Learning from siblings? Family experience, labor market expectations and human capital investment in South Africa

Mame Fatou Diagne*

16 June 2011

Abstract

This paper analyses the impact of the labor market experience of household members (siblings) on young adults' wage expectations and their investment in human capital. It empirically tests predictions of a dynamic learning model of education and labor market entry decisions given subjective beliefs about wages. Heterogeneous beliefs are associated with differences in ambitions and dropping out behavior. Youths with limited information put a large weight on the experience of their siblings when forming expectations of the typical wages in the labor market. This social learning process influences their reservation wages, and their education and labor market entry decisions.

JEL-Classification: D83, I21, J24, O15

Keywords: Labor market expectations, Learning, Social networks, Subjective Expectations, Schooling Decisions, South Africa

* University of California, Berkeley. 508-1 Evans Hall #3880, Berkeley, California 94720-3880, USA. Present Address: Université Gaston Berger, Saint Louis, Senegal and The World Bank, 1818 H Street, NW, Washington, DC 20433, USA. Email: mfdiagne@worldbank.org. I am grateful to Ted Miguel, Jerome Adda, Bryan Graham, Ethan Ligon and Emmanuel Saez for guidance and advice. I thank David Card, Maitreesh Ghatak, Jeremy Magruder, Orazio Attanasio, Oriana Bandiera, John Bellows, Paulina Oliva Vallejo, Gabriela Dobrescu, Eva Arceo Gomez and seminar participants at Berkeley and the LSE for helpful comments and remarks. The Cape Area Panel Study Waves 1-2-3 were collected between 2002 and 2005 by the University of Cape Town and the University of Michigan, with funding provided by the US National Institute for Child Health and Human Development and the Andrew W. Mellon Foundation.

1. Introduction

With imperfect information, individuals may form expectations of the returns to investment in human capital or job search based on the experience of other members of their household or neighborhood. These expectations may in turn influence their choices. While approximating one's own expected returns to effort by those observed among other household members may be rational and efficient if family endowments indeed determine individual outcomes, it could induce large departures from optimal actions if individual endowments and choices are also important. This social learning of the returns to effort (from a limited sample of relatives and friends) could be an additional reason (besides permanent family characteristics such as wealth, parents' education, school quality or segmented labor markets) why "discouragement" is perpetuated within certain families and neighborhoods, as well as unemployment and low educational attainment.

Understanding how individuals get their information and form their expectations about employment and wages, returns to education and to job search effort can shed light on the concentration of unemployment, the formation of "discouraged workers" and the mechanisms of the intergenerational reproduction of inequalities and poverty. These issues are particularly salient in South Africa where unemployment is very high (at 25%[1] in March 2011), as well as the number of "discouraged workers" (i.e. persons who report wanting to work but are not included in the official definition of unemployment because they did not take active steps to look for a job in the past four weeks), who accounted for 6.9% of the working-age population. They are also relevant for many developing countries, notably in Africa where there have been changes in actual and perceived returns to education, in particular as civil service recruitments and public sector wages collapsed in the 1990s. In high unemployment contexts, expectations of low returns to education or simply high uncertainty can have deleterious effects by discouraging human capital accumulation among youth. It is striking that in the sample considered in this chapter, over 40% of Black youth were unsure whether obtaining the high school leaving exam generally made it more likely to get a better job (as opposed to only 11% of young Coloreds and 9% of young Whites, see figure 1). With uncertainty, negative "herd

[1] During the period of the study, unemployment declined, going from 27.7% in March 2002 to 24.2% in March 2005.

externalities" in learning (Banerjee, 1992) can discourage youth in high unemployment families or neighborhoods from investing in education, leading to a vicious circle or poverty trap, for reasons other than low income, even when the government provides the poor with social grants. Conversely, a possible policy implication of these externalities is that employment programs for youth in high unemployment areas can have multiplier effects on education and job search effort and help address the skills shortage problem. More generally, the provision of information about the labor market and, where necessary, incentives for human capital investment could be beneficial if many young adults under- or over-invest in education due to imperfect information about the returns.

The objective of this paper is to determine the importance of the labor market experience of other household members on young adults' own beliefs about the wages they can expect and their investment in human capital. This entails addressing two main questions. First, how do young adults form their expectations of returns to investment in human capital or, more precisely, can their expectations be mostly explained by their own characteristics (including individual ability and family endowments) and the objective average labor market conditions in their area of residence, or does the experience of members of their closest reference group matter significantly? Second, can variations in wage expectations explain the amount of investment in human capital?

These questions can be examined empirically by using a rich panel dataset from urban South Africa, the Cape Area Panel Study (CAPS), which has the advantage of combining education, job search and employment data with subjective variables about the beliefs and expectations of young adults, across time. Siblings are a particularly useful reference group to study local learning effects because they are a close and well-defined social network surrounding adolescents. In the CAPS data, up to three young adults were intensively interviewed in every household and, in addition, education, job market status and wages were collected for all household members. Neighborhood and school identifiers also allow studying learning within other potential information networks, although these are not as well-defined as the household.

An important identification concern when relating the decisions of a cohort to the outcomes of an adjacent cohort is that both may be caused by an unobserved factor. Specifically, a correlation between the educational attainment of younger siblings and the

labor market outcomes of older siblings could indicate that students learn about their own abilities and prospects; but, alternatively, it could be that learning does not play a significant role, and siblings simply share characteristics which are unobserved by the researcher (and were known from the beginning by the family). This analysis uses a rare feature of the CAPS dataset, namely the availability of individual stated beliefs (young adults' expectations about typical wages for someone with their education level; their reservation wages; the education level they expect to attain; etc.) to identify social learning from the experience of siblings. The extent of learning is therefore measured by changes in stated subjective beliefs (about wages for a given education level) in response to new information from older siblings ("signals").

Exploiting the panel structure of the data is, in itself, insufficient to address the identification problem, as both the decisions of younger siblings (e.g., deciding to drop out) and the outcomes of older siblings (e.g., an older sibling dropping out and finding a well-paid job without completing high school) may be caused by unobserved time-varying household level shocks (e.g., the family getting a new "connection" which facilitates finding an unskilled job). However, using stated beliefs about the typical wage for a level of education can directly shed light on learning from the experience of others. As for the possible omitted variable bias problem when relating the outcomes of different cohorts, it can be alleviated by using the wealth of information on observable household level-shocks (including new employment and wages or grants).

The findings support the hypothesis that social learning and the uncertainty caused by insufficient labor market information play a large role in the education decisions of certain groups. Education and labor market entry decisions are modeled given beliefs about ability and wages. The predictions of the model are then tested empirically. Heterogeneous beliefs are found to be associated with differences in ambitions and dropping out behavior. The estimates imply that youths put a large weight on the experience of their siblings in the job market when estimating the typical wages in the labor market for someone of their age and education. This also influences their reservation wages and their education decisions, though the role of beliefs seems lower than that of individual and family endowments.

A large literature provides empirical evidence on the effects of family background or

neighborhood characteristics on youth choices and behavior. But when analyzing education or labor market entry decisions, economists often assume that expectations are rational and homogenous (Manski, 1993). This is partly justified by the scarcity of subjective data on expectations and preferences. In their structural estimation of a sequential model of high school attendance and work decisions, Eckstein and Wolpin (1999) tackle this missing data problem by treating (unmeasured) preferences, abilities and expectations as part of a set of time-invariant "initial traits" that define a fixed number of discrete types of youths[2] . Still in the United States, the heterogeneity of such "traits" has been underscored by Wilson (1987), who has stressed the role of "collective socialization" within neighborhoods to explain the deteriorating condition of disadvantaged youth in inner cities characterized by growing social isolation. Segmented reference groups can indeed produce different subjective beliefs and expectations. An example is provided by Piketty (1995) who explains how differences in perceptions about social mobility can generate differences in political views and attitudes towards redistribution by modeling rational agents who try to learn from their family's dynastic income mobility experience the relative importance of effort and predetermined factors in the generation of income inequality.

A few recent papers have shown the relevance of such beliefs or expectations for educational choices in developing countries. Using a randomized experiment in the Dominican Republic, Jensen (2010) finds that perceived returns to completing secondary school were on average much lower than measured actual returns and that providing information significantly increased the number of students who remained enrolled after one year. Similarly, Nguyen (2008) shows that increasing perceived returns to education (through a randomized provision of statistics or the presentation of a role model) produces significant improvements in test scores and student attendance, especially when students and the role model have a similar background.

However, the relative importance of material endowments on the one hand and beliefs

[2] Their estimates (from the 1979 youth cohort of the US National Longitudinal Surveys of Labor Market) imply that youths who drop out are different from those who graduate (in their ability, motivation, labor prospects, etc) and therefore that mandatory attendance or legislation preventing them from working would have little impact on graduation rates if initial traits were left unaltered.

and expectations on the other hand (shaped by learning, social norms, role models, etc) remains uncertain, though it determines the effectiveness of policies aimed at addressing low educational attainment and inequality. Using data on the subjective earnings expectations of youths (measured at one single point in time but under different alternative scenarios: with completed high school and with completed college), Kaufmann (2007) and Attanasio and Kaufmann (2009) find that expected returns and perceived risk from education are important determinants of college attendance but that, consistent with credit constraints, this link is weaker for poorer students. Yet, in the absence of longitudinal data, these papers do not provide information on the evolution of beliefs and expectations in response to variations in information, and the implications for the decisions to complete school or enter the labor market.

2. Data

2.1. Survey description

The empirical analysis uses the Cape Area Panel Study (CAPS), a longitudinal survey of 4,800 young adults in Cape Town, South Africa (Lam, Seekings, and Sparks, 2006). This urban survey consists of four rounds over five years (2002-2006) and is a representative sample of youth aged 14 to 22 in 2002. Only the first three rounds are used in this analysis. It is particularly suited to identifying social learning effects in education, as it contains yearly information on education (enrollment, attainment, exam results, time spent on homework, reasons for dropping out, etc), employment and job search but also beliefs and expectations (expected schooling attainment, expectations about chances of unemployment, of having a good job, reservation wage, etc). Up to three young adults were intensively interviewed in every household and the survey also provides information on household members beside young adults (including their education, labor market status, wages and receipt of government grants). Summary statistics are presented in table 1.

School and neighborhood identifiers allow considering other reference groups. The survey also contains information on school quality and neighborhood characteristics (including mean unemployment, income, education).

Attrition between wave 1 and wave 3 was mostly due to migration. It was larger among

Whites and Blacks than Coloureds. White graduates migrated to attend university and/or go abroad. Many Black youth migrate between Cape Town and other areas within its Eastern Cape hinterland. Such migratory patterns are likely to correlate with perceived labor market returns and actual education and employment opportunities. This attrition problem was attenuated by information collected from other household members or neighbors about the current location and occupation of the young adult. After using this information, data on the occupational status of respondents was missing for 16% of young adults by wave 2 and 24% by wave 3 (after three years).

2.2. Measurement of subjective beliefs

In all periods, all young adults were asked about their monthly (as well as weekly and daily) reservation wage:

> *What is the absolute lowest monthly take-home wage that you would accept for any full-time work?*

In waves 2 and 3, they were also asked questions about a "typical wage":

> *What is the typical take-home monthly wage for other people like you (same age, education, and skills) who have full-time jobs?*

Subjective expectations about the probability of unemployment over the next few years were only elicited in the first wave. However, a more qualitative question was asked in all periods, about the probability of getting a well-paid job:

> *What do you think are the chances that you will ever get a job that pays well?*

3. Education and the labor market in South Africa

3.1. Educational system and returns

Education in South Africa is compulsory from age 7 (grade 1) to age 15 or the completion of grade 9 (cf. South African Schools Act, 1996). Primary school ranges from grade 0 to 7. High school spans grade 8 to 12 and is completed by the matriculation (also known as "matric" or senior certificate) exam, which provides a national qualification, required for university attendance. In this paper, I define as "dropouts" youths who leave high school without passing the matric exam. This focus on the matric exam is based on the sharp discontinuity identified at this level by the literature on returns to education in

South Africa. They are generally found to be very low for much of primary and secondary education but very high with completed secondary education and tertiary education (Keswell and Poswell, 2004). The returns to tertiary education were found to be even higher for youths from historically disadvantaged backgrounds (Mwabu and Schultz, 1996) due to a skills shortage combined with affirmative action policies (access to government jobs and Black Economic Empowerment).

3.2. Dropping out of high school in Cape Town

Examining the CAPS data, a lower educational attainment and a high dropout rate for Black youths point to the persistence of past inequalities (see table 2). There could be many reasons why some young adults do not complete high school, including differences in preferences, low ability or poor quality of education, high monetary costs combined with credit constraints, heterogeneity in returns to education, or "psychological" costs related to low expectations, the absence of role models and uncertainty. This paper focuses on such psychological costs, particularly the role of uncertainty. It is worth noting that monetary costs and credit constraints are unlikely to be important in the education decision of the young adults in this study: monetary costs of attending primary and secondary school are very low, distance to school is not a concern in Cape Town (the second largest metropolitan area in the country), and the opportunity cost of attending school is low due to very high youth unemployment. This is supported by Edmonds (2006) who provides evidence of credit constraints in education in rural areas in South Africa but not in urban areas.

In light of the institutional characteristics described above, one would expect most rational young adults to leave school either at age 15 (when compulsory education ends) or at least with the matric exam (when they start earning positive returns to education). But fewer than 10% of dropouts leave school at age 15 and many Black students delay this decision until their early 20s (cf. figure 1.5).

3.3. The role of uncertainty in human capital investment decisions

Uncertainty could explain why so many Black young adults remain in school for so long without obtaining the matric exam and with low effort, if they are remaining in school

until they have sufficient information to make a decision. This could take the form of uncertainty about their own ability and their chances of success, which Lam et al. (2010) describe as very high due to a great degree of randomness in the grade advancement process in low quality schools. In this paper, I also hypothesize substantial uncertainty about the returns to education: adolescents, especially those from disadvantaged backgrounds, have insufficient information about the labor market. Evidence of this is provided by the very high proportion (over 40%) of Black young adults in CAPS who answered "don't know" or were not sure when asked whether obtaining the matric exam allows getting a better job. By contrast, most White and Coloured respondents either agreed or disagreed that there were positive returns to completing high school, with less than 10% of uncertain answers (see figure 1). This high uncertainty about returns to education among historically disadvantaged youth can be interpreted in light of two facts. First, South Africa has been undergoing profound structural changes, both in the political and the economic arenas. With the end of apartheid in the early 1990s, new opportunities have emerged for the majority non-White population, meaning that youths face a different set of opportunities than their parents, increasing the need to update beliefs about returns to education. Also, youth living in areas or families where few have obtained education and where unemployment is high are likely to have poor information about wages and employment perspectives. Another reason for this high level of uncertainty may be a large heterogeneity in individual returns to education due to wide differences in school quality or mechanisms other than education attainment that determine returns (for example, Magruder (2010) shows the importance of family networks in finding employment).

4. Theoretical model

4.1. Basic setup

In every period t, individuals consider the decision to enroll in school or not. It is assumed that the decision to leave school is irreversible: once a student drops out or graduates, he does not return to school. The stopping rule is therefore described by the control variable d_t, which can take values:

$d_t = \begin{cases} S & \text{if enroll in school for an additional year} \\ W & \text{if stop attending school (and enter the labor market)} \end{cases}$

Since this paper focuses on the decision to drop out or graduate from high school, the analysis is restricted to young adults initially aged 14-22 and enrolled in grades 7 to 12. Let Y denote the number of years of completed education; $Y \in (7, 12)$.

Value functions

The value of stopping school and working depends on the intertemporal discount rate β and the natural logarithm of the expected wage w. To reflect the nonlinear returns to education, the wage function is assumed to be a step function of completed education Y, such that every young adult needs to form expectations about only two wage levels L, depending on whether (s)he dropped out of school (w^D) or graduated from high school with the matriculation exam (w^M):

$$w^L = \begin{cases} w^D & if\ Y < 12 \\ w^M & if\ Y \geq 12 \end{cases}$$

Assuming an infinite horizon and that, for each individual, the wage remains constant across time, the value of working with Y completed years of education is:

$$V_t^W(Y) = \begin{cases} w^D + \beta V_{t+1}(Y) = \dfrac{w^D}{1-\beta} & if\ Y < 12 \\ w^M + \beta V_{t+1}(Y) = \dfrac{w^M}{1-\beta} & if\ Y \geq 12 \end{cases} \qquad (1)$$

The value of enrolling after having completed T years of education depends on an immediate individual-specific utility noted c (which includes the consumption benefit, financial and opportunity costs, etc.) and on future choices. Indeed, enrolling to complete an additional year of education will still yield wage w^D if the person drops out; it also carries the option value of continuing education to ultimately complete 12 years of education and earn wage w^M.

$$V_t^S(Y)=\begin{cases} c+\pi\beta max\left[V_{t+1}^S(Y+1),\ E_w\frac{w^M}{1-\beta}\right]+(1-\pi)\beta max\left[V_{t+1}^S(Y),\ E_w\frac{w^D}{1-\beta}\right] & if\ Y=11 \\ c+\pi\beta max\left[V_{t+1}^S(Y+1),\ E_w\frac{w^D}{1-\beta}\right]+(1-\pi)\beta max\left[V_{t+1}^S(Y),\ E_w\frac{w^D}{1-\beta}\right] & if\ Y<11 \end{cases} \quad (2)$$

π is the probability of passing a grade. Because there are only two possible wages for any individual, the option value is likely to be a large component of the value of enrollment for those who have completed less than 11 years of education: enrollment is valuable to them if it gives them the option to ultimately reach matric level while they are improving their information on the benefits of education and labor market entry[3].

Uncertainty, subjective beliefs and expectations. Individuals are uncertain about: (i) their ability and (ii) their future wages. Regarding the uncertainty about their ability, it is just assumed that they have a constant belief p about π, which is a function of past repetition, and initial literacy and numeracy skills (the effort level is not endogenous).

The true (population) distribution of the log wage for education level L is assumed to be normal with mean μ^L and variance σ_L ($L=M$ or D). This assumption reflects the approximately normal distribution of the actual log wages observed in the CAPS data (see figure 3). The distribution of wages for individual i in family f depends on the average wage, his/her family background and individual characteristics. Let w_{fi}^L be the wage distribution for individual i with education level L in family f:

$$w_{fi}^D \sim N(\mu^D+\alpha_{fi},\ \sigma_D^2) \text{ and } w_{fi}^M \sim N(\mu^M+\alpha_{fi},\ \sigma_M^2)$$

where $\alpha_{fi}=\delta_f+\epsilon_i$

α_{fi} is an individual specific parameter that reflects the effect of family endowments and individual market ability (including experience and skills). Young adults' uncertainty about the wage they can expect comes from their uncertainty about the mean wages μ^D and μ^M.

Given individual characteristics (age, skills, etc), each individual needs to form a subjective expectation of the average wage for the two education levels. These

[3] See Stange (2009) on the option value of college enrollment in the United States.

expectations depend on their past and present information. Let ˜ denote these subjective expectations (conditional on the information signals X_t received by individual i up to period $t-1$):

$$E_{it}(\mu^L | X_{t-1}, \ldots, X_0) = \tilde{\mu}^L_{i,t}$$

At time t, the prior distribution of μ^L is $N(\tilde{\mu}^L_t, \sigma^2_{L,t})$. $\tilde{\mu}^L_{i,t}$ is the individual's subjective expectation of the mean wage for education level L ("typical" wage for someone with their age, education and skills) at time t. It is the prior mean of μ^L at time t, conditional on past information. The reservation wage w^*_α is the wage w such that the individual is indifferent between remaining enrolled and leaving school, i.e. such that $V^S_t(Y) = V^W_t(Y)$.

4.2. Learning: Information and signals

This part of the model is adapted from Jovanovic's (1979) matching model and Sargent (1987).

Individuals cannot observe their own wages unless they leave school and go on the labor market. They do not know μ^D nor μ^M but they get noisy signals from their reference groups. In particular, they observe the realized labor market outcomes (employment/unemployment status and wages) of other persons who leave school.

These outcomes result from their (known) characteristics (such as age, education attainment, gender, ...) and "luck". They are therefore trying to make inferences about μ^D and μ^M based on signals X^L_t (wage observations or employment status). In this paper, siblings are the main object of interest but neighbors and former school peers are also considered. Assuming that young adults take these signals as unbiased:

$$X^L_t \sim \begin{cases} N(\mu^L, \gamma^2_L) & \text{for signals from siblings} \\ N(\mu^L, \eta^2_L) & \text{for signals from neighbors} \\ N(\mu^L, \psi^2_L) & \text{for signals from school peers} \end{cases}$$

Let $\overline{X}_{it} = \bar{x}_{it}$ be the mean signal over period t (the mean wage or mean employment status of siblings, neighbors or peers who have just entered the labor market). While most young adults would observe at most just one wage signal from a sibling in any given period, they are likely to observe multiple signals from other reference groups such as neighbors or former school peers. In those cases, the "signal" can be viewed as a sample

mean of wages. Let n be the size of the sample from which they are drawing these wage observations.

They use Bayes's law to update their beliefs. Accordingly, they calculate a posterior probability distribution of μ^L conditional on the average signal $\bar{x}_{it}$ from their siblings and their prior belief $\tilde{\mu}^L_{i,t}$. This posterior distribution is normal, with mean $\tilde{\mu}^L_{i,t+1}$ and variance $\sigma^2_{L,t+1}$:

$$\tilde{\mu}^{L}_{i,t+1}=E_{i,t+1}\left(\mu^{L}\middle|\bar{x}_{i,t}\right)=\frac{\sigma^2_{L,t}}{\sigma^2_{L,t}+\frac{\gamma^2_L}{n}}\bar{x}_{i,t}+\frac{\frac{\gamma^2_L}{n}}{\sigma^2_{L,t}+\frac{\gamma^2_L}{n}}\tilde{\mu}^{L}_{i,t} \qquad (3)$$

and

$$\sigma^2_{L,t+1}=\frac{\sigma^2_{L,t}\gamma^2_L/n}{\sigma^2_{L,t}+\gamma^2_L/n}$$

The updating equation is similar in the case of signals from neighbors (with η^2_L replacing γ^2_L) and school peers (with ψ^2_L replacing γ^2_L).

4.3. Model implications Distribution of subjective expectations

From equation (3), in the simple case where young adults receive only one signal from their siblings and because the random variable $\bar{x} - \tilde{\mu}$ is also normal with mean zero and variance $(\sigma^2_{L,t}+\gamma^2_L)$, we know that $\tilde{\mu}^L_{i,t+1}$ (the subjective expectations of the mean wage, i.e. the "typical wages" quoted by young adults in period t+1) is also normally distributed:

$$\tilde{\mu}^L_{t+1}\sim N\left(\mu^L,\ \frac{\sigma^4_{L,t}}{\sigma^2_{L,t}+\gamma^2_L}\right)$$

$\frac{\sigma^4_{L,t}}{\sigma^2_{L,t}+\gamma^2_L}<\sigma^2_{L,t} \Rightarrow \tilde{\mu}_{t+1}$ and $\tilde{\mu}_t$ have the same mean but $\mathrm{Var}(\tilde{\mu}_{t+1}) < Var(\tilde{\mu}_t)$.

It should be noted that the true wage distribution is assumed to remain constant but, besides learning, changes in the variance of $\tilde{\mu}$ could also reflect a change in the true distribution of wages (a widening or reduction in wage inequality)[4].

[4] The true wage distribution was assumed to be time invariant but there was wage inflation during the period under consideration (2002-2005). Unit labor costs in the non-

Updating of subjective expectations

It can be seen from equation (3) that the bayesian estimator of the typical wage is a linear combination of the prior and the signal mean. Equation (3) has several implications for the extent of updating beliefs on the basis of signals:

(i) The higher the uncertainty about the prior (the greater the prior variance), the more weight the individual puts on signals. This would imply that those who have less information (for example, those who are still in school, or younger, or live in high unemployment contexts) update their beliefs in response to signals to a greater extent than those who have more labor market information.

(ii) Conversely, if the signal is a less reliable indicator of the true mean (i.e. if it has high variance), the weight put on the prior is larger. One would therefore expect a single signal from siblings to influence posterior beliefs about the mean wage more than a signal from a neighbor or a former schoolmate.

(iii) As the number of signals increases, the weight on the prior decreases relative to the weight on signals: the number of signals can compensate for the lower reliability of individual signals.

From equation (3), we can derive the updating equation for the "typical wage" relating posterior beliefs to prior beliefs and signals from siblings:

$$\tilde{\mu}^{L}_{i,t+1} - \tilde{\mu}^{L}_{i,t} = \frac{\sigma^2_{L,t}}{\sigma^2_{L,t}+\gamma^2_L/n}(\bar{x}_{i,t} - \tilde{\mu}^{L}_{i,t})$$

The extent of updating based on signals received from siblings can therefore be estimated by running the regression equation:

$$logtypicalwage_{t+1} - logtypicalwage_t = \rho\ (logsignal_{t+1} - logtypicalwage_t) + \varepsilon \qquad (4)$$

The "signal" is measured as the wage of a sibling who found employment in period t+1; $\rho < 1$ is interpreted as the "updating coefficient". It reflects learning (and not income effects) because income effects should not have an impact on subjective expectations of

---

agricultural sectors increased by a total of 33.9% between 2002 and 2005, i.e. an annual average of 7.6%. However, because wages are all expressed in logs, a uniform change in all wages (due to inflation) would not imply a change in the variance of $\tilde{\mu}$.Wage inequality increased only slightly over the period. The wage premium for workers with a high school degree (matric) as their highest completed level of education over the wage for those with some secondary school was 26.2% in 2001 and 30.8% in 2005 (Hlekiso and Mahlo, 2008).

the typical wage (they can have an effect on the reservation wage w_α^*). It is possible that a "signal" is due to an event that changes expected wages for all household members (such as a new family employment connection, as in Magruder, 2010). But while this could change the reservation wage, it should not affect the expectation of the "typical wage", for someone of similar age, education and skills.

Education decisions and the reservation wage

The value of enrolling, V_t^S, increases in the difference between the graduate and the dropout wage $w^M - w^D$. Consequently, the higher the difference in the subjective expectation of these wages $(\tilde{\mu}_{i,t}^M - \tilde{\mu}_{i,t}^D)$, the more likely it is that the individual will complete high school. Also, when these differences are high, the reservation wage w_α^* that makes the person indifferent between staying in school and joining the labor force will be high. The option value of enrollment that appears in equation (2) may explain why many young adults faced with high uncertainty delay the dropping out decision.

5. Empirical results

5.1. Distribution of subjective expectations

Table 3 presents the dispersion of young adults' subjective beliefs about the mean wage (Var($\tilde{\mu}$)). As predicted by the learning model, the variance of the typical wage diminishes across time. It is also larger for respondents whose education level is below matric than for those who already have matric. The variance of young adults' expectations of the mean wage is higher in high unemployment neighborhoods and families, where one would expect that less information is available about the labor market. Also consistent with learning (and with prolonged enrollment due to the option value created by uncertainty about returns to education), young adults who are still enrolled in school have more variable estimates of the typical wage than those who have entered the labor force and are either working or unemployed.

5.2. Learning: updating of subjective expectations

Updating from wage signals

Table 4 presents the updating coefficient, which measures the degree of updating of respondents' beliefs about mean wages for their level of education in response to the

signals they receive. “Signals” are defined as new information and measured as the wage observed for someone from their age group (14-30), with a similar education level (below matric or with matric), who has left school and found a job. Wage signals are defined only when someone in a young adult’s information network (themselves, a sibling, a former school mate or a neighbor) finds employment and earns a positive salary. In each of these regressions, the dependent variable is the difference between the log typical wages between wave 3 and wave 2 (posterior - prior) and the independent variable is the difference between the signal (log wage of the person who found a job) and the prior. If young adults were just replacing their prior belief about the mean wage with the signal (the wage they have observed between waves 2 and 3, either because they started working or a sibling did), then this updating coefficient would be equal to 1. If they updated in a Bayesian way, taking into account that every signal is a draw from the underlying distribution, equation (3) of the model combined with the variances of actual wages would imply an updating coefficient of 0.56 for young adults whose education level is below matric and 0.40 for those who have at least matric.

The results presented in table 4 provide tests of the model implications outlined in section 4.3: as expected, young adults whose education level is below matric have larger updating coefficients in response to signals than those who have at least matric. This is consistent with the higher uncertainty they face: the variance of their priors about the mean wage is larger and they are likely to have less information about the labor market (because they are younger, come from backgrounds with less information, etc.).

Also, the extent of updating beliefs about the mean wage on the basis of one’s own experience (when one finds a job) is greater than with a sibling’s experience, though very similar. Young adults therefore seem to consider signals from siblings to be almost as reliable as those they would get from their own wage experience in the labor market. The degree of updating in response to their own experience matches the updating coefficient implied by the model and the variance of actual wages for this group.

The (total) updating coefficients for signals from school peers and neighbors are large, though we expect young adults to put less weight on any single signal. Indeed, the overall degree of updating on the basis of all these signals can be expected to be higher than the updating coefficient that would apply to a single signal if the greater amount of

information (more signals) counterbalances the lower reliability of the information. While most young adults received at most only one wage signal from their siblings or themselves, they observe (potentially) many signals from former school mates and neighbors. In this case, the dependent variable is the difference between the average signal (mean wage of peers or neighbors who found a job) and the prior. Contrary to signals from siblings and own experience, not all signals from school peers or neighbors are observed in the data: for every young adult, information is available on the wages of all those who live in the same household but, at the school or neighborhood level, only a subset of young adults were sampled and interviewed.

In summary, the updating coefficient in response to a sibling's experience is only slightly below what would be implied by the Bayesian updating model, given the variance of actual wages. The extent of updating from the experience of a single former school peer or neighbors is much less than the updating based on signals from a sibling or own experience. This could reflect the fact that wage signals from peers and neighbors are measured with greater error (are less reliable) and young adults know less about the unobservables of peers and neighbors. Alternatively, the neighborhood (as defined by the census) or the school may not be the relevant unit of observation for adolescents.

Updating from binary signals

When a young adult leaves school but does not find employment, his siblings do not observe a wage draw for a level of education. Binary signals (employment status of a sibling who dropped out or graduated from school) are likely to be less informative about mean wages than actual wage signals. Still, they can carry information about the probability of employment and the likelihood of finding a well-paid job. Table 5 presents the effect of such binary signals on several measures of beliefs and expectations. The results suggest that young adults update their expectations of the mean wage to a greater extent when their sibling finds employment, i.e. when they receive a wage signal from a sibling (as opposed to when they do not receive a wage signal because their sibling remained unemployed).

An effect on their reservation wage is only found if the sibling graduated with matric but remained unemployed. This is consistent with a reduction in the expected opportunity cost of enrollment (if the young adult infers from the experience of his sibling that there

are no jobs available, even with matric) or a high option value of enrollment due to the lack of information about the wage that can be expected with matric.

Lastly, young adults whose siblings graduated with matric and found employment become more optimistic about their own likelihood of getting a well-paid job.

5.3. Beliefs, Ambitions and dropping out Subjective beliefs and ambitions

Subjective beliefs are strongly correlated with initial ambitions. Young adolescents were asked in every period about their educational goals ("As it stands now, how much education do you think you will complete?") and about a likelihood that they would get a well-paid job[5]. As reported in table 6, beliefs, ability and race have a large and highly significant effect on maximum desired education. Ability is measured by the initial literacy and numeracy test (administered at the beginning of the survey) and past school performance (proportion of grades passed between age 9 and the beginning of the survey). Household income is also highly significant in this regression but not parent's educational attainment. This strong association between beliefs about the likelihood of getting a well-paid job and desired educational attainment is robust to the inclusion of household fixed effects: within a household, young adults who have more optimistic beliefs about the labor market and higher ability also desire to gain more education.

Considering changes in the desired educational attainment, they are also positively associated with changes in subjective beliefs about the likelihood of getting a well- paid job and ability. Again, this association is robust to the inclusion of household fixed effects. Within a household, an adolescent who increases by 25% his perceived likelihood of getting a well-paid job (for example, from "about 50/50" to "high") increases his maximum desired education by almost one year.

It is worth noting that Blacks have higher stated ambitions than Coloureds and are also more likely to raise these ambitions between 2002 and 2005. This is at odds with their lower actual educational attainment but could explain their tendency to delay the labor market entry decision and drop out only in their 20s.

Subjective beliefs and dropping out behavior

[5] Possible answers in the survey were categorical: "very low, low, about 50/50, high, very high" and were recoded as numerical probabilities in the analysis (0, 0.25, 0.5, 0.75, 1).

Similar to ambitions, actual education decisions are strongly correlated with changes in beliefs: young adults who revise upwards their perceived likelihood of getting a well-paid job are significantly less likely to drop out of school, controlling for their ability, age and household income (see table 7). This effect is strong, but less than the effect of ability (and in particular new information about their ability, measured by grades they have passed over the period) and household income. Again, Black youths are slower to drop out than Coloureds and Whites, which could reflect the greater amount of uncertainty that they face in making the education or labor market entry decision, or their lower opportunity cost of additional years in school. The negative association between increases in the reservation wage and dropping out is highly significant and consistent with the model: youth who place a high value on enrollment (because they have a low opportunity cost, or a high expected wage following the additional year of schooling, or a high option value of enrollment if they think returns to obtaining matric are high) are less likely to drop out of high school. But this effect is lower than that of new information about their ability. Youth from poorer households are more likely to drop out, but changes in household income over the period are not associated with a greater likelihood of dropping out. This is consistent with a small role for credit constraints in the decision to complete high school. Predictably, there is no significant relationship between changes in the expected typical wage and the likelihood of dropping out. Indeed, it is the difference between the expected wage with matric and the expected dropout wage which matters in the decision.

6. Conclusion

Local social learning is an important determinant of adolescents' and young adults' beliefs and expectations about the distribution of wages in the labor market. Young adults who get more information about the dropout wage than about the high school graduate wage and those who are more likely to observe graduates getting low wages or remaining unemployed after graduating will have more pessimistic beliefs which, in turn, will lead to a reduction in their ambitions and make them more likely to drop out, even after accounting for the effects of ability and family income.

While there is evidence that young adults' beliefs about the labor market are influenced

by the experience of their school peers and neighbors, it is the large weight that they place on the experience of their siblings that is most striking, as it is based on only one observation in most cases. Changes in beliefs associated with such family experiences have a large impact on ambitions and dropping out, though lower than the effect of ability and family income.

The consequent under-investment in human capital among able individuals (and the inefficiency of delaying the labor market entry decision due to a lack of information among those who will not eventually graduate because of their low ability or low (true) expected returns) would warrant targeting labor market information provision interventions to high unemployment areas. Provided they are credible (as in Nguyen (2008) and Jensen (2010)), interventions that would inform adolescents about wages for high school graduates and dropouts could greatly reduce uncertainty and its associated costs. They could provide a means to reduce the high school dropout rate and increase effort and ambitions that would combine the advantages of cost-efficiency and direct targeting of adolescents enrolled in school.

References

Jerome Adda and Russell Cooper. *Dynamic economics: quantitative methods and applications*. The MIT Press, 2003.

Orazio Attanasio and Katia Kaufmann. Educational choices, subjective expectations, and credit constraints. NBER working paper, July 2009.

Abhijit Banerjee. A simple model of herd behaviour. *Quarterly Journal of Economics*, 107(3):797–817, Aug 1992.

Pranab Bardhan and Christopher Udry. *Development Microeconomics*. Oxford University Press, 1999.

Zvi Eckstein and Kenneth Wolpin. Why youths drop out of high school: the impact of preferences, opportunities and abilities. *Econometrica*, 67(6):1295–1339, November 1999.

Eric V. Edmonds. Child labor and schooling responses to anticipated income in South Africa. *Journal of Development Economics*, 81(2):386–414, December 2006.

Thami Hlekiso and Nthabiseng Mahlo. Wage trends and inequality in South Africa: a comparative analysis. *Labour market frontiers*, South African Reserve Bank, October 2008.

Robert Jensen. The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics*, 125(2): 515-548, May 2010.

Boyan Jovanovic. Job matching and the theory of turnover. *The Journal of Political Economy*, 87(5):972–990, October 1979.

Katja Maria Kaufmann. Understanding the income gradient in college attendance in Mexico: the role of heterogeneity in expected returns to college. Job market paper, November 2007.

Malcolm Keswell and Laura Poswell. Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence. *South African Journal of Economics*, 72(4):834–860, 09 2004.

David Lam, Jeremy Seekings, and Meredith Sparks. The Cape Area Panel Study: Overview and technical documentation for waves 1-2-3, December 2006.

David Lam, Cally Ardington, and Murray Leibbrandt. Schooling as a lottery: racial differences in school advancement in urban South Africa. *Journal of Development Economics*, 2010.

Jeremy Magruder. Intergenerational networks, unemployment, and persistent in- equality in South Africa. *AEJ: Applied Economics*, Vol.2 (1): 62-85, 2010.

Charles Manski. Adolescent econometricians: how do youth infer the returns to schooling? In Charles Clotfelter and Michael Rothschild, editors, *Studies of supply and demand in higher education*, pages 43–57. University of Chicago Press, 1993.

Germano Mwabu and T. Paul Schultz. Education returns across quantiles of the wage function: Alternative explanations for returns to education by race in South Africa. In *The*

American Economic Review, volume 86, pages 335–339, 1996. Papers and Proceedings of the Hundredth and Eighth Annual Meeting of the American Economic Association.

Trang Nguyen. Information, role models and perceived returns to education: Experimental evidence from Madagascar. Job market paper, 2008.

Thomas Piketty. Social mobility and redistributive politics. *Quarterly Journal of Economics*, 110(3):551–584, 1995.

Thomas Sargent. *Dynamic macroeconomic theory*, chapter 2. Harvard University Press, 1987.

Kevin Stange. An empirical investigation of the option value of college enrollment. Working paper, 2009.

William Julius Wilson. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University Of Chicago Press, 1987.

Figure 1: Uncertainty about returns to graduating from high school (passing **"Matric")**

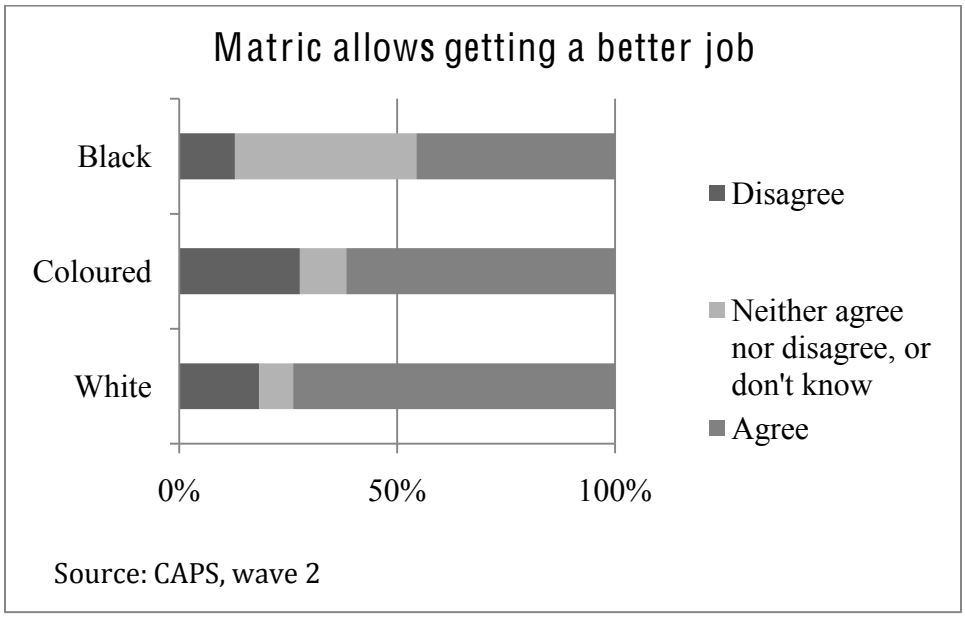


Figure 2: Mean wage by years of completed education

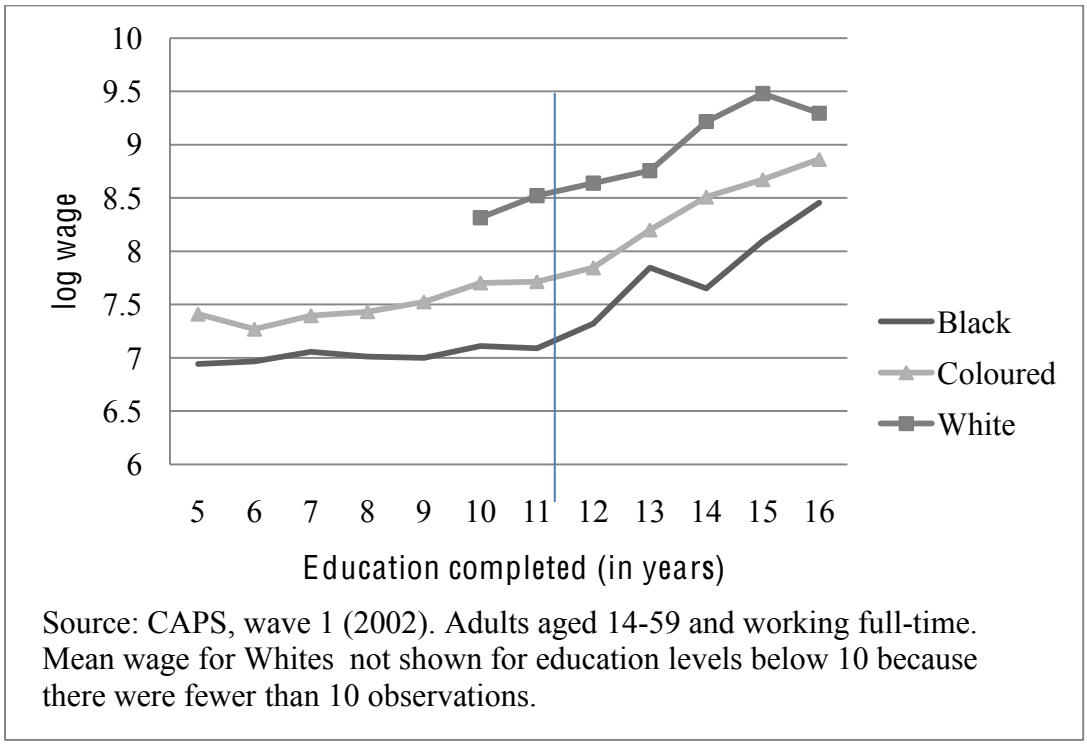
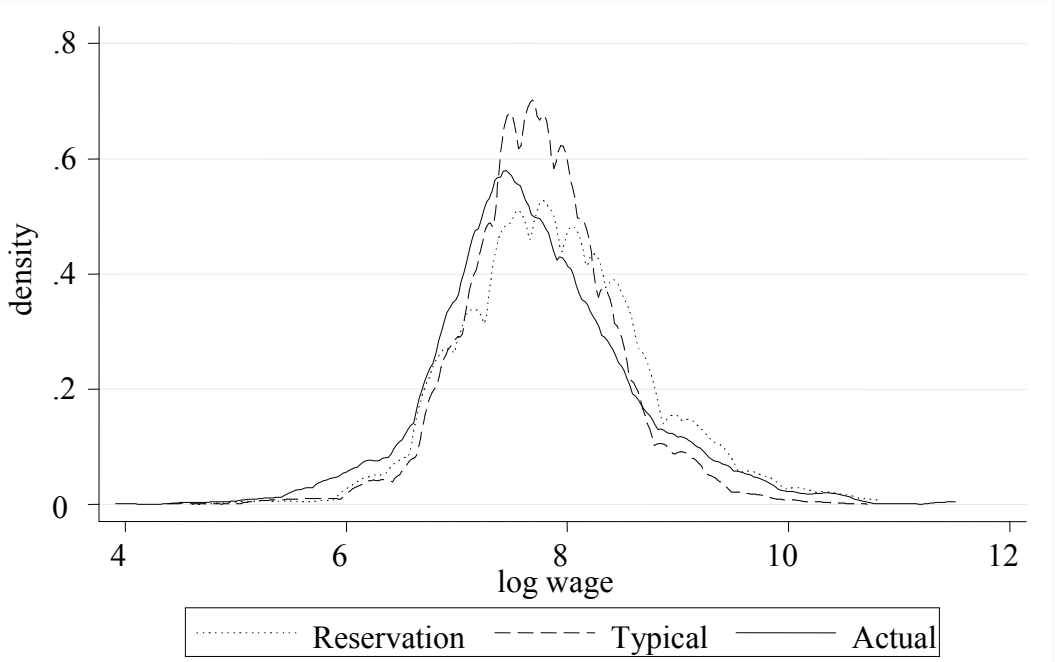
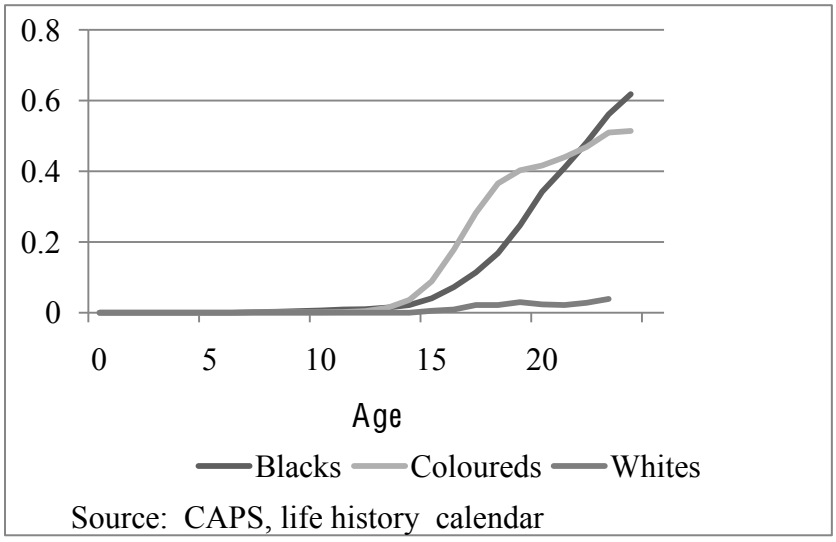


Figure 3: Distribution of wages (actual and subjective)



Distribution of young adults' reservation wages, "typical" wage estimates for a full-time job, and actual wage distribution for working adults. Kernel densities. Source: CAPS, wave 3.

Figure 4: Cumulative proportion of high school dropouts – by age and race



Source: CAPS, life history calendar

Table 1: Summary statistics (means)

	2002	2004	2005
Demographic			
Male	0.450	0.455	0.459
Age	17.8	19.2	20.5
Coloured	0.422	0.431	0.475
Black	0.453	0.464	0.43
White	0.125	0.105	0.095
Occupation			
Enrolled	0.644	0.511	0.378
Work	0.126	0.183	0.336
Unemployed searching	0.072	0.063	0.087
Unemployed discouraged	0.093	0.102	0.058
Inactive	0.044	0.023	0.005
Ever worked	0.34		
Age first worked	16.7		
Ever pregnant (females)	0.171		0.363
Education			
Years of education completed	9.2	10.1	10.5
Maximum education wish to attain	13.9		12.9
Beliefs and expectations			
Reservation wage	1374	2822	3943
Typical wage		2630	2795
Probability of getting a well paid job	0.62	0.64	0.63
N	4752	3932	3536

Source: CAPS, young adult sample

Table 2: Educational attainment and exam results by race

	Black	Coloured	White	Total
Average educational attainment, youth aged 20-29	10.4	10.6	12.39	10.7
% youth aged 20-24 who have passed matric	43.6	55.2	97	55.6
Expected matric pass rate ("Do you expect to pass matric the first time?")	86.2	82.1	95.2	85.3
Actual matric pass rate	75.4	91.1	99.1	87.1
incl. passed with exemption	16.2	23.2	70	31.8
incl. passed without exemption	59.2	67.8	29.1	55.3

Source: CAPS, waves 1, 2, 3

Table 3: Dispersion of estimates of the mean wage ("typical wage")

	Education	
	Below matric (D)	Passed matric (M)
1. Across time		
Wave 2	0.418	0.359
Wave 3	0.353	0.362
2. By occupation (wave 3)		
Enrolled	0.438	0.406
Unemployed	0.234	0.310
Working	0.242	0.311
3. By unemployment rate in neighborhood		
Lowest quartile (% HH head unemployed was <8.7% in 2001 census)	0.261	0.333
Highest quartile (% HH head unemployed was >30.2% in 2001 census)	0.350	0.402
4. By employment rate in household		
Below median (<40%)	0.400	0.399
Above median (>40%)	0.294	0.334

This table displays the variance of the log of subjective estimates of he mean wage by different groups of respondent.
Source: CAPS

Table 4: updating expectations of the "typical wage" in response to new information

	Respondent's update of the "typical wage" (posterior – prior)	
	Below matric (D)	With matric (M)
A. Siblings		
Signal – prior	0.414**	0.298***
	[0.185]	[0.070]
Average number of wage obs.	1.02	1.04
R-squared	0.13	0.27
B. Own experience		
Signal – prior	0.505***	0.291
	[0.134]	[0.182]
R-squared	0.26	0.10
C. School peers		
Signal – prior	0.560***	0.266***
	[0.070]	[0.066]
Average number of wage obs.	1.33	1.53
R-squared	0.29	0.16
D. Neighbors		
Signal – prior	0.542***	0.296***
	[0.037]	[0.051]
Average number of wage obs.	3.89	4.59
R-squared	0.28	0.15

Robust standard errors in brackets. Ordinary Least Squares regressions
* significant at 10%; ** significant at 5%; *** significant at 1%.
The reported coefficients correspond to the estimation of ρ in equation (4).
The independent variable is the difference between the respondent's posterior and prior stated "typical wage"' (in log). The dependent variable is the difference between the wage signal(s) received and the prior stated "typical wage". Wage signals are defined as the actual wage received by a newly employed sibling, neighbor or peer, aged 14-30 and with the same level of education. The coefficients in panels C and D correspond to the cumulative effect of several information signals.

Table 5: Updating expectations (binary signals)

	(1) Δ Typical wage	(2) Δ Reservation wage	(3) Δ Well-paid
Number of siblings who dropped out and became employed	0.345*	-0.137	-0.0777
	(0.195)	(0.176)	(0.0699)
Number of siblings who dropped out and remained unemployed	0.0405	0.0253	-0.00968
	(0.0767)	(0.0718)	(0.0281)
Number of siblings who graduated with matric and became employed	-0.00472	0.0183	0.0397**
	(0.0451)	(0.0451)	(0.0163)
Number of siblings who graduated with matric and remained unemployed	-0.00640	0.0891*	-0.000
	(0.0486)	(0.0473)	(0.0179)
Passed matric	0.0537	0.259***	0.0302
	(0.0938)	(0.0941)	(0.0317)
Number of siblings aged 14 to 30	-0.0301*	0.00680	0.00358
	(0.0173)	(0.0177)	(0.00613)
Constant	0.0584	0.962***	-0.0117
	(0.0366)	(0.0364)	(0.0124)
Observations	1,167	2,463	1,865
R-squared	0.005	0.005	0.005

Standard errors in parentheses. OLS regressions. The dependent variable is a change in the respondent's beliefs over time: change in the log of the "typical wage" (1), change in the log of the reservation wage (2), or change in the perceived likelihood of getting a well-paid job (3).
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Subjective beliefs and ambitions

	Maximum desired education		Change in maximum desired education	
	(1)	(2)	(3)	(4)
Belief about likelihood of getting a well-paid job	0.418***	0.564		
	(0.104)	(0.357)		
Change in belief about likelihood of getting a well-paid job			0.364***	0.915*
			(0.140)	(0.505)
Years of completed education in 2002	0.126***	0.0587	0.0144	0.00645
	(0.0145)	(0.0477)	(0.0278)	(0.0979)
Ability test score result (standardized)	0.354***	0.242*	-0.0527	0.0115
	(0.0383)	(0.135)	(0.0736)	(0.280)
Proportion of grades passed between age 9 and 2001	1.781***	2.090*	1.992***	1.584
	(0.341)	(1.071)	(0.650)	(1.615)
Household per capita income ('000Rs)	0.0974***		0.0225	
	(0.0172)		(0.0272)	
Mother's education	0.0239*		0.0515**	
	(0.0123)		(0.0247)	
Father's education	0.0235**		0.0363*	
	(0.0107)		(0.0191)	
Male	-0.0866	-0.0962	0.0756	0.206
	(0.0573)	(0.160)	(0.102)	(0.349)
Black	1.008***		0.366***	
	(0.0722)		(0.130)	
White	-0.156*		0.356**	
	(0.0944)		(0.160)	
Constant	9.873***	11.08***	-3.606***	-2.297
	(0.353)	(1.010)	(0.723)	(1.818)
Observations	2,381	3,578	1,536	2,313
R-squared	0.180	0.839	0.049	0.844
Household fixed effects	No	Yes	No	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Subjective beliefs and dropping out of high school

	(1)	(2)	(3)
Change in belief about likelihood of getting a well-paid job	-0.841***		
	(0.214)		
Change in log reservation wage		-0.195***	
		(0.0688)	
Change in log typical wage			-0.0101
			(0.138)
Maximum desired education	-0.103**	-0.0648	-0.0363
	(0.0519)	(0.0528)	(0.0693)
Log household per capita income	-0.503***	-0.392***	-0.341**
	(0.109)	(0.122)	(0.152)
Change in log household per capita income	0.149*	0.0745	0.0877
	(0.0905)	(0.0983)	(0.112)
Ability test score result (standardized)	-0.256***	-0.226**	-0.221
	(0.0992)	(0.109)	(0.145)
Proportion of grades passed between age 9 and 2001	-1.921**	-1.680*	-1.755
	(0.790)	(0.872)	(1.199)
Proportion of grades passed from 2002 to 2004	-1.816***	-1.714***	-1.727***
	(0.243)	(0.275)	(0.372)
Constant	-2.537	-4.011**	-2.868
	(1.603)	(1.784)	(2.329)
Observations	1,209	920	491

Robust standard errors in parentheses. Logistic regressions. Include controls for age, race and gender.

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

In all columns, the dependent variable is a dummy variable indicating whether the respondent has dropped out of high school by 2005. The sample includes young adults who were enrolled in high school in 2002.