Report No: AUS19668

Kingdom of Cambodia
Schooling, Skills, and Success

Trends and linkages in schooling and work among Cambodian youth: A cohort panel analysis

March 28, 2017

GED02
EAST ASIA AND PACIFIC

Trends and linkages in schooling and work among Cambodian youth: A cohort panel analysis

March 2017

Haijing Huang[1] Deon Filmer[2] Tsuyoshi Fukao[3]

Introduction

Cambodia's education sector has faced and overcome a number of challenges in recent history. Several decades of political and social unrest caused by the Khmer Rouge regime of the 1970s and Vietnamese occupation in the 1980's dealt a severe blow to the education system and left it in a state of disintegration. Primary and secondary enrollment through to the 1980's fell, with school attendance dramatically lower for individuals who were teenagers in 1975 compared to previous or subsequent cohorts (de Walque, 2004). There were improvements by the following decade: the Paris Agreements and beginnings of UN sponsored elections ushered in a renewed focus on building and reconstructing schools and increasing the national budget allocation toward education, which reached 15.7 percent in 2001 (GAD/C 2002).

Since the 90's, the country has made education a cornerstone of their long-term development strategy. The Cambodian government initiated a series of reforms including large-scale school construction and several scholarship programs to increase education access for students from disadvantaged backgrounds. Starting 2000's, the Ministry of Education, Youth and Sport (MoEYS) prioritized expanding secondary schooling to address the stark discrepancy between number of secondary schools compared to primary schools – a ratio of about 7.6 to 1 (Benveniste et. al. 2008). In this time the number of secondary schools grew by 16 percent from 1998 to 2004, with a faster growth rate compared to primary schools. Many were built in districts which previously did not have one: in 1998 32 out of Cambodia's 183 districts did not have a lower secondary school and 81 did not have an upper secondary school. By 2004, this was reduced to 14 and 45 respectively (Benveniste et. al. 2008).

Figures on education spending reflect these efforts: from 2000 to 2014, primary education spending as a proportion of GDP has decreased slightly, while secondary education spending has increased (Figure 1). Education expenditure per student in terms of PPP$ have increased steadily since 2000, with a much

[1] Pardee RAND Graduate School & RAND Corporation (chuang1@rand.org)
[2] Development Research Group, World Bank (dfilmer@worldbank.org)
[3] Education Global Practice, World Bank (tfukao@worldbank.org)

faster increase for lower secondary spending[3] (Figure 2). Education inputs have also improved: indicators of school quality, such as the number of classrooms with better quality materials, numbers of facilities with drinking water and toilets also increased between 2003 and 2004 for both rural and urban areas (Beneviste and Araujo, 2008).

Figure 1. Education Spending as Proportion of GDP, 2000-2014

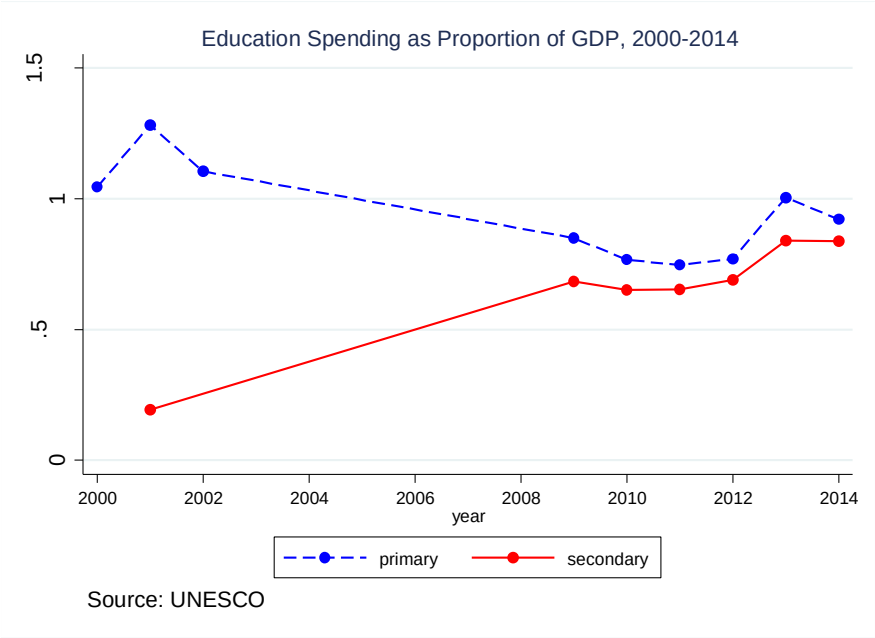
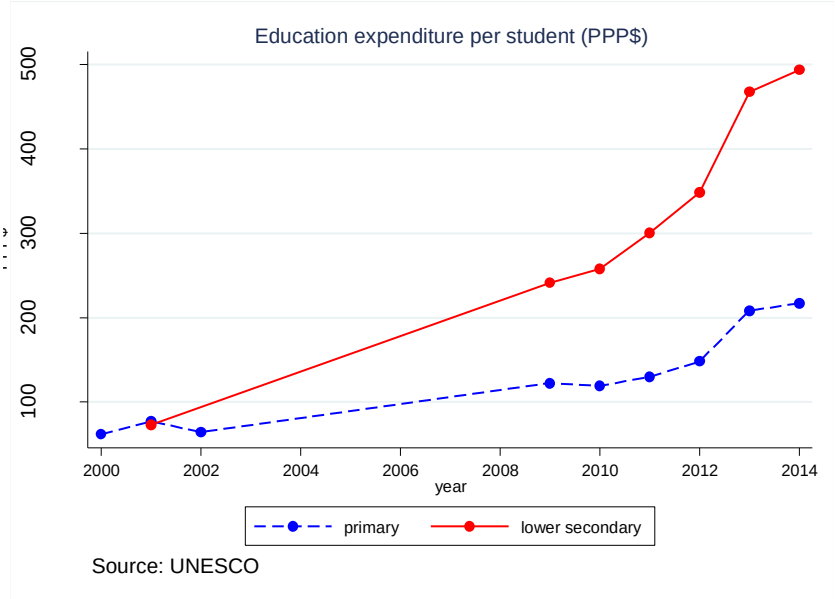


Figure 2. Education expenditure per student (PPP$), 2000-2014



[3] UNESCO data on expenditures per upper secondary student is only available for one year (2001) and hence is not included here.

Recent generations of youth enjoy greater access to schooling than previous ones; 49 percent of youth finish their education at a level higher than their father and 63 percent finish at a level higher than their mother (ILO, 2013). Net primary enrollments increased from 84 percent in 1992 to 96.4 percent in 2012, and net secondary enrollments from 16.6 percent in 2000 to 35 percent in 2012 (Tandon and Fukao, 2015). The labor market in Cambodia also went through a transformation in part due tostrong economic growth for the last two dacades: salaried employment rose one-third in this period from 23 percent in 2004 to 30 percent in 2011 (ILO, 2013). As of 2015, Cambodia has attained the lower-middle-income status, with gross national income (GNI) per capita reaching US$1,070.

Though Cambodia has come a long way since the 70s and 80s in terms of education and employment, several challenges remain. Only 30 percent of youth complete secondary schooling, and there remain notable gender and regional disparities in education attainment. Among urban villages, only 22 percent have a lower secondary school and 13 percent have an upper secondary school in 2011; this number is lower for rural areas (16 and 5 percent respectively). Children living in rural areas also face greater barriers to access: the closest primary and lower secondary schools in rural areas are more than three times the distance in urban areas; and for upper secondary schools it is almost five times the distance of urban villages (Table 1). Gender differentials also remain: the share of females who never attended school is double that of males (20.5 versus 10.9 percent, ILO 2013); and fewer women complete schooling at higher than primary level compared to males.

Table 1. Village Education Indicators, by Region 2011

	Primary school in village	distance to nearest primary school (km)	lower secondary school in village	distance to nearest lower secondary school (km)	upper secondary school in village	distance to nearest upper secondary school (km)
Rural (n=224)	58%	2.79	16%	4.6	5%	10.3
Urban (n=136)	49%	0.81	22%	1.6	13%	2.0

Source: CSES 2011

While unemployment is low, challenges to youth employment include difficulties in finding stable work: majority of youth are unpaid family (51 percent) or own account workers (16 percent). Only a negligible few are able to move from informal work to stable employment (ILO 2013), with rural youth facing greater difficulties in doing so. In terms of employment, more females than males are contributing (unpaid) family workers (27.2 and 1.5 per cent, respectively) and more males than females are wage and salaried workers (51.5 and 39.9 per cent, respectively).

Research objectives

Against the backdrop of Cambodia's evolving economic and political landscape and its present-day challenges, this paper has two main objectives. Our first objective is to gain a descriptive understanding of how individuals born in different years between 1980 and 1998 experienced schooling and work throughout their youth and early adulthood. During the several decades of rapid societal change, to what extent has schooling attainment and work outcomes differed across cohorts born within 19 years of each other? Which outcomes improved, have any declined, and by how much? Given the regional and gender disparities in these outcomes, have more disadvantaged subgroups experienced improvements or declines at a similar, faster or slower pace?

To this end, we construct a pseudo-panel of birth cohorts using eleven years of cross-sectional household surveys from 2003 to 2014, and follow these cohorts through their schooling and labor market trajectories. This method allows a richer dynamic analysis compared to approaches previously used which rely on cross-sectional snapshots of individuals observed at the same age during different time periods. Pseudo-panel analysis has commonly been used to estimate the life-cycle profile for savings (Attanasio 1998; Deaton and Paxson 1994), and has been applied to studying youth unemployment in the UK (Burgess 2003), the persistence of youth informality in Brazil and Argentina (Cruces et. al. 2013) and idle youth in Latin America and the Caribbean (Szekely and Karver, 2014).

Our second objective is to exploit the variation in schooling and work experiences across cohorts to provide evidence on how youth experiences translate to outcomes later in life. Specifically, have changes in education attainment across cohorts during their youth translated to better work outcomes into early adulthood? Does the *type* of work done by youth early in their work life – unpaid or agriculture work– affect what they do as adults? The latter question speaks to the broader literature on the persistence of informality. Previous research on youth labor market outcomes suggest that young people tend to earn lower wages and have less access to formal jobs than adults . In the context of Cambodia, we aim to understand whether these early labor market experiences are temporary spells that characterize only the beginning of youth work experience, or have lasting effects into adulthood.

We use cohort-panel analysis in two ways each appropriate to the goals of the different questions. The specific method for each empirical exercise is detailed in their respective sections. Section 1 presents the data, a general overview of the cohort panel approach, and initial visual evidence on and discussion of cross-cohort patterns. Section 2 presents methods and results from our empirical exercise disentangling cohort versus age and time effects. Here we are able to get a more accurate insight into our first objective.

Section 3 uses a regression analysis framework to present evidence on the relationship between early cohort experiences and labor adult labor market outcomes, our second objective. Section 4 concludes.

I. Data and Cohort-Panel Approach

Survey Data

The Cambodia Socio-Economic Survey (CSES), conducted by the National Institute of Statistics of the Ministry of Planning, collects information from a nationally-representative cross-section of households. The first survey was done in 1994. Since then households across Cambodia have been surveyed in 1996, 1997, 1999, 2004, and annually since 2007. "Large" CSES survey rounds were implemented in 2004, 2009 and 2014, with around 50,000 individuals annually. "Regular" survey rounds were in 2007, 2010, 2011, 2012 and 2013, with around 17,000 individuals annually.

We use data from the years 2004, 2007-2014[4], to maintain the continuity in survey years in constructing our dataset. In CSES round 2004, individuals had interview years from 2003 to 2005; hence our analysis contains consecutive survey years from 2003 to 2014, with the exception of 2006. The CSES contains information about schooling and work for each individual. Since a number of questions in these two areas repeat across survey round, we can harmonize the variables used for our analysis: age, gender, economic activity, region, and schooling.

Cohort Panel Approach

The questions we are typically interested in require panel data following specific individuals over several years through their life cycle. In Cambodia, as in many countries, there is a lack of individual panel data but an availability of repeated cross-sections – the CSES – which are drawn from random samples from the population in each year they are fielded. In such cases, an alternative "pseudo-panel" method developed by Angus Deaton in his seminal 1985 paper can be used (Deaton, 1985). The method posits that individuals sharing immutable characteristics, most commonly birth year and gender, can be grouped into cohorts. Individual outcomes are then averaged to the cohort level. Because the variables defining cohort membership are present in every survey year, we are able to follow the average behavior or

[4] We use consecutive years from 2004 to 2014 (barring 2006) to ensure that the variables we construct are derived from the same questions asked across all years, and to have continuity through time. We omit the earlier years because some variables we are interested in are not available in these years or are asked differently.

experience of a cohort as they age. These cohort-level averages through time comprise a pseudo, or "cohort" panel.

There are advantages and disadvantages to using cohort panel data over individual panel data. The biggest advantage is overcoming selection issues that typically occur when a lengthy period of time passes between interviews. In these cases attrition rates may be substantial, leading to a non-random or non-representative sample of remaining respondents (Antman and McKenzie, 2007). The cohort method avoids this by construction, as it uses repeated cross-sections drawing from a new random and representative sample from the same underlying distribution every year. By avoiding the attrition issue, the analysis can also be extended for longer time periods than traditional panel data. Of course, there are also limitations to this approach. First, as a product of constructing a cohort panel, the main limitation is that we cannot delve into intra-cohort dynamics. This means we cannot evaluate changes which occur to subgroups based on time-varying characteristics- for example, how outcomes may change among the more educated within a cohort. The second limitation is that since we are aggregating means and estimating relationships between them at the cohort level, this aggregation may contain measurement error and hence produce bias. This point will be further explained in the paragraph on sample sizes below.

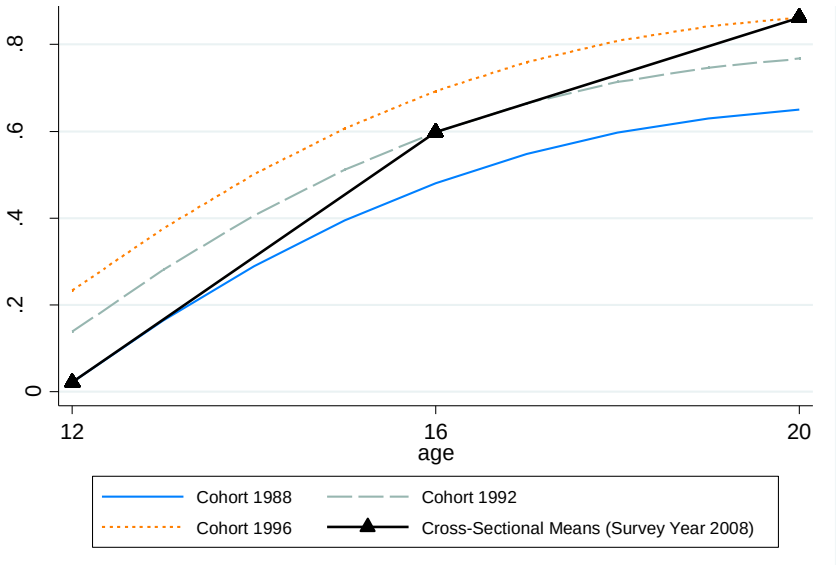
Determining what constitutes a "cohort" requires a consideration of sample size. A cohort panel is comprised of a series of `group-wise' means; hence, the larger the sample size within a pre-defined cohort the less error is likely contained in the means and the closer they are to the true population average (discussed in detail in Verbeek 2008). A few papers group cohorts by two or three-birth years to increase the within-cohort sample size (Szekely and Karver, 2014; Burgess 2003). These papers tend to use survey data that span several decades and have fewer sample sizes in each survey round relative to our data. Because our data span a little over one decade and grouping multiple birth years would greatly decrease our number of total cohort-by-time observations, we define a cohort to be single year of birth. Appendix Table A1 shows the cell sizes for each cohort-by-time observation. The majority of cell sizes fall between 100 and 500, with the large survey years producing cell sizes greater than 1,000. Our average cell size is 567. This is comparable to the cohort sizes discussed in Verbeek 2008, where he shows that many empirical studies have averages that are similar to ours: Browning, Deaton and Irish (1985) have an average of 190; Bloundell Browning and Meghir (1994) have cohort sizes around 500, for example. Following Deaton (1985), we correct for the differences in sample size per cohort by weighting cell means by the square root of the cell size.

Our final pseudo-panel includes a total of 19 cohorts from birth-years 1980 to 1998, and followed for 11 survey years. Since we are interested in outcomes that occur during youth and into early adulthood, we restrict our analysis to those between age 9 and 30. Hence, the first cohort is aged 23 in 2003 and 30 in 2010. The last cohort is aged 9 in 2007 and 16 in 2014. Appendix Table A2 shows the aging pattern for each cohort and the degree of overlap in ages across cohorts. Note this represents the overall range for our analysis; the exact range of cohorts and ages used may vary slightly depending on the empirical exercise to be described in their respective sections.

Patterns from cohort by age graphs

The main novelty of using a cohort approach is that it can provide a richer picture of outcomes of interest than simply using estimates from cross-sections alone. For example, if we are interested in how a person's employment status may change from youth to adulthood, the "snapshot" offered by a single cross-section can mask important information. If there are pronounced differences across cohorts, cross-sectional means would confound the cohort effect of individuals born a certain year with the experience of a particular individual of that age during that survey year. Consider the stylized example in Figure 1. There is a level shift in primary completion across the three cohorts: 1988, 1992 and 1996. Here, we are using smoothed means for a clearer illustration but the raw means produce a similar story. If we were to use estimates from survey year 2008 alone at age 12, 16 and 20, one can see that the resulting age profile picks up means across multiple birth cohorts. Cohort-panel analysis allows one to control for cohort as well as contemporaneous time effects to back out the age profile; conversely, one can control for age and time effects to trace trends across cohorts, which we do in Section II.

Figure 3. Primary Completion, Cross-section vs. Cohort illustration



We first present some observations from basic figures following cohort means over time. Figures 4-5 shows birth-years 1980 to 1997 grouped into 3-year cohorts: the first cohort comprises those born between 1980 and 1982, the second between 1983 and 1985, and onwards. It is only in this section that we aggregating into 3-year birth intervals to prevent cluttering the graphs while maximizing data use. We plot cohort averages at different points in time against the median age corresponding to the interval (e.g. the median age for those born between 1980 and 1982 is 22 in 2003). Graphically, this allows us to follow the average behavior of each cohort as it ages. The lines do not stack at all ages because in our survey years, earlier cohorts are observed at later ages while later cohorts are observed at earlier ages. All means are weighted by household sampling weights.

Several patterns emerge from these figures:

The proportion of those currently in school between ages 9 and 30 has remained fairly constant across cohorts. At first glance, this seems somewhat surprising given efforts since 2008 by the Ministry of Education to expand education access to marginalized students. However, current schooling may not be an accurate reflection of improvements in education attainment, since it masks high levels of late entry and grade repetition among children in the school system. For instance, grade repetition among first-graders was about 40 percent in the 1990s (MEYS 1999). This means that many students who may currently be in primary school are taking classes for the second or third time, and have a lower chance of completing it.

Comparing two people of the same age who are in school, the one born in more recent years is likely to be at a higher grade level. A more in-depth look at current grade and education attainment reveal that there are significant differences across cohorts. Among those attending school, later cohorts are more likely to be in a higher grade and this difference is quite large. The mean grade for 17 year olds born between 1986 and 1988 is 7.8. This increases steadily with later cohorts so that those 17 year olds born in 1995 to 1997 have an average current grade of 10.3 – an increase of 2.5 grades in about a decade. Note that according to the official grade that 17 year olds should be in the Cambodian education system requirements is Grade 12. Hence, results suggest youth belonging to recent are moving towards more age-appropriate schooling levels.

Recent cohorts also have greater education attainment than earlier ones (Figure 4). There are substantial improvements across cohorts for completing primary and secondary education. Primary completion is higher at almost all ages for recent versus older cohorts. Again for 17 year olds, primary

completion increased from just over half to three-quarters in less than 10 years, comparing the 1986-1988 cohort to the 1995-1997cohort. Primary completion increased by more than half from 16 percent to 37 percent among 13 year olds born in the earliest cohort compared to the latest. Lower secondary completion also increased across cohorts – from 14 percent in the earliest cohort to 39 percent in the latest for 17 year olds.

Few young people are attending the official age-appropriate grade level, but this is improving across cohorts. The graphs plotting current schooling level against age (Figure 4) show that there is a left-ward shift in age for the distribution of children currently attending schooling at primary, lower and upper secondary levels (Figure 4). A much higher proportion of 9-13 year olds were attending lower secondary school at the time of the survey in the 1995-97 cohort compared to the 1989-1991. The proportion of 14 year olds in lower secondary school (the official age for the last year of this level) increased from 23 to 36 percent. Among 17 year olds (the official age for the last year of upper secondary), the proportion of youth currently attending upper secondary rates increased from just 10 percent for the 1986-88 cohort to 27 percent in the 1995-97 cohort.

Current economic activity has not changed significantly across cohorts, but there is a substitution away from unpaid and agriculture work towards paid work. Economic activity starts in Cambodia at a young age, with more than 50% of 13 year olds engaging in work. While there is little change across cohorts for current work status there seem to be a substitution from unpaid work[5] to paid work. There is a consistent decline in agriculture work at all ages, with lower rates of agriculture work among those born in more recent years.

Males and females differ in terms of schooling completion and employment. A fewer proportion of women across cohorts are currently working past the age of 19 compared to men, likely due to marriage (which starts to increase rapidly at that age) and childbearing duties. There are no clear differences between genders for type of employment (Figure 6). While schooling outcomes appear to be improving across cohorts for both men and women, there are some differences to note.
Women seem less likely to be currently in school than men, with the divergence beginning around age 15. However, it is unclear what is driving this difference. There are two separate forces that determine the means at each age. There is an age effect or age profile, the shape of which may differ for females compared to males for all cohorts. There is also a cohort effect, where a 15-year old female born in more

[5] Unpaid work here is classified as unpaid family work or own-account work

recent years is as likely to be in school as a 15-year old male in the same cohort, but the difference in schooling is greater for 15-year olds born in earlier years. There is also a time effect, which corresponds to shocks that in occur in the survey year the outcome is being measured. Since these graphs show raw means, we cannot distinguish between these three effects. The next section will attempt to separate out the age and cohort effects, while controlling for time effects, and estimate whether females are indeed "catching up" to their male counterparts in terms of schooling, or whether these graphs simply reveal differences in age profiles that persist across cohorts.

In the next section we will attempt to simplify these graphs in a way that allow us to more clearly and convincingly visualize cohort effects and the age profiles for schooling and work outcomes. As described using the example of current schooling between males and females, we can see that there are three factors that simultaneously affect the cohort mean: the age effect, cohort effect and time effect. To recap, the *age effect* refers to variations along the life-cycle for a cohort; i.e. differences in schooling or work outcomes that follow a particular shape as a person ages, regardless of which cohort they belong to. In the case of being in school for example, it is high at younger ages and starts to drop off around age 14, declining rapidly and smoothing out at 0 for later ages. The *cohort effect* refers to differences in structural conditions that different generations face, which can account for some of the cross-cohort differences we note above. Hence, while it is tempting to interpret the patterns presented so far as a pure cohort effects, it is important to bear in mind that there is a simultaneous *time effect* which corresponds to shocks that occur in a particular year. For example, the cohort of 1996 reaches age 12 in 2008; an economic crisis in that survey year would alter schooling or work outcomes irrespective of age or cohort, i.e. it affects all cohorts. This will be explained in greater detail in the following section.

Figure 4. Schooling Outcomes (Current), by Age and Cohort

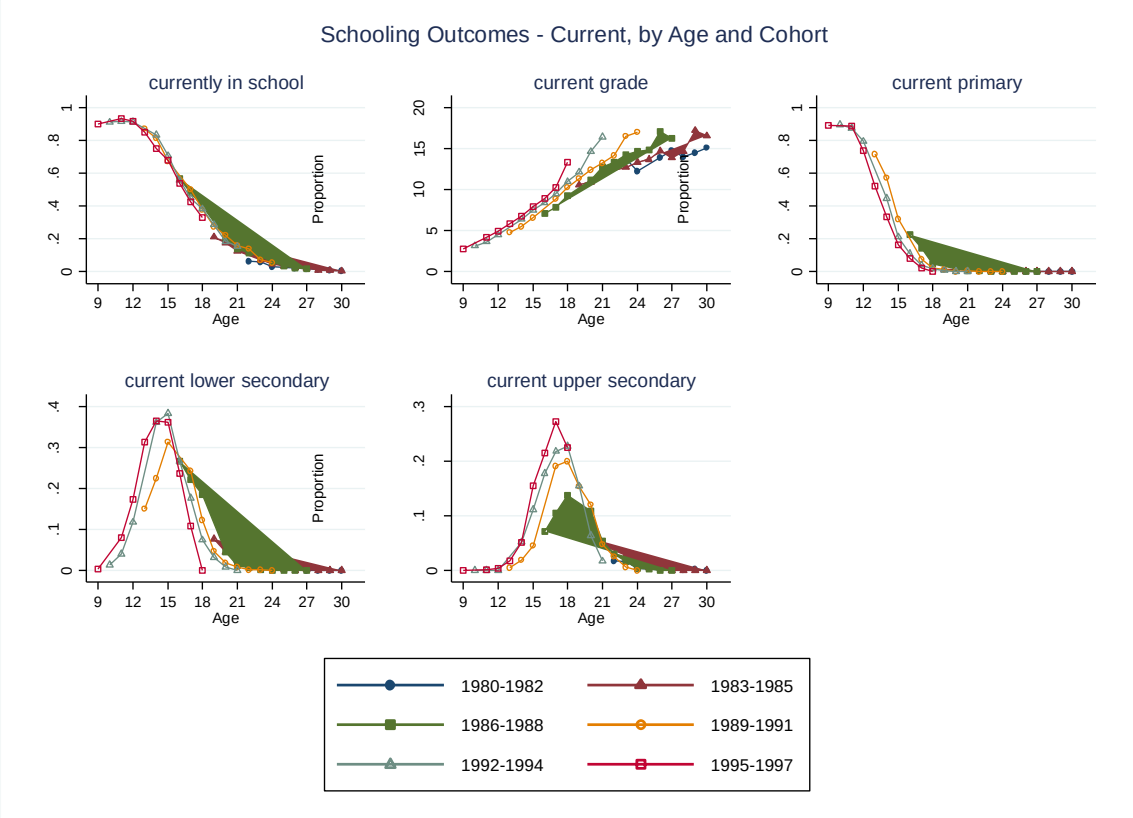


Figure 5. Schooling Outcomes (Highest), by Age and Cohort

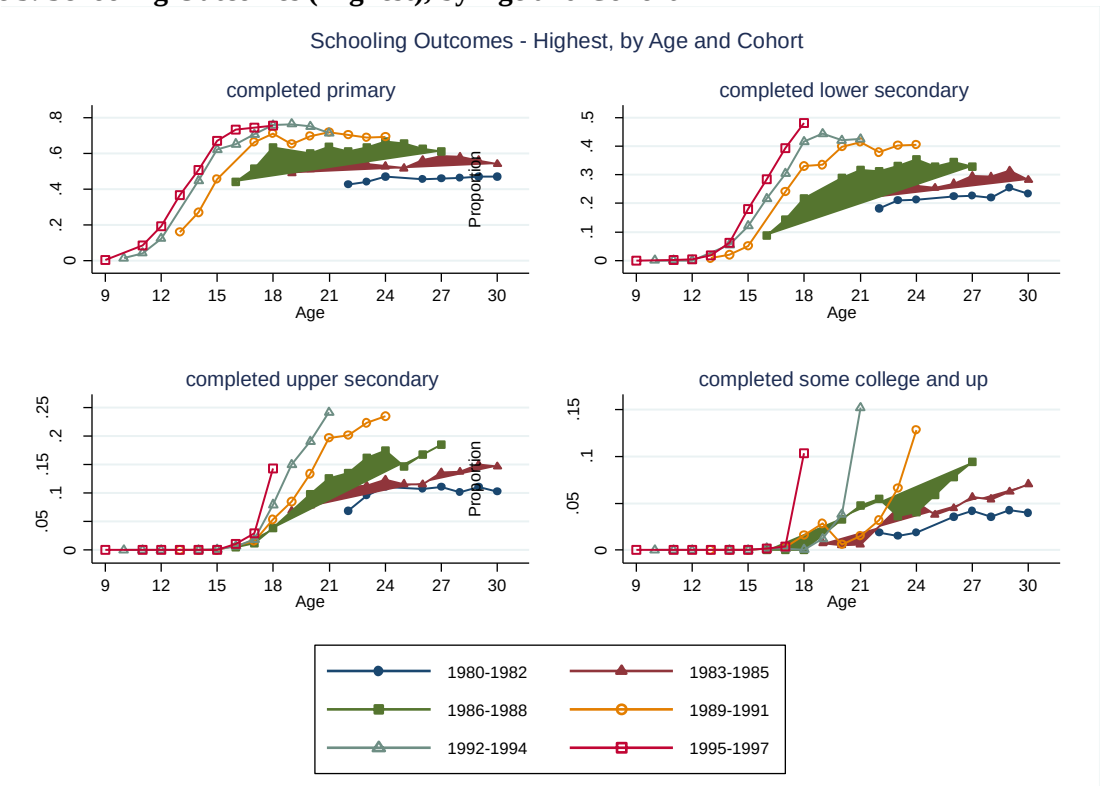


Figure 6. Work and Social Outcomes, by Age and Cohort

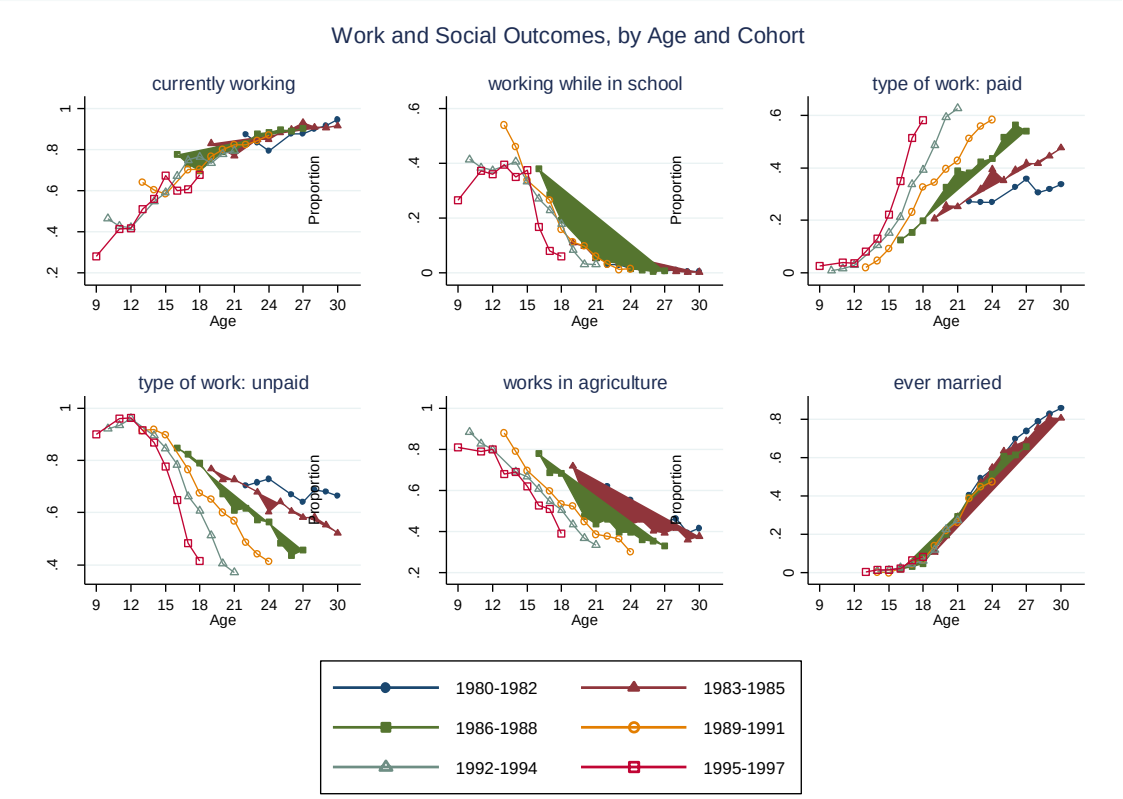
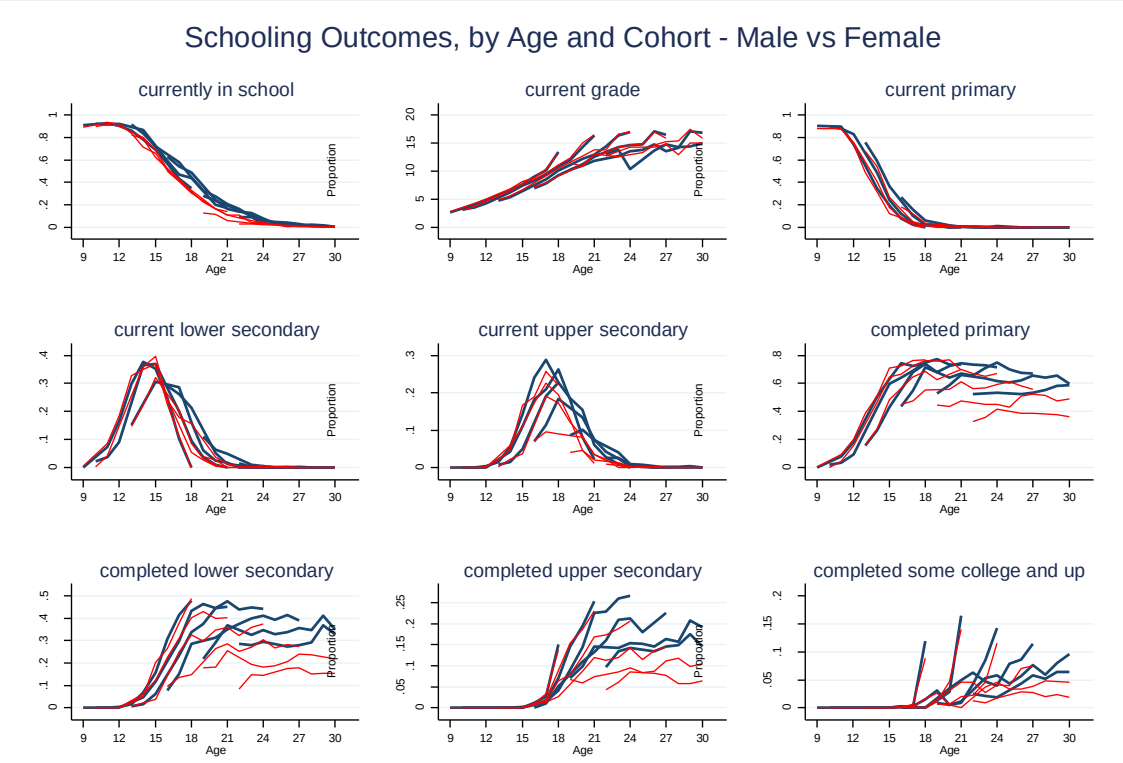
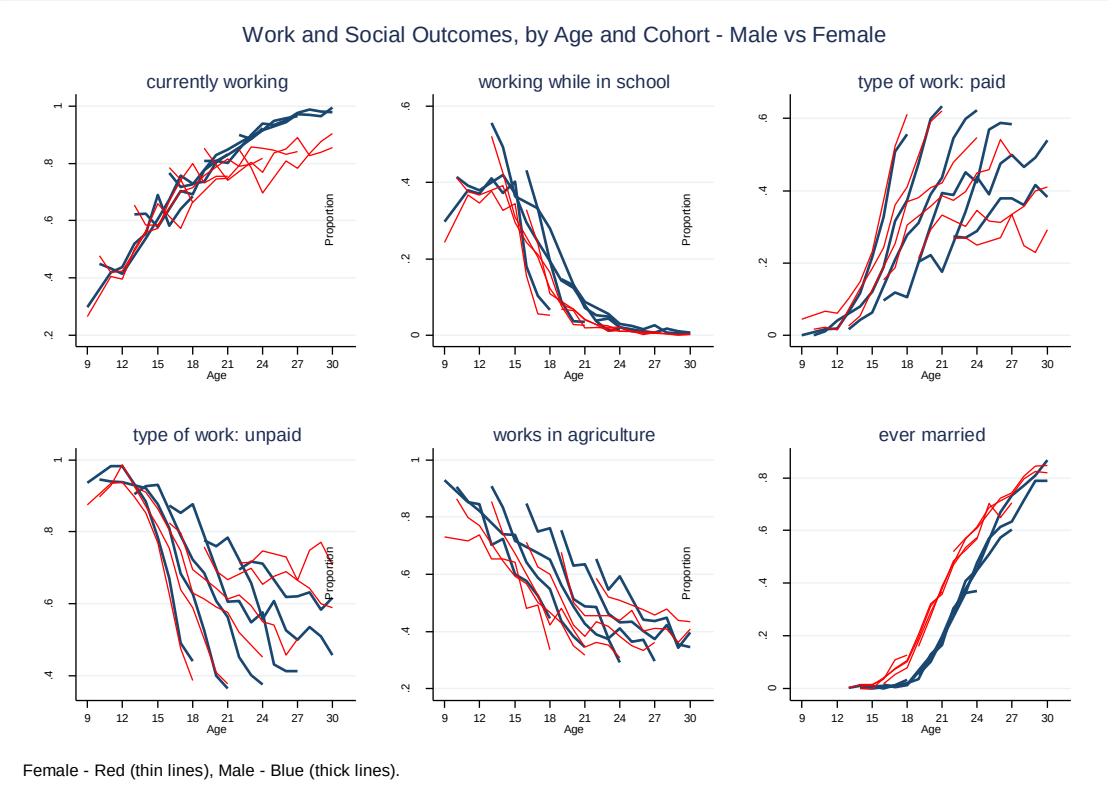


Figure 7. Schooling Outcomes by Age and Cohort- Male vs Female



Female - Red (thin lines), Male - Blue (thick lines).

Figure 8. Work and Social Outcomes by Age and Cohort – Male vs Female



II. School and Work Patterns - Decomposing Cohort/Age effects

So far we have seen that each cohort has their own trajectory over the life cycle, and that there are clear differences across cohorts for some outcomes but not others. In this section, we take this analysis a step further and break down differences in cohort means to an age component and a cohort component, while controlling for time effects. This allows us to remove some of the sampling noise in Figures 4-8 and disentangle age and cohort effects to answer the following:

- For any given cohort, what is the age profile for school and work outcomes? Do these differ by gender and region?

- For any given age, have schooling and work outcomes improved or declined with more recent versus older cohorts? Is the rate of improvement or decline different between males and females, urban and rural youth?

Method

The results in this section will be presented in a series of graphs. Before describing the results, it is important to first discuss the methods and assumptions used in producing them.

Several authors have estimated age (life cycle), cohort (birth year) and time (business cycle) effects, and we adopt similar methods here. The known challenge in estimating all three effects simultaneously is the fact that they are linearly dependent: time = cohort + age. As noted in Attanasio 1998: "the differences between two individuals observed at the same age could be due to time or cohort effects; and the difference between two individuals observed at the same time could be due to age or cohort effects." Hence, estimating all three effects simultaneously requires imposing additional structure to the model.

We follow both Deaton and Paxson (1993) and Attansio (1998)'s approaches to adding structure by normalizing the time variable. This normalization restricts time parameters to measure cyclical variability alone, as deviations from the mean. There can be positive and negative shocks in different years but these "business cycle" effects average to zero (Borsch Supan 2003). All trends in the means can therefore be interpreted as an interaction of age and cohort effects, as time effects are identical across cohorts. This basic identification assumption allows us to estimate a typical age profile and to assess the movements of this profile across the range of cohorts considered.

We estimate the following equation:

$$Y_{ct} = \alpha + C\gamma + \sum_{i=1}^{5} \beta_i A^i + K\varphi + \varepsilon \qquad (1)$$

Y_{ct} is the average outcome for cohort c at time t, $C\gamma$ is a vector of c-1 cohort dummies and age is specified as a 5th order polynomial. We drop the earliest cohort (1983) to be the reference category. $K\varphi$ is a series of t-1 year dummies whose coefficients are constrained to sum up to zero and be orthogonal to linear trends, as discussed above (i.e. this is the equivalent of the time normalization). In all regressions we weight each observation by the cell size, such that cells with more individuals are weighted appropriately.

We run separate regressions for males and females, urban and rural. It is important to recognize that, unlike gender, urban status is not strictly a time-invariant characteristic. Those with certain traits in urban areas may have moved from rural areas. As such, the trends we find for "urban" dwellers may actually reflect the behavior of some other time-invariant trait that determines the choice to pursue work or schooling in urban areas, rather than the regional effect alone. In addition to our main analysis using all years and the full sample, we also do a robustness check using the sub-sample of those who have not moved since birth in the years where this information is available (CSES years 2003-05, 2007, 2009, 2010 and 2011) In 2011 for example, 66 percent of rural dwellers and 36 percent of urban dwellers have

not moved locations since birth. For the most part, the robustness check corroborates findings from the full analysis. We include the results in the appendix and discuss any differences in the main text. It would also be interesting to also look at the intersection of urban status and gender; however, splitting the sample four ways would result in cell sizes that are infeasibly small given our data.

By estimating equation 1, we can plot the predicted values of our indicators for age, cohort and time against age, cohort and survey year (respectively), to separately identify these effects.

Results

Age effects

Figures 9 and 10 are visual representations of the age effect coefficients, which show how outcomes evolve over a life cycle, for a given cohort. In other words, they show the shape of the age relationship across time and cohort. We plot the smoothed age profile by gender, and by urban status. Since the figure shown is for the arbitrary cohort of 1984, the focus for interpretation should be on the relative shape and position of the profile at different ages, and not necessarily the absolute position of each (y-intercept).

The age profiles for schooling for males and females follow roughly the same shape but with level differences. Females are much less likely to be in school at almost every age, when compared to males. For both groups, current attendance in lower secondary school peaks around age 15 and upper secondary peaks around age 18. The proportion of females in school tend to track their male counterparts at early ages, but then start to diverge slightly starting age 12 for lower secondary and 15 for upper secondary.

The economic activity of young women follows an inverse-"u": which peaks around age 20 before declining, whereas it continues to increase for young men. Before age 20 women are also working more than men. This may be because younger women work instead of attending school while older women take care of their families instead of work. Gender differences also emerge for paid labor and marriage. Fewer women after age 20 engage in paid work and more likely to engage in unpaid work and more women after the age of 16 are married than men.

There are large differences in the shape of the age profiles between those living in rural versus urban areas. In terms of schooling, rural dwellers are less likely to be in formal schooling at almost all ages and those who are in school have a lower current grade. More urban youth are in lower and upper secondary between ages 12-20 compared to rural youth. A much greater proportion of rural dwellers below the age of 20 who are in school also work at the same time, compared to urban dwellers.

The most striking difference across rural and urban youth is the difference in age profile for current work. A substantially greater proportion of rural dwellers less than age 25 engage in some form of labor within the last week, compared to urban dwellers. Of those who are working, a lower proportion of rural dwellers get paid and a higher proportion work in agriculture; for these two outcomes there is a level difference at all ages. For urban youth, the prevalence of unpaid work decreases faster the older they get compared to rural dwellers.

Figure 11. Age Effects of School and Work Outcomes – Males versus Females

Note: This is the age profile for an arbitrary cohort (1984). Age is modeled as a 5th order polynomial.

Figure 12. Age Effects of School and Work Outcomes – Urban versus Rural

Note: This is the age profile for an arbitrary cohort (1984). Age is modeled as a 5th order polynomial.

Cohort effects

While the age profile allows us to see how outcomes change over the course of a life *averaged over* cohorts, it does not say anything about whether outcomes are improving or worsening *across* cohorts. One way to visualize cohort effects is to stack multiple age profiles for the range of cohorts in our sample. Here, we plot 8 age-profile curves in total for cohorts between 1980 and 1998 in two-year intervals (Figures 11 to 14). For each cohort we show the predicted values only on ages for which data is available for that cohort, reducing out of sample predictions. Overlapping curves suggest little between-cohort variation; highly dispersed curves suggest greater between-cohort variation. If the curves are highly dispersed for one group and less dispersed for the comparison group, this suggests the cohort effects may be greater for the more dispersed group.

To present a clearer picture of how some outcomes are changing across cohorts, we also show stylized graphs for meaningful outcomes, plotting the predicted values for only the earliest cohort (1984) and the latest cohort (1998). These are in Figures 152 – 22. Solid lines are for predictions on

ages where survey data is available for that cohort and dashed lines are for predictions where it is not (out of sample predictions). Age is held constant at some meaningful value indicated on the graph, and we compare the predicted difference in age effect between subgroups belonging to the latest cohort (marked with circles) and the earliest cohort (marked with triangles).

The results show that while level differences in schooling prevail for males and females, females are "catching up" to their male counterparts in terms of current schooling and primary completion. As seen in Figures 8 and 9, for young women there is more pronounced dispersion across cohorts for being currently in the school system and for primary completion, compared to young men. These differences become more apparent in Figures 15 - 18. The gap between males and females currently in school has reduced significantly comparing those born in 1984 to those born in 1988 (Figure 15), and there has been a corresponding reduction in female work at younger ages (Figure 18). Females are also catching up to males in terms of lower secondary schooling (Figure 17).

A similar "catching up" seems to also be occurring among rural versus urban youth for primary but not secondary schooling. There is greater between-cohort variation for rural dwellers than urban dwellers in terms of current grade and primary school completion (Figures 19 and 20). This suggests that while improvements in age-appropriate schooling and primary completion are occurring for both regions, they are occurring at a slightly faster rate among rural youth, who started off in a more disadvantaged position. Figures 196 and 20 show that the gap in primary schooling completion and current grade between urban and rural youth have closed by more than half in 15 years. However, this "catching up" trend does not seem to extend to lower and upper secondary schooling, where cohort effects actually seem to be slightly greater in urban areas (Figure 21).

In terms of work outcomes, there is a pronounced cohort-level decline in current work status in rural areas compared to an ambiguous effect in urban areas, though this does not hold when we only consider those who remained in their region of birth (Figure 14and Appendix Figure X). We do not find differences between rural and urban cohorts for current work in the robustness check using only those who have not moved since birth in the subsection of years. This suggests that for this variable, the results in the main analysis are likely driven by selection into migrating from urban to rural areas. Among those who are working however, there is faster increase in paid work and a corresponding decline in unpaid work among rural compared to urban youth and this is corroborated by the robustness check. Figure 22 also shows that the gap in paid work between rural and urban youth has fallen between the 1984 and 1998 cohorts.

A more direct way to visualize how outcomes are changing across cohorts is to plot the cohort effects against year of birth. These graphs abstract away from age-specific profiles and instead show the overall cohort experience, controlling for age and time effects. We visualize these cohort effects by plotting the predicted values for the cohort indicators relative to a base cohort – here we use 1983, the earliest cohort for our age range. Increasing or decreasing trends mean that, holding age and time effects constant, later cohorts are doing better or worse compared to the 1983 cohort. Again, we disaggregate by gender and region.

Figures 23 and 24 show the "pure" cohort effect graphs. It is interesting to first note that relative to the earliest cohort of 1983, all outcomes have more or less linear trends through subsequent years. Over time there is a clear increase in current grade, primary completion and lower secondary completion. Upper secondary completion does not appear to have changed much across cohorts. Current work, working while in school and unpaid or agricultural work are all on the decline. Since our sample comprises only youth between 10 to 30, the decline in current work may not necessarily be a sign of greater unemployment but rather a shift from working at young ages to greater schooling (as reflected in positive cohort effects on schooling).

These graphs corroborate the insights from the age effect and stylized graphs. Rural dwellers experienced relatively greater cross-cohort gains in schooling for average current grade and completion of primary schooling, though there are no differences for higher levels of education. For work outcomes, the main analysis suggests thatthe decline in current work, unpaid and agriculture work is more pronounced in rural areas, while paid work is increasing. Again, the robustness check corroborates the findings for paid and unpaid work, but does not find a difference across regions for current work and agriculture work. For males versus females, there are fewer differences to note. The trends seem to track closely for most outcomes, except current schooling, primary completion, and current work, which show a stronger trend for females again corroborating the previous figures.

Figure 11. Age effects of school outcomes across cohorts – males versus females

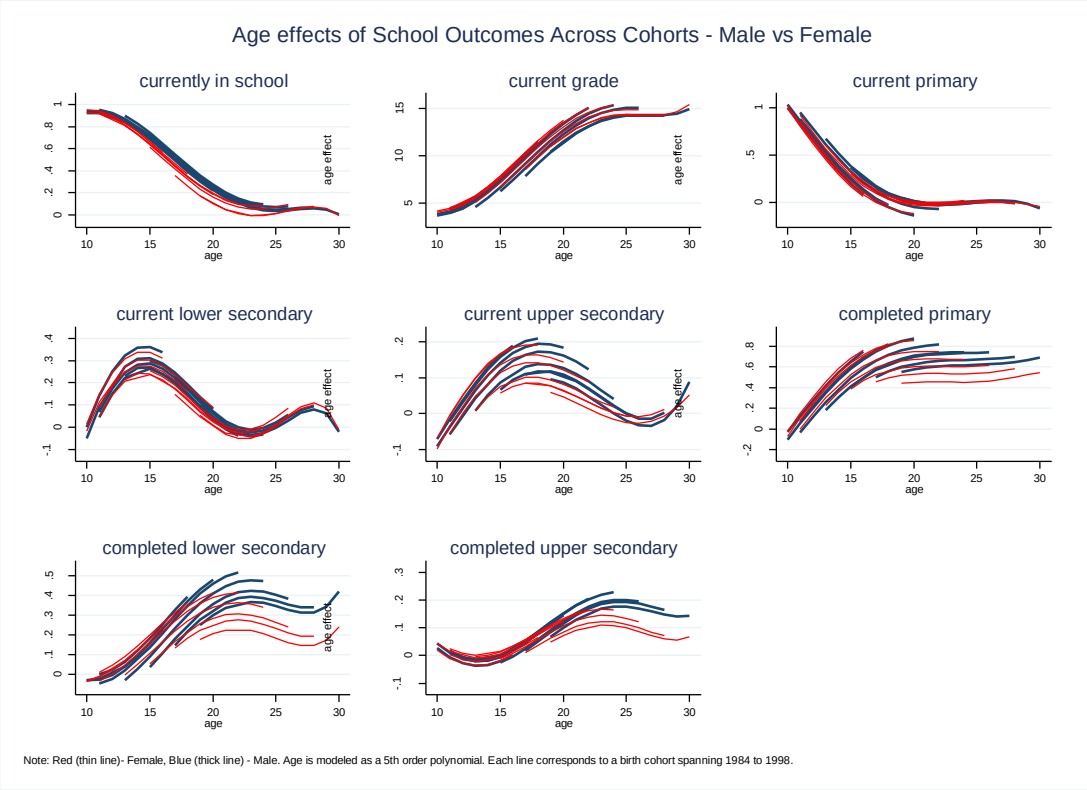


Figure 12. Age effects of work and social outcomes, across cohorts – males versus females

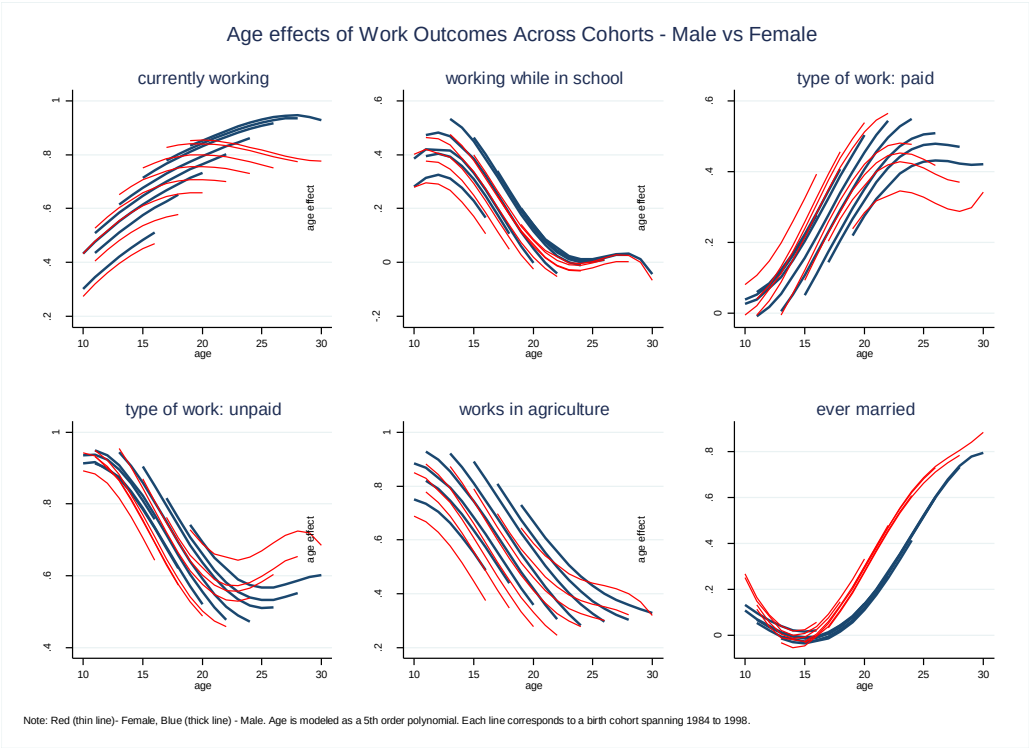
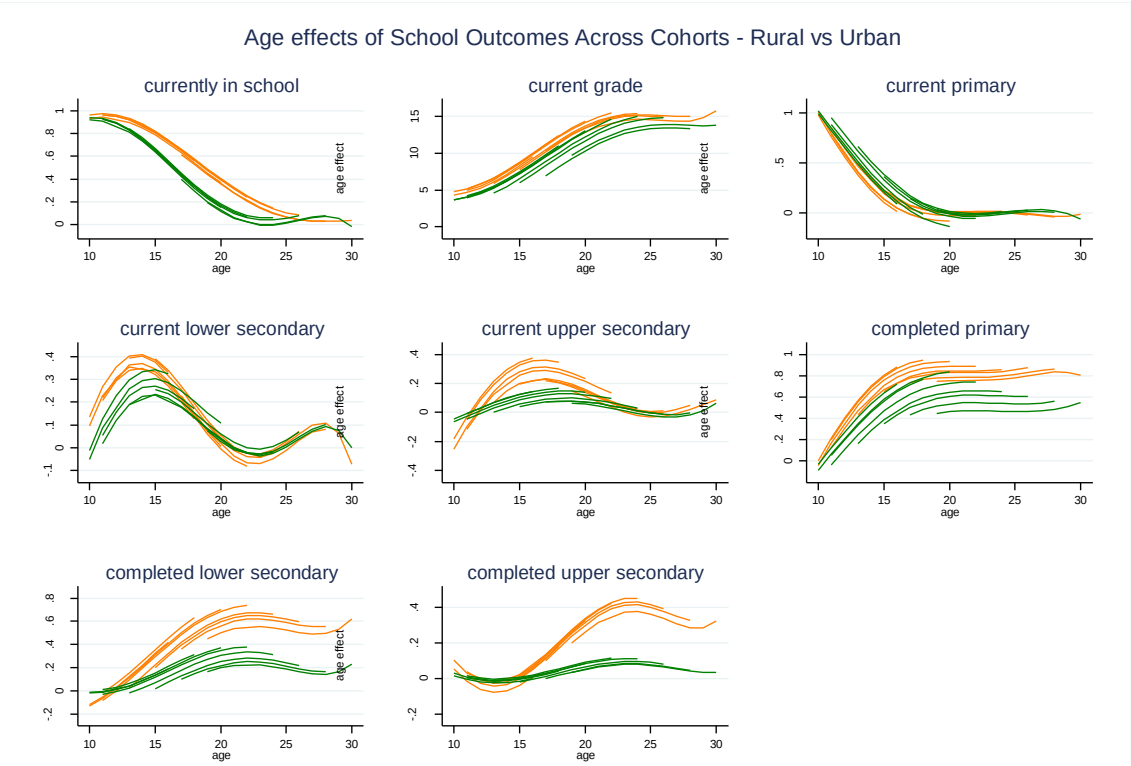
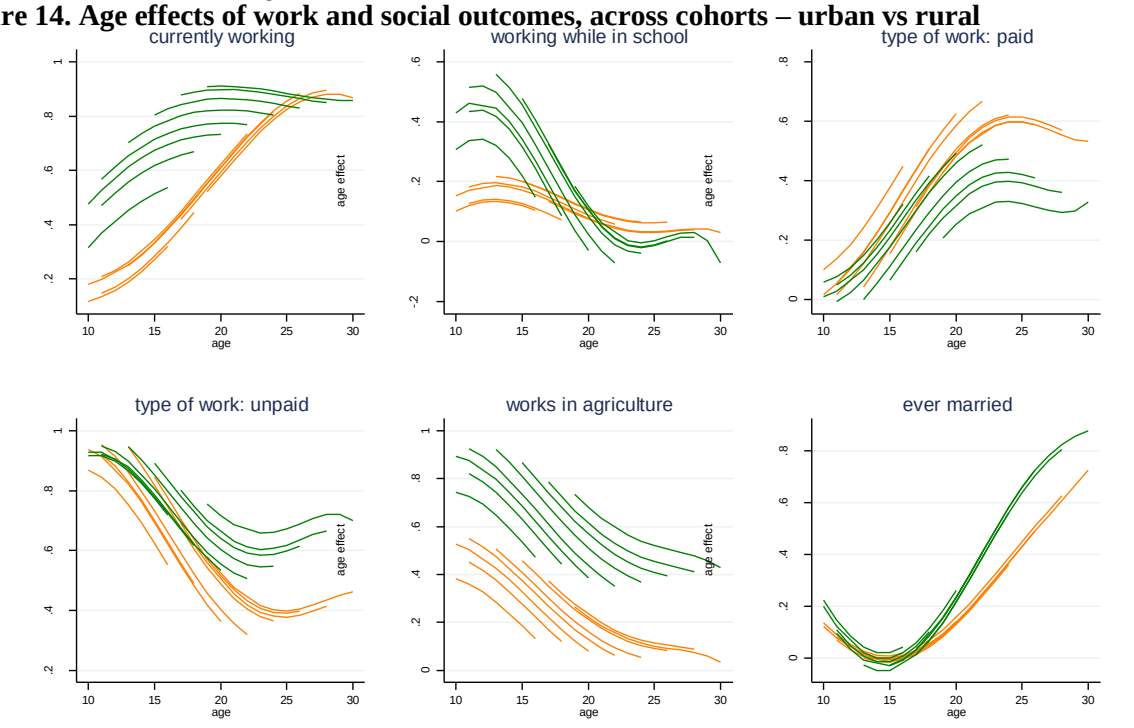


Figure 13. Age effects of school outcomes, across cohorts – rural versus urban



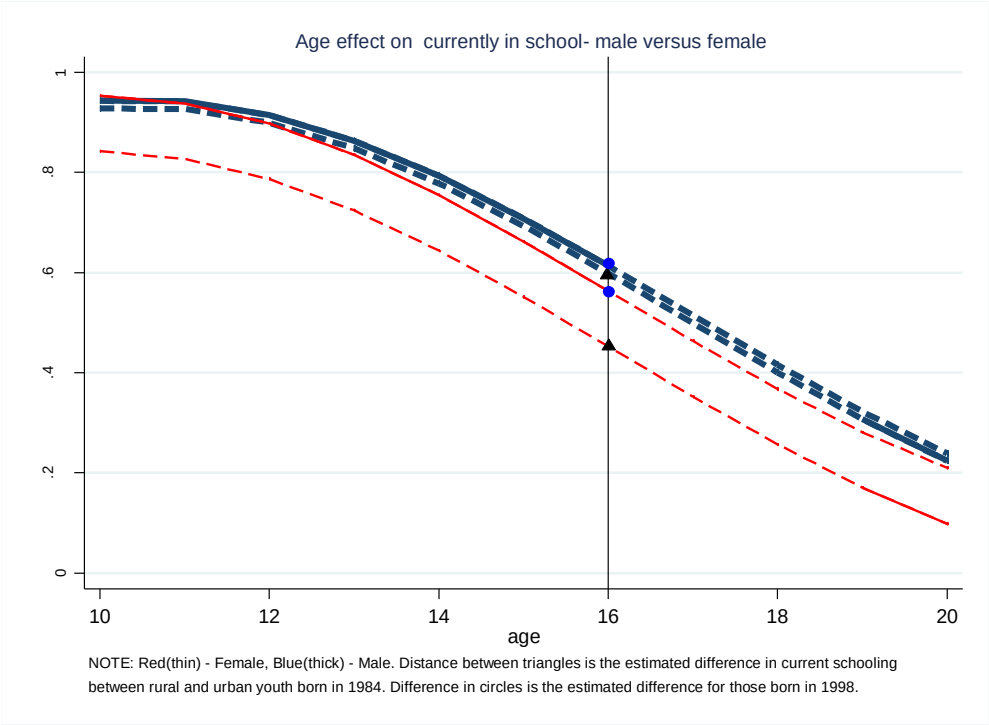
Note: Green - Rural, Orange - Urban. Age is modeled as a 5th order polynomial. Each line corresponds to a birth cohort spanning 1984 to 1998.

Figure 14. Age effects of work and social outcomes, across cohorts – urban vs rural



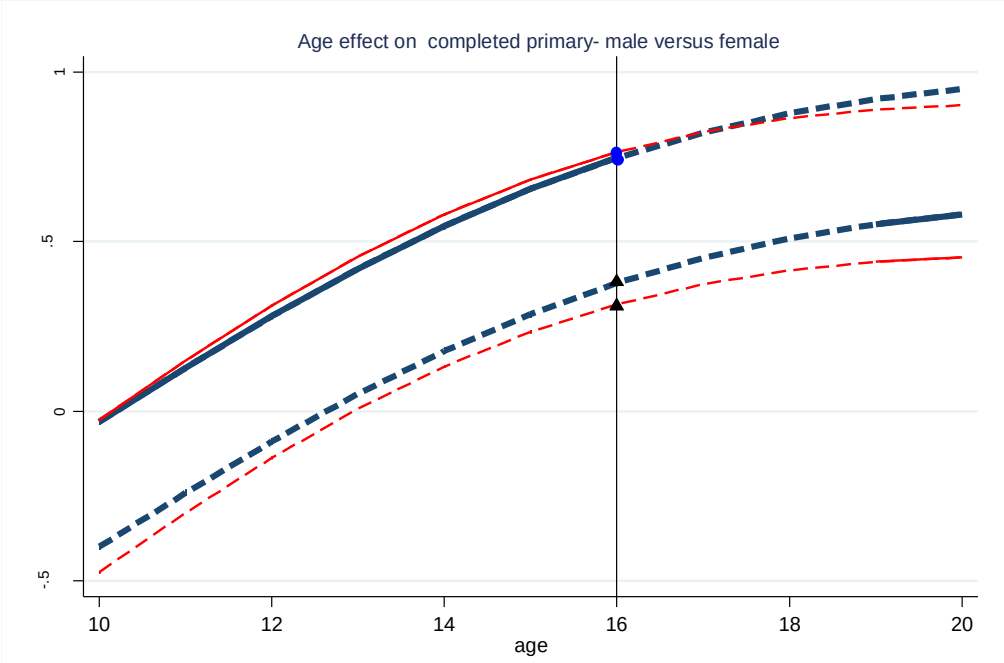
Note: Green - Rural, Orange - Urban. Age is modeled as a 5th order polynomial. Each line corresponds to a birth cohort spanning 1984 to 1998.

Figure 15. Difference in current schooling between males and females born in 1984 versus 1998



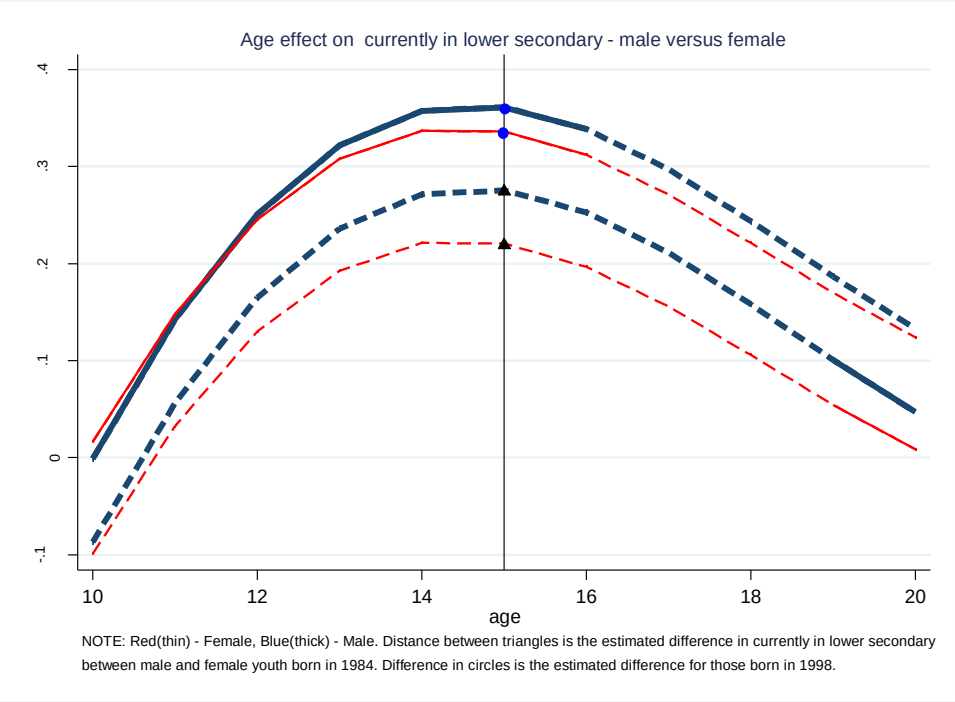
NOTE: Red(thin) - Female, Blue(thick) - Male. Distance between triangles is the estimated difference in current schooling between rural and urban youth born in 1984. Difference in circles is the estimated difference for those born in 1998.

Figure 16. Difference in primary completion between males and females born in 1984 versus 1998



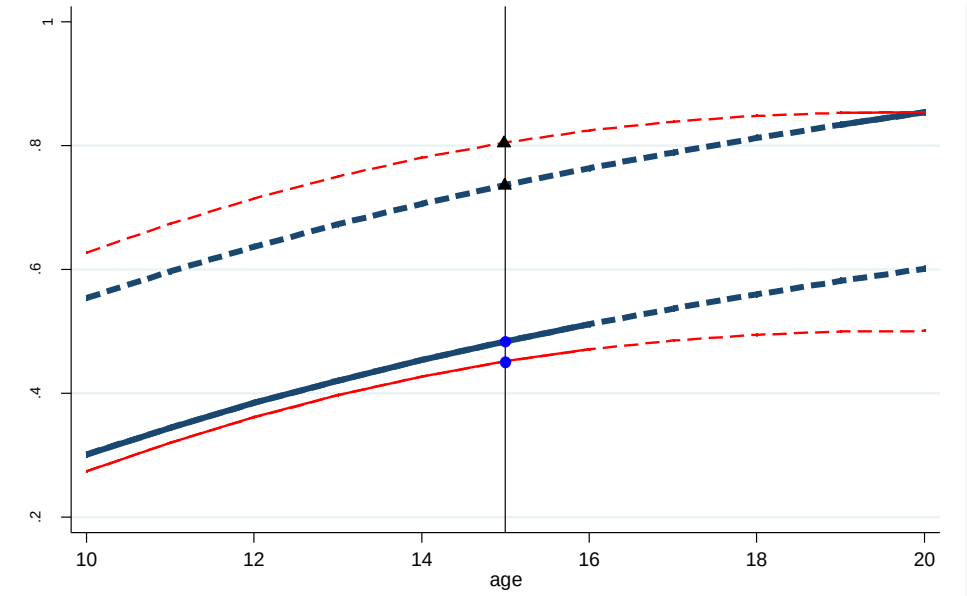
NOTE: Red(thin) - Female, Blue(thick) - Male. Distance between triangles is the estimated difference in primary schooling completion between male and female youth born in 1984. Difference in circles is the estimated difference for those born in 1998.

Figure 17. Difference in current lower secondary schooling between males and females born in 1984 versus 1998



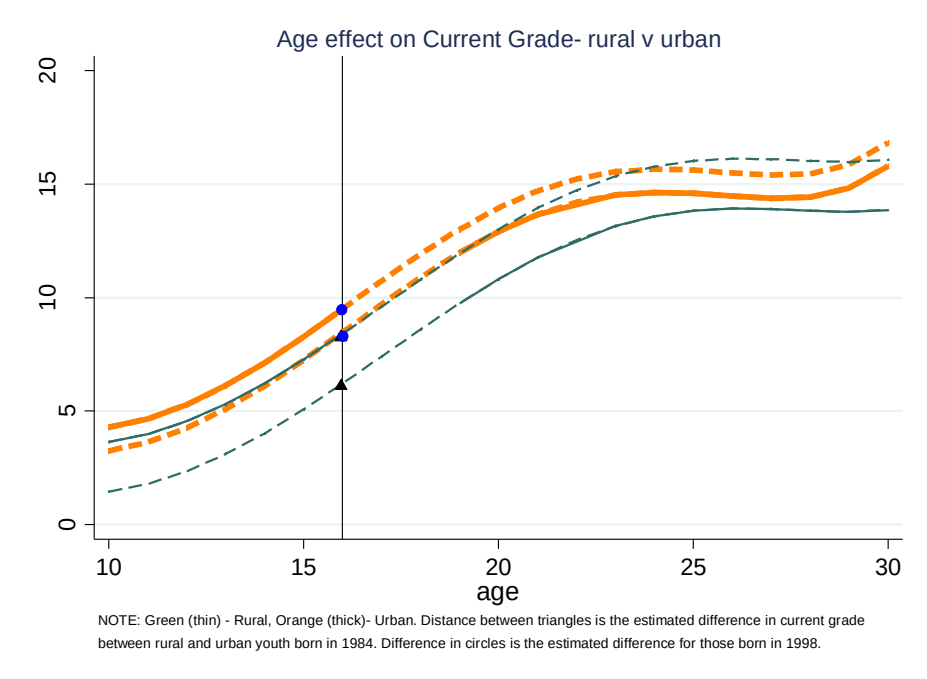
NOTE: Red(thin) - Female, Blue(thick) - Male. Distance between triangles is the estimated difference in currently in lower secondary between male and female youth born in 1984. Difference in circles is the estimated difference for those born in 1998.

Figure 18. Difference in current work between males and females born in 1984 versus 1998



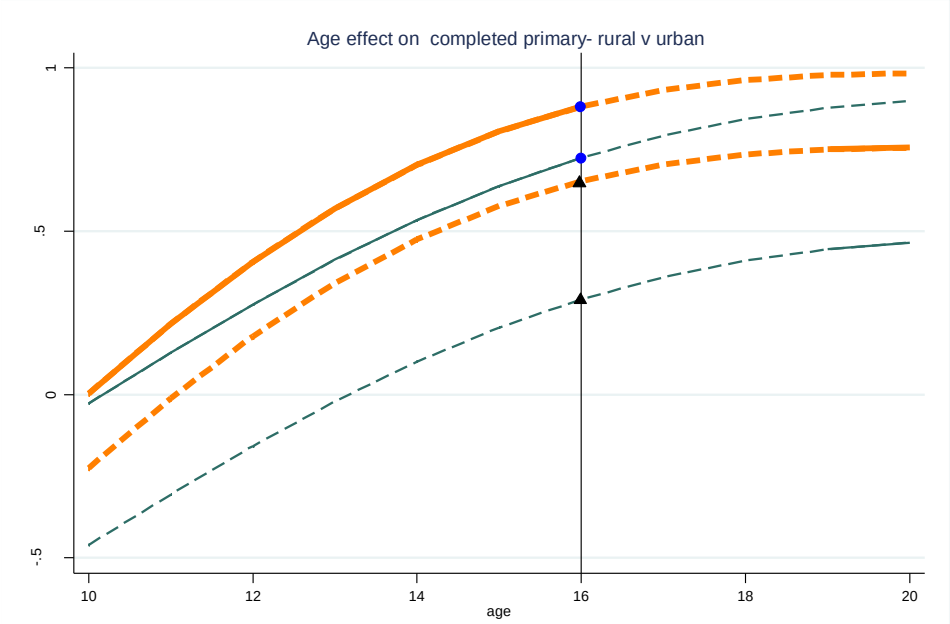
NOTE: Red(thin) - Female, Blue(thick) - Male. Distance between triangles is the estimated difference in currently working between male and female youth born in 1984. Difference in circles is the estimated difference for those born in 1998.

Figure 19. Difference in current grade between rural and urban youth born in 1984 versus 1998



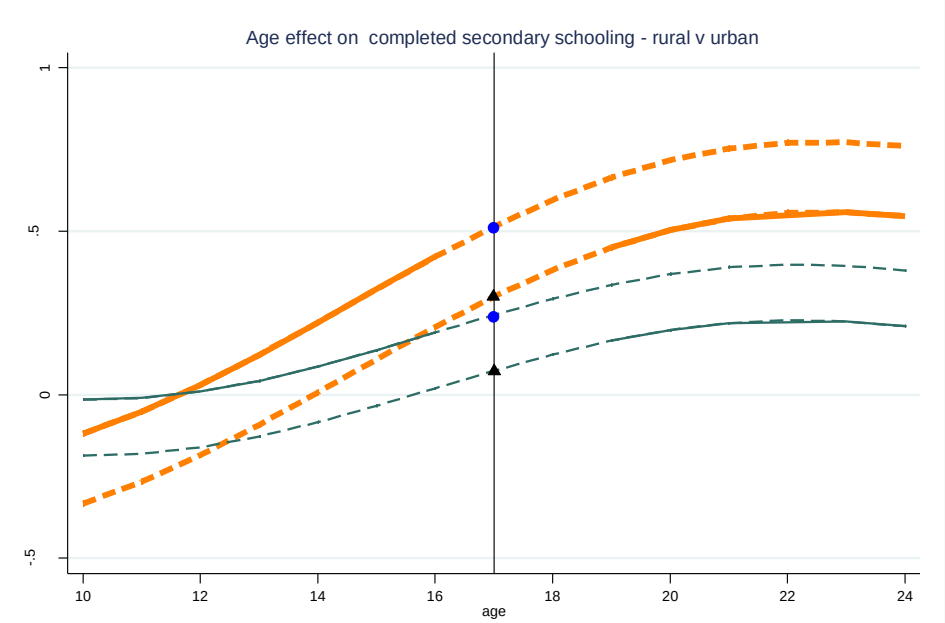
NOTE: Green (thin) - Rural, Orange (thick)- Urban. Distance between triangles is the estimated difference in current grade between rural and urban youth born in 1984. Difference in circles is the estimated difference for those born in 1998.

Figure 207. Difference in primary school completion between rural and urban youth born in 1984 versus 1998



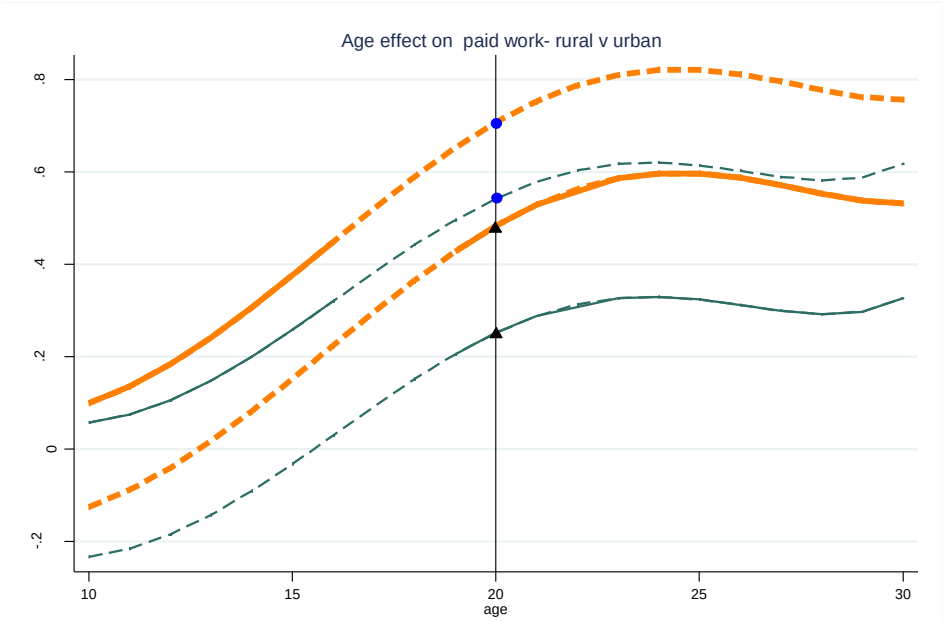
NOTE: Green(thin) - Rural, Orange(thick) - Urban. Distance between triangles is the estimated difference in primary school completion between rural and urban youth born in 1984. Difference in circles is the estimated difference for those born in 1998.

Figure 21. Difference in secondary school completion between rural and urban youth born in 1984 versus 1998



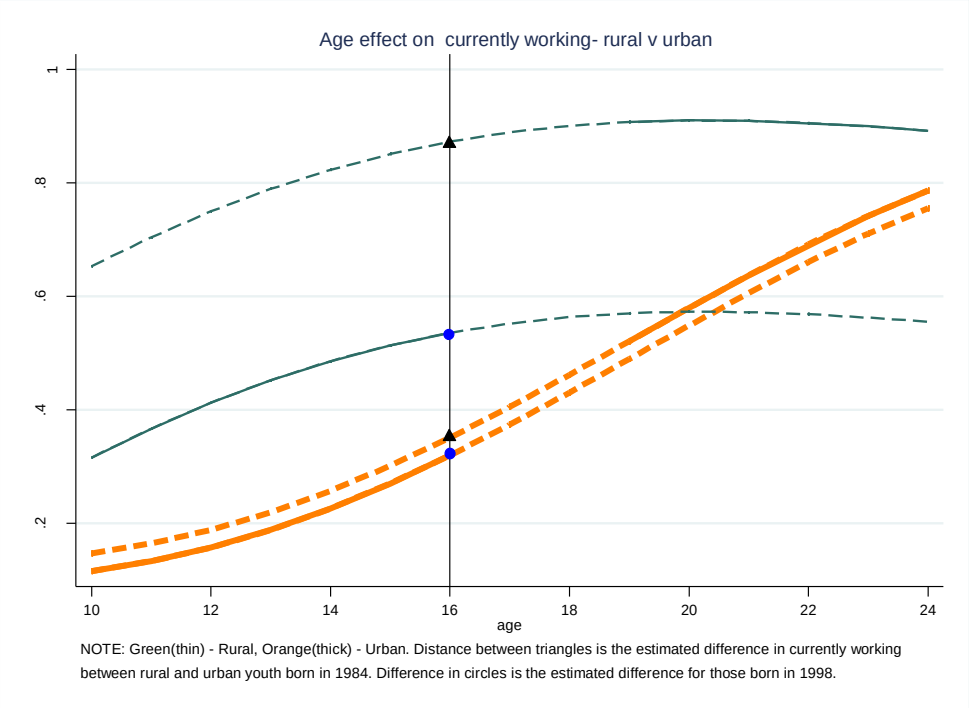
NOTE: Green(thin) - Rural, Orange(thick) - Urban. Distance between triangles is the estimated difference in completed secondary schooling between rural and urban youth born in 1984. Difference in circles is the estimated difference for those born in 1998.

Figure 22. Difference in paid work between rural and urban youth born in 1984 versus 1998



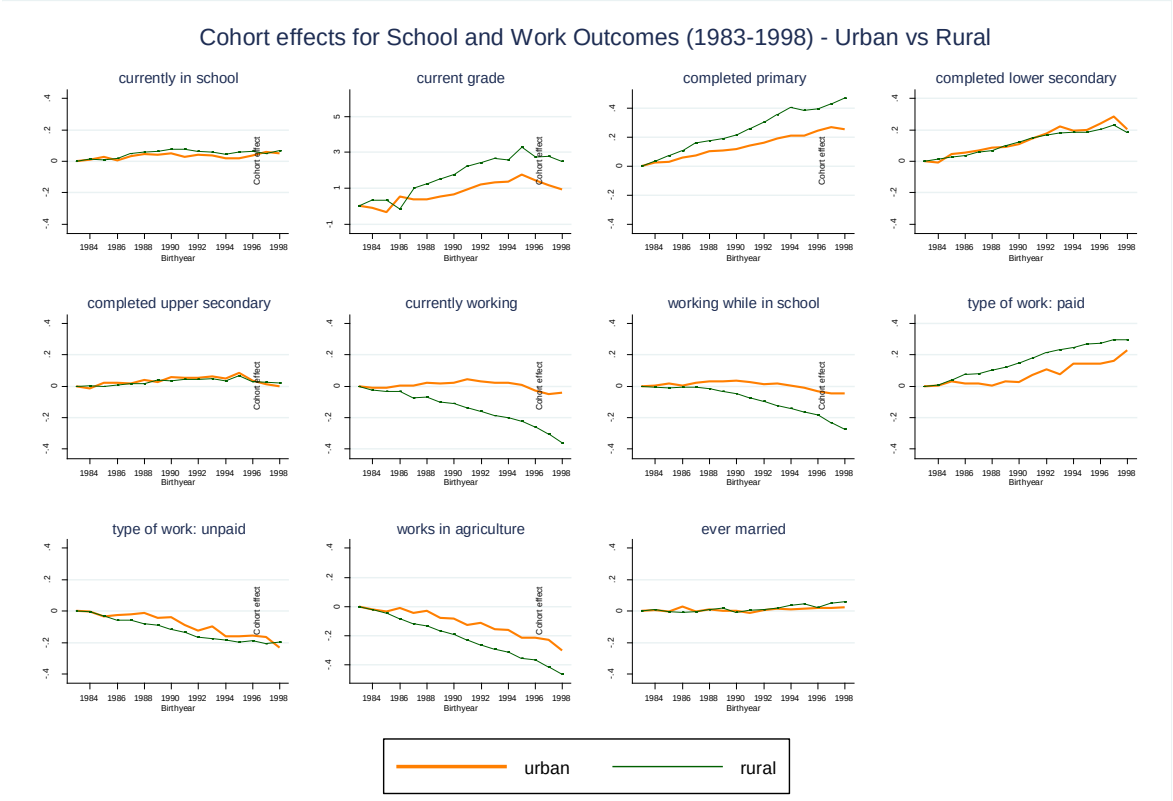
NOTE: Green(thin) - Rural, Orange(thick) - Urban. Distance between triangles is the estimated difference in doing paid work at age 20 between rural and urban youth born in 1984. Difference in circles is the estimated difference for those born in 1998.

Figure 239. Difference in current work between rural and urban youth born in 1984 versus 1998



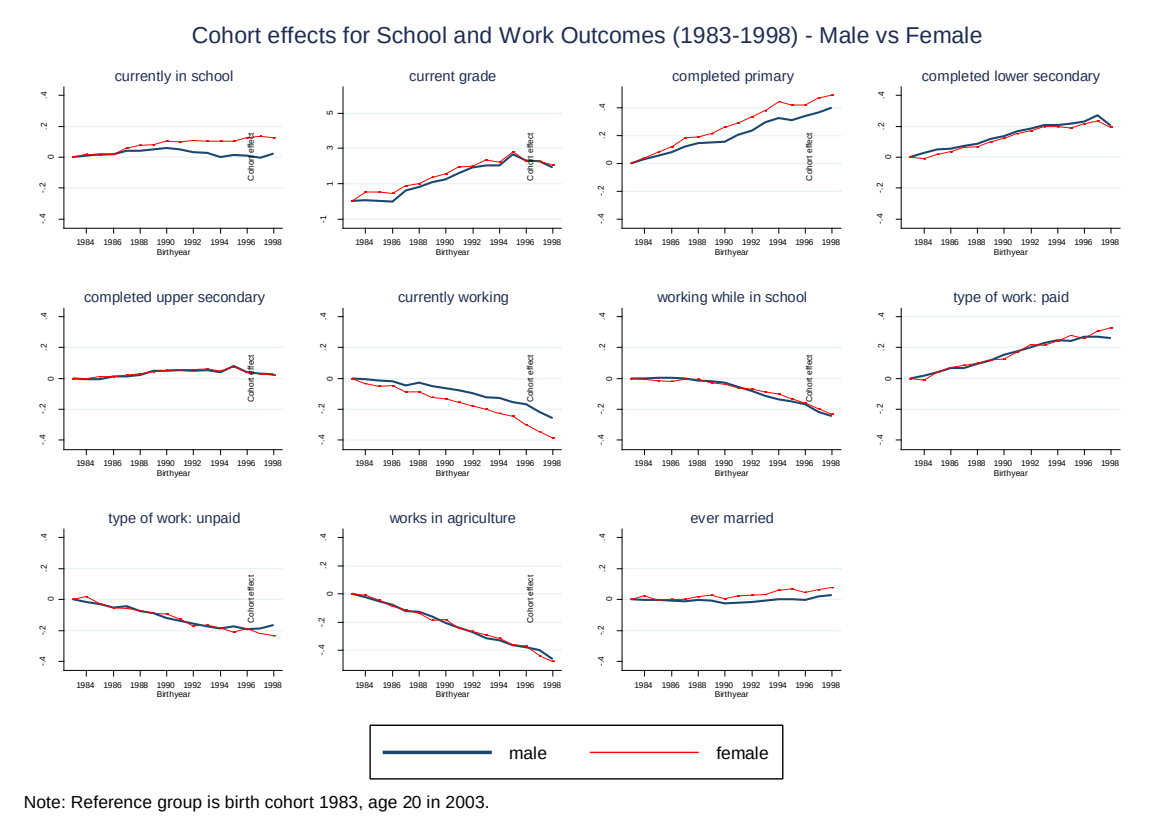
NOTE: Green(thin) - Rural, Orange(thick) - Urban. Distance between triangles is the estimated difference in currently working between rural and urban youth born in 1984. Difference in circles is the estimated difference for those born in 1998.

Figure 24. Cohort trends in school and work outcomes of youth born in 1983-1988, comparing urban vs rural



Note: Reference group is birth cohort 1983, age 20 in 2003.

Figure 25. Cohort trends in school and work outcomes of youth born in 1983 to 1988, comparing males and females



Note: Reference group is birth cohort 1983, age 20 in 2003.

III. Linking Cohort Youth Experience to Later Outcomes: Regression Analysis

In the previous two sections, we have shown that the level of education attainment and type of employment Cambodian youth experience may differ substantially depending on which year they were born in. Notably, youth born in more recent years have higher levels of primary and secondary completion, are more likely to engage in paid work and are less likely to do agriculture and unpaid work. The overall cohort effects follow the same trends for males versus females and for urban versus rural youth, though we found differences in some outcomes across these subgroups.

In this section, we move beyond looking at these variables in isolation and instead consider the dynamic relationship between them. We use the cross-cohort differences found in Part 2 as a source of variation in a regression analysis framework, to understand how youth labor market or schooling experiences impact later work experiences. Specifically, we want know:

- What is the effect of secondary school completion (lower and upper) on labor outcomes at older ages?
- Has the effect of schooling on labor outcomes changed over time?
- Does doing informal work (unpaid or agriculture) at younger ages predict doing similar work at older ages?
- How persistent is informal work? Does the effect of doing informal work during youth fade over time (comparing immediate versus long-run outcomes)?

Empirical Strategy

We continue using the cohort-panel data set-up from Parts I and II, aggregating the cross-sectional data to the cohort-by-time level and creating a "panel" of 11 cohorts, following them as they age.

The "youth cohort experience" is our independent variable of interest, which we define to be the average experience of a cohort at a particular age.[6] Hence, it is a time-constant variable. For early-cohort schooling experience, we use lower and upper secondary schooling completion at 17 and 20: these are three years after the official age of being at the last grade at each level (age 14 for grade 9 and age 17 for grade 12, respectively). For early-cohort labor experience, we define it at age 16, which is based on the average age of leaving school in Cambodia (ILO 2013).

In order to fully utilize all cohort-by-time observations, we make a small adjustment accounting for the lack of survey data in 2006. Because the 1989 cohort does not have an observation at age 17 (lacking survey year 2006), we use age 18 (from survey year 2007). Similarly because the 1986 cohort does not have an observation at age 20, we use age 21. In so doing we are assuming that the schooling experiences of 17 and 18 year olds are comparable, as with 20 and 21-year olds. We then have 11 cohorts for our analysis: 1986 to 1996 when youth experience is defined at 16, and we observe outcomes from ages 18 to 27; 1985 to 1995 when youth experience is defined at 17, and we observe outcomes from ages 19 to 28; 1982 to 1992 when youth experience is defined at 20, and we observe outcomes from ages 22 to 31.

The basic equation we would like to estimate is the following, taking the effect of upper secondary completion on unpaid work as an example:

[6] While the choice of age is somewhat arbitrary, it is necessary to pick one age, rather than a range of ages. If a range of ages is selected (for example, the average of unpaid work in a cohort between ages 12-17) the age bracket may not be available for all survey year and birth year combinations. Later birth years will have younger ages. Hence, if unpaid family work is correlated with age (the younger people are the more likely to engage with these types of works), earlier birthyears will have a higher mean due to data availability alone.

$$Unpaid\ Work_{c,a_{22-31}} = Completed\ Secondary_{c_{20}}\delta + \sum_{i=1}^{5}\beta_i A^i + n_c + \varepsilon_{ca} \quad (1)$$

Where $Unpaid\ Work_{c,a_{22-31}}$ is the outcome of interest of cohort c at age 22 to 31; $Completed\ Secondary_{c_{20}}$ is the time-invariant cohort experience during "youth" at age 20; age is specified as a 5th order polynomial; and $n_c + \varepsilon_{ca}$ represents the cohort unobserved heterogeneity and idiosyncratic error, which we assume is normally distributed with zero mean and is uncorrelated with age and secondary completion.

Estimating this equation presents two challenges. The first is the difficulty in accounting for cohort fixed effects (n_c). Because we are tracking cohorts as they age over time, accounting for cohort fixed effects is important in allowing us to control for characteristics that are time-constant within each cohort. However, because our variable of interest is also observed at the cohort-level, it is perfectly correlated with cohort fixed effects, leaving no additional variation with which to estimate n_c. The second challenge is the linear dependency of age, time and cohort effects which we described in Section II. These two problems mean we cannot include all three effects in our regression equation. Instead, we must make some adjustments and assumptions to account for them in a different way.

To resolve the linear dependency issue, we account for time effects following Burgess (2003) and Cruces et. al. (2013), by normalizing the outcome variables relative to the aggregate mean at that time period:

$$\tilde{y}(a,c) = \frac{Y(a,c,t)}{Y(t)} \quad (2)$$

Y(t) is the aggregate outcome at time *t*. $\tilde{y}(a,c)$ then measures the deviation for cohort *c* at age *a*, from the average outcome of all cohorts and ages at time *t*. Similar to the normalization we implemented in Part II, this assumes that time-effects impact proportionally on all age/cohort cells. Note that normalizing requires taking the average of the outcome for the same age range at each survey year. For work variables, our aggregate outcome is measured for those between ages 16 and 65. For schooling variables, it is based on those between ages 6 and 30. We replace the outcome in equation 1 with this new normalized outcome:

$$\widetilde{Unpaid\ Work}_{c,a_{22-31}} = Completed\ Secondary_{c_{20}}\delta + \sum_{i=1}^{5}\beta_i A^i + n_c + \varepsilon_{ca} \quad (3)$$

The coefficient δ is then interpreted: "A one unit increase in *secondary completion at age 20* leads to a δ unit change in normalized score, meaning cohorts that have more schooling do better/worse relative to other cohorts in the same time period for that outcome."

The above step adjusts for any contemporaneous factors that may affect the outcome. To account for cohort effects, we do the following. First, we estimate equation (3) by assuming that unobserved cohort heterogeneity is uncorrelated with youth cohort experience. That is, $(Z_c, n_c + \varepsilon_{ct}) = 0$. This is a strong assumption. If this does not hold, then we expect our estimates to be upwardly biased, since we expect that the correlation between the cohort effect and outcomes (unpaid work for instance) to have the same sign as the correlation between cohort effects and youth cohort experience (here, completing secondary school by age 19). This is the "pooled" model.

We estimate a second equation in a "two-step method", which accounts for cohort effects indirectly, following the procedure described in Burgess (2003) and Cruces et. al. (2013). The first step is a regression of the outcome on a 5th-order parametrization of age and cohort fixed effects, but excluding the independent variable of interest:

$$Unpaid\ \widetilde{Work}_{c,a_{22-31}} = \sum_{i=1}^{5} \beta_i A^i + \sum_{c=1982}^{1992} \delta_c\ Cohort_c + \varepsilon_{ca} \qquad (4)$$

where $\widehat{\delta_c}$ captures the variation in outcome that is accounted for by time constant cohort factors. In the second step, we regress the estimated values of $\widehat{\delta_c}$ on the cohort youth experience:

$$\widehat{\delta_c} = \gamma Completed\ Secondary_{c_{20}} + v_c \qquad (5)$$

γ now isolates any component of differences between cohorts that is correlated with completion of secondary school at cohort's youth. In other words, it gets at the part of the relationship between cohort effects and outcome that is explained by youth cohort experiences.

To summarize, we present results following the regression model from equation (3) as our main results, and (5) as a robustness check. In addition to estimating these equations for the full range of cohorts and ages, we also examine possible heterogeneity. To assess the question of whether early work experiences has long-run persistence or simply affects the beginning of the adult working life, we run separate regressions for "immediate" versus long-run effects, comparing outcomes at ages 18 to 21, versus 22 to

27[7]. We also split the cohorts into pre and post-1989 to compare whether the association between schooling and work outcomes has changed or remained constant.

We weight all regressions by the square root of the number of observations in each cohort-by-age cell. We present separate results for males vs females and urban versus rural.

Results

Tables 1 and 2 shows the results from the pooled regression for schooling outcomes, for the full sample of cohorts. Tables 3-6 show the results from the regressions that separate cohorts into pre- and post-1989 (for lower secondary schooling) and pre- and post- 1986 (upper secondary schooling). We show only the coefficients for the cohort youth experience variables.

More schooling during youth translates to less informal work and more paid work, with the effect stronger for upper secondary education. Higher rates of completion of both levels of schooling during youth are associated with doing less agriculture, less unpaid work, and more paid work in later years. The coefficients on these regressions can be interpreted as such:

> "**Comparing two people from different cohorts, controlling for age and labor market conditions of that year – the person from the cohort with higher education is doing better/worse than the person in the cohort with lower education, relative to the overall performance in that year."**

We find that a 10 percentage point increase in lower secondary schooling by age 17 leads to a 6.5 percentage point decrease agriculture labor for females and 6.1 for males, relative to the average agriculture work in each year and controlling for age. Similarly, normalized unpaid work decreases by 6.8 and 7.6 for males and females respectively. These are all significant at 1 percent. For upper secondary completion, the effect is even stronger – more than double that of lower secondary completion, across the significant outcomes. Note that since we are measuring outcomes at older ages for upper secondary compared to lower secondary due to the data structure, we cannot make a direct comparison of the effect sizes since it may be confounded with an upwards age trend. Hence, we do the same analysis but restrict the follow-up ages to a common range (22 to 28), allowing us to compare the later work outcomes of

[7] The split is 4:6 to ensure that the sample size is similar, since there are fewer observations at later ages.

those who complete upper secondary school by age 20 to those who completed lower secondary school by age 17 for same period in their life. We find very similar results; the magnitude of the upper secondary effect is more than double that of lower secondary (Appendix table 3): for agriculture work for example, the effect size is -1.483 for upper secondary and for lower secondary it is -0.676.

Higher upper secondary completion is associated with a larger decline in informal work for males compared to females.. The coefficients on upper secondary education are greater for males than females for almost every outcome. A 10 percentage point increase in upper secondary schooling corresponds to a 21.7 percentage point decline in agriculture work and a 20.1 percent decline in unpaid work for males, relative to the average yearly performance. This drops to 14.4 and 15.8, respectively, for females. Both lower and upper secondary schooling have a positive and significant effect on paid work for males, but an insignificant effect for females. Completing upper secondary schooling also has a negative effect on being married for both genders, with a stronger association for males than females.

The effect of secondary completion on informal and paid work is more pronounced for later birth cohorts compared with earlier ones. When we divide our sample to five years pre- and post- 1989 (or1986), holding constant age, higher rates of lower secondary schooling completion is not significantly associated with work outcomes for pre-1989(86) cohorts, but does have a significant effect for those who were born in 1990(87) onwards. Females born in both periods that have completed lower secondary school do less agriculture work, and while the magnitude is higher for the post-1989 cohort the result is not significant at conventional levels (p-value borders .10). There remains no significant effect of lower secondary schooling on paid work for females in both birth periods, although a positive sign emerges in the later period. The same general pattern is found for upper secondary schooling. Since earlier cohorts contain observations at older ages, we conduct a robustness check where we cap the observations in the pre-period to the maximum age in the post-period, in order to have a common pool of comparison between the two periods. The results match those found in the full sample (Appendix table 4 and 5).

Turning to the questions on persistence of youth informality, Table 4 shows the results from the pooled regression of youth work experience on later work outcomes, for the full sample. Tables 5 and 6 show results comparing long-run persistence of youth work experience to near-term effects.

For both males and females, there is a persistent overall effect of doing informal work at age 16 on similar work outcomes later in life. For female cohorts, a 10 percentage point increase in agriculture work at age 16 is associated with an 8.5 percentage point increase in normalized agriculture work in later

years and an 8.9 percentage point increase in normalized unpaid work, holding constant age. Unpaid work has a larger effect on later agriculture and unpaid work– 11.0 and 13.2 respectively. For females, there is no effect of informality on paid work. For males, the persistence of informality is comparable to females, but males are further penalized with doing significantly less paid work after doing agriculture or unpaid work at youth.

For both genders, we find that doing agriculture work during youth has a strong persistent effect through young adulthood. Youth engaged with agriculture work at age 16 are almost equally likely to continue doing agriculture work at ages 18-21 as ages 22-27. The effect on doing unpaid family or own-account work decreases, but is still positive in the long-term. For males, agriculture work is associated with less paid work in the near term, but this association disappears in the long-term. For females, doing agriculture work at younger ages is associated with doing more paid work at ages 22-27.

Unpaid work has a persistent effect on work outcomes in the short run but this fades in the long-run. Unlike doing agriculture work, doing unpaid family work or own-account work at youth does not seem to have a lasting effect. A 10 percentage point increase in unpaid work at 16 leads to a 16.8 percentage point increase in normalized agriculture work at ages 18-21. At 22-27, though the coefficient's sign is still negative the magnitude is less than a third and the effect is no longer significant. Similarly, the effect of youth unpaid work on later paid and unpaid work becomes much smaller in magnitude and insignificant at later ages.

Table 7 shows the results from the two-step estimation technique described in the methods section, which accounts for cohort fixed effects. The coefficients presented correspond to the youth experience variable from the second stage (equation 5). Both the pooled and two-step estimation techniques yield the same conclusions. The magnitudes of the coefficients are slightly larger than the pooled estimation, though the signs and significance are the same.

IV. Conclusion

Cambodia went through major socio-political and economic changes within a few decades. In this paper, we use a dynamic cohort approach to generate new insights into how various cohorts of youth experienced schooling and employment differently. Using 11 years of cross-sectional surveys we construct a cohort panel following the life cycle trajectories of 19 cohorts as they aged from youth into

early adulthood. This approach allows us to control for contemporaneous time effects and trace out the smoothed cohort and age trends in schooling and work outcomes. Using the variation in experiences across cohorts, we are additionally able provide evidence on how early youth experiences translate to outcomes later in life.

Our analysis reveals several key insights:

More recent cohorts have higher levels of primary and secondary completion than earlier cohorts. Grade progression rates and education attainment increased with each younger cohort, suggesting progress towards greater quality and inclusivity. Comparing two people of the same age who are in school, the one born in more recent years is likely to be at a higher grade level and have completed more schooling. This is a promising outcome reflective of MoEYS' expansions in school construction, cash and meal scholarship programs, and the increase in "complete" primary schools which offer all six grades (from 56.8 percent in 2001-02 to 86.2 percent in 2013-2013, UNESCO report).

We also find significant increases in paid work and declines in unpaid and agriculture work at all ages, across cohorts. While Cambodian youth start working at a young age, economic activity among school-aged children is declining with more recent cohorts, likely as they attain more schooling. Among working youth, there is a substitution away from unpaid work towards paid work, as well as a consistent decline in agriculture work. Findings from our regression analysis suggest that cohort youth school and work experiences have a significant on their adult labor market outcomes. More schooling during youth translates to less informal work for both young men and women, though schooling only has a positive return on paid work for men. Agriculture work tends to be persistent: high proportions of agriculture work during one's youth leads to more of the same down the line. Unpaid work, on the other hand tends to be confined to youth and early labor market experiences; the effect decreases or disappears with time for both men and women.

Subgroup differences seem to be diminishing over time along some outcomes: the difference between male and females' school enrollment and primary school completion is notably smaller among the most recent cohort (1998) compared to the earlier one (1984). Among rural and urban school-goers, the gap in current grade closed by more than half in a span of 15 years. Gains in schooling correspond to a faster decline in working at young ages in rural areas relative to urban areas. Among those who are working, there is also a faster increase in paid work among rural cohorts, compared to urban cohorts.

In addition to these positive developments, our results also point to remaining challenges in the education and employment landscape in Cambodia. Despite average cohort-wide gains, gender and regional disparities remain: women and rural youth complete less schooling than men and urban youth, respectively. Rural youth are substantially more likely to be working when they should be in school, and women are more likely to be working at younger ages compared to men, though this reverses around age 20, likely as duties of childbearing and housework take over.

Despite improvements in primary schooling and grade progression among females and rural youth, such "catch-up" effects do not extend to secondary schooling. Rural cohorts are improving faster than their urban counterparts for primary completion, but pronounced differences remain for lower and upper secondary completion and they are not diminishing over time. Between males and females, the gap in lower secondary completion closed by more than half from the oldest to youngest cohort, yet the catch-up does not extend to upper secondary schooling. These results suggest that more targeted policies are needed to address discrepancies among disadvantaged subgroups.

Cambodia has undergone a significant transformation to its education system in the past decade. The empirical analysis presented here demonstrates that education attainment and access have increased and these changes are associated with some positive developments in employment and social inclusion, though there remains room for improvement. While we our paper explored novel ways of using cross-sectional data longitudinally, we are limited by the lack of genuine panel data and certain variables to answer other pertinent research questions. For instance, a more detailed dataset would allow examining the changing nature of informal work- are there meaningful compositional changes within agriculture and unpaid work that the catch-all outcome is masking, perhaps a movement towards higher-value agriculture production? Likewise, since we are only able to disaggregate analysis based on time-invariant variables in our dataset, another question remains of whether differences in intertemporal gains are also found in other socio-economic subgroup measures, such as parental education and household wealth.

Table 1. Effect of completing secondary schooling on later work and social outcomes - Males vs Females

	works in agriculture		type of work: unpaid		type of work: paid		ever married	
	Female	Male	Female	Male	Female	Male	Female	Male
(1) Lower secondary by age 17	-0.658***	-0.613***	-0.684***	-0.764***	-0.294	0.744**	0.007	-0.068
	(0.130)	(0.186)	(0.118)	(0.214)	(0.199)	(0.343)	(0.064)	(0.086)
(2) Upper secondary by age 20	-1.439***	-2.168***	-1.580***	-2.010***	-0.279	1.303***	-0.381***	-0.688***
	(0.230)	(0.336)	(0.170)	(0.336)	(0.670)	(0.436)	(0.122)	(0.248)
Observations	62	62	62	62	62	62	62	62
R-squared (1)	0.397	0.587	0.461	0.446	0.548	0.452	0.969	0.969
R-squared (2)	0.340	0.291	0.657	0.305	0.660	0.201	0.898	0.919

Robust standard errors in parentheses

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

Table 2. Effects of completing lower secondary schooling by age 17, pre and post 1989 cohort – male vs female

	works in agriculture		type of work: unpaid		type of work: paid	
VARIABLES	Pre-1989	Post-1989	Pre-1989	Post-1989	Pre-1989	Post-1989
Male	-0.146	-1.270***	-0.135	-1.437***	-0.001	1.669***
	(0.186)	(0.282)	(0.121)	(0.146)	(0.219)	(0.267)
Female	-0.361**	-0.616	-0.196**	-1.182***	-0.101	0.179
	(0.160)	(0.391)	(0.073)	(0.247)	(0.355)	(0.315)
Observations	41	27	41	27	41	27
R-squared (male)	0.729	0.846	0.728	0.870	0.726	0.791
R-squared (female)	0.342	0.619	0.204	0.664	0.604	0.373

Robust standard errors in parentheses

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

Table 3. Effects of completing upper secondary schooling by age 20, pre and post 1986 cohort– male vs female

	works in agriculture		type of work: unpaid		type of work: paid	
VARIABLES	Pre-1986	Post-1986	Pre-1986	Post-1986	Pre-1986	Post-1986
Male	-0.604	-1.682**	-0.657	-1.277*	1.012	0.729
	(1.033)	(0.631)	(0.642)	(0.647)	(1.361)	(0.772)
Female	-0.406	-0.801*	-0.708*	-1.697***	-0.443	0.060
	(0.422)	(0.432)	(0.418)	(0.219)	(1.273)	(0.600)
Observations	41	21	41	21	41	21
R-squared (male)	0.263	0.551	0.196	0.290	0.162	0.138
R-squared (female)	0.324	0.139	0.276	0.723	0.647	0.511

Robust standard errors in parentheses

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

Table 4. Effect of Agriculture and Unpaid at Age 16 on Later Work Outcomes - Male vs Female

	works in agriculture		type of work: unpaid		type of work: paid	
	Female	Male	Female	Male	Female	Male
(1) Agriculture work at age 16	0.848***	0.879***	0.891***	1.024***	0.074	-1.358***
	(0.149)	(0.173)	(0.132)	(0.139)	(0.313)	(0.290)
(2) Unpaid work at age 16	1.096***	1.236***	1.323***	1.446***	-0.007	-2.026***
	(0.208)	(0.253)	(0.188)	(0.229)	(0.368)	(0.353)
Observations	62	62	62	62	62	62
R-squared (1)	0.538	0.705	0.451	0.607	0.381	0.583
R-squared (2)	0.536	0.701	0.557	0.601	0.380	0.602

Robust standard errors in parentheses

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

Table 5. Medium and long-run effects of agriculture work at 16 –male vs female

	works in agriculture		type of work: unpaid		type of work: paid	
	age 18-21	age 22-27	age 18-21	age 22-27	age 18-21	age 22-27
Male	0.902***	0.895**	1.117***	0.911***	-1.765***	-0.629
	(0.218)	(0.323)	(0.176)	(0.246)	(0.318)	(0.402)
Female	0.990***	0.826***	1.110***	0.653**	-0.643	1.088**
	(0.230)	(0.205)	(0.163)	(0.249)	(0.395)	(0.421)
Observations	35	27	35	27	35	27
R-squared (male)	0.497	0.493	0.529	0.514	0.579	0.218
R-squared (female)	0.546	0.386	0.512	0.367	0.264	0.517

Standard errors in parentheses

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

Table 6. Medium and long-run effects of doing unpaid work at 16 – male vs female

	works in agriculture		type of work: unpaid		type of work: paid	
	age 18-21	age 22-27	age 18-21	age 22-27	age 18-21	age 22-27
male	1.683***	-0.528	1.736***	0.319	-2.293***	-1.013
	(0.211)	(0.733)	(0.189)	(0.579)	(0.385)	(0.637)
female	1.328***	0.798*	1.604***	0.639	-0.332	1.184
	(0.269)	(0.432)	(0.207)	(0.542)	(0.491)	(0.725)
Observations	35	27	35	27	35	27
R-squared (male)	0.742	0.298	0.708	0.158	0.585	0.193
R-squared (female)	0.632	0.187	0.760	0.142	0.215	0.431

Robust standard errors in parentheses

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

Table 7. Two-Step vs. Pooled, early youth experience on later work outcomes

	works in agriculture		type of work: unpaid		type of work: paid	
	Female	Male	Female	Male	Female	Male
Completed lower secondary by age 17						
Pooled	-0.658***	-0.613***	-0.684***	-0.764***	-0.294	0.744**
	(0.130)	(0.186)	(0.118)	(0.214)	(0.199)	(0.343)
N	62	62	62	62	62	62
R-squared	0.397	0.587	0.461	0.446	0.548	0.452
2-step (Second Stage)	-0.818***	-0.888**	-0.913***	-1.165***	-0.294	1.325**
	(0.186)	(0.302)	(0.200)	(0.324)	(0.200)	(0.435)
N	11	11	11	11	11	11
R-squared	0.682	0.491	0.698	0.590	0.194	0.507
Works in agriculture at age 16						
Pooled	0.848***	0.879***	0.891***	1.024***	0.074	-1.358***
	(0.149)	(0.173)	(0.132)	(0.139)	(0.313)	(0.290)
N	62	62	62	62	62	62
R-squared	0.538	0.705	0.451	0.607	0.381	0.583
2-step (Second Stage)	1.129***	1.111***	1.238***	1.285***	0.120	-1.771***
	(0.313)	(0.276)	(0.376)	(0.286)	(0.351)	(0.393)
N	11	11	11	11	11	11
R-squared	0.591	0.643	0.547	0.692	0.013	0.693
Does unpaid work at age 16						
Pooled	1.096***	1.236***	1.323***	1.446***	-0.007	-2.026***
	(0.208)	(0.253)	(0.188)	(0.229)	(0.368)	(0.353)
N	62	62	62	62	62	62
R-squared	0.536	0.701	0.557	0.601	0.380	0.602
2-step (Second Stage)	1.307***	1.616***	1.519***	1.864***	0.190	-2.622***
	(0.280)	(0.349)	(0.303)	(0.353)	(0.368)	(0.453)
N	11	11	11	11	11	11
R-squared	0.708	0.705	0.736	0.756	0.029	0.789

Robust standard errors in parentheses

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

References

Antman F, McKenzie DJ. Earnings Mobility and Measurement Error: A Pseudo-Panel Approach. Economic Development and Cultural Change. 2007; 56(1):125-61.

Attanasio, O. "Cohort Analysis of Saving Behavior by U.S. Households" Journal of Human Resources, vol. 33 no. 3 (Summer 1998), pp. 575-609, 1998.

Bassi, M. and S. Galiani (2009). Youth Labor Market in Latin America: A descriptive analysis, *Unpublished*, Inter-American Development Bank, Washington DC.

Benveniste, Luis, Jeffery Marshall, and M. Caridad Araujo. 2008. Teaching in Cambodia. Washington, DC: World Bank and Royal Government of Cambodia.

Blundell, R., M. Browning and C. Meghir (1994), Consumer Demand and the Life-Cycle Allocation of Household Expenditures, Review of Economic Studies, 61, 57–80.

Browning, M., A. Deaton and M. Irish (1985), A Profitable Approach to Labor Supply and Commodity Demands over the Life Cycle, Econometrica, 53, 503–543.

Burgess, S., C. Propper, H. Rees and A. Shearer (2003) The class of 1981: the effects of early career unemployment on subsequent unemployment experiences, *Labour Economics*, 10(3), pp. 291-309.

Deaton, A. (1985) Panel Data from Time Series of Cross-Sections. *Journal of Econometrics*, vol. 30, pp. 109-126.

Deaton, A. and Ch. Paxson (1994). Saving, Growth, and Aging in Taiwan. NBER Chapters, in: *Studies in the Economics of Aging*, pp. 331-362, National Bureau of Economic Research.

De Walque D. The long-term legacy of the Khmer Rouge period in Cambodia. World Bank Policy Research Working Paper. No. 3446. November 2004.

Freeman, R.B. and D. Wise (1982). The youth labor market problem: Its nature, causes and consequences, in: R.B. Freeman and D. Wise (Eds.), *The youth labor market problem: Its nature, causes and consequences* (Chicago, IL: The University of Chicago Press), pp. 1-16.

Gender and Development in Cambodia (GAD/C) Gender Scorecard. Phnom Penh: GAD/C. 2002.

Kanol, H.; Khemarin, K.; Elder, S. 2013. “Labour market transitions of young women and men in Cambodia”, Work4Youth Publication Series No. 2, Sep. (Geneva, ILO).

Szekely, M and Karver J. Youth Out of School and Out of Work in Latin America: a Cohort Approach” World Bank Policy Research Working Paper. No. 7421. September 2015.

Tandon P, Fukao T. Educating the next generation: Improving teacher quality in Cambodia. World Bank Publications; 2015 Jan 8.

Verbeek M. Pseudo-Panels and Repeated Cross-Sections. In: Mátyás L, Sevestre P, editors. The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 369-83.

Weller, J. (2007). Youth employment: Characteristics, tensions and challenges, *CEPAL Review*, 92, pp. 61-82.

APPENDIX

Table A1. CSES Sample Size by Age and Cohort

	survey year										
Year of Birth	**2003**	**2004**	**2005**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
1980	233	1,242	111	360	335	1,092	367	313	331	303	915
1981	191	1,282	104	331	340	1,036	338	346	260	316	945
1982	221	1,284	89	363	324	1,115	297	294	371	275	952
1983	233	1,231	107	398	409	1,112	316	315	305	352	872
1984	261	1,358	96	350	341	1,200	324	331	318	262	1,048
1985	246	1,284	109	392	368	1,088	360	315	363	295	859
1986	231	1,522	94	320	352	1,154	289	361	332	321	984
1987	282	1,463	138	411	319	1,212	319	318	396	290	992
1988	274	1,482	109	400	332	1,164	351	302	317	351	983
1989	264	1,571	111	425	367	1,239	316	342	358	346	1,112
1990	337	1,658	126	402	440	1,242	372	325	365	362	1,016
1991	272	1,859	126	412	400	1,413	367	377	342	350	1,167
1992	266	1,721	146	461	423	1,336	403	347	414	355	1,115
1993	326	1,669	123	458	430	1,401	392	402	366	416	1,048
1994	247	1,722	108	456	438	1,413	393	343	422	350	1,216
1995	266	1,522	149	442	417	1,328	377	367	381	421	1,035
1996	299	1,469	91	381	394	1,358	360	351	344	359	1,156
1997	205	1,491	111	392	352	1,265	351	274	353	328	962
1998	233	1,386	107	342	380	1,188	347	347	317	296	974

Table A2. Age of Cohorts in Different CSES Years

	survey year										
Year of Birth	**2003**	**2004**	**2005**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
1980	23	24	25	27	28	29	30	31	32	33	34
1981	22	23	24	26	27	28	29	30	31	32	33
1982	21	22	23	25	26	27	28	29	30	31	32
1983	20	21	22	24	25	26	27	28	29	30	31
1984	19	20	21	23	24	25	26	27	28	29	30
1985	18	19	20	22	23	24	25	26	27	28	29
1986	17	18	19	21	22	23	24	25	26	27	28
1987	16	17	18	20	21	22	23	24	25	26	27
1988	15	16	17	19	20	21	22	23	24	25	26
1989	14	15	16	18	19	20	21	22	23	24	25
1990	13	14	15	17	18	19	20	21	22	23	24
1991	12	13	14	16	17	18	19	20	21	22	23
1992	11	12	13	15	16	17	18	19	20	21	22
1993	10	11	12	14	15	16	17	18	19	20	21
1994	9	10	11	13	14	15	16	17	18	19	20
1995	8	9	10	12	13	14	15	16	17	18	19
1996	7	8	9	11	12	13	14	15	16	17	18
1997	6	7	8	10	11	12	13	14	15	16	17
1998	5	6	7	9	10	11	12	13	14	15	16

Table A3. Effect of completing secondary schooling on later work and social outcomes (age 22-28) - Males vs Females

	works in agriculture		**type of work: unpaid**		**type of work: paid**	
	Female	Male	Female	Male	Female	Male
(1) Lower secondary by age 17	-0.676***	-0.583**	-0.421**	-0.551**	-0.854***	0.226
	(0.170)	(0.272)	(0.170)	(0.228)	(0.249)	(0.326)
(2) Upper secondary by age 20	-1.483***	-2.447***	-1.624***	-2.115***	-0.356	1.246**
	(0.264)	(0.329)	(0.184)	(0.367)	(0.694)	(0.507)

Observations (1)	35	35	35	35	35	35
Observations (2)	53	53	53	53	53	53
R-squared (1)	0.308	0.266	0.229	0.222	0.611	0.171
R-squared (2)	0.298	0.355	0.601	0.315	0.494	0.110

Robust standard errors in parentheses

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

Table A4. Effects of completing lower secondary schooling by age 17, pre and post 1989 cohort

	works in agriculture		**type of work: unpaid**		**type of work: paid**	
VARIABLES	Pre-1989	Post-1989	Pre-1989	Post-1989	Pre-1989	Post-1989
Male	-0.220	-1.064***	-0.195	-1.435***	0.145	1.466***
	(0.280)	(0.316)	(0.159)	(0.204)	(0.287)	(0.324)
Female	-0.457**	-1.121**	-0.238**	-1.493***	0.111	0.558
	(0.209)	(0.469)	(0.094)	(0.271)	(0.298)	(0.385)
Observations	27	21	27	21	27	21
R-squared (male)	0.691	0.761	0.720	0.797	0.768	0.663
R-squared (female)	0.374	0.611	0.447	0.708	0.289	0.531

NOTE: We restrict ages to those common in both periods, as a robustness check (ages 19 to 24).

Robust standard errors in parentheses

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

Table A5. Effects of completing upper secondary schooling by age 20, pre and post 1986 cohort

	works in agriculture		**type of work: unpaid**		**type of work: paid**	
VARIABLES	Pre-1986	Post-1986	Pre-1986	Post-1986	Pre-1986	Post-1986
Male	-1.496	-1.682**	-0.926	-1.277*	1.277	0.729
	(1.293)	(0.631)	(0.916)	(0.647)	(1.809)	(0.772)
Female	-0.392	-0.801*	-0.192	-1.697***	-2.277	0.060

	(0.703)	(0.432)	(0.602)	(0.219)	(1.788)	(0.600)
Observations	27	21	27	21	27	21
R-squared (male)	0.199	0.551	0.187	0.290	0.132	0.138
R-squared (female)	0.164	0.139	0.169	0.723	0.520	0.511

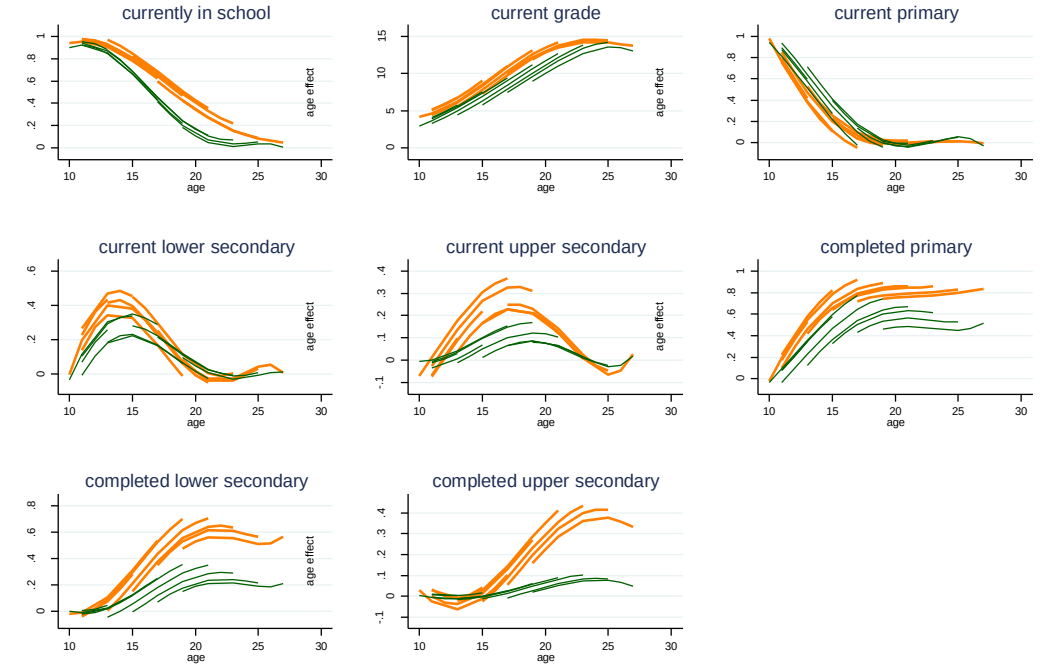
NOTE: We restrict ages to those common in both periods, as a robustness check (ages 22 to 27).

Robust standard errors in parentheses

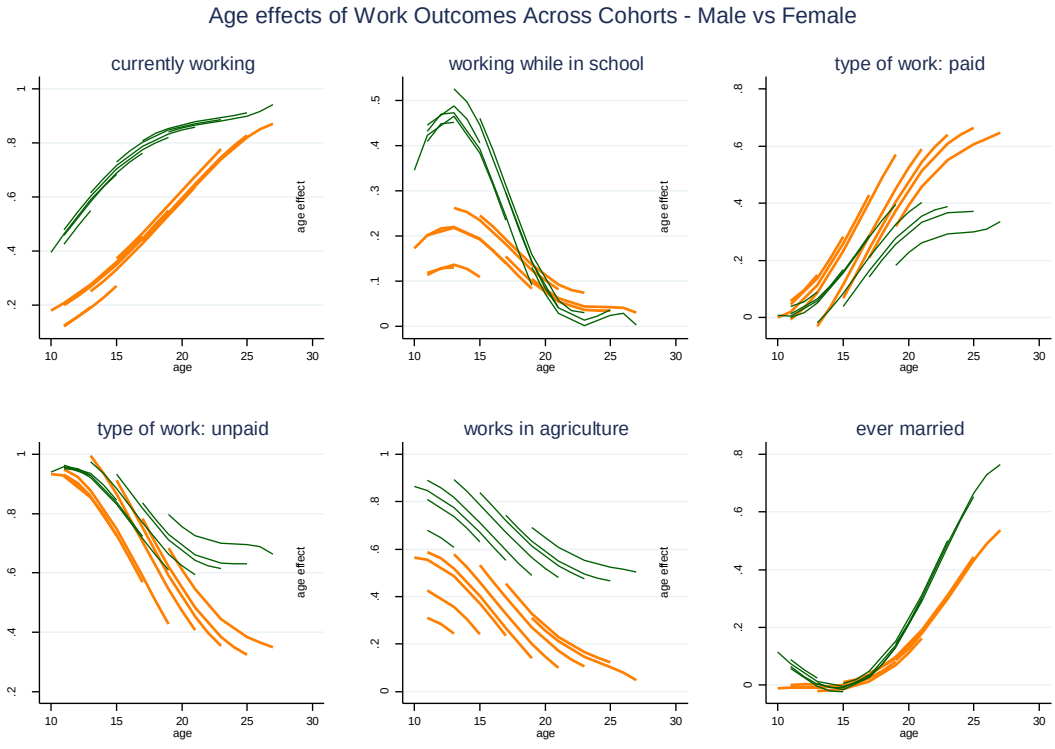
*** p<0.01, ** p<0.05, * p<0.1

Robustness checks for urban versus rural using sub-sample of those who have not moved since birth, years 2003-2004, 2005, 2007, 2009, 2010 and 2011

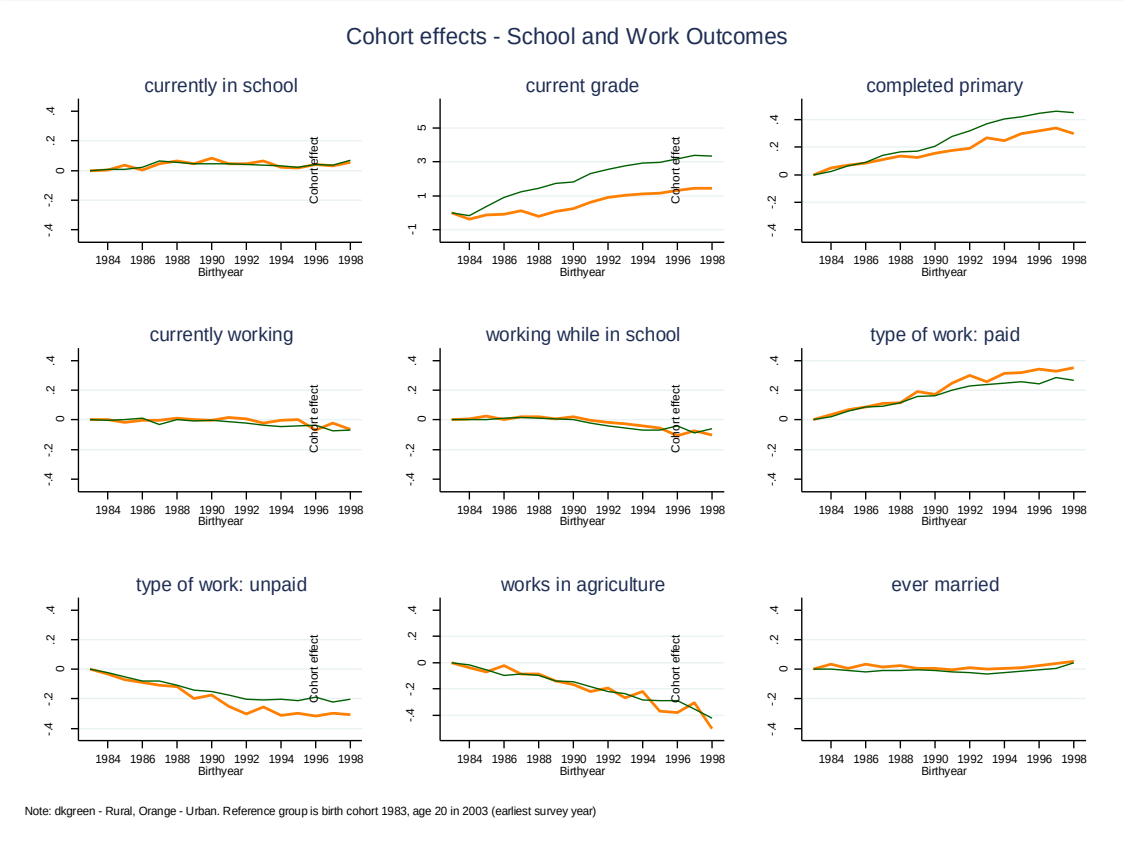
Note: This is the age profile for an arbitrary cohort (1984). Age is modeled as a 5th order polynomial.



Note: dkgreen - Rural, Orange - Urban. Age is modeled as a 5th order polynomial. Each line corresponds to a birth cohort spanning 1984 to 1998.



Note: dkgreen - Rural, Orange - Urban. Age is modeled as a 5th order polynomial. Each line corresponds to a birth cohort spanning 1984 to 1998.



Note: dkgreen - Rural, Orange - Urban. Reference group is birth cohort 1983, age 20 in 2003 (earliest survey year)