

Zhang, Xuelian

Working Paper

Long-term health and socioeconomic consequences of child labor: Evidence from Brazil

Working Paper, No. 2012-05

Provided in Cooperation with:

Department of Economics, Rutgers University

Suggested Citation: Zhang, Xuelian (2012) : Long-term health and socioeconomic consequences of child labor: Evidence from Brazil, Working Paper, No. 2012-05, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:

<https://hdl.handle.net/10419/94229>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Long-term Health and Socioeconomic Consequences of Child Labor: Evidence from Brazil*

Xuelian Zhang[†]

Rutgers University

First Version: October 2010

August 14, 2012

Abstract

Child labor may impose positive and negative, direct and indirect effects on the long-term development of an individual. This study employs the Brazil Living Standards Measurement Study Survey to examine the long-run consequences of child labor on an adult's income, health and educational attainment. When possible, the estimated models take into account of possible endogeneity and measurement error problems. The results suggest that early working has a negative and substantial income impact on rural residents and no impact on urban residents when controlling for schooling and health conditions. Child labor is associated with a worse self-assessed health index for rural adults and higher probability of getting health problems for both urban and rural adults. As for the schooling effect, the later one enters the labor market, the more years of schooling he obtains. While both the income and health impacts of child labor on rural residents are greater than those on urban residents, urban residents suffer greater adverse schooling impacts than their rural peers. My findings make a strong call to reduce child labor in Brazil and other developing countries. Moreover, the different effects of early working on urban and rural adults should be taken into account when child labor policies are proposed.

JEL Classification: J31, I10, I21

Keywords: Child labor, Long-term, Consequence, Brazil

*The author is grateful to Carolyn Moehling for her guidance and numerous helpful discussions. The author would also like to thank Roger Klein and Mark Killingsworth for their useful comments.

[†]Department of Economics, Rutgers, The State University of New Jersey, United States; Email: xzhang@econ.rutgers.edu

1 Motivation

Child labor is one of the most controversial issues in recent years. According to International Labor Organization (ILO)'s estimation (ILO, 2010), in 2008, there were about 215 million children aged 5 to 17 years old working worldwide, with approximately 115 million engaged in hazardous work. Such alarming figures have attracted substantial attention, leading to calls for actions to be taken to deal with child labor. Two of the most prominent international conventions concerning child labor are ILO Convention No. 138 which establishes the minimum age for children to enter the labor market and ILO Convention No. 182 which calls for the elimination of the worst forms of child labor¹.

Although sometimes child labor is an ethical issue seeming beyond discussion, identifying the long-run health and socioeconomic consequences of child labor is essential, since early entry into the labor force may affect an individual's income, health and education in his adulthood. However, much of the research on the consequences of early working emphasizes the short-run effects (see Graitcer and Lerer, 2000; Ray and Lancaster, 2005; Milcent, Huguenin and Carusi-Machado, 2005, for example), and the interactions between child work and adult health and socioeconomic status have not been widely explored. This is due to the limited data linking child work experience and adult outcomes.

Theoretically, one cannot determine the direction and magnitude of long-run effects of child labor, since the relationships between early working and an adult's health and socioeconomic conditions are complicated. On the one hand, early exposure to dangerous machinery, chemicals and heavy lifting during working may not only bring hazards to a child's body today but also retard her development in the long term; stress and negative emotions facing a young working child may lead to psychological problems which need time to manifest themselves. This illustrates the adverse impacts of early working on health. As for the schooling effects, early entry into the labor force may not only lower a child's schooling achievement (Psacharopoulos, 1997; Heady, 2003; Gunnarsson, Orazem

¹For the purpose of the convention, "child" refers to all persons under the age of 18. The term "the worst forms of child labor" comprises:

(a) all forms of slavery or practices similar to slavery, such as the sale and trafficking of children, debt bondage and serfdom and forced or compulsory labor, including forced or compulsory recruitment of children for use in armed conflict; (b) the use, procuring or offering of a child for prostitution, for the production of pornography or for pornographic performances; (c) the use, procuring or offering of a child for illicit activities, in particular for the production and trafficking of drugs as defined in the relevant international treaties; (d) work which, by its nature or the circumstances in which it is carried out, is likely to harm the health, safety or morals of children (ILO, 1999).

and Sánchez, 2006;), but also affect her decision to attend or drop school (Assaad, Levison and Zibani, 2001; Ray and Lancaster, 2005). On the other hand, however, child labor may positively affect adult health. It is well known that child labor is closely related to poverty, and usually child work is an important source of income for a poor household which could improve the living standard of the whole household including the child herself. This could lead to an improvement in the child's nutritional status and consequently help benefit her health in the long run. In addition, child labor could be an important way to finance education, either for the child herself or her siblings. Moreover, a young laborer could acquire work experience which could positively affect her future income. Therefore, how child labor decision affects an adult's education, health and income cannot be determined by theory and needs to be addressed empirically.

The objective of this study is to analyze the interactions between participation into the labor market during childhood and an adult's income, health and education level in Brazil. Data from Brazil Living Standards Measurement Study Survey (*Pesquisa Sobre Padrões De Vida*, PPV—1996/97) are employed. The key feature of this survey is that it asked each respondent at which age he started to work for the first time. This enables me to correlate early working to an adult's current health and socioeconomic conditions. However, instead of employing a dummy variable indicating whether the respondent was ever a child laborer, I include the age at which the individual started the first job in the model. This is because there's no agreement upon the definition of child labor, i.e., under which age we define a worker as a child laborer, and the results are sensitive to the definition of child labor if I include the dummy indicating whether the person ever worked in childhood. By including the variable age started to work, I can explore the long-run effects of one year earlier of entry into the labor market, and thus provide an implication about the effects of child labor.

I study the impacts of child labor on three dimensions of an adult's status: income, health and schooling. To my knowledge, few studies have ever looked at the multidimensional long-term consequences of child labor before. Researchers explore either just the linkages between early working and adult income (see Ilahi, Orazem and Sedlacek, 2001; Emerson and Souza, 2007), or the relationships between child labor and adult health (e.g., Kassouf, Mckee and Mossialos, 2001; Lee and Orazem, 2009). Beegle, Dehejia, and Gatti (2005) examine the income, health and schooling effects of child labor in Vietnam, but their findings are limited to examining the outcomes only 5 years after child working and thus cannot provide the consequences of early working in the long-term view. Few existing

literature provide a full picture of the long-run impacts of child labor, because most of the analyses to date pertain to different samples, different data sets and even different countries. However, this paper complements the existing literature by studying the long-term income, health and schooling effects of child labor simultaneously with the same data set from Brazil. This enables the analysis to be done in the same framework and provides a coherent story about the possible long-term consequences if one enters the labor market early in her childhood. This is useful for policy discussion, when the policy makers determine whether and the extent to which we should reduce the incidence of child labor in Brazil, this study provides a good reference about the multidimensional long-term effects of early working. In addition, I try to deal with the possible endogeneity problem by applying the instrumental variable method which helps to explore the causal effects of early working. Furthermore, as will be shown below, while most of the previous studies analyzing child labor in Brazil pool the urban and rural samples, I find great distinctions of child labor effects on urban and rural residents, which suggests the care needed for implementing child labor policies in different areas.

The present study finds that one year later entry into the labor market is associated with higher incomes, better self-assessed health indexes for rural adults and lower probabilities of getting health problems for both urban and rural adults. As for the schooling effect, the later one enters the labor market, the more years of schooling he obtains. While both the income and health impacts of child labor on urban residents are smaller than those on rural residents, urban residents suffer greater adverse schooling impacts than their rural peers. The main findings for the health and schooling impacts of early working are consistent when either the working sample (including workers with valid income information) or the full sample (including all individuals with valid, missing or zero income data) is employed for estimation.

The rest of this paper is organized as follows. Section 2 takes a brief survey of recent literature about the linkages between child labor and schooling, health and income. Section 3 provides information about the data and some descriptive statistics. Section 4 describes the empirical strategy employed in this study. Section 5 presents the main results. Section 6 provides robustness checks. Section 7 provides conclusions and policy implications.

2 Child Labor in Literature

My research is built upon a growing literature about the short-term and long-term consequences of child labor. A selective review of literature would help to identify which have been analyzed by scholars and which remain to be explored.

2.1 Child Labor and Schooling

Most of the current literature on child labor and schooling focuses on the relationships between early working and the contemporaneous schooling attendance and educational achievement.

Some evidence suggests that early entry into the labor force is negatively correlated to school attendance. Psacharopoulos (1997) observes that child labor makes working children receive 2 fewer years of schooling than their non-working peers in Venezuela. Based on data from Ghana in the late 1980's, Boozer and Suri (2001) conclude that there is a significant trade-off between working and attending school: one more hour of child work is associated with 0.38 fewer hour of school attendance. Assaad, Levison and Zibani (2001) also find a strong association between early working and dropping out of school in Egypt.

But some other researchers argue that working and schooling can be compatible rather than exclusive to each other (Ravallion and Wodon, 2000). One possible explanation is, within an extremely poor household, parents may not be able to afford the schooling fees, but if the child can work part time to support part of her tuition then her likelihood of attending school would be higher. As a matter of fact, a child's working may even help to support her siblings' education (Weiling, 2003).

The weight of evidence suggests that cognitive attainment from schooling is lower for working children, probably because working takes up part of the children's time and leaves children tired and less able to study effectively. Heady (2003) explores the linkages between early working and children's learning achievement with GLSS2 data set from Ghana. He suggests that working outside the household adversely affects children's results on reading and mathematics tests. Gunnarsson, Orazem and Sánchez (2006) estimate that child labor reduces math and language scores by 7.5% and 7% respectively, on 3th and 4rd graders in 11 Latin American Countries.

2.2 Child Labor and Adult Health

The long-run interactions between child labor and health are likely to be dynamic, in the sense that health status is more likely to be correlated to working history, rather than current work activity, since many health risks caused by early working need time to manifest themselves. For example, children exposed to dangerous materials like asbestos or molten glass may not suffer adverse health effects today, but they may get cancer or other serious health problems twenty years later. The stress or negative emotions facing young laborers today may not have an immediate impact, but lead to depression or other psychological problems in their later life. On the other hand, however, the long-run health consequences of early working can be positive as well, because child work may be crucial to an extremely poor household, and children's income contributions to the family may improve their living standards and nutritional status, and hence impose a positive impact on the long-run health development.

Previous research has examined the long-term health consequences of child labor. Kasouf, Mckee and Mossialos (2001) using a Brazilian data set find that the likelihood an individual reports less than good health in adulthood rises as the age of entry the labor market falls. Based on an analysis of the Brazil PNAD data set, Lee and Orazem (2009) argue that early entry into labor market and decreasing schooling time jointly increase the probability of reporting physical ailments in adulthood.

2.3 Child Labor and Adult Income

Early exposure to work may affect a child worker's future income through human capital investment. Education provides skills that raise an individual's productivity and in turn raise his earnings. Therefore, how early working affects a child's education will have a link with his future income. Also, if child work leads to physical injury or psychological stress which may survive through adulthood, or if there is any health benefit arising from the improved nutritional status or living standard owing to young worker's income, such a health effect will affect future earnings in adulthood. In addition, when a child works early in his life, he is able to accumulate working experience which may have pecuniary benefits.

The linkages between child labor and subsequent labor market outcomes have been examined empirically but still many questions remain. Ilahi, Orazem and Sedlacek (2001) explore a national survey in Brazil (PNAD) and find that early exposure to child labor significantly reduces adult earnings and gives rise to an increase in the probability of

being in poverty. However, their study doesn't take into account the possible endogeneity problem: there may exist some unobservable factors like an individual's ability that affect both the child labor decision and income. Hence, Ilahi, Orazem and Sedlacek's work only provides suggestive results. Also based upon the PNAD data set from Brazil, Emerson and Souza (2007) examine whether child labor is harmful to an individual in terms of adult earnings. They employ the GMM IV method to address possible endogeneity and provide a disturbing result: child labor significantly reduce adult earnings for males even after controlling for schooling. However, Emerson and Souza do not control for the individual's health which may affect her income, and it would be of interest to consider whether early working could affect an adult's income when controlling for both schooling and health status, i.e., whether child labor has an income impact other than through its impacts on education and health. In addition, Emerson and Souza only focus on the income impact and omit other impacts of child labor. As will be shown below, my work complements their study by analyzing multidimensional consequences of early entry into the labor market.

2.4 Child Labor in Brazil

There is a long tradition of child labor in Brazil. The first registered child labor dates back to the 16th century, when children helped adults extract pau-brasil (the native Brazilian tree) (Ferreira, 2001). Along with the industrialization in the 20th century, there existed a great demand for child labor, and child employment became very serious through the whole century. According to Moura (1982), in 1912, 30% of the labor force in the four major textile factories was made up of children and adolescents, and this proportion even increased to 40% by 1919.

Although a sharp decrease of child labor occurred in the second half of 1990s, owing to the government's efforts to reduce child labor (such as "Bolsa Escola" which is a cash transfer program conditional on school attendance), there are still a large number of children involved in working. According to the estimation of Instituto Brasileiro de Geografia e Estatística (IBGE, 2007), there are about 5.4 million children aged between 5 and 17 years old in the labor market, of whom 40.7% are under 14 years old despite of the prohibition of child laborer younger than 14 years old from Federal Constitution of Brazil. Among those working children between 5 and 17 years old, one third work 40 hours or more per week. To be more specific, 13.6% of the 10-14 age group and almost one half of the 15-17 age group work more than 40 hours per week (ICFTU, 2004).

3 Data and Descriptive Analysis

The main data used for analysis come from the Living Standards Measurement Study Survey (*Pesquisa Sobre Padrões De Vida*, PPV—1996/97) of Brazil. The PPV was undertaken by the Instituto Brasileiro de Geografia e Estatística (IBGE) and the World Bank jointly from March 1996 to March 1997.

The PPV covered information from urban and rural areas in Northeast and Southeast of Brazil. The living standards in the Northeast Region are the lowest, while the Southeast is the richest region in Brazil. Hence the PPV provided two typical regions with respect to living standards and employment in Brazil. Although the PPV was restricted in geographic coverage, it was representative of about 70% of the national population in Brazil (Turra, 2000). The survey interviewed 4940 households in total, collecting detailed information about household composition, migration, education, health, economic activity, fertility, etc. The key feature of the data set is that it asked each respondent about the age at which he started working² and the working sector of his first job³. The PPV provided detailed information about an individual's health and socioeconomic status as well as his first job, meeting the requirement for my research question. It is noteworthy that the PPV has been little explored in this direction. While most studies on child labor in Brazil employ the Pesquisa Nacional por Amostra de Domicílios (PNAD) (e.g., Ilahi, Orazem and Sedlacek, 2001; Emerson and Souza, 2007; Lee and Orazem, 2009) data set, my analysis using the PPV could supplement the literature.

However, one potential weakness is that the information about a person's first job comes from a recall question and may be subjected to recall errors. I would prefer to have a longitudinal survey in which the same persons are followed from their childhood to adulthood, as well as more detailed information about their first jobs are interviewed, like the working hours, working conditions, working and schooling, etc. But such data are rare, especially in the developing economies where child labor is prevalent.

The sample is composed of individuals aged between 18 and 55 years old with valid

²A person who has worked previously is understood as someone who:

(1) has exercised an economic activity paid in money, merchandise, products or only in benefits (housing, food, clothing, etc.);

(2) has exercised an economic activity with no payment for at least 1 hour per week for the purpose of helping a member of the household unit who has an economic activity, or as an apprentice, trainee, etc. (IBGE, 1997).

³The main working sectors include: agriculture, services, manufacturing, construction, textile, transportation and some other industries.

information on earnings, health and education. I restrict the analysis to individuals old than 18 years since I want to analyze the impacts of early working on adults, and also to individuals younger than 55 years old, since 55 is the retirement age in Brazil and most people older than 55 don't have regular monthly earnings. I restrict the sample to people with valid information on earnings, schooling and health conditions. Furthermore, my sample is selected to include people who entered the labor market between 5 and 31 years old, those people who started to work younger than 5 years old or older than 31 years old are treated as outliers. This does not greatly reduce the number of observations compared to the original survey data, since 98.91% of people started to work at this age range⁴. The sample size is 3901 after this selection process⁵. Owing to the distinct differences in urban and rural areas, all analyses are conducted separately for these two areas, with 3235 and 666 individuals respectively.

The summary information for the variables is presented in Table 1 with the definitions of variables shown in Table A.1 at the appendix. Urban people typically have higher monthly earnings than rural residents⁶. As for the self-assessed health index, it equals 1 if the individual rates his own health condition as "poor" or "average", equals 2 if the individual rates health as "good" and equals 3 if the individual rates health as "very good" or "excellent"⁷. From the summary statistics, urban and rural residents report very close and high health index: about 2.3, implying people evaluate their health conditions as more than good on average. In both urban and rural areas about two out of ten adults report to have health problems. Noticeably, there is a big gap of education levels between the urban and rural sample. While the average years of schooling in the urban sample is 8.33 years, implying people in urban areas complete upper primary education⁸, rural adults receive less than 5 years of education. On average, urban residents entered the labor market at 15

⁴My results are robust to modest changes in the range of age started to work.

⁵The sample size is also reduced due to the unavailability of data for instrumental variables in some years.

⁶While rural residents earn about 281 Reais each month on average, the average income of urban residents is 658 Reais, more than twice of that of rural adults.

⁷The categories "poor" and "average" are combined because few individuals report poor health. The categories "very good" and "excellent" are combined because few rural individuals report excellent health.

⁸Brazilian education system: primary education (1st grau) consists of 1st-8th grade; high school education (2nd grau) consists of 9th-11th grade; undergraduate education typically consists of 4 years of schooling; graduate education differs according to degrees and fields. In this paper, I split the primary education into lower primary (1st-4th grade) and upper primary (5th-8th grade) education as Emerson and Souza (2007) do.

years old, almost 3 years later than rural residents did. The composition of race for the urban and rural samples are quite similar. As for parental education levels, individuals in rural areas typically have less educated parents than those in urban areas.

Figure 1 presents the distribution of age started to work for the urban and rural samples. Note that rural individuals typically entered the labor market earlier than urban individuals. The most common ages of labor market entry for urban individuals are 10, 12, 14, 15 and 18 years old, while the common ages to start working for rural persons are 7, 8, 10, 12 and 15 years old.

Figure 2 and 3 show the average of log-earnings and the average of the self-assessed health indexes by the age of labor market entry for urban and rural individuals respectively. In these figures, I collapse the individuals who started to work before 7 years old into one group and after 20 years old into another group, since from Figure 1, it is noteworthy that there are very few individuals starting their first jobs before 7 or after 20 years old, 91% of my sample entered the labor market between 7 and 20 years old. Figure 2 and 3 exhibit roughly linear relationships between log-earnings and starting age, and between the health index and starting age respectively.

The averages of years of schooling by age started to work are presented in Figure 4a. Again, the individuals who started working before 7 or after 20 years old are collapsed into two groups separately. The increase in the years of schooling associated with the increasing starting age is notable. However, this trend becomes complicated when I consider different quantiles of years of schooling. From Figure 4b, it is clear that the distributions of years of schooling differ conditional on different starting ages, for both urban and rural residents. This suggests that quantile regression is necessary for analysis.

4 Methodology

I examine the impacts of child labor on adult income, health and schooling separately.

4.1 Child Labor and Adult Income

The model to be used for analyzing the long-run effect of child labor on adult income is:

$$m \ln inc_i = \pi_1 + \alpha_1 * startage_i + \beta_1 * sch_i + \gamma_{1a} * ehealth_i + \gamma_{1b} * ghealth_i + x'_{1i} \delta_1 + \varepsilon_{1i} \quad (1)$$

where $m \ln inc$ is the log of monthly income, $startage$ is the age at which the person

started the first job⁹, sch is years of schooling, $ehealth$ and $ghealth$ are the dummies for reporting "excellent health" and "good health" respectively, and x_1 is a vector of exogenous variables, including age, age-squared, gender, race, GDP per capita of the individual's residence state at the interviewing year, and parental education levels which are a proxy of the individual's family background.

It is likely that a person's decision to work, years of schooling and current health status are correlated to the unobserved components of income in model (1). For instance, an individual with higher ability tends to achieve higher level of education and earn higher income (biasing β_1 upward); an individual's unobserved health endowment is not only correlated to her current health status, but also affects her earnings and in turn biases the health coefficient. As for the coefficient on the age started to work, higher ability may lead to later entry into the labor market for an individual since she has the capacity to acquire higher levels of schooling (biasing α_1 upward). A higher unobserved health endowment, nonetheless, makes the child more likely to be sent into the labor market early and thus biases α_1 downward. Meanwhile, measurement error is another source of potential bias, making the directions of biases for the coefficients unpredictable. My data on people's age started to work come from a recall question, and thus would probably be subjected to recall bias. The possible measurement error in the self-reported health status will be discussed in the next section.

One way to address the possible endogeneity and measurement error is to employ instrumental variable technique. Variables qualified to be instruments must be sufficiently correlated to people's decisions to enter the labor market and schooling and current health conditions, but not correlated to the unexplained components of income.

One possible set of instruments for the age of labor market entry and years of schooling include parental occupations, the availability and quality of local education systems and the economic conditions in local labor markets when the individual was a child. Parikh and Sadoulet (2005) argue that children of employers or self-employed persons are more likely to work than children of employees. Thus parental occupations could affect people's decisions to work in his childhood. The weight of evidence suggests that school quality is an important determinant of an individual's schooling decision (Alderman, Orazem and Paterno, 2001; Bedi and Edwards, 2002), and the conditions in local labor markets will

⁹I include the linear specification of age started to work because the relationship between log-earnings and age started to work is roughly linear from the raw data. I also tried the model including startage and startage-squared, and got qualitatively similar results.

affect the supply of and demand for child laborers directly. Therefore, the instruments I use for age started to work and schooling are the parental occupations when the individual was 15 years old (denoted by z_1), the number of teachers per school for the state where the individual lived when he was 7 and 11 years old, since age 7 and 11 are the typical ages for a child to enter the lower primary and upper primary school in Brazil¹⁰, and the GDP per capita of the state where the individual lived when he was 12 years old (denoted by z_2), since age 12 is the legal minimum employment age in Brazil until 1988¹¹. The choice of instruments about the state-level schooling and labor market conditions is guided by Emerson and Souza (2007) and Lee and Orazem (2009)¹². But they all apply the birth state's information for instruments, while in the present study, I take advantage of both

¹⁰Here, "typical" means the individual enters school at 7 years old and there's no delaying or repeating of grades.

¹¹85.41% of my sample consist of individuals whose current state of residence is the birth state. I assume that these individuals were not migrants and thus the birth state's figures of teachers and GDP per capita are used as instruments.

For an individual whose current residence state is not the birth state, I can identify the last state he lived before moving to the current state. If the last state the individual lived before he moved to current state was the birth state, then I assume that he just migrated from the birth state to the current state directly and migrated once in total; if the last state the individual lived was not the birth state, then he migrated at least twice in total.

I can identify how long an individual lived in the birth state. If the individual lived in the birth state for longer than 11 years, then no matter how many times he migrated, all the instruments use the birth state's information.

For those people who migrated from the birth state to the state of current residence directly, if he lived in the birth state for 11 years, then the state GDP per capita when the individual was 12 years old employs the figure from the current state and all the other instruments use figures from the birth state; if the individual lived in the birth state for longer than 6 years but shorter than 11 years, then the number of teachers per school when the individual was 11 years old and the state GDP per capita when the individual was 12 years old use figures from the current state while the rest instruments use figures of the birth state; if the individual lived in the birth state for shorter than 7 years, then all the instruments employ figures of the state of current residence.

For those people who migrated at least twice and lived in the birth states for shorter than 12 years, I cannot determine in which states they lived before coming to the current state and when they came to the current state, so information from the birth state are used as instruments for simplicity. However, there are only 33 individuals (0.85% of the whole sample) migrating at least twice and living in the birth states for shorter than 12 years, such a small portion of the sample should not affect my main results. I run a robustness check in which I give these persons current states' information as instruments, the results are quantitatively and qualitatively similar.

¹²For the instruments of child labor and schooling decisions, Emerson and Souza (2007) employ the number of schools per 1000 children and the number of teachers per school in the birth state when the individual was 7, 11 and 15 years old, and the birth state's GDP per capita when the individual was 12 years old; and Lee and Orazem (2009) employ the number of schools per 1000 children and the number of teachers per 1000 children in the birth state when the individual was 7 years old, and the state-specific average wage rates for low-skilled workers in the year when the individual was 12 years old.

the birth state's and the current living state's data to construct the instruments which makes my instruments more informative. It is a challenge to find instruments which are exogenous to the unexplained components of income¹³. Once controlling for the family background and the current labor market conditions, parental occupations, the variations of schooling quality and local labor market conditions when the individual was a child should be uncorrelated to the error term and satisfy the exclusion restriction.

Considering the instruments for health condition, I use the availability and quality of local health systems when the individual was a child which are represented by the numbers of hospitals, beds and doctors per 1000 inhabitants of the state where the individual lived when she was 7 years old (denoted by z_3). Controlling for all regressors, including the family background and current local labor market conditions, the availability and quality of local health systems when the individual was a child should not have independent influence on adult earnings. Furthermore, as I will demonstrate below, the relevance of the instruments for child labor and schooling decisions as well as the health conditions are checked through the tests of excluded instruments in the first-stage regressions, and the validity of instruments are checked through overidentification tests in the second stage.

Here, data on the parental occupations come from the PPV survey directly. Data on the number of schools and teachers, the number of hospitals, beds and doctors by state and year come from the IBGE online resource "Statistics of the 20th Century"¹⁴. And data on the GDP and population by state and year are taken from the IPEA historical series¹⁵.

To estimate the income model (1), I first run OLS regressions, and then employ 2SLS technique to rule out the possible endogeneity and measurement error. Comparing coefficients from OLS and 2SLS tells us the direction and magnitude of bias in the coefficients due to the endogeneity problem and measurement error, if any¹⁶.

¹³Some people may argue that there may exist a persistency of occupation across generations and this causes parental occupations to be correlated to the unexplained components of income. However, I think that such intergenerational persistency of occupation mainly comes from the effect of parental education upon kids, and after controlling for parental education and other covariates in the model, parental occupations should satisfy the exclusion restriction.

¹⁴These series are available on line at <http://www.ibge.gov.br/seculox/default.shtm> (accessed on 09/11/2010).

¹⁵These series are available on line at <http://www.ipeadata.gov.br/ipeaweb.dll/ipeadata?65370046> (accessed on 09/11/2010).

¹⁶All models in this paper allow for clustering on the birth year and state.

4.2 Child Labor and Adult Health

I employ two health models with two health indicators: one is the self-assessed health index and the other is the incidence of health problems. Self-reported health status has been shown to be a good proxy for a person's true health condition. Kalpan and Camacho (1983) and Mcgee et al. (1999) find persistent associations between self-reported health ratings (like poor, fair, good, excellent, etc.) and mortality, and self-reported health status is a strong prognostic indicator for subsequent mortality. Miilunpalo et al. (1997) reinforce this view and further show that the perceived health is inversely associated with the number of physician contacts per year. However, measurement error may exist in the studies employing self-reported health measures, since how people evaluate his health may depend on his education level, working status, etc. An individual with a higher education level is more likely to take care of himself and may have more information on his health condition. By examining the relationship between a self-reported health measure and a simulated clinical measure with the tetrachoric correlation coefficient, Butler et al. (1987) find the existence of biased reporting. In particular, non-working persons are more likely to report incorrect health conditions, probably due to the need of justification of unemployment.

The purpose for employing two health indicators is to mitigate the possible measurement error in the self-assessed health index since the incidence of health problems is relatively more accurate and objective than the health index. However, the incidence of health problems may not reflect the overall health condition as the health index does. Furthermore, I could check the consistency of results across models with two health measures.

The first health measure is an individual's self-reported health index. This ordered measure comes from the individual's self-assessed health status, that is, respectively, poor or average ($chealth=1$), good ($chealth=2$) and very good or excellent ($chealth=3$). I use the following ordered probit model to capture how the child labor decision affects an individual's health in her adulthood:

$$chealth = \begin{cases} 1 & \text{if } health^* \leq \zeta_1 \\ 2 & \zeta_1 < health^* \leq \zeta_2 \\ 3 & health^* > \zeta_2 \end{cases}$$

$$health_i^* = \alpha_{2a} * startage_i + \beta_{2a} * sch_i + x'_{1i} \delta_{2a} + z'_{3i} \eta_{2a} + \varepsilon_{2ai} \quad (2a)$$

When the latent health status variable $health^*$ crosses a cutoff point, the observed category of the health index changes. x_1 and z_3 consist of the same variables as in the income model (1).

The incidence of health problems is used as the second health measure. Formally, the health problems reported in this survey include flu/cold/pneumonia, infection, accident/injury, digestive problem, pain, infarction and some other problems. Child labor may adversely (or positively) affect the incidence of health problems by affecting a young laborer's health capital and making him more (or less) likely to get a health problem in adulthood. And a probit model is employed to estimate the effect of child labor on the incidence of health problems here:

$$hproblem = \begin{cases} 1 & \text{if } hproblem^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$hproblem_i^* = \pi_{2b} + \alpha_{2b} * startage_i + \beta_{2b} * sch_i + x'_{1i}\delta_{2b} + z'_{3i}\eta_{2b} + \varepsilon_{2bi} \quad (2b)$$

where $hproblem$ implies the incidence of health problems and $hproblem^*$ is the latent variable. x_1 and z_3 are the same sets of variables as in the income model (1).

However, I am still faced with an endogeneity problem caused by unobservable health endowments. Only healthy children are qualified for employment which induces a positive relationship between health endowment and early working. Additionally, measurement error still remains a problem and makes the direction of bias on the coefficient estimates unpredictable. As a result, an IV ordered probit specification for model (2a) and an IV probit specification for model (2b) are employed to rule out potential bias and investigate the causal health effects of early working. The instruments for age started to work and years of schooling consist of parental occupations when the individual was 15 years old (z_1), the number of teachers per school for the state where the individual lived when she was 7 and 11 years old and the GDP per capita for the state where the individual lived when she was 12 years old (z_2)¹⁷. These instruments should be correlated to a person's child labor and schooling decisions, but uncorrelated to her unobservable health endowments, once her demographic characteristics, family background and the current labor market economic conditions are controlled for. Again, I will test the relevance and validity of

¹⁷The procedure to construct instruments and the data source of instruments here are the same as described in the previous section.

instruments via tests of excluded instruments in the first stage and overidentification tests in the second stage respectively.

4.3 Child Labor and Adult Schooling

In this study, an adult's education level is captured by the years of schooling he obtained. We already notice from Figure 4b that the effects of the child labor decision on achieved education level are quite different for different quantiles of years of schooling. Therefore, I estimate quantile regression. The standard linear conditional quantile regression model treats the conditional distribution of the response variable as a linear function of covariates. To be more specific, let $Q_q(sch|x)$ denote the q th standard linear conditional quantile function of the response variable years of schooling given covariates x (including age started to work, x_1 , z_2 and z_3). Then for the q th quantile ($0 < q < 1$), the model can be written as:

$$Q_q(sch_i|x_i) = \pi_{3q} + \alpha_{3q} * startage_i + x'_{1i}\delta_{3q} + z'_{2i}\sigma_{3q} + z'_{3i}\eta_{3q} = x'_i\theta_q \quad (3)$$

Note that the parameters θ_q (including π_{3q} , α_{3q} , δ_{3q} , σ_{3q} , η_{3q}) are allowed to vary across quantiles. The q th quantile regression estimator $\hat{\theta}_q$ minimizes over θ_q the objective function

$$\sum_{i: y_i \geq x'_i\theta} q |sch_i - x'_i\theta_q| + \sum_{i: y_i < x'_i\theta} (1 - q) |sch_i - x'_i\theta_q|$$

where $0 < q < 1$. In this study, I estimate the schooling equation (3) at quantiles 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9.

Although one's early working decision may be correlated to the unexplained components of schooling, I don't have valid instruments for age started to work, hence I will just report results from OLS and quantile regressions¹⁸.

¹⁸I tried the parental occupations when the individual was 15 years old as instruments for the child labor decision in the schooling model, but they cannot pass the overidentification test, indicating the invalidity of these instruments.

5 Empirical Results

5.1 Child Labor and Adult Income

Classical analyses of income models estimate separate models for men and women, since usually there exist substantial gender differences on the wage effects. However, in this study, I test and can not reject the null hypothesis of the equality of all coefficients (except the intercepts) in the male and female models under the 5% significance level, in either the urban or the rural sample. Hence, I pool the men and women samples and include a gender dummy in the income model (1).

In order to estimate the effects of early labor market entry on current adult earnings, I begin by treating the child labor decision, education and health as exogenous¹⁹. Table 2 reports the OLS coefficient estimates of the income model (1). The control variables are age started to work, years of schooling, self-reported health status, demographic factors including age, age-squared, gender and race, parental education levels and the GDP per capita of individual's current residence state at the interviewing year. When I treat the individual's child labor decision, education and health conditions as exogenous, whether an individual worked during childhood does not affect her current income, holding other factors constant. This is true for both the urban and rural adults. The more educated the person, the more income she earns. And an urban resident's health condition is positively correlated to her earnings.

As described above, the 2SLS method is employed to deal with the possible issues of endogeneity and measurement error. I use parental occupations when the individual was 15 years old, the number of teachers per school of the state where the individual lived when he was 7 and 11 years old and the GDP per capita of the state where the individual lived when he was 12 years old, as well as the number of hospitals, beds and doctors per 1000 inhabitants of the state where the individual lived when he was 7 years old as instruments. The first-stage regression results for the urban and rural samples are presented in Table 3a and 3b. For the age started to work, years of schooling and dummies for health condition, the F test of excluded instruments all indicate the joint significance of instruments. Males enter the labor market about two years earlier, receive one year less of schooling and have

¹⁹I firstly estimate the income equation by quantile regression method with and without considering the endogeneity problem for quantiles 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. I cannot reject the null hypothesis that the coefficients of age started to work across quantiles are equal to each other. Therefore, OLS and 2SLS coefficient estimates are reported as the final estimates of the income equation (1).

higher probability to report excellent health. Compared to white people, urban black and other-raced individuals start working earlier and receive less education, and report almost the same health status. Parental occupations do affect child labor decision: compared to individuals whose parents were employees, individuals with fathers who did not work or were self-employed or mothers who worked without a payment when the individual was 15 years old, enter the labor market at younger ages. This is consistent with our expectation. When father doesn't work or mother is unsalaried, the household may face a credit constraint and need the child to work to supplement the household income. Besides, when father is self-employed, such as working on the own farm or factory, the child may need to enter the labor market early to help his father.

Table 4 presents the second-stage regression results of the income equation. I cannot reject the null hypothesis of overidentification test of all instruments, indicating the validity of my instruments for child labor decision, years of schooling and health condition, for both the urban and rural samples. Early entry into the labor market has no significant impact on adult earnings for urban residents, but has a negative and substantial income impact for rural residents, after controlling for the schooling level and health condition. Entering the labor market one year later increases monthly earnings by 16.7% for a rural resident, which is indeed a sizable effect. An early rural labor market entrant suffers a lower income during adulthood since early working may adversely affect the schooling quality which will in turn impose a negative impact on adult income.

The comparison between the OLS and 2SLS estimates is also of interest. While in neither the OLS or 2SLS models is the estimated income impact of child labor significantly different from zero for the urban sample, the OLS estimate of starting age lies below the 2SLS estimate for the rural sample. This implies that the possible endogeneity and measurement error bias the effect of early working on adult earnings downward.

The coefficients other than age started to work have the expected signs. For an urban individual, the higher of education level, the more income he earns, and excellent health brings in higher income than poor health. Income rises as he ages, probably owing to the accumulation of working experience, but the return to aging falls. There is a gender gap in earnings: males typically get higher earnings than female workers. And when the GDP per capita of the residence state increases indicating a better macroeconomic environment, individuals get higher earnings.

5.2 Child Labor and Adult Health

I use model (2a) and (2b) to estimate the long-term impacts of early entry into the labor force on adult health. The self-assessed health index and the incidence of health problems are the dependent variables, and the controls include the individual's age, age-squared, gender, race, parental education levels, the GDP per capita of the residence state at the interviewing year, the number of health facilities per 1000 inhabitants of the state where the individual lived when she was 7 years old²⁰.

I start by estimating the ordered probit model (2a) without considering the possible endogeneity problem. The marginal effects rather than the coefficient estimates of the model (2a) are reported in Table 5. Column 1, 2, and 3 correspond to the marginal effects on the probability that the health index equals 1, 2 and 3 for the urban residents, while column 4, 5 and 6 correspond to the marginal effects for the rural residents. From Table 5, there is no significant effect of child labor on adult health for the urban sample, but in the rural areas, an early labor market entrant is less likely to report very good or excellent health, and more likely to report poor, average or good health.

When I take the endogeneity and measurement error issues into account, an IV ordered probit specification of the health model (2a) is estimated, with the parental occupations when the individual was 15 years old (z_1), the availability and quality of local education system and the fluctuations of local labor market represented by z_2 being employed to identify the child labor decision and years of schooling. Table 6 presents the first-stage regression results. The first two columns of Table 6 correspond to the starting age and schooling equations for the urban sample, while the last two columns are the first-stage estimates for the rural sample. From those results, we can find that the instruments are correlated to the age of entry into the labor market and schooling decision and jointly significant.

The second-stage estimates of the relationship between the health index and child work activity are presented in Table 7. Column 1, 2 and 3 correspond to the marginal effects on

²⁰I separate the health models into men and women and test the equality of all coefficients (except the intercepts) of the male and female models, for the urban and rural samples respectively. I cannot reject the null hypothesis of the equality of all coefficients (except the intercepts) of male and female models for the urban sample under the 5% significance level, indicating that that it's not necessary to separate the urban health model into men and women. And I reject the null for the rural sample. But this is not a strong indication of different models for men and women, since there are only 464 and 202 observations for the rural male and female samples respectively, the rural male and female models are poorly estimated. Hence, I pool men and women for the rural health model for simplicity.

the probability that the health index equals 1, 2 and 3 for the urban residents, and column 4, 5 and 6 correspond to the marginal effects for the rural residents. The comparison between the ordered probit and IV ordered probit estimates indicate that the endogeneity causes the estimated effect of child labor to be smaller than it really is. Results from Table 7 suggest that there is no significant health consequence for an urban early labor market entrant, but there exists a significant negative effect of early working on rural adult health: the probability of reporting very good or excellent health falls and the probability of reporting poor or average health rises as one enters the labor force earlier.

One thing worthy of notice is the opposite effects of schooling on the health status of urban and rural adults. An increase in the schooling years benefits urban adult health but harms rural adult health, due to the twofold impacts of education. On the one hand, as one achieves higher level of education, she would gain access to more knowledge about health care, and usually would take care of herself more carefully, which is expected to have a positive impact on health condition. On the other hand, the more educated of an individual, the more likely she would pay attention to her own health condition and to recognize and report health problems, which suggests the negative effect of schooling on reported health condition. Back to my sample, from the descriptive analysis in Section 3, urban residents receive 3.5 more years of schooling than rural residents on average. Hence, when rural residents are relatively low educated on average, one additional year of schooling may be more effective on recognizing and reporting health problems leading to a negative effect of schooling on reported health condition, while the average urban residents finish the upper primary education, the impact of schooling may focus more on getting access to knowledge about health care and consequently has a positive effect on reported health. Additionally, the probability to report very good or excellent health decreases as persons age and urban men report being healthier than women.

An additional probit model with the incidence of health problems as the dependent variable is estimated. The probit estimates, the first-stage and second-stage regression results of the IV probit specification of model (2b) are presented in Table 8, 9 and 10 respectively. The main results from health model (2b) are consistent with those derived from model (2a): child labor is associated with worse adult health. However, although Table 7 shows child labor only affects future health in the rural sample, Table 10 exhibits significant negative linkages between early working and adult health for both the urban and rural samples. One year earlier of entry into the labor force leads to an increase of 2.1% and 9.1% in the probability of getting health problems in adulthood, for an urban and

rural resident respectively. The probit estimates of child labor are centered over zero for both urban and rural samples, compared to the IV estimates, again, implying the existence of endogeneity. Urban males are about 9% less likely to get health problems than females.

While it's clear that an early labor market entrant suffers worse health outcomes, I find big area differences of child labor effects: whether an individual worked as a child laborer does not affect how he evaluates his health condition in urban areas but does in rural areas; meanwhile, as one enters the labor market one year earlier, the probability of getting health problems increases by 2.1% and 9.1% for an urban and rural resident respectively.

These substantial area differences in the health impacts of early exposure to work may be due to the different working environments and conditions for the first job in the urban and rural samples. In the sample, among those rural residents who started to work before 18 years old, more than 70% were employed in the agriculture sector, while among those urban residents who entered the labor market as a child, most of them worked in the service (24%), retailing (15%), manufacturing and construction industry (19%). As one may notice that the agriculture sector is ranked as one of the most hazardous sectors in terms of morbidity and mortality (Fassa, 2000). Fassa (2000) points out that children employed in the agriculture sector are easily injured by dangerous machinery, exposure to strenuous labor, chemicals and adverse weather (e.g. heat). Also, agriculture is among the less regulated sectors where the laws protecting children are very difficult to enforce. In contrast, child workers in the urban sample who are involved in manufacturing, retailing, services and other industries may not suffer from as adverse working conditions and poorly regulated working environments as do their rural peers.

5.3 Child Labor and Adult Schooling

From Figure 4b, the distributions of years of schooling conditional on starting ages are quite different. Consequently, for both the urban and rural samples, I use the quantile regression approach to capture the associations between early working and schooling. Model (3) is employed to explore the relationship between child labor and adult schooling²¹. I report

²¹I separate the schooling quantile regression models into men and women for the urban and rural samples, and test the equality of all coefficients (except for the intercepts) of the male and female models. In most (except for the 30th quantile in the urban sample) quantile regressions for either the urban or the rural sample, I cannot reject the null hypothesis of the equality of all coefficients (except the intercepts) of the male and female models.

I then estimate the urban schooling model for the 30th quantile, with the sample being separated into men and women. The schooling effects of early working from the male and female models are quite close

OLS estimates first, followed by quantile regression results. A major difference between OLS and quantile regression is that OLS characterizes the mean of the distribution whereas the quantile regression explores the full shape of the conditional distribution of the dependent variable. Relatively speaking, quantile regression provides a more precise estimation and a more complete picture of the conditional distribution of years of schooling.

As explained above, although there may exist some unobservable factors affecting both child labor decision and educational attainment, I do not have suitable instruments that are sufficiently correlated to the child labor decision but not correlated to the unexplained components of schooling, hence my estimates of the impacts of child labor on adult schooling are suggestive but not causal.

Table 11a and 11b exhibit the OLS estimates first, followed by the quantile regression estimates of the schooling model at the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th and 90th conditional percentiles, for the urban and rural samples respectively. Of great interest are the coefficients on the age an individual started to work. These parameters estimate the changes in specific conditional percentiles of years of schooling caused by one unit change in the starting age. The child labor decision imposes significant negative effects on an adult's educational attainment for almost all the quantiles. The later one enters the labor market, the more years of schooling she attains. The marginal changes associated with one year later of entry into the work force in the median conditional quantile of years of schooling are an increase of 0.284 years and 0.178 years, for the urban and rural samples respectively. The coefficients of age started to work vary considerably across quantiles. For instance, there is an 80 percent difference between the starting age coefficients for the 0.5 quantile and 0.1 quantile in the urban sample, while in the rural sample the starting age coefficient for the 0.7 quantile is close to 134 percent above that of the 0.2 quantile. I also test and reject the equality of coefficients of age started to work across quantiles. Most noticeably, starting age has much greater impacts at the middle conditional quantile of schooling for the urban sample and at 0.7 quantile of schooling for the rural sample than those in tails of the schooling distribution. As a matter of fact, quantile regression estimates exhibit an inverse U-shaped trend which rises over the percentiles until around middle quantile and then falls for the urban sample, suggesting that the schooling effect of early working is greater for an individual with middle level of education than one in the tails of the schooling distribution. A similar trend can be found in the rural sample

to each other. Hence, I pool men and women for the schooling model for simplicity, for both the urban and rural samples.

estimates.

Another thing worthy of notice is that child labor imposes greater effects on schooling for urban residents than rural residents (except for the 90th percentile). For some quantiles of years of schooling (e.g., the 10th and 20th quantiles), the impact of child labor on urban residents is twice or even more than twice that on rural residents. Similar to the health model, this large area difference may be due to the different working environment and conditions for the first job in the urban and rural samples. Most rural residents were involved in agricultural work in the first job, and agricultural work is often seasonal work and may be more compatible with schooling than working in urban areas.

Clearly, the quantile regression estimates are different from the OLS estimates. According to the linear regression model, an urban individual's schooling level would increase by 0.227 years if he started working one year later. However, the quantile regression results indicate larger impacts of child labor on the 30th, 40th, 50th, 60th and 70th quantiles of schooling years for the urban sample. For instance, entering the labor market one year later causes the 50th conditional quantile of schooling to increase about 0.284 years for a urban resident. Similar results can be found in the rural sample: the linear regression model underestimates the effects of child labor at the 40th, 50th, 60th, 70th, 80th and 90th quantiles of years of schooling.

Male workers typically receive less schooling than female workers. Urban black and other-raced residents' educational attainments are lower than those of white people. Parental education levels impose positive effects on the individual's schooling level, the more educated of parents, the higher schooling level the individual gets.

6 Robustness Check

All the above results indicate the adverse long-term consequences of early working on adult earnings, health and educational attainment. However, all the analyses are based on the working sample with valid income information, and therefore may be potentially biased. In this section, I re-estimate the health model and the schooling model using the full sample including all individuals with valid and non-valid (missing or zero) income data to work as a robustness check²².

Look at the health model (2a) first. The ordered probit estimates and the first-stage

²²My income model is probably subjected to the sample selection bias. However, sample selection is not the main research problem of interest in the present study, so I do not correct for it here.

regression estimates for the IV ordered probit specification with the full sample can be found in Appendix Table A.2 and Table A.3. My instruments are jointly significant in the first-stage regressions. I report the second-stage regression results for the IV ordered probit specification in Table 12. Age started to work cannot be statistically differentiated from zero in the urban full sample. And in the rural case, one year later of entering the labor market reduces the probability for an individual to assess her health as "good" at a very small magnitude but increases her probability to report "very good or excellent" health significantly. This is consistent with our main finding from Table 7: early working does not affect an urban resident's health but imposes an adverse health effect on a rural resident.

Table 13 presents the second-stage regression results from the full sample for the IV probit specification of the health model (2b)²³. My instruments are jointly significant in the first stage and pass overidentification tests in the second stage. Again, early working exhibits great adverse health effects: as one enters the labor market one year earlier, the probability of reporting health problems in adulthood rises by 1.9% and 5.7% for an urban and rural individual respectively. A great difference of child labor effects between the urban and the rural sample shows up again: while early working does not affect how one evaluates his health condition in urban areas but does in rural areas, one year earlier entry into the labor market increases the probability of reporting health problems much more greatly for a rural resident than an urban resident.

Table 14a and 14b exhibit the OLS and quantile regression results from the full sample for the schooling model. A comparison between Table 11 and 14 shows that the coefficients of age started to work estimated from the full urban sample are close to those estimated from the working urban sample, and the coefficients of age started to work have the same signs for the working and full sample in the rural case, though there exist some differences in magnitudes. The fact that the adverse schooling impacts of child labor are greater on urban adults than their rural peers is true for most quantiles of schooling (except for the 90th percentile) in both the working and the full samples.

In either the health model or the schooling model, most coefficients other than age started to work have the same signs for the working sample and the full sample, and for those estimates which have opposite signs, most of them are insignificant, although there exist some differences in the magnitudes of the coefficients of the working and full samples,

²³The probit estimates and the first-stage regression results for the IV probit specification of the health model are shown in Appendix Table A.4 and Table A.5.

especially in the rural case.

In sum, the story we get from the full sample is consistent with the one from the working sample: child labor negatively affect adult health and schooling, and early entry into the labor market imposes different effects on the urban and rural residents.

7 Discussion and Policy Implication

This study investigates in great detail the long-run effects of working as a child laborer on an individual's health and socioeconomic conditions. It explores the Brazilian PPV data set and analyzes the long-term income, health and schooling effects of early working for the urban and rural samples separately. In order to deal with the possible endogeneity and measurement error problems, I employ instruments to estimate the income model (1) and the health model (2a) and (2b), with the 2SLS method used for the income model, the IV ordered probit method used for the health model (2a) and the IV probit method used for the health model (2b). However, due to the fact that it is too difficult to find a suitable instrument which can decompose the effect of child labor decision from schooling decision, my work does not take into account of the endogeneity problem in the schooling model. The quantile regression technology is used to capture the different effects of early working on schooling across quantiles. It would be informative to find a valid instrument for child labor decision in the schooling model, to examine the causal relationship between early working and adult schooling.

The results presented in this study suggest that early exposure to work for a rural resident leads to lower earnings when controlling for schooling and health conditions, and a worse self-assessed health index when controlling for schooling. Also, an urban/rural adult has a higher probability to get health problems if she worked during childhood. As for schooling, the later one enters the labor market, the more years of schooling she obtains. While both the income and health effects of child labor on rural residents are greater than those on urban residents, urban residents suffer greater adverse schooling impacts than their rural peers. Although early working may help young laborers to accumulate working experience and finance the household or schooling, the combined final effects of child labor on a person's future development are negative, i.e., a child who starts to work early suffers adverse health and socioeconomic consequences in the long run. The findings for the health and schooling models are robust when the full sample including all individuals with valid and non-valid (missing or zero) income information are used for estimation.

My findings have important implications: all the aforementioned negative effects of child labor on adult outcomes make a strong call to reduce child labor in Brazil and other developing countries. In addition, the different effects of early working on urban and rural adults should be taken into account when child labor policies are proposed. Given that rural children are more vulnerable to the adverse consequences of early working at many aspects, we should pay special attention to tackle the issue of child labor in the rural area.

References

- [Alderman:2001] Harold Alderman, Peter F. Orazem, and Elizabeth M. Paterno. School quality, school cost, and the public/private school choices of low-income households in pakistan. *The Journal of Human Resources*, 36(2):304–326, 2001.
- [Assaad:2001] Ragui Assaad, Deborah Levison, and Nadia Zibani. The effect of child work on school enrollment in egypt. *Economic Research Forum, Working Papers, No.111*, 2001.
- [Bedi:2002] Arjun Singh Bedi and John H. Y. Edwards. The impact of school quality on earnings and educational returns—evidence from a low-income country. *Journal of Development Economics*, 68:157–185, 2002.
- [Beegle:2005] Kathleen Beegle, Rakeev Dehejia, and Roberta Gatti. Why should we care about child labor? the education, labor market, and health consequences of child labor. *World Bank Policy Research Working Paper 3479*, 2005.
- [Boozer:2001] Michael A. Boozer and Tavneet K. Suri. Child labor and schooling decisions in ghana. *Working Paper, Yale University*, 2001.
- [Butler:1987] J. S. Butler, Richard V. Burkhauser, Jean M. Mitchell, and Theodore P. Pincus. Measurement error in self-reported health variables. *The Review of Economics and Statistics*, 4:644–650, 1987.
- [Emerson:2007] Patrick M. Emerson and Andre Portela Souza. Is child labor harmful? the impact of working earlier in life on adult earnings. *IZA Discussion Paper No. 3027*, 2007.

- [Fassa:2000] Anaclaudia G. Fassa, Luiz Augusto Facchini, Marinel M. DallaAgnol, and David Christiani. Child labor and health: Problems and perspectives. *International Journal of Occupational and Environmental Health*, 6(1):55–62, 2000.
- [Ferreira:2001] Eleanor Stange Ferreira. Trabalho infantil: historia situacao atual. *Canoas: ULBRA*, 2001.
- [Graitcer:2000] PL Graitcer and LB Lerer. The impact of child labor on health. *Report of a Field Investigation in Egypt, Washington DC, World Bank*, 2000.
- [Gunnarsson:2006] Victoria Gunnarsson, Peter F. Orazem, and Mario A. Sanchez. Child labor and school achievement in latin america. *The World Bank Economic Review*, 20(1):31–54, 2006.
- [Heady:2003] Christopher Heady. The effect of child labor on learning achievement. *World Development*, 31(2):385–398, 2003.
- [IBGE:1997] IBGE, DPE, and DEPIIS. Interviewer manual. *Living Standards Measurement Study Survey*, 1997.
- [IBGE:2007] IBGE. Child labor 2001. *IBGE*, 2007.
- [ICFTU:2004] International Confederation of Free Trade Unions. Internationally recognised core labour standands in brazil. *Report for the WTO General Council Review of the Trade Policies of Brazil*, 2004.
- [ILO:1999] International Labour Organisation. Convention 182: Convention on the worst forms of child labour. *Geneva, ILO*, 1999.
- [ILO:2010] International Labour Organisation. Accelerating action against child labour. *Report of the Director-General, International Labour Conference, 99th session*, 2010.
- [Ilahi:2001] Nadeem Ilahi, Peter F. Orazem, and Guilherme Sedlacek. The implications of child labor for adult wages, income and poverty: Retrospective evidence from brazil. *World Bank, unpublished working paper*, 2001.
- [Kalpan:1983] George A. Kalpan and Terry Camacho. Perceived health and mortality: A nine year follow-up on the human population laboratory cohort. *American Journal of Epidemiology*, 117:292–298, 1983.

- [Kassouf:2001] Ana Lucia Kassouf, Martin Mckee, and Elias Mossialos. Early entrance to the job market and its effect on adult health: Evidence from brazil. *Health Policy and Planning*, 16(1):21–28, 2001.
- [Lee:2009] Chanyoung Lee and Peter F. Orazem. Lifetime health consequences of child labor in brazil. *Research in Labor Economics, Forthcoming*, 2009.
- [McGee:1999] Daniel L. McGee, Youlian Liao, Guichan Cao, and Richard S. Cooper. Self-reported health status and mortality in a multiethnic us cohort. *American Journal of Epidemiology*, 149(1):41–46, 1999.
- [Miilunpalo:1997] Pekka Oja Matti Pasanen Seppo Miilunpalo, Ilkka Vuori and Helka Urponen. Self-rated health status as a health measure: The predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *Journal of Clinical Epidemiology*, 50(5):517–528, 1997.
- [Milcent:2005] Carine Milcent, Jack Huguenin, and Danielle Carusi-Machado. Children in brazil: Health, education and work. *Anais do XXXIII Encontro Nacional de Economia [Proceedings of the 33th Brazilian Economics Meeting]*, ANPEC, No. 173, 2005.
- [Moura:1982] Esmeralda Blanco B. de. Moura. Mulheres e menores no trabalho industrial: os fatores sexo e idade na dinamica do capital. *Petropolis: Vozes*, 1982.
- [Parikh:2005] Anokhi Parikh and Elisabeth Sadoulet. The effect of parents’ occupation on child labor and school attendance in brazil. *UC Berkeley: Department of Agricultural and Resource Economics, UCB. CUDARE Working Paper No. 1000.*, 2005.
- [Psacharopoulos:1997] George Psacharopoulos. Child labor versus educational attainment: Some evidence from latin america. *Journal of Population Economics*, 10(4):377–386, 1997.
- [Ravallion:2000] Martin Ravallion and Quentin Wodon. Does child labor displace schooling? evidence on behavioral responses to an enrollment subsidy. *The Economic Journal*, 110:C158–C175, 2000.

- [Ray:2005] Ranjan Ray and Geoffrey Lancaster. The impact of childrens work on schooling: Multi-country evidence based on simnoc data. *International Labour Review*, 144(2):189–210, 2005.
- [Turra:2000] Cassio M. Turra. Contabilidade das geracoes: riqueza, sistema de transferencias e consequencias de mudancas no padrao demografico. *Master's thesis, CEDEPLAR/UFMG*, 2000.
- [Weiling:2003] Elizabeth Weiling. Do returns on investment for educating children in oaxaca de jurez, mexico, pay off? a qualitative analysis. *International Journal of Qualitative Studies in Education*, 16:817–834, 2003.

Table 1. Summary Statistics

Variables	Urban (N=3235)				Rural (N=666)			
	Mean	Std. Dev	Min	Max	Mean	Std. Dev	Min	Max
mlninc	5.910	0.993	2.303	10.003	5.178	0.831	2.708	9.297
chealth	2.368	0.693	1.000	3.000	2.224	0.702	1.000	3.000
ehealth	0.491	0.500	0.000	1.000	0.383	0.486	0.000	1.000
ghealth	0.385	0.487	0.000	1.000	0.458	0.499	0.000	1.000
phealth	0.124	0.329	0.000	1.000	0.159	0.366	0.000	1.000
hproblem	0.194	0.396	0.000	1.000	0.194	0.395	0.000	1.000
sch	8.325	4.132	0.000	21.000	4.856	3.581	0.000	17.000
startage	15.001	4.376	5.000	31.000	12.285	4.337	5.000	31.000
age	32.511	9.585	18.000	55.000	30.937	9.241	18.000	55.000
agesq	1148.805	662.261	324.000	3025.000	1042.369	628.206	324.000	3025.000
male	0.571	0.495	0.000	1.000	0.697	0.460	0.000	1.000
white	0.486	0.500	0.000	1.000	0.434	0.496	0.000	1.000
black	0.068	0.252	0.000	1.000	0.071	0.256	0.000	1.000
otherrace	0.446	0.497	0.000	1.000	0.495	0.500	0.000	1.000
currqdp	5.908	2.871	1.670	10.649	5.469	2.745	1.670	10.649
fschmiss	0.289	0.453	0.000	1.000	0.458	0.499	0.000	1.000
filliterate	0.029	0.168	0.000	1.000	0.063	0.243	0.000	1.000
flowprimary	0.466	0.499	0.000	1.000	0.414	0.493	0.000	1.000
fupprimary	0.098	0.297	0.000	1.000	0.029	0.167	0.000	1.000
fhighschool	0.072	0.258	0.000	1.000	0.023	0.148	0.000	1.000
fcollge	0.047	0.212	0.000	1.000	0.014	0.116	0.000	1.000
mschmiss	0.324	0.468	0.000	1.000	0.514	0.500	0.000	1.000
milliterate	0.030	0.171	0.000	1.000	0.062	0.241	0.000	1.000
mloprimary	0.437	0.496	0.000	1.000	0.353	0.478	0.000	1.000
mupprimary	0.099	0.298	0.000	1.000	0.036	0.187	0.000	1.000
mhighschool	0.086	0.280	0.000	1.000	0.030	0.171	0.000	1.000
mcollege	0.024	0.153	0.000	1.000	0.006	0.077	0.000	1.000
fnwork	0.114	0.318	0.000	1.000	0.059	0.235	0.000	1.000
femployee	0.508	0.500	0.000	1.000	0.535	0.499	0.000	1.000
fselfemploy	0.274	0.446	0.000	1.000	0.357	0.480	0.000	1.000
femployer	0.054	0.227	0.000	1.000	0.020	0.138	0.000	1.000
focumiss	0.050	0.218	0.000	1.000	0.030	0.171	0.000	1.000
mnwork	0.564	0.496	0.000	1.000	0.506	0.500	0.000	1.000
memployee	0.249	0.433	0.000	1.000	0.224	0.417	0.000	1.000
mselfemploy	0.122	0.327	0.000	1.000	0.117	0.322	0.000	1.000
memployer	0.010	0.099	0.000	1.000	0.002	0.039	0.000	1.000
munsalary	0.043	0.204	0.000	1.000	0.135	0.342	0.000	1.000
mocumiss	0.011	0.106	0.000	1.000	0.017	0.128	0.000	1.000
age12_gdp	4.099	3.204	0.333	12.983	4.042	2.967	0.440	12.983
age7_teacher	5.271	4.235	1.543	27.243	5.327	3.923	1.543	23.968
age11_teacher	6.060	4.636	1.566	27.390	6.088	4.339	1.566	23.968
age7_hospital	0.039	0.014	0.007	0.121	0.042	0.015	0.007	0.121
age7_bed	3.378	1.733	0.107	7.868	3.407	1.677	0.335	7.295
age7_doctor	0.809	0.616	0.050	3.006	0.796	0.565	0.050	2.934

Notes: There are only 2 individuals reporting father's occupation type as "unsalaried", thus I collapse the type "unsalaried" and "self-employed" for father's occupation.

Table 2: OLS of Income Model

Variables	mlninc_u	mlninc_r
startage	-0.002 (0.004)	0.008 (0.007)
sch	0.097*** (0.004)	0.093*** (0.012)
ehealth	0.189*** (0.042)	0.094 (0.077)
ghealth	0.087** (0.042)	0.120 (0.084)
age	0.096*** (0.010)	0.074*** (0.021)
agesq	-0.001*** (0.000)	-0.001*** (0.000)
male	0.617*** (0.026)	0.662*** (0.062)
black	-0.131*** (0.050)	-0.014 (0.123)
otherrace	-0.066** (0.028)	0.041 (0.060)
curr GDP	0.061*** (0.005)	0.035*** (0.011)
fschmiss	-0.069* (0.036)	0.014 (0.071)
filliterate	-0.071 (0.066)	-0.156 (0.118)
fupprimary	0.021 (0.048)	0.151 (0.135)
fhighschool	0.193*** (0.060)	0.120 (0.258)
fcollge	0.330*** (0.077)	0.639* (0.369)
mschmiss	0.008 (0.033)	-0.070 (0.065)
milliterate	0.066 (0.067)	0.052 (0.119)
mupprimary	0.056 (0.045)	-0.121 (0.140)
mhighschool	0.105* (0.054)	-0.087 (0.281)
mcollege	0.041 (0.092)	-0.014 (0.543)
Constant	2.227*** (0.179)	2.430*** (0.386)
Observations	3,235	666
R-squared	0.479	0.318

Notes: Clustered robust standard errors are reported in parentheses.

*** Significant at 1% level** Significant at 5% level, * Significant at 10% level.

Table 3a: IV Estimates - First-stage Regression of Income Model (1) for Urban Sample

Variables	startage_u	sch_u	ehealth_u	ghealth_u
age	0.337*** (0.085)	0.505*** (0.060)	-0.012 (0.011)	0.021** (0.010)
agesq	-0.005*** (0.001)	-0.006*** (0.001)	0.000 (0.000)	-0.000* (0.000)
male	-1.986*** (0.152)	-0.877*** (0.116)	0.057*** (0.017)	-0.028* (0.017)
black	-0.619** (0.248)	-1.128*** (0.240)	-0.053 (0.036)	0.054 (0.038)
otherrace	-0.433*** (0.156)	-1.071*** (0.129)	-0.029 (0.021)	0.032 (0.022)
currgdp	-0.215*** (0.046)	-0.223*** (0.040)	0.008 (0.005)	-0.002 (0.005)
fschmiss	-1.025*** (0.202)	-1.183*** (0.174)	-0.075*** (0.023)	0.051** (0.024)
filliterate	-1.254*** (0.455)	-2.232*** (0.390)	-0.052 (0.051)	0.049 (0.052)
fupprimary	0.396 (0.249)	0.903*** (0.208)	0.021 (0.032)	-0.002 (0.031)
fhighschool	1.089*** (0.272)	1.528*** (0.251)	0.043 (0.039)	-0.031 (0.036)
fcollge	1.695*** (0.343)	2.572*** (0.278)	0.077 (0.048)	-0.014 (0.046)
mschmiss	-0.783*** (0.179)	-2.049*** (0.158)	-0.030 (0.022)	0.004 (0.021)
milliterate	-1.136*** (0.437)	-1.596*** (0.356)	-0.103* (0.053)	0.037 (0.055)
mupprimary	0.905*** (0.231)	1.040*** (0.208)	0.021 (0.028)	-0.006 (0.029)
mhighschool	1.487*** (0.288)	1.962*** (0.198)	0.055 (0.035)	-0.017 (0.033)
mcollege	1.593*** (0.515)	2.955*** (0.388)	0.150** (0.060)	-0.122** (0.060)
focumiss	-0.178 (0.343)	0.082 (0.287)	0.053 (0.043)	-0.053 (0.042)
fnwork	-0.710*** (0.219)	0.273 (0.204)	0.050* (0.029)	-0.036 (0.028)
fselfemploy	-0.399** (0.181)	0.016 (0.154)	0.020 (0.021)	-0.027 (0.021)
femployer	-0.190 (0.347)	0.850*** (0.260)	0.019 (0.040)	-0.019 (0.039)

(Continued) Table 3a: IV Estimates - First-stage Regression of Income Model (1) for Urban Sample

Variables	startage_u	sch_u	ehealth_u	ghealth_u
mocumiss	0.617 (0.609)	0.599 (0.509)	0.131 (0.088)	-0.121 (0.086)
mnwork	0.689*** (0.161)	0.649*** (0.140)	0.042** (0.020)	-0.010 (0.021)
msselfemploy	0.063 (0.256)	0.504** (0.205)	0.066** (0.030)	-0.063** (0.030)
memployer	-0.348 (0.803)	0.479 (0.543)	0.161** (0.081)	-0.070 (0.073)
munsalary	-1.651*** (0.448)	-0.453 (0.341)	0.033 (0.046)	-0.046 (0.047)
age12_gdp	0.079 (0.063)	0.284*** (0.048)	0.010 (0.008)	-0.012 (0.007)
age7_teacher	-0.049 (0.051)	-0.042 (0.039)	-0.001 (0.008)	0.001 (0.007)
age11_teacher	-0.017 (0.047)	-0.075* (0.040)	0.003 (0.005)	-0.004 (0.005)
age7_hospital	8.895 (9.075)	7.492 (6.350)	-1.301 (1.103)	2.300** (1.064)
age7_bed	-0.148 (0.138)	-0.053 (0.107)	0.026* (0.015)	-0.016 (0.015)
age7_doctor	1.267*** (0.441)	0.641** (0.279)	-0.173*** (0.047)	0.149*** (0.042)
Constant	11.146*** (1.660)	-0.075 (1.133)	0.792*** (0.205)	-0.066 (0.194)
Observations	3,235	3,235	3,235	3,235
R-squared	0.208	0.378	0.058	0.023
Test of excluded instruments	F(15, 457)	F(15, 457)	F(15, 457)	F(15, 457)
P-value	7.42	5.82	3.12	3.48
	0.000	0.000	0.000	0.000

Notes: Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 3b: IV Estimates - First-stage Regression of Income Model for Rural Sample

Variables	startage_r	sch_r	ehealth_r	ghealth_r
age	0.224 (0.176)	-0.157 (0.133)	0.005 (0.023)	-0.012 (0.023)
agesq	-0.003 (0.002)	0.002 (0.002)	-0.000 (0.000)	0.000 (0.000)
male	-2.439*** (0.359)	-0.989*** (0.259)	0.104*** (0.040)	-0.082* (0.046)
black	-0.898 (0.577)	-0.152 (0.434)	-0.035 (0.075)	-0.017 (0.091)
otherrace	-0.607* (0.356)	-0.337 (0.269)	-0.163*** (0.042)	0.106** (0.043)
curr GDP	-0.257 (0.157)	0.039 (0.124)	0.024 (0.019)	-0.009 (0.020)
fschmiss	-0.530 (0.367)	-0.934*** (0.285)	-0.133*** (0.044)	0.119*** (0.044)
filliterate	0.534 (0.752)	-1.446*** (0.451)	-0.211** (0.084)	0.224*** (0.082)
fupprimary	1.670** (0.780)	1.750*** (0.644)	-0.041 (0.127)	-0.011 (0.116)
fhighschool	2.369** (1.167)	3.574*** (0.579)	-0.131 (0.107)	0.195 (0.142)
fcollge	3.361** (1.537)	4.864*** (1.055)	0.127 (0.170)	0.033 (0.189)
mschmiss	-1.111*** (0.385)	-1.907*** (0.276)	0.025 (0.048)	-0.057 (0.050)
milliterate	-1.415** (0.686)	-2.263*** (0.489)	0.088 (0.097)	-0.267*** (0.092)
mupprimary	2.003** (0.841)	1.304* (0.705)	0.016 (0.110)	-0.017 (0.117)
mhighschool	2.777** (1.168)	1.632** (0.808)	0.192* (0.113)	-0.291*** (0.092)
mcollge	-1.614 (2.441)	0.313 (1.665)	0.207 (0.245)	-0.101 (0.260)
focumiss	-0.194 (1.032)	-0.287 (0.786)	0.095 (0.115)	-0.034 (0.128)
fnwork	-0.252 (0.574)	0.352 (0.472)	0.072 (0.076)	-0.072 (0.080)
fselfemploy	-0.030 (0.358)	0.380 (0.257)	-0.065 (0.050)	0.029 (0.050)
femployer	-0.106 (0.897)	0.957 (1.054)	-0.025 (0.138)	-0.029 (0.152)

(Continued) Table 3b: IV Estimates - First-stage Regression of Income Model for Rural Sample

Variables	startage_r	sch_r	ehealth_r	ghealth_r
mocumiss	-0.263 (0.892)	0.437 (0.745)	-0.043 (0.126)	-0.183 (0.155)
mnwork	0.138 (0.427)	0.631** (0.311)	0.090* (0.050)	-0.082 (0.053)
mselfemploy	0.285 (0.535)	0.922** (0.461)	-0.036 (0.069)	-0.084 (0.072)
memployer	-0.196 (0.606)	-0.942** (0.398)	-0.242*** (0.078)	0.392*** (0.079)
munsalary	-1.535** (0.607)	-0.422 (0.372)	-0.048 (0.066)	0.002 (0.072)
age12_gdp	0.175 (0.144)	0.072 (0.129)	-0.017 (0.020)	0.003 (0.022)
age7_teacher	0.123 (0.096)	-0.058 (0.066)	0.012 (0.012)	0.001 (0.012)
age11_teacher	-0.010 (0.098)	0.029 (0.066)	-0.028*** (0.011)	0.023** (0.010)
age7_hospital	4.582 (17.658)	-37.066** (15.121)	-1.813 (2.274)	0.446 (2.220)
age7_bed	-0.446 (0.318)	-0.242 (0.232)	0.002 (0.039)	0.019 (0.039)
age7_doctor	0.445 (0.726)	0.192 (0.575)	0.134 (0.105)	-0.207** (0.105)
Constant	12.076*** (3.384)	11.198*** (2.655)	0.469 (0.441)	0.713 (0.438)
Observations	666	666	666	666
R-squared	0.276	0.413	0.117	0.065
Test of excluded instruments	F(15, 269)	F(15, 269)	F(15, 269)	F(15, 269)
P-value	1.77	4.26	5.24	8.23
	0.039	0.000	0.000	0.000

Notes: Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 4: IV Estimates - Second-stage Regression of Income Model

Variables	mlninc_u	mlninc_r
startage	0.018 (0.041)	0.167** (0.071)
sch	0.083* (0.049)	0.111* (0.064)
ehealth	1.640** (0.812)	0.591 (0.659)
ghealth	0.219 (0.839)	0.895 (0.737)
age	0.084*** (0.017)	0.046 (0.033)
agesq	-0.001** (0.000)	-0.000 (0.000)
male	0.567*** (0.078)	1.089*** (0.183)
black	-0.033 (0.082)	0.169 (0.151)
otherrace	-0.018 (0.057)	0.134 (0.103)
currgdp	0.051*** (0.010)	0.050** (0.020)
fschmiss	0.025 (0.065)	0.118 (0.124)
filliterate	-0.004 (0.123)	-0.246 (0.254)
fupprimary	0.010 (0.071)	-0.130 (0.222)
fhighschool	0.150 (0.105)	-0.476 (0.387)
fcollge	0.231* (0.126)	-0.074 (0.544)
mschmiss	0.034 (0.089)	0.167 (0.145)
milliterate	0.213* (0.112)	0.478* (0.272)
mupprimary	0.037 (0.070)	-0.490** (0.234)
mhighschool	0.045 (0.093)	-0.475 (0.405)
mcollege	-0.118 (0.145)	0.173 (0.684)
Constant	1.325* (0.698)	-0.391 (1.046)
Observations	3,235	666
R-squared	0.033	-
Overidentification test of all instruments		
Hansen J-statistic	6.703	15.300
P-value	0.823	0.169

Notes: Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level. R² is not reported for the rural sample since it's negative.

Table 5: Ordered Probit Estimates of Health Model (2a)

Variables	Urban_1	Urban_2	Urban_3	Rural_1	Rural_2	Rural_3
startage	-0.001 (0.001)	-0.002 (0.001)	0.003 (0.002)	-0.005* (0.003)	-0.003* (0.002)	0.008* (0.004)
sch	-0.007*** (0.001)	-0.008*** (0.001)	0.015*** (0.003)	-0.005 (0.003)	-0.004 (0.002)	0.009 (0.006)
age	0.006 (0.004)	0.007 (0.005)	-0.013 (0.009)	-0.000 (0.011)	-0.000 (0.008)	0.000 (0.019)
agesq	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
male	-0.038*** (0.008)	-0.041*** (0.009)	0.079*** (0.017)	-0.080*** (0.024)	-0.044*** (0.012)	0.124*** (0.034)
black	0.010 (0.016)	0.010 (0.016)	-0.020 (0.031)	0.031 (0.051)	0.018 (0.023)	-0.049 (0.074)
otherrace	0.003 (0.009)	0.003 (0.010)	-0.005 (0.020)	0.077*** (0.024)	0.053*** (0.018)	-0.131*** (0.040)
currqdp	-0.007*** (0.002)	-0.008*** (0.003)	0.015*** (0.005)	-0.004 (0.007)	-0.003 (0.005)	0.007 (0.012)
fschmiss	0.020* (0.011)	0.021* (0.011)	-0.040* (0.022)	0.052** (0.025)	0.035** (0.016)	-0.087** (0.040)
filliterate	0.002 (0.023)	0.002 (0.025)	-0.003 (0.049)	0.076 (0.057)	0.031*** (0.012)	-0.107 (0.067)
fupprimary	-0.005 (0.014)	-0.005 (0.017)	0.010 (0.031)	0.041 (0.077)	0.021 (0.027)	-0.062 (0.104)
fhighschool	-0.006 (0.018)	-0.007 (0.021)	0.014 (0.039)	0.061 (0.088)	0.027 (0.021)	-0.088 (0.108)
fcollege	-0.028 (0.019)	-0.036 (0.029)	0.064 (0.047)	-0.068 (0.060)	-0.079 (0.111)	0.147 (0.171)
mschmiss	0.001 (0.011)	0.001 (0.012)	-0.002 (0.023)	-0.022 (0.027)	-0.015 (0.019)	0.037 (0.046)
milliterate	0.046 (0.029)	0.039** (0.019)	-0.085* (0.048)	0.013 (0.055)	0.008 (0.032)	-0.021 (0.088)
mupprimary	-0.002 (0.013)	-0.002 (0.015)	0.004 (0.028)	-0.023 (0.051)	-0.019 (0.049)	0.042 (0.101)
mhighschool	-0.014 (0.016)	-0.016 (0.020)	0.030 (0.035)	-0.020 (0.065)	-0.016 (0.060)	0.035 (0.124)
mcollege	-0.036 (0.023)	-0.051 (0.041)	0.087 (0.064)	-0.085 (0.075)	-0.116 (0.181)	0.200 (0.255)
age7_hospital	0.389 (0.422)	0.432 (0.472)	-0.821 (0.894)	0.420 (1.197)	0.294 (0.836)	-0.714 (2.031)
age7_bed	-0.015** (0.006)	-0.016** (0.007)	0.031** (0.013)	-0.019 (0.021)	-0.013 (0.014)	0.032 (0.035)
age7_doctor	0.068*** (0.015)	0.075*** (0.018)	-0.142*** (0.033)	0.020 (0.047)	0.014 (0.032)	-0.033 (0.079)
Observations	3,235	3,235	3,235	666	666	666

Notes: Marginal effects rather than ordered probit estimates are reported. Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level. Column 1, 2 and 3 correspond to the marginal effects on the probability that the health index equals 1, 2, and 3 for urban residents, column 4, 5 and 6 correspond to the marginal effects for rural residents.

Table 6: IV Ordered Probit Estimates - First-stage Regression of Health Model (2a)

Variables	startage_u	sch_u	startage_r	sch_r
age	0.341*** (0.086)	0.495*** (0.060)	0.235 (0.180)	-0.159 (0.134)
agesq	-0.005*** (0.001)	-0.006*** (0.001)	-0.003 (0.002)	0.002 (0.002)
male	-1.986*** (0.151)	-0.876*** (0.115)	-2.461*** (0.350)	-0.974*** (0.252)
black	-0.619** (0.247)	-1.127*** (0.239)	-0.840 (0.567)	-0.193 (0.430)
otherrace	-0.433*** (0.156)	-1.070*** (0.129)	-0.597* (0.348)	-0.343 (0.265)
currgdp	-0.216*** (0.046)	-0.219*** (0.040)	-0.199 (0.154)	-0.002 (0.111)
fschmiss	-1.021*** (0.202)	-1.194*** (0.172)	-0.551 (0.363)	-0.920*** (0.276)
filliterate	-1.255*** (0.453)	-2.229*** (0.388)	0.517 (0.730)	-1.433*** (0.443)
fupprimary	0.396 (0.248)	0.904*** (0.207)	1.716** (0.760)	1.722*** (0.624)
fhighschool	1.089*** (0.271)	1.529*** (0.250)	2.453** (1.129)	3.512*** (0.565)
fcollege	1.695*** (0.341)	2.572*** (0.277)	3.581** (1.499)	4.708*** (1.027)
mschmiss	-0.781*** (0.177)	-2.054*** (0.156)	-1.061*** (0.377)	-1.942*** (0.271)
milliterate	-1.135*** (0.434)	-1.600*** (0.355)	-1.403** (0.671)	-2.271*** (0.473)
mupprimary	0.903*** (0.230)	1.046*** (0.207)	1.938** (0.806)	1.349* (0.690)
mhighschool	1.483*** (0.288)	1.973*** (0.196)	2.988** (1.174)	1.482* (0.777)
mcollege	1.586*** (0.513)	2.972*** (0.383)	-1.513 (2.283)	0.247 (1.695)
age7_hospital	8.983 (9.036)	7.298 (6.296)	-1.957 (18.744)	-31.567** (14.792)
age7_bed	-0.149 (0.137)	-0.050 (0.107)	-0.368 (0.332)	-0.307 (0.236)
age7_doctor	1.289*** (0.446)	0.581** (0.287)	0.920 (0.709)	-0.131 (0.543)
focumiss	-0.200 (0.341)	0.141 (0.287)	-0.106 (0.888)	-0.347 (0.658)
fnwork	-0.721*** (0.219)	0.304 (0.193)	0.159 (0.519)	0.065 (0.418)
fselfemploy	-0.400** (0.178)	0.021 (0.150)	0.098 (0.326)	0.290 (0.247)

(Continued) Table 6: IV Ordered Probit Estimates - First-stage Regression of Health Model (2a)

Variables	startage_u	sch_u	startage_r	sch_r
employer	-0.161 (0.354)	0.775*** (0.263)	0.459 (0.980)	0.563 (0.916)
mocumiss	0.565 (0.608)	0.737 (0.518)	-0.153 (0.923)	0.363 (0.637)
mnwork	0.673*** (0.169)	0.694*** (0.130)	0.557* (0.334)	0.340 (0.335)
mselfemploy	0.049 (0.254)	0.541*** (0.194)	0.538 (0.508)	0.748 (0.478)
memployer	-0.458 (0.823)	0.772 (0.558)	-0.743* (0.447)	-0.554 (0.431)
munsalary	-1.659*** (0.444)	-0.430 (0.336)	-1.131* (0.671)	-0.703** (0.335)
age12_gdp	0.085 (0.066)	0.268*** (0.053)	0.101 (0.147)	0.125 (0.095)
age7_teacher	-0.049 (0.051)	-0.041 (0.043)	0.056 (0.105)	-0.011 (0.065)
age11_teacher	-0.022 (0.049)	-0.063 (0.041)	-0.023 (0.078)	0.039 (0.050)
Constant	11.082*** (1.668)	0.089 (1.136)	11.750*** (3.491)	11.306*** (2.704)
Observations	3,235	3,235	666	666
Test of excluded instruments				
$\chi^2(12)$	78.48	79.77	35.47	26.38
P-value	0.000	0.000	0.000	0.010

Notes: Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level. The first two columns correspond to the first-stage estimates of the starting age and schooling equations for the urban sample, and the last two columns correspond to the rural sample.

Table 7: IV Ordered Probit Estimates - Second-stage Regression of Health Model (2a)

Variables	Urban_1	Urban_2	Urban_3	Rural_1	Rural_2	Rural_3
startage	0.010 (0.011)	0.009 (0.008)	-0.018 (0.019)	-0.082*** (0.021)	-0.003 (0.007)	0.085*** (0.018)
sch	-0.032** (0.013)	-0.028*** (0.006)	0.060*** (0.018)	0.100*** (0.033)	0.004 (0.009)	-0.105*** (0.029)
age	0.013** (0.006)	0.011** (0.005)	-0.024** (0.011)	0.035 (0.022)	0.001 (0.003)	-0.037* (0.022)
agesq	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)
male	-0.039** (0.019)	-0.033* (0.018)	0.073** (0.036)	-0.129 (0.083)	-0.002 (0.008)	0.131 (0.084)
black	-0.013 (0.019)	-0.012 (0.018)	0.026 (0.037)	-0.032 (0.067)	-0.002 (0.005)	0.034 (0.071)
otherrace	-0.021 (0.015)	-0.019* (0.012)	0.040 (0.027)	0.027 (0.056)	0.001 (0.004)	-0.028 (0.061)
currqdp	-0.009*** (0.003)	-0.008*** (0.003)	0.017*** (0.006)	-0.022* (0.012)	-0.001 (0.002)	0.023* (0.013)
fschmiss	0.002 (0.017)	0.002 (0.015)	-0.004 (0.033)	0.075 (0.052)	0.003 (0.007)	-0.078 (0.056)
filliterate	-0.037 (0.028)	-0.041 (0.034)	0.078 (0.062)	0.229*** (0.072)	-0.013 (0.013)	-0.217*** (0.061)
fupprimary	0.013 (0.019)	0.011 (0.014)	-0.023 (0.033)	-0.013 (0.106)	-0.001 (0.006)	0.014 (0.111)
fhighschool	0.021 (0.028)	0.017 (0.019)	-0.038 (0.046)	-0.129 (0.156)	-0.015 (0.032)	0.144 (0.179)
fcollege	0.016 (0.041)	0.013 (0.029)	-0.029 (0.070)	-0.205 (0.145)	-0.034 (0.068)	0.239 (0.192)
mschmiss	-0.039* (0.023)	-0.038** (0.019)	0.077* (0.041)	0.109 (0.081)	0.005 (0.010)	-0.114 (0.083)
milliterate	0.014 (0.034)	0.012 (0.027)	-0.026 (0.062)	0.133 (0.097)	-0.002 (0.008)	-0.131 (0.090)
mupprimary	0.014 (0.020)	0.012 (0.015)	-0.026 (0.034)	0.019 (0.105)	0.001 (0.003)	-0.020 (0.107)
mhighschool	0.017 (0.027)	0.014 (0.019)	-0.030 (0.046)	0.083 (0.133)	0.000 (0.007)	-0.083 (0.128)
mcollege	0.006 (0.041)	0.005 (0.033)	-0.012 (0.074)	-0.184 (0.143)	-0.029 (0.063)	0.212 (0.191)
age7_hospital	0.258 (0.482)	0.229 (0.441)	-0.487 (0.921)	3.461 (2.432)	0.147 (0.320)	-3.608 (2.486)
age7_bed	-0.012* (0.007)	-0.011 (0.007)	0.023* (0.014)	-0.002 (0.033)	-0.000 (0.001)	0.002 (0.034)
age7_doctor	0.065*** (0.019)	0.058*** (0.023)	-0.124*** (0.040)	0.084 (0.072)	0.004 (0.008)	-0.088 (0.076)
Observations	3,235	3,235	3,235	666	666	666

Notes: Marginal effects are reported rather than probit coefficients. Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level. Column 1, 2 and 3 correspond to the marginal effects on the probability that the health index equals 1, 2 and 3 for the urban residents, and the column 4, 5 and 6 correspond to the marginal effects for the rural residents.

Table 8: Probit Estimates of Health Model(2b)

Variables	Urban	Rural
startage	-0.005*** (0.002)	-0.010** (0.004)
sch	0.001 (0.002)	0.003 (0.006)
age	-0.003 (0.007)	0.010 (0.016)
agesq	0.000 (0.000)	-0.000 (0.000)
male	-0.039*** (0.014)	-0.016 (0.036)
black	0.016 (0.029)	-0.024 (0.050)
otherrace	0.030* (0.017)	-0.012 (0.036)
currgdp	-0.017*** (0.004)	0.002 (0.010)
fschmiss	-0.007 (0.017)	0.090** (0.039)
filliterate	0.087* (0.050)	0.127 (0.097)
fupprimary	-0.037* (0.023)	0.141 (0.141)
fhighschool	-0.011 (0.030)	0.026 (0.118)
fcollge	-0.037 (0.038)	-0.077 (0.106)
mschmiss	-0.029 (0.018)	0.001 (0.041)
milliterate	0.027 (0.042)	-0.017 (0.069)
mupprimary	0.028 (0.028)	-0.064 (0.084)
mhighschool	-0.010 (0.031)	0.052 (0.112)
mcollege	0.013 (0.057)	0.069 (0.229)
age7_hospital	-0.790 (0.900)	1.268 (1.628)
age7_bed	-0.004 (0.012)	-0.045 (0.028)
age7_doctor	0.061** (0.030)	0.015 (0.068)
Observations	3,235	666

Notes: Marginal effects are reported rather than probit coefficients. Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 9: IV Probit Estimates - First-stage Regression of Health Model (2b)

Variables	startage_u	sch_u	startage_r	sch_r
age	0.332*** (0.085)	0.501*** (0.060)	0.188 (0.332)	-0.138 (0.145)
agesq	-0.005*** (0.001)	-0.006*** (0.001)	-0.002 (0.003)	0.002 (0.002)
male	-1.985*** (0.151)	-0.876*** (0.115)	-2.427*** (0.511)	-0.995* (0.567)
black	-0.617** (0.248)	-1.126*** (0.237)	-0.873 (1.881)	-0.165 (1.981)
otherrace	-0.432*** (0.155)	-1.070*** (0.128)	-0.586* (0.338)	-0.348 (0.442)
currgdp	-0.215*** (0.046)	-0.223*** (0.040)	-0.180 (0.445)	-0.002 (0.126)
fschmiss	-1.031*** (0.202)	-1.187*** (0.173)	-0.534 (0.760)	-0.932*** (0.281)
filliterate	-1.255*** (0.449)	-2.232*** (0.386)	0.518 (1.749)	-1.437 (0.889)
fupprimary	0.397 (0.248)	0.903*** (0.207)	1.639* (0.869)	1.767*** (0.545)
fhighschool	1.093*** (0.272)	1.531*** (0.252)	2.472** (1.100)	3.520*** (0.996)
fcollege	1.696*** (0.340)	2.572*** (0.277)	3.436** (1.457)	4.825*** (0.864)
mschmiss	-0.787*** (0.179)	-2.052*** (0.157)	-1.084* (0.601)	-1.921*** (0.324)
milliterate	-1.137*** (0.433)	-1.597*** (0.351)	-1.404 (1.689)	-2.268*** (0.854)
mupprimary	0.910*** (0.230)	1.044*** (0.207)	1.983* (1.167)	1.314 (0.862)
mhighschool	1.502*** (0.288)	1.972*** (0.196)	2.870 (4.285)	1.584 (2.096)
mcollege	1.609*** (0.514)	2.966*** (0.381)	-1.634 (2.534)	0.323 (2.068)
age7_hospital	9.169 (8.900)	7.687 (6.308)	-0.979 (55.783)	-34.170* (18.965)
age7_bed	-0.152 (0.135)	-0.056 (0.105)	-0.407 (0.352)	-0.262 (0.355)
age7_doctor	1.216*** (0.430)	0.605** (0.284)	0.512 (1.252)	0.157 (0.496)
focumiss	-0.145 (0.337)	0.106 (0.282)	-0.240 (0.847)	-0.264 (0.932)
fnwork	-0.705*** (0.218)	0.277 (0.204)	0.256 (0.742)	0.088 (0.843)
fselfemploy	-0.390** (0.175)	0.022 (0.150)	0.141 (0.290)	0.291 (0.226)

(Continued) Table 9: IV Probit Estimates - First-stage Regression of Health Model (2b)

Variables	startage_u	sch_u	startage_r	sch_r
femployer	-0.225 (0.329)	0.825*** (0.254)	0.270 (1.681)	0.762 (1.329)
mocumiss	0.702 (0.598)	0.659 (0.494)	-0.456 (0.922)	0.537 (1.950)
mnwork	0.721*** (0.160)	0.672*** (0.137)	0.183 (0.478)	0.608 (0.491)
mselfemploy	0.100 (0.249)	0.530*** (0.198)	0.371 (1.322)	0.878* (0.468)
memployer	-0.351 (0.777)	0.477 (0.575)	-2.356 (2.415)	0.182 (1.007)
munsalary	-1.597*** (0.438)	-0.415 (0.338)	-1.414 (0.901)	-0.485 (1.011)
age12_gdp	0.075 (0.061)	0.280*** (0.049)	0.070 (0.380)	0.126 (0.208)
age7_teacher	-0.050 (0.049)	-0.043 (0.040)	0.143* (0.081)	-0.069 (0.148)
age11_teacher	-0.006 (0.046)	-0.067* (0.041)	-0.025 (0.092)	0.037 (0.109)
Constant	11.212*** (1.642)	-0.028 (1.130)	12.743 (8.587)	10.850*** (3.754)
Observations	3,235	3,235	666	666
Test of excluded instruments				
$\chi^2(12)$	80.32	77.10	91.63	33.47
P-value	0.000	0.000	0.000	0.010

Notes: Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 10: IV Probit Estimates - Second-stage Regression of Health Model (2b)

Variables	Urban	Rural
startage	-0.021*** (0.008)	-0.091*** (0.018)
sch	-0.014 (0.013)	0.094 (0.069)
age	0.009 (0.009)	0.036 (0.023)
agesq	-0.000 (0.000)	-0.000 (0.000)
male	-0.085*** (0.021)	-0.128 (0.112)
black	-0.014 (0.029)	-0.061 (0.065)
otherrace	0.004 (0.020)	-0.024 (0.035)
currngdp	-0.022*** (0.004)	-0.018 (0.037)
fschmiss	-0.043* (0.022)	0.072** (0.037)
filliterate	0.020 (0.052)	0.230 (0.274)
fupprimary	-0.016 (0.027)	0.042 (0.158)
fhighschool	0.035 (0.038)	-0.095 (0.232)
fcollge	0.035 (0.055)	-0.151 (0.331)
mschmiss	-0.071** (0.028)	0.092 (0.210)
milliterate	-0.019 (0.042)	0.084 (0.132)
mupprimary	0.059* (0.032)	0.034 (0.149)
mhighschool	0.043 (0.043)	0.156 (0.424)
mcollege	0.079 (0.079)	-0.120 (0.110)
age7_hospital	-0.601 (0.961)	3.212 (6.224)
age7_bed	-0.006 (0.012)	-0.017 (0.050)
age7_doctor	0.082** (0.032)	0.072 (0.073)
Observations	3,235	666
Overidentification test	$\chi^2(10) = 4.427$	$\chi^2(9) = 4.9018$
P-value	0.937	0.843

Notes: Marginal effects are reported rather than probit coefficients. Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level. The Amemiya-Lee-Newey statistics for overidentification test of instruments are reported. The overidentification test statistic for the rural sample is distributed as $\chi^2(9)$ since the instrument “memployer” predicts failure perfectly in the estimation process and is omitted.

Table 11a: Quantile Regression Estimates of Schooling Model - Urban Sample

Variables	OLS	q10	q20	q30	q40	q50	q60	q70	q80	q90
startage	0.227*** (0.015)	0.158*** (0.028)	0.223*** (0.025)	0.258*** (0.022)	0.276*** (0.022)	0.284*** (0.021)	0.254*** (0.019)	0.232*** (0.017)	0.211*** (0.019)	0.162*** (0.025)
age	0.433*** (0.058)	0.143 (0.089)	0.318*** (0.072)	0.319*** (0.072)	0.389*** (0.072)	0.492*** (0.073)	0.594*** (0.072)	0.631*** (0.081)	0.552*** (0.087)	0.540*** (0.091)
agesq	-0.005*** (0.001)	-0.002 (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.005*** (0.001)	-0.006*** (0.001)	-0.007*** (0.001)	-0.007*** (0.001)	-0.006*** (0.001)	-0.005*** (0.001)
male	-0.432*** (0.118)	-0.346** (0.176)	-0.498*** (0.150)	-0.390*** (0.143)	-0.534*** (0.136)	-0.456*** (0.145)	-0.443*** (0.144)	-0.509*** (0.153)	-0.543*** (0.179)	-0.425** (0.177)
black	-1.065*** (0.234)	-0.523* (0.313)	-0.727** (0.331)	-0.948*** (0.309)	-0.916*** (0.246)	-1.158*** (0.266)	-1.406*** (0.303)	-1.143*** (0.378)	-1.149*** (0.339)	-1.343*** (0.378)
otherrace	-1.054*** (0.121)	-0.774*** (0.220)	-0.818*** (0.187)	-1.041*** (0.178)	-1.000*** (0.172)	-1.037*** (0.171)	-1.036*** (0.159)	-1.074*** (0.153)	-1.123*** (0.171)	-1.277*** (0.209)
currgdp	-0.189*** (0.038)	-0.002 (0.055)	-0.047 (0.044)	-0.104** (0.044)	-0.178*** (0.037)	-0.223*** (0.043)	-0.278*** (0.049)	-0.310*** (0.060)	-0.302*** (0.072)	-0.219** (0.091)
fschmiss	-0.952*** (0.164)	-0.807*** (0.226)	-0.728*** (0.228)	-1.030*** (0.223)	-0.901*** (0.244)	-0.989*** (0.224)	-1.191*** (0.204)	-1.263*** (0.198)	-0.914*** (0.230)	-0.625** (0.284)
filliterate	-2.026*** (0.367)	-1.267** (0.534)	-1.716** (0.714)	-1.783*** (0.484)	-1.995*** (0.423)	-1.951*** (0.435)	-2.370*** (0.422)	-2.227*** (0.445)	-2.265*** (0.723)	-1.561** (0.776)
fuprimary	0.864*** (0.204)	1.384*** (0.382)	1.217*** (0.276)	0.959*** (0.273)	0.860*** (0.245)	0.769*** (0.203)	0.598** (0.265)	0.488* (0.272)	0.585** (0.272)	0.420 (0.293)
flhighschool	1.330*** (0.243)	1.496** (0.738)	1.574*** (0.465)	1.349*** (0.293)	1.182*** (0.245)	1.100*** (0.252)	0.940*** (0.276)	1.183*** (0.337)	1.114*** (0.340)	0.899*** (0.319)
fcollge	2.286*** (0.259)	2.790*** (0.549)	2.543*** (0.427)	2.195*** (0.479)	2.293*** (0.405)	2.001*** (0.383)	1.947*** (0.314)	1.831*** (0.356)	1.920*** (0.320)	1.424*** (0.332)
mschmiss	-1.871*** (0.158)	-1.161*** (0.226)	-1.600*** (0.216)	-2.068*** (0.201)	-2.511*** (0.236)	-2.476*** (0.246)	-2.300*** (0.207)	-1.852*** (0.210)	-1.767*** (0.200)	-1.395*** (0.278)
milliterate	-1.342*** (0.360)	-2.015*** (0.537)	-2.302*** (0.788)	-1.543** (0.661)	-1.567*** (0.498)	-1.409*** (0.529)	-0.983** (0.490)	-0.932** (0.386)	-0.646 (0.492)	-0.696 (0.588)
muprimary	0.800*** (0.198)	1.567*** (0.532)	1.371*** (0.358)	1.069*** (0.273)	0.627*** (0.237)	0.425* (0.226)	0.390 (0.284)	0.746*** (0.257)	0.761** (0.318)	0.896*** (0.301)
mhighschool	1.550*** (0.194)	3.187*** (0.499)	2.422*** (0.363)	1.777** (0.271)	1.311*** (0.238)	1.188*** (0.229)	1.008*** (0.203)	1.030*** (0.292)	1.287*** (0.310)	1.299*** (0.353)

(Continued) Table 11a: Quantile Regression Estimates of Schooling Model -Urban Sample

Variables	OLS	q10	q20	q30	q40	q50	q60	q70	q80	q90
mcollege	2.318*** (0.344)	3.277*** (0.564)	2.561*** (0.615)	2.387*** (0.687)	2.049*** (0.515)	1.827*** (0.485)	1.963*** (0.581)	2.257*** (0.624)	2.358*** (0.402)	2.024*** (0.383)
age12_gdp	0.265*** (0.046)	0.082 (0.071)	0.133** (0.061)	0.166*** (0.055)	0.186*** (0.054)	0.232*** (0.052)	0.325*** (0.054)	0.382*** (0.060)	0.460*** (0.075)	0.404*** (0.099)
age7_teacher	-0.030 (0.038)	0.005 (0.080)	-0.048 (0.058)	-0.015 (0.052)	-0.057 (0.053)	-0.065 (0.052)	-0.038 (0.043)	-0.029 (0.045)	-0.054 (0.055)	-0.005 (0.067)
age11_teacher	-0.066* (0.036)	-0.072 (0.067)	-0.060 (0.067)	-0.053 (0.046)	-0.028 (0.051)	-0.023 (0.041)	-0.051 (0.033)	-0.107*** (0.036)	-0.084 (0.051)	-0.045 (0.072)
age7_hospital	6.184 (6.199)	-8.638 (8.429)	-3.857 (9.289)	-3.699 (10.074)	-3.235 (8.339)	7.812 (9.507)	19.009** (9.048)	21.556*** (6.636)	15.379* (8.131)	31.903*** (11.625)
age7_bed	-0.017 (0.101)	0.154 (0.128)	0.029 (0.131)	-0.073 (0.144)	0.035 (0.129)	-0.017 (0.133)	-0.047 (0.132)	-0.049 (0.115)	-0.061 (0.140)	-0.230 (0.176)
age7_doctor	0.332 (0.295)	0.000 (0.565)	0.321 (0.439)	0.458 (0.410)	0.536 (0.417)	0.580 (0.387)	0.394 (0.361)	0.638* (0.346)	0.283 (0.414)	0.076 (0.451)
Constant	-2.177** (1.093)	-0.504 (1.615)	-2.341 (1.443)	-1.566 (1.407)	-2.211 (1.455)	-3.686*** (1.417)	-4.552*** (1.356)	-4.650*** (1.456)	-2.184 (1.602)	-1.008 (1.770)
Observations	3,235	3,235	3,235	3,235	3,235	3,235	3,235	3,235	3,235	3,235

Note: Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 11b: Quantile Regression Estimates of Schooling Model -Rural Sample

Variables	OLS	q10	q20	q30	q40	q50	q60	q70	q80	q90
startage	0.163*** (0.034)	0.046 (0.033)	0.094** (0.039)	0.149*** (0.042)	0.171*** (0.045)	0.178*** (0.051)	0.194*** (0.044)	0.220*** (0.043)	0.203*** (0.057)	0.167*** (0.072)
age	-0.207 (0.131)	0.056 (0.144)	0.018 (0.159)	-0.248* (0.138)	-0.225* (0.119)	-0.181 (0.153)	-0.136 (0.187)	-0.107 (0.202)	-0.271 (0.231)	-0.546* (0.300)
agesq	0.003 (0.002)	-0.001 (0.002)	-0.000 (0.002)	0.003* (0.002)	0.003* (0.002)	0.002 (0.002)	0.001 (0.002)	0.001 (0.003)	0.004 (0.003)	0.008* (0.004)
male	-0.598** (0.262)	-0.710*** (0.261)	-0.597** (0.284)	-0.538** (0.252)	-0.467* (0.267)	-0.431 (0.361)	-0.507 (0.376)	-0.485 (0.351)	-0.392 (0.448)	-0.765 (0.597)
black	-0.090 (0.411)	-0.068 (0.421)	-0.090 (0.443)	0.074 (0.438)	-0.110 (0.463)	-0.120 (0.468)	0.088 (0.440)	-0.478 (0.464)	-0.922 (0.585)	0.041 (1.169)
otherrace	-0.246 (0.264)	-0.449 (0.277)	-0.252 (0.318)	-0.223 (0.343)	-0.377 (0.326)	-0.379 (0.353)	-0.012 (0.335)	-0.107 (0.380)	-0.315 (0.385)	0.161 (0.568)
currgdp	0.076 (0.118)	0.015 (0.118)	0.051 (0.111)	0.115 (0.135)	0.095 (0.142)	0.093 (0.154)	0.149 (0.165)	0.068 (0.161)	-0.005 (0.177)	0.146 (0.303)
fschmiss	-0.907*** (0.279)	-1.036*** (0.393)	-1.263*** (0.379)	-1.048*** (0.263)	-1.034*** (0.270)	-0.763** (0.323)	-0.522 (0.341)	-0.532 (0.362)	-0.662 (0.486)	-0.399 (0.605)
filliterate	-1.586*** (0.431)	-1.687*** (0.495)	-2.254*** (0.551)	-1.599** (0.648)	-1.175** (0.565)	-1.261** (0.509)	-1.401** (0.613)	-1.419* (0.761)	-1.076 (0.947)	-1.170 (1.352)
fuprimary	1.325** (0.645)	1.577 (1.128)	1.456 (0.944)	0.909 (1.012)	1.500 (1.167)	1.751* (1.052)	1.540* (0.810)	0.976 (0.732)	0.376 (0.887)	-0.064 (1.262)
flhighschool	3.281*** (0.607)	6.469*** (1.511)	5.365*** (1.302)	4.971*** (1.226)	4.906*** (1.163)	3.479*** (1.082)	2.442*** (0.924)	2.331** (1.072)	1.730 (1.300)	1.483 (1.613)
fcollge	4.171*** (0.906)	5.126*** (1.912)	6.195*** (1.745)	5.641*** (1.754)	5.067*** (1.908)	2.736 (1.763)	4.124*** (1.536)	3.389** (1.391)	3.384*** (1.261)	3.032* (1.584)
mschmiss	-1.841*** (0.266)	-0.663 (0.415)	-0.867*** (0.320)	-0.935*** (0.261)	-0.927*** (0.319)	-1.458*** (0.363)	-2.289*** (0.469)	-2.869*** (0.459)	-3.159*** (0.536)	-3.673*** (0.721)
milliterate	-2.087*** (0.482)	-1.272** (0.505)	-1.612*** (0.475)	-2.029*** (0.650)	-1.420** (0.692)	-1.841*** (0.630)	-2.403*** (0.708)	-2.606*** (0.785)	-3.041*** (0.956)	-3.763*** (1.701)
muprimary	1.050 (0.647)	-0.633 (1.305)	0.819 (1.193)	0.769 (1.186)	0.670 (1.236)	1.288 (1.264)	1.142 (1.171)	1.592* (0.931)	0.822 (0.823)	0.688 (1.533)
mhighschool	0.961 (0.717)	1.291 (1.541)	0.451 (1.479)	0.528 (1.593)	0.802 (1.703)	2.009 (1.414)	1.767 (1.107)	0.901 (0.810)	0.200 (0.704)	-0.520 (1.124)

(Continued) Table 11b: Quantile Regression Estimates of Schooling Model -Rural Sample

Variables	OLS	q10	q20	q30	q40	q50	q60	q70	q80	q90
mcollege	0.514 (1.660)	0.308 (2.127)	-0.695 (2.125)	0.759 (2.417)	0.553 (2.672)	1.949 (2.706)	1.949 (2.690)	0.675 (2.955)	2.921 (3.073)	1.688 (3.662)
age12_gdp	0.049 (0.125)	0.078 (0.147)	-0.026 (0.136)	-0.049 (0.147)	-0.067 (0.158)	-0.019 (0.188)	-0.029 (0.199)	0.098 (0.183)	0.200 (0.183)	0.110 (0.291)
age7_teacher	-0.055 (0.062)	-0.054 (0.084)	-0.041 (0.067)	-0.047 (0.061)	-0.090 (0.071)	-0.061 (0.100)	-0.043 (0.112)	-0.090 (0.106)	-0.095 (0.132)	-0.048 (0.151)
age11_teacher	0.024 (0.066)	0.019 (0.053)	0.001 (0.061)	0.036 (0.068)	-0.000 (0.066)	0.010 (0.093)	0.049 (0.111)	0.036 (0.117)	0.085 (0.128)	0.079 (0.145)
age7_hospital	-40.534*** (14.243)	-8.684 (14.654)	-6.555 (15.826)	-26.266 (16.192)	-31.472* (17.694)	-35.516* (20.156)	-42.604** (21.206)	-29.642 (20.082)	-36.080* (19.314)	-49.910* (27.315)
age7_bed	-0.181 (0.229)	-0.162 (0.233)	-0.215 (0.238)	-0.030 (0.226)	-0.010 (0.247)	-0.168 (0.307)	-0.263 (0.312)	-0.305 (0.272)	-0.276 (0.298)	-0.089 (0.457)
age7_doctor	0.018 (0.579)	0.772 (0.670)	1.001 (0.677)	0.142 (0.609)	0.445 (0.614)	0.117 (0.660)	0.115 (0.704)	0.138 (0.647)	-0.401 (0.822)	-1.668 (1.312)
Constant	10.290*** (2.525)	1.650 (2.905)	3.088 (3.101)	7.778*** (2.917)	8.555*** (2.664)	9.142*** (3.321)	9.304** (3.768)	9.318** (3.918)	13.672*** (4.542)	20.274*** (5.789)
Observations	666	666	666	666	666	666	666	666	666	666

Note: Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 12: IV Ordered Probit Estimates - Second-stage Regression of Health Model (2a) with Full Sample

Variables	Urban_1	Urban_2	Urban_3	Rural_1	Rural_2	Rural_3
startage	-0.005 (0.007)	-0.003 (0.004)	0.009 (0.011)	-0.032 (0.020)	-0.007*** (0.003)	0.039* (0.023)
sch	-0.028*** (0.010)	-0.018*** (0.005)	0.046*** (0.015)	0.017 (0.051)	0.004 (0.010)	-0.021 (0.061)
age	0.013*** (0.005)	0.009*** (0.003)	-0.022*** (0.007)	-0.002 (0.009)	-0.000 (0.002)	0.002 (0.011)
agesq	-0.000 (0.000)	-0.000* (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
male	-0.068*** (0.012)	-0.043*** (0.009)	0.111*** (0.020)	-0.144*** (0.028)	-0.025*** (0.009)	0.169*** (0.028)
black	-0.011 (0.015)	-0.007 (0.010)	0.018 (0.025)	0.043 (0.047)	0.006* (0.003)	-0.049 (0.049)
otherrace	0.000 (0.012)	0.000 (0.008)	-0.001 (0.020)	0.049* (0.025)	0.011** (0.005)	-0.060** (0.029)
currqdp	-0.008*** (0.002)	-0.005*** (0.001)	0.014*** (0.003)	-0.025*** (0.008)	-0.006*** (0.002)	0.031*** (0.009)
fschmiss	-0.014 (0.015)	-0.009 (0.009)	0.023 (0.024)	0.040 (0.049)	0.009 (0.009)	-0.048 (0.058)
filliterate	-0.021 (0.022)	-0.015 (0.017)	0.035 (0.039)	0.054 (0.086)	0.007*** (0.002)	-0.061 (0.088)
fupprimary	0.003 (0.014)	0.002 (0.008)	-0.005 (0.022)	-0.006 (0.080)	-0.001 (0.020)	0.008 (0.100)
fhighschool	0.020 (0.024)	0.011 (0.012)	-0.031 (0.036)	0.021 (0.144)	0.004 (0.020)	-0.025 (0.164)
fcollege	0.007 (0.033)	0.004 (0.019)	-0.012 (0.052)	-0.108 (0.137)	-0.061 (0.127)	0.170 (0.264)
mschmiss	-0.032* (0.017)	-0.021** (0.011)	0.053* (0.028)	0.034 (0.075)	0.008 (0.015)	-0.042 (0.091)
milliterate	-0.018 (0.024)	-0.013 (0.018)	0.031 (0.041)	-0.000 (0.075)	-0.000 (0.017)	0.000 (0.092)
mupprimary	0.030* (0.018)	0.017** (0.008)	-0.047* (0.026)	-0.028 (0.055)	-0.008 (0.020)	0.037 (0.074)
mhighschool	0.046* (0.026)	0.023*** (0.009)	-0.070** (0.034)	0.058 (0.112)	0.006** (0.003)	-0.064 (0.112)
mcollege	0.021 (0.036)	0.012 (0.017)	-0.033 (0.054)	-0.116 (0.132)	-0.070 (0.132)	0.186 (0.265)
age7_hospital	0.374 (0.331)	0.238 (0.215)	-0.612 (0.545)	-1.575 (1.308)	-0.354 (0.372)	1.930 (1.655)
age7_bed	-0.016*** (0.005)	-0.010*** (0.004)	0.026*** (0.009)	0.019 (0.016)	0.004 (0.003)	-0.023 (0.020)
age7_doctor	0.067*** (0.014)	0.043*** (0.011)	-0.110*** (0.024)	-0.006 (0.034)	-0.001 (0.008)	0.008 (0.042)
Observations	6,439	6,439	6,439	1,573	1,573	1,573

Notes: Marginal effects are reported rather than probit coefficients. Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 13: IV Probit estimates - Second-stage Regression of Health Model(2b) with Full Sample

Variables	Urban	Rural
startage	-0.019*** (0.006)	-0.057*** (0.006)
sch	-0.019* (0.011)	0.035*** (0.013)
age	0.015** (0.006)	0.011 (0.011)
agesq	-0.000* (0.000)	-0.000 (0.000)
male	-0.087*** (0.014)	-0.122*** (0.026)
black	-0.021 (0.023)	0.012 (0.041)
otherrace	0.017 (0.016)	-0.000 (0.025)
currngdp	-0.018*** (0.003)	-0.013 (0.008)
fschmiss	-0.032* (0.018)	0.046 (0.028)
filliterate	-0.021 (0.034)	0.081 (0.056)
fupprimary	0.020 (0.020)	0.038 (0.093)
fhighschool	0.033 (0.032)	-0.040 (0.089)
fcollege	0.051 (0.047)	-0.028 (0.129)
mschmiss	-0.061*** (0.020)	0.024 (0.034)
milliterate	-0.015 (0.031)	0.023 (0.049)
mupprimary	0.063** (0.025)	-0.005 (0.070)
mhighschool	0.096*** (0.037)	0.121 (0.090)
mcollege	0.155** (0.061)	-0.085 (0.110)
age7_hospital	0.502 (0.559)	0.731 (1.167)
age7_bed	-0.015* (0.009)	-0.021 (0.017)
age7_doctor	0.075*** (0.024)	-0.014 (0.045)
Observations	6,439	1,573
Overidentification test	$\chi^2(10) = 7.143$	$\chi^2(10) = 13.690$
P-value	0.712	0.188

Notes: Marginal effects are reported rather than probit coefficients. Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level. The Amemiya-Lee-Newey statistics for overidentification test of instruments are reported.

Table 14a: Quantile Regression Estimates of Schooling Model -Urban Full Sample

Variables	OLS	q10	q20	q30	q40	q50	q60	q70	q80	q90
startage	0.222*** (0.011)	0.142*** (0.013)	0.187*** (0.017)	0.235*** (0.015)	0.257*** (0.017)	0.280*** (0.015)	0.256*** (0.016)	0.242*** (0.014)	0.222*** (0.016)	0.219*** (0.018)
age	0.385*** (0.038)	0.141*** (0.048)	0.194*** (0.054)	0.303*** (0.046)	0.367*** (0.055)	0.430*** (0.054)	0.460*** (0.056)	0.528*** (0.055)	0.486*** (0.064)	0.505*** (0.074)
agesq	-0.005*** (0.000)	-0.002*** (0.001)	-0.003*** (0.001)	-0.004*** (0.001)	-0.005*** (0.001)	-0.006*** (0.001)	-0.006*** (0.001)	-0.007*** (0.001)	-0.006*** (0.001)	-0.005*** (0.001)
male	0.040 (0.085)	-0.115 (0.091)	0.011 (0.094)	0.043 (0.091)	0.024 (0.101)	0.063 (0.106)	0.068 (0.116)	0.101 (0.111)	0.001 (0.129)	0.021 (0.145)
black	-0.946*** (0.170)	-0.268 (0.229)	-0.628*** (0.213)	-0.982*** (0.249)	-0.830*** (0.245)	-0.820*** (0.219)	-1.019*** (0.217)	-0.928*** (0.284)	-0.902*** (0.233)	-1.190*** (0.335)
otherrace	-0.968*** (0.091)	-0.709*** (0.112)	-0.888*** (0.113)	-0.977*** (0.105)	-0.935*** (0.125)	-0.930*** (0.121)	-0.899*** (0.122)	-0.902*** (0.112)	-0.961*** (0.122)	-0.999*** (0.153)
currgdp	-0.107*** (0.024)	0.014 (0.028)	-0.028 (0.027)	-0.065*** (0.029)	-0.100*** (0.027)	-0.117*** (0.028)	-0.134*** (0.027)	-0.151*** (0.032)	-0.149*** (0.040)	-0.142* (0.073)
fschmiss	-1.122*** (0.104)	-0.934*** (0.125)	-0.754*** (0.124)	-0.995*** (0.118)	-1.117*** (0.142)	-1.306*** (0.150)	-1.423*** (0.162)	-1.329*** (0.155)	-1.296*** (0.163)	-1.064*** (0.165)
filliterate	-1.488*** (0.262)	-1.183*** (0.199)	-1.385*** (0.407)	-1.182*** (0.331)	-1.310*** (0.274)	-1.840*** (0.346)	-1.772*** (0.358)	-1.805*** (0.343)	-1.548*** (0.565)	-1.006* (0.560)
fuprimary	0.778*** (0.145)	0.869*** (0.240)	1.060*** (0.211)	0.985*** (0.207)	0.748*** (0.205)	0.658*** (0.183)	0.623*** (0.174)	0.618*** (0.199)	0.637*** (0.183)	0.754*** (0.253)
flhighschool	1.561*** (0.177)	1.851*** (0.376)	1.958*** (0.349)	1.922*** (0.277)	1.574*** (0.225)	1.214*** (0.226)	1.327*** (0.197)	1.392*** (0.228)	1.287*** (0.260)	1.043*** (0.243)
fcollge	2.483*** (0.202)	2.708*** (0.335)	2.694*** (0.371)	2.836*** (0.319)	2.683*** (0.345)	2.423*** (0.284)	2.477*** (0.268)	2.364*** (0.248)	1.963*** (0.228)	1.573*** (0.263)
mschmiss	-1.665*** (0.110)	-1.062*** (0.121)	-1.051*** (0.121)	-1.373*** (0.115)	-1.834*** (0.151)	-2.104*** (0.160)	-2.223*** (0.157)	-2.158*** (0.154)	-1.727*** (0.169)	-1.548*** (0.186)
milliterate	-1.461*** (0.253)	-1.950*** (0.276)	-1.481*** (0.449)	-1.513*** (0.270)	-1.648*** (0.398)	-1.686*** (0.383)	-1.424*** (0.345)	-1.310*** (0.350)	-0.891** (0.405)	-0.725** (0.348)
muprimary	0.917*** (0.158)	1.036*** (0.254)	1.466*** (0.220)	1.168*** (0.198)	1.006*** (0.194)	0.814*** (0.180)	0.553*** (0.205)	0.519** (0.205)	0.736*** (0.236)	0.876*** (0.263)
mhighschool	1.823*** (0.145)	3.156*** (0.297)	2.956*** (0.300)	2.286*** (0.214)	1.894*** (0.196)	1.580*** (0.182)	1.288*** (0.171)	1.230*** (0.214)	1.322*** (0.210)	1.315*** (0.265)

(Continued) Table 14a: Quantile Regression Estimates of Schooling Model -Urban Full Sample

Variables	OLS	q10	q20	q30	q40	q50	q60	q70	q80	q90
mcollege	2.465*** (0.276)	3.292*** (0.350)	3.054*** (0.451)	2.754*** (0.406)	2.317*** (0.455)	2.308*** (0.392)	2.139*** (0.349)	1.995*** (0.326)	2.092*** (0.307)	1.989*** (0.362)
age12_gdp	0.191*** (0.033)	0.091** (0.044)	0.075 (0.046)	0.138*** (0.046)	0.180*** (0.044)	0.185*** (0.043)	0.211*** (0.039)	0.218*** (0.042)	0.292*** (0.049)	0.313*** (0.069)
age7_teacher	-0.011 (0.031)	0.054 (0.046)	0.028 (0.042)	-0.010 (0.042)	-0.013 (0.043)	-0.025 (0.043)	-0.034 (0.038)	-0.010 (0.038)	-0.041 (0.042)	-0.015 (0.042)
age11_teacher	-0.060* (0.031)	-0.076* (0.045)	-0.050 (0.048)	-0.063 (0.049)	-0.059 (0.043)	-0.040 (0.040)	-0.053* (0.030)	-0.053* (0.028)	-0.075* (0.039)	-0.037 (0.046)
age7_hospital	1.899 (4.229)	-0.782 (4.550)	-8.921* (5.044)	-4.737 (5.364)	-3.043 (6.334)	3.691 (5.391)	2.471 (5.732)	7.158 (6.639)	8.674 (6.147)	14.318* (7.811)
age7_bed	0.055 (0.067)	0.153** (0.074)	0.153* (0.080)	0.132* (0.073)	0.059 (0.095)	0.014 (0.086)	0.061 (0.083)	0.025 (0.095)	0.039 (0.107)	-0.085 (0.119)
age7_doctor	0.003 (0.205)	-0.441 (0.289)	-0.271 (0.281)	-0.102 (0.281)	0.133 (0.295)	0.116 (0.270)	0.079 (0.235)	0.063 (0.268)	0.087 (0.289)	-0.100 (0.377)
Constant	-1.643** (0.750)	-0.619 (0.936)	-0.559 (1.056)	-1.875** (0.906)	-2.473** (1.077)	-2.916*** (1.034)	-2.268** (1.074)	-2.707** (1.068)	-1.353 (1.197)	-0.954 (1.430)
Observations	6,451	6,451	6,451	6,451	6,451	6,451	6,451	6,451	6,451	6,451

Note: Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table 14b: Quantile Regression Estimates of Schooling Model -Rural Full Sample

Variables	OLS	q10	q20	q30	q40	q50	q60	q70	q80	q90
startage	0.153*** (0.020)	0.021 (0.022)	0.053** (0.025)	0.097*** (0.022)	0.121*** (0.020)	0.127*** (0.023)	0.150*** (0.036)	0.195*** (0.034)	0.220*** (0.036)	0.244*** (0.044)
age	0.004 (0.074)	0.024 (0.065)	0.039 (0.072)	-0.107 (0.082)	-0.056 (0.066)	-0.004 (0.057)	0.023 (0.081)	0.050 (0.094)	0.022 (0.108)	-0.015 (0.152)
agesq	-0.000 (0.001)	-0.000 (0.001)	-0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	-0.000 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.000 (0.001)	0.000 (0.002)
male	-0.088 (0.130)	-0.183 (0.148)	-0.302** (0.147)	-0.338** (0.150)	-0.039 (0.131)	0.037 (0.130)	0.029 (0.178)	-0.011 (0.191)	-0.077 (0.206)	-0.111 (0.310)
black	-0.493* (0.268)	-0.530** (0.250)	-0.624** (0.278)	-0.616 (0.376)	-0.456* (0.253)	-0.442 (0.277)	-0.165 (0.352)	-0.311 (0.329)	-0.489 (0.397)	-0.537 (0.712)
otherrace	-0.321* (0.168)	-0.396** (0.181)	-0.296 (0.181)	-0.298* (0.176)	-0.283* (0.170)	-0.346** (0.145)	-0.334* (0.200)	-0.144 (0.206)	-0.060 (0.253)	-0.064 (0.351)
currgdp	0.020 (0.071)	0.115* (0.065)	0.186*** (0.068)	0.190** (0.086)	0.139* (0.072)	0.102 (0.068)	0.083 (0.088)	0.035 (0.080)	-0.007 (0.079)	-0.140 (0.139)
fschmiss	-1.054*** (0.155)	-0.661*** (0.192)	-0.889*** (0.190)	-0.966*** (0.166)	-0.941*** (0.135)	-0.848*** (0.125)	-0.903*** (0.184)	-0.785*** (0.211)	-1.280*** (0.317)	-1.113** (0.458)
filliterate	-1.424*** (0.275)	-0.794*** (0.249)	-1.158*** (0.250)	-1.599*** (0.346)	-1.286*** (0.332)	-1.066*** (0.302)	-1.247*** (0.387)	-1.259*** (0.439)	-1.074* (0.575)	-1.130* (0.627)
fuprimary	1.404** (0.563)	0.700 (0.559)	0.665 (0.783)	1.129 (0.753)	1.453* (0.763)	1.539* (0.818)	1.347 (0.877)	1.402 (0.956)	1.114 (1.048)	2.419** (1.017)
flhighschool	2.789*** (0.591)	4.468*** (1.541)	4.176*** (0.954)	3.353*** (1.095)	3.021*** (1.092)	3.054*** (0.993)	2.404** (0.991)	2.516*** (0.799)	1.960* (1.024)	2.798* (1.458)
fcollge	4.042*** (0.932)	7.607*** (2.249)	5.128*** (1.830)	4.139** (1.653)	4.122** (1.647)	4.175** (1.622)	4.296*** (1.649)	3.855*** (1.449)	3.268** (1.498)	1.974 (1.878)
mschmiss	-1.617*** (0.158)	-1.313*** (0.203)	-1.211*** (0.168)	-0.992*** (0.173)	-0.887*** (0.146)	-0.944*** (0.130)	-1.122*** (0.228)	-2.040*** (0.312)	-2.408*** (0.387)	-2.570*** (0.494)
milliterate	-1.498*** (0.283)	-1.468*** (0.243)	-1.533*** (0.297)	-1.246*** (0.431)	-1.055*** (0.397)	-0.819** (0.335)	-0.799** (0.386)	-1.566*** (0.451)	-1.885*** (0.526)	-2.331*** (0.726)
muprimary	0.906** (0.456)	0.639 (0.675)	0.364 (0.454)	0.806 (0.559)	0.751 (0.820)	1.821** (0.900)	1.999** (0.878)	1.310 (0.931)	1.395** (0.634)	0.629 (0.813)
mhighschool	1.905*** (0.614)	0.960 (1.824)	2.552** (1.457)	2.668** (1.307)	2.909** (1.248)	2.806** (1.164)	2.955*** (1.047)	1.628* (0.844)	0.939 (0.721)	0.180 (0.913)

(Continued) Table 14b: Quantile Regression Estimates of Schooling Model -Rural Full Sample

Variables	OLS	q10	q20	q30	q40	q50	q60	q70	q80	q90
mcollege	3.309* (1.813)	-0.408 (2.936)	2.789 (3.121)	2.571 (3.212)	4.751 (2.994)	4.212 (2.738)	4.919* (2.725)	5.004** (2.466)	4.017* (2.402)	4.869* (2.487)
age12_gdp	0.105 (0.078)	0.040 (0.070)	0.035 (0.071)	-0.052 (0.086)	-0.019 (0.079)	-0.001 (0.079)	0.044 (0.097)	0.132 (0.105)	0.203* (0.106)	0.280* (0.156)
age7_teacher	0.001 (0.058)	0.036 (0.049)	0.034 (0.047)	-0.003 (0.053)	-0.031 (0.055)	0.009 (0.066)	0.019 (0.084)	0.056 (0.103)	-0.039 (0.129)	-0.129 (0.160)
age11_teacher	0.041 (0.047)	-0.007 (0.030)	-0.028 (0.040)	0.002 (0.045)	0.011 (0.034)	-0.017 (0.039)	-0.012 (0.056)	-0.017 (0.091)	0.104 (0.117)	0.236 (0.146)
age7_hospital	-7.311 (8.860)	0.762 (7.975)	9.146 (7.947)	1.441 (9.584)	-1.636 (8.982)	0.309 (8.409)	-1.991 (11.198)	-7.254 (11.392)	-1.123 (11.467)	-13.619 (15.675)
age7_bed	-0.202 (0.128)	-0.097 (0.150)	-0.255* (0.136)	-0.085 (0.141)	-0.096 (0.123)	-0.203* (0.111)	-0.227 (0.142)	-0.317** (0.142)	-0.334** (0.159)	-0.163 (0.229)
age7_doctor	-0.275 (0.310)	0.188 (0.375)	0.381 (0.401)	-0.100 (0.328)	0.088 (0.316)	0.223 (0.303)	0.002 (0.344)	-0.174 (0.426)	-0.407 (0.455)	-1.200* (0.633)
Constant	4.830*** (1.471)	1.511 (1.338)	1.613 (1.444)	4.946*** (1.690)	4.234*** (1.413)	3.934*** (1.198)	4.158*** (1.723)	4.793*** (1.891)	5.958*** (2.236)	8.335*** (2.980)
Observations	1,578	1,578	1,578	1,578	1,578	1,578	1,578	1,578	1,578	1,578

Note: Clustered robust standard errors are reported in parentheses. ** Significant at 1% level, *** Significant at 5% level, * Significant at 10% level.

Figure 1: Distribution of Age Started to Work (%)

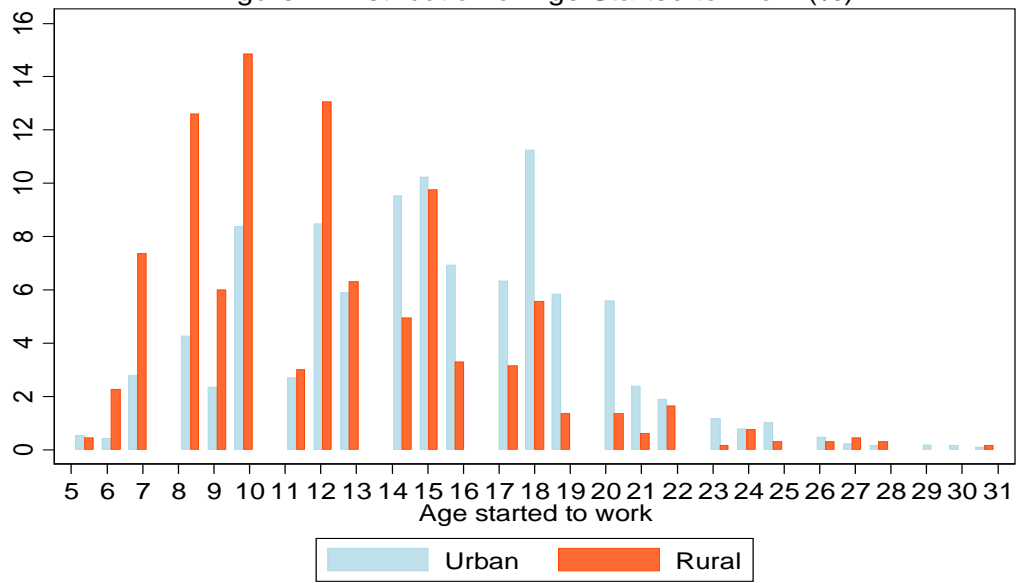


Figure 2: Log-earnings by Age Started to Work

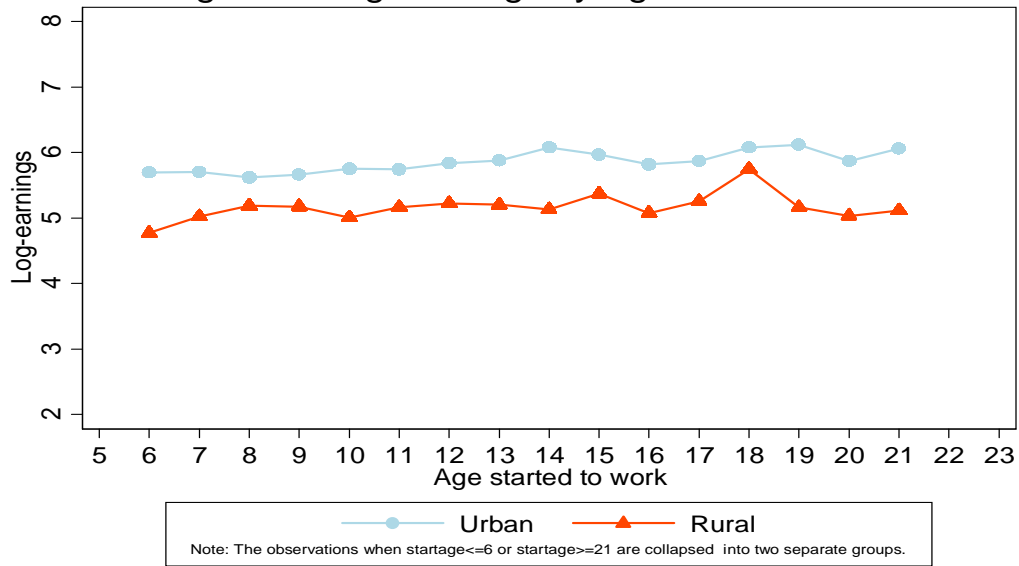


Figure 3: Health Score by Age Started to Work

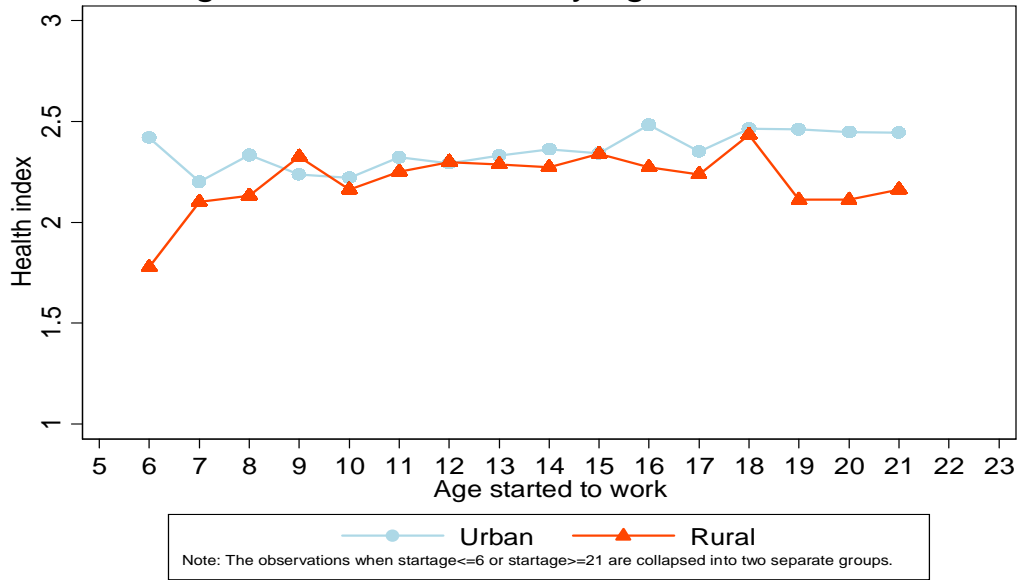


Figure 4a: Years of Schooling by Age Started to Work

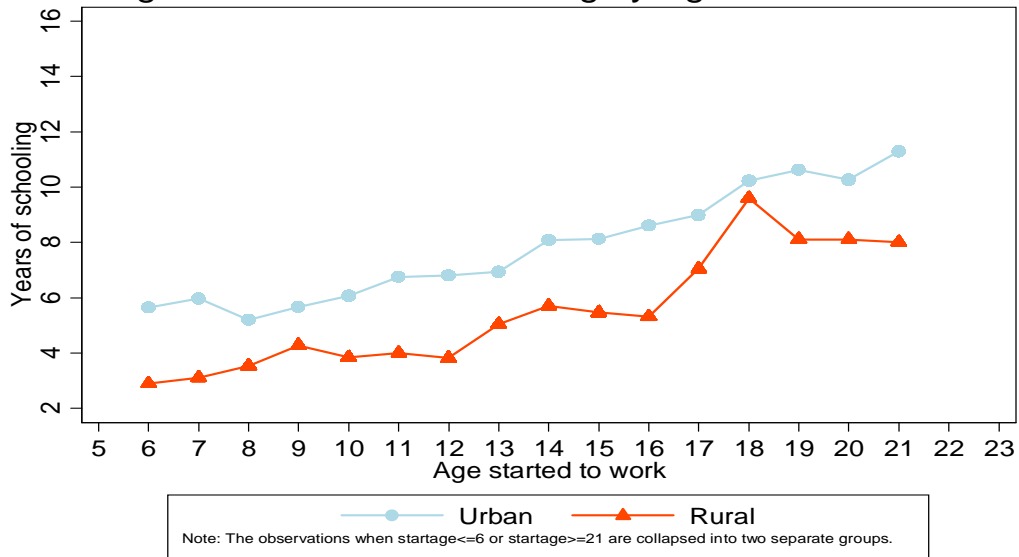
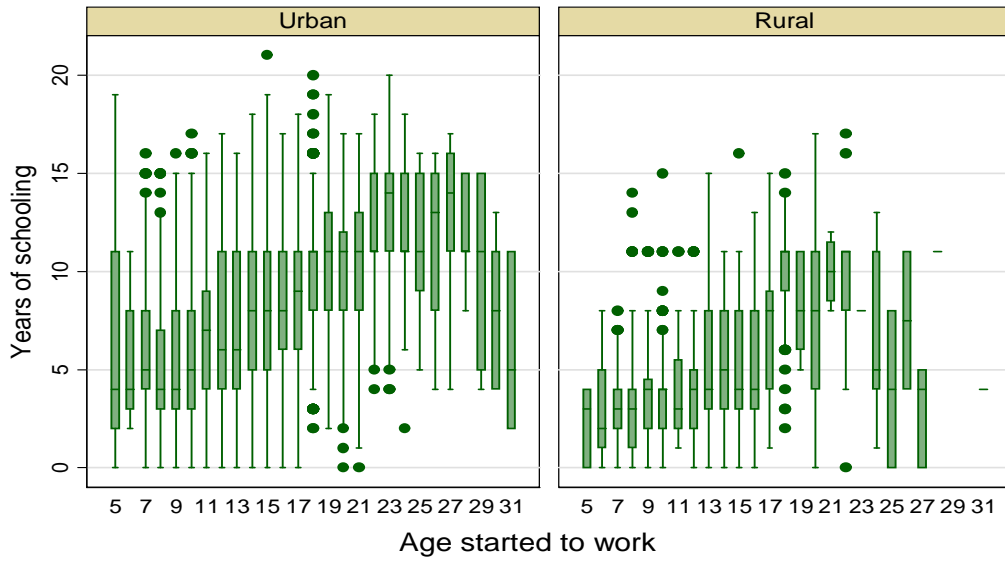


Figure 4b: Years of Schooling by Age Started to Work



Graphs by areas

Appendix

Table A.1. Variable Definitions

Variable	Definition
<i>mlninc</i>	Log of monthly income
<i>chealth</i>	Self-assessed health index
<i>ehealth</i>	1 if report very good or excellent health, 0 otherwise
<i>ghealth</i>	1 if report good health, 0 otherwise
<i>phealth</i>	1 if report poor or average health, 0 otherwise (reference group)
<i>hproblem</i>	1 if report having health problems that require constant monitoring, 0 otherwise
<i>sch</i>	Years of schooling
<i>startage</i>	Age started to work
<i>age</i>	Age
<i>agesq</i>	Age-squared
<i>male</i>	1 if male, 0 otherwise
<i>black</i>	1 if black people, 0 otherwise
<i>othrace</i>	1 if mixed-raced or Indian people, 0 otherwise
<i>white</i>	1 if white people, 0 otherwise (reference group)
<i>currgdp</i>	GDP per capita of the state where the individual currently reside
<i>Father's education</i>	
<i>fschmiss</i>	1 if father's schooling information is missing, 0 otherwise
<i>filliterate</i>	1 if father illiterate, 0 otherwise
<i>flowprimary</i>	1 if father some or completed lower primary, 0 otherwise (reference group)
<i>fupprimary</i>	1 if father some or completed upper primary, 0 otherwise
<i>fhighschool</i>	1 if father some or completed high school, 0 otherwise
<i>fcollege</i>	1 if father some or completed college, 0 otherwise
<i>Mother's education</i>	
<i>mschmiss</i>	1 if mother's schooling information is missing, 0 otherwise
<i>milliterate</i>	1 if mother illiterate, 0 otherwise
<i>mlowprimary</i>	1 if mother some or completed lower primary, 0 otherwise (reference group)
<i>mupprimary</i>	1 if mother some or completed upper primary, 0 otherwise
<i>mhighschool</i>	1 if mother some or completed high school, 0 otherwise
<i>mcollege</i>	1 if mother some or completed college, 0 otherwise
<i>Father's occupation</i>	
<i>focumiss</i>	1 if father's occupation information is missing, 0 otherwise
<i>fnwork</i>	1 if father doesn't work, 0 otherwise
<i>fselfemploy</i>	1 if father is self-employed, 0 otherwise
<i>femployee</i>	1 if father is an employee, 0 otherwise (reference group)
<i>femployer</i>	1 if father is an employer, 0 otherwise
<i>Mother's occupation</i>	
<i>mocumiss</i>	1 if mother's occupation information is missing, 0 otherwise
<i>mnwork</i>	1 if mother doesn't work, 0 otherwise
<i>mselfemploy</i>	1 if mother is self-employed, 0 otherwise
<i>memployee</i>	1 if mother is an employee, 0 otherwise (reference group)
<i>memployer</i>	1 if mother is an employer, 0 otherwise
<i>munsalaried</i>	1 if mother works without a salary, 0 otherwise
<i>age12_gdp</i>	GDP per capita of the state where the individual lived when 12 years old
<i>age7_teacher</i>	# of teachers per school in the state where the individual lived when 7 years old
<i>age11_teacher</i>	# of teachers per school in the state where the individual lived when 11 years old
<i>age7_hospital</i>	# of hospitals per 1000 inhabitants in the state where the individual lived when 7 years old
<i>age7_bed</i>	# of beds per 1000 inhabitants in the state where the individual lived when 7 years old
<i>age7_doctor</i>	# of doctors per 1000 inhabitants in the state where the individual lived when 7 years old

Table A.2: Ordered Probit Estimates of Health Model (2a) with Full Sample

Variables	Urban_1	Urban_2	Urban_3	Rural_1	Rural_2	Rural_3
startage	-0.002*** (0.001)	-0.002*** (0.001)	0.004*** (0.001)	-0.006*** (0.002)	-0.002*** (0.001)	0.008*** (0.003)
sch	-0.008*** (0.001)	-0.006*** (0.001)	0.014*** (0.002)	-0.009*** (0.003)	-0.002*** (0.001)	0.012*** (0.004)
age	0.005 (0.003)	0.004 (0.002)	-0.009 (0.006)	-0.005 (0.008)	-0.001 (0.002)	0.007 (0.011)
agesq	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
male	-0.056*** (0.007)	-0.040*** (0.005)	0.096*** (0.011)	-0.103*** (0.017)	-0.023*** (0.005)	0.125*** (0.020)
black	0.011 (0.013)	0.007 (0.008)	-0.019 (0.022)	0.042 (0.038)	0.007** (0.003)	-0.050 (0.041)
otherrace	0.021*** (0.008)	0.015*** (0.005)	-0.036*** (0.013)	0.045** (0.020)	0.012** (0.006)	-0.056** (0.025)
currqdp	-0.006*** (0.002)	-0.004*** (0.001)	0.011*** (0.003)	-0.020*** (0.006)	-0.005*** (0.002)	0.025*** (0.008)
fschmiss	0.016* (0.009)	0.011* (0.006)	-0.027* (0.015)	0.031 (0.019)	0.008 (0.005)	-0.040 (0.024)
filliterate	0.018 (0.020)	0.011 (0.012)	-0.029 (0.031)	0.012 (0.035)	0.003 (0.007)	-0.015 (0.043)
fupprimary	-0.015 (0.011)	-0.012 (0.009)	0.027 (0.021)	0.003 (0.052)	0.001 (0.013)	-0.004 (0.064)
fhighschool	-0.022 (0.014)	-0.017 (0.013)	0.039 (0.027)	0.035 (0.077)	0.006 (0.007)	-0.041 (0.084)
fcollge	-0.050*** (0.015)	-0.047** (0.019)	0.097*** (0.034)	-0.088 (0.079)	-0.051 (0.078)	0.138 (0.157)
mschmiss	0.008 (0.009)	0.006 (0.006)	-0.014 (0.015)	0.007 (0.021)	0.002 (0.005)	-0.009 (0.026)
milliterate	0.018 (0.022)	0.012 (0.013)	-0.030 (0.035)	-0.014 (0.038)	-0.004 (0.013)	0.018 (0.051)
mupprimary	0.006 (0.013)	0.004 (0.008)	-0.011 (0.021)	-0.040 (0.046)	-0.015 (0.024)	0.056 (0.069)
mhighschool	-0.003 (0.015)	-0.002 (0.011)	0.005 (0.025)	0.016 (0.073)	0.004 (0.013)	-0.020 (0.087)
mcollege	-0.036 (0.022)	-0.031 (0.024)	0.067 (0.046)	-0.062 (0.109)	-0.029 (0.077)	0.090 (0.186)
age7_hospital	0.392 (0.312)	0.279 (0.223)	-0.670 (0.535)	-1.421 (1.040)	-0.374 (0.289)	1.796 (1.323)
age7_bed	-0.018*** (0.005)	-0.013*** (0.003)	0.031*** (0.008)	0.011 (0.015)	0.003 (0.004)	-0.014 (0.019)
age7_doctor	0.065*** (0.012)	0.046*** (0.009)	-0.112*** (0.020)	-0.010 (0.033)	-0.003 (0.009)	0.013 (0.042)
Observations	6,439	6,439	6,439	1,573	1,573	1,573

Notes: Marginal effects rather than ordered probit estimates are reported. Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table A.3: IV Ordered Probit Estimates - First-stage Regression of Health Model (2a) with Full Sample

Variables	startage_u	sch_u	startage_r	sch_r
age	0.247*** (0.060)	0.434*** (0.040)	0.164 (0.102)	0.031 (0.074)
agesq	-0.004*** (0.001)	-0.006*** (0.001)	-0.002* (0.001)	-0.001 (0.001)
male	-1.775*** (0.113)	-0.351*** (0.082)	-2.136*** (0.220)	-0.493*** (0.128)
black	-0.487** (0.203)	-0.953*** (0.171)	-0.760* (0.397)	-0.538** (0.275)
otherrace	-0.253** (0.118)	-0.931*** (0.093)	-0.306 (0.219)	-0.352** (0.177)
curr GDP	-0.167*** (0.032)	-0.131*** (0.025)	-0.259*** (0.090)	-0.014 (0.070)
fschmiss	-0.903*** (0.146)	-1.296*** (0.107)	-0.700*** (0.234)	-1.077*** (0.160)
filliterate	-1.133*** (0.319)	-1.700*** (0.277)	0.302 (0.360)	-1.324*** (0.284)
fupprimary	0.439** (0.187)	0.868*** (0.145)	1.315** (0.660)	1.683*** (0.562)
fhighschool	1.316*** (0.194)	1.827*** (0.182)	2.459*** (0.890)	3.023*** (0.586)
fcollge	1.750*** (0.271)	2.753*** (0.215)	3.101** (1.440)	4.422*** (0.967)
mschmiss	-0.984*** (0.129)	-1.871*** (0.110)	-0.617*** (0.228)	-1.669*** (0.161)
milliterate	-0.917*** (0.290)	-1.660*** (0.254)	-0.977*** (0.361)	-1.642*** (0.285)
mupprimary	0.648*** (0.182)	1.073*** (0.164)	1.210 (0.749)	0.997** (0.490)
mhighschool	1.570*** (0.204)	2.208*** (0.153)	3.733*** (0.890)	2.702*** (0.636)
mcollege	1.782*** (0.377)	3.163*** (0.309)	0.067 (1.669)	3.590* (1.896)
age7_hospital	3.295 (6.290)	2.758 (4.077)	-15.228 (11.207)	-8.149 (9.305)
age7_bed	-0.087 (0.101)	0.035 (0.071)	0.009 (0.189)	-0.218* (0.132)
age7_doctor	0.769** (0.352)	0.193 (0.214)	-0.110 (0.465)	-0.307 (0.351)
focumiss	0.072 (0.270)	0.215 (0.202)	0.046 (0.760)	-0.100 (0.483)
fnwork	-0.444*** (0.169)	0.125 (0.137)	-0.180 (0.361)	-0.196 (0.256)
fselfemploy	-0.717*** (0.139)	-0.040 (0.103)	-0.556** (0.257)	0.106 (0.152)

(Continued) Table A.3: IV Ordered Probit Estimates - First-stage Regression of Health Model (2a) with Full Sample

Variables	startage_u	sch_u	startage_r	sch_r
employer	-0.416*	0.890***	0.232	0.841*
	(0.250)	(0.226)	(0.733)	(0.451)
mocumiss	0.383	0.827**	-0.335	1.470**
	(0.532)	(0.388)	(0.884)	(0.591)
mnwork	0.857***	0.743***	0.437	0.687**
	(0.109)	(0.101)	(0.349)	(0.331)
mselfemploy	-0.007	0.505***	0.188	0.842***
	(0.167)	(0.129)	(0.400)	(0.299)
memployer	-0.067	1.135***	-0.185	3.604*
	(0.495)	(0.439)	(1.021)	(2.042)
munsalary	-1.536***	-0.344	-1.453***	-0.173
	(0.259)	(0.220)	(0.401)	(0.317)
age12_gdp	-0.007	0.187***	0.200**	0.147**
	(0.049)	(0.035)	(0.097)	(0.074)
age7_teacher	-0.022	-0.025	0.163**	0.014
	(0.037)	(0.033)	(0.077)	(0.061)
age11_teacher	0.018	-0.053	-0.147***	0.028
	(0.038)	(0.034)	(0.055)	(0.054)
Constant	12.492***	0.659	12.790***	6.146***
	(1.195)	(0.759)	(2.090)	(1.533)
Observations	6,439	6,439	1,573	1,573
Test of excluded instruments				
$\chi^2(12)$	201.75	121.63	109.30	52.42
P-value	0.000	0.000	0.000	0.000

Notes: Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table A.4: Probit Estimates of Health Model(2b) with Full Sample

Variables	Urban	Rural
startage	-0.005*** (0.001)	-0.005* (0.003)
sch	-0.001 (0.002)	-0.008** (0.004)
age	0.005 (0.005)	0.006 (0.011)
agesq	-0.000 (0.000)	-0.000 (0.000)
male	-0.059*** (0.010)	-0.034 (0.021)
black	0.003 (0.022)	0.015 (0.042)
otherrace	0.039*** (0.012)	-0.009 (0.025)
currngdp	-0.015*** (0.003)	-0.003 (0.008)
fschmiss	0.004 (0.013)	0.047* (0.026)
filliterate	0.028 (0.034)	0.013 (0.044)
fupprimary	-0.004 (0.018)	0.045 (0.100)
fhighschool	-0.023 (0.024)	-0.053 (0.081)
fcollge	-0.031 (0.031)	-0.020 (0.127)
mschmiss	-0.016 (0.013)	-0.014 (0.027)
milliterate	0.029 (0.031)	0.015 (0.046)
mupprimary	0.035* (0.021)	-0.051 (0.065)
mhighschool	0.033 (0.026)	0.003 (0.092)
mcollege	0.074 (0.046)	0.035 (0.206)
age7_hospital	0.453 (0.547)	1.507 (1.224)
age7_bed	-0.017** (0.008)	-0.042** (0.018)
age7_doctor	0.066*** (0.023)	-0.015 (0.046)
Observations	6,439	1,573

Notes: Marginal effects are reported rather than probit coefficients. Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.

Table A.5: IV Probit Estimates - first-stage Regression of Health Model(2b) with Full Sample

Variables	startage_u	sch_u	sartage_r	sch_r
age	0.250*** (0.060)	0.439*** (0.040)	0.146 (0.099)	0.036 (0.073)
agesq	-0.004*** (0.001)	-0.006*** (0.001)	-0.002 (0.001)	-0.001 (0.001)
male	-1.775*** (0.112)	-0.350*** (0.082)	-2.137*** (0.216)	-0.494*** (0.125)
black	-0.485** (0.200)	-0.953*** (0.170)	-0.737* (0.381)	-0.543** (0.263)
otherrace	-0.253** (0.117)	-0.931*** (0.093)	-0.291 (0.211)	-0.355** (0.172)
curr GDP	-0.171*** (0.032)	-0.135*** (0.025)	-0.233*** (0.085)	-0.020 (0.069)
fschmiss	-0.898*** (0.145)	-1.289*** (0.107)	-0.691*** (0.227)	-1.078*** (0.154)
filliterate	-1.132*** (0.320)	-1.698*** (0.279)	0.314 (0.344)	-1.329*** (0.279)
fupprimary	0.440** (0.187)	0.868*** (0.145)	1.273** (0.625)	1.696*** (0.590)
fhighschool	1.319*** (0.193)	1.828*** (0.183)	2.459*** (0.869)	3.025*** (0.561)
fcollege	1.752*** (0.267)	2.756*** (0.214)	3.013** (1.375)	4.452*** (0.917)
mschmiss	-0.985*** (0.128)	-1.869*** (0.110)	-0.627*** (0.221)	-1.666*** (0.160)
milliterate	-0.913*** (0.286)	-1.656*** (0.254)	-0.988*** (0.342)	-1.639*** (0.275)
mupprimary	0.646*** (0.181)	1.068*** (0.164)	1.187 (0.768)	1.002** (0.488)
mhighschool	1.562*** (0.199)	2.192*** (0.153)	3.702*** (0.889)	2.718*** (0.616)
mcollege	1.767*** (0.379)	3.140*** (0.309)	0.009 (1.438)	3.616** (1.740)
age7_hospital	3.537 (6.208)	3.071 (4.069)	-11.209 (10.933)	-9.498 (9.015)
age7_bed	-0.090 (0.101)	0.030 (0.071)	-0.074 (0.182)	-0.193 (0.130)
age7_doctor	0.766** (0.341)	0.195 (0.214)	-0.426 (0.455)	-0.207 (0.336)
focumiss	0.055 (0.259)	0.191 (0.209)	-0.040 (0.647)	-0.084 (0.499)
fnwork	-0.496*** (0.162)	0.057 (0.140)	-0.024 (0.362)	-0.229 (0.246)
fselfemploy	-0.710*** (0.135)	-0.049 (0.103)	-0.356 (0.235)	0.053 (0.155)

Table A.5: IV Probit Estimates - first-stage Regression of Health Model(2b) with Full Sample

Variables	startage_u	sch_u	sartage_r	sch_r
femployer	-0.399 (0.248)	0.927*** (0.225)	0.409 (0.670)	0.800* (0.439)
mocumiss	0.433 (0.502)	0.782** (0.399)	-0.014 (0.707)	1.350** (0.587)
mnwork	0.834*** (0.105)	0.705*** (0.102)	0.236 (0.266)	0.766*** (0.188)
mselfemploy	-0.035 (0.163)	0.481*** (0.134)	0.003 (0.355)	0.897*** (0.284)
memployer	-0.152 (0.471)	1.042** (0.454)	-0.554 (0.819)	3.764*** (1.456)
munsalary	-1.577*** (0.256)	-0.401* (0.220)	-1.652*** (0.298)	-0.099 (0.212)
age12_gdp	0.000 (0.049)	0.196*** (0.035)	0.146* (0.089)	0.161** (0.077)
age7_teacher	-0.021 (0.036)	-0.022 (0.031)	0.235*** (0.070)	-0.006 (0.062)
age11_teacher	0.016 (0.036)	-0.058* (0.033)	-0.115** (0.054)	0.015 (0.046)
Constant	12.453*** (1.191)	0.604 (0.762)	13.049*** (2.001)	6.064*** (1.508)
Observations	6,439	6,439	1,573	1,573
Test of excluded instruments				
$\chi^2(12)$	206.06	120.52	142.27	56.36
P-value	0.000	0.000	0.000	0.000

Notes: Clustered robust standard errors are reported in parentheses. *** Significant at 1% level, ** Significant at 5% level, * Significant at 10% level.