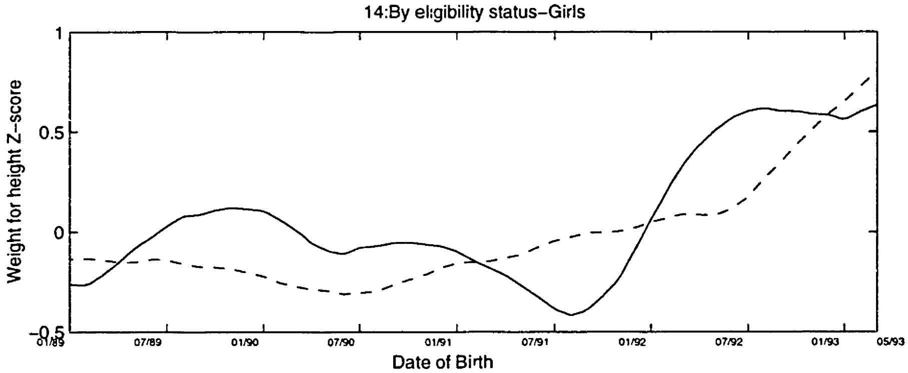
12:By eligibility status-All



13:By eligibility status-Boys



14:By eligibility status-Girls



–– Eligible

– – Non eligible

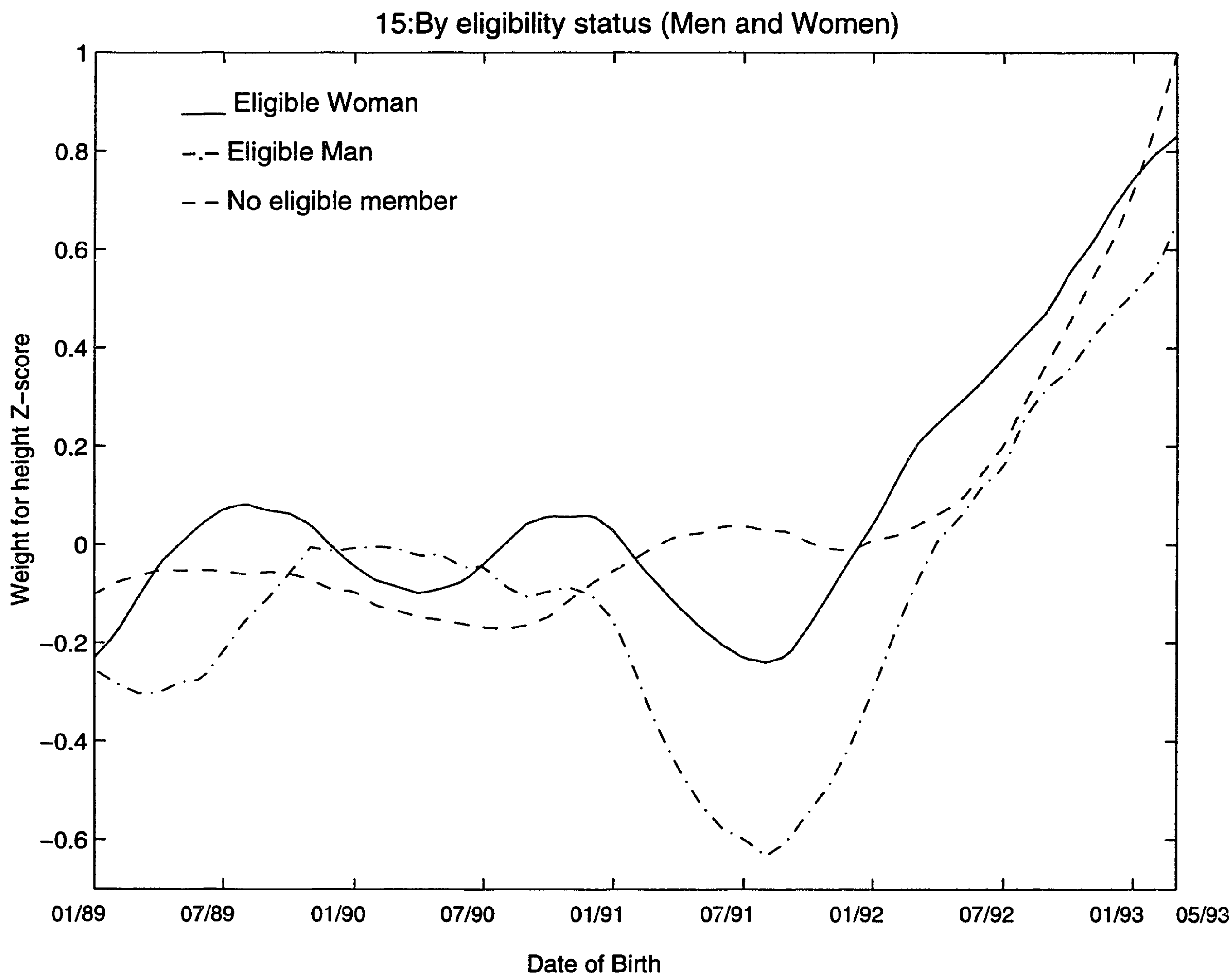

15:By eligibility status (Men and Women)
Eligible Woman
-.- Eligible Man
- - No eligible member
Weight for height Z-score
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
01/89
07/89
01/90
07/90
01/91
07/91
01/92
07/92
01/93
05/93
Date of Birth

Table 1: Descriptive statistics
(standard errors in parenthesis)

	Receipt of pension		Eligibility for pension			Grand parent alive and old	
	Yes	No	Woman	Man	None	Yes	No
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Number of children	1096	2952	1032	373	3183	2296	2096
Mother's education	5.70	5.18	5.72	5.75	5.17	5.06	5.63
	(0.14)	(0.083)	(0.15)	(0.34)	(0.084)	(0.10)	(0.11)
Father's education	4.90	4.51	5.1	4.14	4.52	4.59	4.73
	(0.25)	(0.11)	(0.27)	(0.44)	(0.11)	(0.13)	(0.16)
Household head's	1.84	3.86	1.98	1.62	3.83	3.00	3.71
education	(0.11)	(0.083)	(0.13)	(0.20)	(0.083)	(0.098)	(0.10)
Mother's age	28	30	28	28	30	31	27
	(0.34)	(0.19)	(0.37)	(0.58)	(0.19)	(0.24)	(0.24)
Father's age	36	37	35	41	37	39	33
	(0.87)	(0.37)	(0.78)	(2.04)	(0.37)	(0.42)	(0.59)
Mother absent	0.17	0.077	0.18	0.14	0.076	0.098	0.11
	(0.014)	(0.0058)	(0.015)	(0.022)	(0.0058)	(0.0079)	(0.0088)
Father absent	0.64	0.42	0.67	0.66	0.41	0.41	0.55
	(0.018)	(0.012)	(0.019)	(0.032)	(0.011)	(0.014)	(0.015)
Grand parent in in the household	0.93	0.42	0.96	0.89	0.41	0.56	0.56
	(0.0098)	(0.011)	(0.0079)	(0.021)	(0.012)	(0.014)	(0.015)
Grand parent alive and old	0.86	0.41	1	1	0.35	1	0
	(0.014)	(0.011)			(0.011)		
Eligible for pension	0.83	0.076	1	1	0	0.57	0
	(0.015)	(0.0057)				(0.020)	
Woman eligible for pension	0.73	0.060	1	0.55	0	0.4	0
	(0.018)	(0.0057)		(036)		(0.014)	
Man eligible for pension	0.26	0.025	0.20	1	0	0.14	0
	(0.018)	(0.0038)	(0.018)			(0.0099)	
Household receives pension	1	0	0.82	0.79	0.065	0.46	0.078
			(0.017)	(0.029)	(0.0059)	(0.014)	(0.0083)
Woman receives pension	0.83	0	0.79	0.48	0.042	0.40	0.049
	(0.015)		(0.018)	(0.036)	(0.0049)	(0.014)	(0.0067)
Man receives pension	0.32	0	0.17	0.68	0.028	0.14	0.047
	(0.019)		(0.016)	(0.034)	(0.0040)	(0.010)	(0.0058)
Average amount received (pension-monthly)		0	324	389	22.5	181	27.5
			(9.57)	(19.7)	(2.17)	(6.59)	(3.10)
Average amount received (conditionnal on receiving)	384		394	491	344	390	355.3
	(6.67)		(8.44)	(17.0)	(10.7)	(7.56)	(13.0)
Non-pension income (monthly)	661	930	723	644	909	814	887
	(32.6)	(21.9)	(35.4)	(49.4)	(22.1)	(23.8)	(29.2)
Household size	10	7.86	10.5	10.5	7.67	8.75	7.74
	(0.17)	(0.092)	(0.21)	(0.29)	(0.085)	(0.13)	(0.11)
Total household income per capita (monthly)	122	147	120	122	149	144	142
	(3.84)	(3.80)	(4.45)	(7.07)	(3.76)	(4.20)	(4.56)

Notes: Household averages weigthed by the number of children in each household (multiplied by the survey weights)

Table 2: Descriptive statistics
Height for age and weight for height

	Receipt of pension		Eligibility for pension			Grand parent alive and old	
	Yes	No	Woman	Man	None	Yes	No
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: HEIGHT FOR AGE Z-SCORE							
All children	-1.41	-1.21	-1.38	-1.45	-1.22	-1.29	-1.25
	(0.066)	(0.035)	(0.071)	(0.13)	(0.035)	(0.045)	(0.044)
All boys	-1.51	-1.27	-1.50	-1.67	-1.27	-1.39	-1.27
	(0.087)	(0.048)	(0.093)	(0.16)	(0.048)	(0.061)	(0.058)
All girls	-1.31	-1.15	-1.24	-1.19	-1.17	-1.18	-1.21
	(0.10)	(0.052)	(0.11)	(0.20)	(0.053)	(0.066)	(0.066)
Boys born 01/92 or later	-1.30	-1.08	-1.20	-1.8	-1.09	-1.11	-1.16
	(0.17)	(0.10)	(0.19)	(0.37)	(0.10)	(0.13)	(0.12)
Boys born before 01/92	-1.59	-1.34	-1.62	-1.63	-1.33	-1.49	-1.32
	(0.10)	(0.053)	(0.11)	(0.17)	(0.053)	(0.068)	(0.064)
Girls born 01/92 or later	-0.85	-0.88	-0.62	-0.85	-0.94	-0.75	-0.99
	(0.21)	(0.11)	(0.22)	(0.41)	(0.11)	(0.14)	(0.13)
Girls born before 01/92	-1.48	-1.26	-1.47	-1.29	-1.26	-1.34	-1.30
	(0.11)	(0.059)	(0.12)	(0.045)	(0.059)	(0.073)	(0.076)
PANEL B: WEIGHT FOR AGE Z-SCORE							
All children	0.17	0.16	0.28	0.15	0.15	0.14	0.20
	(0.07)	(0.04)	(0.08)	(0.14)	(0.04)	(0.05)	(0.05)
All boys	0.11	0.18	0.22	0.09	0.15	0.13	0.19
	(0.11)	(0.06)	(0.11)	(0.19)	(0.10)	(0.07)	(0.08)
All girls	0.23	0.15	0.34	0.21	0.14	0.14	0.2
	(0.10)	(0.06)	(0.11)	(0.23)	(0.06)	(0.07)	(0.07)
Boys born 01/92 or later	0.44	0.51	0.54	0.70	0.47	0.5	0.49
	(0.22)	(0.12)	(0.23)	(0.44)	(0.12)	(0.16)	(0.14)
Girls born 01/92 or later	0.65	0.56	0.87	1.11	0.53	0.58	0.58
	(0.22)	(0.11)	(0.23)	(0.60)	(0.11)	(0.15)	(0.13)
Girls born before 01/92	-0.01	0.04	0.10	-0.11	0.03	0.00	0.06
	(0.12)	(0.07)	(0.12)	(0.20)	(0.07)	(0.08)	(0.09)
Girls born 01/92 or later	0.08	-0.02	0.14	-0.12	-0.02	-0.01	0.04
	(0.11)	(0.07)	(0.12)	(0.22)	(0.07)	(0.08)	(0.08)

Table 3: OLS regressions. Impact of treatment variables on height for age

	Treatment=Eligible for pension			Treatment=Receive a pension			Treatment =Grand parent alive and old		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: All children									
Treatment*born after 01/92	0.28	0.27	0.28	0.13	0.13	0.13	0.25	0.27	0.31
	(0.19)	(0.18)	(0.18)	(0.18)	(0.18)	(0.18)	(0.15)	(0.15)	(0.17)
Treatment	-0.24	-0.22	-0.22	-0.23	-0.2	-0.2	-0.098	-0.17	-0.19
	(0.088)	(0.088)	(0.089)	(0.082)	(0.090)	(0.090)	(0.073)	(0.078)	(0.080)
Born after 01/92	0.26	0.27	0.21	0.35	0.32	0.23	0.21	0.20	0.22
	(0.091)	(0.092)	(0.28)	(0.089)	(0.093)	(0.28)	(0.11)	(0.11)	(0.28)
PANEL B: Boys									
Treatment*born after 01/92	0.12	0.11	0.091	0.035	0.035	0.027	0.22	0.26	0.25
	(0.24)	(0.25)	(0.24)	(0.24)	(0.24)	(0.24)	(0.20)	(0.20)	(0.22)
Treatment	-0.28	-0.25	-0.26	-0.26	-0.24	-0.25	-0.17	-0.26	-0.27
	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)	(0.12)	(0.096)	(0.10)	(0.10)
Born after 01/92	0.22	0.22	-0.11	0.24	0.24	-0.099	0.13	0.12	-0.14
	(0.12)	(0.12)	(0.33)	(0.12)	(0.13)	(0.33)	(0.14)	(0.14)	(0.33)
PANEL C: Girls									
Treatment*born after 01/92	0.49	0.47	0.51	0.25	0.21	0.21	0.27	0.29	0.38
	(0.27)	(0.26)	(0.26)	(0.27)	(0.26)	(0.26)	(0.22)	(0.22)	(0.25)
Treatment	-0.21	-0.19	-0.19	-0.21	-0.17	-0.17	-0.026	-0.09	-0.11
	(0.12)	(0.13)	(0.12)	(0.13)	(0.13)	(0.13)	(0.11)	(0.12)	(0.12)
Born after 01/92	0.3	0.31	0.50	0.36	0.39	0.54	0.29	0.29	0.52
	(0.13)	(0.14)	(0.46)	(0.13)	(0.14)	(0.46)	(0.22)	(0.17)	(0.47)
Covariates:									
Year of birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family background variables	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Family background variables *Yob. Dummies	No	No	Yes	No	No	Yes	No	No	Yes

Note: Family background variables: father's age and education, mother's age and education, rural or metro residence.
Standard errors (robust to correlation of residuals within households) in parentheses.

Table 4: First stage regressions

	Household receives pension*born after01/92			Log(pension+1) *born after01/92			Born after 01/92	
							*Woman receives pension	* Man receives Pension
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PANEL A: All children								
Household eligible	0.74			1.15				
*born after 01/92	(0.030)			(0.050)				
Grand parent alive and old		0.39	0.54		0.60	0.83		
*born after 01/92		(0.027)	(0.026)		(0.042)	(0.048)		
Woman receives pension							0.73	0.036
*born after 01/92							(0.034)	(0.023)
Man receives pension							0.018	0.63
*born after 01/92							(0.056)	(0.056)
PANEL B: Boys								
Household eligible	0.75			1.14				
*born after 01/92	(0.037)			(0.061)				
Grand parent alive and old		0.43	0.55		0.64	0.84		
*born after 01/92		(0.036)	(0.037)		(0.056)	(0.061)		
Woman receives pension							0.74	0.043
*born after 01/92							(0.040)	(0.034)
Man receives pension							-0.041	0.61
*born after 01/92							(0.053)	(0.063)
PANEL B: Girls								
Household eligible	0.74			1.16				
*born after 01/92	(0.043)			(0.071)				
Grand parent alive and old		0.35	0.53		0.56	0.83		
*born after 01/92		(0.037)	(0.044)		(0.058)	(0.071)		
Woman receives pension							0.69	0.026
*born after 01/92							(0.054)	(0.030)
Man receives pension							0.11	0.65
*born after 01/92							(0.097)	(0.089)

Notes:Main effect of the instrument+ Full set of covariates also included in all regressions except in column (2) and (5) (no interactions)

Table 5: 2SLS estimates. Effects of Pension receipt on height for age.

	Instruments set								
	Eligible for a pension*born after 01/92			Gd. parent alive and old*born after 01/92			Woman eligible*born after 01/92, Man eligible*born after 01/92		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: All children									
Household receives pension	0.38	0.36	0.38	0.66	0.69	0.58			
*born after 01/92	(0.25)	(0.25)	(0.25)	(0.40)	(0.39)	(0.32)			
Woman receives pension							0.62	0.58	0.59
*born after 01/92							(0.29)	(0.28)	(0.29)
Man receives pension							-0.58	-0.52	-0.51
*born after 01/92							(0.47)	(0.47)	(0.47)
PANEL B: Boys									
Receive pension	0.16	0.15	0.12	0.53	0.60	0.46			
*born after 01/92	(0.32)	(0.32)	(0.33)	(0.48)	(0.48)	(0.41)			
Woman receives pension							0.41	0.40	0.37
*born after 01/92							(0.37)	(0.37)	(0.38)
Man receives pension							-0.75	-0.76	-0.73
*born after 01/92							(0.60)	(0.60)	(0.61)
PANEL C: Girls									
Receive pension	0.66	0.66	0.69	0.79	0.81	0.72			
*born after 01/92	(0.37)	(0.37)	(0.36)	(0.66)	(0.64)	(0.49)			
Woman receives pension							0.86	0.78	0.83
*born after 01/92							(0.43)	(0.42)	(0.43)
Man receives pension							-0.41	-0.26	-0.24
*born after 01/92							(0.77)	(0.76)	(0.77)
Covariates:									
Eligible for pension	Yes	Yes	Yes	No	No	No	No	No	No
Grand-parent alive &old	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Year of birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family background variables	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
Family background variables *Yob. Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Standard errors (robust to correlation within the household) are presented in parentheses

Table 6: Control experiment. (1)

OLS regressions: Impact of treatment variables on weight for height

	Treatment=Eligible for pension			Treatment: Receives pension			Treatment =Grand parent alive and old		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: All children (N=3467)									
Treatment*born after 01/92	0.087	0.099	0.12	-0.008	-0.035	0.0079	0.054	0.066	0.077
	(0.21)	(0.21)	(0.21)	(0.20)	(0.21)	(0.098)	(0.16)	(0.16)	(0.19)
Treatment	0.044	0.022	0.018	0.008	0.011	-0.019	-0.060	-0.061	-0.058
	(0.098)	(0.10)	(0.10)	(0.095)	(0.10)	(0.19)	(0.088)	(0.096)	(0.097)
Born after 01/92	0.53	0.52	0.59	0.54	0.56	0.62	0.52	0.51	0.59
	(0.11)	(0.11)	(0.34)	(0.10)	(0. 11)	(0.34)	(0.13)	(0.13)	(0.34)
PANEL B: Boys (N=1790)									
Treatment*born after 01/92	0.10	0.10	0.13	-0.035	-0.049	-0.037	0.051	0.066	0.13
	(0.29)	(0.29)	(0.29)	(0.29)	(0.30)	(0.30)	(0.24)	(0.24)	(0.28)
Treatment	-0.0023	-0.052	-0.057	-0.03	-0.066	-0.066	-0.054	-0.092	-0.1
	(0.14)	(0.14)	(0.15)	(0.15)	(0.15)	(0.15)	(0.13)	(0.14)	(0.14)
Born after 01/92	0.54	0.54	0.82	0.57	0.58	0.86	0.54	0.53	0.82
	(0.15)	(0.15)	(0.42)	(0.15)	(0.15)	(0.42)	(0.17)	(0.17)	(0.42)
PANEL C: Girls (N=1677)									
Treatment*born after 01/92	0.076	0.090	0.12	-0.032	-0.046	-0.038	0.053	0.060	-0.020
	(0.30)	(0.30)	(0.30)	(0.29)	(0.30)	(0.30)	(0.24)	(0.24)	(0.25)
Treatment	0.095	0.10	0.097	0.10	0.099	0.098	-0.065	-0.031	0.0061
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.12)	(0.13)	(0.13)
Born after 01/92	0.52	0.51	0.29	0.55	0.54	0.31	0.51	0.50	0.29
	(0.16)	(0.16)	(0.53)	(0.15)	(0.15)	(0.53)	(0.18)	(0.18)	(0.52)
Covariates:									
Year of birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family background variables	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Family background variables *Yob. Dummies	No	No	Yes	No	No	Yes	No	No	Yes

Note: Standard errors (robust to auto-correlation of residual within households) in parentheses

Table 7: Control experiment (2).
Effect of the treament on height for age
(Weight, Coloured and Indian)

	Treatment=Eligible for pension			Treatment=Receive a pension			Treatment =Grand parent alive and old		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
All non-african children (N=631)									
Treatment*born after 01/92	-0.24	-0.25	-0.34	0.19	0.065	0.08	-0.51	-0.54	-0.53
	(0.48)	(0.47)	(0.51)	(0.56)	(0.57)	(0.59)	(0.34)	(0.34)	(0.54)
Treatment	-0.29	-0.26	-0.24	-0.66	-0.57	-0.54	-0.11	-0.19	-0.17
	(0.24)	(0.23)	(0.24)	(0.29)	(0.29)	(0.30)	(0.15)	(0.20)	(0.22)
Born after 01/92	0.41	0.37	0.48	0.37	0.34	0.37	0.67	0.64	0.49
	(0.20)	(0.21)	(0.60)	(0.19)	(0.21)	(0.60)	(0.30)	(0.31)	(0.59)
Covariates:									
Year of birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family background variables	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Family background variables *Yob. Dummies	No	No	Yes	No	No	Yes	No	No	Yes

Table 8: Effect of pension and non pension income on height for age.
OLS and 2SLS estimates

	Non-pension income						Pension income		
	OLS			2SLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: All children									
Log(pension income+1)							0.24	0.23	0.20
*born after 01.92							(0.16)	(0.16)	(0.16)
Log(Non pension income+1)	0.17	0.10	0.093	0.32	0.27	0.22			
*born after 01.92	(0.085)	(0.087)	(0.089)	(0.15)	(0.16)	(0.17)			
Log(Non pension income+1)	0.22	0.18	0.18	0.39	0.33	0.34			
*born before 01.92	(0.035)	(0.036)	(0.036)	(0.09)	(0.098)	(0.099)			
Panel B: Boys									
Log(pension income+1)							0.11	0.096	0.079
*born after 01.92							(0.21)	(0.21)	(0.22)
Log(Non pension income+1)	0.093	0.037	-0.011	0.26	0.22	0.18			
*born after 01.92	(0.12)	(0.12)	(0.13)	(0.21)	(0.22)	(0.24)			
Log(Non pension income+1)	0.22	0.18	0.18	0.39	0.32	0.35			
*born before 01.92	(0.045)	(0.047)	(0.048)	(0.11)	(0.13)	(0.13)			
Panel C: Girls									
Log(pension income+1)							0.42	0.40	0.41
*born after 01.92							(0.24)	(0.23)	(0.23)
Log(Non pension income+1)	0.27	0.19	0.24	0.30	0.26	0.20			
*born after 01.92	(0.12)	(0.12)	(0.12)	(0.21)	(0.22)	(0.12)			
Log(Non pension income+1)	0.23	0.18	0.18	0.38	0.33	0.18			
*born before 01.92	(0.053)	(0.054)	(0.055)	(0.12)	(0.13)	(0.054)			
Covariates:									
Year of birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family background variables	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Family background variables *Yob. Dummies	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Standard errors (robust to correlation within households) in parentheses
Instruments in column 4 to 6. Dummies for: head is employed, head holds a regular job, a casual wage job, a job an agricultue, sector of the job employer's type (central or local government, private firm, other), pay type (weekly, fortnightly, monthly)
Instrument in column 7 to 9: eligible*born after 01/92.

Table 9: OLS regressions
Effects of women's and men's eligibility on height for age

	Treatment=eligibility				Treatment=grand parent alive and old		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: All children							
woman treated	0.43	0.40	0.41	0.40	0.31	0.31	0.33
*born after 01/92	(0.20)	(0.20)	(0.20)	(0.21)	(0.16)	(0.16)	(0.17)
man treated	-0.098	-0.31	-0.31	-0.43	-0.069	-0.042	-0.018
*born after 01/92	(0.15)	(0.30)	(0.30)	(0.33)	(0.19)	(0.17)	(0.22)
woman treated	-0.23	-0.22	-0.22	-0.18	-0.10	-0.15	-0.15
	(0.094)	(0.096)	(0.096)	(0.099)	(0.076)	(0.079)	(0.080)
man treated	-0.098	-0.065	-0.064	-0.093	0.095	0.033	0.031
	(0.15)	(0.15)	(0.15)	(0.016)	(0.090)	(0.094)	(0.095)
Born after 01/92	0.27	0.27	0.23	0.24	0.22	0.22	0.23
	(0.091)	(0.092)	(0.28)	(0.28)	(0.10)	(0.10)	(0.28)
PANEL B: Boys							
woman treated	0.27	0.26	0.24	0.25	0.17	0.19	0.17
*born after 01/92	(0.27)	(0.27)	(0.27)	(0.29)	(0.22)	(0.22)	(0.24)
man treated	-0.48	-0.49	-0.46	-0.58	0.095	0.11	0.11
*born after 01/92	(0.37)	(0.37)	(0.38)	(0.45)	(0.27)	(0.28)	(0.29)
woman treated	-0.26	-0.26	-0.27	-0.25	-0.12	-0.19	-0.19
	(0.13)	(0.13)	(0.13)	(0.13)	(0.10)	(0.11)	(0.11)
man treated	-0.15	-0.077	-0.077	-0.064	0.059	0.023	0.018
	(0.17)	(0.18)	(0.18)	(0.19)	(0.12)	(0.12)	(0.12)
Born after 01/92	0.23	0.23	-0.072	-0.065	0.16	0.15	-0.11
	(0.12)	(0.12)	(0.33)	(0.33))	(0.14)	(0.14)	(0.33)
PANEL C: Girls							
woman treated	0.58	0.53	0.56	0.51	0.46	0.44	0.48
*born after 01/92	(0.29)	(0.28)	(0.28)	(0.29)	(0.24)	(0.24)	(0.24)
man treated	-0.039	-0.083	-0.064	-0.13	-0.25	-0.20	-0.14
*born after 01/92	(0.23)	(0.49)	(0.50)	(0.54)	(0.28)	(0.28)	(0.33)
woman treated	-0.20	-0.17	-0.17	-0.097	-0.087	-0.10	-0.11
	(0.13)	(0.14)	(0.14)	(0.14)	(0.11)	(0.11)	(0.14)
man treated	-0.039	-0.07	-0.073	-0.14	0.14	0.026	0.011
	(0.23)	(0.22)	(0.22)	(0.23)	(0.13)	(0.14)	(0.14)
Born after 01/92	0.3	0.32	0.50	0.55	0.29	0.30	0.50
	(0.14)	(0.14)	(0.47)	(0.47)	(0.16)	(0.16)	(0.47)
Covariates:							
Year of birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family background var.	No	Yes	Yes	Yes	No	Yes	Yes
Family background var. *Yob. Dummies	No	No	Yes	Yes	No	No	Yes
Eligible person sick * Y.OB dummies				Yes			

Notes: standard errors (robust to correlation of residuals within households) are shown in parenthese.

Table 10:
Disposition of income

	Savings (formal+residual)		Food expenditures		Alcohol, tobacco, etc… expenditures		Food share		Alcohol, tobacco, etc.. share	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Woman's pension	0.99	0.82	-0.0096	0.049	-0.024	-0.024	0.51	-0.042	-0.28	-0.084
income	(0.093)	(0.16)	(0.054)	(0.086)	(0.016)	(0.025)	(0.24)	(0.38)	(0.11)	(0.16)
Man's pension	0.78	0.53	0.015	0.041	0.027	0.028	0.042	-0.62	0.02	0.19
income	(0.13)	(0.22)	(0.074)	(0.12)	(0.023)	(0.035)	(0.33)	(0.55)	(0.15)	(0.23)
Non pension income	0.53	0.50	0.14	0.13	0.033	0.037	-0.51	-0.61	0.044	0.15
	(0.017)	(0.041)	(0.0092)	(0.022)	(0.0028)	(0.0065)	(0.041)	(0.10)	(0.019)	(0.043)
F. test woman=man	1.49	1.1	0.07	0	3.07	1.43	1.24	0.78	2.47	0.99
(p. value)	(0.22)	(0.29)	(0.78)	(0.95)	(0.08)	(0.23)	(0.27)	(0.38)	(0.12)	0.32

Notes: Standard errors in parentheses
Instruments: Dummies for: head is employed, head holds a regular job, a casual wage job, a job an agricultue, sector of the job
employer's type (central or local government, private firm, other), pay type (weekly, fortnightly, monthly), woman eligible, man eligible.
Coefficients are multiplied by 10,000 in columns (9) and (10).

Chapter 3

Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry

3.1 Introduction

The[1] idea that there are severe limits to what can be achieved through contracting has had an enormous impact on the way economists now think about firms, markets and governments. Correspondingly, there has been a growing emphasis on the role of reputation as a way of counteracting the problems created by the limitations of contracting.[2] While less often emphasized, a view of the world which gives central importance to issues of contracting, reputation and trust, also has important consequences for the process of growth and development. Most importantly, it suggests that the lack of a proper infrastructure for contract enforcement (which makes contracting less effective) and the difficulty of building a secure reputation[3] are po-

[1] This chapter was written joingly with Abhijit V. Banerjee.
[2] See for example Greif (1993), Baker, Gibbons and Murphy (1995).
[3] Stemming from prejudice, or a history of bad performance, as emphasized by Tirole (1996).

tentially important determinants of success in getting out of poverty, along with the more conventional determinants such as human capital and physical infrastructure.

This paper attempts to quantitatively assess the importance of reputation and, by implication, the seriousness of the limits on contracting, in the context of the Indian customized software industry. Customized software is an obvious place to study such effects since the desired end-product tends to be extremely complex and difficult to describe ahead of time in a way that a third party (such as a court) would understand. In fact, typically the parties to the contract themselves do not fully understand what they want till well into the production process. Therefore it seems naive to expect that they could write a contract enforceable by the courts that would fully cover all contingencies that could arise in the production process. Moreover software production does not require very much fixed capital: indeed most firms nowadays simply own a number of PCs (which are cheap and getting cheaper). The rest, including the premises, access to a mainframe and links to a satellite can all be rented.[4] This limits the possibility of the reputation effects that interest us being confounded with the effects of deferential access to capital or the lack of real competition.

The Indian software industry is suitable for such a study for a number of reasons: First, it is an industry which quite large (employing 140000 people with a turnover of $1.75 billion in 1997-98) and growing fast (at an average annual growth rate of 54% over the past six years). Second, its main focus is on exports (more than 60% of its revenue comes from exports) and a large (over 30%) and fast growing share of the exports is customized software. Moreover the industry's current focus is on expanding the export of customized software relative to its other businesses on the grounds that this is likely to be its best bet for the near future. Consequently the limits of contracting are a major issue in this industry and one that everyone is clearly concerned about. Finally, the fact the contracts are typically across long distances

[4]In India the government has actually invested heavily and by all accounts fruitfully, to make sure that firms have the option of renting expensive fixed inputs (such as expensive computers, building space and equipment for satellite telecommunication), in virtual "Software Techonology Parks".

makes contracting more complicated both by making monitoring somewhat harder and, perhaps more importantly, because of the inherent difficulties of international litigation (combined with the deficiencies of the Indian court system). The data we use in this paper comes from interviews of 125 software companies in three major software development centers in India (Bangalore, Hyderabad and Pune)[5]. We collected detailed data on the company and on the two last projects they have completed, including what kinds of contracts were initially arranged between them and how the contract got renegotiated as the project evolved (we have a total of 236 contracts in our data set).

Prima facie, the data supports both the view that contracting is very limited and the view that reputation is important. All contracts in our sample are either fixed price contracts or time and material contracts. In fixed price contracts the software firm - henceforth the firm - gets a fixed price and is supposed to pay for all realized costs. In time and material contracts the software buyer - henceforth, the client- is supposed to pay for all realized costs. A large fraction of the contracts do however get renegotiated ex post: the buyer does not pay the entire cost in almost half the fixed contracts and the client pays less than the full amount in about a quarter of the time and material contracts. There is also a simple pattern in both the kind of contract that get chosen and the sharing of the costs which is a result of the renegotiation. It is shown in Figures 1 and 2. Figure 1 shows the fraction of fixed price projects as a function of the foundation date of the software firm, and Figure 2 shows the share of overrun[6] paid for by the firm as a function of the age of the software firm.[7] Both are sharply increasing with the starting date of the firm. In particular, firms created in 1994 or after (half of the sample) bear a substantially larger share of the overrun than older firm on average, and the share of overrun they bear is increasing more

[5]In each city, we interviewed half of the firms who belong to the software technology park (all exporters do). We selected the firms randomly, but we oversampled the firms that are not fully owned subsidiaries. No firm refused to meet with us and answer the questionnaire. Some appointments could not be arranged to CEO's unavailability at the moment we were interviewing, and these firms were replaced.

[6]The amount of the project cost that goes beyond the initial prediction.

[7]Because the number of firms per year in the sample is small for firms created before 1988, we have grouped all these firms together.

sharply with age over this range. Measured both in terms of the ex ante contract and in terms of ex post outcome, young firms bear a larger share of the 'risk' of each software project.

This effect of age is perhaps the main empirical finding of the paper. We interpret this as an effect of reputation on the grounds that the firms that started in the industry a long time ago and have survived are more likely to be the kinds of firms that clients can trust - the older firms that cannot be trusted are likely to have already gone out of business (since eventually people would have got to know about them).

To provide further support for our interpretation of the age effect as a reputation effect, in section 4 we show that a similar pattern exists when we use other potential measures of reputation such as whether there has been a previous transaction between the firm and the client, whether it is an internal project (i.e. with a client who either owns the firm or has a long-term arrangement with the firm)[8], etc...Further, we show that different kinds of reputation are to some extent substitute. For example, the difference between young and old firms disappear among firms that work for an internal client.

In section 3 of the paper we present a simple model based on our observation of the industry which explains why reputation would have the observed effect on contractual outcomes. The basic idea of the model is that in most cases by the end of the project the firm and the client know who was responsible for cost overruns. While this is not contractible, firms and clients could nevertheless benefit from it if they could commit to always follow a certain norm. The norm we emphasize here - clearly there can be other norms that will also work - is that of being reliable: reliable firms always try very hard to ensure that they do not exceed the cost overrun that they had implicitly promised, and pay for any extra overrun when they fail to do so. The problem is that this is typically not consistent with short-run profit maximization by the firm or the client, and can only be sustained if the firms and clients are either innately reliable or, more conventionally, if the particular equilibrium that they are playing induces them to put some value on their reputation. We look at equilibria where a

[8]We will describe this type of structure below.

certain fraction of firms and clients are reliable and the rest are not and investigate the implications of a change in the fraction of those who are reliable (interpreted as a change in the average reputation of the firms). The basic trade-off that governs what happens is that fixed price contracts are best for protecting reliable buyers from unreliable sellers while the reverse is true of time and material contracts. Therefore there should be more fixed price contracts if the share of buyers who are likely to be reliable is smaller, which is consistent with the evidence we describe above. We also argue that a number of other predictions from this model are consistent with what we observe.

While we do provide some evidence supporting the broad premises of our model, it is clear that we cannot provide sharp enough evidence to rule out alternative reputation models: it is possible for example that the relevant reputation is for honesty or for a different form of reliability. The objective of this paper is not to distinguish among different kind of reputation. However our reputation-based story does rule out many alternative explanations. In particular it rules out models where there are no agency problems as well as models of agency problems where there is no learning about the firm's type.

Of course, this is all conditional on establishing that we are in fact correctly interpreting the data when we impose the reputation model on it. In other words, it still remains possible that what we are picking up here is the effect of some other variable which happens to be correlated with these measures of reputation. In section 5, we consider some of these explanations. They fall broadly in two classes. First, a class of alternative explanations, which rule out agency problems, explain the differences in the contracts either in terms of differences in risk-sharing or in terms of differences in the production technology available to the firm. We first argue that it is very implausible that the contractual variations that we observe are a result of optimal risk-sharing. The basic point is that in our data set, *firms are usually much smaller than their clients and young firms are especially small.* It is therefore very hard to understand why firms bear so much of the risk (57% on average) and why especially the smallest and youngest firms bear the most.[9] In response to the view that there are differences in the production technology (essentially that young firms are more incompetent) we point out hat *the* natural effect of such incompetence should be to *lower the price* the young firms gets paid *rather than to make them bear a lot risk* than them can ill afford. Moreover the evidence does not support the view that the differences in competence between the firms is of a magnitude that can explain the differences in the contracts. For example, we present in figure 3 the average overrun as a function of firm's foundation date. If the high shares of overrun paid by young firms were a way to make them pay for higher overrun, we should see average overrun increasing with age. If anything, the opposite seems to be true. In response to the second class of competing claim, namely that there are agency problems but no learning about the firm, we point out that this conflicts with the evidence on the effects of sources of reputation other than age. Finally, a number of other potential candidates for an alternative explanation of the data are also examined in section 5.

As a final piece of evidence we emphasize the fact that the necessity to build reputation and trust is recognized and is emphasized repeatedly at the industry level as well as by individual firms. For example, the national association of software services companies (NASSCOM) directory of the Indian Software industry has a large section on "quality". The main element they stress is the number of Indian firms that have ISO 9000 certification or are in the process of acquiring it (ISO certified firms have proven that their software development processes follow approved routines, which is a way for firms to establish a reputation). The association provides technical consulting to any member who wants to get ISO certification. The Indian government provides financial incentives for firms who acquire it. At the individual level, effort to develop a reputation are also obvious.[10]

This paper is a part of a small but growing number of papers that study the

[9]There are of course other determinants of the sharing of the risk. We discuss these issues in section 5

[10]20% of the firms in our sample already have ISO certification. 13% are in the process of getting it.

empirics of contractual choice.[11] Among recent papers Crocker and Reynolds (1993) is most closely related to this work. They examine the determinants of the choice between fixed price contracts and more flexible contracts in US Air force engines procurement. In their view, the key trade off is the following: Fixed cost contracts protect the government against ex post opportunism (in particular it makes it useless for the contractors to claim higher costs) but they require the ability to draft an exhaustive list of requirements (a complete contract), which is possible, but costly. Time and material contracts do not require a truly complete agreement ex ante, but open the room for opportunistic behavior by the contractor. Contracts will tend to be fixed costs if the nature of the engine makes them easy to draft (if the engine is well known or the production cycle is short), and if the contractor is more likely to behave opportunistically. Their empirical analysis of a panel of 44 contracts between the government and two contractors confirms these predictions. Their work shares therefore a central intuition with ours: the reputation of the contractor does matter for the choice of contracts.[12] The more reputed a firm is, the less likely it is that the contract will be fixed cost. The central difference is that fixed cost contracts are not associated with any ex-post cost for the contractor, since fixed cost contracts are "truly complete agreements". The government never behaves opportunistically. They don't discuss what happens when there are cost overruns in fixed cost contracts, because that is not an option in their model. In contrast, we recognize the fact that in the software industry the contract is never complete. Fixed cost contracts need not be more precisely drafted than time and material contracts. Overrun happens in both types of contracts. The central trade-off is between containing opportunism by the client and opportunism by the firm.

Lafontaine and Shaw (1996) is another paper that looks at the effect of firm's age on contracts (in the context of franchising) and finds that the franchisor's age has no

[11]Monteverde and Teece (1982), Masten and Crocker (1985) and Joskow (1987) are important early papers on this subject. These papers differ from ours in studying settings where there are huge relationship specific investments and very long term relationships are the norm and where the key trade-off is between tightness of contract (or control) and flexibility.

[12]In their paper, they measure the reputation by the number of litigation conflicts that the contractors had in the past.

effect on the contract. However as they point out (following Mc Affe and Shwartz (1994)), by changing the franchise contract over time a franchisor runs the risk of hurting its early franchisees (who are locked into one contract while their competitors get a different contract that perhaps allows them to be more aggressive). Because of this cross-contract externality, contracts may not change very much over time, even if the market over time becomes more knowledgeable about the franchisor.[13]

The rest of the paper is organized as follows: In section 2, we describe the institutional settings in more detail and present a number of basic facts about the production of customized software. In section 3, we present a model which reflects our understanding of the way the Indian customized software industry functions, based on two-sided asymmetric information and reputation effects. The model gives rise to a number of predictions about how reputation shapes the contractual forms as well as the ex-post renegotiations. In section 4, we provide evidence which support the basic assumption and the implications of this model. In section 5, we discuss alternative explanations of the pattern observed in the data. Section 6 concludes.

3.2 Institutions and basic facts

We begin by describing the sequence of events leading to the off-shore production of a piece of software.[14] The project begins when the client sends a request for proposal to one or more firms. Each interested firm studies the request (this takes the firm 1.25% of the total project cost for the median external project[15]), and submits a proposal, which includes, among other things, a proposed mode of payment and an estimate of how much the client would have to pay. The client chooses a firm, and the firm and the client agree on a contract. The contract specifies an estimate of effort needed to complete the project, a mode of payment, financial details (price, etc...) and a

[13]This paper is also related to papers such as Barron and Umbeck (1984), Shepard (1993) and Genesove (1993) which test the implications of theories based on aymmetric information in industrial contexts (but not the implications for the choice of the contract).

[14]Table 1 shows the descriptive statistics mentioned in this paragraph.

[15]Those projects where the client does not own the firm or does not effectively control the part of the firm which is working towards the completion of the project (see below).

projected schedule for deliverables (which are specific milestones -corresponding to phases of the software development process or to modules of the software- that will be reached in the course of completing the project). The work then starts. The first phase is the writing of specifications. The firm, in collaboration with the user at the client's end, writes the set of functions that the software will execute. For the median project, it takes 10% of the total project effort to complete this phase.[16] At the end of this part of the project, what the client wants and what it would cost is usually clearer to both the client and the firm and the schedule of deliverables is sometimes amended or clarified.

The second phase of the work is the lower level design, coding and testing of the software. When a specified milestone is reached, the firm sends the deliverable to the client. Each time this happens the client can either acknowledge that it has been delivered (by signing off) or request changes. The firms also send regular status reports to the clients (a little less than once a week on average), keeping the clients up-to-date about the progress of the project.

In terms of project outcomes our main focus will be on overrun: overrun in industry parlance is the difference between the amount of effort actually needed to complete the project and the estimated effort given in the contract. It is therefore important to be clear about firms mean by an estimate. A standard textbook on software management (Pressman (1997), has an entire chapter on estimation. He describes the process as follows:

> 'The project planner begins with a bounded statement of software scope and from this statement attempts to decompose software into problem functions that can each be estimated individually. Line of Code or function points (the estimation variable) is then estimated for each function. Alternatively, the planner may choose another component for sizing, such as classes or objects, changes or business processes impacted. Baseline productivity metrics (i.e. line of code per person-month or function point

[16]for some projects, specifications writing and subsequent work are decoupled. One firm -or the client himself- writes the specifications, and another firm completes the project.

per person months) are then applied to the appropriate estimation variable and cost or effort for the function is derived. Function estimates are combined to produce an overall estimate for the entire project.' (Pressman (1997)).

Our interpretation of this and other material in this book (which is also consistent with what we have learnt from industry sources) is that the estimate is the firm's best guess about how much effort will needed to complete the project, *assuming that the firm's current understanding of the project is correct and that the firm adheres to it own productivity norms.*[17]The estimate is therefore clearly not meant to be an unbiased estimate of how much effort the project will actually take. This is important because it tells us that overrun represents the extent of deviation from the firm's initial plan of action.

This also tells us that overruns ought to be quite common: first, because the needs of the client are typically not very clear at the very beginning of a relationship - even to the client himself. Moreover the client may not put enough effort into understanding and explaining what it wants. Not surprisingly then, the firm often does not understand what the client really wants. When, in the course of the project, the needs of the client eventually become clear, changes have to be made and these are costly. Second, the amount of time and effort needed to design and code a piece of software is difficult to evaluate ex ante, even when the set of functions is well-defined (both for the client and for the firm), and will depend on the type of technology being used, the ability and the experience of the staff of the two firms. Third, with the best of staff and the clearest goals there is also the risk that some unexpected problem arises and delays or destroys the project. Finally, not all firms try their hardest to control costs and delays and one would expect that some projects will end up costing much more than they ought to.

Table 1 shows evidence from our interviews confirming that overruns are indeed common: it turns out that 74% of the projects are completed with a positive over-

[17]In other words, the presumption behind the estimate is that the firm has understood perfectly what the client wants and that the firm implements the project at its normal level of productivity.

run. The average overrun amounts to 24% of the initial estimate, and varies a lot (it standard deviation is 34%, and the maximum overrun in the sample is 250%). According to the firms, overruns are due mostly to changes required by the client (these changes cause 48% of the overrun on average). Another 20% of the overrun is due to initial ambiguity in the specifications (i.e. to cases where the firm did not understand what the client really wanted), 8% is due to internal difficulties in the firm (the most frequent one being the loss of the project manager in the middle of the way) and 13% to delay occasioned by the client. Very few projects (less than 5%) are completed with a negative overrun, and that the mean overrun is clearly not zero.

Both firms and clients are, of course, aware of the possibility of overruns.[18] Overruns, apart from being wasteful in themselves (in so much as they could have been avoided by both parties being more diligent), lead to delays which are costly[19] and are a potential source of conflict between the client and the firm (conflicts arise when each side blames the other for the overrun).

Vertical integration and contracts are two ways of limiting the waste due to overrun. Many foreign companies have set up 100% owned subsidiaries in India.[20] These subsidiaries are 100% export oriented, and carry outwork for their mother company and in some cases, for other clients as well. A number of Indian software firms have also entered into arrangements under which the firm dedicates a part of its employees, office space, and computers to a single foreign client. This is what is called an "Off shore software development center" (OSDC). The client sends a steady fraction of his software development need to the firm, and is responsible for making use of the facilities devoted to him. This is a in effect a type of vertical integration: the OSDC becomes virtually a unit of the client for whom it works regularly. In such cases the

[18]For example, the template of a firm's contract specifies that "the effort estimates provided for the conversion and testing phases of this project have been provided by *the software firm* on a best estimate basis. If the scope of the effort changes as a result of discussions during the detailed design phase, *the software firm* will analyze the impact of changes on the project and may present revised schedules and costs. Changes in schedules and costs resulting from such changes will be reflected by an amendment to this contract."

[19]Delays, while rarer than overrun, are far from uncommon in our sample: there are delays in 19% of the cases, and 25% of the cases where there was an overrun.

[20]Including AT&T, IBM, Microsoft, INTEL, ORACLE, Fujitsu and Motorola.

interests of the firm and the client are clearly better aligned and while there may be overrun, there is much less reason why the overrun should be wasteful.[21]

Since we are interested in contracts rather than vertical integration our focus in this paper is mainly on external contracts (i.e. contracts that are performed neither within OSDC, nor for the mother companies of the firm). We observe the following types of external contracts: Under *fixed price contracts*, a fixed price is agreed upon up-front, before the specification analysis. These contracts are by far the most frequent: 58% of external contracts are fixed price contracts. Under *mixed contracts* the price is fixed for the specification phase only at the beginning of the process. The price for the complete project is fixed only when specifications are written and more is known. Typically in such cases the requirement analysis is paid for on a time and material basis, though this is not necessarily the case. Under *time and material contracts*, the entire product is paid for on a time and material basis. This contracts are the least frequent among external contracts (15%). A striking fact is that there appears to be no "intermediate" contracts: all contracts belong to one of these three categories.[22] For example, there are no contracts where the client and the firm agree on sharing the costs.[23]

While these contracts predict extreme outcomes in terms of cost-sharing, we actually do not always observe this. It turns out that a large fraction of contracts get renegotiated ex post. This is evident from Table 2, which shows the fraction of overrun paid for by the firm and the proportion of firms that pay all or nothing of the overrun for the three types of contracts. Even in fixed price contracts, the actual overrun is often shared between the client and the firm (in 46% of the cases) while firms with time and material contracts sometimes pay for overrun (in 22% of the cases). However it is also clear from the figures in table 2 that the initial contract has an clear influence on which party bears the risk of the project: in fixed cost con-

[21]Indeed there may be more overrun in such cases than in general precisely because overrun entails less waste). For example, the client may not need to be very precise about what he wants since he knows that the firm will be happy to do whatever is asked of it.

[22]Or their variants: in some cases property rights in the product substitutes for cash payments.

[23]Such contracts are observed, albeit rarely, among the procurement contracts for airplane engines studied by Crocker and Reynolds (1993).

tracts, firms bear on average 63 % of the overrun, while they bear on average 51.5% in mixed contracts and 15.5% in time and material contracts. Since fixed price contracts dominate our sample, this evidence also implies that firms bear a lion's share of the overrun (57% on average of external project, 76% for the median external project). Since firms are typically much smaller than their clients this is at least somewhat surprising.

There are several potential explanations for the pervasiveness of renegotiation. First, even when a firm faces a fixed price contract it may have some bargaining power because it usually has the option of walking off the job. If it does, it will not get paid for work that it has already done, but it will also avoid the overrun and at least at early stages of the job, the second effect may dominate. Second, the court system in India is extremely inefficient and going to court is very costly. Firms and clients will therefore prefer to make some concessions in order to avoid going to court. In fact, from our conversations with industry people we have the impression that people go to court very rarely and therefore we ought to expect some renegotiation.

Finally, firms and clients may voluntarily pay for any overrun that is of their own making, because they care about their reputation for being reliable. We had a number of conversations where the CEO of the firm told us 'it was our fault and we paid for it'. We also have some more indirect evidence that this is at least sometimes the case: as mentioned above, we asked firms questions about who was responsible for the overrun. In what follows, we assume that the firm is responsible for what it described as changes due to ambiguities and overrun caused by internal difficulties. Changes required by the client and delays coming from the client's side are taken to be caused by the client's responsibility. Table 3 shows the share of overrun paid by the firm when the overrun is entirely due to the client (column (1)), entirely due to the firm (column (3)), or due partly to both (column (2)). In column (4), we present the coefficient of an OLS regression of the share of overrun paid by the firm on the share of overrun which it caused. In all types of contracts, firms always pay more of the overruns entirely caused by their own mistakes compared to the overrun entirely due to the client. Moreover, in all cases but one, the share of overrun paid by the firm lies in between these two numbers when the overrun is partly caused by each side. Furthermore, the OLS regressions indicate that, regardless of the initial contract, the larger the fraction of the overrun that a firm has caused, the larger the share it has to pay (if a firm causes one additional percent of the overrun, it bears approximately 0.20 percent more of it).

In the next section we present a model of the industry which is based on the picture that emerges from the above discussion. The main elements we wish to capture in our model are the following:

- the high levels of overrun,
- the fact that both sides are responsible for overrun,
- the fact that the software firms end up bearing a large part of the overrun,
- the use of simple ex ante contracts,
- the fact that the contracts get renegotiated ex post,
- the fact that the ex ante contract continues to influence the renegotiated outcome,
- the fact that firms and clients care about their reputation for being reliable and will often voluntarily pay for overrun that is of theor won making.
- and the fact that young firms bear, on average, more of the overrun than old firms

3.3 A Model of the Software Industry

The model we propose in this section is an attempt to capture in as simple way as possible what, on the basis of our experience in the industry, we see as the fundamental structures and conflicts in the Indian customized software industry. The contracting outcomes that will be predicted by the model will, as we shall see, match up reasonably well with what is observed in the data. However, one could come up with other models, or at least combinations of other models, which also explain the data. We will discuss some alternative explanations in the next section. In the end, however, it remains plausible that elements of these other models could also be a part of any comprehensive story of the software industry in combination, perhaps, with the story we tell. In this sense, the model is meant to be illustrative rather than definitive.

The premise of the model is that software projects are prone to cost overruns and that the main conflicts are over the apportioning of these cost overruns. Overruns can happen for two reasons. First, the client could have been insufficiently diligent in delineating his requirements or he could have made a mistake. As a result, when the firm comes up with a product he might realize that this is not what he wants and demand changes. The firm is, of course, happy to make the changes – since they are Pareto improving – but only if it is adequately compensated. The issue is whether the client will be willing to compensate it enough. Second, overruns could also happen because the firm was either lazy or unlucky in the way it carried out the project.

Since the overrun could come from either side, when there is an overrun, there is a real possibility that each side will blame the other for it. This would not, of course, be a problem if outsiders and specifically the courts can observe who was really responsible. Our assumption will be that this is not possible in most cases.

This is clearly something of a caricature of reality: firms and clients clearly do try to set up systems to ensure that it is clear, ex post, who was to blame for any overrun. The procedure of defining deliverables and having the client sign off on each deliverable is one such system. Once a client signs off on a deliverable, he is to a large extent committed to admit that at least up to that point the firm had done what it was supposed to do. This clearly limits the scope for future disagreements. Nevertheless, there seem to be lots of disagreements and this is presumably ascribable to the fact that even after many milestones have been reached, there remains substantial ambiguity about what exactly needs to be done.

We capture the possibility of this kind of disagreement as follows. The client (C) wants the firm (F) to build a piece of software that will be worth V to the client. It should normally cost an amount $\overline{y}$ (i.e. the estimate is $\overline{y}$). However with some probability there is an overrun and the total cost is $\tilde{y}$. We adopt the normalization that $\overline{y} = 0$ so that all of the cost of the project is overrun.

Overrun is the sum of overrun caused by the firm (y_F) and overrun caused by the client (y_C). The amounts of the overrun, y_F and y_C, are chosen by the firm and the client respectively. Both firms and clients get some private benefits from generating high levels of overrun - this may be because controlling overrun takes effort[24] or because the firm (or the client) gets to keep a part of the overrun it has generated (cost padding).[25] These private benefits are given by $B_C(y_C)$ for the client and $B_F(y_F)$ for the firm. Both are assumed to have the usual increasing concave shape.

Given these assumptions, the first best outcome has y_C and y_F being chosen to satisfy

$$1 = B'_C(y_C) \text{ for the client,}$$

and

$$1 = B'_F(y_F) \text{ for the firm.}$$

However, we will assume that both y_C and y_F are private information: third parties such as the courts only observe total overrun ($y_C + y_F$). Therefore the only enforceable contracts that do not involve money being thrown away ex post can never give the first best.[26]

It is however possible to improve on this outcome if the behavior of the firms and the clients is at least partly norm-governed. Specifically assume that there are two types of firms and two types of clients. Of these, one type of firm and one type of

[24]It requires that the client puts effort into defining its requirements and the firms puts effort into understanding them.
[25]In other words, the firm's effort should thought of as the effort spent by the management on properly organizing the work on the project and not as simply labor time.
[26]See Holmstrom (1983).

client observes a norm of being reliable. Assume, pending discussion, that reliable firms and clients choose whatever level of overrun that they have ex ante promised to deliver. Assume also that by contrast, unreliable firms and clients always choose y_C (or y_F) to maximize their current profits. We will later make assumptions and that both $\overline{y_C}$ and $\overline{y_F}$ are much higher than the first best level and indeed more than anything that a reliable type would ever choose.[27] Moreover this high level of overrun is accompanied by delays and these delays cost the other party an amount D. We assume that D is large enough that no one will want to contract with someone who is known to be unreliable. Finally let the fraction of reliable firms be denoted by θ_F and the fraction of reliable clients be θ_C and let the actual type of the firm (whether it is reliable or not) be private information.[28]

In this setting, since the reliable firms and clients are going to be self-regulated, the function of the contract is to protect reliable clients against opportunism by an unreliable firms and vice versa. Assume for the time being that in the event of an encounter between an unreliable firm and a client (who may or may not be reliable) or between a unreliable client and a firm, the outcome is governed by the ex ante contract. This is the natural outcome since in this case renegotiation can only lead to a redistribution between the two parties and neither party has any reason to give up something for the benefit of the other. To simplify the analysis we assume that *the ex ante contract is always linear*, i.e. it takes the form of a fixed payment P to the firm and a share s of total overrun that is paid for by the firm. When $s = 1$ we will describe the resulting contract as a fixed price contract and when $s = 0$ we will call it a time and material contract.[29]

[27] In other words, the private benefits of generating a high level of overrun are extremely high for unreliable firms.

[28] We emphasize here the importance of reputation for reliability. Firms, of course, have reputations not just for reliability but also for competence, cooperativeness or honesty. In principle one could build reputation stories based on the idea that it is these characteristics and not reliability that are imperfectly observed that could also explain the data. We focus on this mechanism in writing the model here, because we think it is an accurate description of the most important conflicts in the industry, but it is not to say that we think these other kinds of reputation are unimportant.

[29] The actual time material contracts do not have the fixed price component that we have assumed here. Instead firms get paid a markup on realized costs. Taking this into account complicates the analysis (the convenient transferable utility assumption can longer be made) but the comparative

It remains to say what determines payoffs in a situation where both parties have acted reliably. We will assume that the price P specified in the contract is always paid so that the only issue that remains is to decide how the overrun gets split.[30] If, as we have assumed, reliable firms always follow the norm and deliver the contracted outcome, the contracted division of the overrun is irrelevant in the sense that it has no effect on their choices of y_F and y_C. However it is clearly easier to enforce the norm if the overrun is split in such a way as to give the two parties the incentive to follow the norm. We will presently show that a rule that has this property is one which says that a reliable firm or reliable client always pays for the overrun it has generated. From conversations with industry people it is clear that this is a rule that a lot of firms do follow. Moreover, in the next section, we will provide evidence that this rule is applied.

If we accept that this is the rule for splitting the overrun, it is clear that the ex ante contract will frequently be renegotiated. Moreover renegotiation here will actually enhance ex ante efficiency.

We prefer to remain agnostic about the source of norm-governed behavior that is at the heart of this model. It could be that it is simply an outcome of a repeated game: even in cases where the client and the firm may not expect to transact again they will remain part of the same industry and as long as the information about their past behavior becomes public with some probability, we can expect there to be repeated game equilibria where all sufficiently patient agents honor the norm. Norm-governed

statics remain very similar.

[30]This is not an entirely innocuous assumption. One could imagine a contract where P is set very low and then once it is established that both parties are reliable (i.e. at the end of the production process) they renegotiate the price upwards. This would discourage unreliable firms from bidding for the contract. Alternatively one could set a very high P in order to discourage unreliable clients. In practice we do not observe any instances of this kind of arrangement. This may be because setting a low P also expands the scope for opportunism. In the setting of our model this does not matter because both parties are risk-neutral but if, as is plausible, they are at least somewhat risk-averse, this kind of contract may be too risky. Moreover there is a serious question as to whether courts will enforce such contracts - the courts may feel that the agreed upon price is unfairly low (or high). Finally, and perhaps most importantly, this kind of contracts relies on a much more inclusive notion of what it means to be reliable: reliable clients are now expected to give up a chance to make a huge amount of money (by refusing to raise the price ex post). It is not clear whether this kind of expectation is realistic.

behavior could also be innate (see, for example, Bowles (1998) for an evolutionary model of social norms). Finally it is possible that there is a fraction of innately reliable people and a much larger population of opportunists who in equilibrium imitate the behavior of the reliable agents (as in the reputation literature (Kreps et al. (1982), Fudenberg and Maskin (1986)).[31]

Here we will focus exclusively on equilibria where all the people we have called reliable never deviate from the norm - in other words both those who are innately reliable and those who choose to be reliable will follow the exact same norms. By contrast those we have designated unreliable will always follow the behavior we assigned to them above.

These assumptions allow us to treat the types of the firms and clients as fixed parameters and to analyze what happens in a one-shot interaction between the firm and the client. However a complication still remains: the choice of the contract can be used as a signaling device - a firm that plans to be unreliable will prefer a contract where it pays very little of the overrun and therefore, by choosing to absorb most of the overrun a firm may be able to signal that it is reliable. Given that we are in a signaling environment, we will expect that there will be many equilibria. However all such equilibria will involve pooling since in a separating equilibrium all the unreliable firms and clients will never get a contract and therefore will prefer the contract chosen by the reliable type.

Among the set of pooling equilibria we focus on the contractual outcome where the utility of the client of the reliable type is maximized given that the firm is getting at least its outside option if it is of the reliable type. This is always a Bayesian-Nash equilibrium (sustained by the belief that only opportunists deviate). The fact that it is also Pareto optimal from the point of view of the reliable types makes it an obvious focal outcome.

[31] The simplest way to model this would be to assume that at some point after the transaction between the firm and the client is completed, the fact that one of the parties had acted unreliably becomes public with some probablity (a disgruntled employee reports what really happened or an incriminating document gets to the wrong hands). The future play of the game would then be made contingent on such public information. For modeling of social norms along these lines see Kandori (1990).

3.3.1 Analysis of the Basic model

Given that we are in a transferable utility setting, maximizing the utility of a client of the reliable type under the constraint that a reliable firm is getting at least its outside option amounts to maximizing the total social surplus calculated from the point of view a pair of reliable types. This expression for joint surplus is:

$$W = V + B_C(y_C) + B_F(y_F) - \theta_F y_C - (1-\theta_F)(1-s)(y_C + \overline{y_F}) -$$
$$(1-\theta_F)D - \theta_C y_F - (1-\theta_C)s(y_F + \overline{y_C}) - (1-\theta_C)D.$$

The fifth terms and sixth in this expression give the total surplus that is lost because in a pooling equilibrium a reliable client must allow for the possibility that the firm is unreliable while last two terms are the surplus that is lost because the firm must allow for the possibility that the client is unreliable.

It is immediate that the choice of y_F and y_C that maximize this expression must satisfy, respectively,

$$\theta_C + (1-\theta_C)s = B'_F(y_F), \tag{3.1}$$

and

$$\theta_F + (1-\theta_F)(1-s) = B'_C(y_C). \tag{3.2}$$

The levels of y_F and y_C that will be expected of a reliable firm and a reliable client will therefore satisfy these two equations (recall that reliable firms and clients always do what is expected of them and these are the levels that maximize the joint surplus).

Note that these are also the levels of y_F and y_C that would be chosen by a reliable firm and a reliable client if there were no explicit expectation about the y_F and y_C they chose but they were expected to pay for the share of overrun that they had generated (as long as the other party has behaved reliably). To see this note that under these rules they will choose y_F and y_C to maximize, respectively,

$$W_F = P - \theta_C\left(y_F - B_F(y_F)\right) - (1-\theta_C)\left(s(y_F + \overline{y_C}) - B_F(y_F)\right) - (1-\theta_C)D \quad (3.3)$$

and

$$W_C = V - P - \theta_F\left(y_C - B_C(y_C)\right) - (1-\theta_F)\left((1-s)(y_C + \overline{y_F}) - B_C(y_C)\right) - (1-\theta_F)D. \quad (3.4)$$

Maximizing these expression yields exactly the same expression for y_F and y_C as the maximization of the total surplus.

Claim 1 *If we require firms and clients to pay for any overrun that they have caused and reliable firms follow this rule when they deal with other reliable firms but stick to the initial contract when they are dealing with unreliable firms, the level of y_F and y_C that will be the same as those that maximize the joint surplus defined above.*

On the strength of this last claim we will henceforth assume that reliable firms and clients actually follow the rule of paying for their own misdeeds unless the other party is unreliable. This gives us a specific rule for sharing the overrun following a renegotiation which (as we have seen) is optimal while being empirically plausible. Moreover it explains the frequency of renegotiation: generically there will be some renegotiation of the initial contract whenever two reliable parties get paired.[32]

Recall next that the condition that determines y_F is:

$$\theta_C + (1-\theta_C)s = B'_F(y_F) \quad (3.5)$$

This gives us $y_F(s, \theta_C)$, with $\frac{\partial y_F}{\partial s} < 0$ and $\frac{\partial y_F}{\partial \theta_C} < 0$.

Likewise, we can derive $y_C(s, \theta_F)$ from the equation:

[32] The initial contract described here is therefore an incomplete contract in the sense that it leaves a lot to be determined through ex post renegotiation. This is in fact our understanding of what the actual contracts look like. One could however imagine a complete contract which mimics this contract: it simply has to say that whenever two parties agree to renegotiate, they report the part of the overrun that they have caused and pay for that part. If at least one party does not want to renegotiate, the initial contract is enforced.

$$\theta_F + (1-\theta_F)(1-s) = B'_C(y_C) \tag{3.6}$$

$y_C(s, \theta_F)$ has the property that $\frac{\partial y_C}{\partial s} > 0$ and $\frac{\partial y_C}{\partial \theta_F} < 0$.

We state this as:

Claim 2 *The amount of overrun generated by the firm is decreasing in the share of overrun borne by the firm. The amount of overrun generated by the client is decreasing in the client's share of the overrun. The amount of overrun generated by the client is decreasing with the reputation of the firm, and the converse is true as well.*

This result implies among other things that those clients facing time and material contracts should generate less overrun than in a fixed price contract, while firms in the same situation have the opposite reaction. That ought to be intuitive: time and material contracts give high powered incentives to the client while fixed price contracts give high powered incentives to the firm. It is also worth noting that if $\theta_F = 1$ (all firms are reliable) then the client will choose the first best level of overrun. If $\theta_C = 1$ (all clients are reliable), the firm will choose the first best level of effort. If both parties were reliable, the outcome of the renegotiation is always efficient, which, in a sense, justifies the procedure of renegotiating the contract.

To find the optimal contract we need to maximize the expression for W given above. Differentiating W with respect to s and using the envelope theorem gives us the expression:

$$(1-\theta_C)(\overline{y_F} + y_C) - (1-\theta_F)(\overline{y_C} + y_F)$$

If there is to be an interior optimum for s, we must have:

$$(1-\theta_C)(\overline{y_F} + y_C) - (1-\theta_F)(\overline{y_C} + y_F) = 0$$

and further:

$$(1-\theta_C)\frac{\partial y_C}{\partial s} - (1-\theta_F)\frac{\partial y_F}{\partial s} < 0$$

We have however already seen that $\frac{\partial y_C}{\partial s} > 0$ and $\frac{\partial y_F}{\partial s} < 0$, so this condition cannot be satisfied. This means that s cannot have an interior optimum. Intuitively, the advantage of a high s is that it protects reliable clients by passing off a large part of the overrun onto unreliable firms. Increasing s raises the level of the overrun chosen by reliable clients and therefore increases the advantage of passing overrun to unreliable firms. The disadvantage of a high s is that it forces reliable firms to pay for the overrun generated by unreliable clients. This disadvantage becomes smaller when s becomes larger because the firm itself generates less overrun. For both these reasons, once s is high, the benefit of increasing it even further goes up and therefore an interior optimum cannot exist.

Given that an interior optimum does not exist, the optimum will be either $s = 1$ or $s = 0$. To see which dominates the other, we need to compare:

$$W(1) = V - \theta_C y_F(1, \theta_C) - (1-\theta_C)\left(y_F(1,\theta_C) + \overline{y_C}\right) -$$
$$B_F\left(y_F(1,\theta_C)\right) - \theta_F y_C(1,\theta_F) - B_C\left(y_C(1,\theta_F)\right)$$

and:

$$W(0) = V - \theta_C y_F(0,\theta_C) - B_F y_F(0,\theta_C) -$$
$$\theta_F y_C(0,\theta_F) - (1-\theta_F)\left(\overline{y_F} + y_C(0,\theta_F)\right) - B_C\left(y_C(0,\theta_F)\right)$$

While either of these could be larger, it is clear that the difference $(W(1) - W(0))$ is increasing in θ_C [33] and decreasing in θ_F.[34] This observation gives us the following

[33] The derivative is $\overline{y_C} + y_F(0, \theta_C)$.

[34] The derivative is $-y_C(1,\theta_F) - \overline{y_F}$.

result.

Claim 3 *The optimal contract is always either a fixed price or a time and material contract. It is a fixed price contract when most clients are reliable while firms are more likely to be opportunists, and a time and material contract in the reverse situation.*[35]

This confirms the intuition given in the introduction that fixed price contracts are instituted to protect clients against opportunism, while time and material contracts protect firms. Firms that have high reputation will get time and material contracts while the rest of the firms will not.

3.3.2 Extensions

Introducing some discretion at the level of the court

The model of the previous section implicitly assumes that, when a contract goes to court, the court simply enforces the basic sharing agreement and ignores any other clauses written into the contract. In reality, courts certainly exercise a fair amount of discretion and this gets reflected (for obvious reasons) in out of court settlements as well. This possibility can easily be introduced into our model and, as we will see, does not change any of the results reported so far - though it adds some additional nuances to our analysis of the level and variability of cost overruns. We introduce this possibility by assuming that when the contract goes to court, the court implements an outcome which is a convex combination of the original contract and some "fair outcome" that the court determines.[36] We will assume that the share paid by the firm

[35] *This result clearly depends on the assumption of risk neutrality. The extreme contracts predicted here clearly do not make for optimal risk-pooling and if the firm and the client were sufficiently risk-averse they would surely want a more intermediate contract. In the face of this, the fact that no intermediate contracts are observed suggests either that the people are relatively risk-tolerant or that the contract is also influenced by things that we have not modeled. For example, as soon as the client agrees to pay a share of the overrun it would have to set up a system for determining the total overrun. Likewise, if the firm pays a positive share it would have to invest in a system for measuring overrun. If this kind of measurement has a fixed cost, the optimal contract may well be an extreme contract.*

[36] Or equivalently, the court sticks for the original contract with some probability and chooses some other "fair" repartition in the remaining cases.

in the fair outcome is a decreasing function of the firm's reputation and an increasing function of the client's reputation. In other words, the share of overrun paid by the firm when they go to court is written as:

$$s^*(s, \theta_C, \theta_F) = s(1-\rho) + \rho s_c,$$

where $s_c = s_c(\theta_C, \theta_F)$ with $0 < s_c < 1$ and $\frac{\partial s_c}{\partial \theta_C} > 0$ and $\frac{\partial s_c}{\partial \theta_F} < 0$.

This formulation recognizes that when courts have to decide which of the two disputant is lying they give weight to their past reputation.

It is easily checked that equations (3.1), (3.2), (3.5) and (3.6) remain valid with s^* in place of s. In other words we now have the functions $y_F(s^*, \theta_C)$ and $y_C(s^*, \theta_F)$ which exactly parallel the functions $y_F(s, \theta_C)$ and $y_C(s, \theta_F)$ that we had before. Claim 3 therefore continues to hold. Moreover the problem for the choice of an optimal contract continues to generate only extreme solutions. The only difference is that while the optimal s is still either 1 or 0, the resulting s^* will be $s^*(1) = (1-\rho) + \rho s_c(\theta_C, \theta_F)$ or $s^*(0) = \rho s_c(\theta_C, \theta_F)$.

We now define:

$$\begin{aligned} W^*(s^*, \theta_C, \theta_F) = V + B_C(y_C) + B_F(y_F) - (1-\theta_F)(1 - s^*(s, \theta_C, \theta_F))(y_C + \overline{y_F}) - \\ (1-\theta_F)D - (1-\theta_C)s^*(s, \theta_C, \theta_F)(y_F + \overline{y_C}) - (1-\theta_C)D. \end{aligned}$$

These are the expressions for total surplus for a given value of s^*. Now,

$$\begin{aligned} \frac{\partial (W^*(s^*(1), \theta_C, \theta_F) - W^*(s^*(0), \theta_C, \theta_F))}{\partial \theta_F} = -y_C - \overline{y_F} + \\ \frac{\partial W^*(s^*(1), \theta_C, \theta_F)}{\partial s^*}\frac{\partial s^*}{\partial \theta_F} - \frac{\partial W^*(s^*(0), \theta_C, \theta_F)}{\partial s^*}\frac{\partial s^*}{\partial \theta_F} \end{aligned}$$

Of these the first two terms are of course negative. The third term is also negative, since increasing θ_F reduces s^* and reducing s^* reduces $W_F + W_C$ when s is in the neighborhood of 1 (by the quasi-convexity of the $W_C + W_F$ function). The last term is also negative because increasing θ_F reduces s^* and reducing s^* increases $W_C + W_F$ when s is close to 0. Therefore increasing θ_F reduces $W_C + W_F$ when s is close to 0. Therefore, increasing θ_F favors moving to a time and material contract. A parallel argument shows that raising θ_C favors moving to a fixed price contract.

This is as before. However, the adoption of discretion on the part of the court changes things by allowing the share of overrun in the case of a dispute, conditional on having a time and material or fixed cost contract, to depend on the firm's (and the client's) reputation. The share of the overrun paid by the average firm with reputation θ_F that works for a client of reputation θ_C and has a fixed price contract, is:

$$\theta_F\theta_C\frac{y_F(\theta_F,\theta_C)}{y_F(\theta_F,\theta_C)+y_C(\theta_F,\theta_C)}+(1-\theta_F\theta_C)\left(\rho s_C(\theta_F,\theta_C)+(1-\rho)\right)$$

As θ_F goes up there are three effects: First, $\frac{y_F}{y_F+y_C}$ increases, for two reasons: y_C goes down when θ_F increases, and raising θ_F reduces s^* and thus increases y_F. Second, $s_c(\theta_F,\theta_C)$ goes down. Third, when θ_F goes up, it shifts weight from the second term to the first term in the above expression. Since we think of s^* as being close to s, and therefore close to 1, we expect it to be higher than $\frac{y_F(\theta_F,\theta_C)}{y_F(\theta_F,\theta_C)+y_C(\theta_F,\theta_C)}$ and therefore the net effective should be negative.

The last two effects imply that when θ_F goes up, the share of overrun that the firms pays goes down while the first effect goes the other way. The negative effect will dominate as long as the $y_F(\theta_C,\theta_F)$ and $y_C(\theta_C,\theta_F)$ functions are not too elastic, which seems plausible.

Claim 4 *The firm's share of the overrun (conditionally on having a fixed price contract) is likely to be decreasing in its reputation.*

If we now look at the mean overrun generated by a firm of reputation θ_F and a client of reputation θ_C, this is given by the expression:

$$\theta_F y_F(\theta_F, \theta_C) + (1 - \theta_F)\overline{y_F}.$$

When θ_F goes up, y_F goes up but since $\overline{y_F}$ is much greater than $y_F(\theta_C, \theta_F)$, the shift in weight from $\overline{y_F}$ to $y_F(\theta_C, \theta_F)$ has a negative effect on the overrun. Unless y_F is very elastic with respect to θ_F we would expect the second effect to dominate: this tells us that the average amount of overrun caused by the firm should fall as the firm's reputation gets better. On the other hand, since $\overline{y_F}$ is fixed while $y_F(\theta_C, \theta_F)$ goes up when θ_F goes up, the dispersion of overrun must be lower for more reputable firms and as a result the variance of overrun generated by the firm will also tend to be lower for more reputed firms.[37]

Claim 5 *Mean overrun caused by the firm is lower for more reputed firm. The variance of overrun caused will tend to be lower as well. A parallel result holds for more reputed clients.*

Note however that since $y_C(\theta_F, \theta_C)$ is decreasing in θ_F, the average overrun of clients dealing with more reputed firms, $(1 - \theta_C)\overline{y_C} + \theta_C y_C(\theta_F, \theta_C)$ is lower when the firm is more reputed. Moreover since the dispersion is larger and the mean overrun is lower, the variance of overrun generated with the client will grow with firm's reputation.

Claim 6 *The mean overrun generated by a client goes down with the firm's reputation but its variance goes up. The same hold for the overrun generated by the firm when the client becomes more reputable.*

It then follows that:

Claim 7 *The mean total overrun goes down when the firm becomes more reputable, however the variance can go up or down or remain constant.*

[37]A sufficient condition for this to happen is that θ_F is greater than $1/2$, which seems plausible.

Choice of Projects

The fact that firms with low reputation pay for most of the overrun, should clearly influence their choice of projects. This can be introduced into our model by making the plausible assumption that the most rewarding projects (the ones with the highest V) will also have the highest possibility of large overruns ($\overline{y_C}$ and $\overline{y_F}$ are going to be large). It is easy to show by introducing this assumption into the model of the previous sub-section, that keeping the reputation of the client fixed, firms that are facing a fixed price contracts will be more willing to trade off a lower V for a lower $\overline{y_C}$ than firms which face time and material contracts. Therefore low reputation firms will want projects where client side opportunism is limited even at the cost of a lower V.

It does not immediately follow however that the low reputation firms will get these projects: low $\overline{y_C}$ projects are probably also low $\overline{y_F}$ projects and low $\overline{y_F}$ projects are attractive to clients who are facing time and material contracts (i.e. clients who are working with high reputation firms) for precisely the same reason that the firm wants a low y_C project. Therefore high reputation firms may be able to bid away these projects from low reputation firms. In equilibrium however this is unlikely to be true: the difference lies in the fact that for a fixed θ_C, time and material contracts will only be chosen when θ_F is relatively large and overruns caused by the firm are relatively unlikely. Therefore the saving in cost by choosing a low payoff project is going to be relatively small.

Claim 8 *Low reputation firms will tend to be specialized in projects which have low potential for client-side opportunism.*

In terms of what we observe, this seems to suggest that low reputation firms will choose projects which are simple and well-understood so that the client does not have to do very much work to make clear what he wants. These projects can be either short projects[38], or projects where the main goal is easily defined. Year 2000 projects

[38]The complexity of a software project increases sharply with its size, see Pressman (1997).

are typical in this respect.[39]

Adding the possibility of choosing projects however makes it more likely that the results on the relation between reputation and the mean and variance of overrun will be ambiguous. Because firms with low reputation can choose projects so as to limit the overruns, both the mean and the variance of the overrun generated in the projects of low reputation firms may be lower than that generated in the projects of higher reputation firms.

Reputation building

If seems plausible that most firms with a low reputation will want to build a reputation for reliability. On can imagine them trying to signal their reliability by being extremely careful about not generating any overrun. This would have the effect of reducing the mean overrun generated by low reputation firms and to increase its variance, since the dispersion between the reliable and the unreliable goes up. It would also weaken the relationship between reputation and the share of the overrun generated by the firm.

Incentives for opportunists

We have assumed so far that opportunists cannot respond to incentives. This is clearly extreme, especially since the incentives can be quite strong. If we allow the opportunist firms to reduce the overrun that they cause when s is high, both the mean and the variance caused by the firm will have less of a tendency to fall as the firm's reputation improves.

3.4 Evidence

In this section, we document that the central implications of the model are consistent with the data, by showing that contractual forms as well as the actual sharing of

[39]Other projects where the objectives are relatively easily defined include CAD projects and migration of an existing software from a platform to another.

the overrun vary with characteristics likely to be correlated with the reputation of the firm. We then examine how the other predictions of the model match with the software data. Finally we consider some obvious alternative explanation of the patterns observed in the data.

3.4.1 Sources of reputation

We begin by describing alternative sources of reputation for the firm. There are a number of ways in which a firm can acquire a reputation for being reliable. We list them below (the relevant data is presented in table 4:[40]

First, we have already argued (in the introduction) that we will think of age as an important source or reputation. Firms which manage to survive in the market for a long time are more likely to be "good firms". In the framework of our model, it means that they are less likely to act unreliably. While we do not formalize the process which leads to the elimination of unreliable firms from the pool of firms, it would be straightforward to do so. Essentially, if unreliable firms are discovered with positive probability and nobody wants to deal with them after they are found out, then as firms age, the probability that they are unreliable will decline.

Second, ISO certification can potentially also give a reputation to a firm. ISO certification is awarded by international or Indian agencies, themselves accredited, which examine that the processes of software production in the firm follows some approved routines. In particular, the firm must follow specified procedures to report on the progress of the software and to perform the tests. Consequently, the software development process should be easier to monitor for ISO certified firms. Moreover, ISO certified firms are monitored every once in a while, and lose the certification if they cannot prove that they followed the approved methods. This should give

[40]Note that we think of the reputation as being a attribute of the firm, more than of the individuals who compose it. It could be that an experienced professional leaving his job to create a software firm takes his individual reputation with him. It turns out that individual reputation seems difficult to transport (we asked what the past career of the person who founded the software firm was, and examined whether this was related with sharing of the overrun, but did not find that this was the case). The main reason is that the important input the boss of a software firm has to provide is the management of the team, which may or may not be related to his ability as a software professional.

every incentive to the ISO certified firms to stick to standard procedures and report problems reliably. As we noted in the introduction, firms in the industry are currently very keen to acquire ISO certification, precisely because they think that it will improve their reputation. 19% of the external contracts in the sample were done by ISO certified firms. ISO certified firms tends to be older firms (only 9% of the young firms have ISO certification).

Third, trust established in repeated relationships can play the same role within a relationship that reputation plays on the broader market. When a client has gone through a project with a firm once, he has gathered information about it, and therefore, if he decides to work with the firm again, he must have a better opinion of it than the average opinion it has on a firm of the same type. This will include an opinion about the general ability of the firm (which will certainly affects contracts and the choice of projects in many ways), but also a better idea of its reliability, which will play within the relationship the same role as a reputation. 41% of the contracts in the sample involved a client with whom the firm had worked already. This proportion is roughly the same among young and old firms.

Finally, we contrast internal (projects for OSDC and mother companies) and external projects. The informational problems that we mentioned are greatly mitigated in internal relationships. Therefore we should find the same kind of differences (but perhaps even stronger) between internal and external projects as between projects of old and young firms. However we need to be sensitive to the fact that companies working for internal clients are potentially very different from other companies. In particular, OSDC will be established only after the client has spent a very long time studying the firm. Fully owned subsidiaries are often run by people who had been previously working in the US office of the firm. We therefore restrict the comparisons to firms that perform some internal projects (e.g. subsidiaries that works for their mother company and also for external client). This insures that the selection of firms for internal projects does not invalidate the comparison (since all firms in this subsample have been selected for some internal work).

3.4.2 Choice of contract and sharing of the overrun

Structure of the contracts

An implication of the model is that contractual forms will be restricted to contracts where the ex ante rule is that firms will bear either all or nothing of the cost overruns. This implication rests on the particular assumption we have made (in particular on the fact that we don't allow for risk aversion), but it matches well with the observed pattern.

As we describe in section 2, there are three major types of contracts: fixed cost, time and material and "mixed" contracts. Fixed cost contracts are linear contracts with $s = 1$. As we discuss above time and material contracts are similar to such contracts, with $s = 0$. In mixed contracts, the initial agreement specifies a payment for the specifications only. At the end of the specification phase, another agreement is specified for the development and testing phases. This kind of contracts effectively splits the projects into two subprojects. For each of them, a separate sharing rule is chosen, which is either 1 or 0 (often, time and material for the specification phase and fixed cost for the subsequent work).

In other words, mixed contracts are a juxtaposition of fixed cost and time and material contracts. It is easy to understand why, when the project is broken into these two phases, specifications tend to be written on time and material and the rest of the work tends to be done using a fixed cost contract. In the specification phase, the potential for the client to generate an overrun is extremely large. In particular when the firm first sends the specifications, he can pretend that the specifications written do not correspond to what he wanted. The whole effort of the firm until that point becomes in effect useless. Therefore, it is important to give the client higher powered incentive. On the other hand, at the time the second sub-contract is written, a large part of the uncertainty about what the client really wants is resolved, since it has agreed (in writing) to the specifications. Therefore a fixed cost contract, which give better incentives to the firm, can become optimal from that point on. In practice the choice of the contract for the second phase of the project is often endogenous:

if the firm feels that a substantial amount of uncertainty remains, it can in general insist on getting a second time and material contract. Mixed contracts are therefore ex ante more constraining for the client than for the firm.

Reputation and the choice of contract

The reputation of the firm determines both which contract it will get (choice of s) and what share of the overrun it will end up paying (actual s^*). Firms without a reputation will be more likely to be have fixed cost contracts than time and material or mixed contracts. Conditionally on having fixed cost contracts, firms without a reputation will bear more of the overrun than firms with a reputation. The combined effect of the two is of course that firm without a reputation will bear a larger share of the overrun.

This subsection presents data related to these implications.

We presented evidence that age does matter in the introduction, as a motivation for this project. It is illustrated in figure 1 and 2. The proportion of fixed cost contracts and the share of the overrun borne by the firm are increasing with the foundation date of the firm. Table 5 shows the means of the firm's share of the overrun for each type of firm, and the difference between low and high reputation firms. In column (1), we report the mean for the sample of external firms. In columns (2) to (4), we show the contrast between young firms (created in 1994 or after) and old firms (created in 1993 or after. Young firms are significantly more likely to have fixed cost contracts (the probability is 26% higher). They bear substantially more of the overrun (19%). This holds within the projects with fixed cost contracts (the difference is 13%).

The pattern is less clear for ISO certification: ISO-certified firms are not less likely to get fixed cost contracts and they do not pay for a lower fraction of the overrun in general. However, conditionally on doing fixed cost contracts however, they bear less of the overrun (20.4%).

A relationship with a client has the same effect of a general reputation. Firms engaged in a repeated relationship with their client are about as likely to than other firms to have fixed contracts, but they pay significantly less of the overruns (20% less).

Finally, among firms who have internal contracts, firms pay for more of the overrun when they deal with external clients than when they deal with internal clients. Almost half of their external contracts are fixed cost contracts (a number close to the proportion of fixed cost contracts among old firms), whereas only 23% of the internal contracts are fixed price contracts. They pay a much smaller share of the overrun (20% instead of 47%) in internal contracts than in external contracts. The difference conditional on doing fixed cost projects is not significant, but this is probably due to the small number of fixed price contracts among internal projects.

In summary, it seems that young firms, firms working with a new client and firms working with an external client bear a larger share of the overrun compared respectively to older firms, firms engaged in a repeated relationship and firms working for an OSDC or their mother company. We interpret these results as showing that reputation does influence the way the overruns are shared between the client and the firm. We will address some alternative explanations below, but the first possible caveat to this interpretation is that these firms do different types of projects, which require different types of incentives or entail different types of risk. For example if old firms do mostly project that entails the possibility of very large overruns, they may refuse doing the project unless they know they will be covered in case this happens.[41] In particular, table 4 shows that young firms, non ISO-certified firms, and firms working for external clients do on average smaller and simpler projects than old firms, ISO firms and firms working for internal clients. It is therefore important to check that the simple contrast between the groups is not an artifact of the different composition of their contracts.

In table 6, we show the differences between the overrun paid for by each type of firms in project-size cells (panel B) and complexity cells (panel D).[42] The first panel reproduces the uncontrolled difference of table 5. In the panel C, we show the

[41] We will comment more on the choice of project per se below.

[42] We used the subjective complexity measure given by the firms.

"controlled contrast": this is simply a weighted average of the differences between the young and old firms in the project size cells, where the weights are given by the fraction of projects falling into this project size cell. This is a crude way to take into account the two facts that different types of firms choose different type of projects and that the differences across young and old firms are not necessarily the same for all project sizes.[43]

Firms tend to bear less of the overrun when they do complicated project than when they do simple projects. There is also a weak relationship between the size of the project and the share of the overrun paid for by the firm. Young firms pay a larger share of the overrun than old firms for small and for large projects, but not for medium-sized projects. The controlled contrast between young and old firms is slightly smaller than the simple difference, but still high. The controlled contrast becomes positive, though insignificant, for ISO-certified firms, mainly because that ISO-certified firms doing small projects don't pay any of the overrun.[44] Controlling for project size does not affect the difference between repeated and new clients and between internal and external contracts. Whatever the complexity of the project, young firms bear more overrun than young firms, firms working with new client bear more overrun than firms working with a repeated client, and firms bear less overrun when they do internal projects. The evidence for ISO certification is, once again, mixed.

In summary, even after taking into account the size of the projects, firms with low reputation bear more of the overrun than other firms (although the evidence in favor of ISO certification remains less than overwhelming). A final piece of evidence is presented in table 7. In this table, we examine whether different kinds of reputation are good substitutes. Namely, we ask in panel A whether young firms still bear more of the overrun when they benefit from another kind of reputation. In the table we present

[43]Following is a simple example: if old firms pay the same share of the overrun than old firms for small and for large projects, but both pay more for small projects than for large projects, and if small firms are more likely to do small projects, then the simple contrast would be positive due to the selection, whereas the controlled contrast would be 0.

[44]This number should be taken with caution, as very few ISO firms do small projects.

the difference between young and old firms in the proportion of fixed cost contracts (line 1) and in the share of overrun they pay (line 2) within groups of ISO-certified /non ISO-certified firms, repeated/new clients, internal/external contracts (for firms who do some jobs for internal firms). The contrasts are interesting. Non ISO-certified young firms bear 27% more of the overrun than non ISO-certified old firms, but among ISO-certified firms, there is no difference. Young firms are significantly more likely than old firms to have fixed cost contracts if they work with a new client, but not if they have already worked with this client. Among firms that do some internal contracts, the same contrast appears: young firms are more likely to have fixed cost contracts when they work with an external client, but not when they work with an internal client or an OSDC. Finally, the difference between young and old firms in the share of the overrun the firm pays is larger for contracts with new clients compared to contracts with repeated clients and for contracts with external clients compared to contracts with internal clients. In panel B, we perform the same exercise, but we look at how the difference between the share of the overrun paid for by firms working with a new rather than repeated client varies across different kind of firms. Interestingly, a very different pattern emerges. The difference between new and repeated client subsists for old firms and for ISO certified firms, and does not decline. It suggests that the mechanism of reputation formation is rather inefficient: even after a firm has been in the market for some time, much remains to be learnt about the it.

We have documented systematic differences in the way cost overruns are shared across young and old firms, contracts with repeated and new clients, ISO certified firms and other firms, and firms in internal and external contracts. This evidence is consistent with a model where reputation is an important determinant of the contracts and the sharing of the overrun. In the next subsection, we examine the whether the other empirical predictions of the model also hold.

3.4.3 Further results

Choice of project

A simple extension to our model also predicts that the firms with a low reputation will tend to choose simpler projects, where the objectives are easier to define, which will tend to limit the overrun generated by the client. We present evidence relating to the choice of project in table 4, and figure 5, 6 and 7. Young firms do smaller projects, which have smaller overrun (even expressed in proportion of predicted costs).[45] They also tend to carry more often "simple" projects (Y2K, CAD, Data manipulation), which generate lower overruns, are easily defined, and are easier to monitor. We have also asked them to subjectively rate the complexity of the project, and even according to this subjective measure, young firm do more simple projects. As a result of these two combined facts, the returns to each project (cost multiplied by markup) is smaller on average for young firms than old firms.

This could be at least partly explained by the fact that young firms are on average less competent (and that therefore clients do not want entrust them with large or complex projects). However the same contrast holds between internal and external projects. Since we have restricted the comparison to firms that do at least some internal work, the difference between internal projects and external projects is not tainted by this bias. This confirms that part of this difference between young and old is due to difference in behavior.

Mean and variance of the overrun

As a result of the different choices of projects, the model's predictions about mean and variance of the overrun are not unambiguous. The mean of the overrun generated by the firm and of the overrun generated by the client could decrease or increase with reputation. The basic model implies that overrun generated by the firm should, on average, be larger for firms with lower reputation (and so should total overrun),

[45]Moreover by doing that they keep the share of overrun accounted for by each project more or less similar across young and old firm: this could therefore be explained by adding risk aversion to the model.

but this could be compensated by the fact that firms with low reputation choose on average simpler projects. The variance of the overrun generated by the firm is more likely to decrease with the reputation of the firm (especially if reliable firms with low reputations try to signal their type), but the model does not make any prediction about the variance of the total overrun. As we will see below, alternative models would on the other hand have strong predictions on the mean or the variance of the overruns.

In table 8, we present the mean and the variance of the overruns caused by the firm and of the total project overrun, unconditional and controlling for project size. In table 10 and 11, we present them by project size categories. The mean and the variance of the total overrun do not seem to vary with firms reputation in any systematic way. If anything, overruns are actually smaller for young firms, non-ISO firms, and external projects (but except for the last one, these differences are not significant). The variance of the total overruns are similar. If we restrict attention to the share of the overrun due to the firm (internal difficulties and ambiguities), young firms do seem to cause slightly more of these overruns. The difference between overrun caused by young and old firms, controlling for project size, is 2.6 %. (the sample mean of overrun caused by the firm is 7.67 %). Overruns caused by the firm are also smaller for projects realized for a repeated client. For a given project size, he variance of the overrun caused by the firm is also larger for young firms and firms working with a repeated client (cf tables 10 and 11). These difference are significant. These two facts (small differences in mean of the overrun caused by the firm but large difference in variance) are consistent with our model, extended to take into account reputation building concern and choice of project.[46]

The data seems consistent with our model of how reputation determines contractual outcomes. In the next section, we examine the most obvious alternative explanation to the observed pattern.

[46]Once again, we find that ISO firms do not perform particularly well: they generate large overruns than non ISO firms (although the variance is similar).

3.5 Alternative interpretations of the data

This section reviews alternative explanations to the pattern observed in the data (in particular to the main result that young firms bear a larger share of the overrun than old firms).

3.5.1 Pure Risk Sharing

One possible interpretation of what is going on this industry is pure risk sharing. However, as explained in much greater detail in a previous version of this paper, this interpretation very quickly runs into trouble. In the case where we assume CRRA preferences (which is standard in cases, like this, where there is substantial variation in the size of the contracting parties) we showed in the previous version of this paper that

Claim 9 *If the firm and the client have CRRA preferences, for a fixed project size, the share of the risk that they each bear will be approximately in the inverse proportion of their coefficients of relative risk-aversion, keeping fixed the ratio of their total revenues. It will also be approximately in the direct proportion of their total revenues, keeping fixed the ratio of their risk-aversions.*

Given that the client's revenue are much bigger than that of the firm,[47] an implication of his proposition is that the client should bear most of the risk unless the client is much more risk averse than the firm. In fact, the firm bears on average more than half of the cost overrun, suggesting that the client's coefficient of risk-aversion must be very large relative to the firm's. It is however difficult hard to think of a basis for such differences in risk aversion.

[47] More than half of the contracts in the sample are with "large" clients, 26% are with medium sized client, and 19% are with small clients. Large clients are in general fortune 500 companies or equivalent. Small clients are firms with turnover below \$10 million. In contrast, the median turnover of the software companies in the sample is only \$1.2 million, and the largest firm had a turnover of \$47 million. Only 27% of the firms have a turnover above \$ 10 million. Among the firms engaged in contracts with small clients, the median firm has a turnover of \$ 0.5 million.

Moreover, this result has systematic predictions about the relationship between firm size and the share of the risk that it bears, controlling for client size and project size.

Table 9 presents the share of overrun paid by the firm by client size, project size and firm age.[48] In all project size-client size cells, old firms bear less of the overrun than young firms. Since old firms are on average larger (this is shown in figure 4 and in columns 2 to 4 in table 4: old firm's turnover is larger by $ 3.7 million, or more than 100%, than young firms), this contradicts the basic implication of the risk-sharing model.

One might also speculate that old and new firms generate different risk profiles and that this explains why old firms bear less risk: perhaps old firms simply generate less risk. However, the evidence on the standard deviation of total overrun presented in table 8 shows that this is not the case. The standard deviations of total project overrun are very similar across all types of firms. There is therefore no evidence to support the view that young firms are systematically more risky to deal with. Another possibility, however, would be that the underlying distribution of overrun is different for young and old firms (despite the fact that mean and standard deviation are different), and that the particular form the the risk faced by old firms made this particular risk sharing rule optimal. We examined the entire distribution of overrun generated for both young and old firms. The two distributions are very similar, except for four old firms which generated very large overrun (150% and higher). These four old firms are however not driving the results, since all of them paid 100% of the overrun. Moreover, as we have shown above, the difference between old and young firms is maintained when we control for project type (complexity or size), which are presumably good indicators of project-specific risk.

The evidence we give above strictly only applies to the case of CRRA preferences. There are, of course, many classes of risk preferences which do not fall into this

[48]To maintain legibility and avoid constructing cells few observations, we have divided the projects into larger and greater than median, grouped small and medium clients, and separated the firms between "young" (created in 1994 or later) and old (created in 1993 or before).

category. However, there are two basic intuitions which suggest that these other preferences will not work particularly well either: on the one side, if the coefficient of *absolute* risk aversion falls faster than a CRRA, it is very hard to explain why the firm bears any risk at all. On the other side, if the coefficient falls slower than a CRRA (so that the preferences approach the CARA model), it can be shown that project size and client size effects also become smaller and this leaves very little to explain the inter-firm differences.[49]

3.5.2 Varying Levels of competence

Our model has assumed that both clients and firms are risk-neutral. Suppose we now assume that firms and clients make mistakes which lead to overrun but that these mistakes can be contracted upon. In this case, one possible first best contract is one in which firms take the full responsibility of any mistake that they make.[50] Now if young firms[51] are on average less competent, then it is to be expected that they would pay on average for more of the overrun.

The first point against this explanation is simply that risk-neutrality is a very extreme assumption. If the firm was at all risk averse then the optimal contract would try to insure the firm against all sources of risk that are beyond its control. Therefore, since young firms do not choose to be incompetent, they should be insured against overrun that results from their mistakes. Of course, the extent of such insurance may well be limited by the client's willingness to bear risk, but as we have already argued, the client is in a much better position to bear risk than the firm, and in particular small firms should only bear a small part of the risk. Of course, this assumes that the mistakes are not made deliberately. The case of varying levels of moral hazard will be examined below.

There are also some simple empirical arguments against this view: first (and most importantly), firms pay much more of the overrun than the share for which they are

[49] In the extreme case of CARA preferences neither project size nor client size affects risk-sharing.

[50] Of course the actual contracts do not say anything about dividing the overrun. Therefore what we are referring to here is a fully efficient implicit contract.

[51] Or more generally, firms that we have called so far "low reputation" firms.

responsible: as shown in table 3, even when the client is fully responsible for the overrun, the firms still pay on average 51% of it. Second, to explain the differences in the share of overrun paid by the firms entirely by differences in the share of overrun that is caused by the firm, it would have to be the case that young firms cause substantially more overrun than old firm do. Recall that the total overrun are, if anything, larger for old firms than young firms. The difference between the share of the overrun due to young firms and the share of overrun due to old firms is clearly not large enough to explain difference in the sharing of these additional costs. Differences in sources of the overrun are shown in table 8. If we look at the difference between the fraction of overrun due to young firms and that due to old firms, it is only 1.18 (and not significant)%[52] , but they pay 20% more of the overruns. Finally, even within firms that do some internal work (and are therefore more homogeneous), it is the case that firms bear more of the overrun in external contracts than in internal contracts. Therefore it does not seem to be the case that the differences in the share of the overrun borne by young firms can be explained by systematic differences in competence between young and old firms.

Underbidding by young firms

One could imagine that even in a world where contracts are effectively complete, young firm might systematically underbid (quote a price based on intentionally low estimates) to win the project. Of course the client knows this, and in the optimal contract corrects for it by holding them responsible for the extra overrun resulting from the underbidding.[53] Young firms therefore end up paying for a higher share of the overrun. However it should be easy to see that the same objections that we list above to the competence-based explanation also apply in this case.

[52]This is obtained by adding the difference in the share of overrun due to ambiguity and the share of overrun due to internal difficulties.

[53]Again we are referring here to a fully efficient implicit contract.

Varying level of moral hazard

Of course, this assumes that the mistakes (or prediction errors) are not made deliberately. Once we allow for such moral hazard, the client may well not be willing to insure the firm. Old firms could be less prone to moral hazard than young firms, and therefore bear less risk than young firms. The evidence we have presented on meand and variance of the overrun cannot be used to answer to this possible criticism, since they are endogenous: young firms could be generating the same level of overruns as old firms do precisely because they face higher punishments. Note however that this would be a real alternative explanation to the facts only if the levels of moral hazard were common knowledge (in the opposite case, it is just a restatement of what we assume in our reputation model). In this case, it is not clear why it should make any difference that firms are working with a new or a repeated client. Their level of moral hazard would not have the time to change between the first and the second contract realized for a client. The fact that firms are treated differently the second time around must therefore indicate that there is learning going on about the characteristics of the firm.

Varying levels of honesty

Young firms could also differ from old firms in their propensity to report costs honestly. They could be more prone to try report inflated costs, or to pretend that changes due to their own incompetence are due to the client changing his mind. If the client could not tell the cheaters apart, then the analysis of such a model would be similar to the analysis of the model we propose, and lead to the same conclusion (the reputation of old firms would be a reputation for honesty instead of a reputation for reliability). As we mentioned earlier, our modeling choice was to model a reputation for reliability, but it is clear that the analysis could be carried out with a different reason for the importance of reputation.

Note however that if clients could tell apart cheaters and honest firms, and punish cheaters by imposing them to pay more of the overrun, then we would also observe that young firms would pay on average more of the overrun (but this would not result in any social cost, unlike in our model or a version of the model with a reputation for honesty). Assuming that firms report in the questionnaire what they have reported to the client, then the evidence that young firms pay more often than old firms overruns reportedly caused by the client would simply reflect the fact that they are lying more often then old firms. Because this argument rests on the fact that firms are lying in what they report to us as well as in what they say to their client, it is not easily verified or invalidated in the data. Note however that this argument implies that the clients never make any mistake in telling apart cheaters and honest firms. It is therefore self-defeating: why would firms cheat in the first place if they know that they are going to be found out?

Moreover three facts are difficult to reconcile with this explanation:

First firm pay on average 50% of the overrun when they report that the client is fully responsible for it. The suggested explanation would therefore implies an implausibly high fraction of cheaters among Indian software companies (young and old).

Moreover, if the client has perfect information and can enforce any sharing ex-post, there should be no variation in the contractual form, or at least it should not be related to the final outcome. However, firms pay more of the overrun when they have fixed costs contracts then when they have time and material contracts. Furthermore, young firms have more often fixed cost contracts. Therefore the ex-ante contracts seems both to be relevant and to be used by the clients, which is not consistent with the world we just described.

Finally, note that such a model would not explain the difference between contracts with repeated and new clients, or the difference between internal and external contracts: if the client has perfect information, then it is not easy to explain why firms would behave differently when dealing with different types of clients.

3.6 Conclusion

We set out in this paper to look for evidence that reputation plays an important role in determining contractual outcomes. We find that the evidence seems to strongly support this view, though given that the evidence is indirect (we do not actually observe people looking at reputation when deciding on contracts) and there are important firm characteristics that are potentially correlated with our measures of reputation, some doubts clearly remain.

The conclusion that reputation matters is of course important in itself: it gives support to a range of theories that are based on limitations of contracting. Moreover it might suggest an explanation of why the Indian software industry is not much larger (Indian Software export were only worth 3.4% of the 1995 worldwide outsourcing business) given its obvious labor-cost advantage[54] and the fact that this is a very labor-intensive industry. Or, to state the same point differently, why is it an equilibrium for software professionals in India to get paid so much less than their US counterparts? Reputation at the firm level is one possible explanation: most Indian firms are simply not trusted enough to be given important contracts. While our evidence cannot directly substantiate this view, the fact that reputation is important within the Indian industry suggests that it also ought to be important when an American client is deciding whether to go to a firm in India or to one in the US.

To add support to this view, our results also suggest that the process of reputation formation is rather inefficient. This is reflected in the fact that after controlling for age, whether or not a firm is dealing with a repeat buyer still makes a substantial difference to the contract. In other words, repeat buyers clearly know much more about the firm than the market does. In other words, the fact that a firm performed well in the past vis a vis one firm takes time to become public information. This is of course consistent with rational behavior on the part of the client but it clearly hurts the firm.

[54]The U.S. imports a very large number of Indian software professionals for short-term assignments at a cost of more than twice what they would earn in India.

The policy implication of this view is that a credible system for rating firms modeled on credit rating systems may play an important role in the evolution of industries like the software industry where contracting is inherently problematic, by making it possible for the market to efficiently aggregate all that is known about each firm.

Bibliography

[1] Aghion P. and J. Tirole (1994) "On the management of Innovation" *Quarterly Journal of Economics*109:1185-1207.

[2] Baker G., R. Gibbons and K. Murphy (1997) *Implicit contracts and the theory of the firm*

[3] Barron J. and John Umbeck (1984) "The effect of different contractual arrangements: the case of retail gasoline market" *Journal of Law and economics* 27:313-328.

[4] Bowles S. (1998) *Egalitarianism and the evolution of Group-beneficial norms*

[5] Confederation of the Indian Industry (1997) *Financing the software Industry*

[6] Crocker K., and Scott Masten (1988) "Mitigating contractual hazard: unilateral option and contract length" *RAND Journal of Economics* 19.3:327-343.

[7] Crocker K. and K. Reynolds (1993) "The efficiency of incomplete contracts: an empirical analysis of air force engine procurement" *RAND Journal of Economics* 24.1:127-146.

[8] Genesove D. (1993) "Adverse selection in the wholesale used car market" *Journal of Political Economy* 101-4:644-665.

[9] Greif A. (1994) "Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies" *Journal of Political Econony* 102.5:912-950.

[10] Heeks R. (1996) *India's Software Industry. State Policy, liberalisation and industrial development.* Sage Publication, New Dehli, Thousand Oaks, London.

[11] Holmstrom, Bengt (1982) "Moral Hazard in Teams" *Bell Journal of Economics* 13.2:324-40.

[12] Joskow, P., (1987) "Contract duration and relationship specific investment: empirical evidence from coal markets" *American Economic Review* 77-1:168-185.

[13] Kandori, M. (1992) "Social norms and community enforcement" *Review of Economics Studies* 59:63-80.

[14] Kreps, D. and R.Wilson (1982) "Reputation and Imperfect Information" *Journal of Economic Theory* 27-2:253-79.

[15] Lafontaine F. (1992) "Agency theory and franchising: some empirical results" *RAND Journal of Economics* 23.2:263-283.

[16] Lafontaine F. and K. Shaw (1996) *The dynamics of franchise contracting: evidence from panel data* NBER WP 5585.

[17] Lafontaine F. and M.Slade (1998) *Incentive contracting and the franchise decision* NBER WP 6544.

[18] Lerner J. and R. Merges (1997) "The Control of techonology alliances: an empirical analysis of the biotechnology industry" MIMEO Harvard Business School.

[19] Maskin E. and J. Tirole (1998) *Unforeseen contigencies and incomeplete contracts* MIMEO Harvard and IDEI.

[20] Masten S. and K. Crocker (1985)"Efficient adaptation in long-term contracts: Take-or-pay provisions for natural gas" *American Economic Review* 75.5:1083-1093.

[21] Mc Affe and Shwartz (1994) " Multilateral vertical contracting: opportunism, nondiscrimination and exclusivity" *American Economic Review* 84:210-230.

[22] Monteverde K. and D.J. Teece (1982) "Supplier switching costs and vertical integration in the automobile industry" *Bell Journal of Economics* 13.206-213.

[23] NASSCOM (1997) *Indian Software Directory* CD-ROM.

[24] Pittman R.(1991) "Specific investment, contracts and opportunism: the evolution of railroad sidetrack agreements" *Journal of Law and Economics* 34:565-589.

[25] Pressman Roger S. *Software Engineering, A Practitioner's Approach*, 4d edition, Mac Graw Hill International editions, Software Engineering Series.

[26] Shepard, A. (1993) "Contractual form, retail gasoline, and asset characteristic in gasoline retailing" *RAND Journal of Economics* 24.1:58-77.

[27] Siwek and Furchtgott-Roth (1995) *International Trade in Computer Software* Quorum Books Wesport, London.

[28] Subramanian (1992)*India and the Computer* Oxford University press, New Dehli.

[29] STPI (1997) *Software Technology Parks of India*, MIMEO, Bangalore (1997).

Figure 1
Proportion of fixed cost contracts

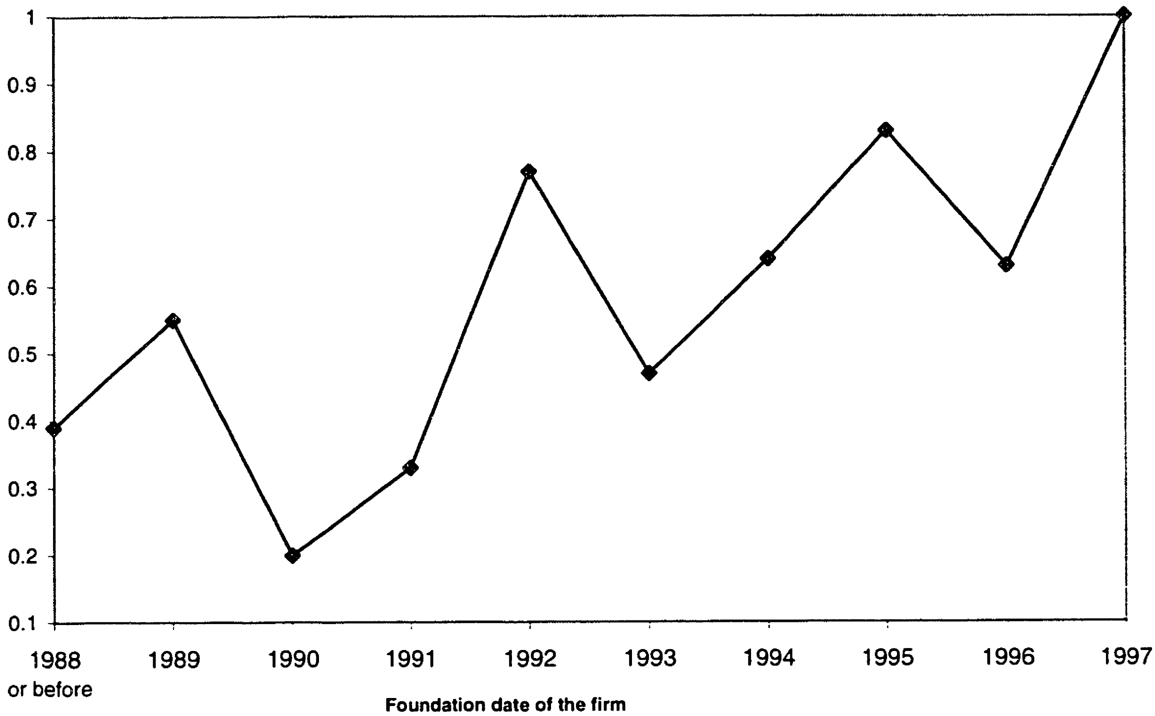


Figure 2
Share of overrun paid for by the firm

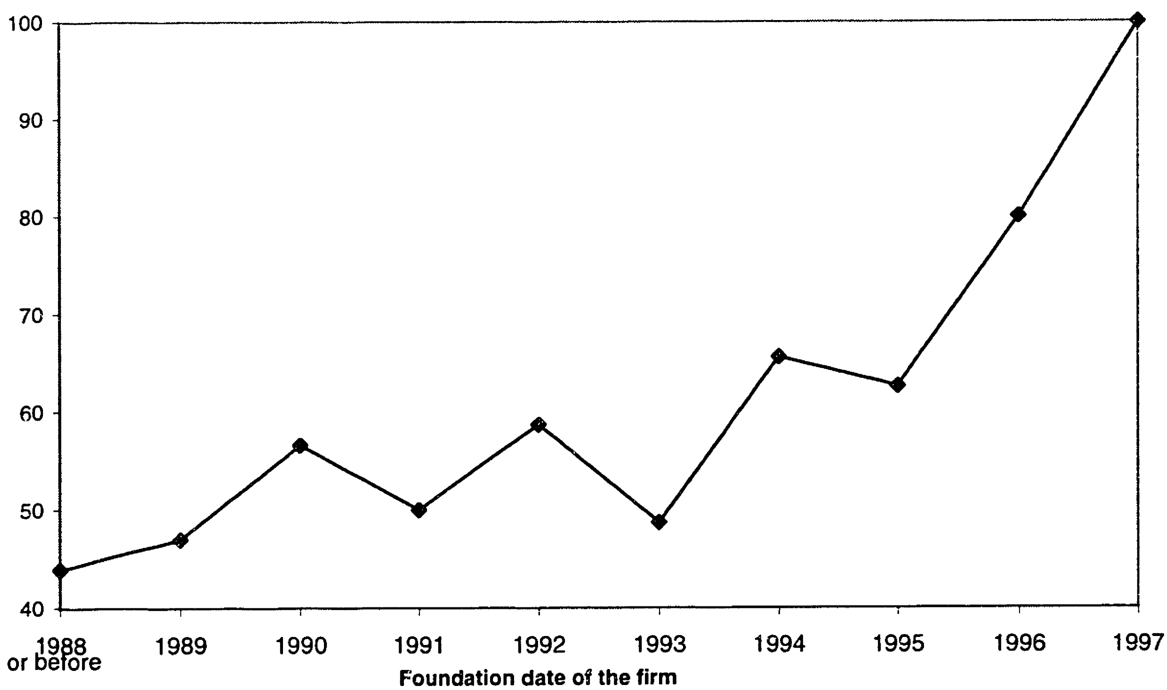


Figure 3:
Mean of project overrun
(percentage of initial evaluation)

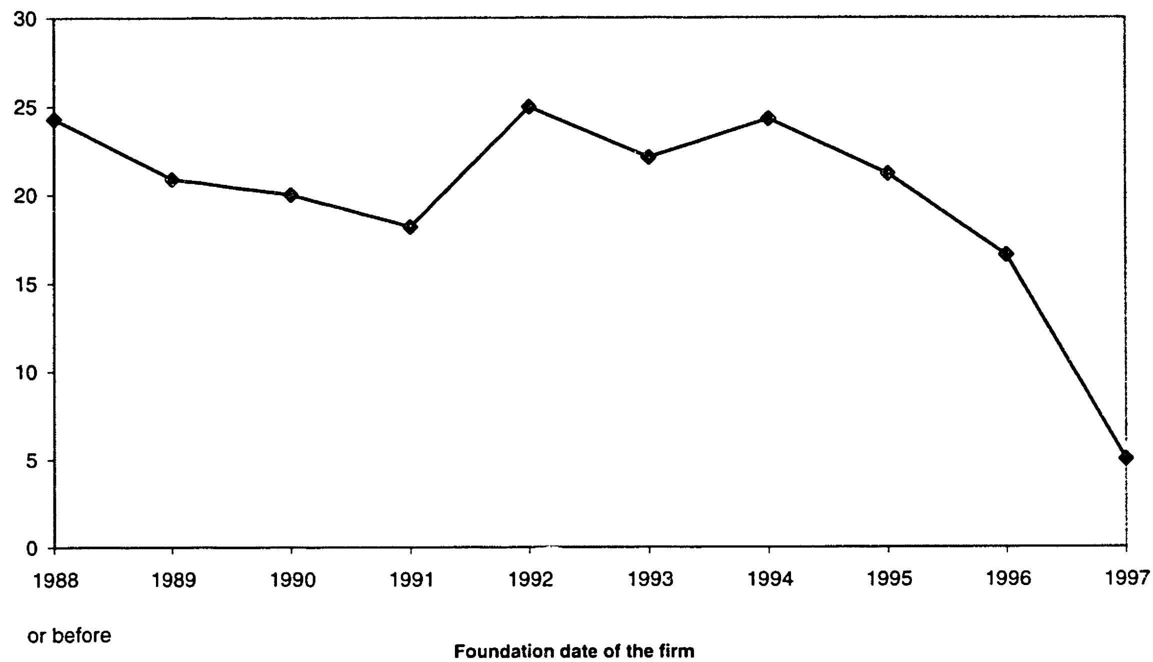


Figure 4
Average Firm Turnover in 1997/98 (Millions $US)

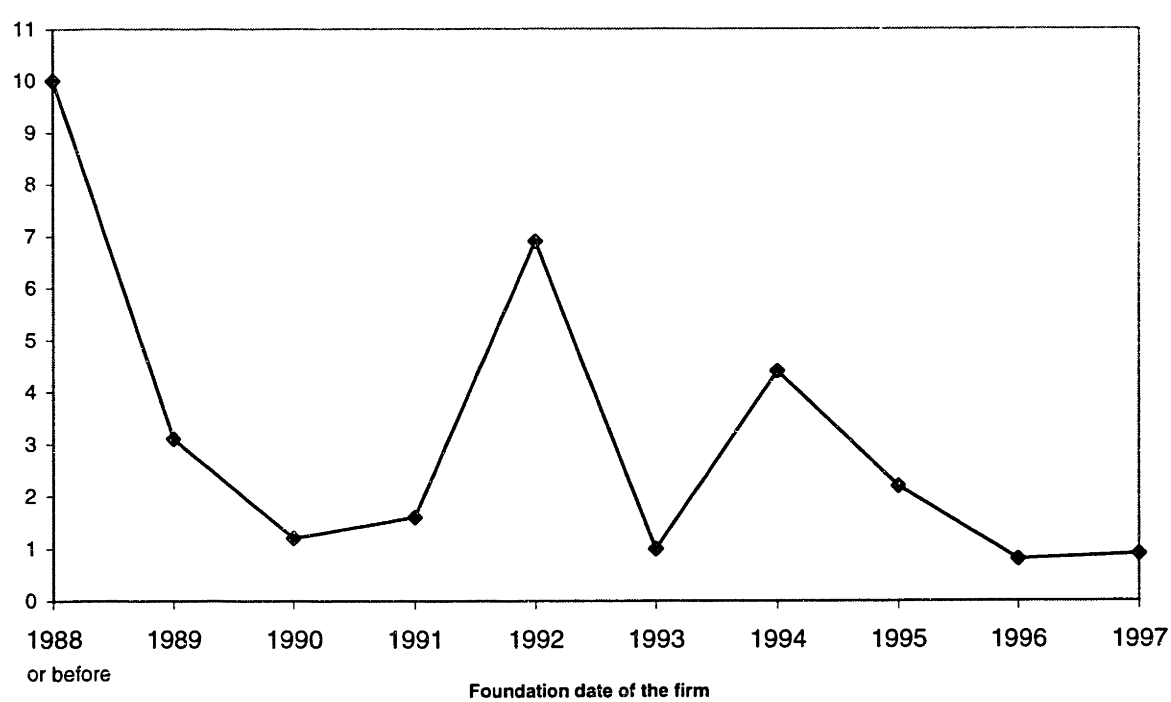


Figure 5:
Subjective complexity measure

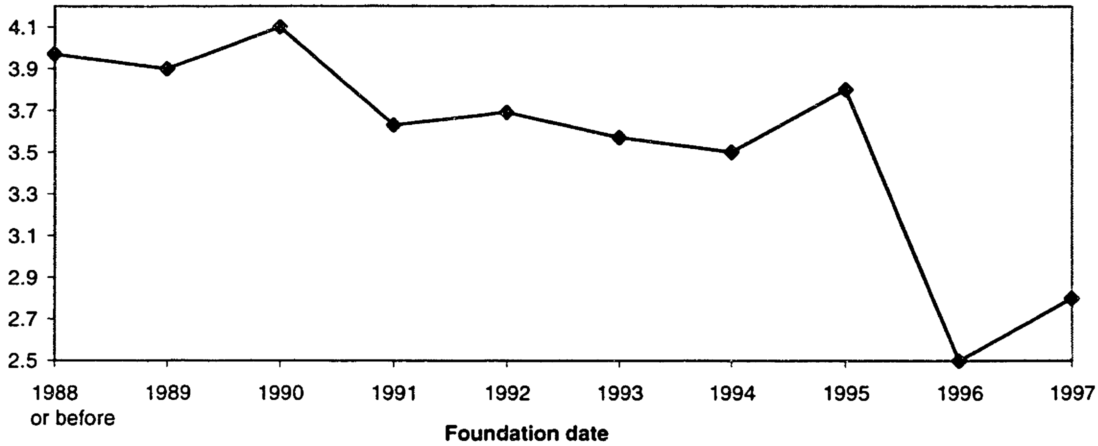


Figure 6:
Size of the project (in man-months)

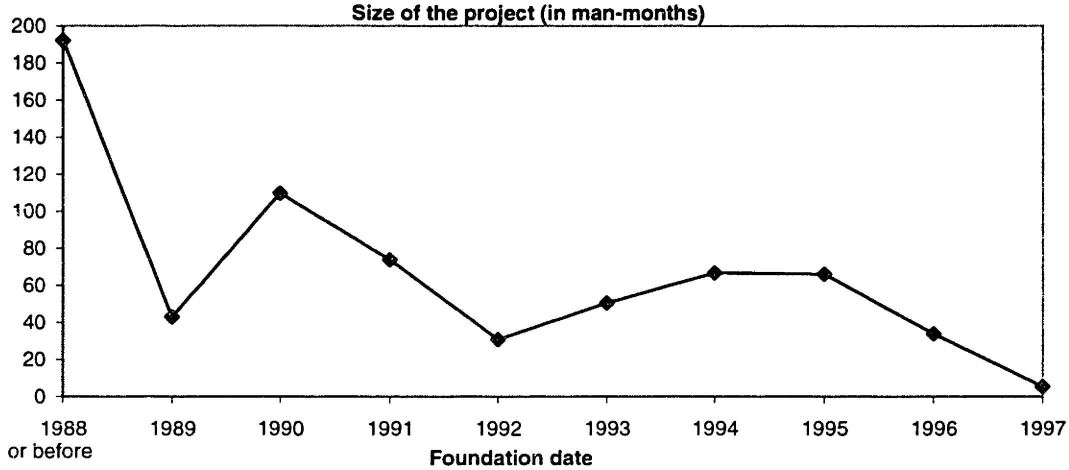


Figure 7:
Proportion of "simple" project
(cad, y2k, web pages, data manipulation)

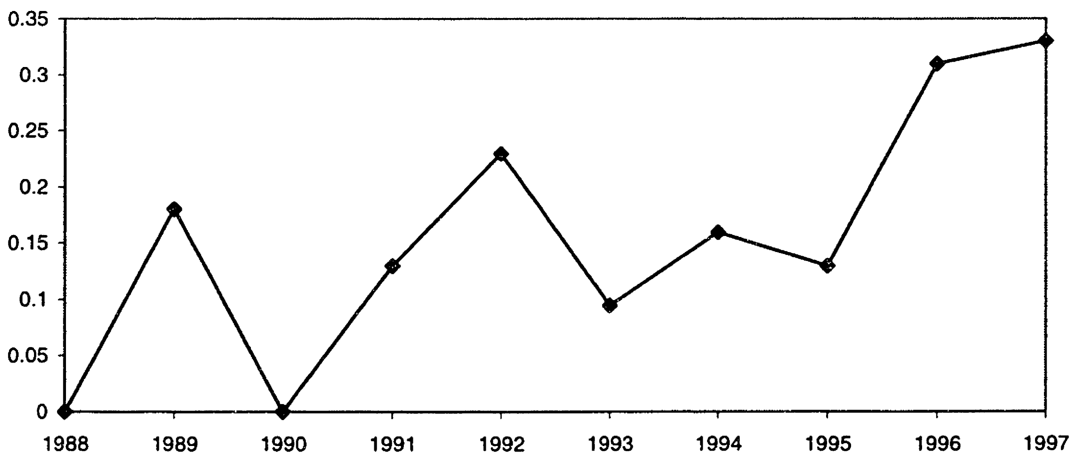


Table 1
Descriptive statistics

	Complete Sample (N=236)					External Projects (N=167)				
	Mean	Median	Standard deviation	Minimum	Maximum	Mean	Median	Standard deviation	Minimum	Maximum
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PANEL A: PROJECT CHARACTERISTICS										
Project size (man month)	113	26	403	0.38	5400	74.7	27	141	0.38	960
Proposal effort (% of project effort)	2.11	0.63	4.5	0	50	2.55	1.25	4.96	0	50
Specification effort (% of project effort)	13.8	10.0	13.0	0	100	13.5	10.0	14.0	0	100
Project cost >estimate (overrun)	0.74	1	0.44	0	1	0.75	1	0.43	0	1
Project cost =estimated cost	0.22	0	0.41	0	1	0.22	0	0.42	0	1
Project cost <estimate (underrun)	0.046	0	0.21	0	1	0.023	0	0.15	0	1
Total overrun (% of project cost)	24	15	34.4	-25	250	22.7	15	30	0	200
Causes of overrun:										
% due to ambiguity	20.4	0	33.5	0	100	19.5	0	33.6	0	100
% due to changes	48.1	50	41.9	0	100	45.4	50	42.7	0	100
% due to internal difficulties	7.8	0	22.4	0	100	9.06	0	25	0	100
% due to delay by the client	12.7	0	25.8	0	100	12.8	0	26.2	0	100
% due to other reasons	9.92	0	26.3	0	100	11.9	0	29.1	0	100
Project was delivered late	0.19	0	0.4	0	1	0.2	0	0.4	0	1
PANEL B: CLIENT										
Project is for the mother company	0.23	0		0	1	0				
OSDC project	0.063	0				0				
External project	0.71	1		0	1	1				
PANEL C: CONTRACT										
Time and material contract	0.3	0	0.46	0	1	0.15	0	0.36	0	1
Mixed contract	0.21	0	0.41	0	1	0.26	0	0.44	0	1
Fixed cost contract	0.48	0	0.5	0	1	0.58	1	0.49	0	1
Share of overrun paid for by the firm	47.4	38.3	46.3	0	1	57.1	76.5	45.1	0	1

Table 2
Share of overrun for paid by the firm as a function of initial contract

	Proportion paying 100%	Proportion paying between 0 and 100 %	Proportion paying 0%	Average share of the overrun paid for by the firm
	(1)	(2)	(3)	(4)
PANEL A: ALL CONTRACTS				
All contracts	39.29	17.85	42.86	47.4
				(3.58)
Time and material contracts	11.1	11.12	77.78	15.6
				(4.92)
Mixed contracts	34.29	25.71	40	51.6
				(7.69)
Fixed cost contracts	54.12	18.82	27.06	63.1
				(4.83)
PANEL B: EXTERNAL CONTRACTS				
All contracts	47.58	20.97	31.45	57.1
				(4.06)
Time and material contracts	17.65	23.53	58.82	28.2
				(9.56)
Mixed contracts	40.62	25	34.38	50.9
				(8.08)
Fixed cost contracts	56.76	18.92	24.32	65.8
				(5.07)

Table 3
Share of overrun paid by the firm and source of the overrun

	Share of the overrun paid for by the firm			OLS coefficient (share paid by the firm on share due to the firm)
	Client fully responsible	Client and firm partly responsible	Firm fully responsible	
	(1)	(2)	(3)	(4)
PANEL A: ALL CONTRACTS				
All contracts	41.1	49.4	52.6	0.17
	(5.41)	(6.47)	(5.24)	(0.095)
Fixed cost contracts	52.2	64.5	70	0.21
	(7.56)	(9.15)	(7.19)	(0.13)
T&M and mixed	27.4	36	39.8	0.21
contracts	(7.60)	(8.40)	(6.81)	(0.12)
PANEL B: EXTERNAL CONTRACTS				
All contracts	50.9	58.4	62.5	0.18
	(6.63)	(7.42)	(5.75)	(0.10)
Fixed cost contracts	53.2	69.1	73.8	0.28
	(8.14)	(9.67)	(7.30)	(0.14)
Other contracts	46.5	44.6	50.3	0.19
	(11.7)	(10.8)	(8.50)	(0.15)

Table 4
Firms' and projects' characteristics

	External projects (for all firms)										Firms which have internal projects		
	Sample	Firm foundation date			ISO 9001 certification			Relationship with the client			Type of project		
	Mean	> 1993	<=1993	Difference	No	Yes	Difference	New client	Repeated	Difference	External	Internal	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Yearly turnover 97/98	4.71	2.56	6.24	-3.66	2.71	13.9	-11.2*	5.21	3.82	1.39	8.11	2.2	5.92*
(million $US)	(0.96)	(0.87)	(1.89)	(2.35)	(0.81)	(4.65)	(2.75)	(1.56)	(1.72)	(2.46)	(2.5)	(0.87)	(2.87)
Number of employees	137	84	188	-104*	72	453	-391*	160	113	-46.4	273	73	200
	(27.2)	(17.8)	(50.12)	(52.9)	(8.73)	(137)	(63.8)	(46.6)	(7.4)	(54.6)	(92)	(12)	(86)
Project size	74.7	53.9	90.1	-36.8	48.9	183	-134.8*	85.67	58.7	26.9	88.5	206.7	-118.2
(in man-months)	(10.9)	(10.8)	(17.1)	(21.9)	(6.54)	(45.5)	(25.7)	(16.2)	(12.59)	(22.2)	(22.5)	(85.1)	(89.1)
Share of yearly turnover	18.5	16.2	20.2	-4.0	20.4	10.6	9.81*	16.5	21.33	-4.78	19.3	28.6	-9.26
coming from this project	(1.92)	(2.53)	(2.77)	(3.88)	(2.32)	(1.85)	(4.81)	(1.84)	(3.91)	(3.91)	(3.56)	(3.84)	(5.25)
Client is "big"	0.55	0.55	0.55	-0.051	0.54	0.59	-0.05	0.44	0.71	-0.22*	0.61	0.28	0.31*
	(0.039)	(0.054)	(0.059)	(0.11)	(0.043)	(0.088)	(0.14)	(0.050)	(0.056)	(0.11)	(0.059)	(0.054)	(0.11)
Proportion of "simple"	0.14	0.20	0.094	0.10*	0.13	0.16	0.023	0.10	0.19	-0.09	0.1	0.072	0.032
projects (y2k, cad, data)	(0.027)	(0.048)	(0.030)	(0.051)	(0.029)	(0.065)	(0.068)	(0.030)	(0.048)	(0.054)	(0.037)	(0.031)	(0.049)
Project return	92.6	59.8	120	-59.8*	58.2	235.4	-177.2*	101.7	78	23.7	123.4	297.1	-173.7
(size*markup)	(14.8)	(11.7)	(24.8)	(29.42)	(8.11)	(62.3)	(34.5)	(21.0)	(18.6)	(30.5)	(33.0)	(139.4)	(134.0)
Firm created before 1993	0.57	0	1	-1	0.47	0.22	0.25*	0.44	0.4	0.047			
	(0.038)				(0.043)	(0.074)	(0.096)	(0.050)	(0.060)	(0.078)			
Firm is ISO certified	0.19	0.098	0.26	-0.16*	0	1	-1	0.18	0.21	-0.024			
	(0.43)	(0.036)	(0.045)	(0.061)				(0.038)	(0.050)	(0.062)			
Contract is with a repeated	0.41	0.38	0.43	-0.047	0.4	0.44	-0.038	0	1	-1			
client	(0.038)	(0.051)	(0.051)	(0.077)	(0.042)	(0.089)	(0.097)						

Table 5
Contracts

	External projects (for all firms)										Firms which have internal projects		
	Sample	Firm foundation date			ISO 9001 certification			Relationship with the client			Type of project		
	Mean	> 1993	<=1993	Difference	No	Yes	Difference	New client	Repeated	Difference	External	Internal	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Proportion of fixed cost	0.58	0.73	0.47	0.26*	0.59	0.53	0.061	0.62	0.53	0.087	0.45	0.23	0.22*
contracts	(0.038)	(0.053)	(0.052)	(0.075)	(0.042)	(0.090)	(0.097)	(0.049)	(0.061)	(0.078)	(0.061)	(0.051)	(0.080)
Share of overrun	57.1	68.6	49.3	19.4*	56.9	57.8	-0.98	64.2	44.2	20*	46.7	20.2	26.5*
paid for by the firm	(4.06)	(5.16)	(6.25)	(8.12)	(4.59)	(8.84)	(10.01)	(4.94)	(6.73)	(8.32)	(6.36)	(5.82)	(8.67)
Share of overrun	62.1	71.8	59	12.8	69.9	49.5	20.4	73.4	49.9	23.57*	58.7	44.8	13.9
paid for by the firm	(5.34)	(6.83)	(7.48)	(10.1)	(5.55)	(11.63)	(12.46)	(5.86)	(9.09)	(10.54)	(9.67)	(14.7)	(17.2)
(fixed cost contracts)													

Table 6

Share of overrun paid for by the firm, by project size

	External projects (for all firms)									Firms which have internal projects		
	Firm foundation date			ISO 9001 certification			Relationship with the client			Type of project		
	> 1993	<=1993	Difference	No	Yes	Difference	New client	Repeated	Difference	External	Internal	Difference
	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(10)	(11)	(12)
PANEL A: Unconditional												
	68.6	49.3	19.4*	56.9	57.8	-0.98	64.2	44.2	20*	46.7	20.2	26.5*
	(5.16)	(6.25)	(8.12)	(4.59)	(8.84)	(10.01)	(4.94)	(6.73)	(8.32)	(6.36)	(5.82)	(8.67)
PANEL B: By project size												
<16 man months	72.1	48.6	23.4	60.4	0.0	60.4	69.8	46.1	23.7	43.5	8.3	35.1
	(10.9)	(9.8)	(14.7)	(7.47)	(0.00)	(46.6)	(9.64)	(11.07)	(14.6)	(13.1)	(8.33)	(15.5)
16-49 man months	45.2	50.2	-5.0	44.2	66.7	-22.5	62.7	27.8	34.9	46.4	27.5	18.9
	(16.0)	(9.69)	(18.1)	(8.76)	(21.1)	(20.4)	(10.3)	(11.3)	(15.6)	(10.6)	(15.9)	(19.5)
>40 man months	70.6	48.5	22.1	57.2	55.1	2.1	58.4	51.2	7.01	47.9	22.5	25.4
	(10.3)	(8.31)	(13.5)	(8.5)	(10.8)	(13.9)	(7.79)	(13.4)	(15.3)	(11.6)	(8.69)	(14.20)
PANEL C: Controlled contrast												
			15.6			15.3			22.1*			26.3*
			(9.3)			(17.4)			(8.84)			(13.2)
PANEL D: By project complexity												
simple	76.2	56.6	19.6	64.1	77.8	-13.76	71.6	56	15.6	56.7	47.04	9.26
	(9.52)	(8.9)	(13.0)	(7.45)	(10.23)	(19.3)	(8.11)	(11.2)	(13.6)	(12.6)	(15.4)	(19.9)
complicated	63.17	45.9	17.23	52.3	51.9	0.45	60.13	37.4	22.7	43.8	11.2	32.6
	(8.27)	(6.34)	(10.5)	(5.8)	(10.8)	(11.8)	(6.20)	(8.30)	(10.4)	(7.37)	(5.10)	(9.18)

Table 7
Interaction of different kind of reputation

	Foundation date		External projects (for all firms)				Firms which have internal projects	
			ISO 9001 certification		Relationship with the client		Type of project	
	> 1993	<=1993	No	Yes	New client	Repeated	External	internal
			(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: DIFFERENCE BETWEEN YOUNG AND OLD FIRMS								
Proportion of fixed cost contracts			0.27*	0.23	0.32*	0.17	0.31*	0.12
			(0.082)	(0.22)	(0.094)	(0.12)	(0.12)	(0.10)
Share of overrun paid for by the firm			25.6*	-9.71	17.1*	18.9	7.9	-1.52
			(8.85)	(23.95)	(9.89)	(14.1)	(13.6)	(11.8)
PANEL B: DIFFERENCE BETWEEN CONTRACTS WITH NEW AND REPEATED CLIENTS								
Share of overrun paid for by the firm	17.1	18.9	18.3	26.1			-15.2	NA
	(13.4)	(10.4)	(9.45)	(18.2)			(13.0)	

Table 8
Overruns and delays

	External projects (for all firms)										Firms which have internal projects		
	Sample	Firm foundation date			ISO 9001 certification			Relationship with the client			Type of project		
	Mean	> 1993	<=1993	Difference	No	Yes	Difference	New client	Repeated	Difference	External	Internal	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Mean of overrun	22.7	19.9	21.6	-1.77	18.7	30	-11.3	23.3	17.3	6	15.9	23.9	-8.02
(% of initial estimate)	(2.33)	(2.84)	(2.65)	(3.93)	(1.89)	(6.10)	(4.88)	(2.59)	(2.85)	(3.94)	(2.55)	-4.21	(4.87)
Mean of overrun				-1.83			-5.86			7.1			-9.48
(controlling for size)				(4.03)			(6.71)			(4.18)			(5.18)
Std. deviation of overrun	30.0	23.7	25.6	F=1.16	21.7	33.9	F=2.45	25.6	23.1	F=1.23	20.8	33.7	F=2.6
(%of initial estimate)				p=0.50			p=0.00			p=0.38			p=0.00
Causes of overrun:													
(1) % due to ambiguity	19.5	18.3	20.3	-2.00	19.4	19.9	-0.51	22.9	13.5	9.39	16.9	22.8	-5.93
	(3.01)	(4.87)	(3.86)	(6.17)	(3.47)	(6.36)	(7.43)	(3.99)	(4.34)	(6.27)	(4.41)	(5.11)	(6.72)
(2) % due to internal	9.06	11.0	7.78	3.18	9.73	6.53	3.19	6.79	13.19	-6.41	11.8	4.28	7.56
difficulties	(2.25)	(3.92)	(2.68)	(4.58)	(2.64)	(3.99)	(5.53)	(2.38)	(4.6)	(4.68)	(4.18)	(1.83)	(4.71)
(3) % due to changes	45.4	46.8	44.4	2.38	46.0	42.9	3.13	44.3	47.0	-2.61	36.2	55.9	-19.7*
	(3.83)	(6.01)	(5.01)	(7.84)	(4.31)	(8.47)	(9.45)	(4.79)	(6.45)	(8.04)	(5.82)	(5.93)	(8.31)
(4) % due to delay	12.8	11.3	13.8	-2.57	12.3	14.5	-2.17	13.1	12.2	0.88	17.1	12.5	4.59
by the client	(2.35)	(3.63)	(3.63)	(4.82)	(2.62)	(5.43)	(5.81)	(2.90)	(4.1)	(4.94)	(4.80)	(3.70)	(6.14)
Overrun caused by	7.67	8.00	5.69	2.32	5.74	10.6	-4.82	8.09	4.58	3.51	4.61	7.46	-2.85
the firm ((1) &(2))	(1.58)	(2.34)	(1.28)	(2.51)	(1.32)	(3.24)	(3.11)	(1.81)	(1.49)	(2.52)	(1.64)	(2.86)	(3.28)
Overrun caused by the				2.63			0.23			3.9			-1.64
firm (controlling for size)				(2.6)			(4.84)			(2.2)			(3.51)
Standard deviation	16.4	12.4	19.6	F=2.5	15.2	18.3	F=1.5	17.9	12.1	F=2.18	13.4	23.1	F=2.96
of overrun caused by firm				p=0.00			p=0.18			p=0.00			p=0.00
Overrun caused by	12.5	9.44	14.4	-4.41	10.7	19.8	-9.01	12	13.2	1.28	8.80	19	-10.2*
the client ((3) &(4))	(1.47)	(1.58)	(2.26)	(2.96)	(1.33)	(5.03)	(3.66)	(1.75)	(2.56)	(3.00)	(1.73)	(3.54)	(3.91)
Project was delivered late	0.20	0.15	0.23	-0.074	0.18	0.28	0.10	0.22	0.17	0.06	0.18	0.20	-0.02
	(0.031)	(0.043)	(0.043)	(0.062)	(0.03)	(0.081)	(0.078)	(0.042)	(0.045)	(0.063)	(0.047)	(0.049)	(0.068)

Table 9
Share of overrun paid
by the firm, by project and
client size

	Young firms			Old firms		
	Size of project			Size of project		
	All	<median	>median	All	<median	>median
	(1)	(2)	(3)	(4)	(5)	(6)
All clients	68.7	72.5	63.3	46.6	54.7	45.7
	(5.15)	(8.17)	(9.82)	(6.25)	(8.34)	(6.60)
Small or medium clients	77.5	79.9	73.2	65.4	61.1	68.2
	(7.93)	(10.0)	(13.7)	(7.5)	(12.7)	(9.5)
Big clients	59.8	63.5	55.8	37.6	50.2	29.3
	(9.49)	(13.5)	(13.9)	(6.56)	(11.2)	(7.72)

Table 10
Mean and standard deviation of overrun caused by the firm

	External projects (for all firms)									Firms which have internal projects		
	Firm foundation date			ISO 9001 certification			Relationship with the client			Type of project		
	Young	Old	Difference	No	Yes	Difference	New client	Repeated	Difference	External	Internal	Difference
PANEL A: MEAN OVERRUN CAUSED BY THE FIRM												
PANEL A1: Unconditional												
	8.00	5.69	2.32	5.74	10.56	-4.82	8.09	4.58	3.51	4.61	7.46	-2.85
	(2.34)	(1.28)	(2.51)	(1.32)	(3.24)	(3.11)	(1.81)	(1.49)	(2.52)	(1.64)	(2.86)	(3.28)
PANEL A2: by project size												
<16 man months	5.53	3.96	1.57	4.86	0	4.86	7.3	2.26	5.04	2.14	2.79	-0.65
	(3.95)	(1.23)	(3.92)	(2.00)	(0.00)	(10.89)	(3.54)	(1.73)	(1.3)	(1.53)	(3.49)	(3.88)
16-49 man months	10.1	6.14	4.0	7.64	7.75	-0.11	11.4	2.75	8.64	11.1	5.67	4.43
	(6.38)	(1.82)	(5.46)	(3.07)	(3.8)	(7.24)	(4.41)	(1.25)	(5.18)	(4.73)	(5.16)	(7.53)
>40 man months	8.92	7.22	1.58	6.18	10.5	-4.34	7.94	7.58	-0.39	1.83	12.0	-10.2
	(3.13)	(2.86)	(3.92)	(2.46)	(3.94)	(4.40)	(2.72)	(3.27)	(4.72)	(1.93)	(5.71)	(6.43)
PANEL A3: Controlled contrast												
			2.63			0.23			3.9			-1.64
			(2.6)			(4.84)			(2.2)			(3.51)
PANEL B: STANDARD DEVIATION OF OVERRUN CAUSED BY THE FIRM												
PANEL B1: Unconditionnal												
	12.4	19.6	F=2.5	15.2	18.3	F=1.5	17.9	12.1	F=2.18	13.4	23.1	F=2.96
			p=0.00			p=0.18			p=0.00			p=0.00
PANEL B2: by project size												
<16 man months	20.9	6.98	F=8.9	15.3	0		19.1	9.64	F=3.9	6.99	16.4	F=5.5
			p=0.00						p=0.00			p=0.00
16-49 man months	23.8	8.74	F=7.4	17.1	9.3	F=3.4	20.2	4.94	F=16.5	20.6	16.3	F=1.6
			p=0.00			p=0.10			p=0.00			p=0.43
>40 man months	14.32	17.4	F=1.5	14.8	18.5	F=1.6	17.4	13.5	F=1.67	9.49	30.2	F=10.1
			p=0.33			p=0.26			p=0.23			p=0.00

Note: the F stat in panel B is for the hypothesis that the variances are different.

Table 11

Mean and variance of total overrun

	External projects (for all firms)									Firms which have internal projects		
	Firm foundation date			ISO 9001 certification			Relationship with the client			Type of project		
	Young	Old	Difference	No	Yes	Difference	New client	Repeated	Difference	External	Internal	Difference
PANEL A: MEAN OF TOTAL OVERRUN												
PANEL A1: Unconditional												
	19.9	21.6	-1.77	18.7	30	11.3	23.3	17.3	-6	15.9	23.9	-8.02
	(2.84)	(2.65)	(3.93)	(1.89)	(6.10)	(4.88)	(2.59)	(2.86)	(3.94)	(2.55)	-4.21	(4.87)
PANEL A2: by project size												
<16 man months	15.1	18.4	-3.27	16.6	25	-8.45	19.2	14.6	4.58	14	16.8	-2.8
	(4.12)	(3.19)	(5.15)	(2.56)	(25)	(14.3)	(4.03)	(3.21)	(5.12)	(4.15)	(7.04)	(8.2)
16-49 man months	21.9	16.3	5.58	18.1	21.2	-2.11	21.8	13.9	7.86	20.7	26.5	-5.80
	(6.88)	(2.52)	(6.23)	(3.47)	(5.49)	(8.28)	(4.19)	(4.16)	(6.02)	(5.11)	(11.4)	(10.8)
>40 man months	23.2	28.3	-5.13	24.4	29.9	-5.56	29.0	19.8	9.27	13.7	30.9	-17.2
	(4.74)	(5.90)	(8.54)	(4.57)	(8.00)	(8.54)	(4.92)	(7.33)	(9.12)	(4.60)	(6.80)	(8.44)
PANEL A3: Controlled contrast												
			-1.83			-5.86			7.1			-9.48
			(4.03)			(6.71)			(4.18)			(5.18)
PANEL B: STANDARD DEVIATION OF TOTAL OVERRUN												
PANEL B1: Unconditional												
	23.7	25.6	F=1.16	21.7	33.9	F=2.45	25.6	23.1	F=1.23	20.8	33.7	F=2.6
			p=0.50			p=0.00			p=0.38			p=0.00
PANEL B2: by project size group												
<16 man months	21.8	18.1	F=1.46	19.5	35.4	F=3.28	21.7	17.9	F=1.48	19.0	33.0	F=3.01
			p=0.32			p=0.49			p=0.30			p=0.02
16-40 man months	25.7	12.1	F=4.54	19.3	13.5	F=2.06	19.2	16.7	F=1.33	22.3	36.0	F=2.62
			p=0.00			p=0.31			p=0.56			p=0.02
>40 man months	21.7	35.4	F=2.65	27.4	36.7	F=1.79	31.5	29.3	F=1.15	22.4	35.5	F=2.51
			p=0.02			p=0.15			p=0.74			p=0.03

Note: The F statistic in panel B is for the hypothesis that the variances are different in the two groups.

THESIS PROCESSING SLIP

FIXED FIELD: ill. ______ name ______

index ______ biblio ______

▶ COPIES: Archives Aero Dewey Eng Hum

Lindgren Music Rotch Science

TITLE VARIES: ▶☐ ______

NAME VARIES: ▶☑ Caroline ~~[illegible]~~

IMPRINT: (COPYRIGHT) ______

▶ COLLATION: ~~[illegible]~~ 197 P.

p. 184-197 unnumbered

▶ ADD: DEGREE: ______ ▶ DEPT.: ______

SUPERVISORS: ______

NOTES:

cat'r: date:

page:

▶ DEPT: ~~[illegible]~~ Econ ▶ J191

▶ YEAR: 1999 ▶ DEGREE: Ph.D.

▶ NAME: DUFLO, Esther