Social inequalities in China: which reality?

Kelly Labart

Université d'Auvergne Clermont-Ferrand I
Faculté des Sciences Economiques et de Gestion
Centre National de la Recherche Scientifique (CNRS)
Centre d'Etudes et de Recherches sur le Développement International (CERDI)

Social inequalities in China: which reality?

Thèse Nouveau Régime
Présentée et soutenue publiquement le 28 Mars 2008
Pour l'obtention du titre de Docteur en Sciences Economiques

Par
Kelly LABAR

Sous la direction de
M. le Professeur Jean-Louis ARCAND

Membres du jury :

Président du jury	Mary-Françoise RENARD	Professeur à l'Université d'Auvergne (CERDI)
Rapporteurs	François GARDES	Professeur à l'Université Paris I Panthéon Sorbonne
	Albert PARK	Professeur à l'Université d'Oxford (Royaume-Uni)
Suffragant	Elisabeth SADOULET	Professeur à l'Université de Berkeley (Californie)
Directeur	Jean-Louis ARCAND	Professeur à l'Université d'Auvergne (CERDI)

Contents

List of Tables

List of Figures

Je tiens avant tout à remercier mon directeur de thèse, M. le Professeur Jean-Louis Arcand pour avoir accepté de diriger ma recherche. De cet accord a découlé une collaboration professionnelle et humaine qui a contribué largement à mon épanouissement scientifique et personnel. Je le remercie pour m'avoir transmis ses connaissances dans des domaines très diversifiés, et tout particulièrement en microéconomie. Je le remercie pour son soutien et sa gentillesse qui ont accompagné les évènements stimulants mais aussi quelques fois difficiles liés à la réalisation de cette thèse.

Je remercie également Mme le Professeur Emérite Sylviane Guillaumont-Jeanneney, pour la pertinence toujours renouvelée de ses commentaires, pour la rigueur de sa démarche intellectuelle qu'elle a bien voulu me transmettre.

Je tiens à remercier les membres du jury pour avoir accepté d'accorder du temps à la relecture de cette thèse. C'est un honneur pour moi que de recevoir leurs commentaires qui contribueront largement à améliorer la qualité de ma recherche.

A M. le Professeur Patrick Guillaumont, je souhaite témoigner ma profonde reconnaissance pour son écoute dans les moments où le doute se manifestait, mais aussi pour son dynamisme communicatif et pour la transmission de sa passion pour les problématiques liées à l'économie du développement. Je le remercie ainsi que Mme le Professeur Emérite Guillaumont-Jeanneney, de m'avoir permis d'avoir accès à un cadre de travail stimulant où les connaissances de professeurs de haut niveau peuvent être transmises dans un lieu agréable. Je remercie ainsi les enseignants chercheurs du CERDI ainsi que les intervenants extérieurs pour leur disponibilité et la qualité de leur enseignement. Au personnel administratif du CERDI et de la faculté de Sciences Economiques et de Gestion : Annie(s), Bernard, Brigitte, Bruno, Denis, Dominique, Jacko, Marie-Michelle, Michelle, Odette, Patrick Doger, Solange, Vincent, je dis un grand merci pour leur gentillesse et pour m'avoir aidée à arriver là où je suis dans les meilleures conditions de travail et d'ambiance.

Je remercie également toutes les personnes qui m'ont soutenue tout au long de ce travail : Elisabeth Alma, Yu Por Chen, Bert Hofman, Mary-Françoise Renard, Grégoire Rota-Graziosi, Benjamin Williams.

Merci à Christophe pour son aide d'urgence dans l'utilisation de Latex ainsi qu'à Fouss pour avoir pensé au CD qui pouvait sauver mon travail. Mes remerciements vont également à mes amis, qui m'ont accompagnée tout au long de ma thèse: Christelle, Eloïse, Florent, Hélène, Jacky, Julien, Laure, Ludivine, Susana, Xiao Xian.

Un grand merci à Jean Pierre pour son aide précieuse en informatique. Je le remercie également pour son écoute et son soutien.

Merci à Bertram, pour avoir su me soutenir avec une telle justesse et une telle gentillesse. Merci à Stéphanie, pour son écoute en toutes circonstances, pour ses relectures qui ont contribué à la qualité de mon travail, et simplement pour son amitié. Je remercie particulièrement Thierry, pour nos discussions qui m'ont aidée à avancer. A ma sœur, qui a su être là même de loin, je dis un grand merci pour ses mots justes et bienveillants.

Je dédie enfin ce travail à mes parents, sans l'amour de qui je n'aurais pas eu la force de parvenir où je suis aujourd'hui. Merci...

Introduction

"Adressing problems affecting people's wellbeing is a basic requirement in order to promote social fairness and build a harmonious socialist society. The biggest and most fundamental role of public finance, as the distributor of public resources, is to ensure social fairness and justice".

This quotation from the report on the implementation of the Central and Local Budgets for 2006 and on the draft Central and Local Budgets for 2007, presented by the Chinese Ministry of Finance at the Fifth Session of the Tenth National People's Congress in March, 5, 2007, can at first appear consistent with the communist egalitarian philosophy which underpinned Chinese society's socio-economic development since the 1949 Communist Revolution. However this statement moreso represents a new guideline to follow for the future as the movement of reforms launched in China at the end of the 1970's changed the political as well as the economic orientations of China leading to a disregard for the original egalitarian philosophy.

Amongst others, changes in wages determinants, in social promotion mechanisms as well as in access to health and education imply the need for research to evaluate the present situation in these dimensions. The present thesis contributes to this research. It provides answers to some of the questions that have emerged from the observation of the evolution of this country and its widescale political actions, as well as from empirical evidences underlined by previous research: How have social inequalities in income, education and health in China been affected by the reform process? And what will be expected for the future?

The evolution of the Chinese labour market has been the subject of many studies as deep modifications have occured following the opening of the economy and the goal fixed by the government to make China a player in the world economy (Meng and Kidd (1997), Meng and Zhang (2001) and Dong and Bowles (2002)).[1] Political decisions have had a major impact relative to two dimensions: access to different types of jobs and access to different levels of wages. During the Mao Tse Dong period, the egalitarian philosophy behind the operation of the society lead to concentration of wages in a narrow range of payments. Individuals were differenciated through their participation to the communist party activities but not through their wealth (Bian, Breiger, Davis, and Galaskiewicz (2005)). Collectivization of land as well as an economy centered on the public sector impeded any divergence from the idea of equal income status among the population. However, mismanagement of entreprises' budgets as well as unsatisfactory agricultural policies weakened economic development of the country and even sometimes had catastrophic consequences on the agricultural production and the population's well-being.[2] Subsequently, after the death of Mao Tse Dong, and based on Deng Xiaoping's idea to establish a socialist market economy, a strong movement of reforms was initiated in order to stimulate growth which in turn would increase the wellbeing of the Chinese population. Deep reforms concerning the access to the labour market as well as to the determination of wages were introduced. In terms of access to the labour market, huge differences that exist between the rural and the urban population have been reinforced (Dong and Bowles (2002)). Even if the Household Responsibility System, launched in 1986, has allowed farming families to manage and make use of their land, the income gap which persists between rural and urban populations has induced workers to move to

[1] The economic decisions taken in order to allow the country enter the WTO are the best examples of such objectives (Daniel (2004), Gilbert and Wahl (2003)).

[2] I can quote for instance the Great Leap Forward which had as an objective to promote the development of agriculture but which lead to the great famine involving the death of at least 15 millions of people (Peng (1987)).

the cities in order to find better paid jobs. However, the possibility of moving from the countryside to the cities is, in fact, very low, due to the Hukou place of living registration system which binds population involved in agricultural activities to stay on their farm. This impedes poorer population's chance of access to more diversified labour markets. Moreover, when migrants are permitted to look for a job in the city, they face discriminations in terms of the kind of employment they will be accepted for and of the level of wages they are likely to receive (Fan (2004)).

The rise of economic competition due to the government permitted development of the private sector as well as the development of foreign investment in China, implies the need for public enterprises to be more competitive and more strictly managed. Consequently, wages have to a lesser and lesser extend, been based on an egalitarian structure, with the arrival of a system of bonuses which has been established in order to offer higher wages to the best workers (Chen, Demurger, and Fournier (2005), Coady and Wang (2000)). Managers in public enterprises have also benefited from higher wages which represented incentives for them to adopt good management practices. Disequalizing forces occured consequently following the reforms. White collar workers have replaced the former privileged blue collar workers of the old system. A new hierarchy based on the level of employment has been put in place. Discriminations, penalyzing women and migrants have also become more common (Dong and Bowles (2002), Zhang, Brauw, and Rozelle (2004)). These discriminations are linked to advantages accorded to urban dwellers when hiring employees, but also to differences in the level of education of the two penalised groups as they also suffer, as I will present below, from a lower access to good education, a factor increasingly influencing access to the labour market and the level of wages.

This evolution of the labour market, has consequently lead to a higher weight given to individual productivity influenced by the level of human capital, i. e. to education and health.

Education has become an increasing factor determining the level of wages as well as the involvment of individuals in communist party activities.

Returns to education in China are at the center of a large scope of research (see Zhang and Kanbur (2005) for a survey), which have highlighted the increasing role of education in the determination of wages, reaching a highest obtained value of 10.2% for one more year of schooling in 2001. The link between education and party membership is also an important factor influencing the evolution of Chinese society as the network linked to the participation in party activities (the *guanxi*) highly influences access to high social status and to better paid jobs. As the party reformed itself, it emphasized the role of its members level of education given the objective of becoming open to the world economy which required highly qualified policy makers. Young intellectuals have consequently been given high levels of responsibility and in turn a high social status.

Subsequently, the ways in which individuals have access to a good level of education has become central in the determination of social status. But the reforms have contributed to modify education delivery in China. Whereas some laws underlining the importance given to education has been promulgated, the disengagement of the State from the financing of educational services has lead to the development of inequalities in access to education given the family budget constraints and their place of living (Gustafsson and Shi (2004)). The new system under which good colleges and universities charge entrance fees has restricted poorer families from sending children to theses schools. As a result children from poorer families must finish their education early while children from rich families have the means to continue to a higher level. Moreover, geographical inequalities exist penalising once again individuals from the countryside. As explained above, their lower level of education impedes their access to better paid jobs when they have the opportunity to move to cities. Inequalities in access to education can subsequently generate inequalities between individuals in terms of education and, in consequence, in terms of wages.

The other component of human capital, health, has also suffered from reforms which lowered equalities in access to health care (Gustafsson and Shi (2004)). Here again, the increased costs of health care for individuals has generated more difficulties for the poor to obtain access to good quality services. This has resulted in the poor suffering from poor health. This has negatively consequences on their working ability and their income both if they are wage earners or self-employed.

Given the modifications operating regarding the three dimensions which are wages, education and health, I propose to develop the analysis of the interactions between these economic factors over three chapters. The first, dealing with the evolution of returns to human capital, will result in the two other chapters presenting problematics relative to social inequalities.

Regarding the increasing role played by human capital in the determination of wages, I will investigate returns to human capital in China over the last number of years. To do so, I use the China Health and Nutrition Survey which covers the periods from 1991 to 2004. I will distinguish between returns to health and to education. In both cases, the panel structure of the database I use will allow me to control for elements usually omitted by the empirical literature on returns to human capital in China but which are central to the rigourous estimations of returns to human capital using wage equations. Particularly, using the Semykina and Wooldridge (2005) estimator will give me the opportunity to control for elements such as individual heterogeneity and selection on the labour market, which can be variable with time as well as endogeneity of the human capital variables. Moreover, noting important measurement errors relative to the education variable as well as one of the measures of health I use, I will correct it using the Dagenais and Dagenais (1997) instrumental strategy, testing for its validity in the specific empirical context I encounter.

In the situation where the higher role of human capital in the determination of wages in China is highlighted, the analysis of the distribution of human capital assets among the population will be important to describe the social forces underpinning the operation of the Chinese society. This will be done through two types of analysis. Firstly, I present a study of social status transmission between parents and children. Doing so, I ask if there can be persistent inequalities in wages and education between generations. Secondly, I will pursue a multidimensional analysis of social inequalities focusing on the three major dimensions I focused on up to now, i.e. income, education and health.

Concerning intergenerational transmission of social status in China, the historical and cultural background will show how different political decisions have affected the potential transmission of social characteristics from parents to children. Having demonstrated that the Chinese context contains elements which can either promote or deteriote social mobility, I will adress this topic for the two dimensions of education and wages. Given that different quantiles of the population, in terms of educational attainment or wages, can transmit in different ways their social status to their children, I will use both a linear and a non linear approach to intergenerational mobility.

The multidimensional analysis will be implemented in order to give a broad picture of social inequalities looking at three major dimensions which can influence the wellbeing of the Chinese population (i.e. income, education and health). The study of wellbeing inequalities considers heterogenous elements, such as wages and education. This leads to measurement issues with which I will firstly deal using a dimension by dimension analysis of inequalities and secondly a multidimensional vision. The former method will link the evolution of inequalities in each dimension to the socio-economic context. The latter method will use an aggregated indicator which will allow me to take into consideration correlations existing between the dimensions considered. The results obtained can mirror the evolution of social inequalities which differ from the evolution of the traditionally described inequalities of income.

In the conclusion, results obtained from the following analysis will allow me to explore political recommendations which could contribute to the social objectives pursued by the Chinese government.

Chapter 1

Growing returns to human capital

1.1 The determinants of wages in China

Before the reform process, wages were essentially determined by an overarching concern for equity. Set by the government, salaries were low but homogeneous, with change being largely a function of seniority. Wage earners also received numerous in-kind benefits, such as subsidized housing and medical care. Worker mobility was very limited, leading to important disequilibria in various segments of the labour market (Chen, Demurger, and Fournier (2005), Li (2003)).

Conventional wisdom has it that the reforms have radically modified the structure of Chinese labour markets. The managers of state-owned enterprises have seen their managerial autonomy increase. Cash bonuses have appeared, and often reward employees perceived as being particularly productive (although the structure of many bonus-schemes is still largely driven by concerns of equity) (Cao (2004), Groves, Hong, McMillan, and Naughton (1994), Aron and Olivella (1994)). Concomitantly, a private sector has emerged. Heterogeneity has also appeared in terms of ownership structures, leading to significant differences in wages. For example, Chen, Demurger, and Fournier (2005), who identify six distinct ownership structures (state-owned enterprises (SOEs), local publicly owned enterprises, urban collective enterprises, private or individual enterprises and foreign enterprises), have highlighted the heterogeneity in employee characteristics and wages, based on the ownership structure of the employer. Employees in SOEs, for example, tend to be older, more experienced and better qualified than their brethren working for other types of enterprises. Foreign enterprises, in contrast, provide the highest level of wages, but demand significantly lengthier workhours. These facts characterize what is usually called "a segmented labour market" (Dong and Bowles (2002)) which persists and emphasizes wage gaps among employees.

As ownership structures have evolved, gender differences have emerged, and significant cleavages have also appeared between recent migrants from the countryside and native city dwellers. Maurer-Fazio, Rawski, and Wei (1999) show that, despite constitutional guarantees, the fundamentally patriarchal nature of Chinese society has led to women being increasingly shunted into low wage occupations, the result being a 50% gender wage gap, which also varies according to ownership structure. Interestingly, the gender wage gap is not systematically lower in publicly-owned enterprises (Dong and Bowles (2002)). Collectives do however, continue to have the most homogeneous wage structure. Migrants, for their part, suffer from a 50% wage gap with respect to native city-dwellers and work 14 more hours per week, as shown by Meng and Zhang (2001). The liberalization of the labour market has however given new opportunities for migrants in terms of the type of contract they have access to. But in parallel to this evolution, this new environment has led to many lay offs among the urban population which has, as a consequence, led to a higher degree of competition on the labour market between urban dwellers and rural migrants (Seeborg and Zhu (2000), Knight and Yueh (2004)).

As a result of the reforms and the increasing role played by the private sector, greater competition should lead to various forms of human capital being more faithfully rewarded in terms of wages. This intuition is confirmed by Liu (2001), who compared the returns to education in a market-oriented province (Guangdong) with those in a province (Liaoning) where central planning still plays a major role. The societal downside is that inequality in education will tend to be increasingly translated into inequality in terms of wages.

Consequently, the role of human capital, which can be divided between health and education, in the determination of wages is expected to increase with the movement of the economy toward a structure dominated by market forces.

1.2 Nutrition and wages in China

1.2.1 The relationship between nutrition and wages: empirical evidence

The link between nutrition and wages

The literature on the relationship between nutrient consumption, productivity and salaries is threefold: experimental, quasi-experimental and econometric. The first type consists of the observation of the consequences on productivity or on physical strength of nutritional programs applied to a treatment group, the results of which are compared to these of a control group. Thanks to this method it is possible to compare the impact of improved micronutrient consumption –such as lead– to that of macronutrient consumption –calories and proteins. It is consequently possible to focus health policies on the resources that are the most beneficial to physical well-being.

The second type is based on geographical disparity in the implementation of public health policies. Comparing the consequences of different policies also enables one to draw conclusions as to their efficiency. The major problem that arises in this literature is that political decisions are not random. The quasi-experimental approach can therefore lead to biased estimates.

Finally, several econometric studies have focused on the relationship linking productivity, farm production, salaries and nutrient consumption.

Doing a survey on experimental and econometric studies before 1993, Behrman (1993) specifies that the results of the first studies were inconclusive, essentially because of attrition problems. However, in most cases, research showed a positive impact of nutrient consumption, and especially of micronutrients, on productivity. Most of these studies were carried out on rural populations, and identified a positive link between caloric consumption, productivity and salaries.

Nevertheless, Behrman (1993) underlines that the above-mentioned results often depend on the estimation technique being used: with or without instrumentation, with or without accounting for time invariant unobserved heterogeneity and controling or not for selection bias.

Thomas and Strauss (1997) studied the link between nutrition and salaries in urban Brazilian. The authors use caloric and protein consumption, Body Mass Index (BMI) and height and find a positive non-linear impact of these variables on salaries. Their work reveals particularly important marginal impacts at low values of caloric and protein consumption. As such,improvements in nutrition can be particularly beneficial to undernourished populations. It corroborates the authors' previous result (Strauss and Thomas (1995)), which estimated the relationship between caloric and protein consumption and salaries non parametrically.

The measurement of nutrition

The estimation of the relationship between nutrient intake and wages, is potentially plagued by a strong measurement error problem. This can have two sources. First, different methods are used to quantify nutrient intake which has consequences in terms of the degree of measurement error we can find in the data[1]. The first method is to measure calorie (protein) availability. It is obtained by converting food quantity (purchases and consumption from own production) into nutrient intake using standard food composition tables. This has the advantage of being easy to carry out but has several drawbacks: first, it makes the hypothesis that no food is wasted; second, food eaten by guests or away from home is not taken into account. It is also assumed that the latter type of meals have the same nutrient composition as meals eaten at home. The second method of measurement is potentially more precise. It consists of measuring what is actually consumed by the members of the household. Nutrient intakes are obtained from the usual conversion into nutrients of ingredients consumed but the latter are weighed prior

[1] See Strauss and Thomas (1998) for a detailed explanation of health measurements and their consequences on the quality of data and wage equations estimations.

to each meal and wastage is weighed afterwards. But meals eaten away from home are still not considered and this type of survey is particularly costly. The last method consists of asking the respondants to recall which ingredients they consumed during the previous 24 hours. This measurement has the big advantage of excluding leakages (such as meals for guests, meals eaten away from home and wastage of food). But this method does not take into account the variation of eating habits. Extending the period of recall would lead to less accurate informations. Strauss and Thomas (1998) show that the impact of nutrient intake on wages is biased when we use calorie availability because the rich tend to waste more food than the poor which can lead to an upward bias. Bhargava (1997) uses the household mean of a 24-hour recall which tends to diminish the measurement error.

The second potential source of measurement error resides in the level at which nutrient intakes are measured (Behrman (1993), Strauss and Thomas (1998)). If we use household data, we have no possibility of knowing the intra-allocation of food among household members. Workers in the household may consume more than non productive individuals. Consequently, it is important to have an individual measure of nutrient consumption.

The survey we are using in this study describes precisely the methodology used to measure nutrient intakes as these data are also used in medical studies. As we shall see in the section dedicated to the variables we are using measurements from the China Health and Nutrition survey carried out with particular care which should lead a relatively low level of measurement errors.

1.2.2 Empirical issues

Estimating a wage equation that includes caloric consumption and BMI as proxies for the health component of human capital, and the identification of the marginal impact of these variables is subject to the same statistical problems that are associated with the estimation of the impact of educational output.

The first problem, which is probably the most important one, is individual unobserved heterogeneity. Just as the academic level of achievement can be correlated with capabilities that have not been observed in individuals, so is BMI defined by genetic heritage to a large extent. In order to interpret the effects of this variable on salaries as the effect of present health, the estimates must be ridden of the genetic contribution. This is done in some studies using samples of monozygotic twins (Behrman and Rosenzweig (2001)).

The panel sample structure of the China Health and Nutrition Survey is used here. Surprisingly unobserved heterogeneity of individuals has not been taken into account in the study on health output in developing countries, weakening the credibility of available results –see, for example, Thomas and Strauss (1997) and Schultz and Tansel (1997). When individual effects are purged, the results sometimes differ from those based on the pooling estimator. In India, Deolalikar (1988) controlling for the unobserved heterogeneity, corroborates and reinforces the pooling results. In the Philippines, Haddad and Bouis (1991) show that the effect of health on salaries is no longer statistically significant after having controlled for individual effects.

A second problem lies in the potential existence of selection bias. This will be the case if not participating in the labour market depends on an individual's physical specificities, such as weight and caloric consumption as well as non observables. Moreover, the "within" estimates of wage equations of people who have been earners over several periods of time does not generally solve this problem, since the selection process might not be invariant through time. This is particularly the case in China, where hiring methods have changed since the beginning of the reforms. Indeed, the increasing importance of human capital could lead companies to select workers with the most robust health and the best education. This potential bias seems all the more likely since our sample of workers in 1991 weighs 1.30 kilos less than the workers in 1993 (t-statistic associated with the drawn hypothesis of equality in weights = 1,90), and displays a BMI that is 0.37 lower (t-statistic = 2,00). A "within" estimation of the wage equation is therefore not sufficient to control for both the selection and the shift in the selection process through time (Semykina and Wooldridge (2005)).

The third problem is linked to the joint determination of measures of human capital and salaries. This question has been considered in depth in the labour economics (see Card (2001) for a summary). This problem is important as regards BMI since weight varies greatly according to the quantity and the quality of food and thus according to variations in wages. The impact of the latter variations on caloric consumption is a direct one. Thomas and Strauss (1997) use the price of food and non-wage income as instruments, whereas Schultz and Tansel (1997) add the distance to health and education facilities, occupation –which can be endogenous- and parental education. In both cases, the tests of the over-identifying restrictions are either rejected or do not give significant results. Moreover, none of the studies using data from developing countries account simultanously for selection bias, endogeniety and unobserved heterogeneity.

In this paper, we control for all these sources of bias using a method developed by Semykina and Wooldridge (2005).

1.2.3 Econometric strategy

My basic specification is given by a standard Mincerian wage equation:

$$\ln w_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it} \tag{1.1}$$

Following Mundlak (1978), the selection equation is specified as follows:

$$s_{it} = 1[z_{it}\gamma + z_{i.}\psi + \nu_{it} > 0] \tag{1.2}$$

where $i = 1, ..., N$ stands for the individuals, $t = 1, 2$ stands for the periods, $1[.]$ is an indicator function equal to 1 when the inequality in square brackets holds and the individual i is an employee working at time t and is equal to zero otherwise, w_{it} stands for the hourly wage rate, x_{it} is the matrix of explanatory variables in the wage equation, z_{it} is the matrix of exogenous variables containing the instruments and exogenous variables that are included in x_{it}. This set of variables will be useful for identifying the selection process and define endogenous variables during the estimate of the wage equation. γ is the vector of coefficients associated to z_{it}, which can vary depending on the period, $z_{i.}$ is the average which is specific to an individual, of explainatory variables in the selection equation. α_i are the unobserved individual effects in the wage equation, and ε_{it} and ν_{it} are the error terms assumed to be distributed according to a common normal bivariate density. Note that Mundlak's approach, in terms of correcting for the existence of individual effects that are time-invariant, is based on including explainatory variables as individual averages, and that it is *not* appropriate to replace equation (1.2) by a probit in fixed effects specification because of the biased nature of this estimator (see the standard discussion on this question —the "incidental parameters problem", in Hsiao (2003)).

It is assumed that there are valid exclusion restrictions for the wage equation (x_{it} is a subset of z_{it}) so that identification will not only be based on non-linearity of the inverse Mills ratio. We consider the individuals that were under study during both sub-periods, whether they are earners or not.

Correcting the selection bias, the unobserved individual effects and endogeneity in panel data

In order to estimate β in equation (1.1) in a consistent manner, while controlling selection bias, unobserved individual effects and endogeneity, Semykina and Wooldridge (2005) propose first estimating equation (1.2) independently for the two sub-periods.

The estimated inverse Mills ratio (IRM) λ_{it}, is then used and the following equation is estimated in two-stage least squares for the pooled data:

$$\ln w_{it} = x_{it}\beta + z_{i.}\delta + \lambda_{it}\rho + \varepsilon_{it} \tag{1.3}$$

in which the instruments are z_{it}, $z_{i.}$ and λ_{it} . Finally the associated standard errors being biased, they are estimated by bootstrap.

1.2.4 Data and descritive statistics

The China Health and Nutrition Survey database

The database used in the whole thesis presented stems from the China Health and Nutrition Survey (CHNS). This survey is an on going longitudinal study that includes 8 provinces (Guangxi, Guizhou, Henan, Hubei, Hunan, Jiangsu, Liaoning and Shandong), mostly situated in the east of the country.[2] This can be a first caveat to the economic or socio-economic results obtained using this database. Actually, as we will specify in chapter 3, the huge level of inequalities in China is highly linked to geography, the western regions being disadvantaged in terms of policy, subsidies etc. Consequently, we cannot totally generalize the conclusions of this thesis for the whole of China. Nevertheless, though the survey is not representative on a national scale, the provinces have been selected to give a sufficient geographical variability, in economic development and for health indicators, in order for them to be seen as representing all Chinese provinces. This is to be noted with caution.

The CHNS began in 1989 and provides panel databases for the years 1989, 1991, 1993, 1997, 2000, 2004 and 2006. As an on-going survey, it continues to provide databases every two or three years. The Chinese Ministry of Public Health is involved in the project coordinated by Barry Popkin of the Carolina Population Center at the University of North Carolina. As this survey has been carried out to examine the effects of health, nutrition, and family planning policies and programs implemented by national and local governments, it is the fruit of a huge colaboration between the University of North Carolina and important health research centers or institutions: the National Institute of Nutrition and Safety, the Chinese Center for Disease Control and Prevention. But next to the initial goals relative to health, the survey gives importance to the socio-economic context influencing the households and individuals behavior. Consequently, important information is available concerning income, education, living standards, some

[2]See the map of surveyed regions included in the Appendix 1-23.

kinds of expenditure (food in particular, which is related to nutrition). That is why economists are using this database to analyse socio-economic issues.

As a microsurvey, the sample size is relatively large. The CHNS included 3,795 households in 1989, and respectively 3,616, 3,441, 3,875, 4,403 and 4,387 for 1991, 1993, 1997, 2000 and 2004. The number of individuals surveyed is respectively of 15,917, 14,778, 13,893, 14,426, 15,648 and 16,126 for 1989, 1991, 1993, 1997, 2000 and 2004.

Concerning the sample design, a multistage random-cluster sampling procedure was used to draw the sample from each of the provinces. Counties in the eight provinces were stratified by income (low-, middle- and high-income tertiles) with per capita income figures from the State Statistical Office, and a weighted sampling scheme was used to randomly select four counties in each province (one low income, two middle income, and one high income). Probability-proportional-to-size sampling was used to select the sample from these units. In addition, urban areas initially not included within the county-strata were later incorporated by including the provincial capital and a low-income city from each province. Within each county, the township capital was selected and three villages were chosen randomly. Within each city, urban and sub-urban neighborhoods were randomly selected. The same random selection procedure was used to choose the neighborhoods for townships and villages.[3]

Next to the doubts relative to the representativeness of the data (due to geographic choices for the survey), a second caveat has to be take into account. Concerning data relative to income, some issues of comparability between the different years database appear mostly due to some changes in the questions asked during the different surveys. For example, the way agricultural incomes are defined (if there is information on auto consumption, on food wasted or not etc.), the inclusion or not of subsidies such as food coupons make difficult a rigourous comparison between different periods non wage

[3]For more precise details relative to the database, please refer to the survey's web site: www.cpc.unc.edu/projects/china.

incomes. Given that I am mainly interested in the changes operating in the labour market (Chapter 1 and 2), this does not constitute a limitation for most of the researches done. Nevertheless, as I will specify in Chapter 3, this can constitute a limitation once I am interested in income inequalities evolution.

Other measurement issues appeared once we decided to analyse various topics. I will explain how I decided to control for them when it will be necessary. This will become an econometric problem in the second section of this chapter.

Sample description and variables used

In this section, I am using the data published in the 1991 and 1993 surveys.

My explained variable is given by wages received by enployees from enterpresis in different sectors, i.e. public sector, private sector or foreign enterprises. They do not include bonuses as those are registered only at the yearly level in the survey.

Anthropometric data was gathered from all the participants in the survey, and only a small part of the available characteristics will be used. Measures were done by a health care staff according to protocol and technical standards.

Caloric consumption was calculated on the basis of surveys on consumed products and quantities during three consecutive days for each individual. Consumption during the 24 hours prior to the test was also registered. Cross-checking tests were also undertaken in order to reduce the number of possible errors. This methodology is highly relevant in terms of accuracy of measurement and regarding the issues stressed in the section 1.2.1. This is mainly due to the initial goal of the survey which was the study of health problematics in China.

The basic sample consisted of men over 16, the anthropometric data of whom were seen as reasonable by nutritionists within the Chinese context. Our attention will be focused on the individuals that were tested during both cycles of the survey.

As presented earlier, in the studies linking health and productivity, several measurements are usually made: height, body mass index (BMI –which equals the ratio of weight in kg and squared height in meters[4]) and the consumption of nutrients (kilocalories, grams of proteins, micro-nutrients).

I will focus on the last two. BMI is a middle-term indicator in so far as it depends both on height –essentially defined by the consumption of nutrients during childhood– but also on weight, which varies on a shorter basis according to nutrition. It also depends on both present and past health care (Strauss and Thomas (1998)). Low BMI –less than 21– indicates under-nutrition: people are more fragile and more easily prone to illness. High BMI –more than 28– indicates obesity, a growing mortality factor in both developed and developing countries (Popkin, Paeratakul, Ge, and Fengying (1995), Popkin (1998), Doak, Adair, Bentley, Fengying, and Popkin (2002), Schultz (2001)). Low BMI, often less than 18, characterizes developing countries, with higher rates of under-nutrition than in developed countries, which remains an important cause of mortality. The Chinese economic transition tends to generate a nutrition transition which leads to both a diminishing state of under-nutrition and growing state of obesity, linked as much to the qualitative than to the quantitative changes in food habits (Zhai, Fu, Du, Gue, Chen, and Popkin (2002), Du, Mroz, Zhai, and Popkin (2004), Popkin, Paeratakul, Ge, and Fengying (1995)). As for nutrients consumption, it represents an energy-free proxy of the consumption, linked to productivity.

Figure 1-1 presents a histogram of BMI values concerning the most important sample group of this study, amounting to 3184 observations.

The BMI is less than 21 in 1662 observations (190 are less than 18), and more than 29 in 69 observations. The median and the average BMI are both about 21. Although obesity has been growing in China since the beginning of the 1990s, a moderate and even grave under-nutrition problem remains.

I also consider, as a measurement for nutrition, calory consumption which raises the maximum oxygen absorption level, which, in turn, makes physical efforts easier. Its values vary between 1000 kilocalories per day –which corresponds to a severe case of

[4]BMI (Body Mass Index) corresponds to the French IMC.

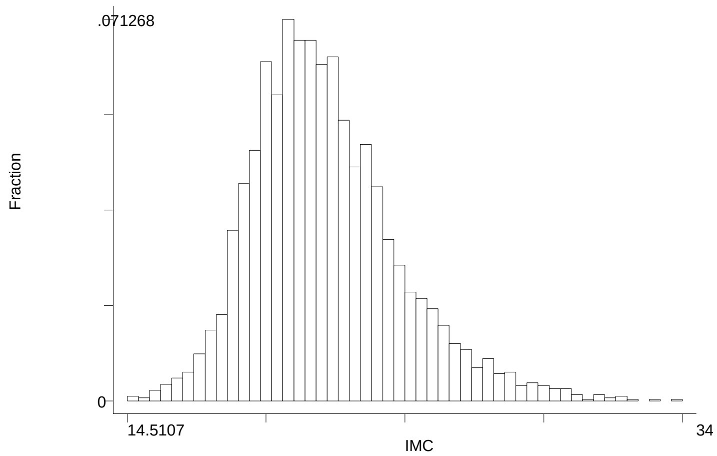


Figure 1-1: BMI distribution in the largest sample of the study.

under-nutrition–, and 6500 kilocalories – associated to high-level sportsmen or obese individuals. Figure 1-2 shows the histogram for the values of caloric consumption, in the same sample.

The average value is of 2789 kilocalories, with 704 kilocalories standard deviation.

I recall that in studies on economic development, the relationship between BMI and salaries is considered positive and often non linear (Strauss and Thomas (1998), Schultz (2001)).[5][6]

Following Thomas and Strauss (1997), I will use in my estimates, as exclusion restrictions for my instrumentation of BMI and the consumption of kilocalories, those that are taken into account in the selection equation.

They include dummy variables that depend on who is cooking at home, who is planning the meals, the amount of food obtained using subsidized vouchers from State-owned shops and on the free market, and finally a dummy variable stating whether the individual

[5] Strauss and Thomas (1998) show with a BMI value above 27, that the link between BMI and salaries tend to be weaker.

[6] Some authors use the weight to height ratio (height being not squared as for BMI). A number of articles report that the link between this indicator and salaries is positive (see Deolalikar (1988), Haddad and Bouis (1991)).

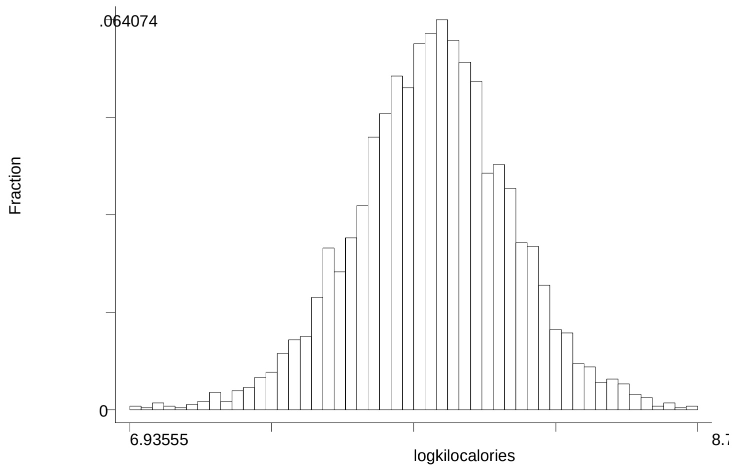


Figure 1-2: The distribution of daily caloric consumption in the largest sample of the study

is a smoker or not.

It is likely that these variables have an effect on the chances of being employed. The planning and cooking of meals may have an impact on the chances of being employed for reasons of time allocation, whereas the availability of food in State-owned shops at subsidized prices may also have an effect on the sense of urgency with which the search for employment is carried out. Lastly, the smoking/non-smoking variable, which is usually used as a proxy in the psychological discount rate, is likely to have an impact on the expenses allocated by an individual to their job-hunting, and thus on the probability of actually finding a job.

As for the instrumentation of endogenous health variables, there is good reason to believe that these instrumental variables could be correlated to my variables for human capital. For example, the smoking dummy variable would be a potentially suitable proxy for an individual's discount rate, a key-factor in the process of decision making concerning health investment. As for the dummy variables for the planning and preparation of meals, they may have an influence on the individual's weight, and thus on their BMI and nutrition –both in terms of quality and quantity.

The same logic can be applied to variables linked to food purchased –which are proxies specific to each individual, prices and availability of products.

The control variables are constituted by the individual's age, them being an urban or rural dweller, and the sector of employment. I distinguish between the latter variable, the state sector, which is made of companies managed by the central government, and the public sector, which includes both state-owned companies and companies managed by local collectivities. I also add dummy variables accounting for the province in which the study is carried out and for the month during which the surveys are done. These control variables, added up with the instruments that are presented below, compose our group of variables z_{it}. The variables $z_{i.}$ will therefore be the individual average for the two given periods for the z_{it}.

Descriptive statistics

Table 1.4 displays the statistics for each variable for two samples: the total sample (earners and non-earners) is made of 3184 individuals, and the sub-group of workers made of 1344 individuals.

It is to be noticed that the earners are younger than the non-earners, and that they are more likely to work in towns and in the public sector. In this latter case non-wage earners are employed in public sector but their remuneration does no take the form of wages.[7] As for the health output variables, earners are sturdier than non-earners but their caloric consumption is not significantly different. As for the instruments, there does not seem to be any major difference between the global sample and the group of earners as to who is planning the meals and cooking within the household. On the other hand, there is a great difference between the places where earners and non-earners buy their food: the first ones are more likely to use State vouchers or buy on the free market. It is consistent in the Chinese context, since earners tend to live in towns and benefit from other advantages granted by policies for urban development. Moreover, they have

[7] This can be particularly the case for employees of collective farms.

access to the free market.

Upon taking the point of view of the intra-individual variability, a large part of the total standard deviation appears to be intra-individual. Individuals' nutrition therefore shows great variability in time, which enables the identification of its effect on salaries.

1.2.5 Results

According to Semykina and Wooldridge's procedure, the results of the first stage selection probit on the total sample group are presented in columns (1) and (2) of Table 1.6.

In accordance to equation 1.2, all the variables z_{it} are also included as individual averages ($z_{i.}$); however, we only bring forward the coefficients that are associated to z_{it}. Including explainatory variables as $z_{i.}$ will correct the presence of unobserved individual effects, without inducing the biases that would follow a conditional probit specification.

It is noted that age is a crucial element for entering the labour market, as is the issue of place of residence, with urban dwellers being privileged. The ways of buying food on the different markets also have an impact on the chances of finding a job.

Besides, we also noted great variation in the effects of different variables through time. The test on the equality of coefficients between 1991 and 1993 shows that the selection process varies through time. The using Semykina and Wooldridge's procedure, which takes this problem into account, is thus justified.

In column (1) of Table 1.7, I made an estimate of the wage equation using Semykina and Wooldridge's estimator as presented in equation 1.3, so as to control simultaneously the selection bias, correlated individual effects and endogeneity. Since correlated individual effects can also create a problem in the selection equation, I have estimated the selection probit of Mundlak type that is specific for each period beforehand.

Concerning the total sample, the results show that none of the given variables has a significant effect on salaries. However, the coefficients of the variables show signs in accord with the literature and the Chinese context. Employment in the public sector would lead to lower wages than in the private sector. It corroborates Chen, Demurger, and Fournier (2005) results. Age is a factor for increased salaries, which would mirror wage evolution based on seniority.

My variables for health capital seem to have no effect on salaries and, therefore, productivity does not seem to be taken into account for defining wages. My results correspond to those found in Behrman and Rosenzweig (2001), showing that checking for individual effects leads to a loss in significance of BMI as defining salaries in an American context. Behrman and Rosenzweig (2001) argue that the indicator is not a good proxie for health at t time, and that it should be seen rather as a measurement for an individual's genetic heritage, which could, indeed, play a significant part in the definition of wages. On the other hand, once the individual effects have been checked, BMI is cleared of its genetic component and has then no impact on salaries.

However, given the specificity of the Chinese context –salaries in the public sector being set by the government[8]–, I chose to continue with my approach by separating earners into two different sub-groups: one for earners in the public sector, and one for the private sector.[9] Once again, the probits of both sub-samples for the first stage and the test on the equality of coefficients for both periods are done. These tests show that, in both the private and the public sectors, the selection varies greatly in time. The results of the estimate on the wage equation for both sub-groups are presented in columns (2) and (3) of Table 1.7.

The results of this distinction are particularly interesting. The variables for health capital are statistically significant in the case of wage-earners from the private sector. BMI has a positive and significant impact on salaries. Calorie consumption is also meaningful and has a non-linear impact, with a noted U curve, the reversal measure of which amounts to 1791.3 Kcal. 50 % of the individuals in our sample are on the decreasing part of the curve and 50 % on the increasing part. It therefore seems that the impact of

[8]The income I am considering in this study does not include bonuses. Consequently, the base wages are still fixed by the government. The income differentials between enterprises reside principally in the bonuses distributed by the managers of collective enterprises.

[9]We recall at this point that the public sector includes both the State sector and the firms managed by provinces or local structures.

nutrition on salaries is positive and increases beyond the nutrition limit. Consequently, it appears that, beyond a given limit that can be seen by the employer, the more robust individuals are, the more productive they are and the more they earn.

As a result, the distinction between private and public sectors points to a significant fact: salaries in the private sector are more likely to mirror individuals' productivity than in the public sector.

1.2.6 Conclusion

In this section, I have studied the returns to health capital, measured using BMI, and nutrition, measured using caloric consumption, in a sample of Chinese workers from the 1991 and 1993 China Health and Nutrition Surveys. I have used the econometric procedure from Semykina and Wooldridge (2005), which simultaneously takes into account the unobserved individual heterogeneity, the selection bias and endogeneity. I have shown that the outputs of health and nutrition in China are relevant in the private sector only, and that there are none in the public sector. With the advancement of reforms, wages in the public sector are more likely to mirror workers' productivity. It is therefore possible that the effects of nutrition and health would also appear in this sector.

1.3 Education, health and wages

1.3.1 The returns to human capital in China: evolving results

The link between human capital and wages can also be measured by returns to education and long term health (which covers other elements than nutrition).

The consequences of the reforms on returns to education are considered by Li (2003), who studies differences in returns to education between cohorts who started working during different periods. For those who began working before 1979, the returns to an additional year of education was equal to 4.7%, while the corresponding figure for those who began working between 1980 and 1987 was 7.3%, the difference being statistically significant; for those who began working between 1988 and 1995, the figure is 6,5% (the difference between these two last values was not found to be statistically significant). These results are confirmed by Zhang, Zhao, Park, and Song (2005). Maurer-Fazio, Rawski, and Wei (1999) find that returns to education are higher in the private sector. The literature also notes that the disparities between men and women, as well as between migrants and native city-dwellers are also important in terms of returns to education (Maurer-Fazio and Ngan (2002), Hauser and Xie (2005), Appleton, Song, and Xia (2005)).

While the educational component of human capital would appear, in light of the available empirical evidence, to play an increasingly important role in determining wages in China, little is known concerning other forms of human capital, the most important among these being health.

The mechanisms whereby health affects the productive capacity of a worker and wages are not difficult to fathom (particularly if individuals are paid according to their marginal productivity). Nevertheless, in order to study the importance of the phenomenon empirically, it is necessary to measure health as accurately as possible. In the preceding section, I have used calorie intake and the BMI to measure short and middle term health indicators.

In this section, I restrict my attention to heaight as the health measure. Nobel Laureate Robert Fogel (2004) refers to this indicator as an element of "physiological capital". Height, thought to be particularly sensitive to early childhood nutritional and health status, conveys information that may proxy more general living standards (Schultz (2001), Strauss and Thomas (1998), Steckel (1995), Fogel (2004)). It is used by some authors as an indicator of the physiological modifications that arise as a result of changes in individual behaviour over time (Steckel (1995), Fogel (1994)). It is also a good proxy for physical strengh, which may be important in certain sectors.

In the development economics literature, the relationship between height and wages has been shown to be positive and statistically significant in Brazil (Thomas and Strauss (1997)), Ghana and Ivory Coast (Schultz (2001)), as well as in the Philippines (Haddad and Bouis (1991)), the causal link running from height to physical strength, and thence from perceived physical strength to wages, at least in occupations requiring significant inputs of physical effort.

In this section, I pursue my objective to study the changes which are happening in the Chinese labour market analysing the returns to height and education during the 1991-2004 period in urban areas.

1.3.2 The added problem of measurement error

To do so, I use the most accurate estimators given the Chinese context, the structure and the quality of the database I am using.

As in the above section relative to returns to nutrition, I am going to estimate a wage equation taking into account the selection process of participation in the labour market. Once more, this leads me to consider the selection process as non constant over time. To cope with this problem, I am consequently going to use the Semykina and Wooldridge procedure which allows me to take into account individuals' heterogeneity in the estimation of both the selectivity equation and the wage equation.[10] Moreover,

[10]The selection process varies over 5 years in the present section as I will consider data from the 1991,

this method has given me the possibility to control for the endogeneity of the nutrition variables.

Concerning the estimation of returns to education and height using the CHNS database, an additional issue emerges: the question of measurement error. In fact, as often observed when dealing with survey data, individuals' declarations relative to their wages or education are often biased.

When we take a closer look at the education variable provided by the CHNS database, we note a high degree of measurement error. Some individuals see their educational level decrease by a number of years over time whilst others see an increase in their schooling attainment of 7 years, between 1991 and 1993 for example. The same phenomenon is apparent for the measurement of height. In Table 1.8, I report the descriptive statistics which give the mean, total standard deviation and within standard deviation. Given that individuals who will be included in my sample should have finished their schooling and their growth, the within standard deviation should be quite low. Nevertheless, we note that it represents around 22 percents of the total standard deviation for education and 14 percent for height. These measurement errors are sources of bias in estimations when I do not correct for it.

In particular, the measurement error can be a source of bias for key parameter estimates if measurement error is not random (aside from the usual attenuation bias stemming from a classical error in variables problem). In this section, I make use of the internal instrument set suggested by Dagenais and Dagenais (1997), which is geared explicitly towards dealing with measurement error, in the absence of additional exogenous instruments.

The higher moments instrument set of Dagenais and Dagenais.

The estimator used here in order to deal with the measurement errors problem is inspired by Dagenais and Dagenais (1997), where the matrix of feasible instruments,

1993, 1997, 2000 and 2004 surveys.

denoted by $Z_{it} = (z_{1it}, z_{2it}, z_{3it}, z_{4it}, z_{5it}, z_{6it}, z_{7it})$, is given by:

$$\begin{aligned}
z_{1it} &= x_{it} * x_{it}, \\
z_{2it} &= x_{it} * y_{it}, \\
z_{3it} &= y_{it} * y_{it}, \\
z_{4it} &= x_{it} * x_{it} * x_{it} - 3x_{it}\left(\frac{x'_{it}x_{it}}{N} * I_r\right), \\
z_{5it} &= x_{it} * x_{it} * y_{it} - 2x_{it}\left(\frac{x'_{it}y_{it}}{N} * I_r\right) - y_{it}\left[\iota'_r\left(\frac{x'_{it}y_{it}}{N} * I_r\right)\right], \\
z_{6it} &= x_{it} * y_{it} * y_{it} - x_{it}\left(\frac{y'_{it}y_{it}}{N}\right) - y_{it}\left(\frac{y_{it}x_{it}}{N}\right), \\
z_{7it} &= y_{it} * y_{it} * y_{it} - 3y_{it}\left(\frac{y'_{it}y_{it}}{N}\right),
\end{aligned} \tag{1.4}$$

where y_{it} is the dependent variable, and where the symbol $*$ designates the Hadamard element-by-element matrix multiplication operator, I_r is an r-dimensional identity matrix, and ι_r is an $r \times 1$ vector of ones. Detailed proofs of the orthogonality of these instruments with respect to the disturbance term are provided in Dagenais and Dagenais (1997).

The resulting "higher moments" estimator, which we shall denote by β_H, is consistent when there are EV (errors in variables) and is much less erratic than other estimators based on sample moments of order higher than two heretofore suggested in the literature.

Note that various implementations of the proposed instrument set are possible. These include Fuller, GMM (the argument taken in an earlier paper by Dagenais and Dagenais (1994)), Nagar (or bias-adjusted 2SLS, see Donald and Newey (2001)), or general k-class estimation. In my situation, I am going to combine the Dagenais and Dagenais method with the Semykina and Wooldridge estimator. Doing so, I will control simultanously for non constant selection, individual heterogeneity, endogeneity and measurement error. The last two problems will be taken into account using the Dagenais and Dagenais set of instruments.

Nevertheless, some doubts can be enounced relative to my choice. At first glance, it is obvious that I am able to estimate the returns to education and height using a panel structure only due to the measurement errors on these two variables, errors which are evolving with time. One though must wonder if I am not estimating white noise instead of the returns to education and height.

1.3.3 Montecarlo Simulations

The measurement error with which I have to contend with here is not one which is traditionally encountered by Dagenais and Dagenais (1997) (i.e. the variable should be stable but varies too much over time).

Therefore, in order to test the consistency and efficiency properties of the method in the present context I am going to employ Montecarlo simulations. I am going to simulate a measurement error on a variable which should not vary considerably over time and compare the results (in terms of bias in the coefficient and of type I error) of different estimation methods.

Following a step by step procedure, I consider at first only a sample with a panel structure and a variable measured with errors, and I compare the results according to the two criteria enounced above of the pooling, panel and Dagenais and Dagenais methods. This last method consists of using the within estimator with instrumentation of the variable measured with error.

Secondly, to bring me closer to the situation I am facing in my final regressions, I introduce a selection process which varies with time and therefore I compare the performances of the pooling, the panel and the Semykina and Wooldridge methods. I will present this last method in two ways: with Dagenais and Dagenais (1997) instrumentation and without, to control for both endogeneity and measurement error. In the following paragraphs, I will use the following notation: when I use the Semykina and Wooldridge estimator without instrumentation, I call it the SW estimator, and when it includes instrumentation, I call it the SWDD estimator.

Montecarlo simulations of the Dagenais and Dagenais method

We consider in this first step a random sample of 800 observations shared among 5 time periods and 1000 replications.

The dependant variable is built in the following way:

$$y_{it} = 1 + 1 \times \widetilde{x}_{it} - 0.5 \times h_{it} + 7 \times \alpha_i + m \times \varepsilon_{it} \tag{1.5}$$

where $h_{it} \rightsquigarrow N(1,1)$ for each time period and is exogenous, $\alpha_i = \widetilde{x}_{i.}$ represents the fixed effects and ε_{it} the residuals which follow a normal distribution for the whole period. We will explain the use of the parameter m below.

The variable of interest is x_{it} which will be measured with error ϵ_{it}.

Without error, this variable is characterized by: $\widetilde{x}_{it} \rightsquigarrow N(8.69, 4)$ as this variable is the parallel to our education variable but with a standard deviation that is lower than the one truly observed (see the mean and standard error of our education variable (Table 1.8)). Given that the error I observe on my education variable is too high in relation to within standard error, the mean of x_{it} will be the same as that for $\widetilde{x}_{it}$ but the standard error will be slightly higher and above all, the within standard error will be large. $\widetilde{x}_{it}$ is made of different elements. At first, I build a variable $\widehat{x}_{it} \rightsquigarrow N(8.69, 3.5)$ for which $\widehat{x}_{it+1} = \widehat{x}_{i1}$ for the periods 2 to 5 to make this variable constant with time. Then, to continue the parallel of our constructed variable of interest and our education variable, I consider that $\widehat{x}_{it}$ can change with time but only in possible ranges, and for a small share of the sample. Given that I have a five year panel, the maximum increase of $\widehat{x}_{it}$ is of one year for the second time period, two years for the third, three years for the fourth and four years for the fifth. I consequently add a variable called u_{it} which follows a uniform distribution of 1 for the second time period, 2 for the third, 3 for the fourth and 4 for the fifth. These permitted variations are only attributed to 5 percent of the sample.[11]

[11] The variable measured with error constructed here is closer to our education variable than to our height variable. The small within variations introduced would not exist for height. So the levels of error

This kind of variability with time is considered as realistic and does not represent a measurement error. The variable measured without error becomes $\widetilde{x}_{it} = \widehat{x}_{it} + u_{it}$. The coefficient on this variable is 1 in equation 1.5, which will consequently be the value toward which the estimations should tend to go once the error is included and if the econometric methodology is appropriate. Moreover, the variables measured with error in our real data (health and education) are also subject to endogeneity. Consequently, I introduce a correlation between $\widetilde{x}_{it}$ and the error terms of equation 1.5, α_i and ε_{it}. $\widetilde{x}_{it}$ becomes, with endogeneity $\widetilde{x}^e_{it}$. x_{it} is then the variable measured with error. I will also differenciate between the case where it is exogenous from the case where is it endogenous, the variable being called x^e_{it} in this last situation.

The following table sums up the characteristics of variables constructed in order to differenciate between the variable measured without error (called "true variable") and the one measured with error (called the "observed" variable).

A) True variable without endogeneity:	$\widetilde{x}_{it} = \widehat{x}_{it} + u_{it}$
B) A + endogeneity linked to the fixed effects:	$\widehat{x}_{it} + u_{it} + \alpha_i$
C) B + endogeneity linked to the random error:	$\widetilde{x}^e_{it} = \widetilde{x}_{it} + 0.9 \times \alpha_i + 0.1 \times \varepsilon_{it}$
D) "Observed" variable without endogeneity:	$x_{it} = \widetilde{x}_{it} + \epsilon_{it}$
E) "Observed" variable with endogeneity:	$x^e_{it} = \widetilde{x}^e_{it} + \epsilon_{it}$

ϵ_{it} is the error term defined as:	$\epsilon_{it} = 0$ if $t = 1$
	$\epsilon_{it} \rightsquigarrow N(0, \sigma_\epsilon)$

As specified above, we see that we have two kinds of variables measured with and without error: $\widetilde{x}_{it}$ and $\widetilde{x}^e_{it}$ which lead to x_{it} and x^e_{it}. As I will explain later, this is done in order to carry out a step by step analysis of the impact of error on the variable without endogeneity, endogeneity without error and endogeneity AND error. To do so, the variable of interest introduced in equation 1.5 will be $\widetilde{x}_{it}$ or $\widetilde{x}^e_{it}$ depending on the situation.

considered in the simulations are actually higher than the real one on height.

On the other hand, the value of σ_ϵ will allow us to consider different levels of error. This will be measured by the ratio $\frac{\sigma_\epsilon}{\sigma_x}$. [12] In order to see the evolution of the efficiency and consistency of the different estimation methods with the level of error, we will use different values of $\frac{\sigma_\epsilon}{\sigma_x}$.

The Montecarlo simulations will estimate:

$$y_{it} = 1 + \beta \times x_{it} - \gamma \times h_{it} + \alpha_i + \varepsilon_{it} \tag{1.6}$$

using ordinary least squares (OLS), fixed effects (FE) and Dagenais and Dagenais (DD) methods. The coefficient of interest is β which must converge to 1. Here again, wether or not I introduce endogeneity on the variable of interest, I will introduce x_{it} or x_{it}^e in equation 1.6.

Concerning the DD method, I will present the results using the set of instruments including z_1 and z_4 on the one hand, and z_3 and z_7 on the other. I do so in order to highlight interesting phenomena that appeared when I tested the different sets of instruments which can be constructed using the Dagenais and Dagenais method.

Moreover, doing the simulations using different values of the coefficients in front of the variables built, the results showed high sensitivity to three key elements: the degree of endogeneity, the size of the error term, and their dependancy on the fixed effects. The two phenomena can be characterized by one parameter: m. Firstly, it gives the range of fixed effects in the total error term, influencing the ratio $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$. This range will influence the quality of the results of the panel estimator which controls for the fixed effects. But as the endogeneity of $\widetilde{x}_{it}^e$ is directly linked to α_i and to ε_{it}, a variation of m also has consequences on the degree of endogeneity of $\widetilde{x}_{it}^e$. Consequently, I choose to consider the sensitivity of the Montecarlo results to m, presenting the dependency of the ratio $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$ given by the chosen configuration on the one hand and estimated from the

[12] As underlined before, this is not exactly the same kind of measure of the error as for height in our database. Nevertheless, the true error contained in our height variable is lower than this one.

	Instruments set		
Scenarii	z_2, z_4	z_3, z_7	Endogeneity
S_{11}^A	X		No: $\widetilde{x}_{it}$
S_{12}^A	X		Yes: $\widetilde{x}_{it}^e$
S_{21}^A		X	No: $\widetilde{x}_{it}$
S_{22}^A		X	Yes: $\widetilde{x}_{it}^e$

Table 1.1: Montecarlo simulations scenarii without considering selection

panel simulations on the other, on the evolution of this parameter. The variations of $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$ (determined by m) will influence the quality of the results obtained due to the different estimators whereas the range of m (in absolute terms) will influence the scale of R^2. I have chosen the range of m in order to obtain reasonable values for R^2 (coherent with the real results). m will be between $[1; 2]$. In the results given below, the key parameter is m but I present the ratio $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$ and R^2 as a function of m to show their inter-dependency.

The various scenarii which stem from the sensitivity of results to the parameter and to the set of instruments being used are summed up in Table 1.1.

Next to these scenarii, the ratio $\frac{\sigma_\epsilon}{\sigma_x}$ will take the values: 0%, 4.4%, 9.7%, 14%, 18.3%, 22.4% and 26.5%. The case for which the ratio is equal to 0% corresponds to the situation of a panel configuration with our variable of interest measured without error. In this context and for the scenarii for which $\widetilde{x}_{it}$ is exogenous, we expect the panel estimator to be the most accurate in terms of bias on the coefficient and which produces the lowest type I error.

The results will be relative to two indicators: bias from the real coefficient (which is 1) and type I error. The nominal size of the test is set at the 10% confidence level. For the Montecarlo simulations to be useful for my empirical application, I also present the evolution of the R^2 with the different levels of error and scenarii, of the pooling, panel and DD estimators. Using this scale and the scale relative to the ratio $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$, I will be able to specify which situation corresponds to my true sample and variables. It will then allow me to give preference to the results of the method which appears, in the Montecarlo simulations, to give the best quality results in terms of bias on the coefficient or in terms of type I error.

I present the results through graphic representations in order to clearly see the sensitivity of results on the different scenarii and the level of error.

Firstly, I am going to have a closer look at the two indicators R^2 and $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$. They are used to show the parallel between the Montecarlo results and potential real estimations with true samples. In truth, this situation is not strictly the same as mine as I must also include selection (which will be done in the following section). Nevertheless, I have chosen a step by step analysis and consequently begin with the situation without selection. For the R^2 and the $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$ ratio, I will only interpret the results obtained from the estimations using the $[z_{1,}z_4]$ set of instruments as it gives the same results as using $[z_3, z_7]$ (see the Appendix 1.A.2). As previously mentioned, I make the distinction between results obtained with $\widetilde{x}_{it}$ exogeneous and $\widetilde{x}_{it}^e$ endogeneous (respectively noted Sx1 and Sx2 on the graphs). Looking at Figures 1-3 and 1-4, a first observation is that the higher m, the lower the explainatory power of the models. The second important element to note is that the estimated R^2 is negative or nul for the panel estimator and the within estimator used with the Dagenais and Dagenais instruments. We know that R^2 are often low when panel estimators are used. The fixed effects for which we control in this way represent an important part of the explaination in the model and consequently, keeping them in the residuals, the R^2 becomes very low. This situation would not appear if we were to use dummy variables to take the fixed effects into account. The pooling estimator appears consequently the best in terms of R^2.

Results relative to the ratio $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$ are presented in Figures 1-5 and 1-6. Different phenomena must be noted. The relationship between m and the ratio is decreasing. As the weight on the random error term rises, the fixed effects are given less weight and the ratio of the standard error of both terms decreases. The constructed ratio and the estimated one are quite similar even if a larger gap appears in the case of exogeneity of $\widetilde{x}_{it}$. On the other hand, I must underline that the ratios are very high as the fixed effects standard error are three to seven times as large as the random residual standard error. This is done in order to give more weight to the fixed effects in the principal equation and to the endogeneity of $\widetilde{x}_{it}$ linked to its correlation with the fixed effects.

Now that I have investigated the common links between usual indicators, I focus on the coefficients and type I errors estimated using the different estimators. Firstly, if we consider the situation in which the error is nul and there is no endogeneity (left column Figures 1-7), we see that the DD and the panel estimations of the coefficient of interest are very unstable even if this instability has a very narrow range. Nevertheless, as expected, the panel estimator does quite well in terms of type I error as does the DD. This error increases marginally with m. For the same reason as mentioned above, it is logical that the fixed effects estimator loses its accuracy once the weight on the fixed effects decreases.

The panel estimator seems consequently to perform well in the situation where there is no error and no endogeneity.

Nevertheless, as soon as endogeneity is introduced, we can see, observing the right hand column of Figure 1-7, that the type I error for the panel estimator shoots up, as does the error for the pooling estimator. Whereas the DD method still performs quite well considering this indicator.

When we consider the different estimators including error, different results emerge. At first, we note that the panel estimation of the coefficient for the model without endogeneity diverges quickly from the true value as soon as the error is at about 9.7% (left column of Figure 1-9). The type I error linked to these estimations reaches a maximum of 100% at the same time. Consequently, the panel estimator performs poorly as soon as a serious econometric issue arises. Considering the same scenario, it is undoubtedly the pooling estimator which performs best regarding the bias on the coefficient and regarding the type I error for a level of error superior to 14%. The DD method using the set of instruments $[z_1, z_4]$ is not regarded as the best estimator in terms of bias on the coefficient when we consider exogeneity but performs best in terms of type I error until an error level of 18.3% is reached.

Once endogeneity is included (right column of Figure 1-9), firstly, the panel estimator performances are still quite poor, and secondly the DD method becomes the best estimator in terms of bias for levels of error inferior to 14% and in terms of type one error in all cases, even if it is obvious that the situation deteriorates with the increase of the error. For the type of error I built in this Montecarlo, the DD estimator with the $[z_1, z_4]$ set of instruments appears to be of no use to correctly estimate the coefficient on the variable measured with error once high level of error are suspected.

Once I consider the set $[z_3, z_7]$ (Figure 1-10), the huge variability on the coefficient leads to a very low level of type I error whatever the situation, but no clear conclusion can be drawn for the coefficient.

Montecarlo simulations of the SW and the SWDD estimators

In these estimations, I am building variables which allow me to consider a first step selection which varies with time and a second step equation which includes the variable measured with errors.

The first step equation is defined for each time period and is built as:

$$\begin{aligned}
s_{i1} &= 1 + 0.92 \times x_{it} - 0.5 \times k_{it} - 0.7 \times h_{it} - 1 \times b_i - 0.3 \times \nu_{it} \qquad (1.7)\\
s_{i2} &= 1 + 0.82 \times x_{it} - 0.7 \times k_{it} - 1.5 \times h_{it} - 1.2 \times b_i - 0.2 \times \nu_{it}\\
s_{i3} &= 1 + 0.72 \times x_{it} - 0.5 \times k_{it} - 0.9 \times h_{it} - 1.1 \times b_i - 0.4 \times \nu_{it}\\
s_{i4} &= 1 + 1.08 \times x_{it} - 1.2 \times k_{it} - 1.7 \times h_{it} - 1.3 \times b_i - 0.3 \times \nu_{it}\\
s_{i5} &= 1 + 0.95 \times x_{it} - 0.5 \times k_{it} - 1.8 \times h_{it} - 1 \times b_i - 0.2 \times \nu_{it}
\end{aligned}$$

where k_{it} is the identification variable of the first step equation, h_{it} an exogenous explainatory variable (similar to the one defined in the above Montecarlo simulation without selection), b_i the fixed effects and ν_{it} the random error term.[13] The coefficients associ-

[13] In order to follow the incremental procedure which consists of adding issues one by one, I have kept

ated with the different variables are fixed to obtain a selection process which will have a significance which corresponds to our real data specification (this will be given by the significance of the IMR introduced in the second stage of the SW procedure). Moreover, they have an impact on the proportion of 1 and 0 inside the sample (see later). These coefficients are different given the time period in order to have a variability of selection with time.

The precise definition of these variables are linked to the form of the second equation which is defined as: .

$$y_{it} = 1 + 1 \times \widetilde{x}_{it} + 0.2 \times h_{it} + 1 \times \alpha_i + m \times \varepsilon_{it} \tag{1.8}$$

Some restrictions must be imposed on the different variables: k_{it}, to be an acceptable identification variable, must not be correlated with α_i and ε_{it}; α_i must not be correlated with ε_{it}; b_i must not be correlated with ν_{it}. On the other hand, the level of correlation between ε_{it} and ν_{it} will determine if there is a significant selection process or not. Consequently, these variables are built from a correlation matrix between the variables $\widehat{k}_{it}$, $\widehat{\alpha}_i$, $\widehat{b}_i$, ν_{it}, ε_{it} which is: $\begin{pmatrix} 1 & 0 & 0.2 & 0 & 0 \\ 0 & 1 & 0.2 & 0.6 & 0 \\ 0.2 & 0.2 & 1 & 0 & 0.3 \\ 0 & 0.6 & 0 & 1 & 0.7 \\ 0 & 0 & 0.3 & 0.7 & 1 \end{pmatrix}$

The five variables $\widehat{k}_{it}$, $\widehat{\alpha}_i$, $\widehat{b}_i$, ν_{it}, ε_{it} follow a normal distribution for the whole sample. I have chosen to modify the mean of $\widehat{k}_{it}$ which becomes $k_{it} = \widehat{k}_{it} + 1.5$. To reinforce the selection process, I also introduce a correlation between α_i and ν_{it}. α_i is constructed as $\alpha_i = \widehat{\alpha}_i + 0.4 \times \nu_{i.}$ (where $\nu_{i.}$ is the average of ν_{it} over all periods) for the first period and then $\alpha_{it+1} = \alpha_{i1}$ for the following periods.

$\widetilde{x}_{it}$ is our variable of interest. It has the same characteristics as defined in the pre-

variables with the same characteristics between the Montecarlo with and without selection where it was possible. This is the case for h_{it}.

ceding Montecarlo simulations and will consequently also be a function of the parameter m influencing the degree of endogeneity. Our hypothesis is that $\widetilde{x}_{it}$ is subject to measurement error. I will consequently consider the variable x_{it} including error as defined in the preceding section (see equation 1.3.3). Once again, I will differenciate between exogeneity and endogeneity of $\widetilde{x}_{it}$. In the former case, $\widetilde{x}_{it}$ remains as it is and leads to a variable measureb with error given by x_{it}. In the latter case, $\widetilde{x}_{it}$ becomes $\widetilde{x}_{it}^{e}$ leading to x_{it}^{e}.

As b_i represents the fixed effects of the selection equation, which will be taken into account in the SW estimator, it has to have characteristics which justify the use of the Mundlak methodology. This means that b_i will depend on the total averages of the explainatory variables included in the selection equation. Consequently, b_i is defined as $b_{i1} = \widehat{b}_i + 0.5 \times x_{i.} + 1.5 \times k_{i.} + 1.2 \times h_{i.} + 0.5 \times \varepsilon_{i.}$ (the dependance on the average of ε_{it} is made to reinforce the selection process). $x_{i.}, k_{i.}$ and $h_{i.}$ are the individual means of x_{it}, k_{it} and h_{it} over all periods, as noted in the SW method described in section 1 of this chapter. For periods greater than one, $b_{it+1} = b_{i1}$.

Given these variables, and the selection process, we build an indicator dummy variable d_{it} which will be equal to 1 when $s_{it} > 0$ and equal to 0 when $s_{it} \leq 0$. The coefficients on the different variables included in the selection equations are chosen so as to have a share of 1 values of 55% and therefore a share of 0 values of 45% which is close to our true database which includes, we will see later, 57% of workers among our total sample.

In the Montecarlo procedure, we now estimate the following equation:

$$y_{it} = 1 + \beta \times x_{it} + \gamma \times h_{it} + \alpha_i + \varepsilon_{it} \tag{1.9}$$

The coefficient of interest is β whose estimation has to lead to 1.

The estimation procedures analysed here are the OLS, the FE, the SW and the SWDD methods.

As in the preceding section, to take into account different levels of error, I consider

	Before selection		After selection	
	z_3	z_7	z_3	z_7
Υ_{it}	0.001	-0.033	0.006	-0.040

Table 1.2: Correlation between $[z_3, z_7]$ instruments constructed before and after selection, and total residuals

different values of the ratio $\frac{\sigma_\epsilon}{\sigma_x}$: 0%, 4.4%, 9.7%, 14%, 18.3%, 22.4% and 26.5%. I also consider different values of the parameter m, and I distinguish the situation where $\widetilde{x}_{it}$ is endogenous or not. Without taking selection into account, I used both sets of Dagenais and Dagenais instruments ($[z_3, z_7]$ and $[z_1, z_4]$). Once I consider the selection process, an added issue appears. As presented in equation (1.4), z_3 and z_7 are based on y_{it} which in the true situation exists only for $d_{it} = 1$. The sample of definition for z_1 and z_4 and for z_3 and z_7 will consequently be different. z_1 and z_4 will be built based on the explainatory variable x_{it} for the whole sample whereas z_3 and z_7 will, in the true situation, be built only on the sub-sample of selected individuals (who are wage earners). Consequently, one must wonder if the selection process has an impact on the construction of the set z_3 and z_7. I will not be able to test this hypothesis on the real sample but we can test it through the use of the Montecarlo simulations. Actually, the variable of interest, y_{it} is built in the first step for the whole population. But I then focus my attention on the sub-sample for which $d_{it} = 1$. So I am able to compare the results obtained from the construction of z_3 and z_7 before the selection process, to the results obtained after the selection process. A first clue is given by the correlation of z_3 and z_7 with the total error term (which I denote by Υ_{it}). It gives the results presented in the Table 1.2.

We can see from these numbers that there is no clear difference between the two sets of instruments. Moreover, the Montecarlo results obtained from the SWDD with the two sets of instruments are very similar.[14] Consequently, I have chosen to keep the results from the set of $[z_3, z_7]$ instruments that are built after the selection process in order to be closer to my own true situation.The new scenarii are presented in Table 1.3.

[14]In order to limit the number of unuseful graphs, the results linked to the set $[z_3, z_7]$ determined before the selection are not included in the thesis. Results are available upon request.

	Instruments set			
Scenarii	$[z_2, z_4]$	$[z_3, z_7]$	$[z_3, z_7]$	Endogeneity of $\widetilde{x}_{it}$
		Before selection	After selection	
S_{11}^B	X			No: $\widetilde{x}_{it}$
S_{12}^B	X			Yes: $\widetilde{x}_{it}^e$
S_{21}^B			X	No: $\widetilde{x}_{it}$
S_{22}^B			X	Yes: $\widetilde{x}_{it}^e$
S_{31}^B		X		No:$\widetilde{x}_{it}$
S_{32}^B		X		Yes:$\widetilde{x}_{it}^e$

Table 1.3: Montecarlo simulations scenarii considering selection

Once more, as an introduction, I will describe the results relative to the R^2 and to the ratio $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$. I will also specify the level of significance of λ_{it}, the inverse Mills' ratio used by the SW and the SWDD estimators. This indicator will help us to compare the Montecarlo simulation model with our true situation.

At this step, it is important to stress that the linkages which exist between all the parameters and the equations can lead to compensation effects. In this section, I am taking into account 4 different issues: selection, endogeneity, individual heterogeneity and measurement error. I use the Montecarlo simulation in order to highlight the total effect of these issues on the results of the different estimators.

In Figures 1-11 and 1-12, we see the R^2 for the scenarii S_{11}^B and S_{12}^B (the results for the scenarii S_{21}^B and S_{22}^B are in the Appendix 1.A.3, as they give similar results). The first remark is that in most of the cases, the panel R^2 is once again very low or negative. The reasons are the same as those given in the preceding section. Concerning the other estimators and whatever the set of instruments, the highest R^2 values, which are practically the same, are given for the SW and the SWDD estimators. They are between 0.6 and 0.9.

Concerning the ratio $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$, represented in Figures 1-13 and 1-14 for the $[z_1, z_4]$ set of instruments (see the Appendix 1.A.3 for the $[z_3, z_7]$ set of instruments), here again we note a negative relationship between $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$ and m. Moreover, the same larger gap as noted in the preceding section between the estimated and the true ratio is observed in the case of exogeneity of $\widetilde{x}_{it}$. It seems that compensation effects influence the evolution of the difference between the constructed and the estimated ratio. This latter ratio is higher than the constructed one when there is zero error and for error greater than 14%. but it is lower when there is a low level of error.

This mirrors the difficulty in predicting results when we consider numerous phenomena. Finally, once again I want to stress the high level of the ratio. This has been done on purpose in order to emphasize the role of the fixed effects both in the principal equation and in the degree of endogeneity.

The level of significance of λ_{it} which can be seen in Figures 1-15 and 1-16, is sensitive to the level of error. The higher the error, the less significant the IMR. This is logical as the selection process is built from the variable measured with error.[15] We also observe that the IMR is more significant once we take into account endogeneity. We will see in the section relative to the true estimations that the non significance is similar to the context of my empirical model.

I am now going to study in detail the quality of the results obtained with the various estimators in terms of bias from the true coefficient (1) and type I error.

Results relative to a zero error are presented in Figures 1-17 and 1-18 for the scenarii S^B_{11} and S^B_{12}, and S^B_{21} and S^B_{22}. Let us focus first on the panel estimator. We see that the estimation of the coefficient is not very stable in the case of exogeneity of $\widetilde{x}_{it}$. This confirms the results obtained when I did not consider the selection. Moreover, we see that the range of the variations are also very narrow. The panel type I error is once again clearly the lowest, as expected in the case of exogeneity and zero error.

The more stable and accurate estimations of the coefficient seem to be given by the SW and the SWDD estimators both in the case of exogeneity and endogeneity of $\widetilde{x}_{it}$. Moreover, as in the preceding paragraph, an important phenomenon is revealed once endogeneity of $\widetilde{x}_{it}$ is introduced. The panel estimator's coefficient estimate shoots up as

[15] This has been done in order to be closer to the true situation where I do not control for measurement error in the selection process.

does the type I error. Clearly, this estimator does not give acceptable results as soon as endogeneity exists. In Figures 1-19 and 1-20, I want to emphasize even more the panel estimator's poor performance when a new problem is introduced. We see that from a level of error of 9.7% the coefficients estimated using the panel estimator are very far from the true value (1) and the type I error rises to 100%. And this is the case both when $\widetilde{x}_{it}$ is exogenous and when it is endogenous.[16] This confirms the poor performance of the panel estimator once econometric issues are introduced.

These results are confirmed when I use the $[z_3, z_7]$ set of DD instruments. But in this latter case, I note a decrease in the performances of the SWDD in terms of bias from the true coefficient compared to the results obtained with the $[z_1, z_4]$ set.

To more clearly see the differences between the two, I choose to focus on the three other estimators, pooling, SW and SWDD. Figures 1-21 and 1-22 give the results for the S^B_{11} and S^B_{12}, and S^B_{21} and S^B_{22} scenarii whithout panel estimations. Looking at the results of the two first scenarii (Figure 1-21), we clearly see that in the situation where $\widetilde{x}_{it}$ is exogenous, the SW and SWDD perform best in terms of bias from the true coefficient when there is a measurement error of up to 18.3% and up to 9.7% in terms of type I error.Then, the pooling estimator is preferred. This appears logical as we recall that the selection IMR becomes less and less significant with the increase of the measurement error. We can see that in the case of $\widetilde{x}_{it}$ endogenous (scenario S^B_{12}), the SWDD method yelds the lowest bias from the true coefficient when there is a measurement error up to 22.4% and then the SW becomes the best estimator. In terms of type I error, the SWDD estimator remains the best whatever the level of error. Nevertheless, it is interesting to see that the pooling type I error decreases with the level of error.

Once we consider the use of the second set of instruments (i.e. $[z_3, z_7]$), we note that the coefficient estimated using the SWDD method is much more unstable but conversely, the type I error is more stable and moreover, this estimator always performs best in

[16] We note a slight convergence when there is endogeneity and high values for m. But this phenomenon is isolated.

terms of type I error. When this set of instruments is chosen in the case of endogeneity of $\widetilde{x}_{it}$, the type I error remains at its lowest but the SW becomes the most effective in terms of bias from the true coefficient, whatever the error level. It appears consequently that the set of instruments based on the dependant variable measured without error gives poorer and more unstable results than the $[z_1, z_4]$ set (based on the variable measured with error) in terms of bias, with this instability leading to lower type I error.

Implications for empirical applications

Following the step by step analysis I have done to compare the performances of different estimation methods, different recommendations are useful for the empirical applications which will follow.

Firstly, the Montecarlo simulations have clearly demonstrated the panel estimator good performances in controlling for the individual unobserved heterogeneity. However, as soon as another empirical issue is introduced, like measurement error or endogeneity on the variable of interest, this estimator gives the poorest results both in terms of bias from the true coefficient and in terms of type I error, whatever the Montecarlo simulation considered. Given my empirical context (including two variables potentially measured with error and endogenous), this estimator cannot be considered as potentially useful.

Secondly, when no selection process exists in an empirical implementation, I have shown that the DD method performs best both in terms of bias from the true coefficient and in terms of type I error for ranges of error lower than 14% (the error being measured by the ratio of the standard error of the error term to the standard error of the variable measured with error). Surprisingly, the pooling estimator performs best when I consider the two indicators of performance for error higher than this value. Consequently, considering a type of error different from that taken into account on the original paper of Dagenais and Dagenais (1997) leads to a decrease in the performances of the DD method with the increase in the level of error.

When selection is introduced, which best corresponds to my own empirical situation, the SWDD estimation method is always the most relevant in terms of type I error but it also appears the best in terms of bias on the coefficient only for errors lower than or equal to 22.4% and in the case of endogeneity of our variable of interest. In the other cases, the SW performs best.

All the results enounced above are given using the $[z_1, z_4]$ set of instruments. The $[z_3, z_7]$ set always gives unstable estimations for the coefficient leading to non conclusive results. The set of instruments based on the variable measured with error lead, consequently, to better estimations of the coefficient on the variable of interest and are the best performers most often in terms of type I error.

Finally, I stressed the importance of the parameter which influences the level of explainatory power (R^2), as well as the level of fixed effects and of endogeneity: which I denoted by m. This parameter will allow me to determine precisely my position on the different graphs in terms of my real empirical situation. I will refer to it again when presenting the results of my true estimations.

1.3.4 The returns to education and height

Variables used

I use the CHNS databases for the years 1991, 1993, 1997, 2000 and 2004.[17] My basic sample consists of urban males, more than 16 years of age, whose anthropometric measurements fall within the bounds of what is considered reasonable by nutritionists in the Chinese context.[18]

My control variables include age, the sector in which the individual works, where we

[17] See the data description made in section 1.2.4 for a detailed presentation of the database.

[18] Measurements were carried out by trained health workers who followed standard protocols and techniques. Weight was measured in light indoor clothing without shoes to the nearest tenth of a kilogram with a beam balance scale. Height was measured without shoes to the nearest tenth of a centimeter with a portable stadiometer. Each of these measurements was done by at least two health workers; one worker took the measurement, and another recorded the reading.

differentiate between the public sector (enterprises owned by all public institutions) and the state sector (enterprises owned by central government only).

Identification of the first-stage probit equation is achieved through two exclusion restrictions. In order to be valid, the variables involved must affect the likelihood of the individual being a wage earner, without having any direct effect on the individual's wage. Following Thomas and Strauss (1997), we use as identification variables commodity prices in the different counties included in the sample, available in the Community Survey carried out in parallel with the household survey.

Descriptive results

Descriptive statistics are given in Table 1.8 for the full sample and the sample of wage earners.

On average, as in the preceding section, wage earners are younger and are more likely to work in the public and state sectors than non-wage earners. Given the definition of the state sector, which is included in the public sector, it is logical that the share of participants in the public sector is higher than the one in the private. The wage earners have a higer education level and they are taller. Let us examine if differences are translated by higher returns to education and height.

Results

Results are presented in Tables 1.9 and 1.10. Table 1.9 presents the first stage probit equations corresponding to each period. Two results are noteworthy. Firstly, while seniority and public sector employment are both significant determinants of wage employment in the earlier periods, their significance diminishes in later periods. This finding reflects the decreasing role played by seniority and public sector employment as the reform process undertaken by the government progresses. Secondly, our identification strategy in the employment probits, based on commodity prices, generally appears to be valid, although the instruments provide relatively weak identification for the 1993 and 2000 periods.

Moving to the results of the estimation of the wage equation, I begin by testing for the presence of selection using the results of the fixed effects estimation which includes the estimated IMR. Statistical significance of this last variable will indicate the presence of non-random selection. The result of this test is presented in the third column of Table 1.10, and indicates that non-random selection occurs in our sample. As such, any estimate of the effect of human capital on wages in urban China must account for selection issues, which leads me to use the Semykina and Wooldridge estimator.

My empirical context is to be compared to the Montecarlo simulations carried out in section 1.3.3. Moreover, given the regressions' results, three characteristics help me to determine exactly which Montecarlo results help to choose the most rigourous estimation method: the R^2, the ratio $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$ and the level of significance of the IMR. Given the values of these constants, I will obtain the value of m which corresponds to my real empirical situation. We can see from Table 1.10 that the R^2 are around 0.65. This corresponds to the values of R^2 estimated by the Montecarlo simulations for SW and the SWDD methods, whatever the value of m. Once I consider the ratio $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$, which is equal to 4.1917 as estimated by the FE estimator, Figure 1-14 leads me to find a value of 1.7 for the parameter m for my empirical situation. This is also confirmed by the level of significance when we look at Figure 1-15, the level of significance obtained from the true estimations being around 1.40.

Given these characteristics, I am in the situation where the SWDD estimator is the best performer in terms of bias from the true coefficient and of type I error, for a level of error lower or equal to 22.4%.[19] I recall that for my two variables which I suspect are subject to measurement error, the within standard deviations represent for education 22% of the total standard deviation whereas it represents 14% for height. Given the information given abive, the results from the SWDD estimator are preferred.

The results corresponding to this estimator can be found in column 5 of Table 1.10.

[19] I use, in this section, the same notation as in the Montecarlo simulations: the Semykina and Wooldridge procedure without the instrumentation is called the SW estimator and with instrumentation, the SWDD estimator.

We note that working in the state sector negatively influences urban wages: this is in accordance with the results of Chen, Demurger, and Fournier (2005) who show that wages are lower in the public and state sectors. Seniority does not have a significant impact on wages. Returns to height are negative but non significant, contrary to the results obtained using the pooling and the SW method. The results relative to education show statistically significant returns to education of about 4.66% using the SWDD, which is the highest estimate given the different estimators.[20] This positive and significant impact of education on wages in urban areas is in accordance with Zhang, Zhao, Park, and Song (2005) and Fleisher and Wang (2005). Nevertheless, controversy remains concerning the appropriate point estimate. In the existing literature, it is common to distinguish between the pre- and post-reform periods.

As my sample covers the entire 1991-2004 period, it could be argued that all individuals are observed after the reforms relative to the labour market. However, given political inertia in the reform process, it is likely that the 1991-1993 period corresponds to the pre- or initial reform period, whereas the 1997-2004 period corresponds to a post- or consolidated reform period. As such, I split my sample into these two periods and consider separate estimates based on these two subsamples. Results are presented in Table 1.11. Even if education does have a statistically significant impact on wages for the 1991-1993 period (with a coefficient of 2.76), this impact is much lower than the one observed for the 1997-2004 period (4.23). It seems consequently that returns to education have increased with time, i.e. follow the movement of reforms. Returns to height remain insignificant for both sub-periods.

In order to investigate the robustness of this result, I consider a different break, with the two subperiods now being 1991-1997 and 2000-2004, respectively. Results are presented in Table 1.12. We now find a return to education of 7.73% for the 1991-1997 period and of 3.81% during the 2000-2004 period. This is surprising as I presented the

[20] We assess the validity of our identification strategy using the usual Hansen test of the overidentifying restrictions. This test does not reject the null hypothesis of instruments admissibility.

argument of increasing returns to education with time. Given that the estimate for the sub-period 2000-2004 is still higher than the one obtained for the sub-period 1991-1993, we can advance the hypothesis that very high returns to education are observed in 1997. This can be either a reverse point in the relationship between education and wage or an absurd estimate. No evidence is available on whether the former or the latter argument is the correct one.

Moreover, the same surprising phenomenon is observed when considering returns to height: they become positive, significant and high for the sub-period 1991-1997. Here again, no clear evidence can explain this cut off in 1997.

1.3.5 Conclusion

Using micro data from the China Health and Nutrition Survey for the period 1991-2004, I study the impact of human capital on urban wages in China.

To do so, I use the number of years of education and height as measures of human capital.

In order to deal with the problems associated with Mincerian wage equations in a panel data context, I use the Semykina and Woldridge estimator which simultaneously accounts for non-random selection, endogeneity and individual heterogeneity issues.

Moreover, given that the variables I focus on (education and height) are subject to important measurement errors, I use the Dagenais and Dagenais instrument set in order to explicitly control both for endogeneity and measurement errors. As the kind of measurement errors encountered are not exactly the same as those presented in the Dagenais and Dagenais (1997) article, I implemented a Montecarlo simulation in order to compare the performances of four estimators: the OLS, the FE, the Semykina and Wooldridge estimator without instrumentation and the Semykina and Wooldridge estimator using the Dagenais and Dagenais set of instruments in order to deal with measurement error. Given the parallel between the Montecarlo simulations and my empirical specification, the simulations using the Semykina and Wooldridge method employing the Dagenais and

Dagenais instrumentation appear to give the best results in terms of type I error and in terms of bias from the true coefficient, leading my empirical results to focus on this methods results.

I stress three major findings. Firstly, I find statistically significant and positive returns to education during the whole period, whereas height is never significant (exception made for the sub-period 1991-1997, which is not truly justified). Secondly, given the emphasis in the existing literature on the evolution over time in returns to education, I consider two structural breaks: 1991-1993 versus 1997-2004; and 1991-1997 versus 2000-2004. The first break leads me to confirm that there has indeed been an increase in the returns to education over time, with a point estimate of approximately 2.76% for the earlier years, which rises to 4.23% during the more recent period. The second break leads to surprising conclusions relative both to returns to education and to returns to height, the former being the highest on the sub-period 1991-1997, and decreasing thereafter, the latter becoming significant for the sub-period 1991-1997, and becoming once again insignificant thereafter. No intuition supports these results.

The transformations in the labour market tend consequently to demonstrate a higher weight accorded to education at the present time and no weight accorded to strengh. Thirdly, my point estimates of the returns to education in urban China are significantly lower than those generally reported in the literature, which may be due to the fact that previous studies have not explicitly dealt with non-random selection, individual heterogeneity and endogeneity in the context of Chinese panel data.

1.4 Conclusion

In this chapter, dedicated to the study of returns to human capital in China, the role played by the reform movements has been emphasized in different ways. Focusing on returns to health and nutrition, I have underlined the significant impact of these factors on wages in the private sector whereas they are not relevant in determining wages in the public sector. Given the evolution of this latter sector toward a more competitive system (through the establishement of bonus pay compensations for more productive workers), the nutritional status as well as the health of individuals should play an increasing role in the determination of wages in this sector.

Considering urban wage returns to education, I also stressed the impact of the reform movement showing increasing returns between the periods 1991-1993 and 2000-2004. Even if these results confirm the time evolution obtained by other studies using different databases, their levels are lower. Given that my estimations take into account different issues relative to estimations of wage returns to education, they can be closer to the true values of returns to education in China.

The increasing returns to human capital linked to the reform movement consequently lead policy makers to face new socio-economic problematics. Firstly, the question of equal opportunities for individuals in terms of access to services (such as education or health services which will in turn affect wages) arises. Regarding this issue, I am going to examine, in the following chapter, intergenerational mobility in education and wages in order to determine if the educational system, as it exists now and as it has evolved since the reform movement, helps to promote equal access to education and allows for mobility between generations. Secondly, given the egalitarian wage structure of the Mao Tse Dong period, the increase in returns to human capital can constitute disequalizing forces leading to higher inequalities. If, moreover, human capital inequalities exist among the population, these inequalities can reinforce the disequelizing forces mentioned above and China could consequently find itself in a vicious circle with which policy makers would have to contend. In order to contribute to the analysis of this last problematic, the study of inequalities both of income and of human capital will be carried out in the last chapter of this thesis, also looking at their evolution with the movement of reforms.

Tables chapter 1

Tables section 1.2

	Total sample			Sub-sample of earners		
Observations	3184			1344		
	mean	standard deviation		mean	standard deviation	
		total	"intra" individus		total	"intra" individual
Hourly wage	1.0349	3.8345	1.0043	1.0345	3.8345	1.0043
Age	42.2082	14.3571	1.0001	38.9586	11.1019	1.0000
Public sector	0.4250	0.4944	0.1538	0.8477	0.3594	0.1743
State sector	0.2806	0.4491	0.1496	0.5743	0.4946	0.1917
Urban	0.3346	0.4719	0.0000	0.5004	0.5002	0.0000
Human capital						
BMI	21.6688	3.1503	1.5046	22.284	3.5465	1.8884
Log kilocalories / day	7.9026	0.2505	0.1456	7.8787	0.2326	0.1449
Log kilocalories / day squared	62.5136	3.9533	2.2945	62.1274	3.6591	2.2736
Instruments						
Individual planning meals:						
man hh head	0.2848	0.4514	0.2713	0.2727	0.4455	0.2678
spouse	0.5310	0.4996	0.2868	0.4967	0.5002	0.2849
Individual preparing meals :						
man hh head	0.1060	0.3079	0.1896	0.1197	0.3248	0.1980
spouse	0.6295	0.4830	0.2619	0.5854	0.4928	0.2640
% purchased food :						
with state coupons	0.1653	0.3133	0.2448	0.2918	0.3748	0.3249
on the state markets	0.0553	0.1773	0.1117	0.0777	0.2113	0.1347
on the free market	0.3535	0.3795	0.2690	0.4634	0.3931	0.3264
Smoker	0.7059	0.4557	0.2210	0.7250	0.4466	0.2184

Table 1.4: Summary statistics for the total sample and the sample of wage earners

	observed in 1991	observed in 1993	observed in 1991 et 1993
$s_{i1991} = 1$	1289	686	686
$s_{i1991} = 0$	1895	1101	1101
$s_{i1993} = 1$	658	1167	658
$s_{i1993} = 0$	1129	674	1129
$s_{i1991} = 1$ or $s_{i1993} = 1$	1289	1167	686
$s_{i1991} = s_{i1993} = 0$	1115	1115	2230
$s_{i1991} = s_{i1993} = 1$	667	667	1344
	3184	1841	3574

Table 1.5: Wage earners observed, 1991 et 1993

	Sample total		By sector			
			public sector		private sector	
	1991	1993	1991	1993	1991	1993
	(1)	(2)	(3)	(4)	(5)	(6)
Age	−0.046	− 0.042	− 0.874	0.378	0.955	− 0.843
	(0.004)	(0.003)	(0.359)	(0.223)	(0.185)	(0.215)
	[−0.016]	[−0.014]	[−0.133]	[0.111]	[0.121]	[−0.116]
Public sector	1.127	1.168				
	(0.256)	(0.256)				
	[0.398]	[0.397]				
State sector	−0.031	0.231	− 0.014	− 0.041		
	(0.262)	(0.245)	(0.335)	(0.311)		
	[0.011]	[0.081]	[−0.002]	[−0.012]		
Urban	− 0.309	0.047	− 0.117	0.249	− 0.372	0.044
	(0.109)	(0.094)	(0.183)	(0.136)	(0.199)	(0.161)
	[−0.106]	[0.016]	[−0.017]	[0.074]	[−0.039]	[0.006]
Instruments						
Individual planning meals:						
man hh head	0.305	− 0.372	0.732	− 0.675	− 0.255	− 0.082
	(0.256)	(0.279)	(0.459)	(0.395)	(0.436)	(0.500)
	[0.113]	[−0.115]	[0.078]	[−0.231]	[−0.027]	[−0.011]
spouse	− 0.044	−0.2077	0.443	− 0.848	− 0.113	0.179
	(0.218)	(0.2205)	(0.407)	(0.356)	(0.318)	(0.318)
	[−0.015]	[−0.0719]	[0.070]	[−0.237]	[−0.0145]	[0.024]
Individual preparing meals:						
man hh head	0.068	0.275	0.376	0.520	0.076	0.328
	(0.288)	(0.274)	(0.581)	(0.403)	(0.434)	(0.470)
	[0.024]	[0.097]	[0.051]	[0.139]	[0.010]	[0.051]
spouse	0.257	0.179	0.119	0.726	0.176	0.207
	(0.284)	(0.276)	(0.544)	(0.408)	(0.436)	(0.471)
	[0.090]	[0.061]	[0.018]	[0.206]	[0.022]	[0.028]
% purchased food :						
with state coupons	0.875	− 2.295	− 0.135	− 2.346	1.803	
	(0.502)	(0.553)	(0.866)	(0.890)	(0.682)	
	[0.308]	[−0.785]	[−0.020]	[−0.689]	[0.228]	
on the state markets	1.353	− 0.069	0.714	− 0.697	0.817	0.780
	(0.501)	(0.471)	(0.949)	(0.919)	(0.685)	(0.767)
	[0.477]	[−0.024]	[0.108]	[−0.204]	[0.103]	[0.108]
on the free market	0.777	− 0.534	0.191	− 0.948	0.983	− 0.535
	(0.303)	(0.306)	(0.724)	(0.748)	(0.390)	(0.379)
	[0.274]	[−0.183]	[0.029]	[−0.278]	[0.124]	[−0.074]
Smoker	− 0.241	0.045	− 0.256	0.229	− 0.340	0.347
	(0.196)	(0.173)	(0.327)	(0.261)	(0.335)	(0.274)
	[−0.087]	[0.015]	[−0.036]	[0.069]	[−0.048]	[0.043]
Test*: χ^2_{26}	76.96		64.22		82.80	
($p-$value)	(0.0000)		(0.0004)		(0.0000)	
Observations	1787 (686)	1787 (658)	733	786	1054	968

Huber-White Standard errors in parenthesis. Marginal effects in square brakets.

*Joint test of equality in coefficient

Table 1.6: First stage probits, 1991, 1993

	Sample	By sector	
	total	public sector	private sector
	(1)	(2)	(3)
Control variables :			
age	0.1686 (0.4497)	0.1435 (0.1360)	0.0816 (0.2424)
public sector	−0.4522 (1.2776)		
state sector	−0.0040 (1.6330)	0.0796 (0.1163)	
urban	0.0854 (0.5963)	0.1392 (0.1843)	0.5547 (0.3196)
Human capital :			
BMI	−0.0627 (0.1920)	0.0351 (0.0442)	0.4857 (0.1275)
log kilocalories	−28.2568 (20.7818)	19.1448 (13.3316)	−243.4696 (16.4528)
log kilocalories squared	1.7805 (1.3160)	−1.2430 (0.8475)	15.3305 (1.0423)
Inverse Mills ratio : $\widehat{\lambda}_{it}$	−0.1990 (1.1292)	−0.1531 (0.2525)	0.0667 (0.3217)
Hansen test of overidentification: $p-value$	0.6571	0.4335	0.6646
Observations	1344	1142	202
Individuals	672	571	101
Observations in	1787 (686)	733	1054
first stage probits	1787 (658)	786	968

Huber-White Standard errors in parenthesis

Table 1.7: Returns to nutrition in China for the years 1991 and 1993 using the Semykina-Wooldridge estimator

Tables section 1.3

	Full sample			Subsample of wage earners		
Observations	3012			1732		
	mean	standard deviation		mean	standard deviation	
		total	"within" individual		total	"within" individual
hourly wage				2.9840	7.1488	4.6136
age	41.8529	12.4916	3.3543	40.4372	10.8865	3.3730
public sector	0.5621	0.4962	0.2263	0.8366	0.3698	0.2538
state sector	0.3735	0.4838	0.2538	0.5583	0.4967	0.3041
Human capital						
height	1.6647	0.0677	0.0094	1.6823	0.0639	0.0096
years schooling	8.6906	4.0538	0.8916	10.0572	3.8100	1.0446

Table 1.8: Summary statistics for urban dwellers differentiating non wage earners and wage earners

	1991	1993	1997	2000	2004
	(1)	(2)	(3)	(4)	(5)
age	−0.1747	−0.1283	−0.1150	−0.0155	−0.0728
	(0.0450)	(0.0345)	(0.0364)	(0.0407)	(0.0598)
	[−0.0631]	[−0.0481]	[−0.0449]	[−0.0058]	[−0.0229]
public sector	2.4699	1.5267	0.4731	1.7256	−0.0584
	(0.5806)	(0.3550)	(0.3065)	(0.3629)	(0.3636)
	[0.7831]	[0.5522]	[0.1839]	[0.5918]	[−0.0175]
state sector	−0.0708	−0.2159	1.0117	−0.3674	0.7948
	(0.4828)	(0.3362)	(0.3510)	(0.3021)	(0.5056)
	[−0.0256]	[−0.0809]	[0.3629]	[−0.1388]	[0.1864]
Identification variables					
Commodity prices					
F-test of joint significance	χ^2_8= 13.24	χ^2_8= 2.64	χ^2_8= 6.05	χ^2_8= 1.00	χ^2_8= 6.43
observations	608	730	664	601	409

Standard errors in parenthesis. Marginal effects in square brakets.

Table 1.9: First stage Mundlak probit estimations for 1991, 1993, 1997, 2000 and 2004

	OLS	FE	Wooldridge test	SW	SWDD
	(1)	(2)	(3)	(4)	(5)
age	0.0082	0.1752	0.1595	−0.0071	−0.0154
	(0.0013)	(0.0043)	(0.0054)	(0.0077)	(0.0122)
D93	0.3761			0.4017	0.4095
	(0.0348)			(0.0380)	(0.0456)
D97	1.3196			1.3922	1.4441
	(0.0360)			(0.0471)	(0.0802)
D00	1.6236			1.7088	1.7699
	(0.0406)			(0.0619)	(0.1130)
D04	1.8461			1.9961	2.0647
	(0.0561)			(0.0902)	(0.1642)
public sector	−0.0922	0.0986	0.2525	−0.0681	−0.0714
	(0.0579)	(0.0577)	(0.0689)	(0.0809)	(0.0807)
state sector	−0.1269	0.0532	0.0165	−0.1721	−0.1803
	(0.0303)	(0.0460)	(0.0502)	(0.0573)	(0.0576)
Human capital					
height	0.6096	1.5131	−1.9011	0.4778	−2.1714
	(0.2274)	(1.3183)	(2.4416)	(0.2451)	(1.3842)
years schooling	0.0200	0.0092	0.0455	0.0216	0.0466
	(0.0041)	(0.0124)	(0.0258)	(0.0044)	(0.0130)
Inverse Mills ratio:					
$\widehat{\lambda}_{it}$			0.6183	−0.1244	−0.1339
			(0.0949)	(0.0892)	(0.0924)
σ		0.5148	0.5111		
R^2	0.6571	0.0385	0.0384	0.6642	0.6373
Statistical tests:					
$F-$test: $\alpha_i = \alpha$		$F^{869}_{1142}= 5.13$	$F^{814}_{909}= 4.47$		
Hausman test		$\chi^2_7= 1224.67$	$\chi^2_8= 851.10$		
Partial r^2 Height					0.0278
Education					0.1305
Partial F_test Height					$F^7_{1700} = 3.27$
Education					$F^7_{1700} = 27.10$
Hansen test of overidentification					$\chi^2_5= 0.0794$
ratio $\frac{\sigma_{\alpha_i}}{\sigma_{\varepsilon_{it}}}$		4.1947			−0.9902
					(0.8389)
observations	2019	2019	1732	1732	1732
individuals		870	815		
observations in first stage probit(s)			3012	3012	3012

Huber-White standard errors in parentheses.

Table 1.10: Returns to human capital in China, 1991, 1993, 1997, 2000 and 2004. OLS, Fixed Effects, Wooldridge test for the presence of sample selection bias, and Semykina-Wooldridge estimator

	1991 – 1993			1997 – 2004		
	OLS	SW	SWDD	OLS	SW	SWDD
	(1)	(2)	(3)	(4)	(5)	(6)
age	0.0129 (0.0016)	0.0476 (0.0105)	0.0756 (0.0125)	0.0085 (0.0022)	0.0617 (0.0086)	0.0625 (0.0097)
public sector	−0.5758 (0.1157)	−0.8443 (0.1907)	−0.8197 (0.1966)	−0.0062 (0.0618)	−0.0392 (0.0802)	−0.0553 (0.0822)
state sector	−0.1882 (0.0453)	−0.0155 (0.0990)	−0.0600 (0.1068)	−0.1966 (0.0434)	−0.1861 (0.0673)	−0.1769 (0.0687)
Human capital						
height period 1	1.1018 (0.3302)	1.0936 (0.3521)	−1.7692 (4.1858)			
height period 2				0.3891 (0.3341)	0.2991 (0.3496)	−1.9022 (1.3778)
years schooling period 1	0.0117 (0.0049)	0.0137 (0.0054)	0.0276 (0.0196)			
years schooling period 2				0.0469 (0.0066)	0.0365 (0.0074)	0.0423 (0.0155)
Inverse Mills ratio:						
$\widehat{\lambda}_{it}$		−0.2473 (0.1554)	−0.2421 (0.1628)		−0.1273 (0.1093)	−0.1178 (0.1094)
R^2	0.2004	0.2540	0.3459	0.1148	0.1897	0.1569
observations	964	791	791	1055	941	941
observations in first stage probit(s)		1338	1338		1674	1674

Huber-White standard errors in parentheses.

Table 1.11: The returns to human capital in China, 1991-93 and 1997-2004 (1); 1991-97 and 2000-04 (2). Semykina-Wooldridge estimator

	1991 – 1997			2000 – 2004		
	OLS	SW	SWDD	OLS	SW	SWDD
	(1)	(2)	(3)	(4)	(5)	(6)
age	0.0116	0.1264	0.1538	0.0084	0.0542	0.0542
	(0.0019)	(0.0102)	(0.0122)	(0.0027)	(0.0140)	(0.0173)
public sector	−0.4806	−1.0346	−0.6048	0.0586	−0.0278	−0.0427
	(0.0937)	(0.1827)	(0.2224)	(0.0702)	(0.0911)	(0.0924)
state sector	−0.1657	−0.0852	−0.0783	−0.1247	−0.1220	−0.1274
	(0.0507)	(0.1014)	(0.1172)	(0.0559)	(0.0781)	(0.0799)
Human capital						
height period 1	1.5326	1.2579	9.1448			
	(0.3677)	(0.3733)	(4.2151)			
height period 2				0.6078	0.5050	−1.0828
				(0.4213)	(0.4253)	(1.4858)
years schooling period 1	0.0296	0.0274	0.0773			
	(0.0057)	(0.0059)	(0.0238)			
years schooling period 2				0.0445	0.0419	0.0381
				(0.0088)	(0.0094)	(0.0188)
Inverse Mills ratio:						
$\widehat{\lambda}_{it}$		−0.7414	−0.3033		−0.0432	−0.0475
		(0.1619)	(0.1852)		(0.1410)	(0.1390)
R^2	0.1559	0.3056	0.0247	0.6438	0.1700	0.1529
observations	964	1139	1139	640	593	593
observations in first stage probit(s)		2002	2002		1010	1010

Huber-White standard errors in parentheses.

Table 1.12: The returns to human capital in China, 1991-93 and 1997-2004 (1); 1991-97 and 2000-04 (2). Semykina-Wooldridge estimator

Figures chapter 1

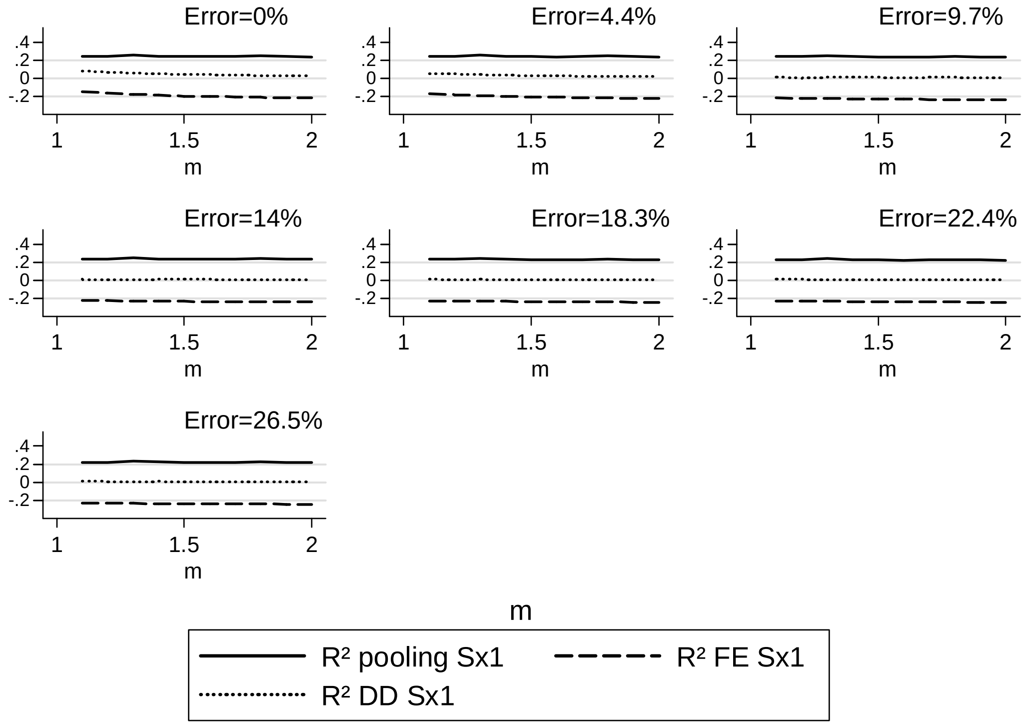


Figure 1-3: Montecarlo simulation of R^2 for the scenario S_{11}^A

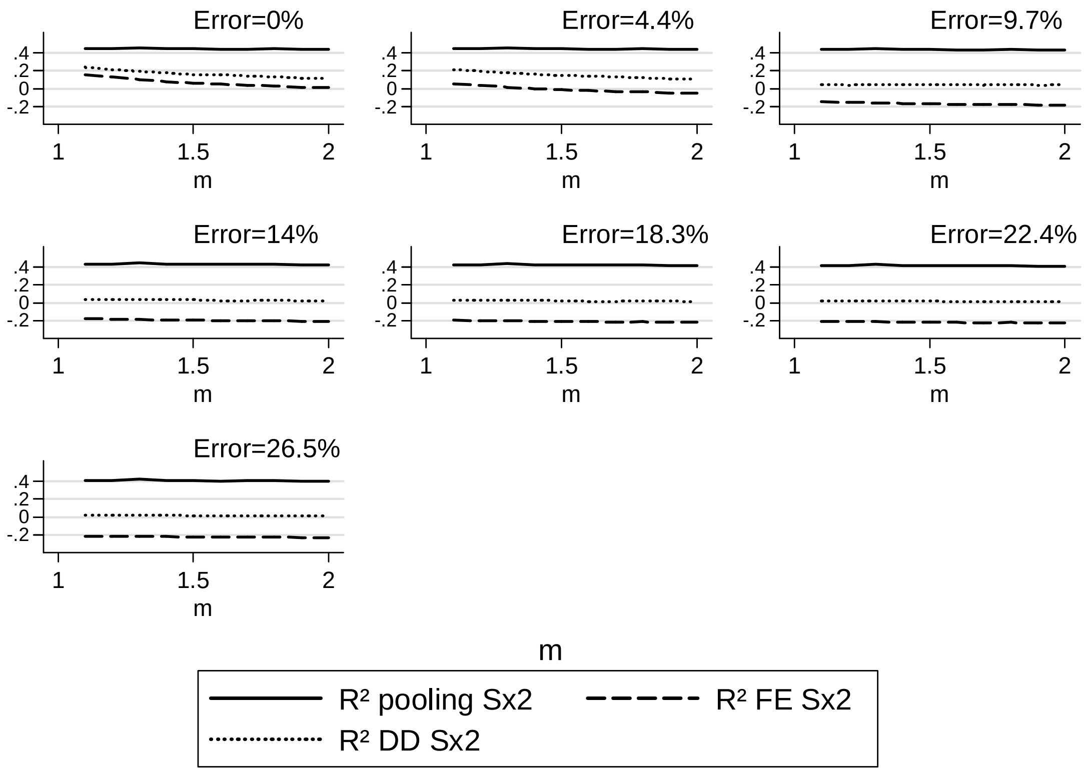


Figure 1-4: Montecarlo simulation of R^2 for the scenario S_{12}^{A}

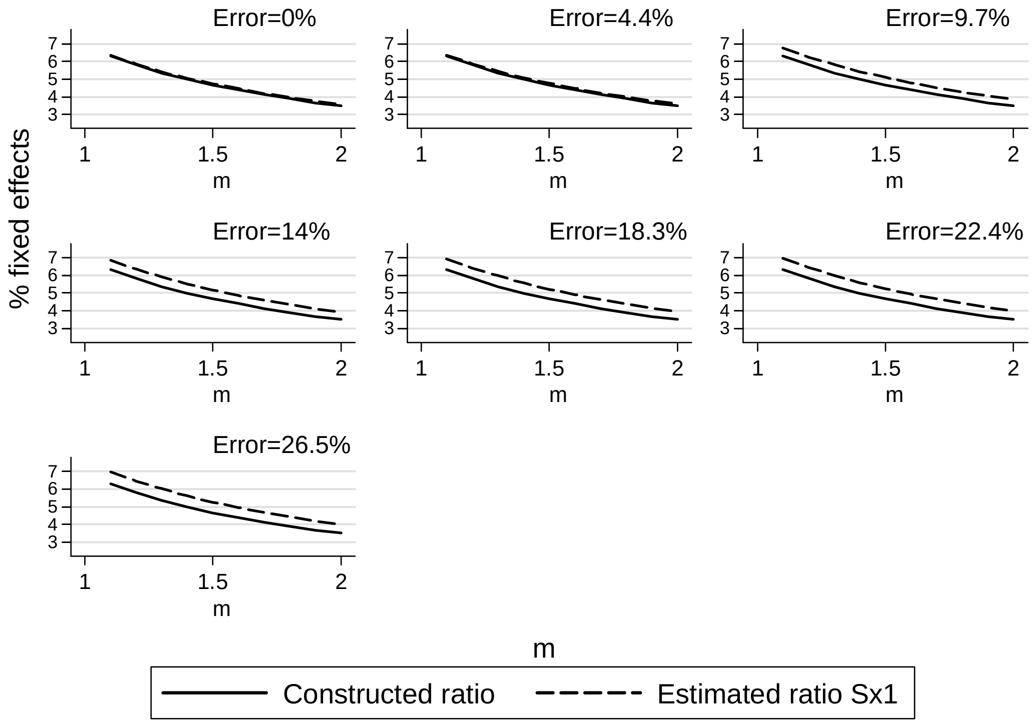


Figure 1-5: Montecarlo simulations of standard deviations ratio for the scenario S_{11}^{A}

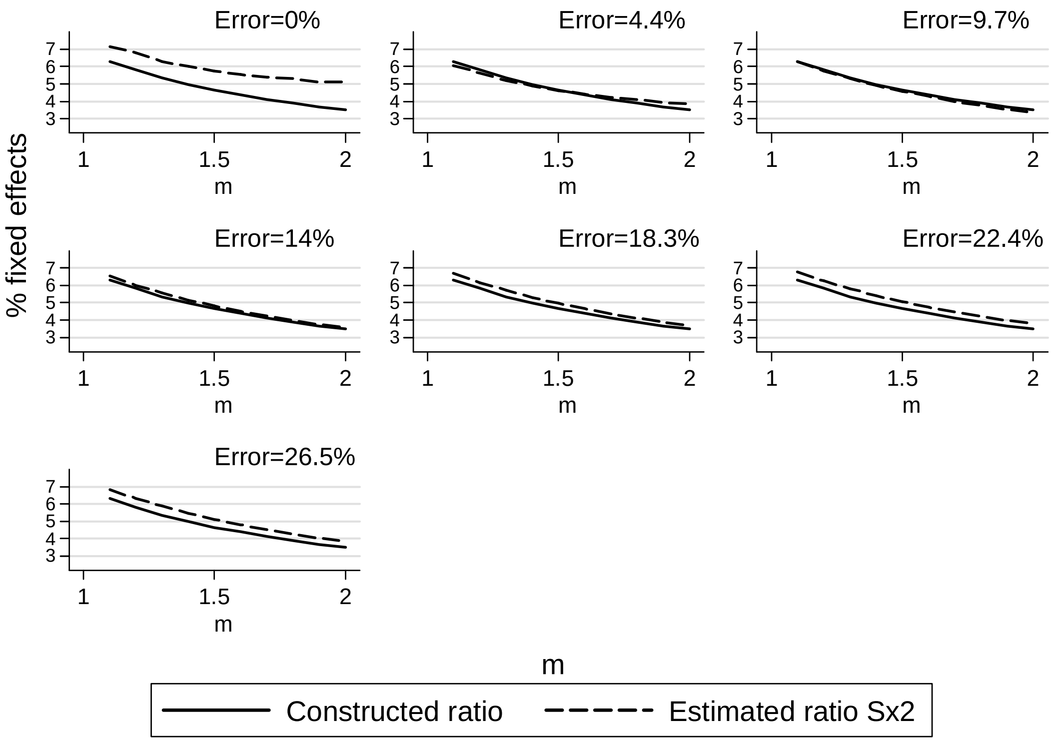


Figure 1-6: Montecarlo simulations of standard deviations ratio for the scenario S_{12}^{A}

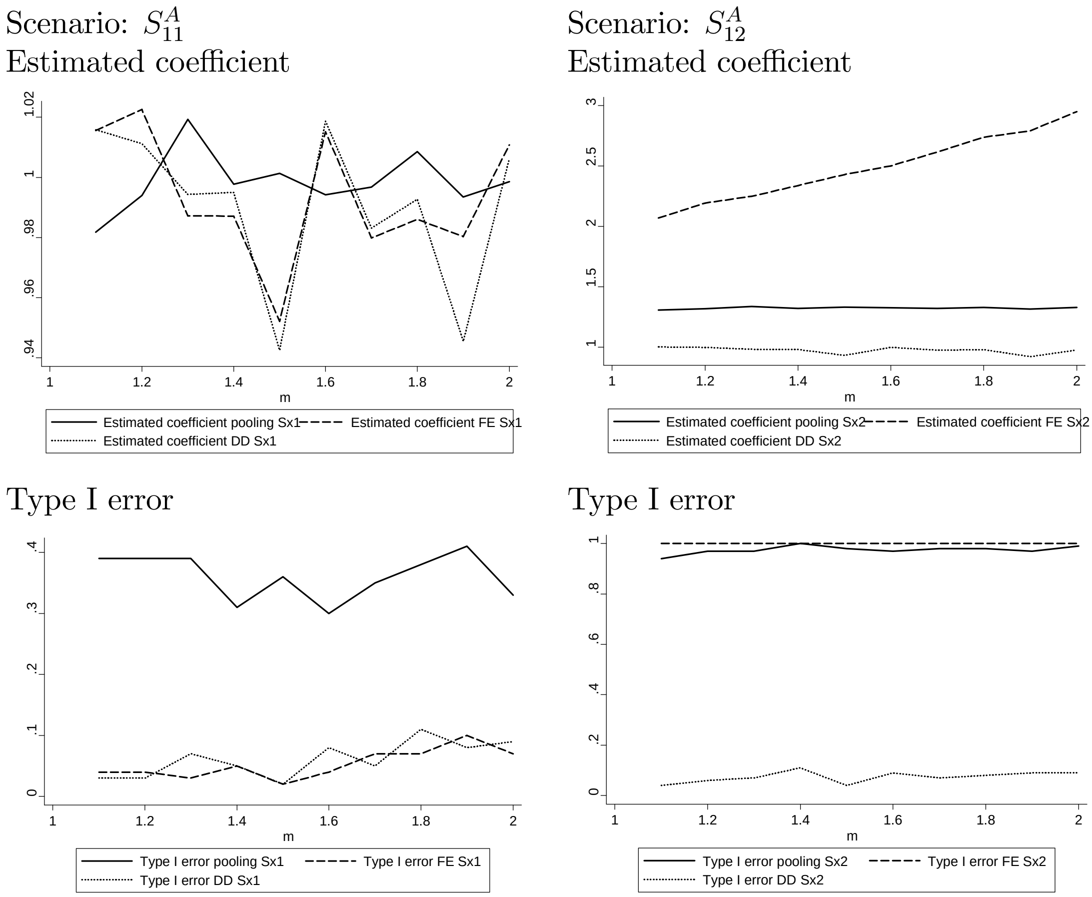


Figure 1-7: Montecarlo simulations of coefficient and type I error for a nul error and for the scenarii S_{11}^A and S_{12}^A

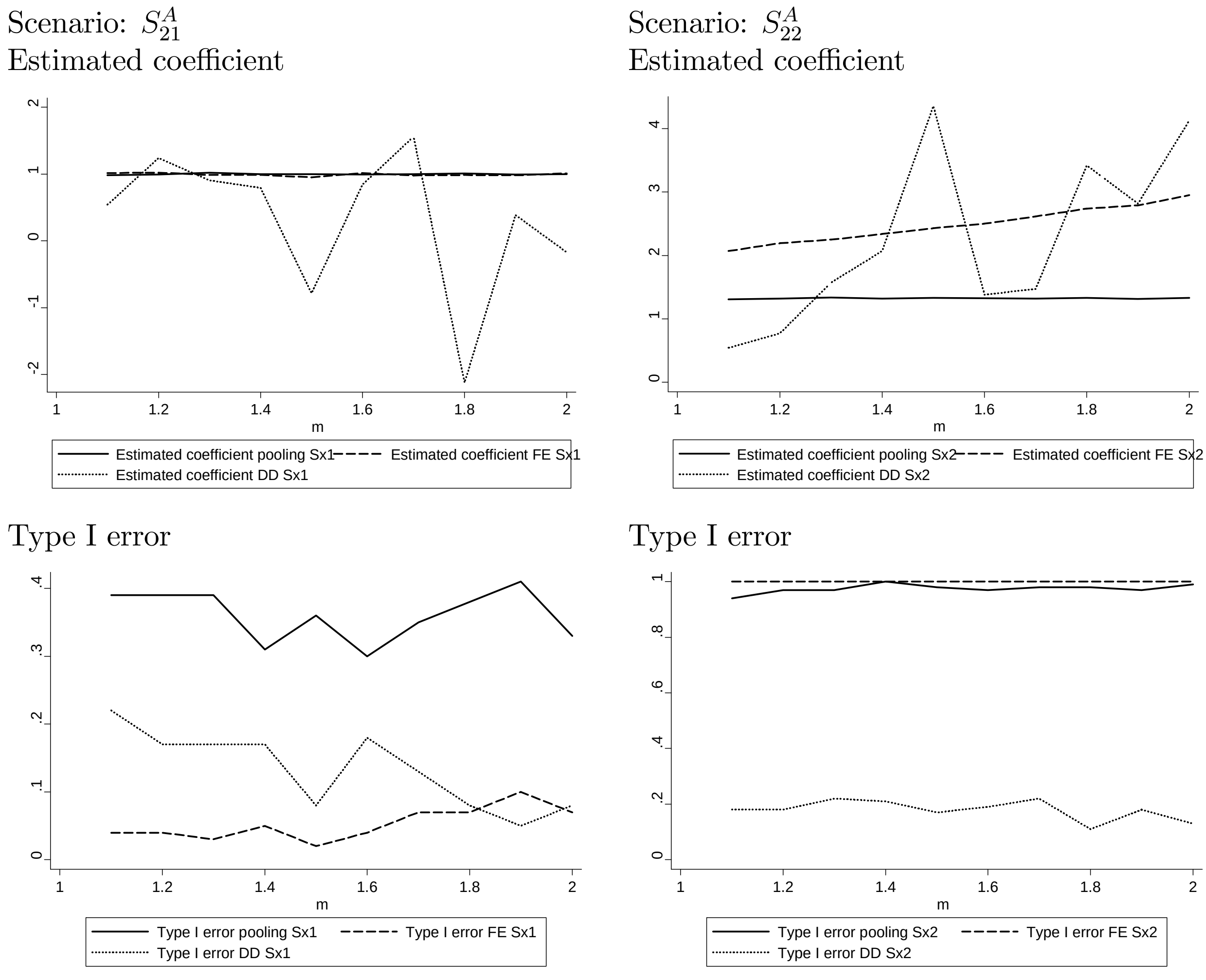
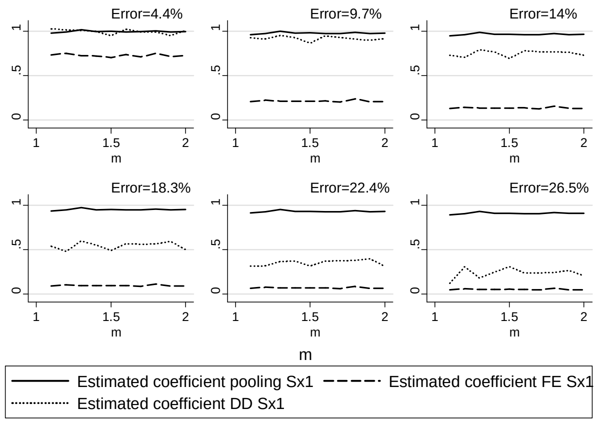
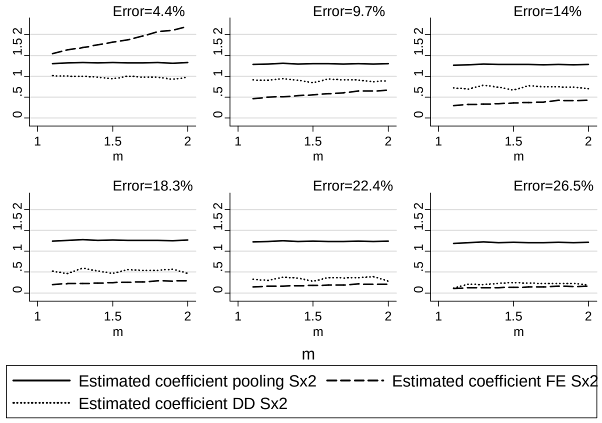


Figure 1-8: Montecarlo simulations of coefficient and type I error for a nul error and for the scenarii S_{21}^A and S_{22}^A

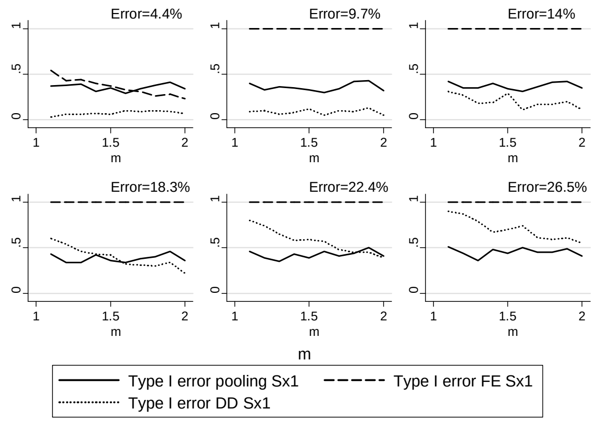
Scenario: S_{11}^{A}
Estimated coefficient



Scenario: S_{12}^{A}
Estimated coefficient



Type I error



Type I error

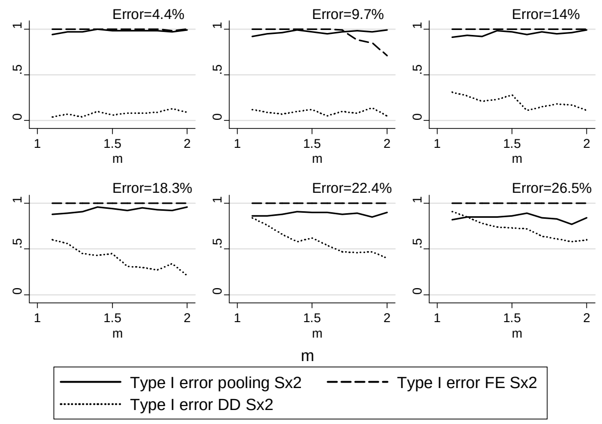
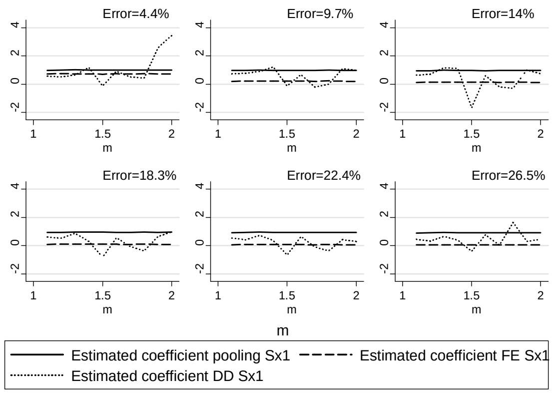
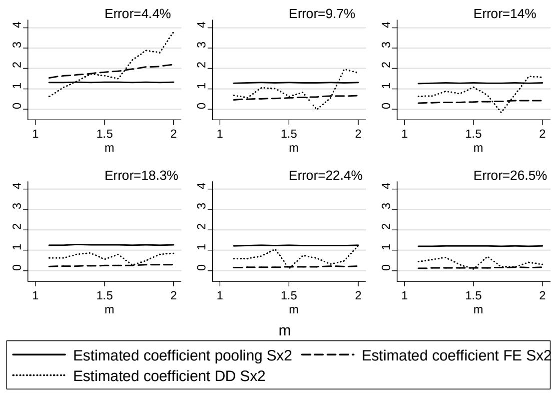


Figure 1-9: Montecarlo simulations of coefficient and type I error for the scenarii S_{11}^{A} and S_{12}^{A}

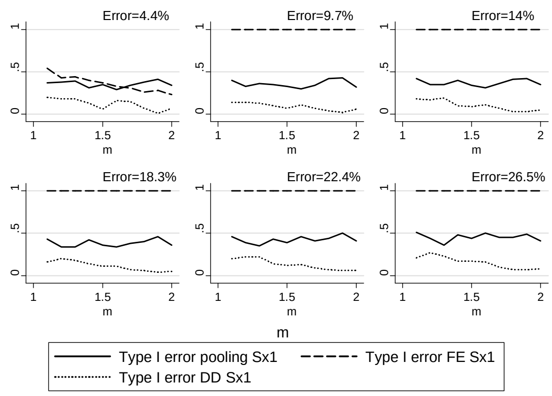
Scenario: S_{21}^{A}
Estimated coefficient



Scenario: S_{22}^{A}
Estimated coefficient



Type I error



Type I error

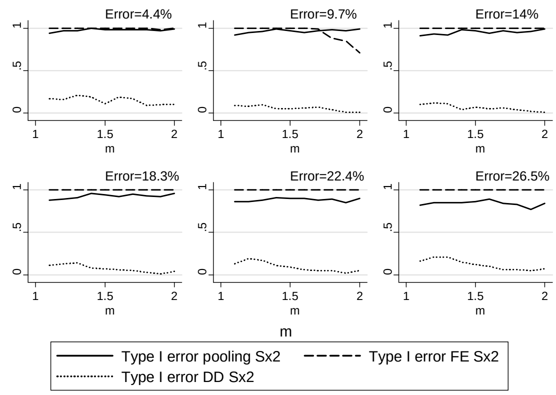


Figure 1-10: Montecarlo simulations of coefficient and type I error, for the scenarii S_{21}^{A} and S_{22}^{A}

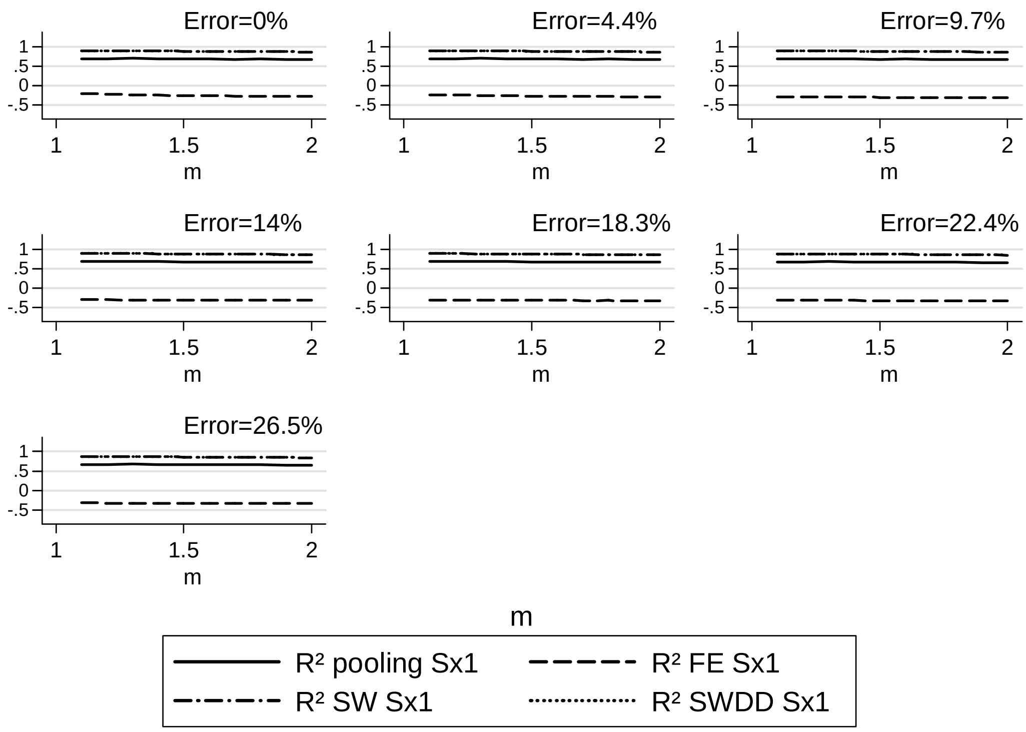


Figure 1-11: Montecarlo simulations of R^2 for the scenario S_{11}^B

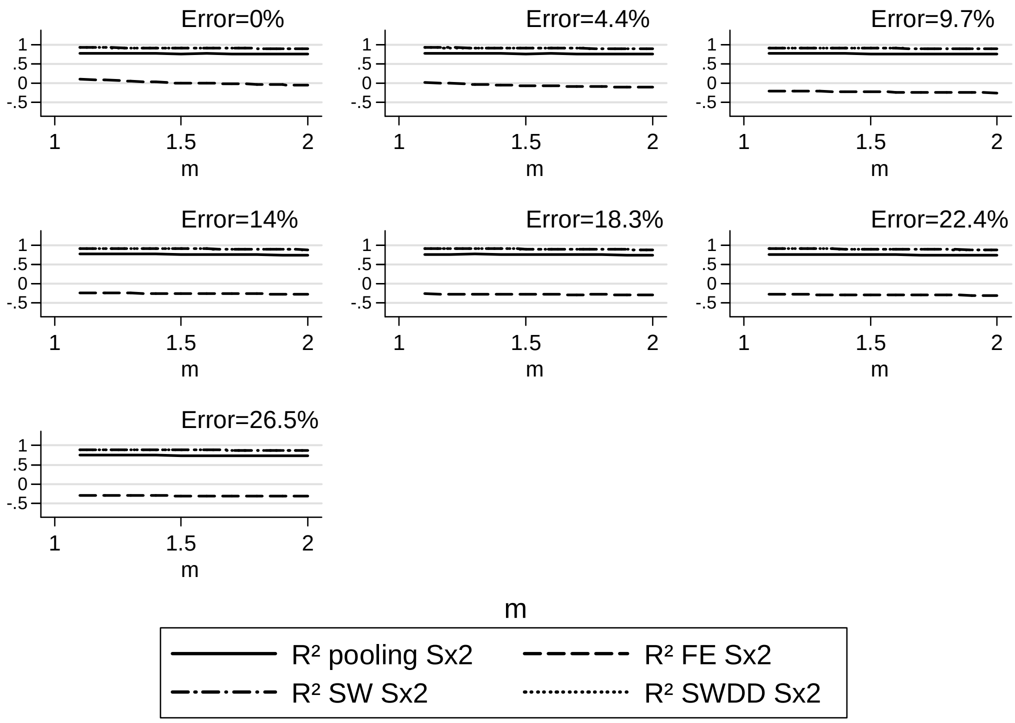


Figure 1-12: Montecarlo simulations of R^2 for the scenario S_{12}^B

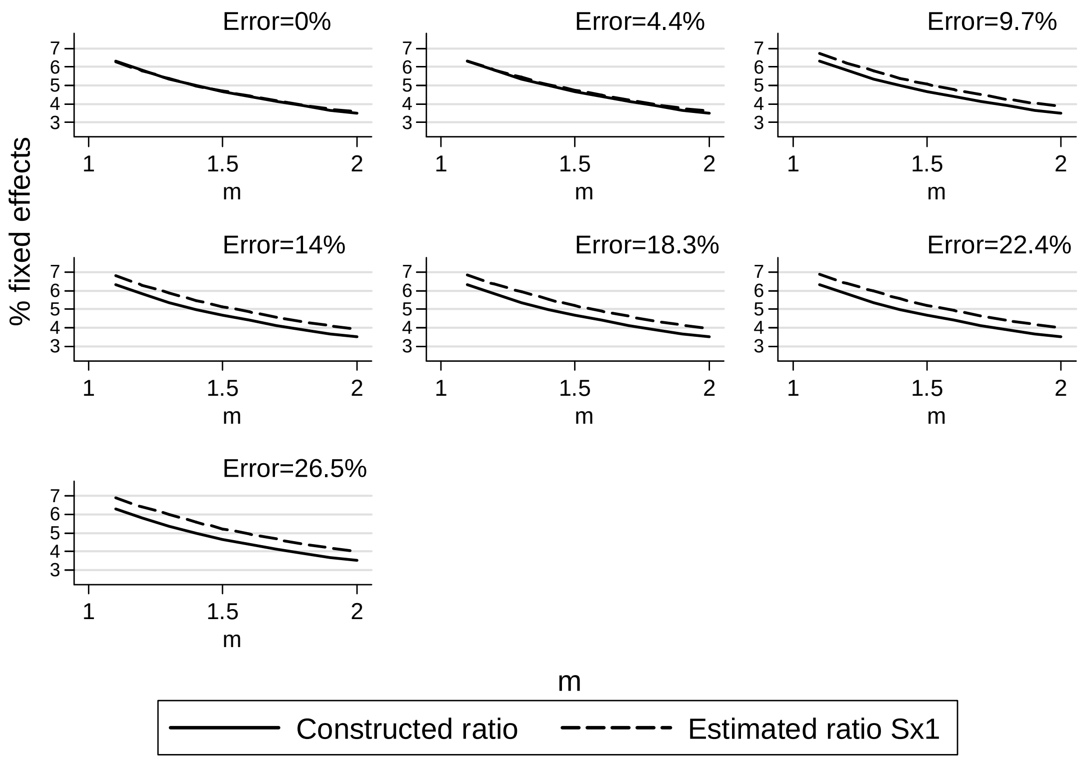


Figure 1-13: Montecarlo simulations of standard deviations ratio for the scenario S_{11}^{B}

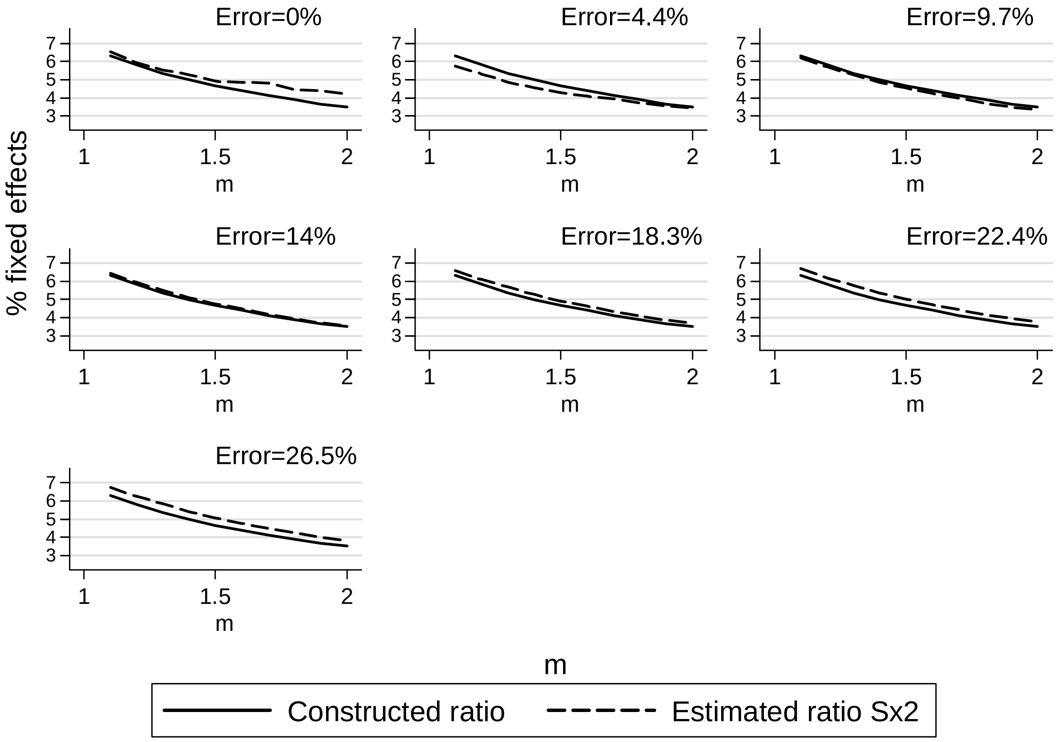


Figure 1-14: Montecarlo simulations of standard deviations ratio for the scenario S_{12}^{B}

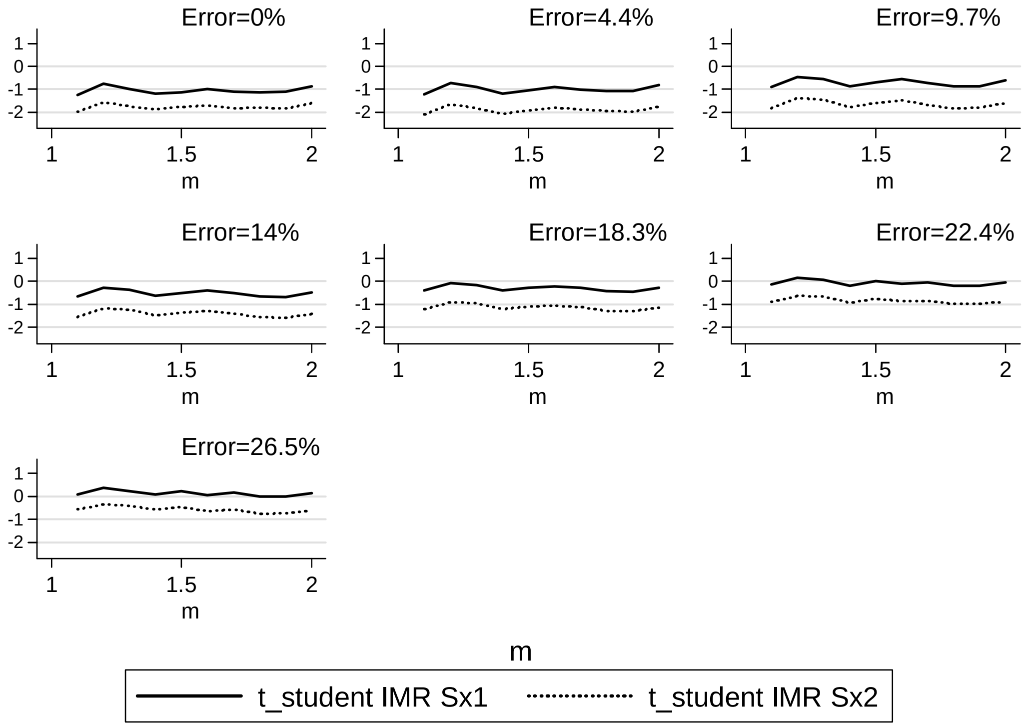


Figure 1-15: Montecarlo simulations of the IMR significance (t-student) for the scenarii S^B_{11} and S^B_{12}

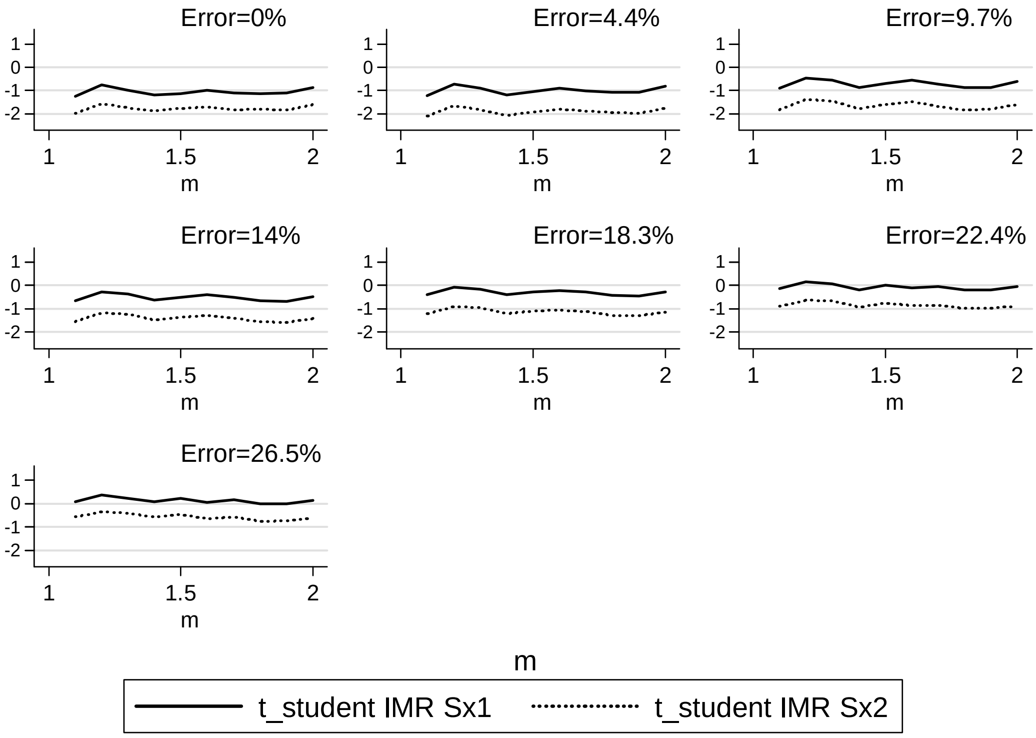


Figure 1-16: Montecarlo simulations of the IMR significance (t-student) for the scenarii S_{21}^{B} and S_{22}^{B}

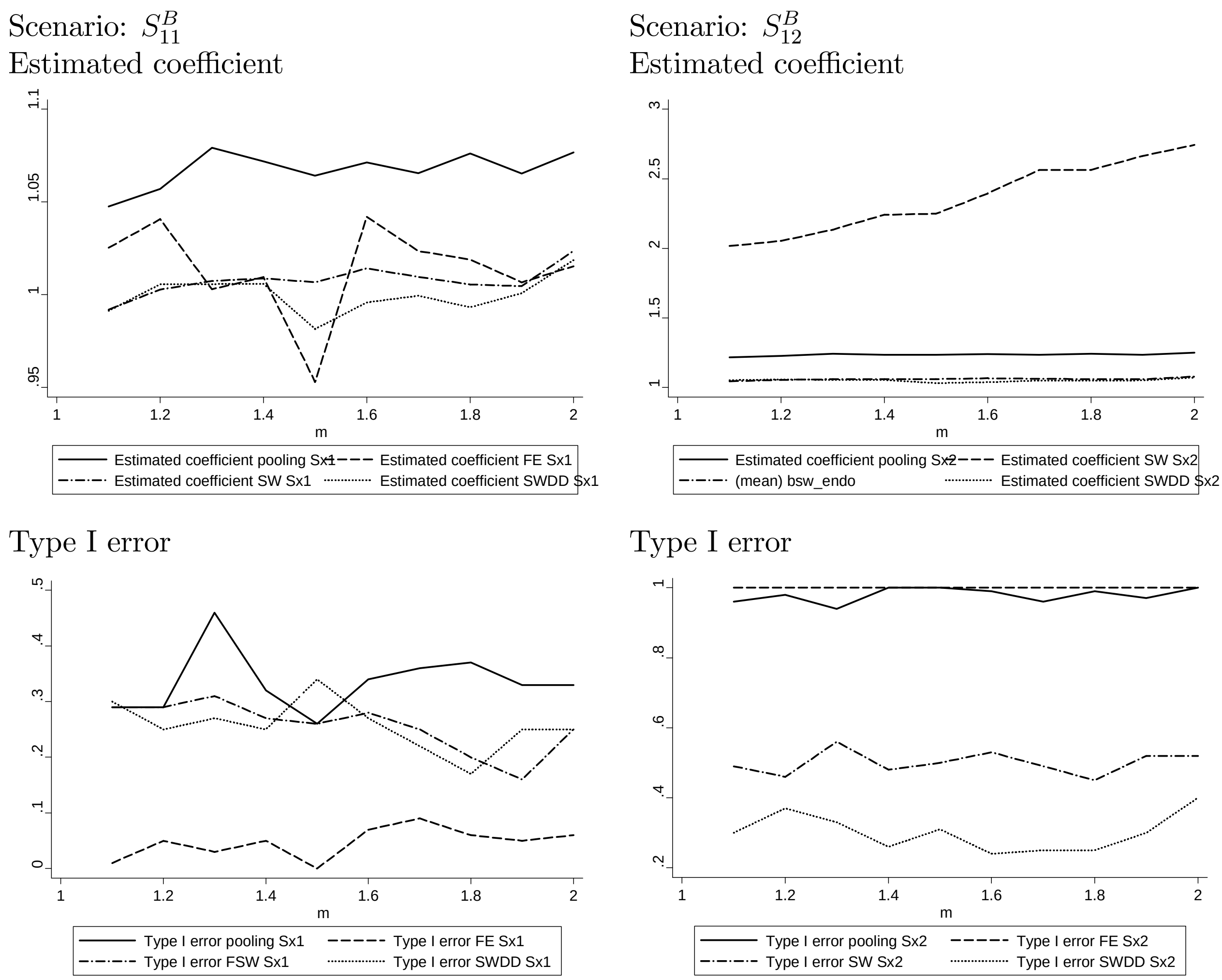


Figure 1-17: Montecarlo simulations of coefficient and type I errors for a nul error and for the scenarii S_{11}^B and S_{12}^B

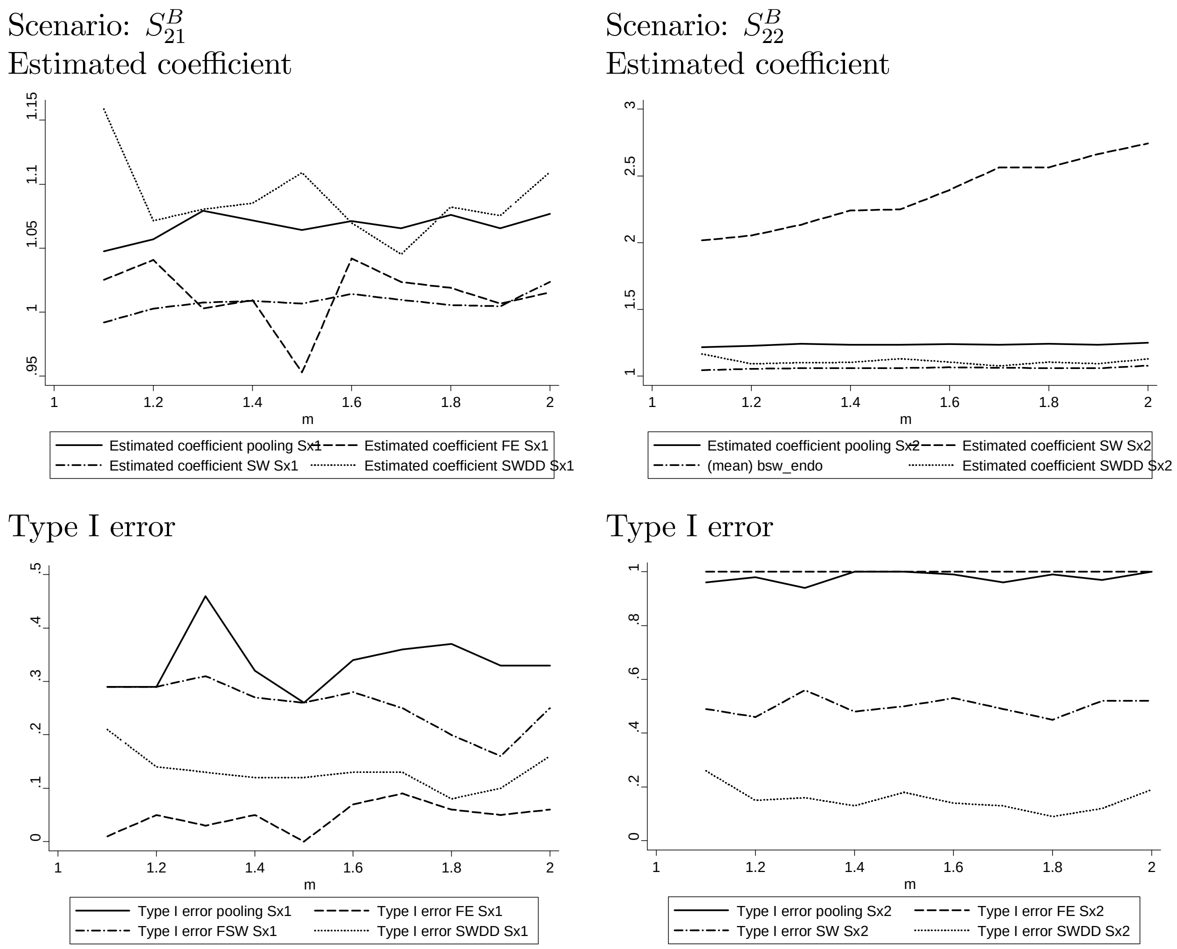
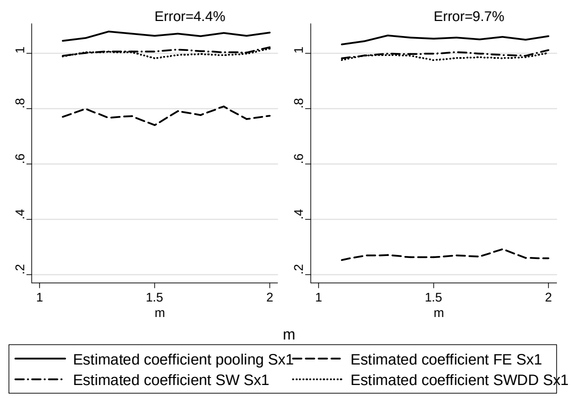
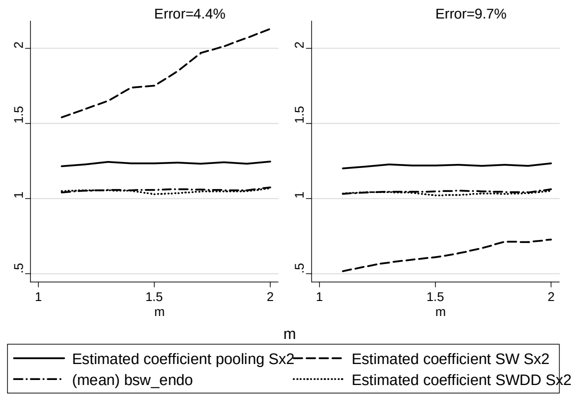


Figure 1-18: Montecarlo simulations of coefficient and type I errors for a nul error and for the scenarii S_{21}^B and S_{22}^B

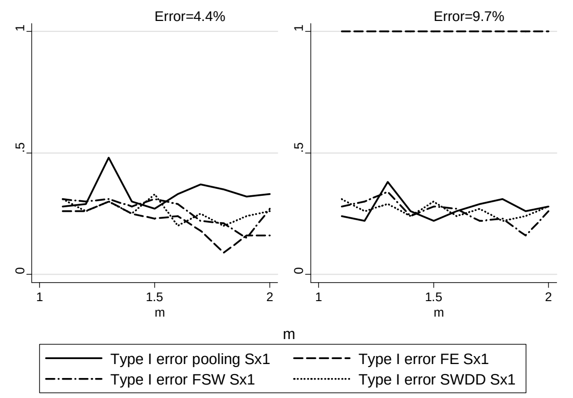
Scenario: S_{11}^{B}
Estimated coefficient



Scenario: S_{12}^{B}
Estimated coefficient



Type I error



Type I error

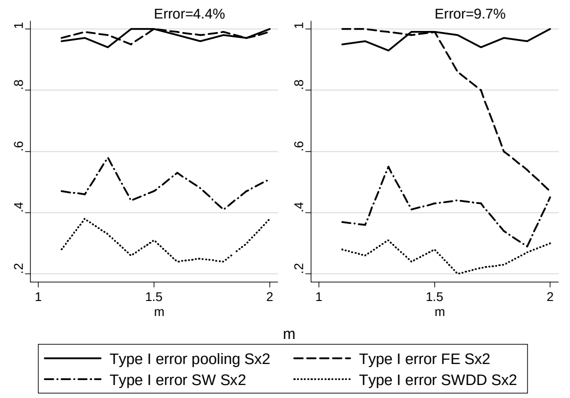
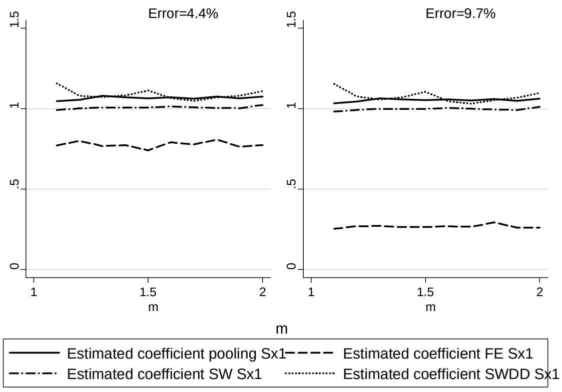
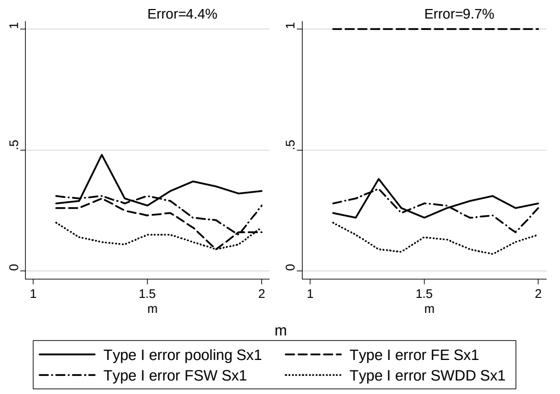


Figure 1-19: Montecarlo simulations of coefficients and type I errors for low error levels and for the scenarii S_{11}^{B} and S_{12}^{B}

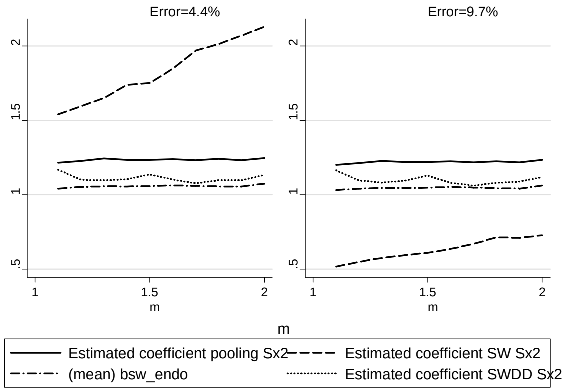
Scenario: S_{21}^{B}
Estimated coefficient



Type I error



Scenario: S_{22}^{B}
Estimated coefficient



Type I error

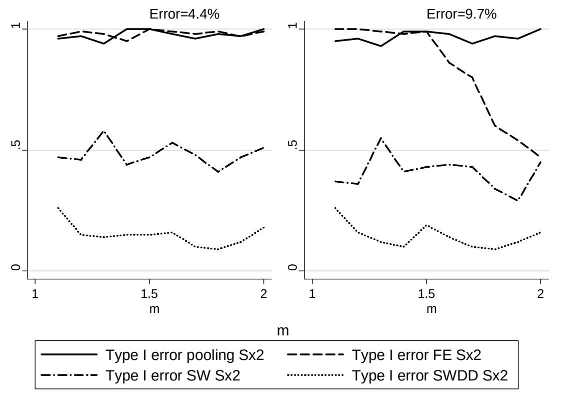
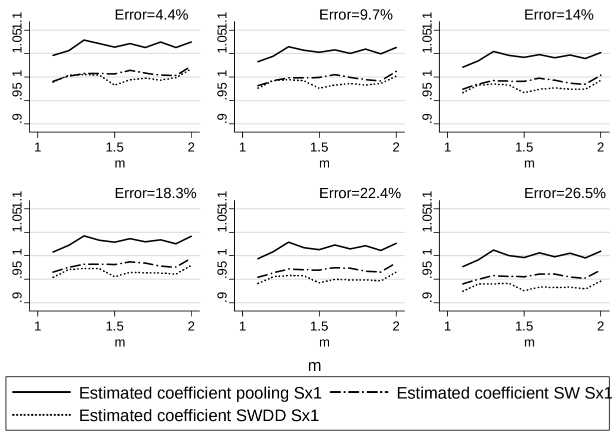
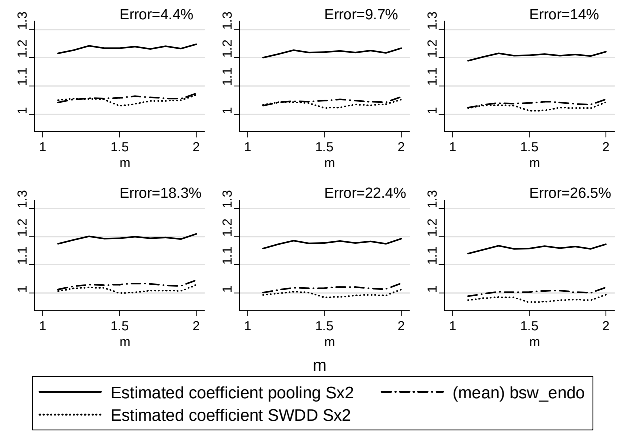


Figure 1-20: Montecarlo simulations of coefficients and type I errors for low error levels and for the scenarii S_{21}^{B} and S_{22}^{B}

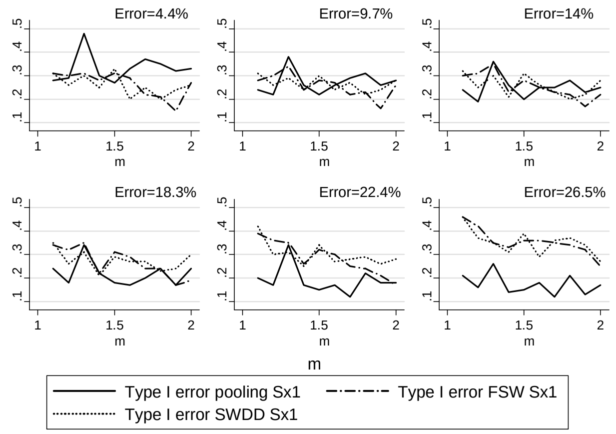
Scenario: S_{11}^{B}
Estimated coefficient



Scenario: S_{12}^{B}
Estimated coefficient



Type I error



Type I error

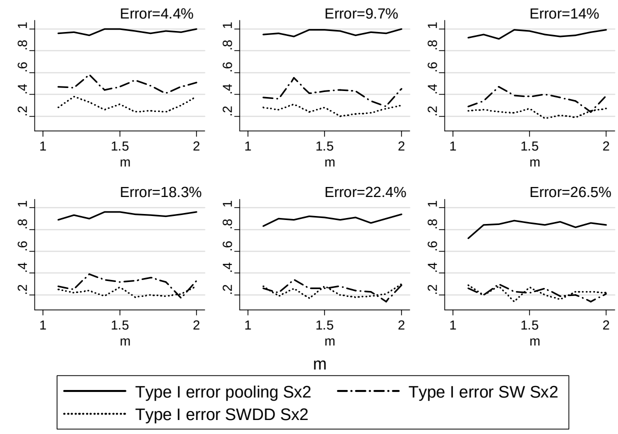
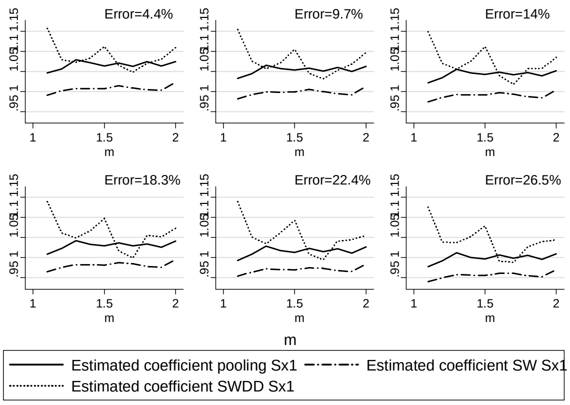
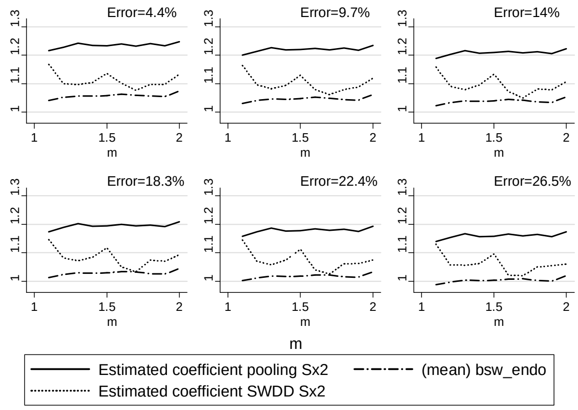


Figure 1-21: Montecarlo simulations of coefficients and type I errors without panel for the scenarii S_{11}^{B} and S_{12}^{B}

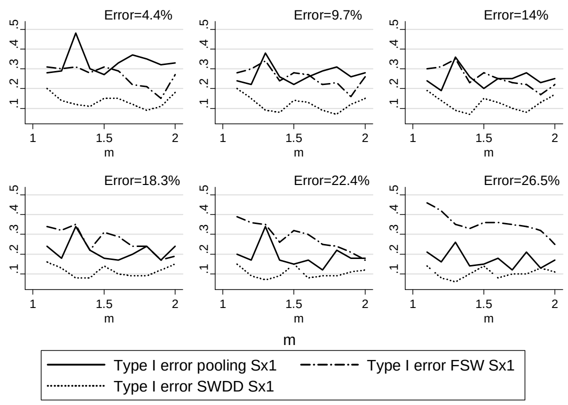
Scenario: S^B_{21}
Estimated coefficient



Scenario: S^B_{22}
Estimated coefficient



Type I error



Type I error

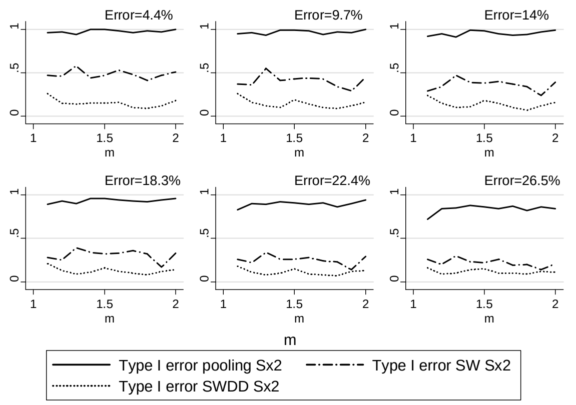


Figure 1-22: Montecarlo simulations of coefficients and type I errors without panel for the scenarii S^B_{21} and S^B_{22}

1.A Appendices: Chapter 1

1.A.1 Map of Chinese provinces included in the CHNS survey

Figure 1-23: Map of the surveyed regions

1.A.2 Results of the Montecarlo simulations for the $[z_3, z_7]$ set of instruments

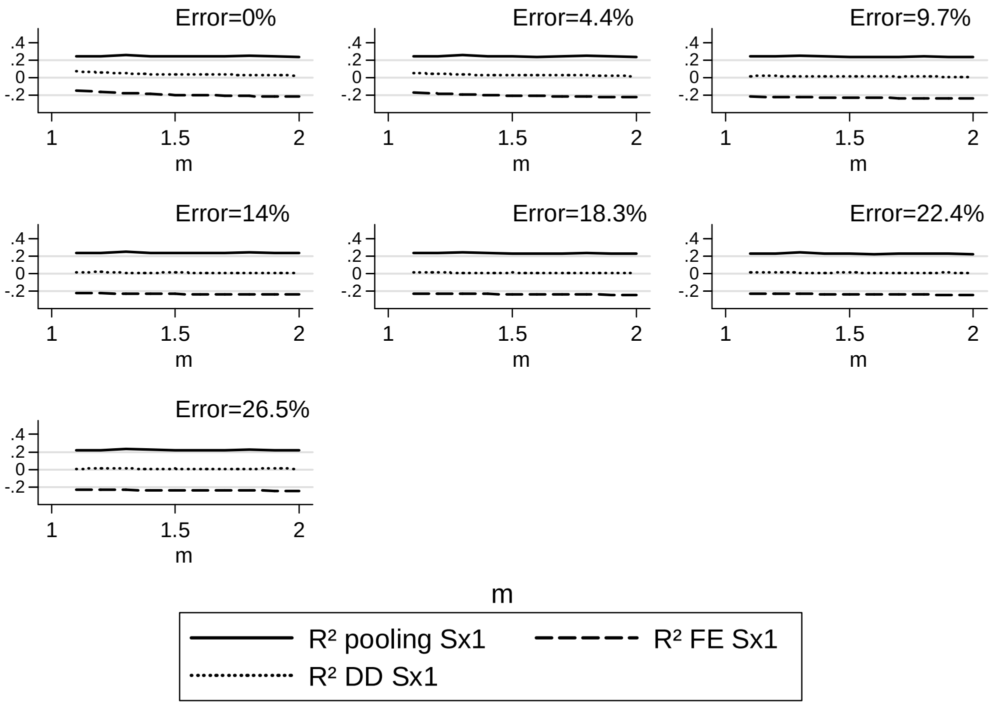


Figure 1-24: Montecarlo simulations of R^2 for the scenario S_{21}^A

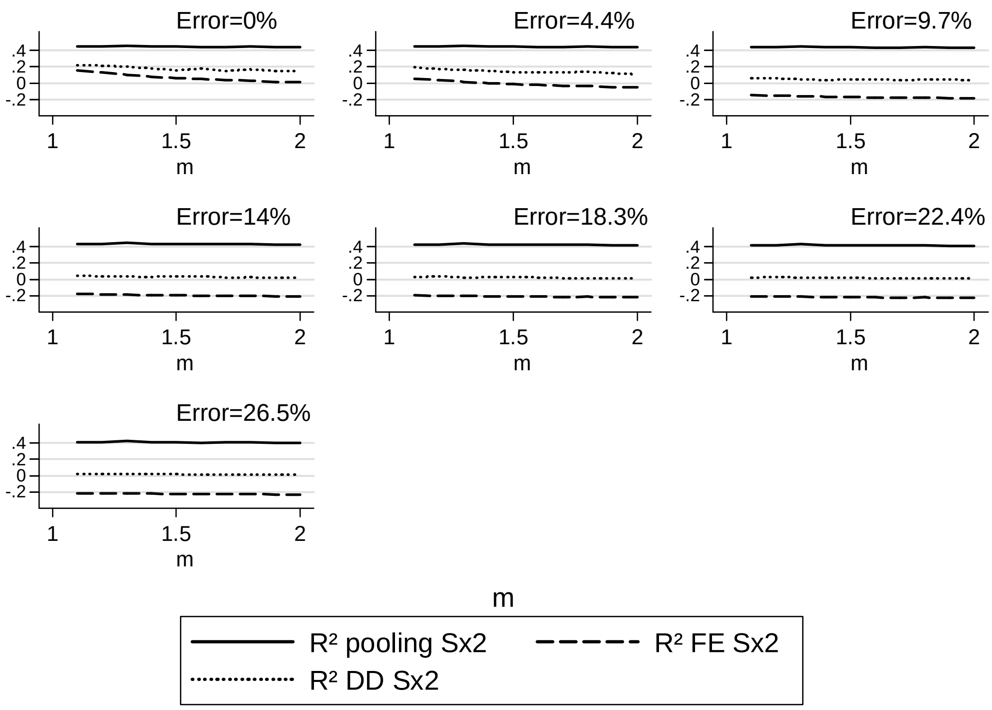


Figure 1-25: Montecarlo simulations of R^2 for the scenario S_{22}^A

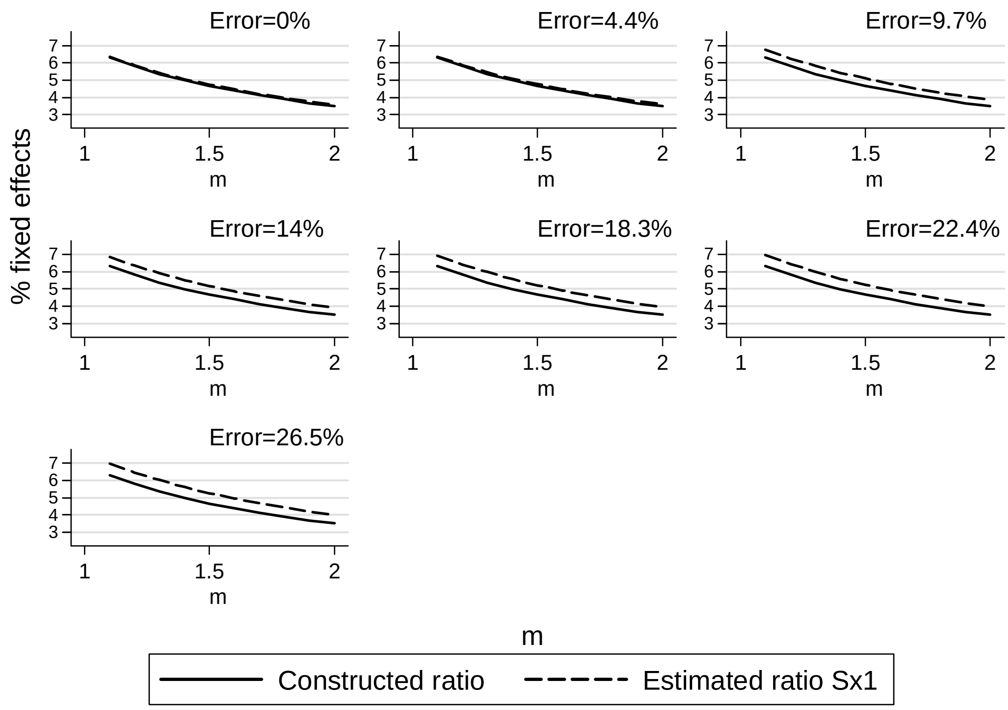


Figure 1-26: Montecarlo simulations of standard deviations ratio for the scenario S_{21}^{A}

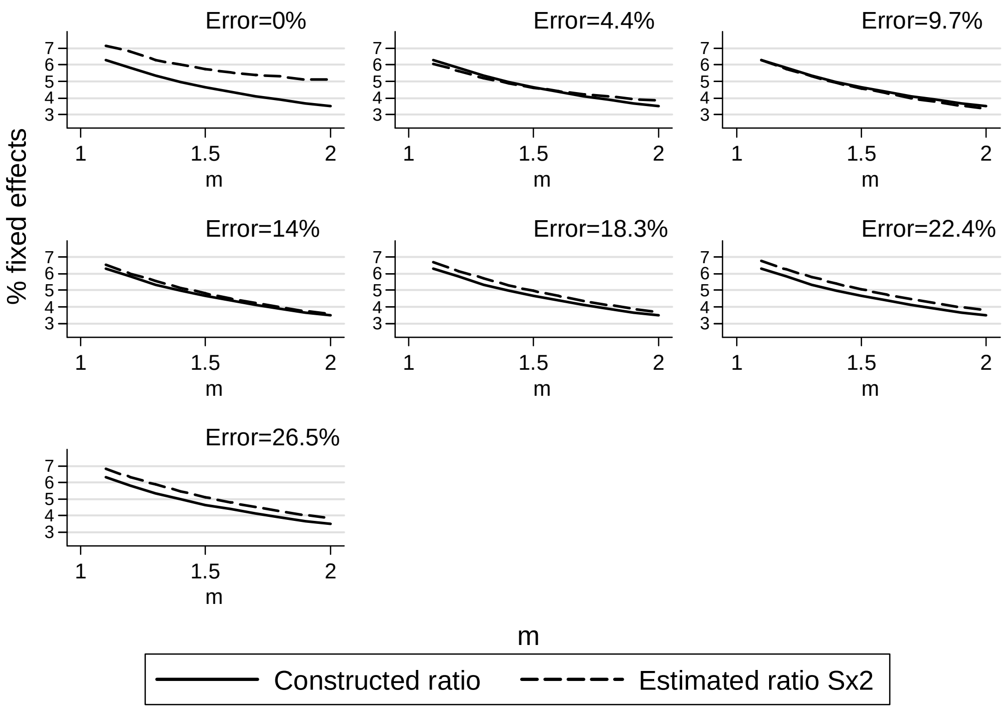


Figure 1-27: Montecarlo simulations of standard deviations ratio for the scenario S_{22}^{A}

1.A.3 Results of the Montecarlo simulations for the $[z_3, z_7]$ set of instruments with selection

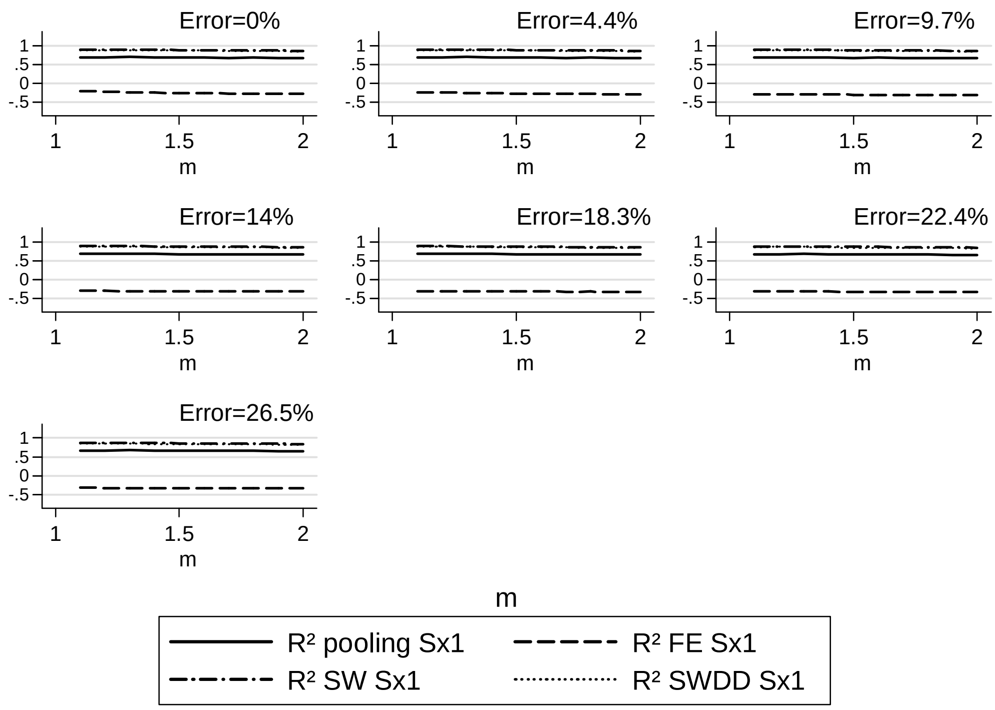


Figure 1-28: Montecarlo simulations of R^2 for the scenario S_{21}^{B}

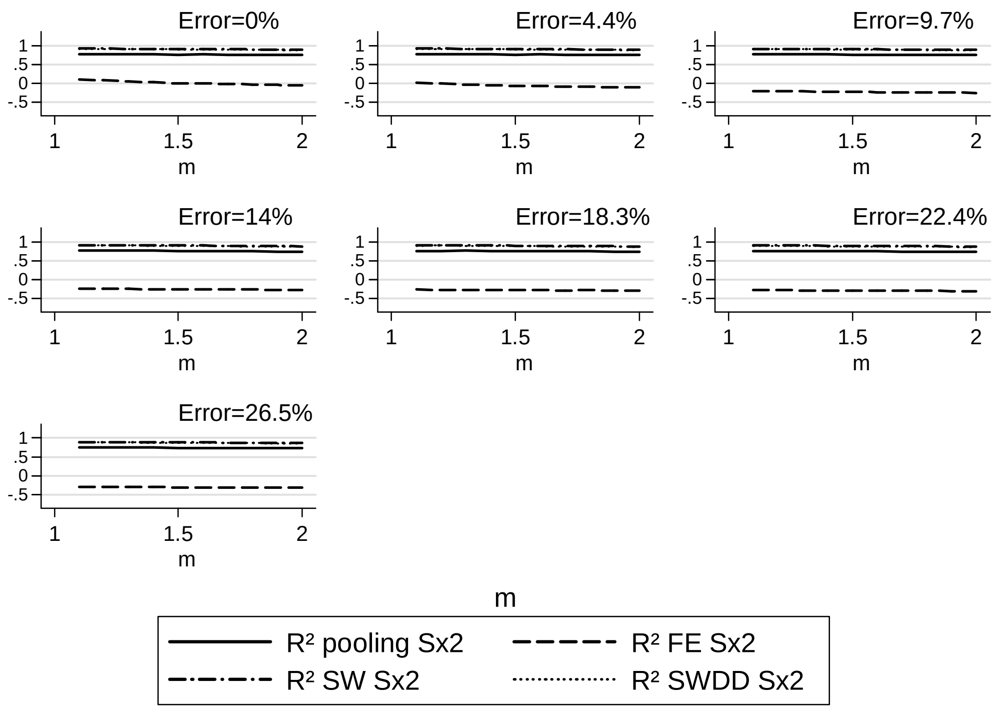


Figure 1-29: Montecarlo simulations of R^2 for the scenario S_{22}^{B}

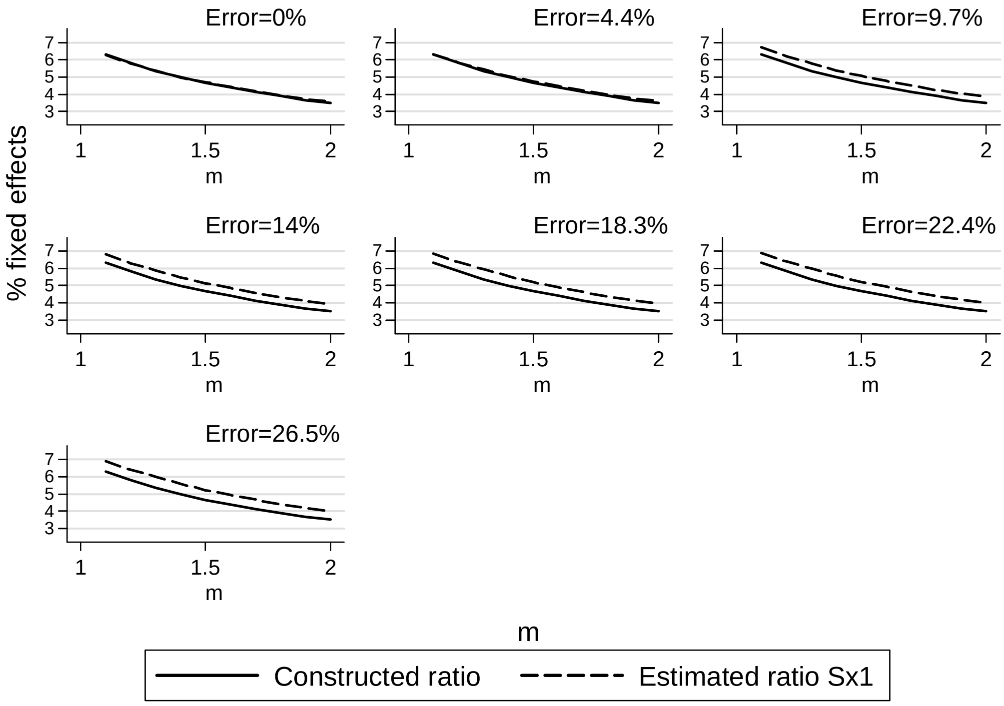


Figure 1-30: Montecarlo simulations of standard deviations ratio for the scenario S_{11}^{B}

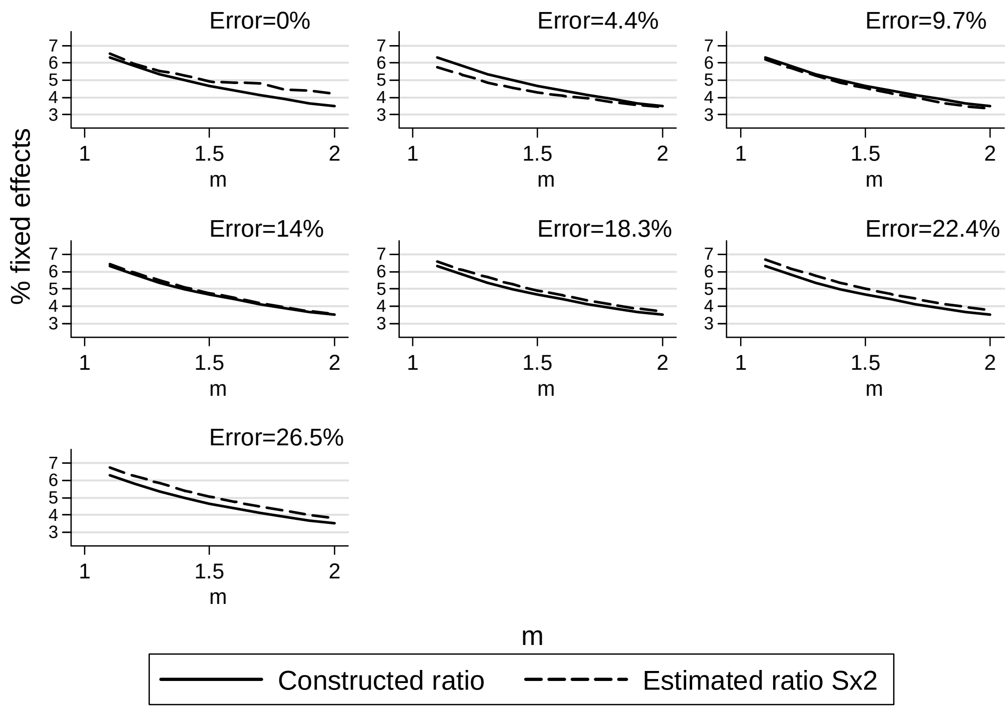


Figure 1-31: Montecarlo simulations of standard deviations ratio for the scenario S_{22}^{B}

Chapter 2

Decreasing intergenerational mobility in China

I have stressed in the preceding chapter the increasing role played by human capital in the determination of wages in China. Consequently, if indiviuals had equal access to education, this evolution would help them to predict or to choose the level of wages they would have access to. Nevertheless, this potential evolution is based on the hypothesis of equal opportunity in the access to schools and to employment. This condition is however not always respected once we take into account the environment in which individuals build their social identity. Particularly, the parents' social characteristics can have an impact on the child's through their level of education, their professionnal network or through their financial or time investment in the child's development. Intergenerational transmission of the social status can consequently occur, reinforcing some pre-existing intragenerational inequalities.

This transmission, which defines no or low social mobility between generations, is also a function of the political and social context. The involvement of the State in the financing of the education system, for example, plays an important role in the level of equality of opportunities in terms of the access to education. This factor is central in China where the government gets more or less involved given the evolution of the political and economic environment. Particularly, the movement of reforms launched by the end of the 70's has highly influenced and still influences individuals' level of equal access to the schooling system and to the labour market, giving more space for a larger intervention of the familly in the child's life. Consequently, it is expedient to wonder if the parental social status will, through their various investment level in their child's education, or through their wages, influence the child's level of education and income.

In order to analyse intergenerational mobility in China, I will at first present in section 1, its historical evolution until now. Indeed, mostly following political movements, Chinese history demonstrates non linearities in the role played by the parents' characteristics in the determination of the child's social status. On the one hand, China is the country which established the competitive examination system based on the concept of meritocracy, but on the other hand, social origins have always influenced the following generation's one, with different classes privileged given the political dominance. I will then focus on the reform movement and its consequences on the determinants of intergenerational mobility.

In section 2 and 3, I study intergenerational mobility relative to two dimensions: education and wages.[1] In section 2, I will distinguish between what I choose to call absolute and conditional mobility. These concepts are based on the difference between an objective mobility of children (they are not in the same quantile as their parents in terms of education or wages), and the fact that the parents' social characteristics are determinant for this absolute mobility. I will give the definitions used in this linear approach of intergenerational mobility and the quantifications of absolute and conditional mobility.

Section 3 will be dedicated to the non linearities which can exist between children and parents' characteristics, presenting non parametric relationships.

[1] The choice of wages as the income dimension is based on data quality and availability. Broader explanations are given in the paragraph dedicated to the data used.

2.1 Historical background

2.1.1 Before the Communist revolution

Intergenerational mobility during the dynasties

Historically, the Confucian dualistic social concept was that inequalities are inherent to any society, but these inequalities could not be based on an ad-hoc hierarchy, rather on individual merit. Ho (1959) underlines that the medieval hereditary aristocracy had been broken throughout the Tang dynasty (618-907 A D.) during which began the competitive examination system. Social mobility started to increase during this era to reach its maximum during the Ming and Qing dynasties (1368-1644 and 1644-1911 A D). Upward mobility relied on the examination system , which allowed individuals to have access to high level functions, whatever their parents' social status. During this period, we could also notice downward mobility. Families which could sometimes not assume their responsibilities were downgraded.

Social mobility does not mean that there was no hierarchical structure in the imperial periods as it is one of the elements underpinning the birth of communist's movements.

The 1911 Republic

During the Republic, proclaimed in 1911 and which lasted until 1949, the country was divided between the nationalist party (Guomindang) lead by Chang Kai Chek and a communist spirit born from both the fight against the traditional Confucian hierarchy (whose symbol can be the anti-imperialist 4th of May Movement) and the neighbouring Soviet revolution of 1917.

The young republic, which decided to move toward a market economy, faced both civil wars launched by the corrupted warlords in the north of China, and the foreign protagonists who invaded the country during this period. In parallel to the fight for national peace, a trade system was developing in China and between China and western countries, and even Japan. The communist party was gaining support in Chinese society.

During this period, it appears that social mobility was low. A corrupted and inefficient government helped the empowerment of some trading actors. A new aristocratic class, with living conditions totally opposed to the miserable countryside livelihood, emerged. Opposed to this new unequal social system, labourers' movements appeared in some rich towns like Shanghai.

The evolution of the familial context also influences the child's mobility. The Republic environment emphasized family links. Most children's life events were decided by key individuals of decisions in the household, who usually were the elders, or the parents. Mobility between social classes was consequently impeded by the richer families who helped their children to have access to higher paid jobs, or chose better schools and same high social class spouses for them (Engel (1984), Lindbeck (1951)). As emphasized in some modern studies on intergenerational mobility (Holmlund (2006)), assortative mating effects plays a negative role on mobility. The fact that the central wedding decision was in the hand of the family could reinforce the low mobility in China during the imperial and republic periods.

2.1.2 Between 1949 and the end of the seventies

When the communist party and its leader Mao Tse Dong arrived at the head of the country, an important campaign against traditional hierarchy and "family individualism" was launched (Lindbeck (1951)), aimed at reducing the historically unequal society and the family power on the Child's decisions.

The transmission of status in terms of wages or education tended to be relatively low during the communist period, emphasizing social mobility. The challenge of the traditional hierarchy gave privileges to usually isolated social classes such as peasants. A deep social transformation occured modifying the hierarchical social class structure and leading to some social promotions. This had consequences on children's equal access to education, thus impacting also equal access to employment. As Walder (1989) explains, two groups were overrepresented in the best high schools: children of party officials and military officers, and children of educated middle-class background. Both of these two groups of children were in a life environment which promoted success at school and hence access to high responsibility jobs.

In parallel, Davis (1992) specifies that a particularly strong downward mobility among children of the Maoist middle class was linked to the Cultural Revolution. Walder (1989) underlines that the mobility was made because of drastic measures (for example, children sent to the countryside).

Underlying that families' ties could be a strong deterrent to social mobility, another goal of the Mao's system was their dissolution. Piketty (2000) specifies that family background and children's cultural environment are two important determinants of intergenerational mobility. Some authors argue that the only possible way to break persistent inequality requires a major conflict between the government and the family (Boudon (1974)). Consequently, a lack of democracy can sometimes favor intergenerational mobility, diminishing future inequalities.[2]

Supporting this idea, Mao Tse Dong assimilated the traditional family into a feudal and barbarous system, justifying the distance introduced between parents and their children. Many new institutions such as youth league organizations were created in order to make the child think by himself. Moreover, different laws allowed him to be more independent from his parents (Lindbeck (1951), Engel (1984)). In 1950, a marriage law established that "the feudal marriage system shall be abolished" (cited in Engel (1984)). Consequently, we see that the government established a system in which the child had more rights and was more independent. In the countryside, the traditional family still played a strong role in the Child's life. Different studies emphasize that these family characteristics were not challenged as they could have undermined the support of the rural population to the party (Walder (1989)).

Consequently, concerning the mobility in wages and education, the Mao Tse Dong

[2] This is in contradiction with the theory presented by Roemer (2002) who argues that democracy can be a better environment than dictatorship to reduce inequalities.

period, both thanks to the restructuration of the society's hierarchy and to the promotion of child independancy, demonstrated high mobility next to the revolution. Then, the established egalitarian system impeded broad movements in terms of wages and education, even if transmission of social status between generations occured more in terms of political and social responsibilities.

2.1.3 What is the new context influencing mobility?

The reforms, launched by the end of the 70's, have had many consequences on the social organization of the Chinese society. On the one hand, a new hierarchical structure was developing due to the marketization of the economy. On the other hand, the family began to regain some of its place in its child's life. Focusing on mobility relative to wages and education, I will present how the reform movement has had an impact on these two dimensions.

Wage mobility

Two elements should be considered when we look at the determinants of job and wages mobility since the reform movement: the parallel evolution of labour both in terms of internal organization (hierarchical structure inside the enterprises, the role of family network etc.), and in terms of wages structure.

Concerning the first point, three factors influence occupational mobility: the challenging of the communist hierarchy, the low level of government approved migrations and the family network ("guanxi").

With the introduction of a competitive economic system whose goal was to increase efficiency, a status re-allocation happened. The traditional communist spirit of equality (which implied some downward mobility for the upper-class) collapsed. As specified in Bian, Breiger, Davis, and Galaskiewicz (2005), the blue collars, seen in the Maoist period as the leading class, began to experience real proletarian life conditions. They have become more and more isolated from the upper classes constituted of managers, white collars and of political elite. This last group has seen their social status re-evaluated. The party members, and particularly the State executives, have benefited from above average compensation packages and have been given privileges such as social promotions and training. A kind of "embourgeoisement" appeared in this class. We can consequently observe mobility in terms of occupation inside enterprises or between enterprises in the same sector.

Nevertheless, we cannot clearly see geographical mobility between occupations. The former drastic system of registration of place of residence, the Hukou, reduced possible migration from the countryside to the city. People could not, as a result, have access to the urban job market which could offer a more diversified occupational structure (Cheng and Dai (1995)). With the easing of the registration system, a higher geographical mobility can be observed and consequently, more occupational mobility can be expected with the development of the reforms. Knight and Yueh (2004) emphasize that a higher job mobility is observed among the migrant population, but they stress that on the one hand, this is not always by choice, and on the other hand, this does not mean that there is wage mobility.

Concerning the role played by the "guanxi" in the intergenerational occupational mobility, Jaeger and Holm (2007) stress that family social capital is particularly and still important in China. Knight and Yueh (2002) underline that this kind of capital has an even more important impact on wage determination than human capital. And this is not only the case in the public sector. The communist party networks have influenced and still infuence access to certain type of occupations and social status (Bian, Shu, and Logan (2001), Bian, Breiger, Davis, and Galaskiewicz (2005)). Nevertheless, the way to have access to this network has evolved with the movement of reforms. The need for more intellectuals have implied that the party desemphasized the role of family social class and made a special effort to attract young intellectuals. Education has consequently played a more and more important role in terms of access to political responsibility. As party membership remains a central factor determining work occupation, particularly for managerial responsibilities (Bian, Shu, and Logan (2001)), monetary and social returns to family investment in their childrens' education are even higher. Consequently, the family social status can nowadays play its role not through the occupational status or network, but through the level of investment in their child's education and as a result, through wealth.

Considering the mobility in terms of wages, the three factors explained above relative to occupational mobility are directly linked to the new way wages are determined. As presented in the Chapter 1, wage is dictated less by central planners but is more a function of human capital and the level of responsibility. Combined with this evolution, mobility in terms of occupation have led to upward wage mobility for the new classes of managers and white collar workers. Given that China left an egalitarian society in terms of wages, this new labour market context has implied intergenerational wage mobility. Even if some parents' characteristics (like party affiliation or social class) have influenced the occupation mobility at the beginning of the reform movement, we cannot say that their wages are a determinant of their children's wage mobility.

Geographic immobility has implied that the individuals living in the countryside were bound to stay in the countryside, not having access to a broader choice of jobs and hence to a larger range of wages. Inequalities in terms of wages, which already existed during the Mao Tse Dong period, between urban and rural areas (Walder (1989)), have increased. The parents' farming characteristic and often the poor income linked to this occupation, has been reproduced due to the low level of mobility allowed until recently. Concerning the rural population, the intergenerational mobility can thus be observed to be lower than the one for the urban population.

Moreover, as education plays an increasing role in the determination of wages both through the higher human capital returns and through access to party membership (which in turns influences the level of wages), investment in childrens' education will become a central determinant of wage mobility, isolating the population which will not be able to cope with educational expenditure. Educational mobility will become a growing factor determining wage mobility.

Educational mobility

Educational mobility is influenced by two major determinants: the place given to the parents involvment in the child's choice, and the legal context of the education system. Concerning the former, this involvment is constrained by their preference (if they want their child to have a good level of education and wages) and by their budget constraint, depending on the necessity to pay for good education. The latter determinant is totally function of the political decisions. Reforms have had consequences for both elements.

Two major policies have influenced the parents' preferences concerning their child's education: the One Child Policy, and the Household Responsibility System. The One Child Policy, launched in 1979, allowing parents to have only one child, implied a new kind of intervention for parents in their childrens' life. As Green (1988) and Short, Fengying, Siyuan, and Mingliang (2001) specify, this has resulted in parents becoming more concerned about the education of their child. The strong presence of family in the child's life could be seen as positive for his development. Nevertheless, the direct transmission of productive ability through the child family environment is one of the determinants slowing down social mobility, as is the increase in financial investment from the parents in the child's education (Piketty (2000)).The stronger involvement of the family in their child's life will consequently lead them to be more likely to pay for good educational services provision, but only the richest will be able to cope with this bugdet constraint and to help the child to be more confident in his abilities (in comparison with his peers).

The Household Responsibility System launched in 1984 allowed rural families to work on their own fields and farms. This can have two opposite effects on the school enrollment of children from these families. If these last are more motivated by the fact that income earned from their own farms will be totally gained by the household, this can lead to an increase in productivity and consequently of agricultural incomes. This revenue effect can be positive for children's school enrollment. However, a substitution effect can also arise from self employment in household farms. Parents can be tempted to employ their children in agricultural activities. These children will subsequently leave school to help their parents on the family farm.

Child labour is one of the determinants in children dropping out of school, even for other reasons. As stressed in Connelly and Zheng (2003), the development of the industrial sector throughout the country implied the hiring of children. This problem has recently been stressed concerning mining activities or even for the production of goods for the 2008 Peking Olympic Games. Child labour clearly has a negative impact on a child's possibility of gaining an education.[3]

The development of child labour emphasizes the need for higher income for the poor population who cannot send their children to school not necessarily because of their preferences but because of budget constraints.

Actually, the movement of reforms has an impact on the educational system and the schooling enrolment with the disengagement of the central government in the financing of the educational services provision (Gustafsson and Shi (2004)). In 1985, the Resolution on Educational System Reform gave total responsibility to local government for primary education financing (Law (2002), Mok (2000), Connelly and Zheng (2003)). Consequently, provinces had to find financial sources to develop the schooling system. On the one hand, new competition between provinces in order to obtain a higher budget from the central government occured. The provinces which have negotiated in the most efficient way have been able to hire more qualified teachers and so the quality of the educational system has been different among provinces. Entrance fee appeared both in the public sector and in the private educational sector, developing to cope with the lack of educational institutions (Mok (2000)). This "marketization" of the education system

[3] This problem is developing in China Bulletin (2007). Moreover, it concerns moreso boys as they are able to produce a higher level of effort. Girls are more likely to be kept at home for house work.

in China lead to inequalities in terms of access to education. Household budget constraints became a central determinant of the participation of children in the schooling process and education expenditures represented a rising percentage of total household expenditure (Gustafsson and Shi (2004)).

The parents now have to invest more in the education of their child. The richer family can give better education both in quality (unequal teaching quality given the geographic areas) and in quantity (unequal availability of schools, colleges and universities given the geographic areas). This is nevertheless reduced by the registration system which prevents massive migration, inducing the rich to invest in local schooling systems. But the young student flows inside the country and between China and foreign countries tend to prove that geographical mobility is possible and increasing.

Concerning the legal context of the educational system. As Holmlund (2006) demonstrates, the increase of the compulsory years of schooling is a positive factor for mobility. In China, as the political leadership has become aware of inequalities in access to education, different measures have been taken to decrease the number of illiterates individuals and to increase the educational attainment of most children. One of them is the Basic Educational Law (1986) establishing nine years of compulsory school (Law (2002)). This law had as a goal to give a minimum of nine years of schooling to all children. Nevertheless, as for important economic reforms, some zones have been priviledged in the application of educational reforms, increasing inequalities between provinces. The political leaders asked the highly developed zones (coastal provinces and some interiors zones) to launch this law in 1990, in 1995 for the semi-developed zones and no target date was imposed to the less developed areas (Hannum (1999)). Due to a lack of monitoring in these last geographic zones, inferior school quality remained.

To sum up, since the movement of reforms, the numerous and sometimes contradictory political measures have had a huge impact on the evolution of intergenerational mobility. Nevertheless, it is not clear by now which force (equalizing or disequalizing) will lead the final results. In the following sections, I shed light on this problematic issue.

2.2 Social mobility since the reform movement: a linear approach

I am interested in two kinds of intergenerational mobility: educational and wage mobility. Moreover, in order to make the analysis clearer, I distinguish two kinds of mobility, which are named, by convention, absolute and conditional mobility[4].

The concept of absolute mobility is linked to the question: are children in the same social category as their parents? If the answer is no, there is, objectively, absolute mobility. It can be the result of parents characteristics but also of environmental political decisions, local policies, economic development etc. I am not looking here at the causal link between the parents' characteristics and the child's one. Instead, I focus on the movements between parents' and child's educational attainment and wages, in absolute terms. As a complement of absolute mobility, conditional mobility answers the question: given their parents' characteristics, do children have the same chance to move to a higher or a lower category? I consequently want to know if the parents level of education or wages is a determinant of their children's social mobility considering a causal link.

2.2.1 Absolute mobility in China

Measurement

To describe the social movements between two generations of individuals, I use mobility matrices. This methodology looks at the proportion of individuals in each educational (wages) attainment category as a function of the category of educational (wages) level of their parents. The hypothesis behind this tool is that no mobility is translated through

[4]The concepts of absolute and conditional mobility used here are different from the concepts of absolute and relative mobility used in sociological literature. Cheng and Dai (1995) use these two last notions studying occupational mobility in China.

high percentages on the matrix's diagonal. On the contrary, if higher percentages are observed above or under the diagonal, this means that children are not in the same category as their parents.

Considering two dimensions of interest, that is education and wages, I will have to define ordered categories as key educational attainment and wages quantile. Moreover, it is important for the mobility matrices relative to wages, to consider wages conditional on the individual's (child or parents) age. This is done to take into consideration differences in experience among parents and children which influence the level of income. This makes the two generations' wages more comparable. Consequently, for absolute mobility, wages are measured by the prediction of the following equation:

$$y_i = \gamma_0 + \gamma_1 Age_i + \gamma_2 (Age_i)^2 + \varepsilon_i$$

where y_{it} is wages, and ε_{it} is the residual.

This kind of tables is commonly used in empirical studies on intergenerational mobility both in economics and sociology (Checchi, Ichino, and Rustichini (1999), Behrman and Taubman (1985), Bourguignon, Ferreira, and Menendez (2001), Cheng and Dai (1995)).

Mobility matrices give a picture of absolute mobility but it is based on an arbitrary cut off between categories. Moreover, less mobility can be seen for the top and bottom categories (what is called floor and ceiling effects), due to the fact that if parents were at the top of the distribution, their child can only move downward and the contrary is true for parents coming from the bottom of the distribution (Corak and Heisz (1999)). This is why the further analysis of conditional mobility will be done in sub sections 2.2.2 and 2.2.3.

Data

As specified in the preceding chapter, the database I am using here is the CHNS database, for the years 1991, 1993, 1997, 2000 and 2004. This survey allows us to study the intergenerational mobility under certain circumstances. A strong selection bias exists as

I only have the parents and child's educational attainment and wages in the event that the child still lives with his parents. This appears restrictive as these children may have personal characteristics which influence their level of education and wages. Particularly, I do not observe children who have moved from their town to study in other provinces or abroad. If I consider that these children come from richer families, I am underestimating the level of mobility in income and education. I will give some elements of understanding in the descriptive statistics in order to characterize individuals who live with their parents.

Even if I will control for this issue in the paragraph relative to conditional mobility, I am not able to take it into account in the descriptive situation of absolute mobility. We consequently have to keep this caveat in mind for the interpretation of absolute mobility results. Moreover, this will be a constraint in terms of the number of observations. This will have different consequences on the quantification of both absolute and conditional mobility and particularly when I want to take into account the temporal dimension. When I study educational mobility, I will be able to distinguish the five time periods and look at the evolution of mobility between 1991 and 2004. But this will not be possible once I study wage mobility. We will see that the samples considered in this case are very small even once I pool the 5 years of observations together. Consequently, I will use this last pooling configuration as soon as I work on wage mobility. Also due to the small number of observations, I cannot use average incomes over the five periods considered. Consequently, I am not measuring a permanent wage. Solon (1992) and Corak and Heisz (1999) have stressed that this constitutes a limit when we study wage mobility. Nevertheless, as these authors have shown, using annual income data can tend to overestimate the degree of mobility. Consequently, we must be cautious in the interpretation of the results, particularly for those relative to conditional mobility analysis.

The variables of interest are given by the child and parents educational attainment on the one hand, and by their wages on the other hand. When both the mother and the father are present in the household, I will use the average of the two levels of education and wages. When I face a single parent household (as there are some in the database, principally with mother only), I will take the single parent educational attainment and wages.[5] Children considered are more than 16 year of age and have finished school. Parents are more than 32 years of age.

As I specified in the first chapter, in the database description, doubts exist relative to the total income data. Agricultural incomes in particular appears questionable. Thus, I focus here on the data which seem the most reliable, specifically wages data. Consequently, as soon as I take into account wages both in education and income mobility, the sample will be restricted to children or parents' who are wage earners. Among wage earners, I am able to consider two different characteristics: the hourly wage and the yearly wage, including bonuses. The former does not take into consideration the quantity of work chosen by or imposed to the individuals. For the latter I am bound to take the annual wages, as bonuses are measured at the annual level.

This last variable will give a larger picture of wages and can be particularly interesting for assessing the impact of the parental wages on the educational status of their child. As the educational expenditures and investment of the parents for the education of their child are usually yearly planned, the annual wages will more closely mirror the total budget available for education expenditures.

Finally, I must specify that parents or individuals who are not included in the sample of wage earners are principally constituted from the agricultural population with small business or handicraft activities remaining marginal.

I will consider five samples.

Sample A includes children for whom I have both their own and their parents' educational attainment. Given that I do not consider any individual's wage there will be

[5] Some studies on intergenerational mobility distinguish the impact of father's and mother's social characteristics on the child's one (Lam (1999), Strauss and Thomas (1995)). The results are not unanimous on the existence of a significant difference between the two. As will be seen in the descriptive statistics, I am bound to the sample size. Consequently, I cannot distinguish the mother's characteristics impact from the father's one.

Samples	Child's Characteristics	Parents' Characteristics	Observations
A	Education	Education	5499
B	Education	Education and yearly wages	1808
C	Education	Education and hourly wages	1720
D	Yearly wages	Education and yearly wages	528
E	Hourly wages	Hourly wages	342

Table 2.1: Description of sample characteristics

included individuals from all social origins and all kinds of jobs. This will give a larger picture of what is happening in terms of educational mobility but the consequence is that very different parents' and child's activities are mixed.

In samples B and C, I focus on the parents' who are respectively hourly and yearly wage earners, as defined above. The first sample is larger than the second as sometimes I have information on the yearly wage but not on the hourly wage (this being calculated from the yearly wage and the number of months worked per year, weeks worked per month, days worked per week and hour worked per day).

Finally, samples D and E contain children for whom I have respectively the yearly wage and the level of yearly wage of their parents, and the hourly wage and the level of hourly wage of their parents.

Descriptive statistics

Before analysing the parents' and chldren's characteristics for sub samples including only individuals who are living with their parents, I would like to make a more detailed description of the characteristics of children who coreside whith their parents considering time evolution. This is done in Table 2.2.

This table concerns the largest sample considered in this study. Individuals come from diverse social origins. As the same conclusions will arise from similar tables done for the samples B and C, these tables can be found in the Appendix. In addition, I cannot perform the same analysis for the samples D and E as the samples' sizes would lead to samples which are too small for each year.

	1991	1993	1997	2000	2004
Average age	18.0924 (1.3735)	19.0496 (1.9123)	20.3414 (2.9330)	21.0829 (3.5873)	22.0551 (4.8062)
% of men	47.65 (49.98)	50.68 (50.02)	56.43 (49.60)	60.75 (48.85)	67.29 (46.94)
Educational attainment	7.7648 (2.6491)	8.1970 (2.5099)	9.0524 (2.4611)	9.3089 (2.6441)	9.9633 (2.4850)
% living in urban areas	26.48 (44.15)	26.03 (43.90)	22.85 (42.00)	30.66 (46.13)	34.87 (47.68)
Observations	808	949	1641	1037	1064

Table 2.2: Characteristics of children living with their parents for the larger sample

I can underline different trends. Among the children living with their parents, the percentage of men increases with time. This can mirror the later marriage of men, or their leaving their parents' house at a later age. This argument is supported by the growing departure age of children as a whole.

The increase in the level of education can explain the later departure of the child from the parents' house. If he/she studies longer studies, he will find a job later and consequently be independant with his own home, later.

No clear trend is emphasized relative to the percentage of individuals living in urban areas even if the last years demonstrate a higher rate of urban children living with their parents. This can simply be due to the higher level of migration observed in China during the last years.

Table 2.4 gives the mean and standard deviation for the principal interest variables for the five samples. At first, it is important to stress the low samples sizes given that the data in this table are pooled for the five years. As specified above, this is and will be a central reason for the exclusion of interesting variables (distinction between parents, parents' occupations etc.).

Mostly due to the restriction of the sample to children still living with their parents, the average age of the sample is particularly young (between 19.6 years old and 21.8 years old) compared to the total population older than 16 years old and to the total potential active population. As Behrman and Taubman (1985) presents, this can have important impact on the results. The usual earning profile shows that income tends to increase

	(1) A / B	(2) A / C	(3) A / D	(4) A / E	(5) B / C	(6) D / E
Child characteristics						
age	6.528	6.254	−8.447	−7.757	−0.131	−0.823
gender	2.827	3.076	2.626	2.665	0.246	0.411
urban sector	−10.430	−10.598	−6.219	−5.608	−0.275	−0.375
years schooling	−14.461	−14.755	−10.666	−10.014	−0.524	−1.052
Parents'characteristics						
age	10.138	10.272	0.261	0.359	0.374	0.145
years schooling	−18.088	−18.341	−8.984	−8.201	−0.479	−0.676

Table 2.3: t-student for tests of equal mean between samples A, B, C, D and E

with time. *A priori*, the level of wage observed will be lower than the future wage the individual will earn. This is confirmed by the average wage of the parents which is, both for the hourly and yearly wage, higher than that of their children. Consequently, what I will observe will be moreso the impact of the parents characteristics on the starting wage in the individual's life than on his/hers average wage for life.

As the observed individuals have finished their education, no bias will be observed when studying educational mobility (other than the one linked to the coresidence selection bias) For all samples, the average educational attainment of the parents is lower than the child's one. This can mirror the effects of the nine years of compulsory school law on the average evolution of school attainment.

Concerning the comparison between the different samples, we note that averages in terms of age (both for parents and child), gender, place of residence (urban, rural) and educational level (both for parents and child) are quite similar between the samples B and C and between samples D and E (see columns 5 and 6 of Table 2.3). This is logical as these two sample groups are very close in terms of the characteristics of individuals they include. However, important differences can be stressed once we compare the sample A with samples B and C and D and E. I recall that sample A includes individuals coming from diverse social origins whereas the other samples include individuals who are living with parents' who are wage earners or who themselves and their parents are wage earners. This distinction is essential and the characteristics of the groups issued from these samples are clearly different. The more diverse sample of children demonstrates a higher percentage of men compared to the two groups of other samples (taking B and C on the one hand and D and E on the other hand). This can be the translation of gender inequality in China and particularly among the agricultural population, as I specified that the complement to the samples of wage earners is principally relative to agricultural population. Moreover, this statement is supported by the percentage of the urban population in sample A which is about 28% whereas it represents 41% in the other samples (the difference being significant, as can be seen from the four first columns of Table 2.3). We note that the level of education of both the children and the parents are lower in the full sample. There is a 2 year difference of education attainment between this samples parents' education level and the two other groups of samples (this difference being significant). As the agricultural population is more present in the former sample, this tends to prove that individuals living in rural areas or/and having an agricultural activity are less educated.

Focusing on wage characteristics, we note that average parent's yearly and hourly wages are higher than those of their children, which may be the result of seniority.[6]

Absolute mobility in China: results

I recall that absolute mobility describes the children's changes of categories of education and wages compared to their parents. The elements of interpretation are relative to the Chinese context but I cannot prove any causal link between these explainations and the movements observed. The establishement of the causal effects of the parents characteristics and the social environment on the child's characteristics will be the object of the paragraphs on conditional mobility.

Results relative to absolute mobility in educational attainment are described in table

[6] The $t - student$ associated with the test of equality of means between the parents' and child's level of wages are of about 5.5949 for the hourly wage and 4.0523 for the yearly wage.

2.5 for the larger sample A and in table 2.6 for the sample B of yearly wages earners parents. The table relative to the hourly wages earners parents (sample C) can be found in the Appendix as the results are very similar to those found for sample B. The cut offs for the categories chosen in education are based on the Chinese educational system. The completion of primary school corresponds to 6 years of education; the lower middle to 9 years of education (which is the minimum number of years of education imposed by the Nine Years of Compulsory School law); the upper middle corresponds to 12 years of education whereas college corresponds to 18 years of education.

We note at first, that the tables' diagonals do not demonstrate low mobility. We see that children whose parents possess only a primary school education experience an upward mobility to the lower middle level whatever the sample considered. I can stress to this point that this phenomenon can be particularly linked to the improvement of the educational system and to the law relative to compulsory schooling. However, the larger sample A demonstrates a higher percentage of children staying in the primary school category. This is linked to the more diversified nature of this sample and particularly to the inclusion of agricultural households. It seems that the environment in rural areas and among the agricultural population do not prone educational mobility for less educated individuals. And this segment of the population appears penalized for other categories of educational attainment. Whilst individuals from sample B experience upward mobility towards upper middle level when their parents have a lower middle level of education, more individuals from sample A tend to remain in the same category of their parents. Moreover, a higher downward mobility affects children whose parents are in the upper middle category when the same children from sample B remain more in the same category as their parents. The more diversified sample which includes individuals from agricultural origins appear penalized in terms of mobility.

For the top educational level category, downward mobility is noted for both samples. This can be due in part to the ceilings and floors effects which I stressed before.

Generally speaking, even if differences are emphasized between the two principal samples studied, mobility exists for education in China. There is not only an upward mobility but also a downward mobility as we can see that for the two highest categories, a high percentage of children are in the category just below that of their parents.

Concerning the mobility matrices for wages, the categories are defined by the quartile of wages for the parents and the child. In order to take into account the evolution of the whole structure of wages (due to wage re evaluation for example) or the fact that wages considered for children are their beginning wages, the quartile are defined separately for parents and children. Absolute mobility will consequently be observed when the child is in a higher or lower quartile of wages than his parents' given the scale of each generation's wages.

Results are presented in tables 2.7 and 2.8, respectively for samples D and E. We note a particulary high percentage of immobility for the two extreme categories. This can be in part due to the ceiling and floor effects enounced before. But considering the whole diagonal, I emphasize a relatively high mobility as the diagonal percentages are between 23 percent and 50 percent at the maximum. Moreover, we note that upward mobility occurs more frequently than downward mobility even if this does exist for the top income quartiles. Concerning the distinction made between the hourly and yearly wages, I do not stress an important difference between the two tables. The percentages for the different quantile considered are very close.

These first results underline a relatively good level of absolute wages mobility which can be due to the higher geographical mobility allowed thanks to the hukou reforms or to the more diversified labour market which gives a broader access to higher paid jobs.

2.2.2 A more sensitive approach: conditional mobility, definition and empirical issues

Theoretical background

I am going to study the transmission of education or wages status from parents to the child using information from two generations.

Let us consider the child's matrice of status characteristics $S_c = \{E_c, W_c\}$ which includes the two vectors Ec and Wc, which represent respectively the educational level of the child and his wage level. In a symetric way, I define $S_p = \{E_p, W_p\}$ as the matrix of the two social characteristics of the parents. I could have used a dynamic model, which would use two time periods as a parallel to the two generations. This would be equivalent to the one chosen here, but the denomination of variables using child's and parent's characteristics appears clearer. Consequently, the impact of the parents' status on the child's one will be studied through the relationship:

$$S_c = f(S_p, C_c, C_p) \tag{2.1}$$

where C_c and C_p represent control characteristics of the child and the parents.

As S_c includes two variables, I will estimate two independent equations for the child's education and wages. Consequently, the two equations which will be estimated are, for individual i:[7]

$$E_{ci} = E_{pi}a + ln(W_{pi})b + C_{ci}c + C_{pi}d + v_i \tag{2.2}$$

where v_i is the residual term.

And,

[7] I underline that no time dimension is considered in these equations because, as explaned earlier, I cannot use a panel structure for the estimations due to data restrictions.

$$ln(W_{ci}) = ln(W_{pi})\alpha + E_{ci}\beta + C_{ci}\gamma + C_{pi}\delta + u_i \tag{2.3}$$

where u_i is the residual term. I do not include in the wage equation the parents' level of education as it will be used to instrument the child's level of education in the estimations.

Using this model, the null hypothesis I am testing is the following:

$$\begin{aligned} H_{o1} &: \text{ no educational conditional mobility if } a \neq 0 \\ H_{o2} &: \text{ parents' wages influences child's level of education if } b \neq 0 \\ H_{o3} &: \text{ no wages conditional wages mobility if } \alpha \neq 0 \end{aligned} \tag{2.4}$$

Actually, if wages or educational level of the parents have an impact on the educational attainment or wages of the child, it is logical to consider that no conditional mobility exists. The size of the coefficients will give the magnitude of the transmission effect. I expect that, if the coefficients are significant, they will be positive. This would translate the fact that the richer or the more educated the parents are, the higher the educational level and income of the child will be.

When I look at equation (2.3), I see that child's educational attainment is included in the explainatory variables. This is done in order to take into account the effect of the child's human capital on his wage. As the parents' level of wages influences both the child's level of wages and the child's level of education, the total effect of parents' wages is given by $(\alpha + b.\beta)$ as it can be seen in the following equation:

$$ln(W_{ci}) = ln(W_{pi}).(\alpha + b.\beta) + C_{ci}.(\gamma + c.\beta) + C_{pi}.(\delta + d.\beta) + u_i \tag{2.5}$$

It is also important to note that in the estimation of equations (2.3), the usual problem of the estimation of the return to education in a wage equation arises. Education is correlated with the error term generating an endogeneity bias. I will have to use instrumentation to take this problem into account. Bourguignon, Ferreira, and Menendez (2001), in their study on intergenerational mobility in Brazil, use the education of the parents as the instrument, excluding it from the explainatory variables. The parents' level of education is an instrument commonly used to instrument an individual's level of education (Levin and Plug (1999), Uusitalo (1999)). I will also choose this strategy.

Sample definition: problems of selection and econometric strategies

I specified in the data description made in the preceding sub-section, that I am facing a problem of selection: I only observe the parents characteristics in the case of co-residence of the child with his parents. Even if I was not able to control for this selection in the description of absolute mobility, I can use econometric strategy to cope with this issue in the conditional case.

Moreover, estimating a wage equation in the case of income mobility, I face another selection bias due to the fact that I observe the wage only for individuals who are employed. This selection is usually linked to individual characteristics which can influence the results of the wage equation.

Consequently, depending on whether I am studying educational or income mobility, I will have to use a single selection process (living or not living with the parents for educational mobility), or a double selection process (living or not living with parents and earning or not earning an income for income mobility).

Concerning the single selection problem, I use the Heckman two stage procedure. The selection probit in the first stage will control for the fact of living or not living with parents. The variables influencing this process are described in the following paragraph. I then introduce the Inverse Mill's Ratio (IMR) $\widehat{\lambda}_{1i}$ in the education equation of the second stage.

Two different strategies will be used to control for the double selection process.

First, I will consider two selection equations: one for the case of living with parents, and one for the case of being employed or not. Then using an extended Heckman procedure, I will take into account this double selection thanks to extended Inverse Mill's Ratio which will be introduced in the wage equation. From here, the question is: are the two selection processes correlated or not?[8]

I encounter two selection equations:

$$I_{1i} = \mathbf{1}[z_{1i}\gamma_{1i} + \epsilon_{1i} > 0] \tag{2.6}$$

$$I_{2i} = \mathbf{1}[z_{2i}\gamma_{2i} + \epsilon_{2i} > 0] \tag{2.7}$$

With I_{1i} being the first selection binary variable equal to 1 if the individual is lving with his parents and 0 if not; I_{2i} being the second selection binary variable equal to 1 if the individual earns a wage and 0 if he does not; z_{1i} and z_{2i} are the exogenous variables explaining respectively the first selection process and the second one; and ϵ_{1i} and ϵ_{2i} are the disturbance terms assumed to be normally distributed with zero mean and variance-covariance matrice Ω of dimension 2.

If ϵ_{1i} and ϵ_{2i} are independent, the two selection equations can be independently estimated. Then, the commonly built IMR are constructed for each selection equation giving $\widehat{\lambda}_{1i} = \frac{\phi(z_{1i}\gamma_{1i})}{\Phi(z_{1i}\gamma_{1i})}$ and $\widehat{\lambda}_{2i} = \frac{\phi(z_{2i}\gamma_{2i})}{\Phi(z_{2i}\gamma_{2i})}$.

The wage equation becomes:

$$ln(W_{ci}) = ln(W_{pi}).\alpha + E_{ci}.\beta + C_{ci}.\gamma + C_{pi}.\delta + \widehat{\lambda}_{1i}.\zeta_1 + \widehat{\lambda}_{2i}.\zeta_2 + u_i \tag{2.8}$$

Nevertheless, if ϵ_{1i} and ϵ_{2i} are correlated, so that $cov(\epsilon_{1i}, \epsilon_{2i}) = \rho$, the wage equation becomes:

[8] See Fishe, Trost, and Lurie (1981) and Corak and Heisz (1999) for a detailed review of the multiple selection procedure.

$$ln(W_{ci}) = ln(W_{pi}).\alpha + E_{ci}.\beta + C_{ci}.\gamma + C_{pi}.\delta + \widehat{\lambda}_{12i}.\vartheta_1 + \widehat{\lambda}_{21i}.\vartheta_2 + u_i \qquad (2.9)$$

$\widehat{\lambda}_{kl}$ has been determined by Henning and Henningsen (2007) and are given by:[9]

$$\widehat{\lambda}_{kl} = \frac{\phi(z_k\gamma_k)\Phi(\frac{z_l\gamma_l - \rho z_k\gamma_k}{\sqrt{1-\rho^2}})}{\Phi(z_k\gamma_k, z_l\gamma_l, \rho)}$$

where $\phi(x,y)$ and $\Phi(x,y)$ are respectively the bivariate cumulative and density functions. These expressions can be evaluated numerically.

Consequently, I will have to check for the existence of correlation between ϵ_{1i} and ϵ_{2i} to determine which technique will be used.[10]

The second strategy to cope with the double selection process consists of doing a common two stages Heckman procedure including the $\widehat{\lambda}_{1i}$ IMR of the first selection equation for coresidence into a tobit model for the second stage equation. Using the tobit for estimating the wage equation comes to do as if individuals who are unemployed earn a nul income, or, more specifically a 1 yuan income making the logarithm nul. The variable is consequently considered as left-censored. The assimilation of unemployed individuals to one monetary unit income earners has been previously used by Couch and Lillard (1998). We also choose to use the tobit regression when we note that unemployed individuals represent an important share of the sample. If it is not the case, the results generated from the first estimation methodology enounced before will be prefered. The results of the tobit estimations will be presented in parallel with the other methods' in order to compare them.

[9] For further information relative to the calculation of the IMR under correlation of residuals, see the Henning and Henningsen (2007) complementary paper available upon request at chenning@agric-econ.uni-kiel.de or ahenningsen@agric-econ.uni-kiel.de.

[10] A likelyhood ratio test of $\rho_{12} = 0$ will allow us to test for this correlation.
Ermisch (1999), studying the determinants of co-residence, underlines the role of unemployment for the fact that the adult child still lives with his parents. Consequently, even if the test rejects the correlation between the two error terms, I will present the results of the second stage equation both with λ_{1i} and λ_{2i} and with λ_{12i} and λ_{21i}.

Variables and other empirical issues

The problem of selections enounced above is the most important issue we encounter estimating conditional mobility.

In this paragraph, I present the set of variables used in the selection and principal equations. Then I stress some limitations linked to the data availability.

For the selection relative to whether a child lives with his parents or not, estimating income mobility of the Canadian population, Corak and Heisz (1999) face the same issue. They control for this selection bias estimating the child's income elasticity to parent's income using a two stage Heckman procedure. They use the Social Insurance Number assigned to individuals who filed an income tax return as an identification variable. They find no significant difference with the results which do not take into account this problem, but the Inverse Mills Ratio is significant. I use the empirical literature developed on the determinants of co-residence. Different studies, done in China, in Great Britain and in the US underline three major dimensions: family size, particularly the number of young children in the household; marital status of the adult child and housing characteristics (price or area) (Francesconi and Nicoletti (2004), Ermisch (1999), Aquilino (1990), Logan, Bian, and Bian (1998), Logan and Bian (1999)). Concerning China, Logan and Bian (1999) and Logan, Bian, and Bian (1998) stress the role played by the parents needs, such as their health status for example. They also emphasize that the marital status can play a marginal role in urban areas if the housing constraint is high. If there is a low housing availability, coresidence can be more frequent due to space constraints.

Given the database I use, I am able to use as determinants the number of young children, the marital status of individuals and the size of the house. I will subsequently consider the number of children under 5 years present in the household as the first identification variable.[11] I will distinguish only between married (the variable is equal to 1) and non married (the variable is equal to 0) for the marital status.[12]

[11] The cut off of 5 years old is made in parallel to the definition of infanto-juvenil population.

[12] I have information on whether the individual is a widow or is divorced. Nevertheless, the variable

It is important to note that all these variables can also influence the probability of a person working or not. Consequently, they will also be introduced in the earner selection equation. Moreover, following Strauss and Thomas (1998) I use as identification variables a set of commodity prices available in the community survey from the CHNS.

I expect that the age of the individual plays a particularly important role in the coresidence selection as it is obvious that the older individual, the less likely it is that he/she lives with his/her parents.

The variables considered in the selection process appear to be good identification variables as they are correlated with the co-residence and with the employment status of the child but should not be correlated with the level of wages.

However, I will not be able to include all the variables introduced in the wage and education equation in the selection equations. Particularly, the variables relative to the parents education, age and income will, obviously, not be included in the selection process as they are only available for children still living with their parents.

Concerning the two econometric models described in the presentation of conditional mobility (2.3, 2.2), our explained variables S_c are the child's wages and educational attainment whereas our variables of interest S_p are the parents wages and educational attainment.

As described in the data presentation above, I also distinguish here the three samples relative to education (even if the individual is not a wage earner), hourly wage and yearly wage (sample A to D).

Concerning our control variables C_c, I will use the urban/rural status of the child (the variable will take the value of 1 for urbans and 0 for rurals) as well as the province where he resides. I do not know the child's province of origin and do not know precisely how long

putting these three characteristics together appears more appropriate in my case.

On the other hand, it is impossible for me to introduce the health status of the parents as I have this information only in cases where the child lives with his parents.

he has lived in this province. This will consequently control for some province effects,such as for example the educational context or policy, but not always for the environment in which the child has grown up and built his personality and choices. I also introduce, in the wage equation, the child's educational level as specified in equation (2.3). This variable appears endogenous once I introduce it in the wage equation as some individual unobserved characteristics are in the residual and can be correlated with it. This is the common bias when one estimate returns to education. As specified earlier, I will use the parents' education variable as an instrument for the child's education. Moreover, following Strauss and Thomas (1998), the identification variables used in the selection equation will be included in the set of instruments.

For the parents' characteristics, C_p, as Grawe (2006) underlines, it will be critical to consider the parent's age in the equation. It is particularly important as it is possible that the parents wage is observed at an underestimated level compared to the highest wage they will receive during their working life. The parents age will consequently be introduced in the circumstances variables. Moreover, even if I cannot control for all the occupations of the parents, due to the small number of observation, I will include one variable which will control for the fact that the mother or the father is involved in a farming activity. This variable will not play the same role given the sample we use. If we consider the sample A, which contains all kinds of parents' activities, this will control for all kinds of farming activities, i.e. thos being self employed in household farms as well as being wage earner in collective farms. While this variable will control only for the latter type of employment once I consider parents' who are wage earners. Interpretations will consequently have to be taken with caution.

Another estimation problem is relative to the gender of the individual. Most of the studies estimating wage equation differentiate between men and women. Nevertheless, as I lose an important number of observations considering only individuals living with their parents, I will not be able to estimate two separate education or wage equations for men and women. I however control for the difference between men and women introducing a gender dummy variable in the estimations.

Some control variables cannot be used as they would strongly and negatively affect the sample size. It is the case for the variable describing if the individual is or is not a party leader, or for the individual ethnicity.

Finally, I have to specify the way I observe the influence of the reform process on my results. Two kinds of data allow me to take time into consideration. The first one is a huge transversal database with information on individuals and their parents for every generation considered in the sample. I cannot consider this kind of structure here as the age range of children for whom we have parents' characteristics is too small.

The second, and the one I can use given my database, is to consider a longitudinal survey in which I have the same information on the parents' education and wages for different years. As I clarified above, I am able to look at the evolution of intergenerational mobility during the time only for educational mobility, differenciating the five points of time observation (1991, 1993, 1997, 2000 and 2004). Figure 2-1 gives a picture of the date of birth and the periods during which individuals were at school for every individual considered in each survey.

As we see, some of them have been more or less affected by the reform process for education, as for the One Child Policy. I consider in this scheme individuals aged between 16 and 30 years of age as we have seen in the descriptive statistics which only include individuals who still live with their parents a decrease in average age of observed individuals. I consider that a child's education begins at five years of age.

The time evolution and the relatively high proportion of individuals considered in each survey will be the only indications which can give information on the sources that influence intergenerational mobility.

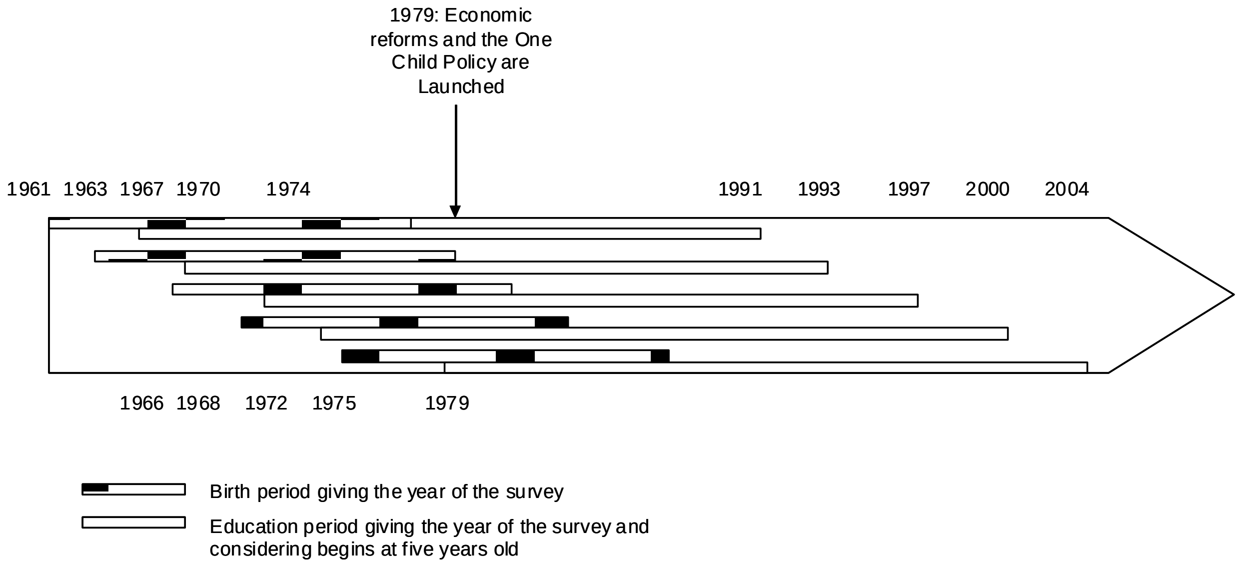


Figure 2-1: Chronology of the most important reforms that have an impact on educational level and incomes.

2.2.3 The causal link between parents' and child's social status

Educational mobility

The results for educational mobility are given in the tables 2.9, 2.10 and 2.11. Concerning the probit regressions results, some precisions must be added. Firstly, being more than 34.8 year olds fully determines the nonresidence status of individuals with their parents.[13] Consequently, I choose to restrict my attention to individuals aged less than 34.8 years old for whom there are individuals both living and not living with their parents. This choice will obviously be done also in the case of double selection (living or not living with parents and being a wage earner or not) for income mobility. Secondly, in the case of educational mobility, I do not include the commodity prices in the selection equation as they highly decrease the number of observations and moreover, they are not

[13] It is interesting to see that there is no adult older than 34.8 years old living with its parents. This can be a clue proving that the tradition relative to the coresidence of children with their parents is not so restrictive anymore.

my principal identification variables relative to the coresidence selection. Finally, the 2004 CHNS survey dataset often produces some questionable results. In the context of the coresidence selection, the number of households who have at least one young child is zero once we consider all the variables included in the regression. Consequently, this variable cannot be introduced in the selection process for the single year 2004.

Table 2.9 presents the selection process for the whole period and for each year and regarding sample A. We see that age always plays a significant and highly negative role in the case of children residing with parents as does the urban area of residence (except for the year 1993). This last factor may be relative to the higher freedom individuals have regarding the choice of leaving the parents' home. Children who migrate from countryside to urban areas can also explain this phenomenon.

Being a man tends to increase the coresidence probability. Moreover, the evolution of this determinant with time shows its increasing role. This is in accordance with the results stressed in the descriptive statistics. We previously mentioned the higher and increasing share of men still living with their parents. As specified before, this can mirror the later marriage age of men, or a later departure of men from their family home (possibly due to longer studies for example). As the correlation between being married and living with parents is negative and significant, this can support the argument of men marrying at a later age in China.[14]

As just enounced, the marital status identification variable has an important well its role and it is highly significant. Nevertheless, the same cannot be said for the number of young children in the household variable which is significant only for the year 1993.

When we look at the marginal effects, we see that being married is the major determinant for leaving the parents' household, the next being living in an urban area and finally increasing age. The positive impact of being a man is nearly as high as the negative one

[14] This effect can also be reinforced with time due to the consequences of the One Child Policy on the gender balance in China. I recall that this policy have had as a direct consequence an increase of children murdering, especially for girls. Men are consequently privilege, and maybe kept longer in the families.

of being married. These two opposite effects seem to be the principal determinants of living or not living with parents in my specification.

I am now going to analyse the intergenerational conditional mobility, both controlling and not controlling for selection, in order to see the impact of taking into account the selectivity process on the level of mobility. As I specified in the description of the model, I consider that there is no conditional mobility where the coefficients on parental characteristics are significant (hypothesis 2.4). Nevertheless, I need to mitigate this fact by considering the empirical literature on intergenerational mobility. At first, Corak and Heisz (1999) find a coefficient relative to intergenerational mobility of income in Canada of about 0.2 and significant. Nevertheless, they consider that this effect is low and that a relatively good income mobility exists among the Canadian population. Holmlund (2006), for Sweden, finds a child income elasticity to parents income of about 0.16 and an increase of 0.4 years of schooling for a parents' year of schooling. For Brazil, Lam (1999) emphasizes a 0.15 to 0.42 year increase of child education for one more year of parents' education and are of 0.22 for South Africa. For the US, Behrman and Taubman (1985) find a child educational response of 0.17 to 0.20 to one more year of parents' education.

Given the preceding results, even if we can say that as soon as parents' characteristics have an influence on child's one, the strict conditional mobility does not exist, it is important to consider that most of the time, this situation occurs. Consequently, I will consider that a coefficient of about 0.2 for the impact of parents' income and education on that of their child corresponds to a relatively good level of intergenerational mobility.

Table 2.10 presents the link between parents' educational attainment and the child's one without including the parents' wages (this corresponds to the sample A). We see that the range of the coefficient on parental education is about 0.25 meaning that an increase of 1 year of parents' schooling attainment raises the child's educational level by 0.25 years. This coefficient is significant at the 1% level of confidence. Nevertheless, we must remember the warning mentioned above concerning the range of the coefficient. Even if the impact is significant, it is not so strong relative to other countries. The educational mobility observed in the description of absolute mobility seems to be stronger considering a more causal link between parents' and child's characteristics. Moreover, we see that the coefficient on the parents' education is stable through time.

On the other hand, we must go back to the sample definition used in this case. We remember that it is the larger sample which considers parents' who are from diverse origins. In particular, it includes parents who come from agricultural origins. I emphasized, using the absolute mobility concept, that a lower absolute mobility exists in this sample compared to the sample which includes wage earners parents. Is this lower absolute mobility supported by the conditional mobility which wants to establish a causal link between parents' and child's characteristics? The lower coefficient on the parents' education once we introduce the parents' income tends to point in the same direction as the results noted by the absolute mobility (Tables 2.11, 2.12 and 2.13). Whereas the impact of parents' education on the child's is about 0.24, the average becomes 0.17 when I introduce wages. This can be due to the differences in the sample characteristics underlined before, stressing the negative impact of the agricultural origin of the parents.

Going back to Table 2.11, I make the comparison between non controlling for selection regressions (columns (1) and (3)) and controlling for it (columns (2) and (4)) both for the yearly and hourly parents' wages (this corresponds to the samples defined before as B and C). At first, concerning the impact of the parents' educational level, we note that it is significant and positive.

Concerning the role played by the parents' wages in the determination of the child's level of education, there is a significant and positive impact. As expected, the parents' yearly wage impact on the child's educational attainment is higher than the impact found with the hourly income (1% more of parents' yearly wage increases the child's level of education by about 0.3812 years of schooling whereas the impact amounts to 0.2449 years of schooling for the hourly wage). This tends to corroborate the hypothesis that the yearly wage has more influence on the level of parents' investment in their child's education. Even if the IMR is significant in both cases, the impact of the two wage variables remain stable once we take selection into account. Consequently, it appears that parental wages influence the child's level of education. Looking at Tables 2.12 and 2.13, interesting phenomena are revealed. Concerning the hourly wage (table 2.13), its impact is not significant before the year 2004. But we see that this impact increases with time and becomes significant with an amplitude of 0.5158 in 2004. But it is more striking to consider the yearly wage (table 2.12) where the coefficient increases with time, it being significant from the year 1993 and reaching a maximum in 2004 with a marginal impact of 0.5640. The effects of the parents' level of education still exists but represents 26.42% to 41.55% of the wage effects. It becomes consequently marginal in comparison to the wages effects.

The principal conclusion drawn from the preceding remarks, is that the child's education tends to be increasingly linked to the parents' wages. This is in accordance with the evolution of the financing system presented in the first section of this chapter: inequalities of access and higher investment by parents' in their children education, allowed thanks to the development of user payment for education provisions, seem to have strong consequences on intergenerational mobility.

Another parental characteristic which can influence the child's educational attainment is their economic activity. As specified in the variables description, I use the farming activities of mother and father as a proxy for this issue. The results are presented in Table 2.14. This table includes in the first column, the results of educational mobility relative to the sample A. Columns (2) and (4) give again the results for educational mobility including wages but not the parents' occupation. Columns (3) and (5) include both occupations and wages. We are hence able to compare the various effects of the introduction of parents' occupation on the child's schooling attainment as well as on the other variables' impacts (due to correlations between variables). Concerning parents' occupation, we note in any case that only the father's farming activity has an impact on the educational attainment of the child. This impact is negative, demonstrating that when the father works as a farmer, his children's level of education is lower. But this variable's impact's range is different given the sample considered and given the manner it is linked with the different wages presented here. When we look at column (3), which includes the yearly wages, we see that this variable is not significant anymore and that the impact of the father's occupation on his child's educational attainment is lower and not significant anymore. I recall that the difference between yearly wage and hourly wage is that the former includes the quantity of hours worked during the year, as well as yearly bonuses. Moreover, as I underlined in the variables description, the farming occupation is relative to employees on collective farms as soon as I consider only parents who are wage earners. The loss of significance of yearly wages can consequently come from the correlation between quantity of work and the bonuses earned from this farming activity which can lead to two major interpretations: employees on collective farms work more hours than other employees or bonuses are principally distributed in the farming activities.[15] Nevertheless, these are intuitions based on the Chinese context which cannot be totally justified. They constitute only a suggestion and need to be further analysed with larger datasets. Moreover, another interesting phenomenon is revealed once the hourly wages are introduced (column 5). The impact of the father's occupation becomes more important than when wages are not included and the hourly wage remains significant, his impact being even stronger than when occupation is not included in the specification.

Two other facts have to be emphasized once the occupational variables are introduced. At first, the variable controling for gender differences is not significant anymore. Actually, without introducing the parents' occupation, the gender variable is significant with the

[15] The way bonuses are distributed in China is a large topic. Controversies on the question of whether they are directly linked to the level of productivity or not are numerous. If it is the case, the productivity is more clearly measurable for farming activities (Foster and Rosenzweig (1994)). This can explain the correlation between my yearly wage variable and the farming activity as well as the loss of significance once this last variable is introduced. No definitive conclusion can be drawn considering this issue.

impact of being a man on the child's educational attainment being negative, particularly since 1997 (as can be seen from Tables 2.10, 2.13 and 2.12). Once the parents' occupation enters in the specification, this variable becomes non-significant, whatever the sample. In the case of the larger sample, which includes parents' who are self employed, this can be translate as the consequences of the Household Responsibility System mentioned in the historical background section. I recall that with this system, the incomes from farming activity are more often kept by the families. And as suggested above, this can have an impact on children working particularly for males. For the samples restricted to wage-earner parents, this loss of significance can also mirror a higher participation of men in farming activities even as employees on collective farms.

The second element which must be stressed is the less important positive impact of living in urban areas once the parents' farming occupation is introduced. From Tables 2.10, 2.13 and 2.12, we can see the high and positive role played by living in urban zones on the child's educational attainment. Moreover, this impact increased with time showing that urban dwellers are privileged regarding the schooling system. This is in accordance with the inequalities development in the education services provision. The rise of the coefficient on the dummy urban tends to underline the negative evolution of this phenomenon. Including the parents' farming occupation decreases the range of this impact but it remains significant. Due to sample restrictions, we cannot look at the evolution of this effect with time when the parents' occupation is controlled for.

Wage mobility

Studying wage mobility, I distinguish results relative to the yearly wage and the hourly wage and also compare them.

Concerning the selection process for hourly and yearly wages (tables 2.17 and 2.15), we clearly see from the coefficients estimated and thanks to the ratio test of residuals correlation, that there is no difference between the results using the biprobit estimations and the two separate probits. Consequently, we can either use the usual IMR and the IMR built from the bi-selection process in the wage equation. Concerning the selection relative to coresidence, I find the same conclusions as found in the preceding paragraph where I made only one selection process. Nevertheless, the specifications here are not exactly the same as before as I introduce this time for both selection equations the commodity prices variables, the sector of activity (public or not) and the individual schooling attainment. As specified earlier, the first variables were not included in the educational mobility study so as not to lose too many observations but are included here as they constitute our identification variables for the other selection process (being a wage earner or not). The second and third variables are added here in order to avoid forbidden regressions as they are included in the wage equation.[16] I find again that age, living in an urban area and being married have a negative impact. Being a man remains a positive determinant of coresidence. The only important difference with the former selection specification for the coresidence resides in the impact of the number of young children in the household, this one being significantly negative. This can be interpreted as a space constraint for homes which do not allow big families to share the same house.

The selection process relative to being a wage earner gives slightly different results for hourly and yearly wages. Living in urban areas, working in the public sector and having a higher level of education constitute positive determinants for those employed as wage earners.[17]

The results relative to wage mobility are given in tables 2.18 and 2.16. In both cases, I distinguish four specifications: 2SLS without controlling for the selection process (columns (1)), 2SLS controlling for the selection thanks to two probit IMRs (columns (2)) and the biprobit IMRs (columns (3)) and tobit. In this last situation, I must underline that for the yearly income, the proportion of censored data (with a logarithm of wages

[16] The individual schooling attainment was not included in the educational mobility selection process as this variable was, for children, the explained variable.

[17] The differences observed between the selection process of the two different wages considered are essentially relative to the number of children which influences the probability of earning an hourly wage but does not influence the probability of earning a yearly wage. Given that the difference in the selection of the two samples are principally determined by observation availability, this can appear surprising.

equal to 0) is 64%. This is very small and this probably explains most of the doubtful results I find using this method for yearly wage mobility. I will consequently focus on the other methods' results in this case.

Results for the two kinds of wages are quite close emphasizing a positive and significant role of the age, synonymous of a seniority evolution of wages which persists. Nevertheless, this is less true for the yearly income including bonuses, possibly translating again as the distribution of bonuses regarding productivity instead of only age considerations. Living in an urban area is also positively correlated with wages. This can be linked to the differences in the kind of activities developed in urban and rural areas. Working in the public sector does not seem to influence the wage level. This can be explained by the fact that we pool the observations for the whole 1991-2004 period, the differentials in wages between sectors can consequently be cancelled. The only exception occurs for the tobit specification for the yearly income. Nevertheless, the range of this coefficient as well as the sign of the coefficient on the gender variable tend to prove the less appropriate use of this specification as mentionned earlier.

The most important variable for my purpose is the parents' income impact on the child's one. This effect is significant and around 0.28 meaning that a 1% increase in parents' wage raises the child's wage by about 0.28%. This tends to support a relatively high level of mobility in terms of wages. The impact of the parents' yearly wage seems a little higher than that of the hourly wage. This can be surprising given that bonuses should on the contrary be more a function of individual productivity. My interpretations will be improved in the future with further research carried out on the way bonuses are distributed in China. Taking a closer look at the total impact of parents' wages on the child's (impact which includes both the direct impact and the indirect impact through the child's education), I note that parents' wages still influence significantly the child's wages. The marginal effects are slightly lower given the negative coefficient for the child's education.

Looking at the tobit results, we see that the estimations of the mobility coefficient is higher each time. Given the low level of censure in the case of yearly wage data, no clear conclusion can be drawn. The results for hourly income show a coefficient of about 0.38 which is much larger than those obtained using the linear 2SLS method. The way I deal with the double selection influences the level of estimated conditional mobility. I cannot definitively conclude on which method is the best given the constraints I have with my database.

Finally, columns (4) of the tables include the father's farming occupation. It has a negative impact on the hourly wage of the child but no impact on the yearly wage. This phenomenon can mirror the negative impact of having agricultural origins on hourly wages through the negative effect of this origin on schooling attainement (as highlighted in the preceding section). But when yearly wages are considered, individuals with agricultural origin can provide a higher number of hours worked compared to the urban dwellers. Doing so, they compensate for the low wages they receive.

Having analysed intergenerational mobility in China using linear methods, three major conclusions can be drawn. Firstly, the impact of the parents educational level is not large compared to other countries but the role played by the parents wages (particularly the yearly one) is strong and increases with time. This can have a very negative implication as the richer families will be more able to pay for the education of their children who will receive, due to the rising returns to education, a higher wages which will be used for their children etc.. This vicious cercle can be a future determinant increasing income inequalities if nothing is done to increase equality of opportunity in terms of access to schooling services provision.

Secondly, our results have also underlined the gap between urban and rural areas, stressing again the possible increase in inequalities of access to education given the parents level of wage but also place of residence.

Finally, I have underlined a quite good level of wages mobility in China compared to other developed or developing countries. This situation can however change in the future given the conclusions relative to educational mobility.

2.3 Non linearities in the relationship between parent's and child social characteristics

In the preceding section, I have estimated intergenerational mobility using linear specifications given by pooling regressions. However, as pointed out by various authors, there is no reason to think that the impact of parents' characteristics on the child's is linear (Corak and Heisz (1999), Ermisch and Francesconi (2004), Couch and Lillard (2004)).

In China, given the economic and social context described above, we also suspect the existance of some non linearities in intergenerational mobility. That is what this section is going to analyse.

2.3.1 Why do I suspect non-linearities? Empirical research and Chinese socio-economic context

Grawe (2004), examining intergenerational mobility in different countries, pays attention to the reasons underpinning non linearities in the transmission of social status between parents' and children. This author uses Roemer (1998) argument to stress two phenomena: first, it is reasonable to expect that higher average earnings will be observed for children from high-earning families, and second, there is however the possibilities that some children may move to other social categories. The former argument is based on the idea that correlations exist between generations relative to preferences for education and effort. Nevertheless, the latter emphasizes that there may be deviations from this general rule. Grawe (2004) coined the term "the exceptional Child" for children who moved from their social category despite their family context. Given this description of the potential movement across social categories, it appears more interesting to study intergenerational mobility considering the different social classes independently. Estimating the average effect of parents' characteristics on the child's is consequently restrictive.

Supporting this idea, different studies on intergenerational income mobility looked at non-linearities of intergenerational mobility (see Solon (1999) for a survey). Most of these studies underline a higher mobility for higher social categories than for lower ones. This means that the social characteristics of parents are more likely to be transmitted to their children for poorer (in various social dimensions) parents than for richer ones (richer parents being more likely to move downwards and the middle classes having the possibility of moving upwards). More recently, using quantile regressions, Ermisch and Francesconi (2004) for Great Britain emphasize the same stronger impact of the bottom social categories on their children characteristics whereas Couch and Lillard (2004) find that both for Germany and the United States, earnings of males appear to be related to those of their fathers when the fathers are relatively high earners.

Using non parametric estimations, Corak and Heisz (1999) demonstrate an inverted U shape for the relationship linking parents' and child's income. The explainations of this phenomenon are linked, for these authors, to the factors underpinning the relatively high mobility depending on parents' income quantiles. Corak and Heisz (1999) quote two principal determinants: the role of budget constraints (enounced by Becker and Tomes (1986)) which would appear to be a heavier weight borne by the poorest, leading to less mobility for the highly constrained families; and the neighbourhood effects (enounced by Durlauf (1996)) which improve the quality and the quantity of the education depending on different places' socio-economic environment. With these two determinants in mind, Corak and Heisz (1999) support the inverted U shape saying that for the poorer populations, parental income is rather low but the child's ability is increasing so that the budget constraint becomes binding for an increasing fraction of the population. On the contrary, for the richer populations, income mobility is increasing given that parents are sufficiently rich to invest in their children education, the budget contraint is relaxed even if the child's ability continue to rise.

For China, as far as I am aware, no studies thus far have examined non linearities in the relationship between parents' and child's characteristics. However, the two factors described above which justify potential non linearities in this relationship are totally applicable to the Chinese context. Even if we cannot clearly anticipate the form of non linearities in mobility for this country, finding an inverted U shape would not be surprising. As presented in the preceding section relative to linear mobility in China, budget constraint becomes an increasingly important determinant of education and in turn income mobility in China. On the other hand, I have also explained that geographical inequalities exist in terms of educational services provision. Neighbourhood effects are also revealed in China. The argument presented by Corak and Heisz (1999) are consequently relevant to the Chinese context.

I will analyse non linearities which can exist in the educational as well as wages mobility in China using two different methodologies which are presented in the following section.

2.3.2 Econometric strategy

Two empirical methodologies will be used to take non linearities into consideration. The first is given by quantile regression models. When the usual linear regression models gives an estimation of the mean conditional on the values of explainatory variables, quantile regressions allow non linearities between quantiles of the effect of explainatory variables on the variables of interest. Using the notation of Buchinsky (1995), quantile regression models are described as:

$$\begin{aligned} y_i &= x_i'\beta_\theta + u_{\theta i} \\ with\ \text{Quant}_\theta(y_i|x_i) &= x_i'\beta_\theta \end{aligned}$$

where y_i is the explained variable for individual i, x_i the explainatory variables and $u_{\theta i}$ the error term. The term $\text{Quant}_\theta(y_i|x_i)$ gives the θth conditional quantile of y given x.

Considering quantiles relative to the child's characteristics (education and wages), I will show the results relative to the first 25 percent, at the median, and for the 75 quantile of educational attainment and wages. The set of explanatory variables will be the same as those described for the linear specification. The variables of interest will be the parents' education and wages for the estimations of educational mobility and parents' wages for wage mobility. As before, I will distinguish results relative to the yearly and the hourly wage, given that they were very similar when considering a linear pattern.

Next to this methodology, I also implement semi-parametric estimations of the coefficients relative to the parents' characteristics using a partially linear model which can be written, using the same notation as in the section 2.2.2, as:

$$E_{ci} = f(E_{pi}) + ln(W_{pi})b + C_{ci}c + C_{pi}d + v_i \tag{2.10}$$

$$E_{ci} = E_{pi}a + g(\ln(W_{pi})) + C_{ci}c + C_{pi}d + v_i \tag{2.11}$$

$$ln(W_{ci}) = h(\ln(W_{pi})) + E_{ci}\beta + C_{ci}\gamma + C_{pi}\delta + u_i \tag{2.12}$$

Equation 2.10 considers a linear relationship between the child's education and variables other than parents' education. Parents' education enters in the equation following a non linear function f which will be estimated semi-parametrically. This translates exactly as non linear effects in intergenerational mobility in education.

As I noted analysing linear mobility, parents' income plays an increasing role in the determination of the child's schooling attainment. Therefore equation 2.11 looks at non linearities which can exist between parents' wages (yearly and hourly) and the child's level of education. This can mirror the role of the budget constraint described as a potential factor explaining non linearities.

Finally, considering wage mobility, equation 2.12 presents the non linear relationship existing between parents' and child's wages, both for yearly and hourly wages.

Concerning the time evolution of the influence of parents' characteristics on the child's, which is taken into account in the preceding section, I will examine this phenomenon more closely considering non linearities. In order not to overoverburden the analysis, I will at first consider the quantile regressions pooling together the five periods of observations. Quantile regressions will give a global picture of non linearities and provide some intuitions guiding the semi-parametric study. It is important however to stress that quantile regressions are based on quantiles in the child's characteristics whereas semi-parametric studies non linearities in the relationship between different levels of parents characteristics with different levels of the child's. Semi-parametric will consequently give graphical results both for the pooled periods but also distinguishing between the five time periods when this will be possible, such as for educational mobility only.

Given the small samples which I obtain when I divide my full sample between the five sub-periods, I will have to pay attention to the smoothing parameter which depends highly on the sample size. A lower value of this parameter will be used, decreasing the level of precision of non linear estimations. Once again, larger samples would render the analysis easier.

Moreover, I will not take into consideration the role played by parents' farming occupation in this section. This variable highly decreases the number of observations in our samples. Moreover, we cannot look at non linearities concerning parents' occupation as I only use the binary variable which tells us whether parents are involved in farming activities or not.

I will take into consideration the selection process described when studying linear specifications including the IMRs for co-residence in the case of educational mobility and co-residence and being a wage earner for wage mobility both in the quantile and in the semi-parametric estimations. Nevertheless, I will not control for endogeneity of the child's educational attainment in the wage equation when considering the quantile regressions analysis. This will be done in the semi-parametric specifications, using the same instruments as specified in the linear case.

2.3.3 Non linearities in intergenerational mobility in China

Quantile regressions estimates

Table 2.19 gives the results of quantile regressions not including the parents' wages. As will be the case for the following tables, the first column contains results obtained from the linear estimations. The impact of the parents' educational attainment variable seems to have a U shape evolution with the child's quantile. It seems that parents' education plays an important role for the two extreme tails of the distribution. And the impact is particularly high for the lower extreme. This shows the importance of parental educational status for individuals coming from poorer backgrounds and maybe less educated social classes. Given that no study, to my knowledge, has examined closely non linearities in educational mobility, no comparison can be drawn. However, this result seems intuitive regarding the Roemer (1998) argument: a higher correlation can appear logical for the two extremes of the distribution considering differences in preferences for schooling or the family's environment.

Another phenomenon which is interesting to note is the role played by the urban living environment which is particularly high for the upper-tail of the distribution. This underlines inequalities in access to higher levels of schooling (particularly university) between rural and urban areas.

Quantile regressions regarding educational mobility considering the parents' level of wages are given in Tables 2.20 and 2.21. I recall that in the present situation, I am looking at individuals whose parents are wage earners. Remember that I differentiated between three samples when considering educational mobility: sample A gave a larger picture of the parents origins, considering parents who are wage earners as well as those self employed in household farming for example; samples B and C included only children coming from families in which at least one parents is a wage earner. I stressed, when making these divisions of the sample, deep differences that exist in terms of mobility between the samples.

Two interesting observations must be made regarding the variables of interest. Firstly, the impact of the parents' education is now regularly increasing, reaching the highest level for the upper tail of the child's educational distribution. This can be explained both by the differences in the individuals included in sample A on the one hand, and in samples B and C on the other, and by the fact that the parents' educational attainment could have translated as differences in parents' incomes when these incomes were not included. A simple way to find out which argument is the most valid is to consider educational mobility without including parents' wages but only for the children whose parents are wage earners (samples B and C). In doing so, I find a regular increase in the elasticity of children's educational attainment to the parents' one, favouring the former argument (results of these estimations can be found in the Appendix). Consequently, it is interesting to note that the highly significant role played by the parents' education for the lower tail of the distribution is closely linked to the definition of the sample. The sample including children from agricultural origins demonstrates a lower mobility for the lower tail of the distribution.

Secondly, concerning the role played by parents' wages on the child's educational attainment, we note an inverted U shape with higher elasticities for the yearly income. This inverted U shape implies a higher impact of the parents income for a child attaining a middle schooling education. This can be interpreted as being close to the conclusions of Corak and Heisz (1999) relative to the more binding budget constraints when the ability of children is increasing. However, this argument, presented above, was obtained from non parametric estimations which allow one to say that the increasing weight of the budget constraint with the increase in the ability of children concerns poorer families. We cannot affirm this claim here as quantile regressions only give information relative to quantile in the child's schooling attainment and not relative to the parents' wage quantiles. The semi-parametric estimations will allow me to consider these non linearities. For now, the hypothesis which I make considering the parallel shown between the increasing binding budget constraint and the inverted U shape obtained here, is that children with the lower levels of educational attainment come from from poorer families, which is supported by the positive and significant link between the child's educational attainment and the parents' wages.

Regarding the hourly wage, parents' hourly wage is non significant for the upper quartile. This is due mainly to two phenomena: the first is that as argued before, the yearly wage can be related more to the educational expenditures which are more likely to be yearly planned. Secondly, this can also show the lower binding budget constraint for the richest families (once again, with the hypothesis that better educated children mostly come from the richest families).

Consequently, studying educational mobility using quantile regressions, non linearities are observed in two different manners: for the larger sample including children who come from different socio-economic origins, the level of education of the parents strongly influences the child's schooling attainment for the lower quantile. This is not the case anymore when restricting our attention to children who come from families with at least one wage earner. Considering these children, the relationship between the child's educational attainment and parents' wages follows an inverted U shape, which shows the higher impact of parents wages for middle educated children.

Focusing now on wage mobility, we note important differences between the results including yearly wages and hourly wages for both parents and children. For the former (Table 2.22), we note a decrease in the impact of the parents' income on the child's with the increase in the child's wages quantiles. Particularly, a strong drop occurs between the median and the upper quantile groups. This tends to show the higher and strong impact of the parents' yearly wage on the child's for poorer groups.

But another phenomenon is emphasized once I consider hourly wages (Table 2.21). The relationship between children's and parents' hourly wage follows a U shape, demonstrating the stronger impact of the parents' wages for the two extremes of the distribution. I recall that the only difference between the hourly and the yearly wages resides in the quantity of hours worked per week, weeks worked per month and months worked per year. Consequently, once I purge for the effect of the quantity of work done, a lower mobility is observed for the two extremes of the distribution which can be explained on the one hand by a higher transmission of wages status for the poorer, correlated with the relationship related to yearly wage, and on the other hand by a higher transmission of wages status for the richer who can offer higher levels of education to their children who in turn will receive a higher hourly wage given the increasing role played by human capital in the determination of wages in China.

Consequently, looking at non linearities in social mobility in China brings about some very interesting conclusions both in terms of educational mobility and in terms of wages mobility.

Quantile regressions only give a preliminary intuition for these non linearities. Further analysis is done using semi-parametric methods.

Semi-parametric results

Concerning educational mobility, let us firstly take a closer look at the semi-parametric estimations for the whole sample (sample A, including children who come from different social origins).

In Figure 2-2, we can see the elasticity of the child's educational attainment relative to the parents'. The horizontal line represents the marginal effect obtained from the linear specifications, standard errors are given by the dot lines on both side of the line which represents the semi-parametric estimation of the elasticity.

Even if the coefficient obtained from the linear specification is always included in the confidence interval, we can clearly see using the semi-parametric estimation a non linear pattern. This demonstrates the higher impact of parents' education on the child's for the lower extreme of the distribution. Consequently, children with less educated parents will more often be less educated compared to the average: there is less mobility. Consequently, it seems that the Chinese context does not promote "exceptional" children, as defined earlier.

To further the analysis, I now consider the evolution of the elasticities estimated for the five periods of time (1991, 1993, 1997, 2000 and 2004).

We see from Figure 2-3 that the higher impact of the parents' education on the child's for the lower bound of the distribution is observed mainly before 1997. After this year, even if a small non linearity for this category can be noted for 2000, this is combined with large confidence intervals and morever, from 1997, the semi-parametric results are very similar to the coefficents obtained by the linear specifications. The higher impact obtained for 1991 for the lower categories decreases from 1997 because of the laws which stipulated levels of compulsory schooling. All the interpretations given for the linear case are still valid once we consider non linearities.

When I focus on the sub-samples of children whose parents are wage earners (samples B and C) I am going to consider both the educational mobility and the impact of the parents wages on the child's educational attainment. The former relationship can be seen in Figure 2-4. We see an increasing link between the parents' educational attainment and the child's one once we consider the sub-samples of children from families with at least one wage earner. We cannot overstress the relationship for the extremes as we note large confidence intervals but we clearly see that the coefficient obtained from the linear specification under-estimates the transmission of the educational status for the upper educational categories (even if the coefficient is always included in the confidence intervals).

Once again, I take a closer look at the time evolution of these marginal effects. At first, I specify that results for the two samples are similar in terms of educational mobility, as can be seen for the whole period in Figure 2-4. Consequently, I only include results relative to the evolution for the sample B. They can be found in the Figure 2-5.

We see that the increasing relationship has a rising amplitude until 2000 whereas a decreasing pattern appears for the year 2004. Consequently, even if I emphasize a lower mobility in education for children who come from higher educated families including at least one wage earner, no huge differences are stressed between the linear results and the non linear ones.

Let us now consider the impact of parents' wages on the child's educational attainment. Semi-parametric estimations results are found in Figure 2-6.

When considering semi-parametric estimations of the non linear impact of the parents' income on the child's educational attainment, it is clear that the linear estimations are strongly and significantly upward biased. The coefficients estimated in the linear case are not on the graph as they are about 0.36 for sample B and 0.23 for sample C (see column 1 of Tables 2.20 and 2.21). Even if the marginal effects remain significantly different from zero, their ranges are between 0 and 0.17. Moreover, the relationship between the parents' wages and the child schooling attainment becomes regularly decreasing, underlining the stronger role of the educational financing constraint for poorer populations. Given the relationship between parents' wages and child's schooling attainment and contrary to the inverted U shape emphasized by Corak and Heisz (1999), it appears that the need for higher parental income is important for the lower tail of the child's education distribution and decreases with the increase in his schooling attainment. I however stress the same pattern as Corak and Heisz (1999) for the upper tail of the distribution. The budget constraint is less binding for the parents' who earn higher wages. Going further in the analysis, I once again add the evolution with time of the derivatives. Here again, I only include the results for the yearly wages as the hourly wages offer the same conclusions. Considering time evolution (Figure 2-7), we note no clear differences between the years of observations when considering semi-parametric estimations. Here again, I stress the overestimations of the linear estimations which cannot be seen on the graphs as their levels are too high compared to those obtained using semi-parametric estimations (see Tables 2.10 and 2.13) We note the same decreasing pattern as observed for the whole period, underlining the higher role of parents' wage on child's education for poorer groups.

Let us now focus on the non linearities in wage mobility. I am not going to analyse the time evolution of the relationship between parents' and child's wages as the number of observations are too low. The distinction I make here is between yearly and hourly wage earners. In Figure 2-8, I report the derivative of the child's wages as a function of parents' wages.

For yearly wages, the coefficient estimated by linear specifications converges with that obtained from the semi-parametric estimations only for a narrow range of parents' wages values. Additionally non linearities also depict a worrying phenomenon. For parents who earn lower yearly wages, I emphasize a negative relationship between parents' and child's yearly wages, showing a significant and strong downward mobility. The marginal negative effects range between -1.7 and -0.5 (including the confidence intervals), which are very high values. For higher yearly wages, the impact of the parents' yearly wages is also very large and positive, emphasizing lower income mobility for the richer groups. Even if these conclusions are significant and precisely estimated, I need however to put them into perspective in terms of the number of observations on which the extremes of the relationship are estimated. The small lines present on the absciss axis give the number of observations used for the estimation. We clearly see that most of them are situated in the range for which the semi-parametric estimations converge with the linear ones. Given the sample size, our conclusions are only preliminary results which underline the necessity for non linear considerations when studying intergenerational mobility in China.

Results relative to mobility in hourly wages present another reality. These seem very much related to the conclusions found using quantile regressions. For the fraction of the semi-parametric estimations for which the confidence intervals are not too large, we observe a U shape emphasizing lower mobility for the poorer and higher wages groups. This can be linked to different factors.

For poorer families, I have just stressed that their income is more closely related to their children's schooling attainment. This is also supported by results linking parents' education to the child's one for the larger sample as I emphasized that for the bottom of the distribution , lower mobility was related to the farming origin of the children's family which is in turn linked to lower income. Consequently, children from poorer families tend to have lower access to a good level of education. This can then negatively influence their access to a high hourly wage.

For richer families, who also demonstrate lower mobility, this can be closely linked to the relationship existing between education, access to responsibilities in the communist party, networks which result from them, and wages. I stressed in the section 2.1 relative to linear mobility the new factors favouring the social promotion of highly educated children (who are mostly from richer families) after the reforms who then invest more in their children's education, in this manner creating a virtuous circle for their income categories but leading to lower mobility as transmission of social status will be high.

We have seen that considering non linearities in intergenerational mobilities are important as the impact of parents' characteristics on the child's can be different given the education or wage quantile which parents and children come from.

2.4 Conclusion

Using data from the CHNS database, which I use to take five periods of time into consideration, I have studied the intergenerational educational and income mobility in China. The specificity of the political as well as the economic context of this country shows the necessity of studying mobility within Chinese society.

Given that I have information on educational attainment and income for both parents and children who reside with their parents, the results of this study have required econometric strategies to help to control for selection processes both relative to co-residence and to being a wage earner when analysing wage mobility. Moreover, I carried out both linear and non linear mobility analysis as some arguments tend to support intergenerational mobility at different levels depending on where parents and children are situated in the distribution of education or wages.

Concerning the linear apporach, in order to improve the understanding of mobility in China, I distinguished between absolute and conditional mobility, the former being a description of movements between educational or wages categories of children compared to their parents, and the latter focusing on the causality link between parents' and child's education and wage characteristics.

I have emphasized the existence of absolute mobility in education using mobility matrices. This absolute mobility is completed by a relatively high conditional mobility in education considering the impact of parents' education (one extra year in the parents' education leads to approximately 0.2 extra years in the child's education). Nevertheless, worrying conclusions have been drawn from the role played by parents' income on the child's educational attainment with the highest impact of 0.5964 extra years in the child's education for an increase of 1 percent in the parents' yearly income. This observation tends to support recent literature developed around the increase in education inequalities due, among others, to the increase in entrance fees for colleges and universities. I have also underlined the role played by parents' occupations and particularly the negative impact of the parents' participation in farming activities. Moreover, this effect appears more important than the effect linked to parents' income. Inequalities in access to education in rural areas seem central for the child's educational attainment.

For income mobility, a one percent increase in parents' incomes implies a 0.25 to 0.29 percent increase in the child's income. These figures are not very high once we consider literature on income mobility in developed or developing countries. However, the role played by parents' wages in the child's schooling attainment could negatively affect the level of wage mobility in China.

Moving to a non linear approach to intergenerational mobility, I stressed the importance of further analysing the relationship between parents' and child's characteristics given that various levels of significance as well as coefficients describing this relationship are found for different quantiles of parents' and child's education and wages.

Using quantile regressions, I stressed a higher impact of the parents level of education for less educated children with this impact being particularly true for children from agricultural families (as noted when comparing the results relative to the larger sample including children with parents who are wage earners and parents with agricultural orignes to the subsamples of children whose parents are wage earners). This non linearity is even more pronounced using semi-parametric estimations as lower mobility is observed for children with less educated parents.

Concerning wage mobility, an interesting pattern appeared when considering semi-parametric estimations: the relationship between parents' and child's hourly wages follows a U shape emphasizing lower mobility for poorer and richer wage earners. For the bottom of the distribution, this can mirror the lower level of education of children from poorer families, which is then translated into lower mobility. For the richer, this can be the consequences of a virtuous circle where the highest paid parents invest more in their children's education, these children then invest heavily in their children's education, leading the richest staying richest, or even leading to an increasing gap between the richest and the poorest.

Regarding the results presented above, I can conclude that the egalitarian political regime during the MaoTse Dong period has succeeded relatively well in re-setting the hierarchical and traditional system where there was in a small intergenerational mobility in the social status of individuals.

Nevertheless, the increasing inequalities in income and in the financing of education following the reforms tend to mitigate this encouraging conclusion. Consequently, political concerns should focus more on the reduction of inequalities to create an "equal playing field".

We see once more with this chapter that mobility in wages is closely relative to education, these two dimensions being particularly and increasingly linked. The polarization which can come from lower mobility noted for the two extreme wage categories and linked to the different levels of possible investment from parents in the education of their children can lead to increasing inequalities in both wages and education.

The following chapter proposes consequently to provide deeper analysis of the evolution of well-being inequalities of income, education and health, given the results underlined in the two first chapters of the thesis.

Tables chapter 2

	Sample A		Sample B		Sample C		Sample D		Sample E	
Observations	5499		1.808		1.720		535		347	
	mean	standard deviation	mean	standard deviation	mean	standard deviation	mean	standard deviation	mean	standard deviation
Child characteristics										
hourly wage									1.572	1.945
yearly wage							3343.6	4722.91		
age	20.259	3.473	19.66	3.010	19.677	3.014	21.58	2.956	21.75	3.018
gender	0.571	0.495	0.533	0.499	0.528	0.499	0.511	0.500	0.497	0.501
public sector							0.672	0.470	0.640	0.481
urban sector	0.277	0.448	0.407	0.491	0.412	0.492	0.405	0.491	0.418	0.494
years schooling	8.940	2.639	9.951	2.388	9.994	2.400	10.22	2.615	10.415	2.707
Parents' characteristics										
hourly wage					1.945	4.048			1.981	2.842
yearly wage			3763.39	3817.1			3848.0	4385.8		
age	48.46	6.695	46.714	5.131	46.659	5.030	48.479	5.465	48.326	5.142
years schooling	5.724	3.501	7.4392	3.487	7.495	3.479	7.160	3.585	7.332	3.782
GINI coefficient										
Child income							0.3917		0.4567	
Parents' income							0.4561		0.4337	

Table 2.4: Summary statistics for the five samples chosen

	Child's educational status				
Parental education	primary	lower	upper	college	Observations
status (% total observations)	school	middle	middle		
primary school	20.46	58.48	16.62	2.44	3,232
lower middle	7.14	53.99	32.84	6.04	1,541
upper middle	3.16	34.74	49.65	12.46	570
college	1.28	8.33	57.69	32.69	156

Table 2.5: Educational mobility matrix for the larger sample

	Child's educational status				
Parental education	primary	lower	upper	college	Observations
status (% total observations)	school	middle	middle		
primary school	9.96	60.92	25.41	3.71	673
lower middle	4.65	46.48	40.93	7.95	667
upper middle	2.71	26.81	55.42	15.06	332
college	0.00	7.35	59.56	33.09	136

Table 2.6: Educational mobility matrix for the sample B

	Child yearly earnings quartile				
Parents's yearly earnings quartile (% total observations)	bottom	2^{nd}	3^{rd}	top	Observations
bottom	52.76	31.50	8.66	7.09	127
2^{nd}	25.20	22.83	35.43	16.54	127
3^{rd}	15.08	16.67	31.75	36.51	126
top	11.20	17.60	27.20	44.0	125

Table 2.7: Yearly wage mobility matrix

	Child hourly earnings quartile				
Parent's hourly earnings quartile (% total observations)	bottom	2^{nd}	3^{rd}	top	Observations
bottom	49.35	37.66	11.69	1.30	77
2^{nd}	23.81	23.81	34.52	17.86	84
3^{rd}	17.11	17.11	34.21	31.58	76
top	15.29	18.82	15.29	50.59	85

Table 2.8: Hourly wage mobility matrix

	Full sample	1991	1993	1997	2000	2004
	(1)	(2)	(3)	(4)	(5)	(6)
age	−0.3088	0.9426	−0.9503	−0.3148	−0.3504	−0.1716
	(0.0062)	(0.0443)	(0.0534)	(0.0133)	(0.0167)	(0.0094)
	[−0.1218]	[−0.0005]	[−0.0969]	[−0.1255]	[−0.1181]	[−0.0666]
urban	−0.3535	−0.3307	−0.2384	−0.5642	−0.4286	−0.3877
	(0.0392)	(0.1283)	(0.1488)	(0.0858)	(0.1099)	(0.0780)
	[−0.1369]	[−0.0001]	[−0.0225]	[−0.2213]	[−0.1500]	[0.1515]
D91	−1.5421					
	(0.0583)					
	[−0.4820]					
D93	−0.8013					
	(0.0577)					
	[−0.2874]					
D00	0.8085					
	(0.0630)					
	[0.3118]					
D04	1.017					
	(0.0613)					
	[0.3840]					
gender	0.5357	0.0835	0.0338	0.3877	0.8034	1.3445
	(0.0366)	(0.1100)	(0.1310)	(0.0782)	(0.1041)	(0.0887)
	[0.2085]	[0.00005]	[0.0034]	[0.1536]	[0.2712]	[0.4897]
Identification variables						
Number of children	0.0670	0.1754	0.7366	0.1331	−0.0292	
	(0.0646)	(0.2506)	(0.2402)	(0.1193)	(0.1408)	
	[0.0264]	[0.0001]	[0.0751]	[0.0531]	[−0.0098]	
Marital status	−0.5636	−0.7448	−1.3300	−0.8872	−0.2655	−0.6171
	(0.0482)	(0.2927)	(0.2525)	(0.0940)	(0.1122)	(0.0889)
	[−0.2168]	[−0.0004]	[−0.1024]	[−0.3415]	[−0.0909]	[−0.2350]
R^2	0.6079	0.7991	0.8350	0.6765	0.6952	0.4860
observations	11809	2453	2092	3135	1921	2207

(Huber-White standard errors in parentheses).
D91, D93, D00 and D04 are dummy variables which take the value of 1 respectively for the years 1991, 1993, 2000 and 2004, and 0 elsewhere.
1997 is taken as the year of reference. All incomes are expressed in natural logarithme.

Table 2.9: Coresidence selection procedure relative to ntergenerational mobility in education

	Full sample	Full sample	1991	1993	1997	2000	2004
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
age	0.0451 (0.0176)	0.1673 (0.0254)	−0.0059 (0.0858)	−0.0233 (0.0515)	0.1197 (0.0374)	0.1819 (0.0369)	0.0974 (0.0360)
urban	1.3361 (0.1632)	1.4754 (0.1643)	1.3128 (0.2640)	1.2499 (0.2310)	1.3024 (0.2061)	1.4799 (0.2359)	1.6519 (0.1923)
D91	−1.0332 (0.1423)	−0.4668 (0.1491)					
D93	−0.7024 (0.0995)	−0.4809 (0.0961)					
D00	−0.0921 (0.1070)	−0.2692 (0.1092)					
D04	0.2509 (0.1240)	0.1355 (0.1256)					
gender	−0.0156 (0.1199)	−0.1897 (0.1217)	0.6313 (0.1959)	0.3563 (0.1687)	−0.2415 (0.1229)	−0.4414 (0.1920)	−0.7361 (0.4449)
Parental characteristics							
age	0.0037 (0.0104)	0.0023 (0.0103)	−0.0012 (0.0171)	0.0107 (0.0157)	0.0010 (0.0141)	−0.0043 (0.0168)	0.0138 (0.0144)
years schooling	0.2418 (0.0213)	0.2412 (0.0212)	0.2419 (0.0320)	0.2685 (0.0291)	0.2431 (0.0252)	0.2429 (0.0364)	0.2091 (0.0295)
IMR: $\widehat{\lambda}_{it}$		−1.3788 (0.2435)	0.1986 (0.3411)	0.1569 (0.2737)	−1.0327 (0.3080)	−1.6631 (0.4924)	0.7720 (0.4449)
R^2	0.2506	0.2581	0.2071	0.2183	0.1759	0.2177	0.2133
observations	5499	5499	808	949	1641	1037	1064

(Huber-White standard errors in parentheses).
D91, D93, D00 and D04 are dummy variables which take the value of 1 respectively for the years 1991, 1993, 2000 and 2004, and 0 elsewhere.
1997 is taken as the year of reference. All incomes are expressed in natural logarithme.

Table 2.10: Intergenerational mobility in education for the larger sample A

	Without selection	With selection	Without selection	With selection
	(1)	(2)	(3)	(4)
age	0.0663 (0.0277)	0.2764 (0.0411)	0.0702 (0.0293)	0.2885 (0.0440)
urban	1.0613 (0.1347)	1.2819 (0.1352)	1.0864 (0.1464)	1.3171 (0.1467)
D91	−0.5523 (0.1587)	0.4277 (0.1988)	−0.6106 (0.1746)	0.3990 (0.2162)
D93	−0.3864 (0.1294)	0.0102 (0.1356)	−0.3755 (0.1422)	0.0326 (0.1484)
D00	0.0266 (0.1465)	−0.2789 (0.1534)	0.0226 (0.1494)	−0.3022 (0.1575)
D04	0.3379 (0.2074)	0.0986 (0.2012)	0.4637 (0.2050)	1.1910 (0.2017)
gender	−0.1511 (0.1282)	−0.4684 (0.1249)	−0.1124 (0.1299)	−0.4435 (0.1263)
Parental characteristics				
age	0.0411 (0.0157)	0.0361 (0.0154)	0.0397 (0.0171)	0.0339 (0.0167)
years schooling	0.1773 (0.0205)	0.1799 (0.0206)	0.1816 (0.0206)	0.1844 (0.0205)
yearly wage	0.3812 (0.0804)	0.3632 (0.0807)		
hourly wage			0.2449 (0.0968)	0.2301 (0.0971)
Inverse Mills ratio: $\widehat{\lambda}_{it}$		−2.5503 (0.3692)		−2.6411 (0.3880)
R^2	0.2575	0.2827	0.2496	0.2761
observations	1808	1808	1720	1720

OLS estimator (Huber-White standard errors in parentheses).
D91, D93, D00 and D04 are dummy variables which take the value of 1 respectively for 1991, 1993, 2000 and 2004, and 0 elsewhere.
1997 is taken as the year of reference. All incomes are expressed in natural logarithme.

Table 2.11: Intergenerational mobility in education for samples B and C including respectively yearly and hourly wage-earner parents

	Full sample	1991	1993	1997	2000	2004
	(1)	(2)	(3)	(4)	(5)	(6)
age	0.2764 (0.0411)	0.0546 (0.1453)	−0.0059 (0.0803)	0.1401 (0.0620)	0.2731 (0.0675)	0.1949 (0.0693)
urban	1.2819 (0.1352)	0.6204 (0.2234)	0.6467 (0.1975)	1.1862 (0.2257)	1.4335 (0.2698)	1.8688 (0.3198)
D91	0.4277 (0.1988)					
D93	0.0102 (0.1356)					
D00	−0.2789 (0.1534)					
D04	0.0986 (0.2012)					
gender	−0.4684 (0.1249)	0.2866 (0.2030)	0.0885 (0.2154)	−0.3099 (0.1706)	−0.3950 (0.2615)	−1.2355 (0.3901)
Parental characteristics						
age	0.0361 (0.0154)	0.0377 (0.0253)	0.0392 (0.0218)	0.0254 (0.0216)	0.0320 (0.0278)	0.0523 (0.0398)
years schooling	0.1799 (0.0206)	0.1713 (0.0346)	0.1983 (0.0347)	0.1798 (0.0294)	0.1863 (0.0454)	0.1490 (0.0495)
Yearly wage	0.3632 (0.0807)	−0.0841 (0.1650)	0.4773 (0.1645)	0.3796 (0.1476)	0.4961 (0.1479)	0.5640 (0.2253)
IMR: $\widehat{\lambda}_{it}$	−2.5503 (0.3692)	0.1241 (0.6802)	−0.2704 (0.6368)	−1.5445 (0.4858)	−2.9617 (1.1966)	−1.2195 (0.8858)
R^2	0.2827	0.1855	0.2225	0.1897	0.2523	0.2705
observations	1808	305	333	549	332	289

OLS estimator (Huber-White standard errors in parentheses).
D91, D93, D00 and D04 are dummy variables which take the value of 1 respectively for the years 1991, 1993, 2000 and 2004, and 0 elsewhere.
1997 is taken as the year of reference. All incomes are expressed in natural logarithme.

Table 2.12: Intergenerational mobility in education. Parents' income measured by the yearly income

	Full sample	1991	1993	1997	2000	2004
	(1)	(2)	(3)	(4)	(5)	(6)
age	0.2885	0.0348	−0.0233	0.1302	0.2904	0.2076
	(0.0440)	(0.1469)	(0.0556)	(0.0626)	(0.0696)	(0.0747)
urban	1.3171	0.5775	0.7464	1.1978	1.5213	1.9013
	(0.1467)	(0.2332)	(0.2236)	(0.2348)	(0.2840)	(0.3337)
D91	0.3990					
	(0.2162)					
D93	0.0326					
	(0.1484)					
D00	−0.3022					
	(0.1575)					
D04	1.1910					
	(0.2017)					
gender	−0.4435	0.3041	0.1745	−0.2851	−0.4253	−1.1919
	(0.1263)	(0.2057)	(0.2233)	(0.1724)	(0.2757)	(0.4218)
Parental characteristics						
age	0.0339	0.0324	0.0533	0.0275	−0.0278	0.0516
	(0.0167)	(0.0273)	(0.0280)	(0.0218)	(0.0283)	(0.0433)
years schooling	0.1844	0.1641	0.2225	0.1782	0.1953	0.1370
	(0.0205)	(0.0355)	(0.0353)	(0.0306)	(0.0498)	(0.0525)
Hourly wage	0.2301	−0.1369	0.1235	0.2669	0.2862	0.5158
	(0.0971)	(0.1907)	(0.2276)	(0.1738)	(0.2499)	(0.2289)
IMR: $\widehat{\lambda}_{it}$	−2.6411	0.2331	0.1544	−1.5606	−3.0054	0.7720
	(0.3880)	(0.7014)	(0.6737)	(0.4919)	(1.1997)	(0.4449)
R^2	0.2761	0.1754	0.2245	0.1865	0.2331	0.2541
observations	1720	293	309	529	320	269

OLS estimator (Huber-White standard errors in parentheses).
D91, D93, D00 and D04 are dummy variables which take the value of 1 respectively for the years 1991, 1993, 2000 and 2004, and 0 elsewhere.
1997 is taken as the year of reference. All incomes are expressed in natural logarithme.

Table 2.13: Intergenerational mobility in education. Parents' income measured by the hourly income

	Full sample	Full sample	Full sample	Full sample	Full sample
	(1)	(2)	(3)	(4)	(5)
age	0.1264 (0.0427)	0.2764 (0.0411)	0.1740 (0.0640)	0.2885 (0.0440)	0.1751 (0.0686)
urban	0.9387 (0.2156)	1.2819 (0.1352)	0.8152 (0.2404)	1.3171 (0.1467)	0.7857 (0.2542)
D91	−0.5432 (0.3279)	0.4277 (0.1988)	−0.2586 (0.4366)	0.3990 (0.2162)	−0.0008 (0.4619)
D93	−0.3469 (0.2054)	0.0102 (0.1356)	−0.2743 (0.2935)	0.0326 (0.1484)	−0.1381 (0.3173)
D00	−0.1777 (0.1976)	−0.2789 (0.1534)	0.1588 (0.2735)	−0.3022 (0.1575)	0.1565 (0.2859)
D04	0.2874 (0.2462)	0.0986 (0.2012)	0.9699 (0.4086)	1.1910 (0.2017)	0.9878 (0.4315)
gender	0.0532 (0.1676)	−0.4684 (0.1249)	−0.3198 (0.1957)	−0.4435 (0.1263)	−0.2337 (0.2088)
Parental characteristics					
age	0.0333 (0.0142)	0.0361 (0.0154)	0.0484 (0.0203)	0.0339 (0.0167)	0.0540 (0.0225)
years schooling	0.2686 (0.0350)	0.1799 (0.0206)	0.2612 (0.0354)	0.1844 (0.0205)	0.2697 (0.0369)
Yearly wage		0.3632 (0.0807)	0.0613 (0.1276)		
Hourly wage				0.2301 (0.0971)	0.3072 (0.1705)
Farming occupation					
Mother	−0.3877 (0.2388)		−0.2600 (0.2442)		−0.2569 (0.2555)
Father	−1.0421 (0.1957)		−0.8810 (0.2815)		−1.0981 (0.3218)
IMR: $\widehat{\lambda}_{it}$	−1.1474 (0.4894)	−2.5503 (0.3692)	−1.5581 (0.6095)	−2.6411 (0.3880)	−1.5935 (0.6331)
R^2	0.3129	0.2827	0.3938	0.2761	0.4070
observations	1805	1808	530	1720	490

OLS estimator (Huber-White standard errors in parentheses).
D91, D93, D00 and D04 are dummy variables which take the value of 1 respectively of 1991, 1993, 2000 and 2004, and 0 elsewhere.
1997 is taken as the year of reference. All incomes are expressed in natural logarithme.

Table 2.14: Intergenerational mobility in education including parents' occupation. Parents' income measured by the yearly and the hourly incomes

	probit			biprobit
	"live with parents"	"wage earner"	"live with parents"	"wage earner"
	(1)	(2)	(3)	(4)
age	−0.3283	0.0012	0.3282	0.0013
	(0.0078)	(0.0053)	(0.0131)	(0.0053)
	[−0.0540]	[0.0005]		
urban	−0.3849	0.3037	0.3852	0.3032
	(0.0498)	(0.0894)	(0.1125)	(0.0895)
	[−0.0572]	[0.1200]		
D91	−1.7837	−0.0708	−1.7843	−0.0715
	(0.0738)	(0.0832)	(0.1113)	(0.0832)
	[−0.1871]	[−0.0278]		
D93	−0.9436	−0.1933	−0.9448	−0.1939
	(0.0683)	(0.0822)	(0.0957)	(0.0821)
	[−0.1100]	[−0.0751]		
D00	0.8497	−0.0053	0.8497	−0.0058
	(0.0704)	(0.0748)	(0.0670)	(0.0748)
	[0.1995]	[0.0021]		
D04	1.2182	0.4288	1.2179	0.4293
	(0.0733)	(0.0909)	(0.1149)	(0.0909)
	[0.0209]	[0.1698]		
public sector	−0.2631	2.3553	−0.2663	2.3558
	(0.0519)	(0.0732)	(0.0716)	(0.0733)
	[−0.0407]	[0.7469]		
gender	0.5829	0.0710	0.5810	0.0728
	(0.0430)	(0.0368)	(0.0662)	(0.0369)
	[0.0956]	[0.0279]		
years schooling	0.0041	0.0923	0.0043	0.0922
	(0.0081)	(0.0084)	(0.0121)	(0.0084)
	[0.0007]	[0.0363]		
Identification variables				
Number of children	−0.1845	−0.0530	−0.1857	−0.0531
	(0.0629)	(0.0543)	(0.0495)	(0.0543)
	[−0.0303]	[−0.0209]		
Marital status	−0.8063	0.0131	−0.8071	0.0125
	(0.0534)	(0.0637)	(0.0690)	(0.0637)
	[−0.1460]	[0.0052]		
Commodity prices: $F-$test of joint significance ($p-value$)	χ^2_2= 2.22	χ^2_2= 4.09	χ^2_2= 4.73	χ^2_2= 4.73
	(0.3303)	(0.1295)	(0.3158)	(0.3158)
R^2	0.6375	0.4787		
observations	10612	10612	10612	10612

(Huber-White standard errors in parentheses).
D91, D93, D00 and D04 are dummy variables which take the value of 1 respectively for 1991, 1993, 2000 and 2004, and 0 elsewhere. Figures in brakets are the marginal effects. 1997 is taken as the year of reference. All incomes are expressed in natural logarithme.

Table 2.15: The "living with parents" and "wage earner" selections probit and biprobit for the sample D

	2SLS	2SLS	2SLS	2SLS	Tobit
	Without selection	Independant selections	Dependant selections	Dependant selections	One selection
	(1)	(2)	(3)	(4)	(5)
age	0.0398 (0.0131)	0.0590 (0.0326)	0.0581 (0.0333)	0.0615 (0.0360)	0.3310 (0.1201)
urban	0.1727 (0.0793)	0.1653 (0.0799)	0.1642 (0.0799)	0.1914 (0.0874)	0.8696 (0.5766)
D91	−1.1905 (0.1100)	−1.0782 (0.1991)	−1.0793 (0.2003)	−1.0576 (0.2179)	−1.8245 (0.8049)
D93	−0.9434 (0.0900)	−0.8816 (0.1152)	−0.8819 (0.1143)	−0.8676 (0.1251)	−2.1074 (0.6184)
D00	0.3405 (0.0794)	0.3120 (0.0866)	0.3085 (0.0859)	0.3416 (0.0959)	−0.5961 (0.5712)
D04	0.6987 (0.0882)	0.5973 (0.1044)	0.6004 (0.1030)	0.6195 (0.1199)	1.8639 (0.8369)
public sector	−0.0391 (0.0714)	−0.3156 (0.3622)	−0.2554 (0.3457)	−0.4952 (0.3559)	8.6244 (0.7495)
gender	0.0870 (0.0458)	0.0495 (0.0762)	0.0538 (0.0775)	0.0336 (0.0805)	−1.0480 (0.3976)
years schooling	−0.0569 (0.0320)	−0.0605 (0.0450)	−0.0536 (0.0453)	−0.0816 (0.0454)	0.4195 (0.2651)
Parental characteristics					
age	−0.0070 (0.0067)	−0.0076 (0.0066)	−0.0079 (0.0066)	−0.0049 (0.0062)	0.0000 (0.0320)
yearly wage	0.2973 (0.0701)	0.2890 (0.0708)	0.2869 (0.0706)	0.2938 (0.0745)	0.4142 (0.0702)
yearly wage indirect effect	0.2909 (0.0685)	0.2809 (0.0691)	0.2704 (0.0648)	0.2517 (0.0702)	
Father farming occupation				0.0977 (0.0940)	
IMR: living with parents		−0.1574 (0.2100)	−0.1680 (0.3325)	−0.4015 (0.3430)	−0.9724 (0.7163)
IMR: wage earner		−0.2323 (0.3548)	−0.1455 (0.2112)	−0.1619 (0.2269)	
R^2	0.6976	0.7007	0.7003	0.6903	
Hansen test of the overidentifying restrictions ($p-$value)		χ^2_2= 11.865 0.0184	χ^2_2= 12.217 0.0158	χ^2_2= 11.086 0.0256	
% censured observations					64.02
observations	528	528	528	492	2857

(Huber-White standard errors in parentheses).
D91, D93, D00 and D04 are dummy variables which take the value of 1 respectively for the years 1991, 1993, 2000 and 2004, and 0 elsewhere.
1997 is taken as the year of reference. All incomes are expressed in natural logarithme.

Table 2.16: Intergenerational yearly income mobility. 2SLS and tobit estimators' results

	probit			biprobit
	"live with parents"	"wage earner"	"live with parents"	"wage earner"
	(1)	(2)	(3)	(4)
age	−0.2958 (0.0083) [−0.0699]	0.0020 (0.0054) [0.0008]	−0.2958 (0.0083)	0.0020 (0.0054)
urban	−0.3942 (0.0576) [−0.0854]	0.2337 (0.0480) [0.0924]	−0.3942 (0.0576)	0.2337 (0.0480)
D91	−1.6010 (0.0991) [−0.2174]	0.0642 (0.0729) [0.0258]	−1.6010 (0.0991)	0.0642 (0.0729)
D93	−0.8807 (0.0841) [−0.1557]	−0.0492 (0.0665) [−0.0194]	−0.8807 (0.0841)	−0.0492 (0.0665)
D00	0.7336 (0.0800) [0.2121]	0.0840 (0.0615) [0.0334]	0.7336 (0.0800)	0.0840 (0.0615)
D04	1.1289 (0.0791) [0.3536]	2.2962 (0.0735) [0.6041]	1.1289 (0.0791)	2.2962 (0.0735)
public sector	−0.2761 (0.0611) [−0.0613]	2.0370 (0.0508) [0.6409]	−0.2761 (0.0611)	2.0370 (0.0508)
gender	0.6410 (0.0497) [0.1494]	0.0395 (0.0401) [0.0159]	0.6410 (0.0497)	0.0395 (0.0401)
years schooling	−0.0046 (0.0094) [−0.0011]	0.0761 (0.0082) [0.0304]	−0.0046 (0.0094)	0.0761 (0.0082)
Identification variables				
Number of children	−0.2775 (0.0758) [−0.0657]	−0.1006 (0.0512) [−0.0401]	−0.2775 (0.0758)	−0.1006 (0.0512)
Marital status	−0.7866 (0.0599) [−0.1951]	0.0393 (0.0576) [−0.0157]	−0.7866 (0.0599)	0.0393 (0.0576)
Commodity prices: $F-$test of joint significance $(p - value)$	χ^2_6= 17.81 (0.0067)	χ^2_6= 50.20 (0.0000)	χ^2_6= 93.61 (0.0000)	χ^2_6= 93.61 (0.0000)
R^2	0.6032	0.4955		
test of ρ= 0			χ^1_6= 24.67 (0.6194)	χ^1_6= 24.67 (0.6194)
observations	7159	7159	7159	7159

(Huber-White standard errors in parentheses).
D91, D93, D00 and D04 are dummy variables which take the value of 1 respectively for 1991, 1993, 2000 and 2004, and 0 elsewhere. Figures in brakets are the marginal effects. 1997 is taken as the year of reference. All incomes are expressed in natural logarithme.

Table 2.17: The "living with parents" and "wage earner" selections probit and biprobit for the sample E

	2SLS	2SLS	2SLS	2SLS	Tobit
	Without selection	Independant selection	Dependant selection	Dependant selection	One selection
	(1)	(2)	(3)	(4)	(5)
age	0.0475 (0.0112)	0.0819 (0.0224)	0.0817 (0.0226)	0.0805 (0.0239)	0.0572 (0.0337)
urban	0.0977 (0.0688)	0.1597 (0.0685)	0.1589 (0.0685)	0.1644 (0.0784)	0.0957 (0.2129)
D91	−1.3098 (0.1087)	−1.1041 (0.1484)	−1.1048 (0.1490)	−1.0899 (0.1606)	−3.5644 (0.2893)
D93	−0.9842 (0.0886)	−0.8706 (0.1036)	−0.8706 (0.1038)	−0.8666 (0.1097)	−0.7707 (0.2181)
D00	0.2340 (0.0771)	0.2008 (0.0790)	0.2005 (0.0790)	0.2338 (0.0842)	0.0637 (0.0856)
D04	0.5995 (0.1140)	0.6667 (0.1186)	0.6663 (0.1185)	0.7408 (0.1251)	0.9996 (0.3883)
public sector	−0.1383 (0.0984)	0.0315 (0.1321)	0.0312 (0.1319)	0.0334 (0.1419)	0.6304 (0.4275)
gender	0.0843 (0.0532)	0.0167 (0.0725)	0.0172 (0.0728)	0.0199 (0.0774)	−0.0089 (0.0769)
years schooling	−0.0318 (0.0331)	−0.0352 (0.0313)	−0.0347 (0.0313)	−0.0535 (0.0355)	−0.0471 (0.1994)
Parental characteristics					
age	−0.0012 (0.0062)	−0.0004 (0.0061)	−0.0004 (0.0061)	−0.0000 (0.0065)	−0.0018 (0.0084)
hourly wage	0.2729 (0.0610)	0.2728 (0.0618)	0.2725 (0.0618)	0.2717 (0.0645)	0.3773 (0.1114)
hourly wage indirect effect	0.2654 (0.0522)	0.2613 (0.0549)	0.2654 (0.0520)	0.2594 (0.0537)	
Father farming occupation				−0.2717 (0.0645)	
IMR: living with parents		−0.2826 (0.1649)	−0.2805 (0.1653)	−0.2999 (0.1752)	−0.0719 (0.1690)
IMR: wage earner		0.1625 (0.1218)	0.1628 (0.1198)	0.1751 (0.1259)	
R^2	0.7683	0.8118	0.8120	0.8053	
Hansen test of the overidentifying restrictions ($p-$value)	χ^2_6= 7.47 0.2794	χ^2_6= 15.73 0.0464	χ^2_6= 15.72 0.0465	χ^2_6= 14.80 0.0632	
% censured observations					78.84
observations	342	342	342	315	1824

(Huber-White standard errors in parentheses).
D91, D93, D00 and D04 are dummy variables which take the value of 1 respectively for the years 1991, 1993, 2000 and 2004, and 0 elsewhere.
1997 is taken as the year of reference. All incomes are expressed in natural logarithme.

Table 2.18: Intergenerational hourly income mobility. 2SLS and tobit estimators' results

	Full sample	Lower quartile	Median	Upper quartile
	(1)	(2)	(3)	(4)
age	0.1673 (0.0254)	0.0673 (0.0195)	0.0440 (0.0102)	0.1477 (0.0122)
urban	1.4754 (0.1643)	1.0110 (0.0806)	0.8415 (0.0438)	1.3065 (0.0525)
D91	−0.4668 (0.1491)	−1.1896 (0.1308)	−0.3509 (0.0709)	−0.2252 (0.0859)
D93	−0.4809 (0.0961)	−0.6544 (0.1133)	−0.2688 (0.0606)	−0.3355 (0.0729)
D00	−0.2692 (0.1092)	−0.0961 (0.1080)	−0.0959 (0.0577)	−0.2239 (0.0692)
D04	0.1355 (0.1256)	0.2021 (0.1117)	0.1133 (0.0598)	0.2213 (0.0715)
gender	−0.1897 (0.1217)	−0.1271 (0.0748)	−0.1121 (0.0404)	−0.3181 (0.0496)
Parental characteristics				
age	0.0023 (0.0103)	−0.0017 (0.0071)	0.0119 (0.0035)	0.0105 (0.0042)
years schooling	0.2412 (0.0212)	0.2451 (0.0118)	0.1297 (0.0061)	0.1788 (0.0079)
Inverse Mills ratio: living with parents	−1.3788 (0.2435)	−0.8788 (0.1898)	−0.4708 (0.1058)	−1.3795 (0.1363)
R^2	0.2581	0.1296	0.0397	0.1557
observations	5499	5499	5499	5499

Table 2.19: Intergenerational mobility in education: quantile regressions (sample A)

	Full sample	Lower quartile	Median	Upper quartile
	(1)	(2)	(3)	(4)
age	0.2764 (0.0411)	0.0419 (0.0189)	0.1568 (0.0214)	0.3647 (0.0305)
urban	1.2819 (0.1352)	0.6864 (0.0681)	0.9192 (0.0804)	1.1207 (0.1135)
D91	0.4277 (0.1988)	−0.0976 (0.1224)	−0.2619 (0.1446)	0.1057 (0.2067)
D93	0.0102 (0.1356)	−0.0466 (0.1020)	−0.0267 (0.1241)	−0.2344 (0.1789)
D00	−0.2789 (0.1534)	−0.0975 (0.0989)	−0.0116 (0.1174)	−0.4591 (0.1685)
D04	0.0986 (0.2012)	0.1132 (0.1097)	0.2410 (0.1297)	0.1983 (0.1801)
gender	−0.4684 (0.1249)	−0.1939 (0.0660)	−0.3539 (0.0799)	−0.5195 (0.1175)
Parental characteristics				
age	0.0361 (0.0154)	0.0180 (0.0077)	0.0300 (0.0085)	0.0379 (0.0115)
years schooling	0.1799 (0.0206)	0.1088 (0.0098)	0.1213 (0.0119)	0.1734 (0.0183)
yearly wage	0.3632 (0.0807)	0.2239 (0.0481)	0.2804 (0.0566)	0.2524 (0.00812)
Inverse Mills ratio: living with parents	−2.5503 (0.3692)	−0.8698 (0.2083)	−1.8792 (0.2455)	−3.3441 (0.3927)
R^2	0.2827	0.0429	0.1532	0.1865
observations	1808	1808	1808	1808

Table 2.20: Intergenerational mobility in education: quantile regressions. Parents' income measured by the yearly wage (sample B)

	Full sample	Lower quartile	Median	Upper quartile
	(1)	(2)	(3)	(4)
age	0.2885	0.0402	0.1699	0.3665
	(0.0440)	(0.0158)	(0.0235)	(0.0309)
urban	1.3171	0.6412	0.9969	1.1763
	(0.1467)	(0.0575)	(0.0879)	(0.1140)
D91	0.3990	−0.0967	0.2840	0.0805
	(0.2162)	(0.1041)	(0.1590)	(0.2125)
D93	0.0326	−0.0428	0.0642	−0.2636
	(0.1484)	(0.0881)	(0.1378)	(0.1839)
D00	−0.3022	−0.0854	−0.0565	−0.4244
	(0.1575)	(0.0824)	(0.1276)	(0.1664)
D04	1.1910	0.1695	0.2679	0.2535
	(0.2017)	(0.0921)	(0.1420)	(0.1802)
gender	−0.4435	−0.1783	−0.3520	−0.4920
	(0.1263)	(0.0557)	(0.0876)	(0.1172)
Parental characteristics				
age	0.0339	0.0143	0.0220	0.0413
	(0.0167)	(0.0067)	(0.0095)	(0.0117)
years schooling	0.1844	0.1160	0.1215	0.1618
	(0.0205)	(0.0081)	(0.0131)	(0.0185)
hourly wage	0.2301	0.1410	0.2381	0.1279
	(0.0971)	(0.0446)	(0.0645)	(0.0856)
Inverse Mills ratio: living with parents	−2.6411	−0.8362	−2.0258	−3.3807
	(0.3880)	(0.1760)	(0.2707)	(0.3904)
R^2	0.2761	0.0363	0.1530	0.1816
observations	1720	1720	1720	1720

Table 2.21: Intergenerational mobility in education: quantile regressions. Parents' income measured by the hourly income (sample C)

	Full sample	Lower quartile	Median	Upper quartile
	(1)	(2)	(3)	(4)
age	0.0581 (0.0333)	0.0634 (0.0188)	0.0361 (0.0222)	0.0172 (0.0269)
urban	0.1642 (0.0799)	0.1787 (0.0529)	0.1304 (0.0667)	0.0507 (0.0869)
D91	−1.0793 (0.2003)	−0.9009 (0.1146)	−1.0336 (0.1414)	−1.3212 (0.1833)
D93	−0.8819 (0.1143)	−0.7827 (0.0805)	−0.8170 (0.0989)	−0.9008 (0.1280)
D00	0.3085 (0.0859)	0.1548 (0.0680)	0.2889 (0.0875)	0.2492 (0.1140)
D04	0.6004 (0.1030)	0.5555 (0.0899)	0.6123 (0.1413)	0.5841 (0.1506)
public sector	−0.2554 (0.3457)	0.1557 (0.1847)	−0.0357 (0.2477)	0.0967 (0.3497)
gender	0.0538 (0.0775)	0.0399 (0.0492)	0.0922 (0.0712)	0.1802 (0.0833)
years schooling	−0.0536 (0.0453)	0.0075 (0.0119)	−0.0131 (0.0150)	0.0133 (0.0193)
Parental characteristics				
age	−0.0079 (0.0066)	−0.0035 (0.0043)	−0.0105 (0.0056)	−0.0108 (0.0073)
yearly wage	0.2869 (0.0706)	0.3802 (0.0316)	0.3772 (0.0413)	0.2269 (0.0676)
Inverse Mills ratio: living with parents	−0.1680 (0.3325)	0.2454 (0.1494)	0.0889 (0.1996)	0.2321 (0.2787)
Inverse Mills ratio: wage earner	−0.1455 (0.2112)	−0.2731 (0.1156)	−0.0621 (0.1424)	0.1194 (0.1784)
R^2	0.7003	0.5498	0.5350	0.4774
observations	528	528	528	528

Table 2.22: Intergenerational mobility in yearly wage: quantile regressions

	Full sample	Lower quartile	Median	Upper quartile
	(1)	(2)	(3)	(4)
age	0.0817	0.0943	0.0598	0.0711
	(0.0226)	(0.0359)	(0.0222)	(0.0262)
urban	0.1589	0.1975	0.0982	0.0808
	(0.0685)	(0.1236)	(0.0720)	(0.0948)
D91	−1.1048	−1.0532	−1.2341	−1.1439
	(0.1490)	(0.2589)	(0.1618)	(0.2043)
D93	−0.8706	−0.8798	−0.9297	−0.8012
	(0.1038)	(0.1787)	(0.1124)	(0.1347)
D00	0.2005	0.0641	0.2023	0.2065
	(0.0790)	(0.1603)	(0.0942)	(0.1119)
D04	0.6663	0.5993	0.5775	0.5179
	(0.1185)	(0.2255)	(0.1413)	(0.1790)
public sector	0.0312	0.2229	−0.1034	−0.1326
	(0.1319)	(0.1991)	(0.1283)	(0.1797)
gender	0.0172	−0.0122	0.0922	0.0344
	(0.0728)	(0.1198)	(0.0704)	(0.0878)
years schooling	−0.0347	−0.0083	0.0074	−0.0008
	(0.0313)	(0.0254)	(0.0143)	(0.0170)
Parental characteristics				
age	−0.0004	−0.0083	−0.0023	0.0031
	(0.0061)	(0.0102)	(0.0066)	(0.0078)
hourly wage	0.2725	0.2614	0.2183	0.2717
	(0.0618)	(0.0931)	(0.0496)	(0.0646)
Inverse Mills ratio: living with parents	−0.2805	0.3125	0.2100	0.1741
	(0.1653)	(0.1851)	(0.1123)	(0.1512)
Inverse Mills ratio: wage earner	0.1628	−0.4648	−0.1758	−0.2544
	(0.1198)	(0.2514)	(0.1528)	(0.1862)
R^2	0.8120	0.5913	0.5594	0.5240
observations	342	342	342	342

Table 2.23: Intergenerational mobility in hourly wage: quantile regressions

Figures chapitre 2

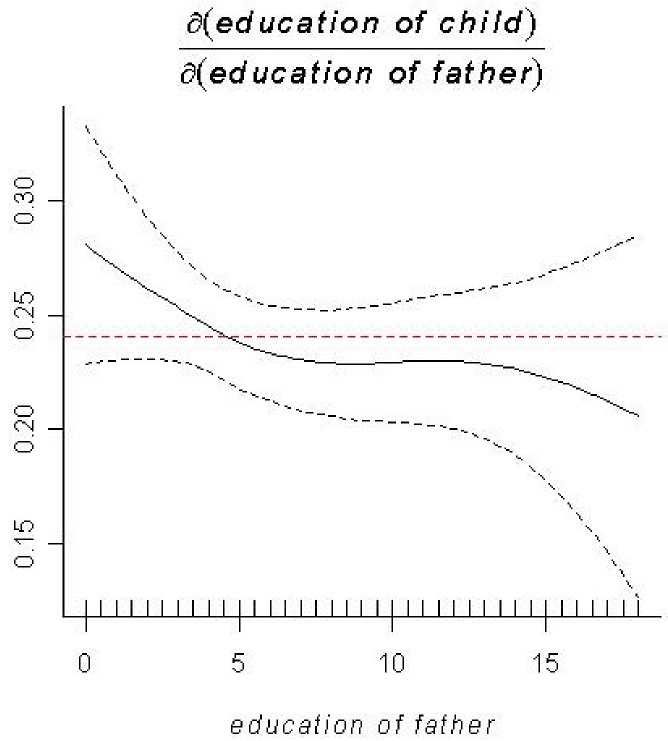


Figure 2-2: Semi-parametric estimation of educational mobility for the larger sample and for the whole period

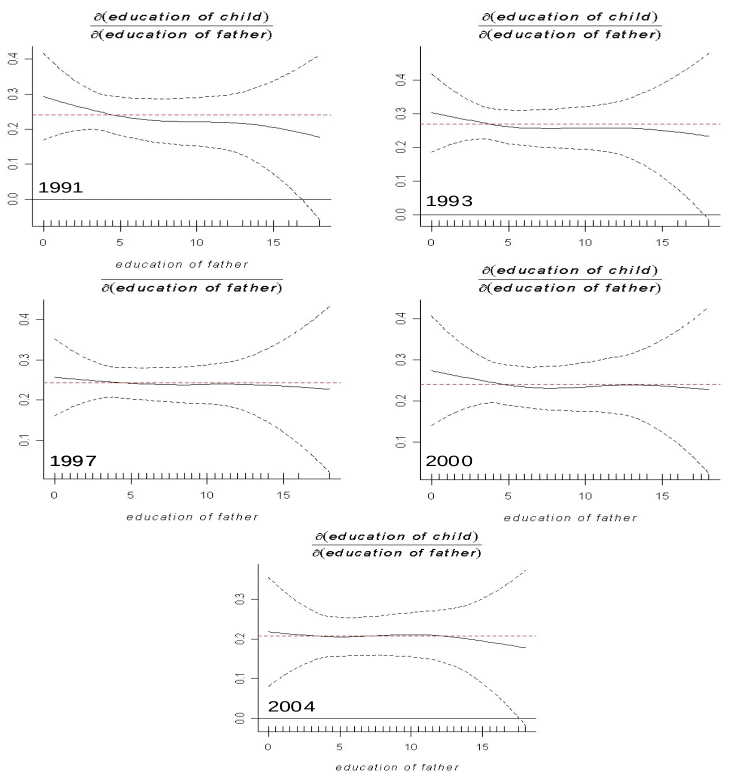


Figure 2-3: Semi-parametric estimation of educational mobility for the larger sample considering time evolution

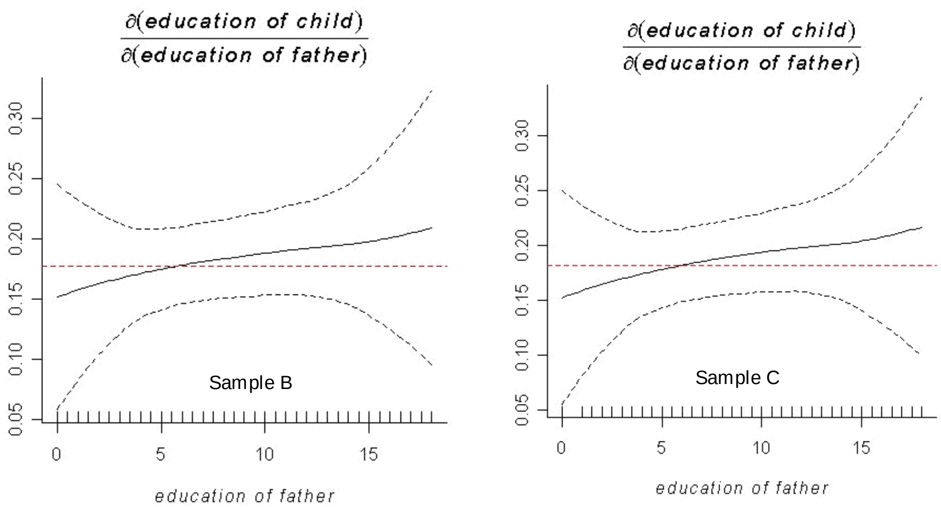


Figure 2-4: Semi-parametric estimations of educational mobility for the sub-samples B and C and for the whole period

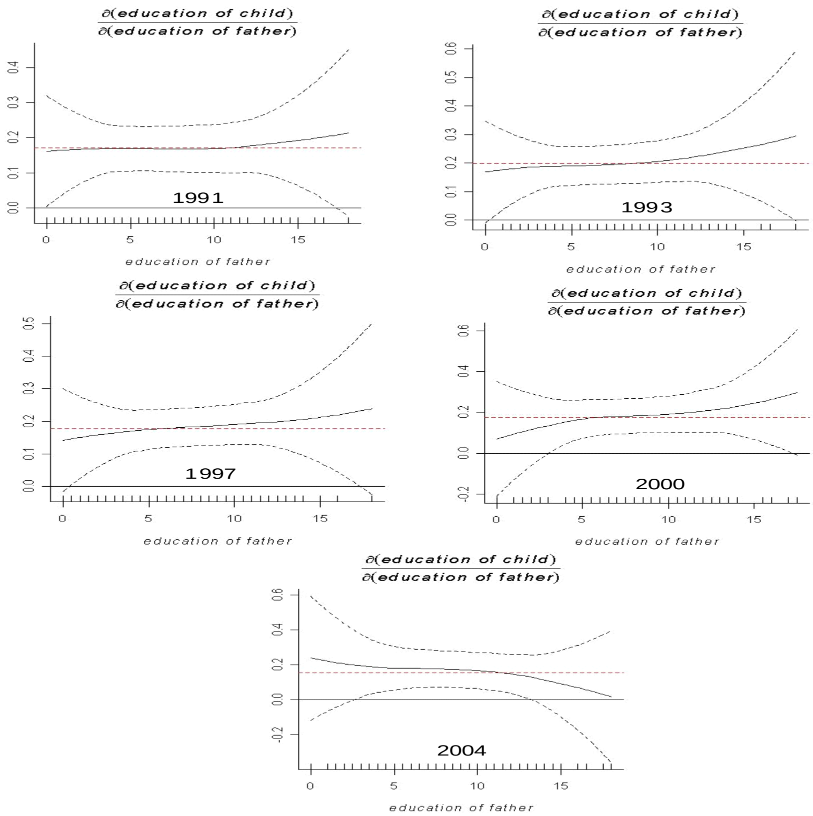


Figure 2-5: Semi-parametric estimation of educational mobility for the sample B considering time evolution

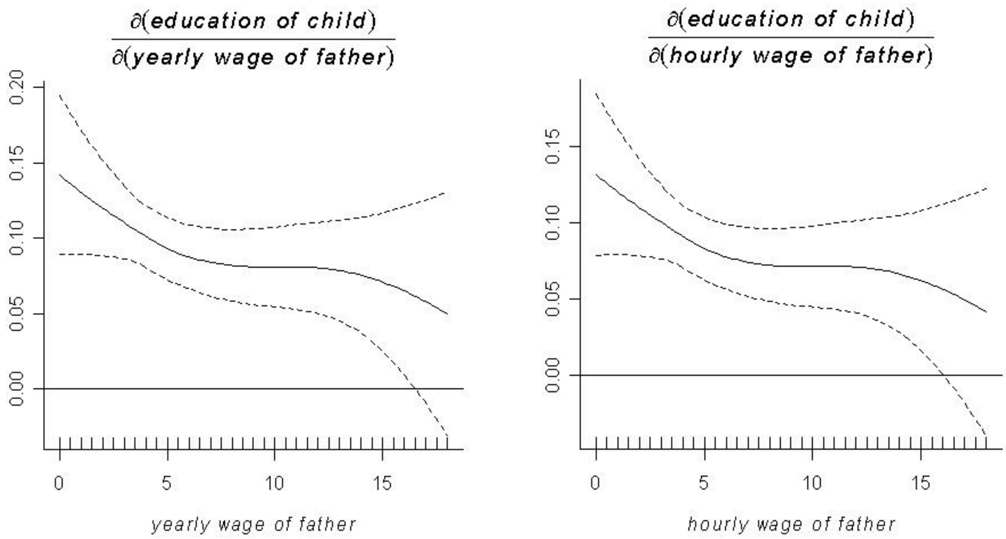


Figure 2-6: Semi-parametric estimations of the impact of parents' wages on the child's educational attainment for the sub-samples B and C and for the whole period

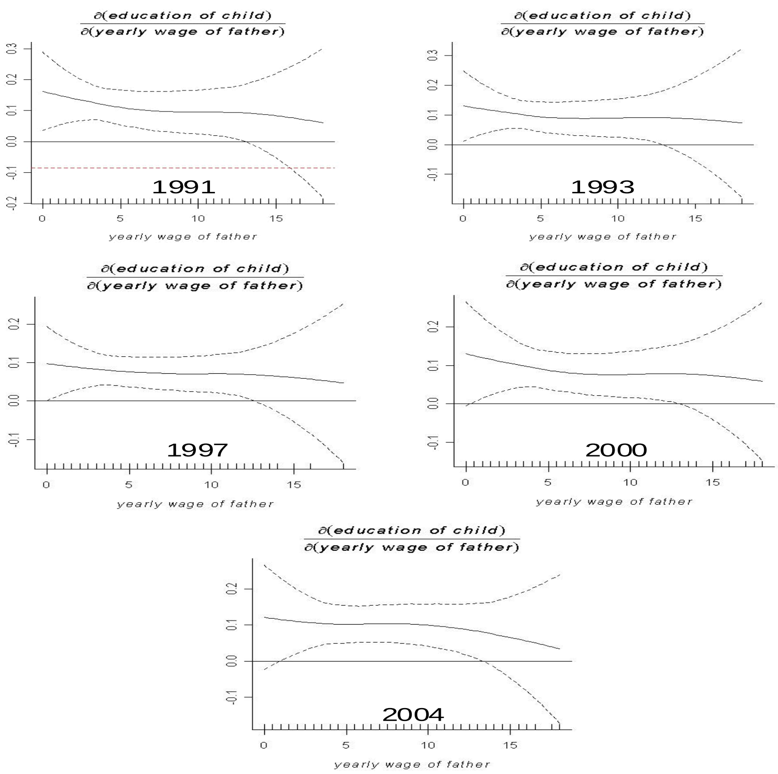


Figure 2-7: Semi-parametric estimations of the impact of parents' wages on the child's educational attainment for the sub-sample B considering time evolution

Figure 2-8: Semi-parametric estimations of wages mobility

2.A Appendices: Chapter 2

2.A.1 Tables of completentary results

	1991	1993	1997	2000	2004
Average age	18.0157	19.0790	19.8351	20.3765	20.9339
	(1.3434)	(1.9097)	(2.7600)	(3.3329)	(4.2373)
% of men	44.92	51.65	53.19	55.72	61.25
	(49.82)	(50.05)	(49.94)	(49.75)	(48.80)
Educational attainment	8.9377	9.3213	10.0346	10.3916	11.0830
	(1.9783)	(2.0946)	(2.2008)	(2.4917)	(2.6784)
% living in urban areas	42.29	41.14	32.60	43.07	51.21
	(49.48)	(48.28)	(46.92)	(49.59)	(50.07)
Observations	305	333	549	332	289

Table 2.24: Characteristics of children living with their parents: evolution with time for the sample of yearly wage earners parents

	1991	1993	1997	2000	2004
Average age	18.0078	19.0239	19.8773	20.4278	20.9602
	(1.3415)	(1.8981)	(2.7928)	(3.3320)	(4.2008)
% of men	45.39	50.48	52.74	55.63	60.59
	(49.87)	(50.08)	(49.97)	(49.76)	(48.96)
Educational attainment	8.9590	9.4045	10.0567	10.4094	11.1784
	(1.9789)	(2.0784)	(2.2132)	(2.5222)	(2.7039)
% living in urban areas	41.64	41.75	32.70	43.44	53.90
	(49.38)	(49.39)	(46.96)	(49.64)	(49.94)
Observations	293	309	529	320	269

Table 2.25: Characteristics of children living with their parents: evolution with time for the sample of hourly wage earners parents

	Child's educational status				
Parental education	primary	lower	upper	college	Observations
status (% total observations)	school	middle	middle		
primary school	9.71	59.71	26.75	3.82	628
lower middle	4.67	46.35	40.75	8.24	643
upper middle	2.53	26.90	55.06	15.51	316
college	0.00	6.77	59.40	33.83	133

Table 2.26: Education mobility matrix for the sample C

	Full sample	Lower quartile	Median	Upper quartile
	(1)	(2)	(3)	(4)
age	0.2251	0.0388	0.1531	0.3613
	(0.0389)	(0.0161)	(0.0191)	(0.0295)
urban	1.1044	0.6417	0.9613	1.1643
	(0.1401)	(0.0588)	(0.0716)	(0.1109)
D91	−0.1366	−0.1946	0.1242	−0.0516
	(0.1799)	(0.1025)	(0.1241)	(0.1933)
D93	−0.2813	−0.1692	−0.0262	−0.4037
	(0.1355)	(0.0867)	(0.1080)	(0.1690)
D00	−0.0393	−0.0474	0.1651	−0.4447
	(0.1399)	(0.0854)	(0.1043)	(0.1617)
D04	0.5194	0.2607	0.5170	0.2144
	(0.1945)	(0.0912)	(0.1123)	(0.1700)
gender	−0.3892	−0.1888	−0.3276	−0.5167
	(0.1227)	(0.0571)	(0.0714)	(0.1142)
Parental characteristics				
age	0.0318	0.0143	0.0217	0.0410
	(0.0157)	(0.0067)	(0.0076)	(0.0112)
years schooling	0.1867	0.1088	0.1234	0.1743
	(0.0209)	(0.0083)	(0.0105)	(0.0177)
Inverse Mills ratio: living with parents	−2.6220	−0.8588	−1.7834	−3.3752
	(0.3952)	(0.1803)	(0.2247)	(0.3748)
R^2	0.2737	0.0361	0.1459	0.1823
observations	1808	1808	1808	1808

Table 2.27: Intergenerational mobility in education for the sample B: quantile regressions without parents' yearly wages

	Full sample	Lower quartile	Median	Upper quartile
	(1)	(2)	(3)	(4)
age	0.2307	0.0437	0.1539	0.3889
	(0.0408)	(0.0174)	(0.0209)	(0.0284)
urban	1.1143	0.6459	1.0032	1.1396
	(0.1485)	(0.0628)	(0.0783)	(0.1046)
D91	−0.1131	−0.1652	0.0828	0.0167
	(0.1798)	(0.1072)	(0.1353)	(0.1821)
D93	−0.2161	−0.1356	−0.0452	−0.3125
	(0.1327)	(0.0919)	(0.1182)	(0.1603)
D00	−0.0745	−0.0613	−0.0714	−0.4553
	(0.1462)	(0.0896)	(0.1133)	(0.1528)
D04	0.5228	0.2586	0.0225	0.3404
	(0.1996)	(0.0994)	(0.0085)	(0.1623)
gender	−0.3546	−0.1663	−0.2737	−0.5191
	(0.1258)	(0.0607)	(0.0781)	(0.1080)
Parental characteristics				
age	0.0338	0.0138	0.0255	0.0384
	(0.0172)	(0.0074)	(0.0085)	(0.0108)
years schooling	0.1874	0.1114	0.1220	0.1696
	(0.0205)	(0.0088)	(0.0116)	(0.0168)
Inverse Mills ratio: living with parents	−2.7133	−0.8877	−1.7892	−3.5778
	(0.4049)	(0.1936)	(0.2471)	(0.3631)
R^2	0.2740	0.0337	0.1503	0.1804
observations	1720	1720	1720	1720

Table 2.28: Intergenerational mobility in education for the sample C: quantile regressions without parents' hourly wages

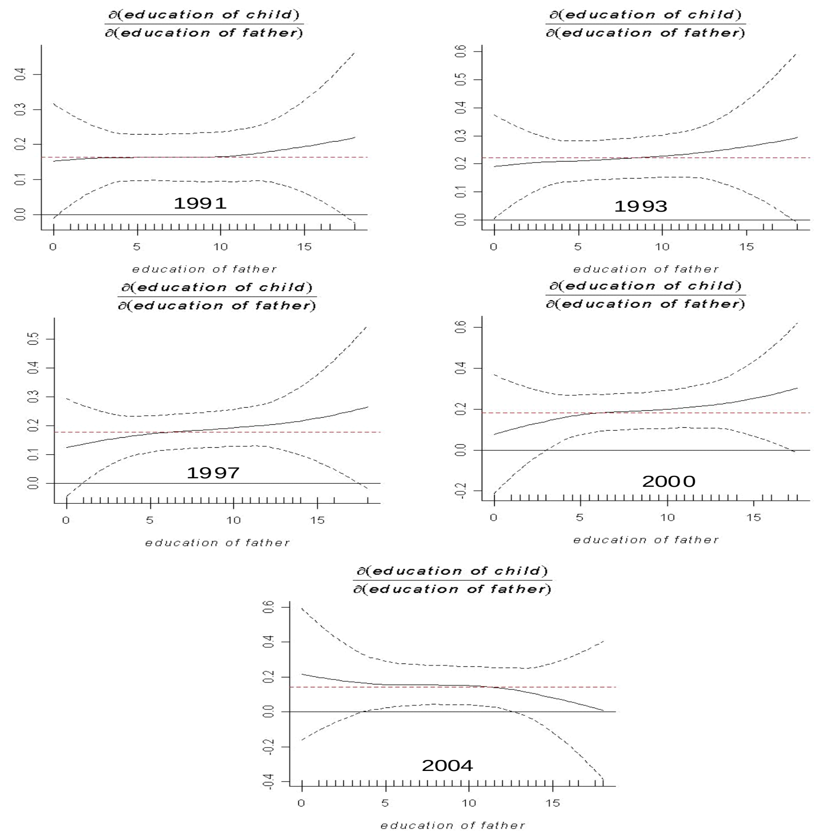


Figure 2-9: Semi-parametric estimation of educational mobility for the sample C considering time evolution

2.A.2 Figures of non parametric estimations for educational mobilty and for the sample C

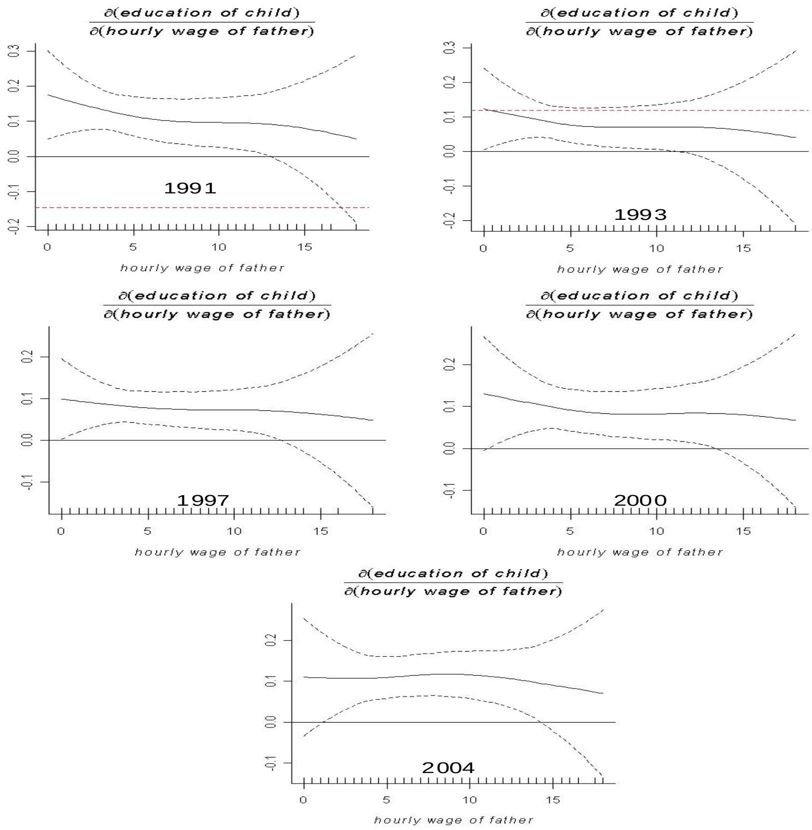


Figure 2-10: Semi-parametric estimations of the impact of parents' wages on the child's educational attainment for the sub-sample C considering time evolution

Chapter 3

Unidimensional and multidimensional views of economic and social inequalities in China

The evolution of China's environment linked to the reform movement has had many consequences for the determination of wages and educational attainment. Regarding the path toward a labour market, returns to human capital including health and education are increasing. Moreover, quite a good level of wage and education intergenerational mobility is observed but future risks of decreasing mobility exist. This is notably due to the growing role of parents' wages in the financing of education which could in turn influence the child's social status. The deep changes occuring in the income, education and health dimensions combined with the analysis of the consequences of reforms leads me to describe the evolution of inequalities in these three dimensions considering them independantly (unidimensional analysis) or aggregately in a single indicator (multidimensional analysis).

Since Rawls (1971) and Sen (1973), the concept of inequalities of well-being has become central when talking about inequalities between individuals (Atkinson and Bourguignon (1982), Quadrado, Heijman, and Folmer (2001)). Wellbeing does not only encompass the income dimension, which strongly contributes to wellbeing, but also takes into account more diversified elements. Consequently, trying to analyse wellbeing inequalities -and therefore multidimensional inequalities- leads to two major issues: which elements constitute wellbeing and how to quantify wellbeing inequalities?

Concerning the first point, income is part of the description of wellbeing as it is a central element which influences a person's quality of life. However, other dimensions such as education, health, the institutional context or environmental quality can bring about wellbeing by themselves even if they are linked to income. Depending on the level of analysis considered, taking individuals, households or regions for example, the dimensions which can be included when studying social inequalities differ. For instance, Quadrado, Heijman, and Folmer (2001), analysing inequalities of wellbeing between Hungarian regions, use variables determined at the regional level and relative to education (enrollment rate) or health (number of hospital beds) dimensions. However, when measuring education or health assets at the household level, other kinds of variables must be used (father's educational attainment, working hours lost due to illness, etc.).

I focus on the individual level of wellbeing to study multidimensional inequalities in China. Doing so, I will need to make hypotheses relative to the income available for each individual included in the household which is not necessarily the individual's income given that reallocations occur inside the household. The other dimensions I will include in the social inequalities analysis are the individual's educational level as well as his health status.

Next to the choice of dimensions which influence well-being inequalities, the second difficulty concerns the quantification of multidimensional inequalities, when one wants to describe inequalities in dimensions which are either monetary (e.g. income) or not (e.g. education). The question of aggregation arises, and with it the question of the weighting scheme. Moreover, using an aggregate indicator does not allow me to identify the sources of evolution of multidimensional inequalities. For instance, an increase of social inequalities, such as a deterioration of the situation in terms of social equality among the population, can come from a strong increase in one dimension's inequalities whereas the other dimensions are slowly decreasing. The multidimensional analysis consequently needs a complementary analysis which describes the sources of the evolutions noted using multidimensional indicators. As such, the commonly used unidimensional analysis of each dimension's inequalities is still necessary.

Nevertheless, the big advantage of using aggregate/composite indicators of wellbeing inequalities is that it gives a synthesis of the evolution of these inequalities. For policy makers, a single multidimensional indicator of wellbeing inequalities leads to an enlarged vision of the real social situation in their country without focusing only on an income deprivation approach. Moreover, an individual who experiences deprivation in one dimension (e.g. income) could also suffer from a shortage in other dimensions (low level of education or poor health). On the contrary, a person with higher values in one dimension could also have a higher level in other attributes.[1] Consequently, when talking about wellbeing inequalities, and quantifying them, it is important to take the correlation between dimensions into account. This is not allowed by the simple unidimensional analysis.

Studying the evolution of social inequalities in China, I will carry out both unidimensional and multdimensional analysis of inequalities. In the first section, I will present a summary of the reasons which suggest that social inequalities in China are meant to change due to the reforms cited in the two preceding chapters. I will then present in section 2 the method I will use to measure unidimensional as well as multidimensional social inequalities. Finally, section 3 will describe the evolution of social inequalities in China. In order to emphasize the impact of the reforms on the evolution of unidimensional as well as multidimensional inequalities, I will differentiate the results between cohorts of individuals which have been to differently affected by the reforms. These reforms have different effects on the dimensions considered and were launched at a different periods. Section 4 will conclude on the evolution of social inequalities in China given the indi-

[1] The words "dimensions" or "attributes" are synonymous when studying multidimensional inequalities.

cators used and given the values of the central parameter influencing multidimensional inequalities quantification.

3.1 Three dimensions of inequalities in China: income, education and health

To assess whether the study of multidimensional inequalities in China is relevant, I will show firstly that important differences exist between individuals in terms of their income and wages, their educational level and their health status. These gaps are mostly the result of changes in Chinese society, following the reform movement.

3.1.1 Wage and income inequalities

Up to now, I have paid attention to the evolution of the labour market and specifically to changes in wage structure. This was done for two principal reasons. The first being that I was studying returns to human capital using wage equations. The second was the quality of data once we consider total income, including agricultural incomes. However, this choice appears questionable when studying inequalities. As wage are distributed at the individual level, studying wage inequalities at the individual level would consequently be based on the strong hypothesis of no intra-household reallocation: the wage earned by the individual earner in the household would be his available income. This is doubtful when we consider the empirical literature developed on the issue of intra-household bargaining and income reallocation (Lundberg and Pollak (1996), McElroy (1990), Thomas (1990)). Consequently, despite the data limitations in terms of quality relative to total income, I will take the intra-household reallocation into account considering this income, as being at the household level. However, in order to pursue the analysis of the evolution of the labour market and the use of the best quality data, I will still consider the wage inequalities.

Khan and Riskin (1998) show that wages, which represent an increasingly important fraction of revenue, are a factor that increases inequality. I will consequently sum up the principal determinants of wages in this country which in turn will influence the level of hourly or yearly wage inequalities in China.[2] Wage inequalities in China evolve due to factors which influence what can be called "justified" and "unjustified" wage inequalities. This distinction, made by Oaxaca (1973), is central for the Chinese context. The country was until recently in a situation where the wage structure was deeply egalitarian. However, disequalizing forces occured when a new labour market developed. These forces can either be "justified", when it concerns returns to higher efficiency (due to a higher level of education for example), or "unjustified", when linked to discriminations for example.[3] Both phenomena have occured since the launching of the reform movement, consequently leading to higher inequalities.[4] Concerning the former, returns to human capital are increasing since the reform movement. Moreover, bonuses, distributed to the most efficient workers, also reinforce this situation where the most efficient workers receive higher payments (Cao (2004), Zhang, Zhao, Park, and Song (2005)).

Concerning the "unjustified" type of inequalities, discriminations in the Chinese labour market concern two groups. One notes gender discrimination penalising women both in terms of the level of hourly wage and in terms of the number of hours worked, influencing the yearly wage (Chu Ng (2007), Li (2003), Maurer-Fazio, Rawski, and Wei (1999)). Differences between native urban dwellers and migrants from the countryside

[2] Sections 1.1 and 1.3.1 of chapter 1 presented respectively the determinants of wages in China and the evolution of returns to human capital in China.

[3] The distinction between "justified" and "unjustified" inequalities is particularly expedient for the Chinese case. Leaving an egalitarian wage structure and creating a labor market increase necessarilly inequalities. However, the type of inequalities which appear are not always negative for individuals. For instance, being more paid for a higher job or a higher level of education can consitute a incentive for working harder, making innovations etc. In turn, these elements can have a positive impact on the individual but also on growth as a whole. Even if I have chosen a microeconomic approach to the Chinese problematics, linking these last to a more global economic point of view is direct. In turn, a higher growth through redistribution plays a role in the individual well-being. The study of inequalities, even if investigated at the micro level, is closely related to macro issues.

[4] This distinction is made to clarify the different types of disequalizing forces which contribute to higher wages inequalities in China. I will however not be able to distinguish between "justified" and "unjustified" inequalities when pursuing the quantification of unidimensional and multidimensional inequalities.

are also important, linked amongst others to the hukou system (Liu (2005), Maurer-Fazio and Ngan (2002), Meng and Zhang (2001)). Urban dwellers are often privileged in terms of job type (they are hired for better qualified and consequently better paid jobs) and in terms of finding new employment in the event of loss of jobs. Finally, inequalities of treatment between the public, private and foreign enterprise sectors exist. Chen, Demurger, and Fournier (2005) show that the best paid workers are employed by foreign enterprises, while urban collective enterprises pay the lowest wages.

Broadening the analysis to total income inequalities in China, two major divisions are to be seen: between urban and rural areas, and between provinces or regions (see Gustafsson and Shi (2002) for a survey of these studies). The gap between urban and rural zones is firstly due to the differences in the type of jobs in the two areas: we found wage earners with a broader and more diversified job in urban areas whereas only Township Village Enterprises offer paid jobs in rural areas where self employment in household farms is the only other source of income. In this latter case, within rural income inequalities exist as the wages for non-farm jobs are higher than income earned from agricultural activities (Tsui (1998), Peng (1999)). The Household Responsibility System has contributed to the growth of agricultural output and income but the shortage of farmland as well as the high number of peasants have negated against the positive evolution of agricultural incomes. Another factor influencing income inequalities between urban and rural areas seems to be inequalities of access to school and higher education. This leads to better educated individuals, who will earn better wages being concentrated in urban zones (Sicular, Ximing, Gustafsson, and Shi (2007)).

Concerning inequalities between provinces and regions, policies which have favoured the eastern and coastal geographic zones generated differenciations in terms of the type of enterprises present in China's provinces. Jones, Li, and Owen (2003) demonstrate that coastal cities experience about a 3% higher annual growth rate than inland cities. This contributes to inequalities in economic development between regions which can lead to a "two-speed" system (Renard (2002)).

To sum up, wages and income inequalities tend to have increased since the movement of reforms due to the modifications of the labour market and to geographical disparities.

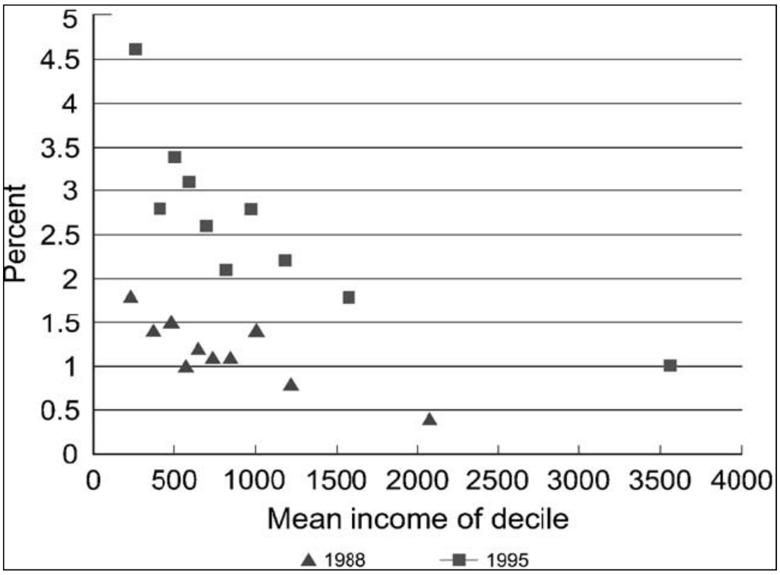
3.1.2 Inequalities in education

Determinants of education inequalities in China have been presented in Chapter 2 when studying factors influencing educational mobility. I present a brief summary of the principal factors in this subsection.

One of the major goals of Mao's policy was to develop an educational system which was accessible to every one and which was egalitarian. Hannum (1999) describes it as a socialist egalitarian model. Some authors suggest that the Cultural Revolution, which had been a catastrophe for education (particularly for secondary and university education) had been a means for poor rural populations to gain access to primary education and even to litteracy (Han (2001)). In reality and even today, the educational system is far from being egalitarian. The reforms have contributed to this through different channels.

In May 1986, a wave of reforms[5] gave total responsibility for primary education to local government (Connelly and Zheng (2003), Law (2002)). This decentralization of decision making and financing created distortions in the provision of education between the different provinces. Even if the aim of 9 years of compulsory school was set by the government, only provinces that had negotiated higher budgetary allocations could attain the objective of providing good educational services with competent teachers with an adequate level of training. For poor provinces, conditions in schools were often precarious. Teachers left such provinces to earn more in richer ones. As a result, access to schooling and to good teachers is highly heterogeneous across the population. This is confirmed

[5] Law (2002) gives six national education laws: Regulation on degrees (1980), Basic education law (1986), Teachers law (1993), Education law (1995), Vocational Education law (1996), Higher education law (1998). Reforms undertaken after 1990 will have an impact on educational level of cohorts younger than those considered in this study. But we will see the impact of the most important one, the decentralization of decision making and financial constraints.



Source: Gustafsson and Li (2004). Figure 2, p. 297.
Disposable income is yearly income expressed in 1998 prices.

Figure 3-1: Education expenditure in rural China (1988 and 1995) as a percentage of disposable income

by Liu (2001) who shows that a part of the inequality in wages stems from inequality in training. A 1997 World Bank report (Hossain (1997)) underlines that the government is aware of this situation. Nevertheless, inequality in education continues to increase (Zhang and Kanbur (2005)).

Households, even if they are aware of the importance of education in the growing economy of China, must deal with the need for higher levels of educational expenditures for their children.[6] Consequently, as demonstrated by Gustafsson and Shi (2004) and Fang, Zhang, and Fan (2002) the proportion of family budgets allocated for expenditures on education has increased. Moreover, this rise is a decreasing function of household income (Figure 3.1.2), resulting in a tigher budget constraint for poorer households.

Households face a new trade-off between sending their children to school, paying for

[6] Liu (2004) quotes a household member talking in relation to schooling: "Sending the child to primary school at a certain age has become a practice that has been internalized in the Chinese people's mind". Moreover, I recall that the One Child Policy can have a positive impact on the investment level of the families in their child's education (see chapter 2).

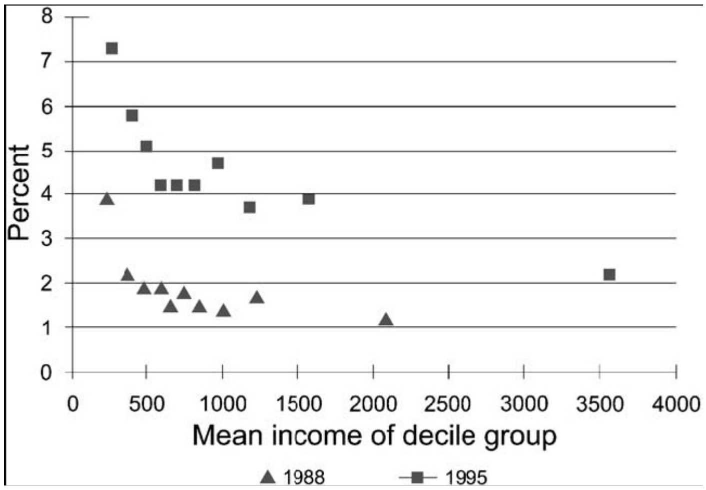
their schooling, or keeping them at home, which can be an added resource when they work on the farm or at home.

Moreover, the dissolution of collective farms gave a new role to the family, especially in rural areas. Since household members have to work more to reach a given level of income, priority is not always given to education. Even if the law states that families must send their children to school for 9 years, Liu (2004) shows that the decision about schooling is, fundamentally, the family's own private decision. Thus, both the supply of and the demand for education has been modified by the reform movement that began at the end of the 1970's.

3.1.3 Inequalities in health status

Inequalities in health have been, as for education and income, highly influenced by the reform process. During the Mao Tse Dong regime, 85 % of villages had a health station staffed by barefoot doctors who provided basic preventive and curative care. Townships had the responsibility for health centers, and counties for county health bureaus (Bloom and Xingyuan (1997)).While the level of health improved during this period, the reform process and the transition to a market economy has had consequences on the health system, especially in terms of its financing (Worldbank (1997)).

Bloom and Xingyuan (1997) identify two major consequences of the reform process. Firstly, the rise in health care costs, principally linked to the decentralization of health services and financing, lowers the opportunity for the poorer to have access to health services. The proportion of health expenditure directly born by the central government declined from 38.2 % in 1985 to 26.6 % in 1990 (Ho (1995)). Primary health care financing is in the hands of the provinces whereas the government encourages more and more hospitals to be financially independent. Users must increasingly pay for medical services. As with education, the proportion of health care expenditure in the household budget rises (see Figure 3-2 and Gustafsson and Shi (2004)) which again results in higher



Source: Gustafsson and Li (2004). Figure 2, p. 297.
Disposable income is yearly income expressed in 1998 prices.

Figure 3-2: Health expenditure in rural China (1988 and 1995) as a percentage of disposable income

percentages for the poorer.

Secondly, disparities between urban and rural areas also emerge in terms of health care delivery. In rural areas, before the dissolution of collective farms, people benefited from cooperative medical care schemes. 80% to 90 % of the rural population were covered in 1979. In 1995, only 20 % continued to benefit from this system (Ho (1995)).

Combined with the increase in the costs of health services due to the reforms, the traditional insurance system collapsed, leading to an increasingly important role for consumer insurance costs.[7] Alongside public insurance, individuals can purchase insurance, but this is essentially restricted to urban areas. Nevertheless, Liu, Zhao, Cai, Yamada, and Yamada (2002) and Henderson, Akin, Hutchinson, Jin, Wang, Dietrich, and Mao (1998) show that there is no evidence of a decrease in access to health care services and that the new insurance model is more equitable in ensuring access to primary care for out-patients.

[7]The traditional scheme is composed of a publicly-funded medical care system which covers civil servants, workers in public agencies, universities; the labor insurance medical care which protects employees in state and collective enterprises and their immediate family members; and the cooperative medical care scheme for rural areas. All these structures increasingly ask their clients to pay for their care.

3.2 How to measure social inequalities

3.2.1 Unidimensional inequality measurement: the Theil index

To measure unidimensional inequalities, many tools can be used. Inequality tends to describe the distribution of revenue (usually) among the population. So measures such as variance or standard deviation seem initially to be good instruments to describe this phenomenon. However, those two indicators are not desirable because they don't fulfil some conditions required to rank inequalities, i.e. to compare different levels.[8] Various indicators have been built from established rules (see Blackorby, Bossert, and Donaldson (1999) for a detailed review). To describe unidimensional inequalities in our sample for the three dimensions (wages, education and health), I will use the Theil index, on which, we will see later, is also based the multidimensional inequalities indicator.

The Theil index is obtained from the generalised entropy class of measures. If we consider N individuals where the i-th has an income y_i, i running from 1 to N, the general situation is given by the following formula :

$$GE(\alpha) = \frac{1}{\alpha^2 - \alpha}\left[\frac{1}{N}\sum_{i=1}^{n}\left(\frac{y_i}{\bar{y}}\right)^{\alpha} - 1\right],\ \forall\, y_i > 0$$

where $\bar{y}$ is the mean of income for the whole sample.

The parameter α represents the weight given to distances between incomes at different parts of the income distribution, and can take any real value. For small (large) values of α, GE(α) is more sensitive to changes in low levels of income. I will use the usual index called Theil's measure given for a value of 1 for α that gives :

[8] For example, variance is not independent of the income scale. If we double all incomes , we would obtain a quadrupling of the estimate of inequality. However, this axiom of scale invariance has been put into perspective by different studies which stress the possible preference of some policy makers for another type of invariance (Kolm (1976), Bresson and Labar (2007)). The translation invariance postulate that inequalities do not change when we add a constant to the income of all individuals (Kolm (1976)). I will however use indicators which correspond to the most broadly accepted axioms.

$$GE(1) = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\bar{y}} \log \frac{y_i}{\bar{y}} \ , \ \forall \ y_i > 0.$$

This indicator can be decomposed into two parts. One describes the inequality between groups, and the other the inequality within groups. Decomposition of income inequalities in China has been carried out for differences between urban and rural population, and for the different provinces (Bhalla, Shujie, and Zongyi (2003), Gustafsson and Shi (2002)). I will use the decomposition property of the Theil index where relevant to complete the analysis.

3.2.2 Maasoumi's multidimensional inequality indicators

The analysis of multidimensional inequalities can be carried out in two different ways: using multidimensional composite indicators or doing a graphical study using multidimensional stochastic dominance. I have chosen in this thesis to concentrate on the former methodology. Given the political objectives of using such multidimensional descriptions of social inequalities and given the time dimension of the study, I focus on indicators which can bring complete ordering of multidimensional inequalities. Moreover, the stochastic dominance does not give information on the level of inequalities. Studying the time evolution of the level of inequalities in China, this method is consequently of no use.

As a background to the construction of multidimensional inequalities, a set of axioms have been derived from the unidimensional situation and some are added in order to take into account the new issues linked to the multidimensional context.[9] Two major problems occur when we examine multidimensional inequality. The first is ordering distributions containing various attributes. One has to find criteria which allow us to indicate the dominance of one distribution over another (Muller and Trannoy (2003), Savaglio (2001)). The second comes from the fact that attributes can interact, especially

[9] For a detailed presentation of the various axioms underpinning multidimensional inequalities, see Weymark (2004) or Tsui (1999).

income with non-income dimensions (this refers to the axiom of Correlation Increasing Majorization (CIM), Tsui (1999), Savaglio (2001)). Indicators built until now, to my knowledge, cope with the latter problem but not systematically with the former. The indicator used in my thesis, Maasoumi's, respects the basic rules for the construction of multidimensional indicators but not the CIM criteria. However, as we will see later, Maasoumi's indicator will deal with the issue of correlation between dimensions. It is important to stress that, as far as I am aware, no published empirical articles exist on multidimensional inequalities. Some unpublished papers also use Maasoumi's indicator, often highlighting its limitations (Justino (2005), Nilsson (2007)). Discussions with these authors lead me to support their choice given the doubts attributed to other kinds of indicators which, at first glance, seem more rigourous regarding the axiomatic rules, but do not seem applicable or coherent.

Moreover, before presenting the construction of Maasoumi's indicator, I must stress the necessity of normalising attributes in order for them to have the same unit of measurement. If I were not to do that, aggregation of dimensions as well as concepts of compensation between them would be meaningless. I will consequently follow the normalisation rule used to build the Human Development Index which leads to a range of values between 0 and 1. Dealing with the unidimensional case, I encounter a vector of income values for n individuals indexed by i. For the multidimensional case, we have a set of matrices whose columns x_j represents attributes j and whose line i represents individual i. If we have k attributes (in this case, $k = 3$), and n individuals, the size of our matrix is $n \times k$.Considering the attributes x_{ij} before normalization, the scaling rule is given by:

$$X_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}} \text{ with } i \in [1, n] \text{ and } j \in [1, k].$$

Given this normalization, attributes X_{ij} will enter in the aggregation process and in the construction of the Maasoumi indicator.

The Maasoumi (1986) indicator is built following a two stage procedure. At first, he uses a welfare aggregation function (S) in order to aggregate the dimensions (also called attributes or arguments) considered. This consists in a weighted sum of attributes and an adjustment due to the covariation of attributes. The second step applies Generalized Entropy (GE) inequality measures to the univariate distribution obtained from the first step.

For the first stage of the indicator construction, the welfare function used can consequently be defined as

$$S_i = S(X_{ij}), i \in [1, n] \text{ and } j \in [1, 3] \tag{3.1}$$

To obtain the "optimal" aggregation function, Maasoumi (1986) minimizes the following multivariate generalization of GE measures, such that $\sum S_i = 1$:

$$\begin{aligned} D_\beta(S, X, w) &= \sum_{j=1}^{j=k} w_j \left\{ \sum_{i=1}^{n} S_i \left[\left(\frac{S_i}{X_{ij}} \right)^\beta - 1 \right] / \beta\,(\beta + 1) \right\}, \beta \neq 0, -1 \qquad (3.2) \\ D_0(S, X, w) &= \sum_{j=1}^{j=k} w_j \left[\sum_{i=1}^{n} S_i \log \left(\frac{S_i}{X_{ij}} \right) \right], \beta = 0 \\ D_{-1}(S, X, w) &= \sum_{j=1}^{j=k} w_j \left[\sum_{i=1}^{n} X_{ij} \log \left(\frac{X_{ij}}{S_i} \right) \right], \beta = -1 \end{aligned}$$

where w_{ij} are the weights given to each attribute and β is the parameter which characterizes the degree of substitution between the dimensions. I will develop the role played by this parameter later.

The "optimal" aggregation function obtained from these measures are defined as:

$$S_i \approx \left\{ \begin{array}{c} \left(\sum_{j=1}^{k} w_j X_{ij}^{-\beta} \right)^{-1/\beta}, \beta \neq 0, -1 \\ \prod_j X_{ij}^{w_j}, \beta \to 0 \\ \sum_{j=1}^{k} w_j X_{ij}, \beta \to -1 \end{array} \right\} \tag{3.3}$$

We see from equation 3.3 that the parameter defining the degree of substitution between attributes, i.e. β, plays a central role in the welfare function determination.

When this parameter increases, there is less substitution between attributes.

When $\beta \to -1$, this corresponds to total substitution between attibutes which means that the population is indifferent to which dimension is privileged. In this case, low levels in one dimension can be fully compensated by high levels in other dimensions. For instance, if individuals have a high level of income, suffering from poor health or having a low level of education is compensated by the income dimension. We can expect, in this situation, lower levels of multidimensional inequalities compared to the results obtained when lower degrees of substituability are considered given the permitted compensation between attributes.

On the contrary, for the limit when $\beta \to \infty$ the included dimensions are assumed to be perfect complements, so there is no possible substitution between attributes. The welfare function corresponds therefore to the Rawlsian approach which favours total equality between dimensions implying that individuals are affected just as much by having a high level of income, as having a good level of education or being in good health. In this situation, S_i will mirror the attribute with the lowest value for every statistical unit, i.e. S_i will consequently mirror the worst performer in terms of attributes. This interpretation in terms of best or worst performer is permitted by the normalisation made in order to have the same unit of measurement for each attribute.

Finally, concerning the social welfare function for $\beta = 0$, I will not consider this situation as the dimensions considered take the value of 0, leading also to a value of 0 for social welfare. Even if it were possible to use the weights given to each dimension, w_j, to see how preferences relative to the various attributes can affect the level of multidimensional inequalities, I consider that the parameter β, which characterizes the degree of substitution between dimensions, manages to scale the preferences itself. Consequently, I will give equal weight, equal to 1, to all dimensions, and I will use β to examine the impact of different preferences on the level of social inequalities in China.

The second step consists in using the common unidimensional inequalities function.

In particular, Maasoumi uses the Theil functional form to define his M_0 indicator. To do so, he defines d_i as the share of population corresponding to the $i-th$ unit, which is usually equal to $1/n$, and $S_i^* = \frac{S_i}{\sum\limits_i S_i}$. Shown in parallel with the Theil index, the indicator is:

$$M_0(S) = \sum_{i=1}^{n} S_i^* \log\left(\frac{S_i^*}{d_i}\right)$$

3.3 Empirical implementation

3.3.1 Selecting the relevant sample and variables

I use the CHNS database to study social inequalities in China. For a broad description of the database, please refer to the section 1.2.4 of Chapter 1.

Dimension measurement: variables of wages, income, education and health

Studying education, health and income inequalities requires some conceptual explainations. For education and health, the major issue concerns the comparability of educational levels and health status between individuals. Subsequently, I will focus my attention on individuals who have completed their education or finished growing and who consequently will be comparable regarding these two dimensions. Of course, these individuals can also improve their qualifications by engaging in on the job training. However, I will not be able to take this into account in the measurement of education. The inequalities I will measure will consequently be relative to the central schooling system from primary school to university.

As underlined in sub-section 3.1.1, I will consider both wages and income to describe income inequalities in China. In order to follow a step by step analysis, I will at first look at the hourly wage, which is at the individual level. Once again, I specify that wages included in the database are from non agricultural activities. It does not include wages earned from work on collective farms, as these wages are not registered (or registered correctly) in the survey. Inequalities in hourly wages will cover differences in human capital and their impact on the wage level, but it will also cover wage discrimination. Next, I will move to the yearly wage which is the information initially given by the CHNS database (the hourly wage being obtained using information on hours worked per week, weeks worked per month and months worked per year). Using this measurement of income, another dimension can be a source of inequality: the quantity of hours worked can be chosen by individuals but this choice is often a function of the individual's level of poverty (the poorer being forced to work more as they are paid less).

The comparison between hourly wage and yearly wage inequalities is possible due to the fact that we consider the same sample. The differences of inequalities which will be noted will be based on inequalities in the quantity of hours worked by individuals.

The next step consists of considering the total income from paid activities but also from self-employment, distributed among individuals of the household. The use of total income is made in order to consider intra-household allocations. The allocation rule involves equal sharing of income between household members, but this does not control for bargaining effects inside the family. Keeping in mind the limits of the allocation rule chosen, the income obtained from the aggregation of all kinds of income in the household and divided between household members can be considered as individual available income. I use the methodology suggested by Deaton (1997) to take into account potential economies of scale in the household. I will divide total household income by n^a, with n being the number of household members and a being the equivalence factor. Following Wan and Zhang (2006), I will use a value of $a = 0.8$.

Moreover, given that we want to compare the situation between taking into account intra-household allocations or not, we base our analysis on the sample of wage earners, which is the smallest in terms of the number of observations. Individuals considered for our study will consequently be wage earners living in households where there can be other kinds of income added to wages. These can be earned by the wage earner, having a secondary occupation (helping in agricultural activities or business etc.) or by other members of the household who are not wage earners and who are consequently not included in the sample. Differences in inequalities obtained from the different kinds of income considered will consequently be related only to the fact that individual available income is not given only by the individual wage but can also comes from incomes of other members of the household.

Educational levels are measured by the individual number of years of schooling. This measure is the one which has been considered in the preceding chapters.

In order to consider a health indicator which gives a global picture of health status, I use the Body Mass Index (BMI) as my measure of health status. As defined in Chapter 1, BMI (equal to the ratio of weight in kg to squared height in meters) reflects short-term variations in nutritional and health status, since it depends on both past and present investments in health. To have a more detailed presentation of the range of BMI, please refer to section 1.2.4 of chapter 1. This variable is continuous. It is comparable between individuals as soon as we consider individuals who have finished growing. This indicator mirrors to some extent individual characteristics (mostly through health, a component which depends in part on genetic elements) but also access to health services or to nutritional food (through the weight component).

Unequal exposure to the reform process: cohort sub-samples Our basic sample consists of individuals older than 16 years of age, whose anthropometric measurements fall within the bounds of what is considered reasonable by nutritionists in the Chinese context. For this sample, individuals have completed their schooling, which allows us to study educational inequalities.

China is an interesting country in which to study multidimensional inequalities because of its reform process. Nevertheless, reforms do not have an impact on the whole population we consider here, especially when education and health are concerned, because of the inertia of the political process.

In terms of education, reforms began with an impact on children aged between 6 to 15 year of age in 1980. Children who were at school earlier were not affected by reforms in the educational system.

In terms of health, given that I use BMI at adulthood to describe health status, I will capture different elements of the evolution of health in the Chinese population. Since the BMI is a medium term indicator of nutrition, we will see the impact of the nutritional transition on adult diet through its impact on weight, and whatever the age of individuals considered in the full sample (from 16 to 73 years of age). The BMI reflects two other phenomena. The first is child care provided in early childhood (0 to 5 years of age). The second is the nutrition of children between 6 and 13 years of age which has been shown to have an impact on the nutritional status of adults, and particularly their BMI (Wang, Ge, and Popkin (2000)). For the two last categories, all the sample population was not affected by reforms, initiated in 1980, which had consequences on both the child care system and nutrition during childhood. This is displayed in Table 3.1.[10]

We see that if I wish to examine the whole population affected in 2000 by reforms concerning child care during early childhood, I have to consider individuals who were between 16 and 25 in 2000 (sample 1).[11] For the impact of the nutrition transition on children, I must take individuals between 26 and 33 years of age in 2000 (sample 2). Finally, individuals between 26 and 35 years of age in 2000 are individuals affected by education reforms. This constitutes the sample 3.

I will therefore study unidimensional inequalities first, followed by multidimensional inequalities, on three separate samples. The first concerns all the individuals in the sample (aged between 16 and 75). The second will be individuals between 16 and 25 for each year (sample 1).[12] And the third will be individuals between 26 and 33 for each year

[10] What we call "targeted cohort" describes the cohort for which all individuals were touched by the reforms in 2000.

[11] I choose to fix the samples refering to the year 2000 in order to have two years of observation for which the whole sample is potentially affected by the reform process.

[12] The precise definition of the age cut off is 26 years of age. Given that the age is determined consid-

(sample 2). A percentage of the population considered each year will be in the category of individuals concerned by the reforms (see Table 3.1), considering that 100 % of the population was affected in 2000. In order not to overcharge the results, I do not give results for the sample 3. However, the sample 2 will cover individuals who are affected both by the reforms relative to the nutrition transition and by education reforms.

Correlation between dimensions

When considering multidimensional analysis of inequalities, it is important to closely examine correlations between dimensions. As stressed before, some individuals can experience combined high levels of deprivation in more than one attribute. Consequently, I present in Table 3.2 and 3.3 Pearson and Spearman correlations between the five variables considered in my study for the whole sample.

What interests me are the correlations between each income dimension and education and health, and the correlation between health and education. However, before commenting, these correlations, I would like to underline the very high correlation between hourly and yearly income (table 3.2). This is logical as the only difference between the two measures are the number of hours worked. It seems that this last element does not make a strong difference. We will see what will happen once I consider unidimensional inequalities of hourly and yearly wages. On the other hand, the correlations between wages and total income are interesting, since between 1991 and 1993 or between 1993 and 1997 (depending on which correlation is used) a huge increase occured. Between these periods, correlations have been multiplied by at least 4. This is in accordance with the increasing share represented by wages in total household income.

Concerning correlations between dimensions, interesting phenomena are to be noted (table 3.3). Firstly, the correlation between different kinds of income and BMI is always

ering the number of months, this variable takes values between 25 and 26 years of age. Consequently, the youngest sample will include individuals less than 26 years old (<26) whereas the middle age sample will include individuals aged 26 or older (>=26) and less or equal to 33 years old.

positive and significant, even if they are not high. As these are only correlations, it does not aim to identify the causality between health and income. Secondly, the correlation between income and education becomes positive and significant from 1997. Since then, it has continued to increase strongly. Given the results enounced in Chapter 1, we see a coherence between these correlations and the conclusion of increasing returns to human capital in China. Finally, the correlation between education and health is mostly negative but not very significant.

As specified earlier, I distinguish two sub-samples which are chosen to mirror the relatively high impact of the reforms on individuals, given their age. As for inequalities, which will be presented later, correlations between dimensions can be different given the cohorts considered. Tables 3.4 and 3.5 give the correlations between incomes for the two sub-samples. Once again, we note the increasing role of wages in total income as the correlations between wages and total income increase strongly between 1991 and 1993 or 1993 and 1997 and continue to grow thereafter.

If conclusions relative to different kinds of income are quite similar between samples, it is not the case when considering correlations between incomes, education and health. Whereas education became positively and significantly correlated with all types of income from 1997 for the whole sample, we note a clear cut off regarding correlations between education and wages in 1997 and 2000 for the two sub-samples. These correlations are insignificant for both sub-samples. However, the correlations between education and total income remain positive and significant for all periods for individuals aged less than 33 (except in 1993 for the 26-33 year old sample). Given the definition of total income, which is household income divided among household members, and the definition of our samples, which only include wages earners, this tends to support the fact that wage earners living in richer households are also better educated. Once again, as we consider correlations, we cannot say if households are richer thanks to better educated individuals or if better educated individuals have a higher level of education thanks to a higher total household income. However, given that the correlations between education and wages are not significant between 1997 and 2000, it seems that education is not linked to wages during these periods, which can seem surprising given that the younger individuals entered the labour market after the reforms which have encouraged higher returns to human capital in the determination of wages. Education becomes nevertheless significantly correlated with all kinds of income for both sub-samples in 2004.

Concerning correlations between the different kinds of income and health, conclusions for the two sub-samples differ significantly from those drawn for the whole sample. These correlations are significant for the sub-sample of individuals aged between 26 and 33 years of age only in the year 1993. Moreover, these correlations are negative which is surprising. For the youngest sub-sample, the correlations are only significant for the year 2004 but are positive for this period. Given that correlations between BMI and the various incomes were significant for the whole sample for all periods, I highlight a higher correlation between health (measured by BMI) and incomes for older individuals. Given that older Chinese people work until an advanced age, or receive meagre retirement pensions, this higher correlation is more intuitive.

Finally, as for the whole sample, in general, BMI and education dimensions are not significantly correlated even if they are positively and significantly correlated for those who are the less than 26 years old for the year 2004.

Given the analysis of correlations carried out above, taking into account correlations between dimensions as well as different degrees of substitution between them is very important for the analysis of multidimensional inequalities.

Descriptive statistics

Descriptive statistics relative to the three samples defined in the preceding sub-section are given in table 3.8.

We can see that except for total income, the level of each attribute has grown between 1991 and 2004. The status of individuals is improving on average as regards income, education and health. Concerning total income, a surprising drop between 1991 and 1993 is observed. This could be due to the problem of coherence between survey rounds as described in Chapter one. After 1993, a slight decrease affects total income but it then increases regularly. The comparison of the three samples leads to some observations already made in Chapters 1 and 2. For example, average educational attainment is increasing whatever the sample considered. Moreover, a clear break appears between the whole sample and the two youngest samples, also translating the improvement in average schooling attainment in the Chinese population.

3.3.2 The evolution of unidimensional inequalities in China

In this section, interpretation of results will be based on figures describing the evolution of inequalities regarding the three dimensions considered independently. The precise numbers can be found in the tables presented in the Appendix.

Wages and income inequalities

As explained earlier, I will distinguish between three kinds of income to quantify income inequalities in China. This is done in order to take into account the potential intra-household income reallocations and to give the broadest picture of income inequalities. We must however keep in mind the caveat specified in the description of total income data relative to their quality.

Before drawing a comparison of the evolution of income inequalities between cohorts, for the different kinds of income, I will present a comparison of results obtained considering the different kinds of income, for the three cohorts.

Figure 3-3 present this latter comparison. For all the samples considered, the evolution as well as the level of hourly and yearly wages inequalities are the same, underlining no clear influence of the number of hours worked by individuals on the yearly wage income inequalities. Consequently, in the following paragraphs, both for the unidimensional inequalities of income and the multidimensional inequalities, I will focus on the results given by the total income and the yearly income as these two incomes are comparable in terms of the period of time for which they are defined. Moreover, these incomes are linked to different kinds of professional activities with respect total income, which is calculated at the household level, includes both wages and income from self employment of individuals considered in our sample (if they have a secondary occupation) or the income of other household members.

I want to now emphasize the differences in income inequalities noted between generations. To do so, I present the results regarding the Theil index. We find in Figure 3-4, the graphical comparison of inequalities of yearly wages and total income between the samples. Concerning yearly wages, I emphasize similar levels of inequality as well as evolutions of inequalities between cohorts. They are quite unstable given the years considered with the values of the Theil index varying between 0.18 and 0.35. Nevertheless, we note a trend of growing inequalities for the whole period with an increase of 74% of yearly wage inequality for the larger sample, 47% for the sample younger than 26 and over 100% for the sample aged between 26 and 33.[13] These increases are very high considering they are over a period of 13 years. For the larger sample, this evolution is in accordance with the results found by Dong (2005) who studied wage inequalities in China during the 1990's. This author stresses an increase of the Theil index of about 73% between 1994 and 2001. Fang, Zhang, and Fan (2002), using an expenditure based inequality approach in urban areas (containing both wage earners and self employed individuals), also emphasize growing inequalities between 1992 and 1998 in these areas.

On the other hand, it is interesting to note the differences between the cohorts considered. We note that the group which has suffered most from an increase in its yearly wage inequality is on average, for the whole period, the middle aged population, which was the first affected by the reform movement, whereas the youngest population, which entered the labour market mostly after the beginning of the reforms, experiences a slower increase. This can be interpreted as the consequences of the disequalizing force which

[13] These figures represent the percentage of increase between 1991 and 2004 and are calculated from the figures of tables 3.10, 3.11 and 3.12.

occured after the beginning of the reforms, induced by inequalities in the structure of wages linked to changes in the factors influencing their determination. The path toward a labour market seems to have most affected the cohorts who entered the labour market during the most important changes in this determination structure.

Total income inequalities show larger variations of the Theil index. I recall that the differences between yearly income and total income reside principally in the fact that total income includes agricultural incomes and this income is calculated from the household total income shared among the household members.[14] Consequently, what we measure here is inequalities in income between households.[15]

It seems that total income inequalities between households have dropped sharply between 1991 and 1997 but have increased since then. The huge decrease between 1991 and 1997 appears surprising and does not corroborate the results of previous studies on Chinese income inequalities. For example, Khan and Riskin (1998), studying total household income inequalities and making a decomposition between rural and urban incomes, emphasize an increase of about 18% in inequalities between 1988 and 1995. Agricultural income inequality rose by 42.5% and urban income inequality by 23%. However, Benjamin, Brandt, and Giles (2003), using the same measurement of inequality (i.e. the Gini coefficient) have noted only a 7% increase in inequalities in rural areas between 1987 and 1995. Nevertheless, the huge decrease displayed by in our samples appears questionable concerning the 1991-1997 period. This could principally be linked to the low quality of total income data underlined before. I recall the dramatic drop of the mean in total income between 1991 and 1993 stressed in the descriptive statistics. With the improvement of the surveys relative to economic information collected, results for 1997 to 2004 appear more reliable.

During this latter period, total income inequalities rise from 8.8% for the youngest

[14] Concerning the weights given to individuals, refer to section 3.3.1 which describes the intra-household income allocation rules chosen.

[15] As specified above, I chose to transform the household income into individual income in order to be at the individual level, i.e. the analysis level chosen for the study of educational and health inequalities.

cohort to 33.7% for the 26 to 33 years of age sample. The larger sample experiences an increase of its total income inequalities of about 46%. Considering that we are pooling together individuals from different provinces and from rural and urban areas, this increase confirms the conclusion drawn concerning the geographical and inter-household disparities following the reform movement. As in the results relative to yearly wages, the cohort which most suffers from income inequality increase is the middle aged one. This mirrors once more the impact of the reforms on the cohort which is most affected by them, confirming their negative impact on income inequality. However, this strong increase is not always supported by the literature. For instance, Sicular, Ximing, Gustafsson, and Shi (2007), using household disposable income, show no clear trend in the evolution of inequalities between 1995 and 2002. Given the quality of their data, issued from the Household Income Survey, more confidence can be given to their conclusions.

Inequalities of education and health

Much less attention has been paid to inequalities in health and education in China. As described in sub-sections 3.1.2 and 3.1.3, studies showing the disengagement of the state in the financing of both educational and health public services, combined with those showing the greater share of household expenditure dedicated to education and health, tend to support the potential increase in disparities in these two dimensions.

Figure 3-5 shows the evolution of educational inequalities. Whereas the larger sample tends to display decreasing educational inequalities, no clear evidence can be drawn for the two sub-samples of individual who are less than 33 years of age from this figure. Consequently, I add, for education inequalities, another graph including only the two sub-samples of younger cohorts. In Figure 3-6, we see that for the sub-sample of the youngest individuals, educational inequalities are also regularly decreasing. For the middle-aged sub-sample, unstable evolution occurs between 1991 and 1997 preceding the same decrease as seen with other samples thereafter.

These results are not in accordance with those found by Zhang and Kanbur (2005) who show an increase in illiteracy rate inequalities, supporting increasing inequalities regarding education. However, their measurement of education is different from mine as they focus on a specific type of education. Given that I put together information relative to all levels of education, my results convey explainatory factors relative both to the supply and the demand side. For instance, being literate can be an objective for a large share of the population, whatever its social origin or living place, whereas some individuals can choose not to attend college or university for various reasons. The results can be consequently driven by a relatively high level of equality inside two extreme sub-groups (at the very minimum) but with a cut-off between them.

To analyse deeper, we could take a closer look at the decomposition of inequalities, not between cohorts, but between individuals who have less and more than a set level of education. I tried different cutoffs given the Chinese educational system. The break which demonstrated an interesting pattern in terms of inequalities is most relevant for 12 years of education, which corresponds to the end of the middle school system. Given that schooling begins at 6 years of age indivuals, who pursue higher levels of education are aged more than 18 years of age. In order to investigate potential non linearities of inequalities as a function of the level of education, I make a decomposition of the Theil index in between and within group inequalities.[16] If the between group inequality share in total inequality increases with time, the schooling attainment of 12 years of education can represent a threshold below and above which within group inequality will have a decreasing influence on total inequalities.[17]

Table 3.9 shows interesting phenomena in terms of the evolution of educational in-

[16] As the Theil indicator is strictly decomposable, the addition of the two kinds of inequalities is equal to the total Theil indicator. The between groups inequalities are measured by, following the same notations as before:

$$T_b = \frac{1}{n} \sum_{j=1}^{k} \frac{n_j}{n} \left(\frac{y_j}{\overline{y}} \right) \log \left(\frac{y_j}{\overline{y}} \right)$$

with j being the sub-samples considered, and n_j the population of the $j-th$ sub-sample.

[17] We check different levels of thresholds for this break. This is the one which emphasizes the most relevant information.

equalities. For the whole sample, whereas within groups inequalities dominated the level of inequalities between 1991 and 2000, between groups inequalities are superior for the year 2004. Furthermore, looking at the two youngest sub-samples, we see that since 2000, between groups inequalities are predominant and represent 70% of the total level of inequality in 2004. This leads us to think that a break in the educational system occured following the reform movement in China. The highest level of education seems to be less accessible for certain groups. This is in accordance with the evolution of the financing of the educational system and the geographical distribution of the different kinds of schooling infrastructure (Hannum and Wang (2006)). This can also be linked to the Vocational Education Law, established in 1996 which has promoted vocational education which exists principally for the years of schooling before the upper schooling years (i.e. until 12 years of schooling identified as the significant break in the evolution of between group inequalities share in total inequalities). This argument is supported by the strong increase in the share of between groups inequalities between 1997 and 2000 for the youngest cohort which is directly affected by this reform given its age range.

Given the preceding conlusions, the decrease of total inequalities can be driven by decreasing levels of within groups inequalities which does not imply that there are none in the Chinese educational system.

Concerning health inequalities (Figure 3-5), all the samples experience a huge increase in BMI inequalities with a rise of 48%, 176% and 95% for respectively the larger, the youngest and the middle-aged samples. The two latter samples are the ones which have been most affected by the reforms and which demonstrate a dramatic increase in their BMI inequalities. I recall that BMI can be translated as nutritional status. As presented earlier, the youngest cohort may have suffered from the reforms relative to the disengagement of the State in the financing of health services, during their early childhood, influencing their whole life health status. These results show a parallel with the results of Zhang and Kanbur (2005) relative to the inequalities existing regarding the availability of health services (the number of health care personnel or hospital beds).

These authors emphasize an increase in inequalities mostly due to the gap between urban and rural areas.

3.3.3 Multidimensional inequalities in China: results considering sensitivity to a subjective parameter

I use Maasoumi's indicator to measure multidimensional inequalities in China for the three dimensions of income, education and health, considering $\beta = -1$ and $\beta \in [0.1, 10]$. Given that β can take values between -1 and ∞, I used values of β between 0.1 and 1000 to fully analyse the impact of this parameter on the results. However, given that once dimensions are normalised, their values are between 0 and 1. Values of β which are too high lead subsequently to high values which STATA has problems dealing with. Consequently, I restrict the value of β to lie between 0.1 and 10 which allows for a relatively good variability of the substitution parameter. In order to simplify the presentation, I will present the results for four values of β: -1, 0.1, 1 and 10.

In the preceding subsection, I emphasized different evolutions of inequalities between the dimensions I have chosen to describe social inequalities. As stressed in this chapter's introduction, the unidimensional analysis of social inequalities gives independent views regarding the different dimensions. If we want to obtain a broader picture of social inequalities, it may be useful to aggregate the dimensions. The indicator described earlier allows me to do this. Results relative to Maasoumi's indicator are presented below.

I will focus on three major elements for the analysis of multidimensional inequalities. Firstly, the range of the values obtained which, as I specified in the definition of the Maasoumi indicator, is expected to be a function of the value of the substitution parameter β. Secondly, I will pay attention to the time evolution of multidimensional inequalities. Moving from a unidimensional view of inequalities to a multidimensional point of view leads me to wonder if compensation effects not only between levels of attributes but also between the evolution of inequalities in each dimension will occur. I will consequently focus on which dimension dominates. This can also be a function of β. Finally, considering multiple well-being attributes can lead to larger variations of inequalities as correlation effects can increase or reduce the amplitude of variability in inequality values.

Multidimensional inequalities using Maasoumi's indicator

Bi-dimensional inequalities I defined in the measurement methodology the Maasoumi indicator $M(0)$, which can be quantified for any number of dimensions. In order to identify the links between dimensions, firstly, I will present multidimensional inequalities considering pairs of dimensions given by i) yearly wage and education ii) total income and education, iii) yearly wage and health, iv) total income and health and finally v) education and health. As I found no difference between the unidimensional inequalities of yearly wages and hourly wages, I noted above that only the former will be developed for the analysis that follows. Results regarding hourly wages are included in the appendix. Moreover, I present the results using graphical representations, which give a clear picture of the time evolution as well as the range of multidimensional inequalities.

Bi-dimensional inequalities of yearly wage and education are presented in Figures 3-7 for the three cohorts. Looking first at the values of multidimensional inequalities as a function of β, we clearly see that the lower the level of substitution, the higher the multidimensional inequalities.[18] This tends to confirm the intuition which states that a higher preference for complementarity leads to higher levels of social inequalities as inequalities are perceived as additive.

Next, regarding the evolution, we clearly see a different evolution between the situation where attributes are perfect substitutes and where they are not. Regarding the former, it is the evolution of educational inequalities which dominates. Additionally this dimension has the highest values once normalization is made (see Tables 3.13, 3.14 and 3.15 in the appendices). Given that perfect substitution exists in the case of $\beta = -1$,

[18] I recall that $\beta = -1$ corresponds to complete substitution between attributes and the degree of substitution decreases with the increase in β.

the prevailing attributes have the highest weight, driving the evolution of bi-dimensional inequalities of education and health. Once the degree of substitution decreases, we see that the worst performer, i.e. the evolution of yearly wage inequalities, dominates (see tables in the appendices), and the amplitudes of the variations underlying this time evolution increases in the value of β (i.e. with the decrease in the degree of substitution). Consequently, the high variabilities of inequalities we noted in the unidimensional case of yearly wage inequalities are also found here.

Moving to the bi-dimensional inequalities of income and education (Figure 3-8) similar patterns are observed. Results for $\beta = -1$ give the same evolutions and values range as in the bi-dimensional yearly wages education situation. However, a slight increase is noted between 1991 and 1993 for the whole sample even if neither total income, nor education inequalities are evolving in this way during this sub-period. The intuition between this phenomenon is relative to the correlations between total income and education for 1991 and 1993. I have underlined a positive and significant correlation between total income and education in 1991 and a negative and significant correlation between them in 1993. The change in the results of the correlation can potentially explain the reverse evolution of inequalities between unidimensional and multidimensional inequalities for the period 1991-1993. This proves the importance of considering multidimensional indicators once their evolutions can be deconnected from those in the unidimensional situation given the level and significance of correlation between attributes.

Concerning the differences observed between the perfect substitution situation and lower degrees of substitution, once again I note an increase in the values of multidimensional inequalities with β. The preference for complementarity between attributes consequently lead to higher values of multidimensional inequalities which result from the addition of the levels of inequalities in each dimension. Moreover, I also draw similar conclusions regarding the attribute's evolution which dominates once lower degrees of substitution are considered, i.e. the worst performer which is total income (see tables in the appendix). We consequently note a sharp drop in inequalities between 1991 and

1997 and an increasing tendancy (less noticeable for the sub-sample of individuals less than 26 years old) between 1997 and 2004.

Similar conclusions can be drawn from the analysis of bi-dimensional inequalities of yearly wage and health (Figure 3-9) and of total income and health (Figure 3-10). The evolution of inequalities in health, which is the best performer in terms of normalized values, most of the time dominates when there is perfect substitutability. Exceptions are noted between 1993 and 1997 for the whole sample and the yearly wage as this last dimension inequalities' evolution dominates, and between 1991 and 1993 for all samples and total income as this last dimension inequalities evolution's also dominates on this period of time. A more interesting situation occurs between 1993 and 1997 for total income in the situation where $\beta = -1$ and for the whole sample: the slight decrease observed corresponds neither to the increase of total income inequalities nor to the increase in health inequalities. It is interesting once again to note the parallel between the evolution of correlations observed during these two periods. It was positive and significant in 1993 whereas it became non significant in 1997. This can be an element of interpretation. Finally, as before, for decreasing values of the degree of substitution, the evolution of the worst performer's inequalities dominates (except between 1991 and 1997 for total income).

Looking at the bi-dimensional inequalities of education and health, some preceding results are confirmed whereas some new elements are brought to our attention. The negative relationship linking the degree of substitutability and the level of the values of multidimensional inequalities is supported once again by these bi-dimensional inequalities. However, a more complex evolution with time occurs which does not always correspond i) to the evolution of the worst performer attribute when considering lower levels of substitution and ii) to the evolution af any dimension. For instance, whereas the evolution obtained for an infinite level of substitution was before that of the dimension with the highest values of normalized attributes, for the youngest sample, it is still the case between 1991 and 1997 (as education, which remains as the dimension with the highest normalized values, has evolution which dominates), but from 1997, the strong increase of bi-dimensional inequalities are related more to the evolution of health inequalities. Similarly, for 26-33 year old individuals, the drop between 1991 and 1993 underlined for values of β higher than -1 corresponds more to the strongest attribute whereas the evolution between 1993 and 2007 mirrors neither the slight increase in education inequalities nor the increase in health inequalities during this period. Moreover, for the whole sample, we clearly see that it is principally the evolution of education inequalities which dominates. Consequently, no prediction could have been made concerning the evolution of these bi-dimensional inequalities. A possible explaination for this phenomenon is the non significance of the correlations between health and education emphasized above. In the case of non correlation between attributes, we clearly see that the multidimensional analysis leads us to new and important conclusions relative to the evolution of inequalities of well-being. This supports the necessity for an aggregating indicator which takes correlations into account.

To draw a first conclusion relative to bi-dimensional inequalities in China, an important phenomenon must be stressed when comparing the level of bi-dimensional inequalities between cohorts. Most of the time, inequalities for the two sub-samples are lower, due to a less diversified population being included in these samples. However, when we look at bi-dimensional inequalities including income dimensions, convergence occurs when considering quite a high level of complementarity between attributes. A more striking result concerns the comparison between the cohort under 26 years of age and 26-33 year olds. We see that inequalities for the former group are most often above those obtained for the latter. Given that the objective of the differenciation between cohorts was to see the impact of reforms on the level of inequalities, this conclusion can be alarming as the youngest individuals are those who have been most and who will be most affected by the reform process. And this is even more the case when considering bi-dimensional inequalities of human capital (combining education and health dimensions). Given that these attributes are increasingly linked to the level of income (through returns to human capital) and bring about by themselves well-being, the rise in well-being inequalities described through cohort differentiation is not a source of optimism for the future.

Tri-dimensional inequalities of income, education and health In this section, the only distinction I will make is relative to the type of income introduced in the multidimensional indicators. Looking at Figures 3-11 and 3-12, it is now obvious that i) the level of multidimensional inequalities increases with the decrease in the level of substitution; ii) for infinite substitution, the evolution of inequalities in the strongest dimensions dominates; iii) as we converge to complementarity of attributes, the evolution of inequalities for the worst performer in terms of normalised values dominates, with increasing amplitudes with the decrease in the degree of substitution.

Moreover, phenomena linked to the higher or lower of correlations between dimensions still occur. We see that for $\beta = -1$, which corresponds to perfect substitution, an evolution similar to that observed in the bi-dimensional education and health situation is obtained. Consequently, not only the evolution of inequalities in the two dimensions with the highest normalized values for attributes dominates, but also the surprising conclusions stressed before for the evolutions of the bi-dimension education and health inequalities are still obtained, potentially linked to the insignificant correlations between these two attributes.

Regarding the level of inequalities, we note once more a higher level for the youngest sub-sample even if convergence occurs when complementarity between attributes is high. This confirms the alarming conclusions drawn in the bi-dimensional case relative to the impact of the reforms on wellbeing inequalities.

3.4 Conclusion

Regarding the analysis which has been done, important implications must be considered when studying well-being inequalities in general. Firstly, I underlined that given the degree of substituability the population or policy makers choose, the evolution of multidimensional inequalities of wellbeing will be different. In the case of perfect substitutability, the stronger dimensions, given for attributes with the highest normalized values, have evolutions of inequalities which dominate. For the case of Chinese data and when considering bi or tri-dimensional inequalities including income (both yearly wage and total income), with human capital dimensions being the stronger dimensions, we note very different patterns of evolution of multidimensional inequalities compared to the unidimensional inequalities of income. We clearly see the importance of moving to the multidimensional analysis in this case.

Moreover, moving from perfect substitutability to perfect complementarity between attributes, I emphasized two important phenomena: the range of the multidimensional indicator values increase, confirming the idea that substitution between attributes tend to decrease the level of wellbeing inequalities, and the poorest performers in terms of normalized values of the dimensions have inequalities' evolution which dominates, confirming the higher weight given to the dimensions in which there is higher deprivation when moving to a more Rawlsian appraoch of well-being inequalities.

Another important phenomenon highlighted by the analysis is the role played by the more or less high, and relative significance of correlations between attributes. It seems that when attributes are not significantly correlated, no prediction can be made relative to the evolution of inequalities as we cannot see a clear polarization of wellbeing dotations between rich and poor individuals in terms of wellbeing attributes. Compensation effects as well as evolutions of neither dimensions are abserved.

Finally, as the focus of this chapter is to describe the evolution of wellbeing inequalities in China, even if they are sensitive to the values of the parameter of substituion, some conslusions can be made. To do so, I principally consider results obtained when a non perfect level of substitution is chosen which can be seen as the most common case. Firstly, wage inequalities are highly unstable, translating as the influence of equalizing and disequalizing forces which can lead to different compensation effects depending on the period considered. A more precise level of disaggregation would maybe bring about better insights to identify the sources of these evolutions. Given that multidimensional inequalities translate this as the evolution of this dimension's inequalities, no clear conclusions can be drawn on multdimensional inequalities of well being when considering yearly wages as the income dimension.

Moving to multidimensional inequalities including total income, which is meant to take into account intra-households allocations, I highlight a strong decrease in wellbeing inequalities between 1991 and 1997 but an increasing trend between 1997 and 2004.

To better analyse the effects of reforms on the evolution of wellbeing inequalities, I distinguished cohorts of individuals who have potentially been to a certain degree affected by the reform process. Given that the evolution of inequalities in incomes most often dominates and that these are very similar between cohorts, no clear difference is observed regarding the evolution of multidimensional inequalities once the income dimension is included. However, lower levels of multidimensional inequalities are observed for the younger cohorts. This is logical as the whole sample includes more diversified individuals. Focusing on the two sub-samples I highlight higher levels of multidimensional inequalities for the youngest sub-cohort of individuals (under 26 years of age) in most of the situations. On the other hand, exceptions are found in the case of perfect substitution between attributes. As soon as complementarity between dimensions is introduced, the youngest cohort suffers the most from wellbeing inequalities whatever the dimensions aggregated. Consequently, this last conclusion highlights the role of reforms in the deterioration of the situation in terms of well-being, leading to alarming anticipations for the future. This is particularly the case when considering human capital inequalities (bi-dimensional inequalities of education and health), which is even more negative for the future evolution of well-being inequalities as these two determinants contribute to the improvement of the income dimension, and as they also contribute to well-being by themselves.

The analysis of well-being inequalities in China has once again underlined the risks of a deterioration of the situation among the Chinese population. This is supported with the conclusions enounced when considering intergenerational mobility which emphasized a potential decrease, in the future, of opportunities in terms of equal access to better education. Chinese policy makers will have to pay a special attention to the education and health dimensions which appear to be negatively affected by the reforms, which in turn could have negative consequences on income and wellbeing as a whole.

	Reforms: 1980	1991	1993	1997	2000
Whole sample		2734	2312	2001	1805
Health reforms					
Child care system					
Sub-sample 1:		16-25	16-25	16-25	16-25
% tot population		21.69% (593)	18.90% (437)	15.74% (312)	9.92% (179)
Cohorts affected	0-5	11-16	13-18	16-22	16-25
% sub-sample 1		3.54% (21)	11.67% (51)	55.77% (174)	100% (179)
Child nutrition					
Sub-sample 2:		26-33	26-33	26-33	26-33
% tot population		22.16% (606)	20.07% (464)	17.59% (352)	17.01% (307)
Cohorts affected	6-13	17-24	19-26	23-30	26-33
% sub-sample 2		0% (0)	7.97% (37)	69.03% (243)	100% (307)
Education reforms					
Sub-sample 3:		26-35	26-35	26-35	26-35
% tot population		28.46% (778)	25.86% (598)	25.79% (516)	23.38% (422)
Cohorts affected	6-15	17-26	19-28	23-32	26-35
% sub-sample 3		9.00% (70)	25.58% (153)	55.43% (286)	100% (422)

Table 3.1: Cohorts affected by the reforms in different fields

Tables chapter 3

	Yearly income		Total income	
	Pearson correlation coefficient	Spearman rank correlation	Pearson correlation coefficient	Spearman rank correlation
1991				
Hourly wages	0.4164	0.8885 0.0000	−0.0129	0.0578 0.0025
Yearly wages			0.0461	0.0827 0.0000
1993				
Hourly wages	0.7893	0.9034 0.0000	0.0317	0.2471 0.0000
Yearly wages			0.0479	0.2474 0.0000
1997				
Hourly wages	0.7785	0.8549 0.0000	0.2302	0.3774 0.0000
Yearly wages			0.2888	0.3938 0.0000
2000				
Hourly wages	0.8187	0.8302 0.0000	0.1886	0.3232 0.0000
Yearly wages			0.2345	0.3887 0.0000
2004				
Hourly wages	0.8221	0.8689 0.0000	0.2217	0.4734 0.0000
Yearly wages			0.2568	0.4936 0.0000

P-values in parenthesis

Table 3.2: Correlations between monetary dimensions for the whole sample

	Education		BMI	
	Pearson correlation coefficient	Spearman rank correlation	Pearson correlation coefficient	Spearman rank correlation
1991				
Hourly wages	−0.0726	−0.0128 (0.5022)	0.0122	0.1223 (0.0000)
Yearly wages	−0.0427	−0.0160 (0.4020)	0.1036	0.1347 (0.0000)
Total income	−0.0040	0.0666 (0.0005)	0.0655	0.0863 (0.0000)
Education			−0.084	−0.074 (0.0001)
1993				
Hourly wages	−0.0411	−0.0513 (0.0072)	0.0321	0.0559 (0.0000)
Yearly wages	−0.0290	−0.0450 (0.0305)	0.0147	0.0400 (0.0542)
Total income	0.0457	−0.0975 (0.0000)	−0.0404	0.2474 (0.0000)
Education			−0.0290	−0.0422 (0.0422)
1997				
Hourly wages	0.0066	0.0588 (0.0085)	0.0149	0.0803 (0.0003)
Yearly wages	−0.0261	0.0166 (0.4572)	0.0035	0.0701 (0.0017)
Total income	0.0722	0.1207 (0.0000)	0.0040	0.0225 (0.3148)
Education			−0.0043	−0.0054 (0.8101)
2000				
Hourly wages	0.0865	0.2120 (0.0000)	0.0386	0.1129 (0.0000)
Yearly wages	0.0962	0.1929 (0.0000)	0.0546	0.1092 (0.0000)
Total income	0.1230	0.2030 (0.0000)	0.0606	0.0973 (0.0000)
Education	1.000	1.0000	0.0339	0.0427 (0.0696)
2004				
Hourly wages	0.1791	0.3981 (0.0000)	0.0540	0.0676 (0.0045)
Yearly wages	0.1540	0.3757 (0.0000)	0.0532	0.0835 (0.0004)
Total income	0.1776	0.2723 (0.0000)	0.0115	0.0662 (0.0054)
Education			0.0381	0.0146 (0.5407)

P-values in parenthesis

Table 3.3: Correlations between welfare dimensions for the whole sample

	Yearly income		Total income	
	Pearson correlation coefficiant	Spearman rank correlation	Pearson correlation coefficient	Spearman rank correlation
1991				
Hourly wages	0.7723	0.9161 (0.0000)	0.0445	−0.0012 (0.9776)
Yearly wages			0.1206	−0.0012 (0.9777)
1993				
Hourly wages	0.9047	0.9273 (0.0000)	0.0338	0.1679 (0.0004)
Yearly wages			0.0233	0.1525 (0.0014)
1997				
Hourly wages	0.7845	0.8993 (0.0000)	0.1531	0.3865 (0.0000)
Yearly wages			0.2947	0.3892 (0.0000)
2000				
Hourly wages	0.8851	0.8463 (0.0000)	0.2107	0.3245 (0.0000)
Yearly wages			0.2198	0.3687 (0.0000)
2004				
Hourly wages	0.8608	0.8713 (0.0000)	0.2776	0.4213 (0.0000)
Yearly wages			0.1259	0.3937 (0.0000)

P-values in parenthesis

Table 3.4: Correlations between monetary dimensions for the sub-sample 1

	Yearly income		Total income	
	Pearson correlation coefficient	Spearman rank correlation	Pearson correlation coefficient	Spearman rank correlation
1991				
Hourly wages	0.5893	0.9131 (0.0000)	−0.0391	0.0478 (0.0025)
Yearly wages			0.0123	0.0803 (0.0506)
1993				
Hourly wages	0.9059	0.9448 (0.0000)	0.2208	0.3897 (0.0000)
Yearly wages			0.2216	0.3790 (0.0000)
1997				
Hourly wages	0.7476	0.8212 (0.0000)	0.2366	0.3075 (0.0000)
Yearly wages			0.3730	0.3265 (0.0000)
2000				
Hourly wages	0.8760	0.8076 (0.0000)	0.3453	0.2672 (0.0000)
Yearly wages			0.4091	0.3531 (0.0000)
2004				
Hourly wages	0.8940	0.8473 (0.0000)	0.2633	0.5043 (0.0000)
Yearly wages			0.3976	0.5307 (0.0000)

P-values in parenthesis

Table 3.5: Correlations between monetary dimensions for the sub-sample 2

	Education		BMI	
	Pearson correlation coefficiant	Spearman rank correlation	Pearson correlation coefficiant	Spearman rank correlation
1991				
Hourly wages	−0.1466	−0.0761 (0.0637)	−0.0531	−0.0045 (0.9123)
Yearly wages	−0.1316	−0.0942 (0.0217)	0.0233	0.0036 (0.9298)
Total income	−0.0007	0.1872 (0.0000)	0.0257	−0.0452 (0.2709)
Education			−0.0778	−0.0923 (0.0244)
1993				
Hourly wages	−0.0726	−0.1002 (0.0363)	−0.0658	−0.0438 (0.3613)
Yearly wages	−0.0745	−0.0999 (0.0369)	−0.0588	−0.0510 (0.2876)
Total income	0.0308	0.1109 (0.0204)	−0.0733	−0.0256 (0.5933)
Education			−0.0687	−0.0763 (0.1112)
1997				
Hourly wages	−0.1122	−0.0102 (0.8565)	−0.0479	0.0331 (0.5579)
Yearly wages	−0.1603	−0.0686 (0.2246)	−0.0438	−0.0165 (0.7707)
Total income	0.0992	0.1326 (0.0186)	−0.0380	−0.0444 (0.4324)
Education			−0.0080	−0.0353 (0.5327)
2000				
Hourly wages	0.1414	0.0957 (0.1564)	0.0357	0.1059 (0.1165)
Yearly wages	0.0778	0.0227 (0.7367)	0.1097	0.0922 (0.1720)
Total income	0.1268	0.2278 (0.0006)	−0.0119	−0.0549 (0.4166)
Education			−0.0719	−0.0596 (0.3781)
2004				
Hourly wages	0.1797	0.3548 (0.0000)	0.1446	0.1616 (0.0307)
Yearly wages	0.0375	0.2420 (0.0000)	0.1348	0.1715 (0.0217)
Total income	0.4011	0.4153 (0.0001)	0.0574	0.0997 (0.1840)
Education			0.1664	0.1519 (0.0424)

P-values in parenthesis

Table 3.6: Correlations between welfare dimensions for the sub-sample 1

	Whole sample		16-25 years old sample		26-33 years old sample	
	Obs.	Mean	Obs.	Mean	Obs.	Mean
Hourly wages						
1991	2734	0.6417	594	0.4963	594	0.5954
		(0.8054)		(0.3661)		(0.4957)
1993	2312	0.8972	437	0.6868	464	0.7893
		(1.0342)		(0.5097)		(0.5571)
1997	2001	1.4127	315	1.1935	352	1.3807
		(1.0911)		(1.2478)		(1.0779)
2000	1805	2.0433	221	1.8717	307	2.0706
		(2.4482)		(3.0196)		(2.3566)
2004	1763	2.7388	179	2.1372	307	2.6310
		(2.6593)		(1.6463)		(2.9510)
Yearly wages						
1991	2734	1367.575	594	1108.822	594	1302.233
		(797.887)		(684.105)		(747.494)
1993	2312	1932.066	437	1556.114	464	1840.362
		(1962.91)		(1111.211)		(1359.578)
1997	2001	2826.169	315	2435.052	352	2658.831
		(2040.081)		(1812.466)		(1544.328)
2000	1805	3870.437	221	3765.658	307	3967.828
		(4212.692)		(5142.425)		(4626.461)
2004	1763	5519.484	179	4778.28	307	5559.498
		(5277.709)		(5203.294)		(6182.683)
Total income						
1991	2734	10, 232.8	594	10, 414.23	594	8754.094
		(16186.14)		(21228.29)		(10.470.95)
1993	2312	3, 739.72	437	4214.374	464	3298.8
		(3883.307)		(4615.974)		(3118.35)
1997	2001	3174.876	315	3262.091	352	2973.579
		(2066.005)		(2272.232)		(1842.719)
2000	1805	4562.936	221	4652.693	307	4316.461
		(3676.794)		(4823.693)		(3488.197)
2004	1763	6393.062	179	5100.835	307	6130.085
		(5389.245)		(3247.867)		(5245.527)
Years schooling						
1991	2734	8.6358	594	9.4233	594	9.7458
		(3.4894)		(2.2158)		(2.5462)
1993	2312	8.8091	437	9.4233	464	9.7263
		(3.5058)		(2.2158)		(2.4346)
1997	2001	9.2279	315	9.9651	352	9.6734
		(3.3291)		(2.3773)		(2.6642)
2000	1805	9.8853	221	10.7692	307	10.3160
		(3.4414)		(2.6591)		(2.8217)
2004	1763	10.9450	179	11.6313	307	11.6710
		(3.3106)		(2.8161)		(3.0107)
BMI						
1991	2734	22.1193	594	20.7787	594	21.7886
		(2.9701)		(2.3705)		(2.5888)
1993	2312	22.1833	437	20.7787	464	21.6929
		(2.9027)		(2.3953)		(2.4854)
1997	2001	22.8141	315	21.1431	352	22.3495
		(3.0245)		(2.7154)		(2.8663)
2000	1805	23.1808	221	21.2048	307	22.5066
		(3.2541)		(3.1336)		(3.0799)
2004	1763	23.6091	179	22.1030	307	22.6823
		(3.9098)		(4.3170)		(3.8845)

Standard deviations in parenthesis. All incomes are in Yuan.

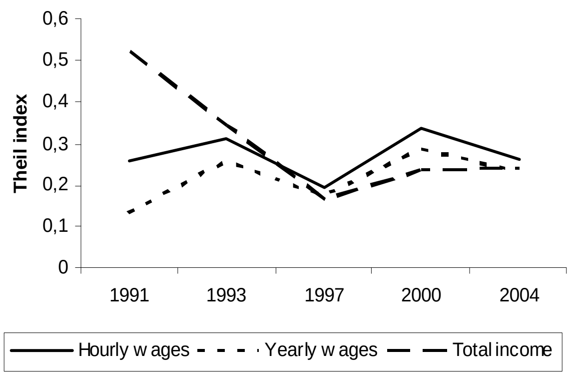
Table 3.8: Summary statistics for the different samples

	1991	1993	1997	2000	2004
Whole sample					
Total	0.1042	0.1027	0.0808	0.0716	0.0512
Between	0.0228 (22%)	0.0230 (22.4%)	0.0199 (24.6%)	0.0257 (36%)	0.0266 (52%)
Within	0.0814 (78%)	0.0797 (77.6%)	0.0609 (75.4%)	0.0459 (64%)	0.0246 (48%)
Sample of 16-25 years of age					
Total	0.0370	0.0307	0.0317	0.0292	0.0293
Between	0.0113 (30.5%)	0.0084 (27.4%)	0.0101 (32%)	0.0185 (63.3%)	0.0203 (69.3%)
Within	0.0257 (69.5%)	0.0223 (72.6%)	0.0216 (68%)	0.0107 (36.7%)	0.0090 (30.7%)
Sample of 26-33 years of age					
Total	0.0387	0.0335	0.0404	0.0385	0.0333
Between	0.0140 (36.2%)	0.0133 (39.7%)	0.0140 (36.4%)	0.0197 (51.2%)	0.0233 (70%)
Within	0.0247 (63.8%)	0.0202 (60.3%)	0.0264 (63.6%)	0.0188 (48.8%)	0.010 (30%)

Table 3.9: Decomposition of educational inequalities between groups of individuals who have more and less than 12 years of education

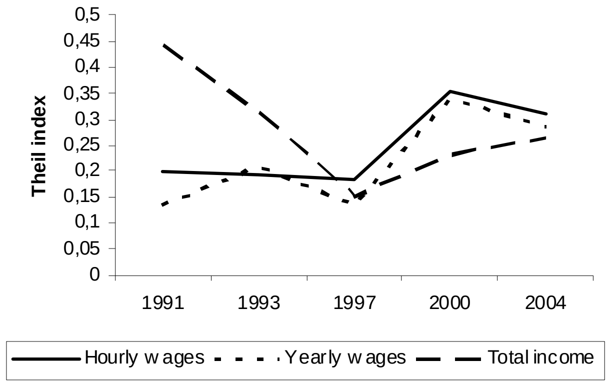
Figures chapter 3

Figures section 3.2.1



Whole sample

Less than 26 year old individuals



26-33 year old individuals

Figure 3-3: Comparison of the evolution of hourly wages, yearly wages and total income inequalities considering Theil index

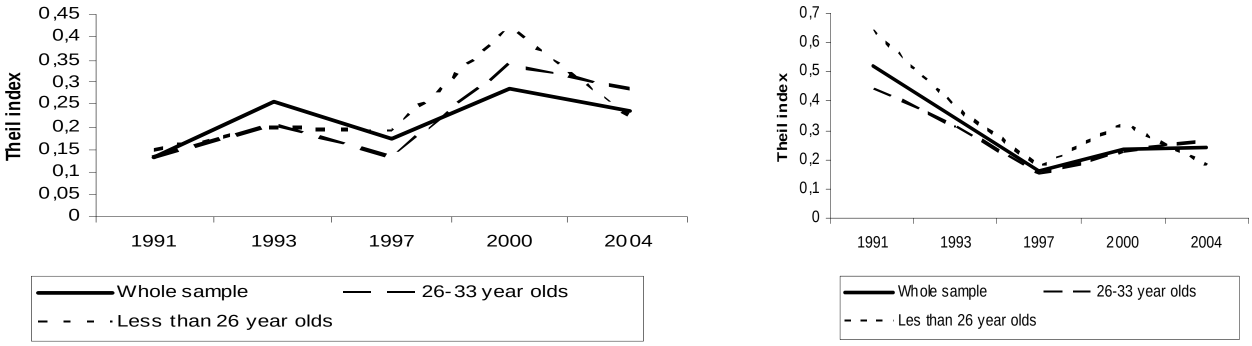


Figure 3-4: Comparison of yearly wages and total income inequalities evolution between cohorts

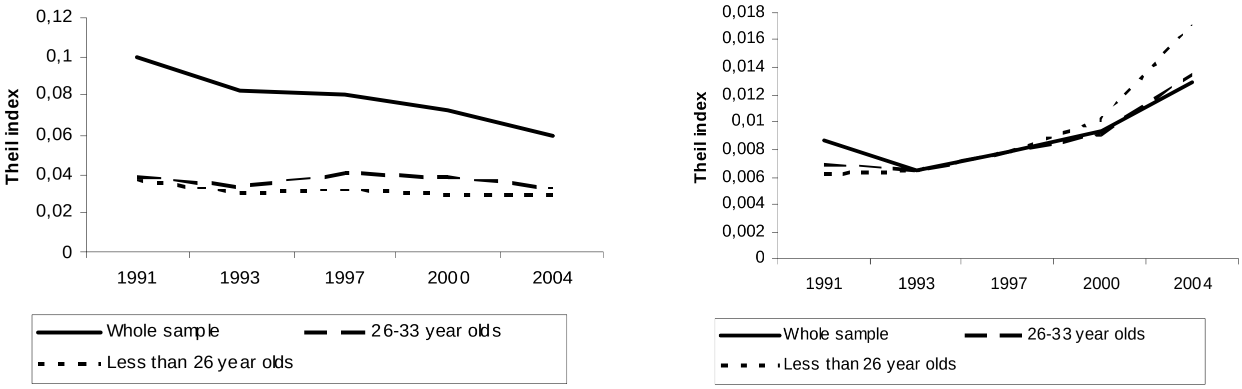


Figure 3-5: Comparison of human capital inequalities evolution between cohortes

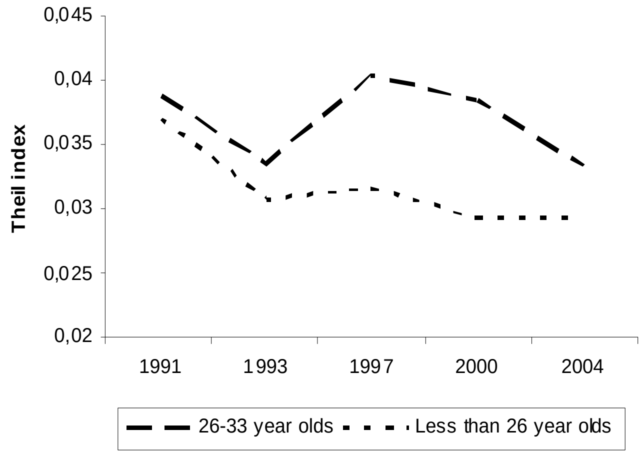


Figure 3-6: Comparison of educational inequalities evolution between the sub-samples 1 and 2

Figures section 3.2.2

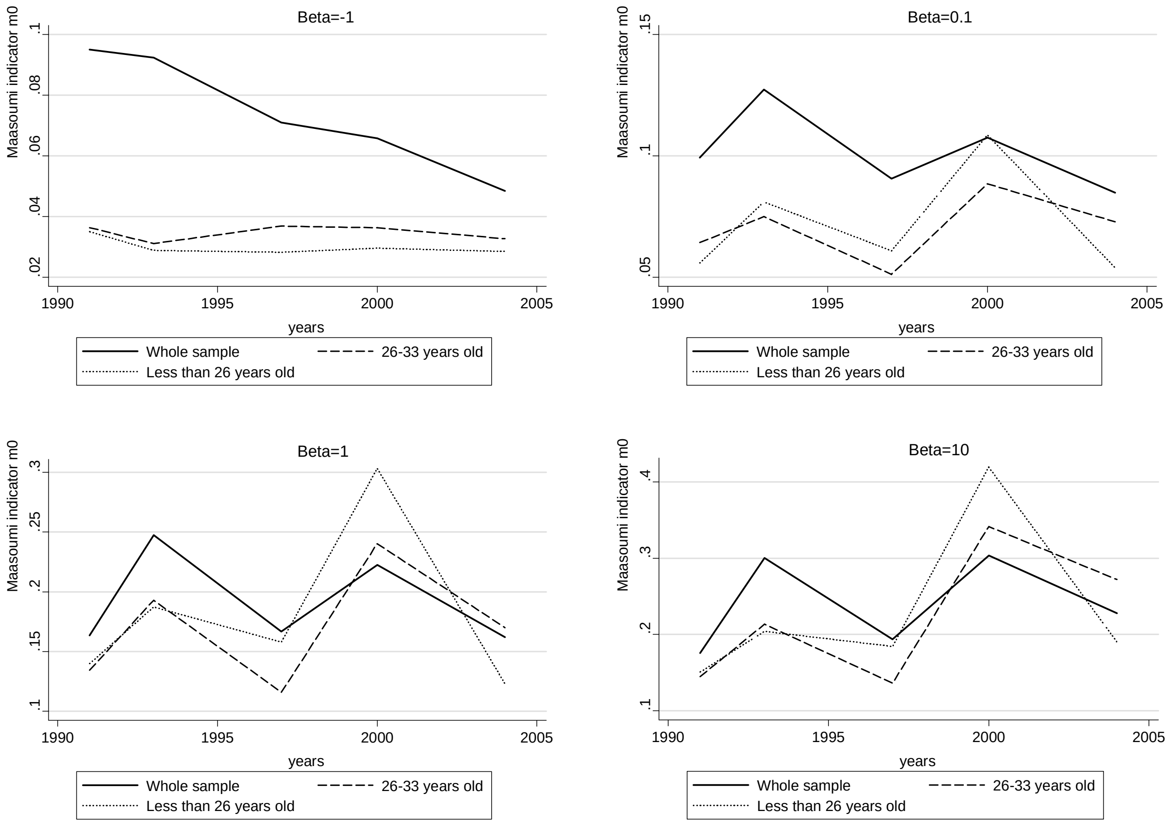


Figure 3-7: Evolution of bi-dimensional inequalities in yearly wages and education as a function of beta

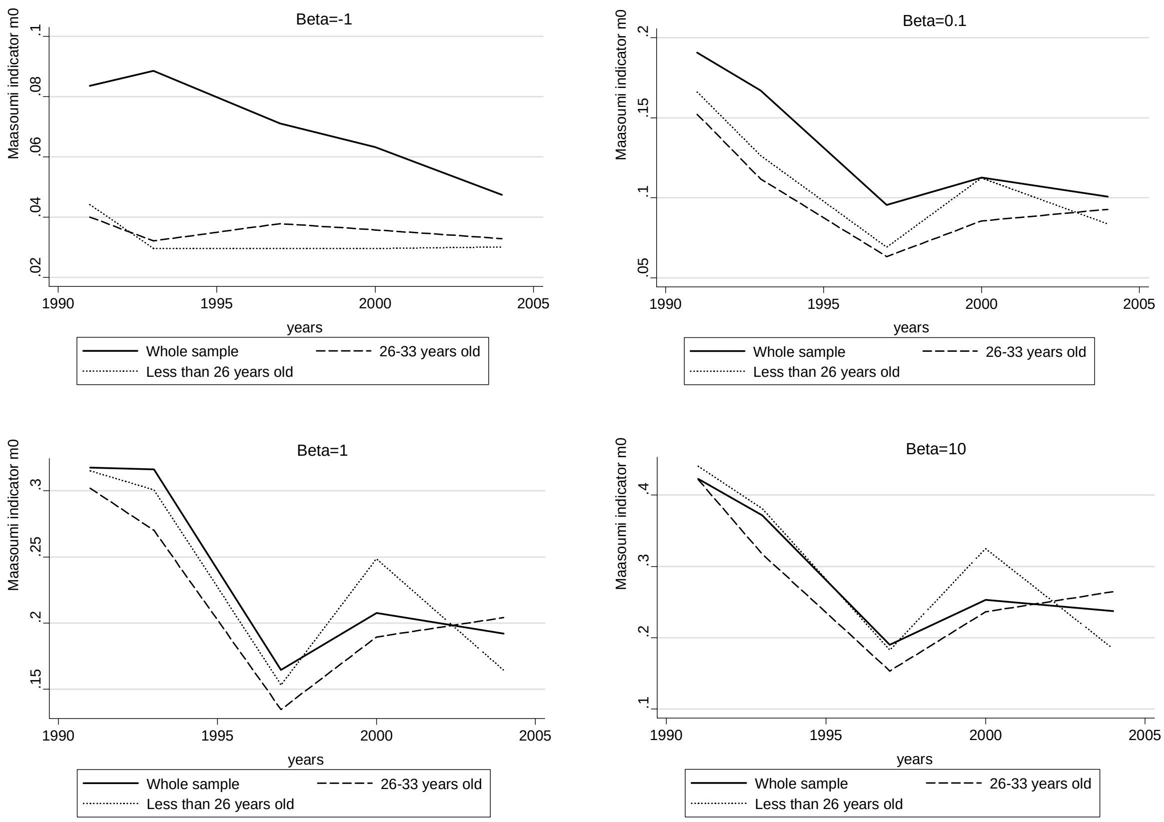


Figure 3-8: Evolution of bi-dimensional inequalities in total income and education as a function of beta

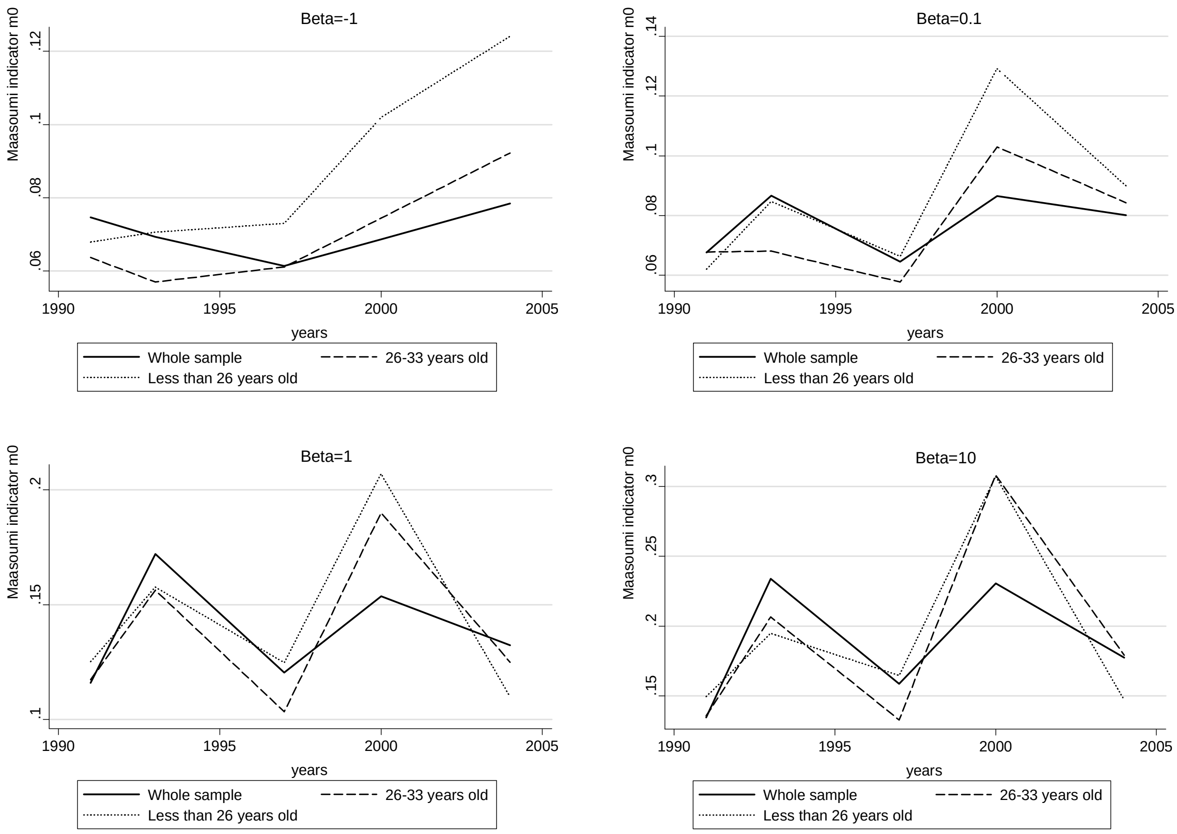


Figure 3-9: Evolution of bi-dimensional inequalities in yearly wages and health as a function of beta

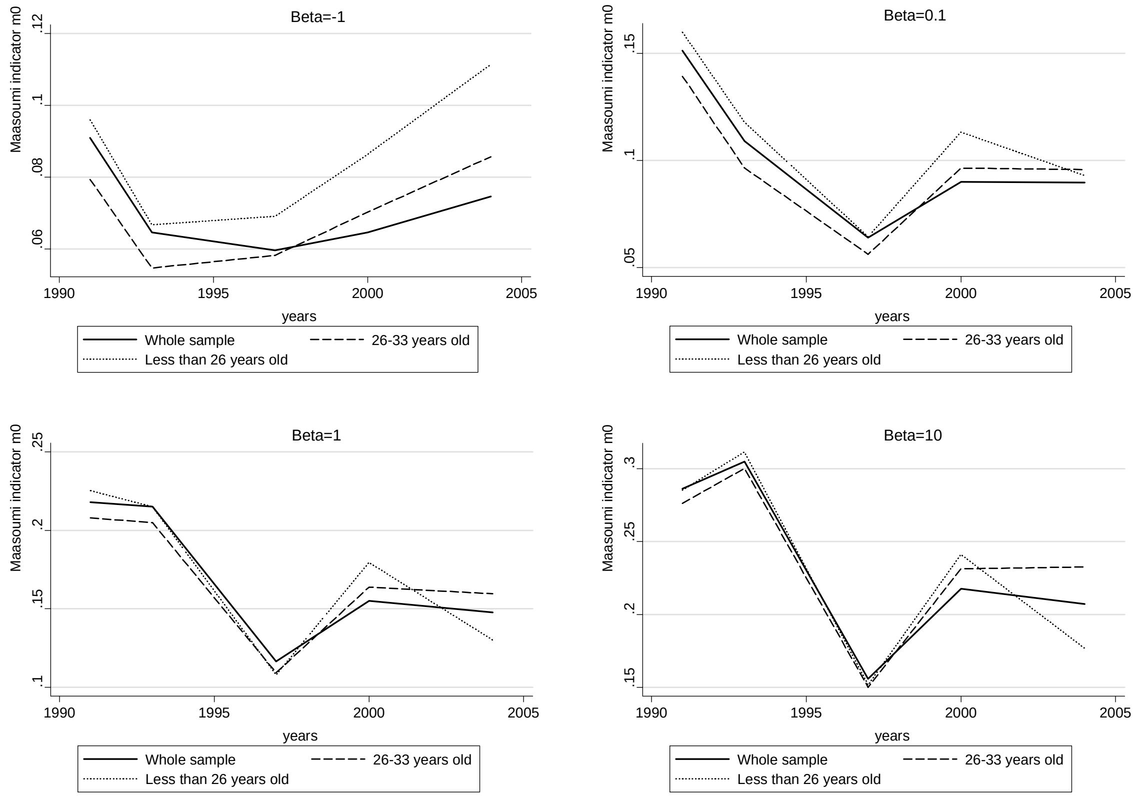


Figure 3-10: Evolution of bi-dimensional inequalities in total income and health as a function of beta

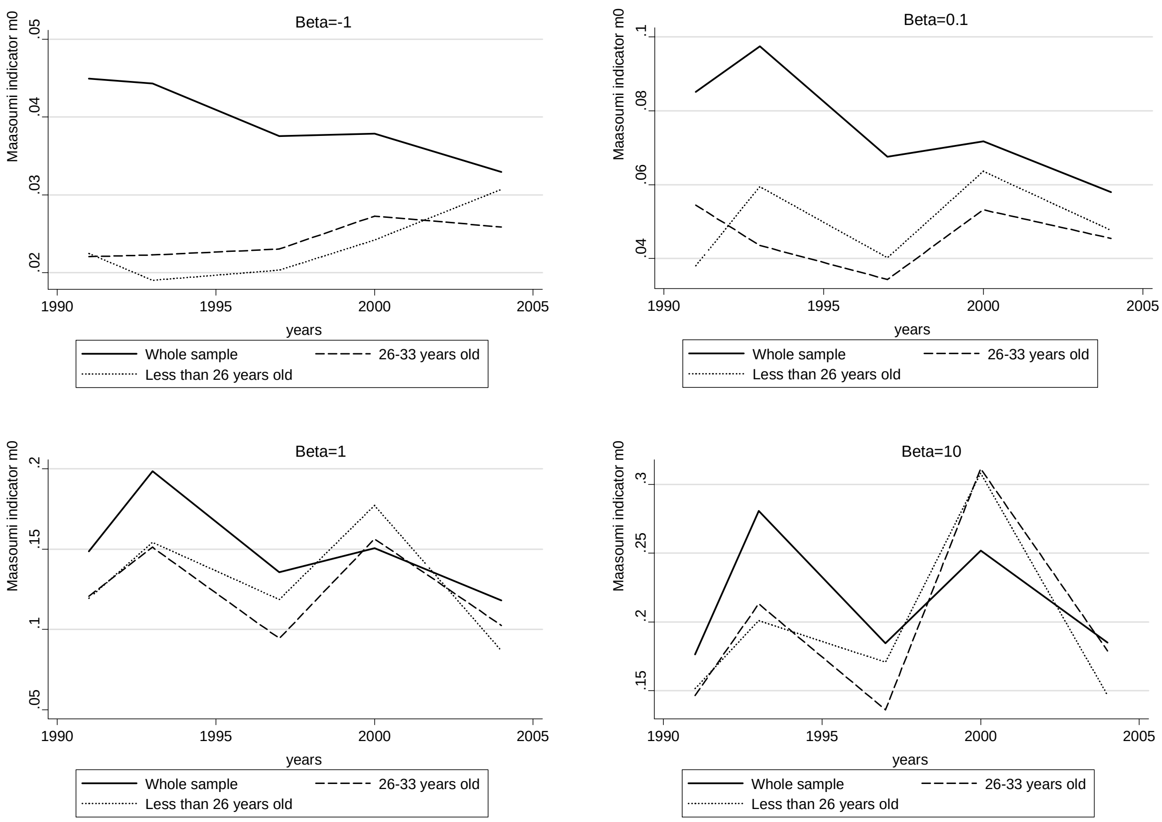


Figure 3-11: Evolution of tri-dimensional inequalities in yearly wages, education and health as a function of beta

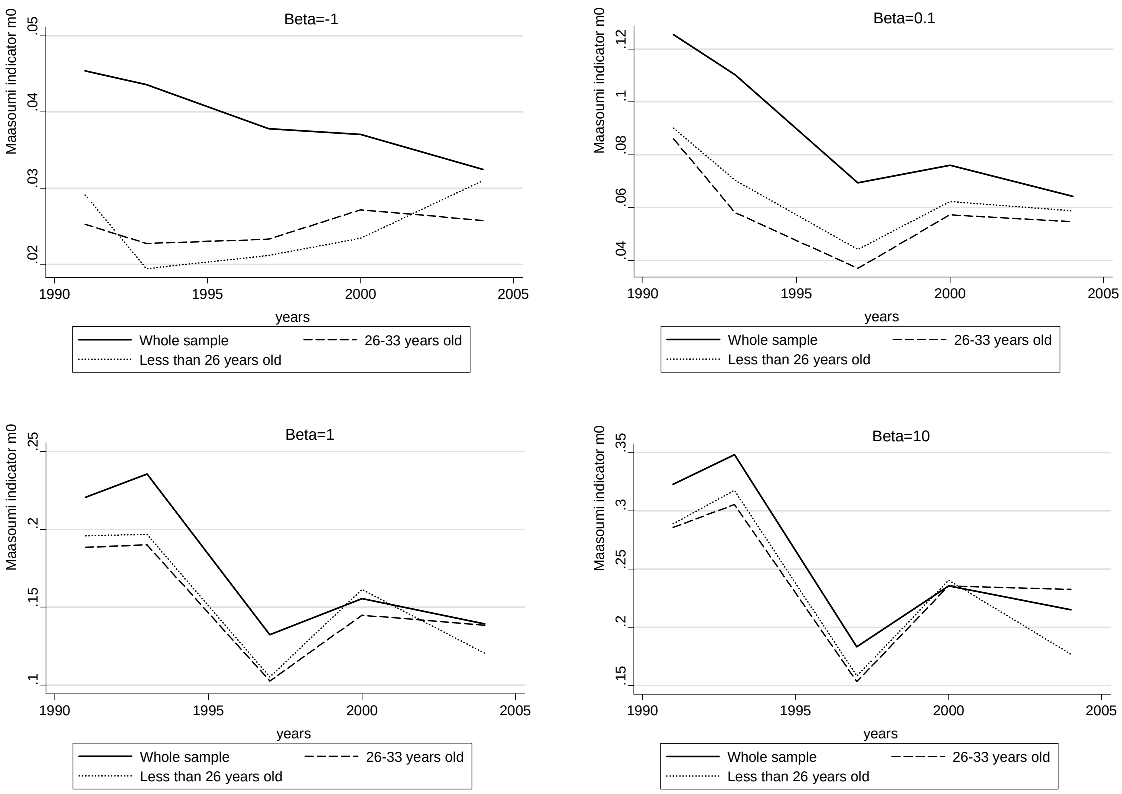


Figure 3-12: Evolution of tri-dimensional inequalities in total income, education and health as a function of beta

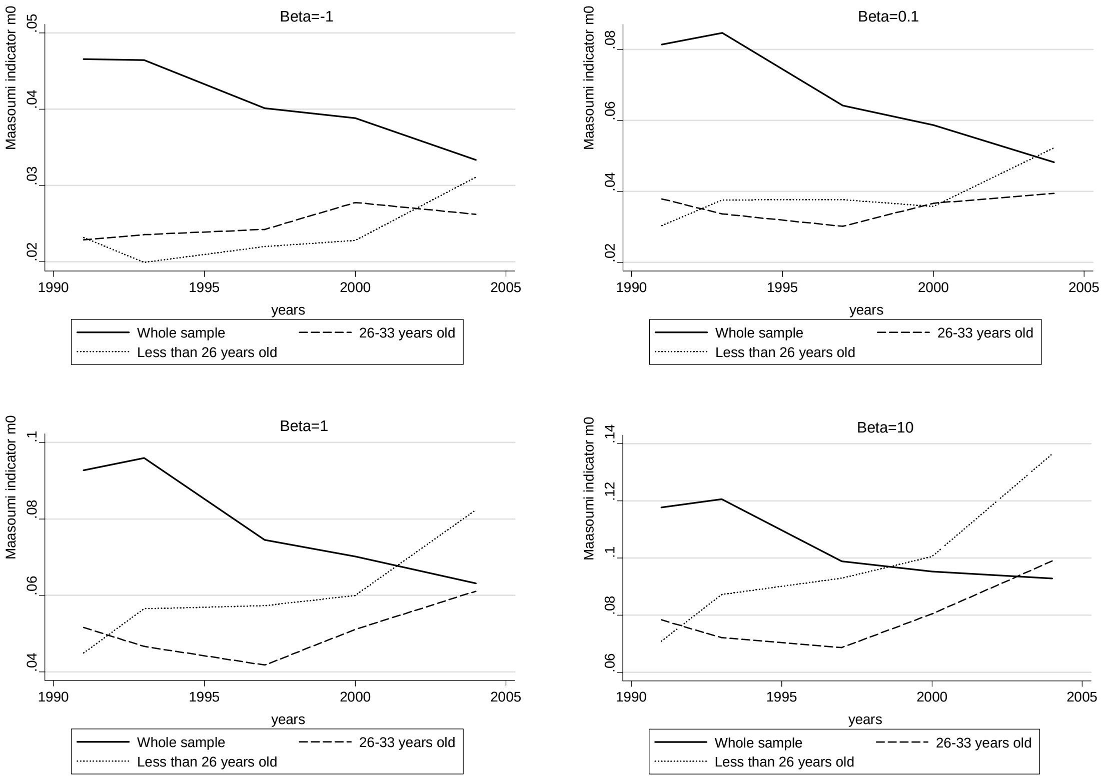


Figure 3-13: Maasoumi indicators for education and health as a function of beta

3.A Appendix Chapter 3

3.A.1 Unidimensional Theil index

	1991	Δ(%)	1993	Δ(%)	1997	Δ(%)	2000	Δ(%)	2004
Hourly wage									
Theil	0.2548	21.58	0.3098	−37.86	0.1925	74.55	0.336	−22.29	0.2611
Yearly wage									
Theil	0.1342	90.09	0.2551	−31.48	0.1748	63.62	0.286	−18.50	0.2331
Total Income									
Theil	0.5229	−34.71	0.3414	−52.08	0.1636	44.25	0.236	1.27	0.239
Education									
Theil	0.1029	−0.87	0.102	−20.98	0.0806	−11.29	0.0715	−28.53	0.0511
Health									
Theil	0.0087	−4.60	0.0083	3.61	0.0086	12.79	0.0097	32.99	0.0129

Table 3.10: Evolution of unidimensional inequalities for the whole sample using Theil index

	1991	Δ(%)	1993	Δ(%)	1997	Δ(%)	2000	Δ(%)	2004
Hourly wage									
Theil	0.1866	10,77	0.2067	29,17	0.2670	82,43	0.4871	-60,03	0.1947
Yearly wage									
Theil	0.1510	31,13	0.1980	-3,23	0.1916	118,95	0.4195	-46,51	0.2244
Total Income									
Theil	0.6422	-41,05	0.3786	-52,69	0.1791	81,63	0.3253	-43,31	0.1844
Education									
Theil	0.0370	-17,03	0.0307	3,26	0.0317	-7,57	0.0293	0,00	0.0293
Health									
Theil	0.0062	4,84	0.0065	21,54	0.0079	27,85	0.0101	69,31	0.0171

Table 3.11: Evolution of unidimensional inequalities for the sub-sample 1 using Theil index

	1991	Δ(%)	1993	Δ(%)	1997	Δ(%)	2000	Δ(%)	2004
Hourly wage									
Theil	0.1981	-2,22	0.1937	-4,18	0.1856	89,87	0.3524	-11,98	0.3102
Yearly wage									
Theil	0.1340	54,40	0.2069	-35,81	0.1328	154,67	0.3382	-15,38	0.2862
Total Income									
Theil	0.4439	-29,67	0.3122	-51,92	0.1501	54,56	0.2320	14,09	0.2647
Education									
Theil	0.0387	-13,44	0.0335	20,60	0.0404	-4,70	0.0385	-13,51	0.0333
Health									
Theil	0.0069	-5,80	0.0065	21,54	0.0079	15,19	0.0091	48,35	0.0135

Table 3.12: Evolution of unidimensional inequalities for the sub-sample 2 using Theil index

3.A.2 Normalization of the dimensions

	1991	1993	1997	2000	2004
Hourly wage	0.0152	0.0212	0.0335	0.0484	0.0649
Yearly wage	0.0118	0.0167	0.0244	0.0334	0.0476
Total Income	0.0846	0.0324	0.0274	0.0394	0.0552
Education	0.4797	0.4893	0.5126	0.5490	0.6080
Health	0.2165	0.2187	0.2376	0.2485	0.2615

Table 3.13: Normalized values for the various dimensions and for the whole sample

	1991	1993	1997	2000	2004
Hourly wage	0.0117	0.0162	0.0283	0.0442	0.0506
Yearly wage	0.0095	0.0134	0.0210	0.0323	0.0413
Total Income	0.0833	0.0364	0.0282	0.0400	0.0440
Education	0.5000	0.5200	0.5531	0.5960	0.6462
Health	0.1765	0.1766	0.1874	0.1890	0.2163

Table 3.14: Normalized values for the various dimensions and for the sub-sample 1

	1991	1993	1997	2000	2004
Hourly wage	0.0141	0.0187	0.0327	0.0490	0.0623
Yearly wage	0.0112	0.0159	0.0230	0.0342	0.0480
Total Income	0.0756	0.0285	0.0257	0.0373	0.0529
Education	0.5414	0.5403	0.5374	0.5731	0.6484
Health	0.2069	0.2013	0.2237	0.2284	0.2337

Table 3.15: Normalized values for the various dimensions and for the sub-sample 2

Conclusion

In this thesis, I have studied the evolution of social inequalities in Chinese society following the reform movement.

I focused on three topics: the role played by human capital in the determination of wages which was expected to increase, giving higher remuneration to more productive individuals than before the reforms; intergenerational mobility of education and wages, which mirror the potential persistence of inequalities between generations and inform us on the future evolution of inequalities; and multidimensional inequalities of income, education and health, which give a broad picture of the evolution of wellbeing inequalities as affected by the reforms.

Concerning the role played by human capital in the determination of wages, I used the panel data structure of the China Health and Nutrition Survey as well as the Semykina and Wooldridge (2005) method to adress common issues relative to the estimation of wage equations, i.e. individual heterogeneity, selection which varies through time and endogeneity of human capital variables. Moreover, as measurement errors were identified in my variables of interest, I used the set of instruments proposed by Dagenais and Dagenais (1997) to correct for these measurement error. Given the specificity of the type of measurement error I encountered, I needed to check if the econometric method chosen was appropriate and would give the most rigorous results concerning the coefficients estimated. The Montecarlo simulations implemented confirmed the good performances of the Semykina and Wooldridge (2005) estimation method as well as of the Dagenais and Dagenais (1997) set of instruments. These two tools combined together lead to precise estimations of the coefficient and to the lowest type I error compared to the commonly used OLS and fixed effects methods. Consequently, the use of these methods in estimating returns to human capital have shown themselves to be appropriate for my situation.

Results obtained using these methods have emphasized increasing returns to nutrition and education following the reform process. Specifically, returns to nutrition are only significant in the private sector, displaying an inverted U shape evolution with the increase in calory consumption (the threshold being found at 1791.3 Kcal). It appears consequently, that below a certain level of nutrition which can be observed by employers, more robust individuals will receive higher payments to compensate for higher effort. Given that the private sector is developing in China and that the operation of the public sector is converging with that of the private sector, we can expect returns to nutrition being higher in the future.

Returns to education also demonstrate a high sensitivity to the evolution of the reform process. Considering a time break differenciating the 1991-1993 and 2000-2004 periods, I noted higher returns to education for the latter period than for the former. Returns to education subsequently reach a value of 4.23%, which is lower than estimations obtained by other studies. However, the increase of these returns with time confirms the trend previously found in the academic literature. Increasing returns to human capital are in accordance with the path towards a true labour market. Disequalizing forces as underlined have stressed the necessity to study issues relative to equal opportunities of access to education and health in order to see if the social context tends to reinforce inequalities created by the dissolution of the Mao Tse Dong period egalitarian wage structure.

The study of intergenerational mobility of wages and education provides a first overview of these questions. Using the database, I had information on both parents' and child's social characteristics only for children still living with their parents. Hence, I needed to control for the co-residence selection process when focusing on educational mobility, and both co-residence and wage earner selections when studying wage mobility. The commonly used Heckman two stage procedure allowed me to control for the selection bias. In the situation of a double selection process, the broadening of this method considering bi-probit estimations helped me to take correlations between the two selection processes into account.

These technical issues being controlled for, I highlighted relatively high social mobility in China regarding both education and wages. However, worrying conclusions have been emphasized which could mitigate against this positive observation.

Firstly, parents' wages play an increasing role in the determination of the child's educational attainment. This tends to lower opportunities for children from poorer families having access to education and thus to better wages. Given that reasons supporting non linearities in the intergenerational mobility have been underlined by previous empirical studies, I also carried out non linear estimations using quantile regressions and semi-parametric estimations. The non linear results have highlighted even more the risk of lower mobility for poorer families showing the higher impact of parents' wages on the child's educational attainment for the extremes of the parents' income distribution. The U-shape relationship noted between children's and parents' wages tends to translate as lower mobility in the two extremes of the wage distribution. This could lead to polarization of the population creating an enlarged gap between the poor and the rich.

Secondly, results linked to the professional activities of parents underlined that children who come from farming families are penalised compared to their counterparts whose parents are wage earners. This can be due to unequal access to educational services as well as to the existence of a trade off for farming families between sending their children to school or keeping them at home to help in farm work.

The two major conclusions mentioned above from the analysis of intergenerational mobility stress potential higher future inequalities as the poor, with lower education will receive lower wages and will consequently find it harder to invest in their children's education (leading to a vicious circle), and the rural-urban gap will persist due to the lower mobility of children from farming origins.

The analysis of social inequalities in China completes the preceding analysis and confirms its worrying conclusions. In order to see the impact of the reform movement on the evolution of inequalities, I differentiate between three samples. The first included the whole population of individuals having completed their education and being wage earners. The second one was restricted to individuals between 26 and 33 years of age, who have been affected by reforms linked to the labour market, and partly by educational reforms and by the health system reforms. Finally, the sample of individuals less than 26 years old best represented individuals affected by the reform process in terms of wages, education and health reforms.

Even if no clear trend of evolution of these inequalities with time could be shown, the higher levels of both unidimensional and multidimensional inequalities of income, education and health obtained for the sub-sample of the youngest group (most affected by the reform process), demonstrate the deterioration of the situation since the reform movement. Concerning methodological issues, the Maasoumi indicator of multidimensional inequality gives results highly dependent on the degree of substitution between the dimensions considered. When the population's preferences move toward higher complementarity between dimensions (i.e. toward a more Rawlsian view of wellbeing), the evolution of inequalities of the dimension with the lower values (once normalization between dimensions has been carried out) dominates. For China, this implies that the evolution of multidimensional inequalities including income followed this last dimension's evolution (this attribute having lower values after normalization). Nevertheless, bi-dimensional inequalities of education and health showed interesting patterns as some of their evolutions were dominated neither by the evolution of education inequalities nor by the evolution of health inequalities. I have argued that this phenomenon occurs when different levels of significance in the correlations between dimensions exist. In this case, multidimensional aggregation, taking into account correlations between attributes, gives a more rigourous picture of wellbeing inequalities than the undimensional analysis of each dimension's inequalities.

The major conclusion of this thesis highlights the evolution of the social context underpinning Chinese society since the reforms. The rising returns to human capital that we have noted cannot be seen as a negative factor when equal opportunities in access to good education and health exist. However, the results relative to the evolution of inequalities both through lower mobility for poorer families and through the higher level of social inequalities for younger individuals tend to show that this access is unequal. Consequently, larger gaps will develop in the future.

Politicy makers will have to tackle these issues by trying to reduce inequalities in their various forms. Concerning geographical inequalities, principally between urban and rural areas, redistributive politics tools would help to lower gaps. Given the decentralization choice made by the government regarding the control of public health and education services, policies will have to be implemented at the local level and local government will, therefore, need financial support. As mentioned in the thesis, provincial budgets are negociated bilaterally between provinces and the central government, leading to inequalities in budget allocations. Consequently, in order to facilitate reforms leading to better access to education and good health, fiscal reallocations will probably be needed. Studies related to fiscal decentralization and its consequences on regional disparities have been available for a number of years. For instance, Yu and Tsui (2005) and Luo, Zhang, Huang, and Rozelle (2007) focus on regional disparities regarding fiscal policies and on the link between elections at the local level and the provision of public services. Given the size of the country, much work remains in order to include remote areas in economic development.

Continuing to describe and measure the evolution of the social conditions in China, research will also have to concentrate on measures the country can take to abandon the disequalizing path it has followed up until now. Studies on budget, fiscal and monetary policies which could help to deal with the mentioned issues are consequently needed.

Bibliography

APPLETON, S., L. SONG, AND Q. XIA (2005): "Has China Crossed the River? The Evolution of Wage Structure in Urban China During Reform and Retrenchment," *Journal of Comparative Economics*, 33(4), 644–663.

AQUILINO, W. (1990): "The Likelihood of Parent-Adult Child Coresidence: Effects of Family Structure and Parental Characteristics," *Journal of Marriage and the Family*, 52(2), 405–419.

ARON, D. J., AND P. OLIVELLA (1994): "Bonus and Penalty Schemes as Equilibrium Incentives Devices, with Application to Manufacturing Systems," *Journal of Law, Economics, Organization*, 10(1), 1–34.

ATKINSON, A., AND F. BOURGUIGNON (1982): "The Comparison of Multidimensioned Distributions of Economic Status," *Review of Economic Studies*, 49(2), 183–201.

BECKER, G. S., AND N. TOMES (1986): "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, 4(3), 1–39.

BEHRMAN, J., AND P. TAUBMAN (1985): "Intergenerational Earnings Mobility in the United States: Some Estimates and a Test of Becker's Intergenerational Endowments Model," *The Review of Economics and Statistics*, 67(1), 144–151.

BEHRMAN, J. R. (1993): "The Economic Rationale for Investing in Nutrition in Developing Countries," *World Development*, 21(11), 1749–1771.

Behrman, J. R., and M. R. Rosenzweig (2001): "The Returns to Increasing Body Weight," Penn Institute for Economic Research Working Paper 01-052.

Benjamin, D., L. Brandt, and J. Giles (2003): "The Evolution of Income Inequality in Rural China," Working Paper.

Bhalla, A. S., Y. Shujie, and Z. Zongyi (2003): "Causes of Inequality in China, 1952 to 1999," *Journal of International Development*, 15(8), 939–955.

Bhargava, A. (1997): "Nutritional Status and the Allocation of Time in Rwandese Households," *Journal of Econometrics*, 77(1), 277–295.

Bian, Y., R. Breiger, D. Davis, and J. Galaskiewicz (2005): "Occupations, Class, and Social Networks in Urban China," *Social Forces*, 83(4), 1443–1468.

Bian, Y., X. Shu, and J. Logan (2001): "Communist Party Membership and Regime Dynamics in China," *Social Forces*, 79(3), 805–841.

Blackorby, C., W. Bossert, and D. Donaldson (1999): "Income Inequality Measurement : The Normative Approach," *Handbook on Income Inequality Measurement*, pp. 133–161.

Bloom, G., and G. Xingyuan (1997): "Health Sector Reform: Lessons from China," *Social Science and Medicine*, 45(3), 351–360.

Boudon, R. (1974): *Education, Opportunity and Social Inequality.* New York: Wiley.

Bourguignon, F., F. Ferreira, and M. Menendez (2001): "Inequality of Outcomes, Inequality of Opportunities and Intergenerational Education," Research Project World Bank.

Bresson, F., and K. Labar (2007): ""Leftist", "Rightist" and Intermediate Decompositions of Poverty Variations with an Application to China from 1990 to 2003," ECINEQ WP 2007-76.

BUCHINSKY, M. (1995): "Quantile Regression, Box-Cox Transformation Model and the U.S. Wage Structure, 1963-1987," *Journal of Econometrics*, 65(1), 109–154.

BULLETIN, C. L. (2007): "Small Hands: A Survey Report on Child Labour in China," *China Labour Bulletin*, Report No.7: CLB Research Series.

CAO, Y. (2004): "Behind the Rising Meritocracy: Market, Politics and Cultural Change in Urban China," *Social Science Research*, 33(3), 435–463.

CARD, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.

CHECCHI, D., A. ICHINO, AND A. RUSTICHINI (1999): "More Equal But Less Mobile? Education Financing and Intergenerational Mobility in Italy and in the US," *Journal of Public Economics*, 74(3), 351–393.

CHEN, Y., S. DEMURGER, AND M. FOURNIER (2005): "Wage Differentials and Ownership Structure in Chinese Enterprises," mimeo, CERDI-CNRS, Université d'Auvergne.

CHENG, Y., AND J. DAI (1995): "Intergenerational Mobility in Modern China," *European Sociological Review*, 11(1), 17–35.

CHU NG, Y. (2007): "Gender Earnings Differentials and Regional Economic Development in Urban China, 1988-1997," *Review of Income and Wealth*, 53(1), 148–166.

COADY, D., AND L. WANG (2000): "Equity, Efficiency, and Labor-Market Reforms in Urban China : The Impact of Bonus Wages on the Distribution of Earnings," *China Economic Review*, 11(3), 213–231.

CONNELLY, C., AND Z. ZHENG (2003): "Determinants of School Enrollment and Completion of 10 to 18 Year Olds in China," *Economics of Education Review*, 22, 379–388.

CORAK, M., AND A. HEISZ (1999): "The Intergenerational Earnings and Income Mobility of the Canadian Men: Evidence from Longitudinal Income Tax Data," *The Journal of Human Resources*, 34(3), 504–533.

COUCH, K., AND D. LILLARD (1998): "Sample Selection Rules and the Intergenerational Correlation of Earnings," *Labour Economics*, 5(3), 313–329.

COUCH, K. A., AND D. R. LILLARD (2004): "Non-Linear Patterns of Intergenerational Mobility in Germany and the United States," in *Generational Income Mobility*, ed. by M. Corak. Cambridge University Press.

DAGENAIS, M. G., AND D. L. DAGENAIS (1994): "GMM Estimators for Linear Regression Models with Errors in Variables," Cahier 9404, Département de sciences économiques, Université de Montréal.

——— (1997): "Higher Moment Estimators for Linear Regression Models with Errors in the Variables," *Journal of Econometrics*, 76(1-2), 193–221.

DANIEL, C. I. (2004): "The Impact on China of its Ascension Into the WTO," *The Social Science Journal*, 41(3), 363–374.

DAVIS, D. (1992): ""Skidding": Downward Mobility Among Children of the Maoist Middle Class," *Modern China*, 18(4), 410–437.

DEATON, A. (1997): *The Analysis of Household Surveys. A Microeconomic Approach to Development Policy.* Johns Hopkins University Press, Baltimore, MD.

DEOLALIKAR, A. (1988): "Nutrition and Labour Productivity in Agriculture," *Review of Economics and Statistics*, 70(3), 406–413.

DOAK, C., L. ADAIR, M. BENTLEY, Z. FENGYING, AND B. M. POPKIN (2002): "The Underweight/Overweight Household: An Exploration of Household Sociodemographic and Dietary Factors in China," *Public Health Nutrition*, 5(1A), 215–221.

DONALD, S. A., AND W. K. NEWEY (2001): "Choosing the Number of Instruments," *Econometrica*, 69(5), 1161–1191.

DONG, X. Y. (2005): "Wage Inequality and Between-Firm Wage Dispersion in the 1990's: A Comparison of Rural and Urban Enterprises in China," *Journal of Comparative Economics*, 33(4), 664–687.

DONG, X. Y., AND P. BOWLES (2002): "Segmentation and Discrimination in China's Emerging Industrial Labor Market," *China Economic Review*, 13, 170–196.

DU, S., T. A. MROZ, F. ZHAI, AND B. M. POPKIN (2004): "Rapid Income Growth Adversely Affects Diet Quality in China - Particularly for the Poor," *Social Science and Medicine*, 59(7), 1505–1515.

DURLAUF, S. N. (1996): "A Theory of Persistent Income Inequality," *Journal of Economic Growth*, 1(1), 75–94.

ENGEL, J. W. (1984): "Marriage in the People's Republic of China: Analysis of a New Law," *Journal of Marriage and the Family*, 46(4), 955–961.

ERMISCH, J. (1999): "Prices, Parents, and Young People's Household Formation," *Journal of Urban Economics*, 45(1), 47–71.

ERMISCH, J., AND M. FRANCESCONI (2004): "Intergenerational Mobility in Britain: New Evidence from the British Household Panel Survey," in *Generational Income Mobility*, ed. by M. Corak. Cambridge University Press.

FAN, C. C. (2004): "The State, the Migrant Labour Regime, and Maiden Workers in China," *Political Geography*, 23(3), 283–305.

FANG, C., X. ZHANG, AND S. FAN (2002): "Emergence of Urban Poverty and Inequality in China : Evidence from Household Survey," *China Economic Review*, 13(4), 430–443.

FISHE, R., R. TROST, AND P. LURIE (1981): "Labour Force Earnings and College Choice of Young Women: An Examination of Selectivity Bias and Comparative Advantage," *Economics of Education Review*, 1(2), 169–191.

FLEISHER, B. M., AND X. WANG (2005): "Returns to Education in China under Planning and Reform," *Journal of Comparative Economics*, 33(2), 265–277.

FOGEL, R. W. (1994): "Economic Growth, Population Theory and Physiology," *American Economic Review*, 84(3), 369–395.

——— (2004): "Health, Nutrition and Economic Growth," *Economic Development and Cultural Change*, 52(3), 643–658.

FOSTER, A. D., AND M. R. ROSENZWEIG (1994): "A Test For Moral Hazard in the Labor Market — Contractual Arrangements, Effort, and Health," *Review of Economics and Statistics*, 76, 213–227.

FRANCESCONI, M., AND C. NICOLETTI (2004): "Intergenerational Mobility and Sample Selection in Short Panels," ISER Working Papers, Number 2004-17.

GILBERT, J., AND T. WAHL (2003): "Labor Market Distorsions and China's WTO Accession Package: An Applied General Equilibrium Assessment," *Journal of Comparative Economics*, 31(4), 774–797.

GRAWE, N., D. (2004): "Intergenerational Mobility for Whom? The Experience of High- and Low-Earning Sons in International Perspective," in *Generational Income Mobility*, ed. by M. Corak. Cambridge University Press.

GRAWE, N. D. (2006): "Lifecycle Bias in Estimates of Intergenerational Earnings Persistence," *Labour Economics*, 13(5), 551–570.

GREEN, L. W. (1988): "Promoting the One-Child Policy in China," *Journal of Public Health Policy*, 9(2), 273–283.

GROVES, T., Y. HONG, J. MCMILLAN, AND B. NAUGHTON (1994): "Autonomy and Incentives in Chinese State Enterprises," *The Quarterly Journal of Economics*, 109(1), 183–209.

GUSTAFSSON, B., AND L. SHI (2002): "Income Inequality Within and Across Counties in Rural China 1988 and 1995," *Journal of Development Economics*, 69(1), 179–204.

GUSTAFSSON, B., AND L. SHI (2004): "Expenditures on Education and Health Care and Poverty in Rural China," *China Economic Review*, 15(3), 292–301.

HADDAD, L. J., AND H. E. BOUIS (1991): "The Impact of Nutritional Status on Agricultural Productivity: Wage Evidence from the Philippines," *Oxford Bulletin of Economics and Statistics*, 53(1), 45–68.

HAN, D. (2001): "Impact of the Cultural Revolution on Rural Education and Economic Development: The Case of Jimo County," *Modern China*, 27(1), 59–90.

HANNUM, E. (1999): "Political Change and the Urban-Rural Gap in Basic Education in China, 1949-1990," *Comparative Education Review*, 43(2), 193–211.

HANNUM, E., AND M. WANG (2006): "Geography and Educational Inequality in China," *China Economic Review*, 17(3), 253–265.

HAUSER, S. M., AND Y. XIE (2005): "Temporal and Regional Variation in Earnings Inequality: Urban China in Transition Between 1988 and 1995," *Social Science Research*, 34(1), 44–79.

HENDERSON, G. E., J. S. AKIN, P. M. HUTCHINSON, S. G. JIN, J. M. WANG, J. DIETRICH, AND L. M. MAO (1998): "Trends in Health Services Utilization in Eight Provinces in China, 1989-1993," *Social Science and Medicine*, 47(12), 1957–1971.

HENNING, C., AND A. HENNINGSEN (2007): "Modeling Farm Households' Price Responses in the Presence of Transation Costs and Heterogeneity in Labor Markets," *American Journal of Agricultural Economics*, 89(3), 665–681.

HO, L. S. (1995): "Market Reforms and China's Health Care System," *Social Science and Medicine*, 41(8), 1065–1072.

HO, P.-T. (1959): "Aspects of Social Mobility in China, 1368-1911," *Comparative Studies in Society and History*, 1(4), 330–359.

HOLMLUND, H. (2006): "Intergenerational Mobility and Assortative Mating Effects of an Educational Reform," Working Paper 4/2006 Swedish Institute for the Social Research.

HOSSAIN, S. I. (1997): "Making Education in China Equitable and Efficient," *World Bank Policy Research Working Paper*, 1814.

HSIAO, C. (2003): *Analysis of Panel Data*, Econometric Society Monographs No. 34. Cambridge University Press, Cambridge, UK, second edn.

JAEGER, M., AND A. HOLM (2007): "Does Parents' Economic, Cultural, and Social Capital Explain the Social Class Effect on Educational Attainment in the Scandinavian Mobility Regime?," *Social Science Research*, 36(2), 719–744.

JONES, D., C. LI, AND A. L. OWEN (2003): "Growth and Regional Inequality in China During the Reform Era," *China Economic Review*, 14(2), 186–200.

JUSTINO, P. (2005): "Empirical Applications of Multidimensional Inequality Analysis," Poverty Research Unit at Sussex, Working Paper.

KHAN, A. R., AND C. RISKIN (1998): "Income and Inequality in China : Composition, Distribution and Growth of Household Income, 1988 to 1995," *The China Quarterly*, 154, 221–253.

KNIGHT, J., AND L. YUEH (2002): "The Role of Social Capital in the Labour Market in China," Department of Economics Discussion Paper Number 121, Oxford University.

——— (2004): "Job Mobility of Residents and Migrants in Urban, China," *Journal of Comparative Economics*, 32, 637–660.

KOLM, S. C. (1976): "Unequal Inequalities," *Journal of Economic Theory*, 12(3), 416–442.

LAM, D. (1999): "Generating Extreme Inequality: Schooling, Earnings, and Intergenerational Transmission of Human Capital in South Africa and Brazil," Research Report Population Studies Center, University of Michigan.

LAW, W. W. (2002): "Legislation, Education Reforms and Social Transformation: The People's Republic of China's Experience," *International Journal of Educational Development*, 22(6), 579–602.

LEVIN, J., AND E. PLUG (1999): "Instrumenting Education and the Returns to Schooling in the Netherlands," *Labour Economics*, 6(4), 521–534.

LI, H. (2003): "Economic Transition and Returns to Education in China," *Economics of Education Review*, 22(3), 317–328.

LINDBECK, J. M. H. (1951): "Communist Policy and the Chinese Family," *Far Eastern Survey*, 20(14), 137–141.

LIU, F. (2004): "Basic Education in China's Rural Areas: A Legal Obligation or an Individual Choice?," *International Journal of Educational Development*, 24(1), 5–21.

LIU, G. G., Z. ZHAO, R. CAI, T. YAMADA, AND T. YAMADA (2002): "Equity in Health Care Access to: Assessing the Urban Health Insurance Reform in China," *Social Science and Medicine*, 55(10), 1779–1794.

LIU, Z. (2001): "The Effects of Economic Reforms on Wage Inequality: Some Evidence from China," *Applied Economics Letters*, 8(4), 285–288.

LIU, Z. (2005): "Institution and Inequality: The Hukou System in China," *Journal of Comparative Economics*, 33(1), 133–157.

LOGAN, J., AND F. BIAN (1999): "Family Values and Coresidence with Married Children in Urban China," *Social Forces*, 77(4), 1253–1282.

LOGAN, J., F. BIAN, AND Y. BIAN (1998): "Tradition and Change in the Urban Chinese Family: The Case of Living Arrangements," *Social Forces*, 76(3), 851–882.

LUNDBERG, S., AND R. A. POLLAK (1996): "Bargaining and Distribution in Marriage," *The journal of Economic Perspectives*, 10(4), 139–158.

LUO, R., L. ZHANG, J. HUANG, AND S. ROZELLE (2007): "Elections, Fiscal Reform and Public Goods Provision in Rural China," *Journal of Comparative Economics*, 35(3), 583–611.

MAASOUMI, E. (1986): "The Measurement and Decomposition of Multidimensional Inequality," *Econometrica*, 54(4), 991–997.

MAURER-FAZIO, M., AND D. NGAN (2002): "Differential Rewards to, and Contributions of, Education in Urban China's Segmented Labor Market," processed.

MAURER-FAZIO, M., T. RAWSKI, AND Z. WEI (1999): "Inequality in the Reward for Holding Up Half the Sky: Gender Wage Gaps in China's Urban Labor Market, 1988-94," *The China Journal*, 41, 55–88.

MCELROY, M. B. (1990): "The Empirical Content of Nash-Bargained Household Behaviour," *The Journal of Human Resources*, 25(4), 559–583.

MENG, X., AND M. KIDD (1997): "Labor Market Reform and the Changing Structure of Wage Determination in China's State Sector During the 1980s," *Journal of Comparative Economics*, 25(33), 403–421.

MENG, X., AND J. ZHANG (2001): "The Two-Tier Labour Market in Urban China. Occupational Segregation and Wage Differentials Between Urban Residents and Rural Migrants in Shangai.," *Journal of Comparative Economics*, 29(3), 485–504.

MOK, K. H. (2000): "Marketizing Higher Education in Post-Mao China," *International Journal of Educational Development*, 20(2), 109–126.

Muller, C., and A. Trannoy (2003): "Multidimensional Inequality Comparisons : A Compensation Perspective," Working Paper.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46(1), 69–85.

Nilsson, T. (2007): "Measuring Changes in Multidimensional Inequality. An Empirical Application," Work in process, presented at the 2nd Meeting of the Society for the Study of Economic Inequality, Berlin, July 2007.

Oaxaca, R. (1973): "Male-Female Wage Differentials in Urban Labour Markets," *International Economic Review*, 14(3), 693–709.

Peng, X. (1987): "Demographic Consequences of the Great Leap Forward in China's Provinces," *Population and Development Review*, 13(4), 639–670.

Peng, Y. (1999): "Agricultural and Nonagricultural Growth and Intercounty Inequality in China, 1985-1991," *Modern China*, 25(3), 235–263.

Piketty, T. (2000): *Theories of Persistent Inequality and Intergenerational Mobility.*

Popkin, B. M. (1998): "The Nutrition Transition and its Health Implications in Lower-Income Countries," *Public Health Nutrition*, 1(1), 5–21.

Popkin, B. M., S. Paeratakul, K. Ge, and Z. Fengying (1995): "Body Weight Patterns Among the Chinese: Results from the 1989 and 1991 China Health and Nutrition Survey," *American Journal of Public Health*, 85(5), 690–694.

Quadrado, L., W. Heijman, and H. Folmer (2001): "Multidimensional Analysis of Regional Inequality : The Case of Hungary," *Social Indicators Research*, 56(1), 21–42.

Rawls, J. (1971): *A Theory of Justice*. Harvard University Press.

Renard, M. F. (2002): "A Pessimistic View on the Impact of Regional Inequalities," *China Economic Review*, 13(4), 341–344.

ROEMER, J., E. (1998): *Equality of Opportunity.* Harvard University Press.

ROEMER, J. E. (2002): "Does Democracy Engender Inequality?," Discussion paper Institute of Economics University of Copenhagen.

SAVAGLIO, E. (2001): "On Multidimensional Inequality : Ordering and Measurement," Center for Operations Research and Econometrics Working Paper.

SCHULTZ, T. P. (2001): "Productive Benefits of Improving Health: Evidence from Low-Income Countries," processed, Department of Economics, Yale University, March.

SCHULTZ, T. P., AND A. TANSEL (1997): "Wage and Labor Supply Effects of Illness in Côte d'Ivoire and Ghana: Instrumental Variable Estimates for Days Disabled," *Journal of Development Economics*, 53(2), 251–286.

SEEBORG, M. C., Z. J., AND Y. ZHU (2000): "The New Rural-Urban Labor Mobility in China: Causes and Implications," *Journal of Socio-Economics*, 29(1), 39–56.

SEMYKINA, A., AND J. M. WOOLDRIDGE (2005): "Estimating Panel Data Models in the Presence of Endogeneity and Selection: Theory and Application," unpublished working paper, Michigan State University.

SEN, A. K. (1973): *On Economic Inequality.* Clarendon Press, Oxford.

SHORT, S., Z. FENGYING, X. SIYUAN, AND Y. MINGLIANG (2001): "China's One-Child Policy and the Care of Children: An Analysis of Qualitative and Quantitative Data," *Social Forces*, 79(3), 913–943.

SICULAR, T., Y. XIMING, B. GUSTAFSSON, AND L. SHI (2007): "The Urban-Rural Income Gap and Inequality in China," *Review of Income and Wealth*, 53(1), 93–126.

SOLON, G. (1992): "Intergenerational Income Mobility in the United States," *The American Economic Review*, 82(3), 393–408.

SOLON, G. R. (1999): "Intergenerational Mobility in the Labour Market," in *Handbook of Labour Economics, Vol. 3A*, ed. by O. C. Ashenfelter, and D. Card. Amsterdam, North Holland.

STECKEL, R. H. (1995): "Stature and the Standard of Living," *Journal of Economic Literature*, 33(4), 1903–1940.

STRAUSS, J., AND D. THOMAS (1995): "Human Resources: Empirical Modeling of Household and Family Decisions," in *Handbook of Development Economics, Vol 3A*, ed. by J. Behrman, and T. Srinivasan. North Holand.

——— (1998): "Health, Nutrition and Economic Development," *Journal of Economic Literature*, 36(2), 766–817.

THOMAS, D. (1990): "Intra-Household Resource Allocation: An Inferential Approach," *Journal of Human Resources*, 26(4), 635–664.

THOMAS, D., AND J. STRAUSS (1997): "Health and Wages: Evidence on Men and Women in Urban Brazil," *Journal of Econometrics*, 77(1), 159–185.

TSUI, K. Y. (1998): "Factor Decomposition of Chinese Rural Income Inequality: New Methodology, Empirical Findings, and Policy Impications," *Journal of Comparative Economics*, 26(3), 502–528.

TSUI, K.-Y. (1999): "Multidimensional Inequality and Multidimensional Generalized Entropy Measures : An Axiomatic Derivation," *Social Choice and Welfare*, 16(1), 145–157.

UUSITALO, R. (1999): "Returns to Education in Finland," *Labour Economics*, 6(4), 569–580.

WALDER, A. G. (1989): "Social Change in Post-Revolution China," *Annual Review of Sociology*, 15, 405–424.

WAN, G., AND Y. ZHANG (2006): "The Impact of Growth and Inequality on Rural Poverty in China," *Journal of Comparative Economics*, 34(4), 694–712.

WANG, Y., K. GE, AND B. M. POPKIN (2000): "Tracking of Body Mass Index from Childhood to Adolescence: A 6-Year Follow-Up Study in China," *The American Journal of Clinical Nutrition*, 72(4), 1018–1024.

WEYMARK, J. A. (2004): "The Normative Approach to the Measurement of Multidimensional Inequality," Working Paper, Department of Economics, Vanderbilt University.

WORLDBANK (1997): "Financing Health Care," *China 2020 Issues and Options for China*.

YU, Q., AND K. TSUI (2005): "Factor Decomposition of Sub-Provincial Fiscal Disparities in China," *China Economic Review*, 16(4), 403–418.

ZHAI, F., D. FU, S. DU, K. GUE, C. CHEN, AND M. B. POPKIN (2002): "What is China Doing in Policy Making to Push Back the Negative Aspects of the Nutrition Transition," *Public Health Nutrition*, 5(1A), 269–273.

ZHANG, J., Y. ZHAO, A. PARK, AND X. SONG (2005): "Economic Returns to Schooling in Urban China, 1988-2001.," *Journal of Comparative Economics*, 33(4), 730–752.

ZHANG, L., A. D. BRAUW, AND S. ROZELLE (2004): "China's Rural Labor Market Development and its Gender Implications," *China Economic Review*, 15(2), 230–247.

ZHANG, X., AND R. KANBUR (2005): "Spatial Inequality in Education and Health Care in China," *China Economic Review*, 16(2), 189–204.

Social inequalities in China: which reality?

Résumé : *Au regard des distorsions sociales qui ont suivi le mouvement de réformes initié en Chine à la fin des années 1970, cette thèse étudie les évolutions récentes relatives à trois principaux sujets : le marché du travail, le niveau d'éducation et le niveau de santé. En utilisant la base de données CHNS, je considère dans un premier temps les rendements du capital humain en Chine, étant donnés les besoins en gains de productivité liés à une économie plus compétitive. Je souligne l'augmentation des rendements de l'éducation et de la nutrition en Chine depuis 1991, soulignant également l'impact des réformes sur la manière dont les salaires sont fixés aujourd'hui. Cette conclusion apparaît dans un premier temps positive pour l'augmentation de la productivité et de la croissance dans l'avenir. Cependant, si les individus ne bénéficient pas d'un égal accès à l'éducation et à la santé, une plus forte rémunération de ces facteurs peut mener à une détérioration en termes d'inégalités. C'est pourquoi je me focalise dans un deuxième temps sur l'évolution relative au niveau d'éducation et de santé depuis le mouvement de réformes à travers deux canaux : la possible transmission du statut social des parents à leurs enfants, i.e. la mobilité sociale ; puis les inégalités de bien-être dans trois dimensions sociales que sont le revenu, l'éducation et la santé. Grâce à l'utilisation de matrices de mobilité ainsi que de stratégies économétriques, je démontre un niveau de mobilité salariale et en terme d'éducation en Chine dans la moyenne en comparaison à celui d'autres pays développés ou en développement. Cependant, l'impact croissant du salaire des parents sur la scolarisation des enfants peut se traduire par une mobilité plus faible dans le futur, dans la mesure où cela renforce les dynamiques inégalitaires. Mis en parallèle avec les résultats de l'analyse multidimensionnelle des inégalités de bien-être, je conclus que les inégalités sociales en Chine sont amenées à augmenter dans les années à venir, nécessitant des mesures politiques en faveur de l'amélioration de l'accès à l'éducation et à la santé.*
Mots clés : Chine, salaires, éducation, santé, mobilité, inégalités.
Codes JEL : D31, D63, I12, I21, J24, J31, J62.

Abstract: *Given the social distortions born from the movement of reforms launched at the end of the 1970's in China, the present thesis investigates evolution regarding three major topics: labour market, the level of education and the health status. Using the CHNS database, I first consider returns to human capital in China, given the requirements of a more competitive economic environment in terms of productivity. I highlight increasing returns to education and nutrition in China since 1991, underlining the impact of the reforms on the way wages are now fixed. The conclusion appears positive for a rise in productivity and consequently for future growth. However, if individuals do not have equal access to education and to a good health, a higher payment of these factors can lead to deterioration in terms of inequalities. I subsequently focus on the evolution of the level of education and of the health status since the movement of reforms through two channels: potential transmission of the parents' social status to their children, i.e. social mobility; and inequalities in wellbeing in three social dimensions which are income, education and health. Using mobility matrices as well as econometric strategies, I demonstrate a comparably high level of mobility in education and wages as in other developed and developing countries. Nevertheless, the increasing impact of parents' wages on the child's schooling attainment can result in future lower mobility. Combined with results found from the analysis of multidimensional inequalities in wellbeing, I conclude that social inequalities in China are expected to increase in the coming years. Policy measures should address them focusing particularly on providing equal access to education and good health.*
Keywords: China, wages, education, health, mobility, inequality.
JEL codes: D31, D63, I12, I21, J24, J31, J62.