Misinformation or mismatch? Decomposing the gap between higher education students' earning expectations and realized wages

Draft, v1.0

Sam Jones, Ricardo Santos, and Gimelgo Xirinda

UNU-WIDER

May 2019

Abstract

Inaccurate expectations regarding future wages have been documented among students in a variety of contexts. However, most studies refer to high income countries and the sources of expectational errors are not well-established. This paper extends the literature in two directions. First, we show how to decompose expectational errors into two main parts: (i) misinformation regarding how worker characteristics are rewarded in the labour market; and (ii) mismatches between the anticipated type of job and the actual position obtained. Second, we apply this procedure using rich longitudinal data from a representative sample of students in a low income country, namely Mozambicans who were finishing post-secondary education in 2017. We find that the gap between expected and first wages are very large, equal to nearly 200% of the realized mean wage. Also, while informational errors are substantial, job mismatches of various kinds are critical and account for around half of the total expectational error. This suggests that both rates and types of investment in higher education may be excessive, at least given current labour market conditions.

Acknowledgements: This research was conducted within the Inclusive Growth Mozambique programme, led, in partnership, by the Ministry of Economics and Finance of the Government of Mozambique, Eduardo Mondlane University, the University of Copenhagen and UNU-WIDER, with the support of the Kingdom of Denmark, the Republic of Finland and the Kingdom of Norway. The authors wish to acknowledge the invaluable support in fieldwork and subsequent research assistance by Felix Mambo, Edson Mazive, Thomas Sohnesen, Vincenzo Salvucci and Yonesse Paris. The research also could not have been conducted without the support of the Ministry of Higher Education, Science and Technology and the Ministry of Labour of the Government of Mozambique, the Catholic University of Mozambique, Pedagogic University, Polytechnic University, St. Thomas Aquinas University of Mozambique and UniZambeze. All errors and omission are the our own.

1 Introduction

The importance of labour market expectations is well recognised. Conventional models of human capital formation presume expectations of future income inform educational investments (e.g., Betts, 1996), not just in terms of the quantity of education accumulated (number of years) but also in terms of its quality and type. For instance, when choosing an area of specialization, Schweri and Hartog (2017) propose that students make predictions about the future wages paid to graduates from different courses, and choose the course that may yield the highest net gain. Thus, if students make educational decisions based on subjective expectations that are incorrect (on average), we may see systematic over- or under-investment in education (Webbink and Hartog, 2004). Furthermore, as Becker (1962) suggested, income expectations are likely to inform decisions about job search, including whether to accept a particular job offer or remain in an existing job. Thus, unless these expectations are correct, individuals may reject job offers they mistakenly consider to be underpaid or accept job positions for which they are overqualified.

Guidance from theory suggests that a systematic gap between expected and realized wages is likely to provoke sub-optimal decision-making and even mis-allocation of talent. Existing empirical evidence also suggests this concern is material. Focussing on university students or recent graduates, who are expected to be fairly well-informed about labour market conditions (relative to younger students), only a handful of studies find wage expectations are broadly accurate (Webbink and Hartog, 2004; Van der Merwe, 2011). And while a small number of studies find early wage expectations are pessimistic versus observed market outcomes (Wolter, 2000; Klößner and Pfeifer, 2019), the majority of studies point to optimistic expectations (e.g., Jerrim, 2011, 2015; Wiswall and Zafar, 2015; Abbiati and Barone, 2017). Nonetheless, this evidence is incomplete. Not only does the magnitude of the expectations gap vary substantially across studies (and samples), but the bulk of research only pertains to rich industrialized countries with (comparatively) high levels of graduate employment and large, diverse formal sectors. Few studies of this sort have been undertaken in middle income countries, and even fewer in low income countries where formal labour markets are typically thin and, arguably, the systematic mis-allocation of talent might be more detrimental.

In addition to the limited geographical coverage of empirical evidence on wage expectations, little is known about the factors that account for such errors. To date, a main explanation has been that prospective workers are poorly informed about the distribution of wages across jobs and, thus, incorrectly estimate the magnitude of returns to different levels of education or how specific individual characteristics (e.g., higher academic ability, language skills) are rewarded.

While the contribution of misinformation to expectational errors is plausible, particularly in developing countries with a small population of graduates, findings from a separate literature suggest that labour market mismatches may also generate a gap between desired and realized wages. To see this, note that subjective wage predictions are always made on a conditional basis – i.e., expectations are elicited on the assumption that a range of job characteristics will hold including, minimally, that the individual has found work. Evidence from high income countries suggests that poor job matches, such as being over-educated for a position or working in a field different from that of your training, often incur a wage penalty (McGuinness et al., 2018; Somers et al., 2019). Thus, where the profile of a given work position does not match earlier expectations, this mismatch may imply realized wages are also below expectations. In developing countries, this kind of error is highly plausible. Difficulties in finding 'good' jobs in the formal sector have been extensively documented, especially for younger workers in the sub-Saharan African region (e.g., Filmer and Fox, 2014). And while the specific issue of job mismatches within the formal sector has not received attention outside of high income economies (for an exception see Sam et al., 2018), it stands to reason this phenomenon also may be large.

The aim of this study is to investigate the gap between labour market expectations and realized early-career incomes in Mozambique, a low income country located in Southern Africa. In addition to measuring the size of the expectations gap across different individuals, we seek to identify relevant explanatory factors, differentiating between errors stemming from being misinformed about labour market returns and errors stemming from job mismatches. Since previous studies have not formally quantified the role of *ex post* job mismatches to expectational wage errors, this decomposition itself represents a novel contribution. Furthermore, and also in contrast to most previous studies, we rely on longitudinal data for a representative sample of final year university graduates (subsequently tracked over time) containing information on wage expectations and realizations for the same individuals. This allows us to compare expected and actual returns to a detailed set of individual and job characteristics; also, it allows us to consider potential bias from unobserved selection effects associated with who gains employment.

A main finding is that expectational errors are positive and large. On average, while around two thirds of the sample report undertaking paid work, their starting salary is less than half of what they had expected in their final year of studies. Decomposing this error, a range of vertical and horizontal mismatches appear material. For instance, on beginning work, the majority of participants had not completed all formal study requirements and thus had not yet graduated. Furthermore, many were working as (paid) interns, or on a part time basis, without a contract, and were continuing to look for another job. Taken together, the wage penalties associated with these mismatches are large and account for around half of the overall (average) expectations gap. The remaining half can be attributed to misinformation, including a substantial overestimation of the returns to higher education in particular.

By way of structure, Section 2 reviews previous studies and explanations for gaps between expected and realized labour market outcomes. Section 3 shows how a conventional Mincerian framework can be extended to account for both informational and matching errors via a form of Blinder-Oaxaca decomposition. Moving to our application, Section 4 provides an overview of the Mozambican context and summarises the survey data used here, reporting descriptive statistics of labour market outcomes (expected and actual) for the sample of university students subsequently tracked as they entered the labour market. Section 5 sets out our decomposition results, and Section 6 considers their implications for private pecuniary returns to higher education. Finally, Section 7 concludes.

2 Wage expectations versus reality

The idea there may be systematic differences between the wages expected by students prior to entering the labour market and their actually-realized wages is not new. In an early study, Smith and Powell (1990) found that while college seniors had reasonable knowledge of the average value of higher education, they showed a strong propensity for 'self-enhancement', raising questions regarding the extent to which job seekers might be treated as well-informed. Since then, a range of published studies, summarised in Table 1, have examined the same issue. Typically, these focus on university and/or high school students – both of which are taken as groups with some notion of the labour market and who face important decisions around whether to continue study or pursue work.

Four broad insights emerge from the previous literature. First, and as already noted, the majority of studies find wage expectations are positive in the sense of being over-optimistic. This finding applies not only on average but also after conditioning on a range of background variables or proximate determinants – i.e., it is not driven by errors among specific subgroups. Second, with only rare exceptions, almost all published studies refer to high income contexts (e.g., USA, Western Europe). This is perhaps natural given the scale of graduate education in such countries, as well as ongoing concerns regarding excessive expansion (and high public costs) of the tertiary education sector (e.g., Becker, 1960). Nonetheless, the selective coverage of past studies leaves

Table 1: Previous studies of expectational errors

Reference	Baseline sample	Expectation metric	Panel?	Outcome metric	Error
Klößner and Pfeifer (2019)	Higher education students, Germany	First salary graduates in same field	No	First salary of recent graduates	-18%
Avitabile and De Hoyos (2018)	Secondary school students, Mexico	Wage of people aged 30-40 years	No	Observed wages in population	+33%
Vasilescu and Begu (2018)	Unemployed Young people between 15-29 years old, Romania	Reservation wages	No	Observed wages in population	+30%
Frick and Maihaus (2016)	Higher education students, Germany	First job salary	No	First salary from early graduated students	+17%
Abbiati and Barone (2017)	Secondary school students, Italy	Salary after graduation	No	Observed wages in population	+32%
Reuben et al. (2017)	Undergraduates, USA	Income at age 30 and 45	No	Observed wages in population at age 30	+36%
Huntington-Klein (2015)	High school junior and senior, USA	Income at age 30	No	Observed wages in population at age 30	+40%
Alonso-Borrego and Romero-Medina (2016)	Junior university students, Spain	Salary after graduation	No	Wages of graduates aged 25 -29 years	+27%
Wiswall and Zafar (2015)	Undergraduate students, USA	Wages of people aged 30 years	No	Observed wages in population	+9%
Jerrim (2015)	Males aged 20, USA	Income at age 30	Yes	Wage income at age 30 predicted from wages observed at age 23-26	+40%
Menon et al. (2012)	Undergraduates, Cyprus	First job salary after graduation	No	Wages of recent graduates	+8%

Table 1: Previous studies of expectational errors

Reference	Baseline sample	Expectation metric	Panel?	Outcome metric	Error
Jerrim (2011)	Undergraduates, UK	First job salary	No	First salary from early graduated students	+17%
Van der Merwe (2011)	First year students, South Africa	First salary on graduation	No	Observed wages in population	∼0%
Van der Merwe (2009)	First year students, South Africa	Salary in first job after graduation	Yes	Observed wage 1 year after baseline	+62%
Rouse (2004)	High school seniors, low income USA	Income at age 30	No	Observed wages in population at age 25-30	+100%
Webbink and Hartog (2004)	University and Higher vocational students, Netherlands	Net starting salary after graduation	Yes	Observed wage 4 years after baseline	∼0%
Orazem et al. (2003)	Senior university students, USA	Salary in first job after graduation	No	Observed wages in population	+4%
Wolter (2000)	High school & University students, Switzerland	Median wage of people aged 30-40	No	Median wage of people aged 30-40	-5%
Carvajal et al. (2000)	Senior college students, USA	First job salary after graduation	No	Wages of recent graduates	+8.4%
Betts (1996)	Undergraduates, USA	Starting salary after graduation	No	Wages of recent graduates	-6%
Smith and Powell (1990)	Final year undergrad1uates, USA	Income in first year of job & after 10 years	No	Wages of graduates at age 18-24 and 30-24	+17%

Source: own elaboration.

open whether similar errors are found in other countries, namely those with small(er) cohorts of university graduates and/or those with very different labour market conditions, such as most developing economies. Third, most previous studies estimate the gap between expected and realized wages using different cross-sectional samples. Longitudinal studies of school-to-work transitions are surprisingly limited in scope, again especially outside of advanced countries. Necessarily, the absence of panel data limits the kind of analysis that can be undertaken; and in most studies expectational errors are thus only estimated, not observed directly.

Fourth, the studies in Table 1 show substantial variation in expectational errors, even within the same country. Nonetheless, what accounts for this is not entirely clear. On the one hand, there is evidence of a range of psychological or preference effects, such as associated with gender (Dawson, 2017; Wiswall and Zafar, 2017; Caliendo et al., 2017), as well as mistakes (especially among younger students) in predicting the final level of education at which they will enter the labour market (e.g., Jerrim, 2011). On the other hand, the literature also hints that students from deprived backgrounds, including those from outside more advanced industrialized nations, tend to make comparatively larger errors in a positive direction (Rouse, 2004; Van der Merwe, 2009; Vasilescu and Begu, 2018). This interpretation is supported by a range of unpublished studies in a broad(er) range of contexts. For instance, in Mozambique, Mendola and Minale (2018) find expectational errors of between 70 and 130 percent among students in two of the major universities. In this sense, variation in the degree of prior exposure of participants to the labour market seems relevant. These findings have inspired important work on *how* individuals use information to form and update their expectations (e.g., Böheim et al., 2011; Wiswall and Zafar, 2015). These studies suggest that expectational errors stem from a combination of incomplete (inaccurate) information, as well as a degree of over-confidence in own abilities (see Clark and Friesen, 2008).

A potential separate explanation for systematic gaps between expected and realized wage outcomes concerns difficulties in obtaining the *type* of job that was anticipated when wage expectations were elicited. That is, alongside (unexpected) unemployment, individuals may accept job offers in organizations or roles that they had not originally desired. Studies of this phenomena, which generally have not explicitly connected to the literature on expectational errors, point to various forms of mismatch (for recent surveys see McGuinness et al., 2018; Somers et al., 2019). These include: *vertical mismatch*, where the individuals' level of education does not meet the formal requirements of the job position; and *horizontal mismatch*, where the employees' area of study (degree) does not correspond to the field of the job position. To these we might add completion or *graduation mismatch*, which refers to cases where individuals begin work without having fully completed (attained) the final level of education they had earlier anticipated.

Studies of various forms of mismatch and their implications have primarily considered experiences in high income countries, particularly those that have witnessed significant expansion in access to higher education, as well as contexts with comparatively high rates of youth unemployment. For instance, in their handbook chapter on over-education, Leuven and Oosterbeek (2011) survey a large number (N=151) of empirical studies of this form of vertical mismatch, of which zero refer to the African continent and just 18 to Asia. Even so, a consistent finding is that mismatches are often associated with substantial earnings penalties versus the counterfactual of being correctly matched. Indeed, among the studies surveyed by Leuven and Oosterbeek (2011), the average penalty associated with being over-educated for one's work position equals around half of the coefficient associated with the required or correct years of schooling (also Dolton and Silles, 2008; Li et al., 2018; Caroleo and Pastore, 2018).

Existing literature related to what we term graduation mismatch has mostly focussed on the determinants and implications of not completing post-secondary education (e.g., Manski, 1989; Light and Strayer, 2000). However, a small group of studies consider the more specific problem of delayed graduation, which occurs where individuals prolong the length of their studies (far) beyond the minimum course duration. As Aina et al. (2011) document, this is a serious problem in some countries and appears to be closely associated with labour market conditions. The notion is that where (graduate) positions are scarce, individuals are willing to 'queue' for these posts while prolonging their studies, sometimes also undertaking occasional paid work to make ends meet. This also has material implications – e.g., in Italy, Aina et al. (2012) estimate that delayed graduation is associated with an earnings penalty equal in value to 7% of the median wage.

3 Analytical framework

The previous section pointed to (often positive) systematic errors in wage expectations across multiple studies. However, existing evidence is somewhat limited – few studies are based on panel studies, and the vast majority refer to experiences from high income countries. Furthermore, the sources of expectational errors are not well established. Alongside informational errors, unanticipated wage penalties driven by job mismatches of either a permanent or temporary nature are highly plausible. However, this source of expectational error has not been widely studied, neither in high income contexts nor elsewhere.

In this section, we develop a framework to distinguish between various forms of expectational error. Our analytical starting point is a conventional Mincerian (hedonic) function for wages of the following form:

$$\ln w_{it} = \mu + \phi_i + \delta t + Z_{it}\beta + H_{it}\gamma + \varepsilon_{it} \tag{1}$$

where w is the wage of individual i observed at time t, Z is a vector of observed individual characteristics, and H is a vector of occupational characteristics. While this type of specification hardly needs further discussion, it is important to note that it is relevant for *both* actual and expected wages. In the latter case, subjective expectations are almost always elicited on a conditional basis, taking into consideration both expected (desired) personal characteristics and a particular job profile. In other words, the point prediction for an individual's future wage can be thought of as a rich conditional expectation: $w_i^e = \mathrm{E}(w_{it} \mid \mu = \mu^e, \phi_i = \phi^e, t = t^e, Z = Z^e, H = H^e)$, where the e superscript refers to a subjective expectation over future wage conditions. This formulation is an elaboration of the framework suggested by Dominitz (1998), who noted future wage expectations are not only individual-specific, but they are also conditional on achieving some form of employment.

In our baseline survey, described further below, we collected information on subjective expectations concerning each individuals' own future wage income in first employment, the desired type of employment, and individual characteristics. Using these as the conditions on which wage expectations are formed, the future wage expected to be realized at time t^e is given from equation (1) as:[1]

$$\ln w_i^e = \mu^e + \phi_i^e + \delta^e t_i^e + Z_i^e \beta^e + H_i^e \gamma^e + \varepsilon_i^e \tag{2}$$

This clarifies that the wage expectation is based not only on the individual's information set regarding rewards in the labour market (e.g., $\mu^e, \phi_i^e, \delta^e, \beta^e, \gamma^e$) but it is also conditional on specific outcomes being realized, such as achieving employment at a specific period (as captured by t^e), and in a given sector.

On entering the labour market, a similar expression can be applied. Here a proportion of

[1] We drop the time subscript hereafter as we focus on just two observation periods per individual, both of which are explicit in the notation.

individuals accept employment offers and in turn report data on their wage income, as well as the characteristics of the job. In keeping with the discussion in the previous section, neither the wage nor the type of job necessarily matches with previously-stated expectations. Thus, the realized counterpart to equation (1) at time t^r is now:

$$\ln w_i^r = \mu^r + \phi_i^r + \delta^r t_i^r + Z_i{}^r \beta^r + H_i^r \gamma^r + \varepsilon_i^r \tag{3}$$

and where r refers to labour market realizations in the first paid position reported following the baseline survey (when students were in university).[2] Using these two expressions, the gap between expected and realized (log.) wage income is then simply:

$$\begin{aligned} \ln w_i^e - \ln w_i^r = {} & (\mu^e - \mu^r) + (\phi_i^e - \phi_i^r) + (t_i^e \delta^e - t_i^r \delta^r) \\ & + (Z_i^e \beta^e - Z_i^e \beta^r) + (H_i^e \gamma^e - H_i^r \gamma^r) + (\varepsilon_i^e - \varepsilon_i^r) \end{aligned}$$

So, errors can come from two sources: misinformation regarding labour market rewards (e.g., $\gamma^e \neq \gamma^r$); and/or labour market mismatches (e.g., $H^e \neq H^r$). To isolate these two sources of error more clearly, the same gap can be equivalently restated as follows:

$$\begin{aligned} \ln w_i^e - \ln w_i^r = {} & (\mu^e - \mu^r) + (\phi_i^e - \phi_i^r) \\ & + t_i^e(\delta^e - \delta^r) + (t_i^e - t_i^r)\delta^r \\ & + Z_i^e(\beta^e - \beta^r) + (Z_i^e - Z_i^r)\beta^r \\ & + H_i^e(\gamma^e - \gamma^r) + (H_i^e - H_i^r)\gamma^r \\ & + (\varepsilon_i^e - \varepsilon_i^r) \end{aligned} \tag{5}$$

Or, in more compact form we have:

$$\ln w_i^e - \ln w_i^r \equiv \Delta \ln w_{it} = e_i^I + e_i^M + \Delta\varepsilon_{it}$$

$$e_i^I = \Delta\mu + \Delta\phi_i + t_i^e \Delta\delta + Z_i^e \Delta\beta + H_i^e \Delta\gamma \tag{6a}$$

$$e_i^M = \Delta t_i \delta^r + \Delta Z_i \beta^r + \Delta H_i \gamma^r \tag{6b}$$

This states that the total error is constituted by three components: information errors, due to differences between actual and realized labour market rewards (e^I); matching errors, due to differences between actual and realized job outcomes (e^M); and a residual. This specification represents a generalization of the decomposition in Webbink and Hartog (2004) who implicitly

[2] Note that t^r varies across individuals as they start work at different points in time.

assume that all labour market expectations other than wages are satisfied in reality. Or, in the language of the above model, they impose the assumption: $e_i^M = 0 \Leftrightarrow (t^e = t^r) \wedge (Z^e = Z^r) \wedge (H^e = H^r)$, which thus allows them to regress the expectational error on the baseline characteristics alone.

Estimation of the above model(s) can be undertaken using conventional techniques, including both linear (OLS) and non-linear methods (quantile regressions). While the objective of the analysis is to provide a rich description of systematic associations in the data – i.e., it is a diagnostic exercise rather than a causal analysis – the presence of omitted variables does remain a challenge. These could bias the coefficient estimates and thereby muddle the distinction between different sources of expectational errors. To address these, we combine two approaches. First, we rely on an extensive range of control variables, collected at the individual level and including proxies for both academic and cognitive ability, as well as background family characteristics. In addition, we attempt to correct for bias in terms of who is selected into employment. To do so, we evaluate the probability of obtaining a job based on initial characteristics and job preferences and then use the generalized residual from this procedure as a additional control in the subsequent wage regressions, as per Heckman (1979). Further details regardin the data and methods are set out in Sections 4 and 5, respectively.

4 Background and data

As a low-income country, Mozambique broadly fits a set of stylized facts regarding the nature of its economy, population and, at the intersection of these two, its labour market. As described in Jones and Tarp (2016a) (also Lachler, 2018), economic activities are predominantly informal and more than 70% of the economically active population, aged 15 or older, occupied in farming activities. Mozambique also has a young and fast-growing population. The most recent 2017 population census counted 27.9 million Mozambicans, growing at an average rate of 3.1% per annum from 20.6 million in 2007, of which 46.6% were less than 14 years old (INE, 2019). As with many other (low-income) countries, Mozambique has also witnessed very rapid growth in access to education at all levels over recent decades, albeit from a low base (see Jones et al., 2018b). With respect to higher education, the number of students graduating from university (across the country) rose from under 700 in 2003 to over 18,000 in 2016 (Jones et al., 2018a), implying an annual growth rate of around 30%. The increasing number of graduates is adding to what nonetheless remains a small group of Mozambicans with higher education, reported to represent just 1.2% of the country's population in 2015 (INE 2015a).

Despite the scarcity of highly skilled personnel, new graduates face what can be politely described as 'challenging' labour market conditions. As indicated, the formal sector remains small – e.g., less than 12% of all workers report receiving a wage and the proportion of wage earners in the urban working population has increased only slowly over time (Jones and Tarp, 2016a,b). Furthermore, competition for jobs is extremely high. More than 300,000 young people enter the job market each year, while opportunities for non-agricultural employment remain thin and found largely in the (informal) services sector. Production sectors that have grown most rapidly over recent years include extractive industries, which are capital intensive and have often relied on foreign workers to fill key technical positions. Put differently, there is little evidence to suggest rapid or sustained growth in demand for workers with a university education.

In view of these labour market challenges, as well as the lack of data regarding labour market conditions, we undertook a representative survey of students in their final year of studies across the six largest public and private universities in the country.[3] The baseline survey was conducted in 2017 and collected data on personal characteristics, educational and professional histories, cognitive abilities and labour market expectations.[4] Starting from early 2018, *after* their studies should have been completed, we re-contacted the same individuals four times by telephone over the subsequent year (i.e., quarterly) and when most had entered the labour market. On each occasion we collected data on their employment situation, including realized wages, type of work undertaken and employment outlook.

Our baseline sample contained 2,176 students and is representative of the population of Mozambican university graduates by university, gender and study area (*viz.*, Education, Humanities, Social Sciences and Law, Natural Sciences, Engineering, Agriculture and Health).[5] A descriptive report of the baseline survey, including further details on the design and implementation, can be found in Jones et al. (2018a). In the total baseline sample, 1,989 participants provided subjective estimates of their expected pre-tax wage after graduating.[6] Of these, we were subsequently able to trace 1,887 individuals at least once, representing an attrition rate of just 5.1%.

[3] Mozambique has no regular labour market survey; and the last household (budget) survey was undertaken in 2014/15.

[4] For details see Jones et al. (2018a).

[5] Sample weights based on the survey are employed throughout. In the presentation of results we do not report results for specific universities. This is to maintain anonymity and was a requirement to gain permission to proceed with the study.

[6] A total of 186 students (110 male and 76 female) from the baseline survey, including the sample design, reported they did not intend to look for a job, and, therefore, did not state an expected wage.

Table 2 reports baseline characteristics for the core sample used herein, split between those that did and did not obtain a paid position during the follow-up year. We note that 1,187 participants reported undertaking some paid work during the follow-up period and a total of 700 participants did not report any paid work in the follow-up period. Of these, 83% (N=587) originally expressed an interest in seeking work after completing their studies, implying 40% of those interested in seeking work were not able to find any work (or changed their mind). We also note that survey participants who did find work tended to be significantly older (2.5 years), more likely to be men, married and with children. Students of (lower cost) public universities are relatively more represented among those that found a job, as are those that studied in the field of Education, while students of Social Sciences are relatively over-represented among those that did not find work.

The same table reveals that the majority of university graduates in Mozambique study education (principally, to be teachers) or social sciences. Around 15% studied courses in natural sciences, engineering or health (medicine). In terms of stated (desired) job expectations, employment in the private sector dominates (45%), followed by the public sector (33%) and then self-employment (16%). The expected starting salary was just over US$ 400, which far exceeds the average monthly cost of education and suggests a very high expected premium from education (see below).[7] Expected salaries were slightly higher among those that actually obtained work. Although, we find no significant difference in the distributions of expected starting salaries between those that did and did not obtain obtain work (see Appendix Figure B1), the overall differences in profiles between these groups raises the possibility of (unobserved) selection bias.

Employment outcomes for the first paid position reported in the follow-up period are summarised in Table 3. In terms of employment organizations, these outcomes would seem to match desired outcomes quite closely (e.g., 46% work in the private sector vs. 45% in the baseline expectation). However, in line with the discussion of Section 2, a closer look reveals a series of what we describe as job mismatches, which refer to domains in which their realized first job does not align with earlier expectations.[8] As shown in the table, a large proportion of individuals stated they: had not yet formally completed their studies (83%); were working in positions outside their field of studies (54%); were working as interns (12%) or on a part-time basis (44%); did not have a fixed/permanent contract (72%); were actively looking for another job (65%); were not

[7] For ease of interpretation, all wage values are stated in constant prices (January 2017 = 1) and, where relevant, converted to US$ at an exchange rate of 60 Meticais to 1 US$.

[8] These mismatches follow directly from the research design and baseline questionnaire. Indeed, wage expectations were *explicitly* elicited on the assumption the individual had completed their studies and they had also obtained the desired type of employer and work sector.

Table 2: Descriptive statistics from baseline survey (2017)

	Obtained work in follow-up period?					
	No		Yes		All	
Individual characteristics:						
Age	24.42	(0.17)	26.93	(0.20)	26.05	(0.14)
Female	0.60	(0.02)	0.36	(0.01)	0.44	(0.01)
Married	0.09	(0.01)	0.18	(0.01)	0.14	(0.01)
Has kids	0.20	(0.02)	0.37	(0.01)	0.31	(0.01)
University attended:						
Public university	0.71	(0.02)	0.85	(0.01)	0.80	(0.01)
Total cost USD/month	73.68	(2.34)	62.34	(1.49)	66.31	(1.28)
Course of study:						
Education	0.24	(0.02)	0.36	(0.01)	0.32	(0.01)
Humanities	0.01	(0.00)	0.02	(0.00)	0.02	(0.00)
Social Sciences	0.51	(0.02)	0.40	(0.01)	0.44	(0.01)
Natural Sciences	0.04	(0.01)	0.04	(0.01)	0.04	(0.00)
Engineering	0.07	(0.01)	0.08	(0.01)	0.07	(0.01)
Agriculture	0.05	(0.01)	0.06	(0.01)	0.05	(0.01)
Health	0.07	(0.01)	0.06	(0.01)	0.06	(0.01)
Job expectations:						
Will seek work	0.83	(0.01)	0.75	(0.01)	0.77	(0.01)
Private sector employee	0.35	(0.02)	0.32	(0.01)	0.33	(0.01)
Public employee	0.44	(0.02)	0.46	(0.01)	0.45	(0.01)
NGO employee	0.05	(0.01)	0.05	(0.01)	0.05	(0.01)
Self employed	0.16	(0.01)	0.16	(0.01)	0.16	(0.01)
Wage (USD/month)	403.24	(7.72)	426.59	(5.91)	418.42	(4.70)
Observations	700		1,187		1,887	

Notes: cells are variable means calculated applying survey weights, with standard errors in parentheses; univerity names are abbreviated.
Source: own estimates.

Table 3: Realized outcomes in first labour market position (N = 1,187)

	Private uni.		Public uni.		
	Male	Female	Male	Female	All
Private sector employee	0.57	0.62	0.42	0.46	0.46
Public employee	0.21	0.12	0.27	0.33	0.27
NGO employee	0.07	0.04	0.09	0.05	0.07
Self employed	0.11	0.16	0.19	0.14	0.17
Study unfinished	0.79	0.78	0.86	0.82	0.83
Job unlike course	0.55	0.63	0.50	0.57	0.54
Intern position	0.13	0.18	0.11	0.11	0.12
Works part time	0.43	0.38	0.48	0.38	0.44
No fixed contract	0.73	0.66	0.74	0.71	0.72
Searching for work	0.69	0.63	0.68	0.58	0.65
Employee mismatch	0.62	0.67	0.68	0.60	0.65
Sector mismatch	0.41	0.39	0.52	0.44	0.48
Mismatch count	3.88	4.02	4.09	3.79	3.98
Realized wage (USD/month)	226.23	196.23	149.29	138.52	155.71
Expected - realized wage (USD)	255.31	228.51	294.16	240.33	270.88
Expectational error (log.)	0.94	0.92	1.30	1.16	1.20

Notes: ....
Source: own estimates.

working for the type of organization stated in the baseline (65%); and were not working in the sector identified in the baseline (48%). Each of these eight types of mismatch, which cover both vertical and horizontal dimensions, are operationalized as dummy variables; on average, the individual-specific sum of mismatches is close to four. This suggests first jobs do not generally match with original expectations.

The bottom part of Table 3 compared realised wages to earlier expectations. The gap is large and positive – on average, individuals in their first paid position after university earn US$ 156 per month, which is US$ 271 below what had been expected. Transformed into natural logarithms, the expectational error, defined as expected minus realized wage (consistent with the literature), equals 1.2 points on average. These distributions are illustrated in Figure 1, where plot (a) is the cross-sectional distributions of expected and realized wages, while plot (b) is the individual-specific differences (in US$). At an individual level, it becomes clear that fewer than 10% of the respondents who obtained a job received a wage that equalled or exceeded their earlier expectations; and close to 80% were being paid at least US$ 100 *less* than expected per month. Overall, this would confirm that university graduates face a tough jobs market, at least compared to their expectations in their final year of studies. In light of earlier studies, while the presence of a positive expectational error is not so surprising, the magnitude of the error in this case seems very large.

5 Results

5.1 Wage determinants

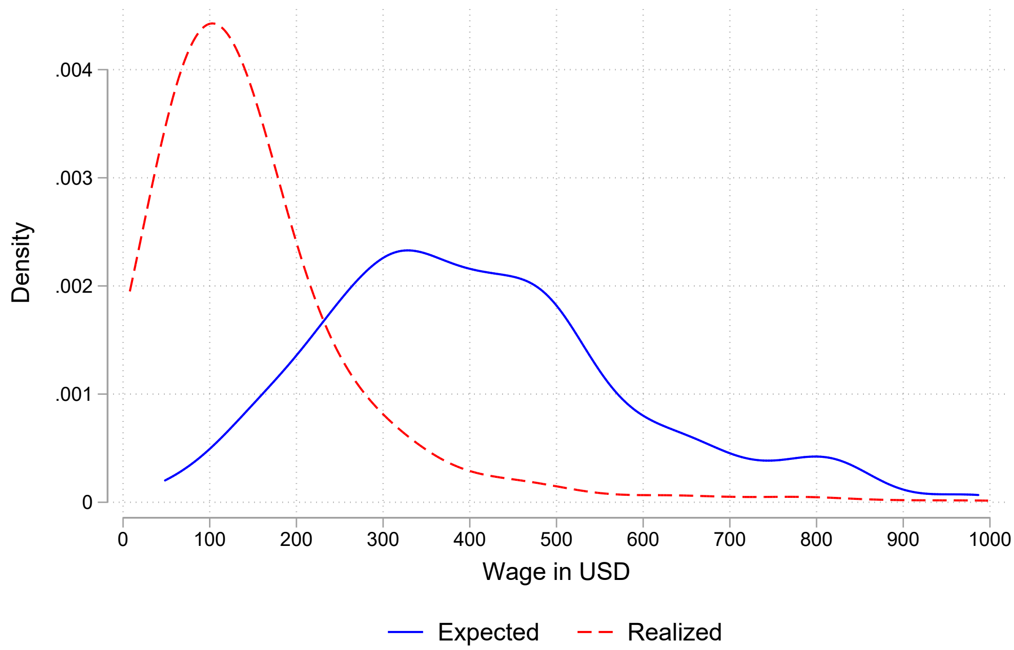
We now move to the formal analysis of expectational errors. To begin, we first consider the determinants of obtaining a paid position in the post-baseline follow-up period and, thus, who subsequently reports a non-zero realized wage. Column (I) of Table 4 summarises estimates from a linear probability model where the dependent variable takes a value of one if the participant reported having a paid job post-baseline, using only baseline individual and future (desired) job characteristics as explanatory variables, including each participants' stated interest in seeking work in the future.[9] These results reveal some important variations by individual characteristics, particularly that females were significantly less likely to find work, and (less surprisingly) that those with greater previous work experience were more likely to report being employed during the follow-up period. At the same time, specific university and expected job characteristics generally provided little predictive guidance as to who reports a first wage. This is despite the fact that around 40% of the baseline sample did not report any paid work in the 12 month follow-up period; and of these, over 80% had originally affirmed an interest in seeking work (see Table 2).

Columns (IIa) and (IIb) regress the natural logarithm of participants' expected first wage against the same baseline characteristics, as per equation (2). The only difference between these estimates is that (IIa) refers to the full sample (N = 1,887), while column (IIb) only contains the sub-sample for which we have a subsequent wage realization (N = 1,187). Comparing the estimated coefficients, we observe only minor differences, implying that the degree of bias from unobservables may not be so large (see also Section 4). Finally, column (IIc) adds the standardized generalized residual from a probit model on the form of column (I) to the sub-sample model. This term is statistically significant at the 5% level and, when included,

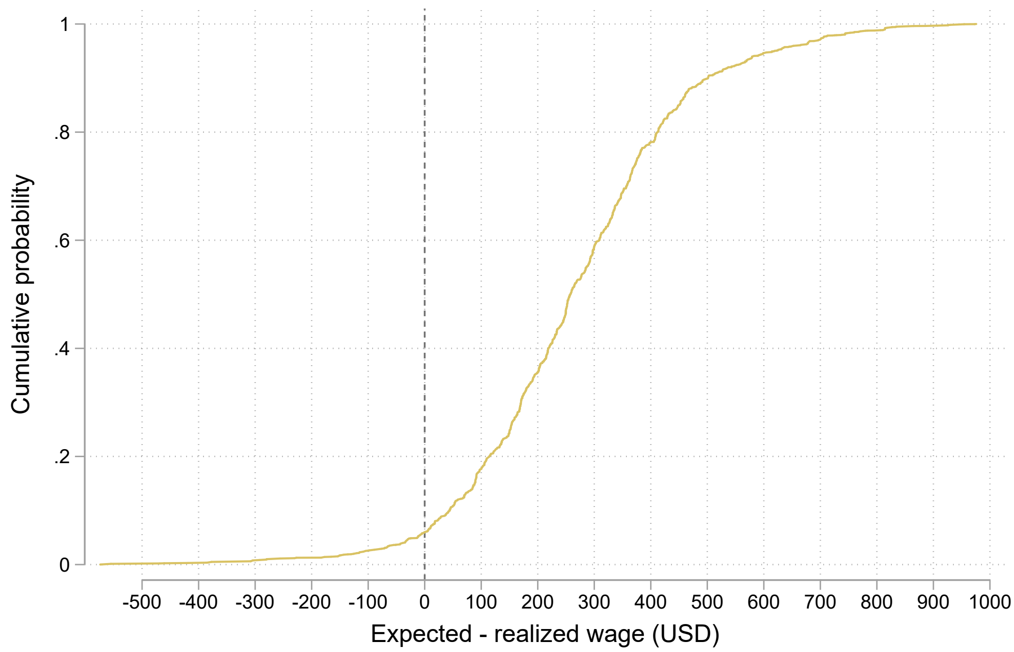
[9] Throughout, the reference group for the regressions are students who attended education at the Universidade de Eduardo Mondlane (UEM) and expected to enter the private sector. Only selected coefficients are shown. Full results are available on request.

Figure 1: Expected vs. realized wages

(a) Cross-sectional distributions



(b) Individual-level differences



Source: own calculations.

Table 4: Linear regression estimates of job expectations and outcomes

	(I) Job?	(II) Expected wage			(III) Realized wage			
	(.)	(a)	(b)	(c)	(a)	(b)	(c)	(d)
Constant	0.60***	10.17***	10.15***	10.22***	8.88***	8.88***	9.43***	9.45***
	(0.10)	(0.13)	(0.16)	(0.17)	(0.19)	(0.20)	(0.18)	(0.19)
Age	-0.00	-0.00	-0.00	-0.00	0.03***	0.03***	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Female	-0.17***	-0.11***	-0.12***	-0.04	0.07	0.06	-0.02	0.03
	(0.03)	(0.03)	(0.04)	(0.06)	(0.07)	(0.10)	(0.06)	(0.10)
Married	-0.02	0.06	0.06	0.07	0.19***	0.19***	0.15***	0.15***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
Prev. work exp.	0.03***	0.02***	0.01**	-0.00	-0.02**	-0.02	-0.02**	-0.02
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Public university	0.13***	-0.04	-0.02	-0.09	-0.39***	-0.39***	-0.36***	-0.39***
	(0.04)	(0.05)	(0.05)	(0.07)	(0.09)	(0.11)	(0.08)	(0.11)
Natural Sciences	-0.05	0.17**	0.18***	0.20***	0.27***	0.27***	0.37***	0.38***
	(0.05)	(0.07)	(0.06)	(0.06)	(0.10)	(0.10)	(0.08)	(0.08)
Engineering	0.03	0.31***	0.28**	0.26**	0.41**	0.41**	0.53***	0.52***
	(0.06)	(0.09)	(0.11)	(0.11)	(0.20)	(0.20)	(0.16)	(0.16)
Agriculture	0.05	0.01	0.00	-0.02	-0.04	-0.04	0.14**	0.13*
	(0.06)	(0.05)	(0.06)	(0.06)	(0.09)	(0.09)	(0.07)	(0.07)
Health	0.04	0.34***	0.31***	0.29***	0.11	0.11	0.11	0.10
	(0.09)	(0.07)	(0.06)	(0.07)	(0.14)	(0.14)	(0.09)	(0.09)
English proficiency	0.07**	-0.04	-0.07	-0.10*	0.16**	0.17**	0.21***	0.19**
	(0.03)	(0.05)	(0.06)	(0.06)	(0.07)	(0.08)	(0.07)	(0.08)
Future wage growth	-0.06	-0.51***	-0.48***	-0.45***	-0.01	-0.01	0.04	0.06
	(0.04)	(0.05)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Public employee	-0.01	-0.06*	-0.07	-0.06	-0.24***	-0.24***	-0.06	-0.05
	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)	(0.07)	(0.07)
Self employed	0.02	0.02	0.07	0.06	-0.04	-0.04	-0.39***	-0.39***
	(0.04)	(0.03)	(0.04)	(0.04)	(0.07)	(0.07)	(0.07)	(0.07)
Nonselection hazard				-0.10*		0.00		-0.05
				(0.05)		(0.08)		(0.08)
Study unfinished							-0.25***	-0.25***
							(0.08)	(0.08)
Works part time							-0.33***	-0.33***
							(0.06)	(0.06)
Intern position							-0.34***	-0.34***
							(0.06)	(0.06)
Searching for work							-0.10**	-0.10**
							(0.04)	(0.04)
Job unlike course							-0.17***	-0.17***
							(0.05)	(0.05)
Obs.	1,887	1,887	1,187	1,187	1,187	1,187	1,187	1,187
R^2	0.20	0.22	0.22	0.22	0.23	0.23	0.34	0.34
Actual outcomes?	No	No	No	No	No	No	Yes	Yes

Notes: in column (I) the dependent variable is whether the individual obtained a job; in columns (II) and (III) the dependent variable is the natural log. of expected and realized wages, respectively; sample in columns II(b)-III(d) are only those that obtained a job; selected coefficients shown; in columns III(c) and III(d) all job outcomes are as realized, else they are as expected; non-selection hazard included in columns II(c), III(b) and III(d); cluster-robust standard errors in parentheses.
Source: own estimates.

results in a shrinkage of the coefficient on being female toward zero while other controls remain largely stable. One interpretation is that unobserved factors associated with gender influence *both* expected wages and the likelihood of obtaining a paid job. As a consequence, accounting for selection effects appears to be material.

The results in column (IIc) are informative. Consistent with results elsewhere [ref??], factors that appear material to obtaining a job are not the same as the determinants of expected wages. In the latter case, the participants' university, course of study and preferred job characteristics become important. For instance, students in engineering, health and natural sciences all expect higher starting salaries than those studying educating (the reference); and participants expect to obtain lower salaries in the public sector relative to the private sector or self-employment. Thus, in line with our analytical framework, this supports the idea that individuals form wage expectations on the basis of both already-given and expected job conditions. It also suggests that individuals expect the labour market to reward different individual and job types differently – i.e., they have some knowledge.

The remaining columns of Table 4 (IIIa-IIId) shift the focus to realized wages in the first job observed in the follow-up period. Columns (IIIa) and (IIIb) replicate the specifications of columns (IIb) and (IIc), using only baseline characteristics as explanatory variables. While the generalized residual no longer appears significant, other coefficients do alter and in some cases even switch sign. For instance, while natural science and engineering graduates correctly expected to obtain a wage premium relative to other courses, the premium actually obtained by these same students appears larger than expected, while the discount associated with public sector work appears more severe than expected (-0.23 log points versus -0.08). The problem with interpretations of the latter sort is that not all students who had expressed a desire to work in the public sector subsequently did so. Thus, the public sector coefficient in column (IIIb) refers to the discount associated with the baseline expectations, not the actual job outcome. To clarify this distinction, columns (IIIc) and (IIId) replace these expected labour market job characteristics (sector and employer) with their realized counterparts, now as per equation (3). In addition, we add controls for the range of mismatches that represent constant conditions under which wage expectations were elicited (see Section 4). The key result is that the new specification makes a difference – compared to the model using baseline controls only, various coefficients shift substantially in magnitude (c.f., columns IIIa and IIIb) and the added mismatch variables are not only statistically significant but also represent material discounts to individual wages. For example, failing to complete all courses (a type of vertical mismatch) is associated with a discount of around a third on realized wages; and having a job outside the field of study (horizontal mismatch) is associated with a 16% wage discount.

5.2 Error decomposition

Table 5 provides the formal decomposition results, applying the specification set out in equation (6a) and assuming individual characteristics, aside from age, remain fixed over time $Z^e = Z^r$. Columns (I) and (II) refer to alternative estimators, where the former is (sample weighted) OLS and the latter is the iteratively reweighted least squares (IRWLS) proposed by Huber (1973). Sub-columns (a) regress the expectation error on the set of baseline characteristics only, assuming all match variables are zero (as in Webbink and Hartog, 2004); sub-columns (b) relax the aforementioned restriction, representing the full specification; and sub-columns (c) add the generalized residual from the selection model (Table 4, column I). As expected, the former results represent the difference in parameters estimated in columns (IIa) and (IIIa) of Table 4; and the remaining results summarise the contributions of information and matching errors, representing differences in parameters and variable values respectively.

Three principal findings merit note. First, as before, the complete specification adds significant explanatory value. Explicitly accounting for labour market mismatches improves the overall fit of the model by around 50%, increasing the R^2 from 0.18 to 0.29 (see columns IIa versus Ia). Furthermore, the complete model modifies the interpretation of some explanatory variables. In particular, while the expected returns to self-employment were not different to those in the private sector in the partial model (columns Ia and IIa), the complete model suggests that are excessively optimistic (by at least 0.30 log points); also, once mismatches are taken into account, the constant term (which corresponds to the informational error for the reference group) drop considerably. In light of this, the second observation is that matching errors account for approximately half of the total observed expectational error. This is indicated in the bottom of the table, which reports sample averages (and 95% confidence intervals) for the two predicted component errors. These are calculated directly from each regression model, applying a simple shrinkage factor to adjust for parameter uncertainty. Concretely, the average information error is calculated from the vector of individual-level errors of this form:

$$\hat{e}_i^I = \sum_{x \,\in\, \mu, t^e, Z^e, H^e} x_i \times \hat{\theta}_x \times [1 - \Pr(\hat{\theta}_x = 0)] \tag{7}$$

which is the sum of the products of each relevant baseline variable (including the constant term, μ), the associated regression coefficient from the decomposition model, and its shrinkage factor.

Table 5: Regression estimates of expectational error

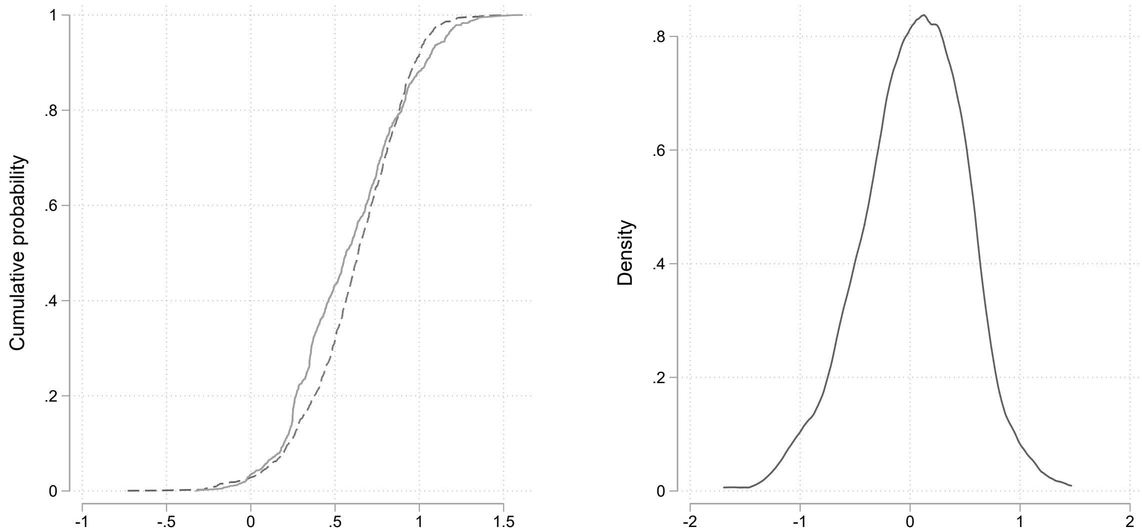
	(I) OLS			(II) Robust [M-estimator]		
	(a)	(b)	(c)	(a)	(b)	(c)
Constant	1.27***	0.52*	0.57**	1.34***	0.64**	0.69**
	(0.20)	(0.26)	(0.28)	(0.25)	(0.29)	(0.29)
Age	-0.03***	-0.02***	-0.02***	-0.03***	-0.02***	-0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female	-0.19**	-0.12	-0.05	-0.16**	-0.10	-0.03
	(0.08)	(0.07)	(0.10)	(0.06)	(0.06)	(0.10)
Married	-0.13*	-0.09	-0.08	-0.15*	-0.09	-0.09
	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)
Prev. work exp.	0.03***	0.03***	0.02	0.04***	0.03***	0.02
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Public university	0.37***	0.33***	0.28***	0.34***	0.29***	0.23**
	(0.08)	(0.07)	(0.10)	(0.08)	(0.07)	(0.10)
Health	0.20*	0.21**	0.19*	0.20	0.21	0.18
	(0.12)	(0.09)	(0.10)	(0.13)	(0.13)	(0.13)
English proficiency	-0.23***	-0.24***	-0.27***	-0.22***	-0.23***	-0.26***
	(0.09)	(0.09)	(0.10)	(0.07)	(0.07)	(0.08)
Future wage growth	-0.48***	-0.53***	-0.51***	-0.54***	-0.58***	-0.55***
	(0.10)	(0.10)	(0.10)	(0.08)	(0.08)	(0.08)
Self employed	0.11	0.44***	0.43***	0.05	0.31***	0.30***
	(0.08)	(0.11)	(0.11)	(0.07)	(0.09)	(0.09)
Study unfinished (Δ)		-0.22***	-0.22***		-0.24***	-0.24***
		(0.08)	(0.08)		(0.07)	(0.07)
Works part time (Δ)		-0.33***	-0.33***		-0.35***	-0.35***
		(0.07)	(0.07)		(0.07)	(0.07)
Intern position (Δ)		-0.35***	-0.36***		-0.39***	-0.40***
		(0.09)	(0.09)		(0.08)	(0.08)
Searching for work (Δ)		-0.10*	-0.10*		-0.10**	-0.10**
		(0.05)	(0.05)		(0.05)	(0.05)
Job unlike course (Δ)		-0.15***	-0.15***		-0.19***	-0.19***
		(0.05)	(0.05)		(0.05)	(0.05)
NGO employee (Δ)		0.21**	0.21**		0.25***	0.26***
		(0.09)	(0.09)		(0.09)	(0.09)
Self employed (Δ)		-0.38***	-0.38***		-0.29***	-0.29***
		(0.07)	(0.07)		(0.07)	(0.07)
Nonselection hazard			-0.08			-0.08
			(0.09)			(0.09)
Obs.	1,187	1,187	1,187	1,187	1,187	1,187
R^2	0.17	0.28	0.28	0.19	0.31	0.31
Courses	0.15	0.13	0.23	0.47	0.25	0.27
Info. error	1.26	0.60	0.61	1.27	0.63	0.62
	[1.0,1.5]	[0.2,1.0]	[0.2,1.0]	[0.9,1.6]	[0.2,1.0]	[0.2,1.0]
Matching error	.	0.55	0.55	.	0.59	0.58
	-	[0.4,0.8]	[0.4,0.7]	-	[0.4,0.8]	[0.4,0.8]

Notes: dependent variable is the log. difference between expected and real wages (reported in real terms); selected coefficients shown; columns I(a) and II(a) refer only to baseline characteristics, remaining column add differences (Δ) between expected and realized outcomes; non-selection hazard included in columns I(c) and II(c); cluster-robust standard errors in parentheses.

Source: own estimates.

Figure 2: Empirical distributions of component errors

(a) Cumulative distributions of component errors (b) Unit-wise difference in component errors



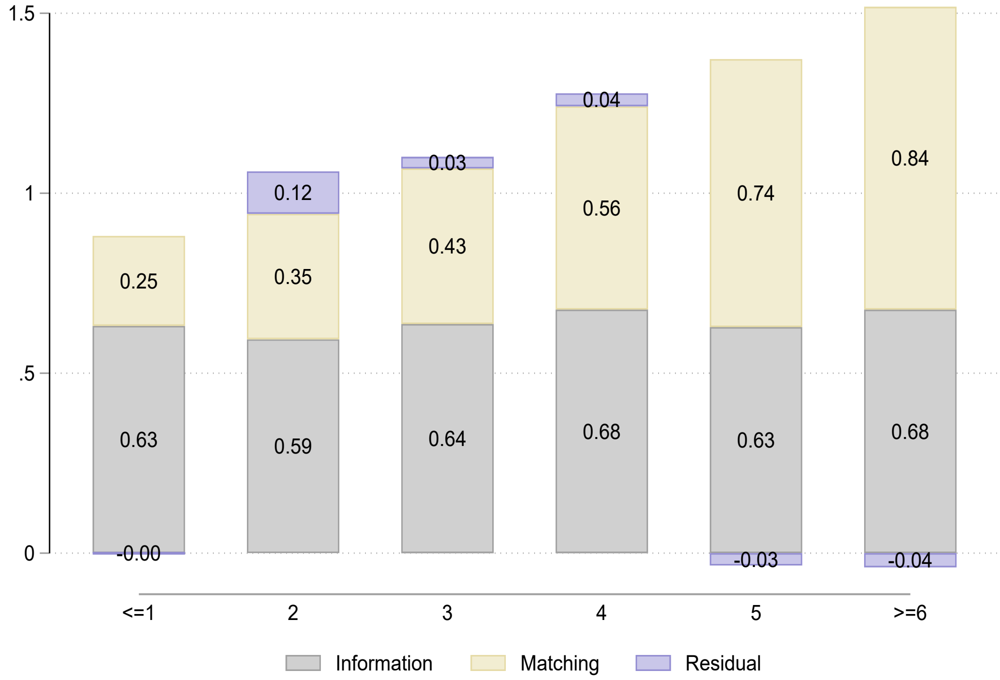
Notes: in figure (a), the dashed line (- - -) is the informational error, and the solid line (—) is the matching error; figure (b) illustrates the distribution of differences between the information and matching errors.
Source: own calculations.

Similarly, the mean matching error is calculated from the difference variables, as follows:

$$\hat{e}_i^M = \sum_{x \,\in\, \Delta t, \Delta Z, \Delta H} x_i \times \hat{\theta}_x \times [1 - \Pr(\hat{\theta}_x = 0)] \tag{8}$$

Figure 2(a) illustrates the cumulative empirical distributions of these two error components, and Figure 2(b) plots the distribution of their unit-wise differences. This serves to confirm the critical contribution of job mismatches to expectational errors that emerge as university students enter the labour market. Another way of seeing this is revealed in Figure 3, which shows the respective contributions of the two component errors (plus the residual) from low to high degrees of *ex post* mismatch. As discussed in Section 4 (Table 3), the latter metric is constructed by summing across a full range of observed mismatch dummy variables, ranging from zero to eight and taking a median of four. Notably, while the informational error is broadly stable across different degrees of mismatch, at the highest levels the magnitude of the mismatch component is very substantial, accounting for expectational errors equivalent to around 150% of the realized starting salary (0.91 log points).

Figure 3: Component errors, by mismatch degree



Notes: horizontal scale is defined from the individual count of labour market mismatches, where a higher value implies a lower job match quality; error components are as described in the text, estimated from the results in Table 5 column IIc.
Source: own calculations.

Third, alongside mismatch errors the informational errors are also large in absolute magnitude, especially in comparison with the overall size of errors observed elsewhere (Table 1). On average, these errors equate to around 80% of the realized starting salary (0.6 log points), which is more than twice the optimism bias found in studies across high income countries. The bulk of these errors are accounted for by the constant term, which aggregates features of the reference group. As such it cannot be interpreted precisely; nonetheless, it is highly suggestive of very high (average) over-estimates of the generic premium to higher education, regardless of individual over-confidence. This view is supported by the study of Mendola and Minale (2018), also undertaken in two of the major universities in Mozambique. At the same time, there is important variation in the size/direction of informational errors across the sample. As shown in Table 6, participants who had previously-attending private universities display substantially lower informational errors on average (see also Figure B2); similarly, older students show both lower matching and informational errors, in line with greater experience of the labour market

Table 6: Error decomposition, by groups

Group	Value	Obs.	Error	Decomposition Info.	Match	Residual
Female	No	734	1.25	0.66	0.60	0.03
	Yes	453	1.10	0.60	0.54	-0.06
Older	No	550	1.30	0.70	0.68	0.01
	Yes	637	1.13	0.60	0.48	0.02
Public uni.	No	205	0.96	0.43	0.58	0.02
	Yes	982	1.23	0.67	0.57	0.01
Mismatch	≤ 1	77	0.95	0.63	0.25	-0.00
	2	137	0.98	0.59	0.35	0.12
	3	197	1.04	0.64	0.43	0.03
	4	281	1.23	0.68	0.56	0.04
	5	280	1.36	0.63	0.74	-0.03
	6	215	1.45	0.68	0.84	-0.04
All		1,187	1.19	0.64	0.57	0.01

Notes: older is above median age for the sample who had obtained a job; mismatch is an ordinal score based on the sum of eight underlying dummy variables.
Source: own estimates.

and contact with other workers.

5.3 Validation

To validate and further probe the previous results we pursue two complementary strategies. First, we review how the determinants of expectational errors vary across the outcome (error) distribution. Using a quantile regression model, Table 7 replicates the full specification in column I(c) of Table 5, estimated from the 5th to 95th percentiles. While the broad pattern of estimates are broadly stable across percentiles, a few insights do stand out. In particular, variation in the predicted (total) informational errors appears to lie behind the variation in the total error – i.e., up to around the 35th percentile, informational errors are less than or not different to zero, but rise steeply as the total error increases. In contrast, matching errors are consistently positive and large throughout the overall error distribution. We also note that the contribution of self-employment varies considerably – at upper percentiles, it is clear that individuals in self-employment has significantly over-estimated the returns to this activity. These results are consistent with being pushed into (occasional) self-employment in the absence of other work.

As a second strategy, we re-run the earlier error decompositions, now constructing the expectational error from the highest wage reported by each individual over the follow-up period as opposed to the first non-zero observation. While this diverges somewhat from a strict interpretation of the original wage expectation, it nonetheless allows for the fact that some of the first reported salaries may have reflected temporary mismatches, such as a short probationary or internship period, or reduced hours to allow participants to complete their studies. This is also relevant, of the 1,187 participants reporting a first wage, just under 50% later report a higher one. The decomposition associated with the latter are reported in Table A1, which replicates Table 5. The parameter estimates do not differ considerably between the two tables; however, the estimated mean matching error has shrunk, consistent with the notion that mismatches are not necessarily permanent – i.e., there is a degree of queuing for the 'right' or 'good' jobs, which was not originally anticipated. The same view is confirmed when we focus on the (smaller) group who report a salary for a full-time job (Table A2).

[Add comparison to IOF?]

6 Implications for educational returns

Before concluding, we reflect on what these expectational errors mean for returns to education.[10] The presence of systematic, large positive gaps between expected and realized wages that emerge on entering the labour market suggests that (early) pecuniary returns to higher education are also lower than anticipated. Considering most university students in Mozambique pay relatively significant fees and other expenses to pursue their studies, especially in private universities, this is a potential concern.

To calculate educational returns using the present data we adopt the simple shortcut formula [ref??], which focuses on the marginal benefit of higher education, given by:

$$b_i = \frac{w_i - \tilde{w}_i}{N(\tilde{w}_i + c_i)} \tag{9}$$

where w_i is the post-university wage (for student i), $\tilde{w}_i$ is the counterfactual wage that would be

[10] We continue to focus here on the sub-group that obtained a job. Obviously, returns to education are negative for those that have yet to work.

Table 7: Quantile regression estimates of expectational error

	Percentile						
	5	20	35	50	65	80	95
Constant	0.82	0.27	0.50	0.48	0.52	0.77*	0.80
	(0.60)	(0.44)	(0.39)	(0.36)	(0.36)	(0.41)	(0.51)
Age	-0.01	-0.02***	-0.03***	-0.02***	-0.02***	-0.03***	-0.02**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female	0.01	0.04	-0.12	-0.15	-0.21	-0.11	-0.35*
	(0.20)	(0.15)	(0.12)	(0.12)	(0.13)	(0.16)	(0.18)
Married	0.10	-0.12	-0.13	-0.05	-0.08	-0.10	-0.22**
	(0.14)	(0.10)	(0.09)	(0.09)	(0.09)	(0.10)	(0.11)
Prev. work exp.	0.00	0.01	0.03	0.05**	0.04**	0.03	0.03
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Public university	0.16	0.21	0.26**	0.33**	0.37***	0.28*	0.26
	(0.20)	(0.14)	(0.13)	(0.13)	(0.14)	(0.16)	(0.17)
Health	-0.57	0.29	0.23	0.26	0.23	0.38	0.48
	(0.36)	(0.20)	(0.20)	(0.17)	(0.17)	(0.23)	(0.29)
English proficiency	-0.28	-0.24*	-0.22**	-0.15	-0.18*	-0.33***	-0.43***
	(0.20)	(0.13)	(0.11)	(0.10)	(0.10)	(0.12)	(0.13)
Future wage growth	-0.52***	-0.49***	-0.51***	-0.46***	-0.55***	-0.56***	-0.45**
	(0.17)	(0.15)	(0.11)	(0.11)	(0.12)	(0.14)	(0.17)
Self employed	0.16	0.29	0.31**	0.27**	0.44***	0.61***	0.81***
	(0.23)	(0.19)	(0.14)	(0.13)	(0.13)	(0.18)	(0.18)
Study unfinished (Δ)	-0.06	-0.24**	-0.24**	-0.21**	-0.24***	-0.29***	-0.13
	(0.14)	(0.11)	(0.10)	(0.09)	(0.08)	(0.11)	(0.16)
Works part time (Δ)	-0.22	-0.22**	-0.35***	-0.31***	-0.30***	-0.42***	-0.56***
	(0.16)	(0.10)	(0.09)	(0.09)	(0.10)	(0.12)	(0.13)
Intern position (Δ)	-0.32	-0.34***	-0.38***	-0.35***	-0.23*	-0.46***	-0.50***
	(0.20)	(0.13)	(0.12)	(0.12)	(0.12)	(0.16)	(0.17)
Searching for work (Δ)	-0.27**	-0.14*	-0.06	-0.13**	-0.10	-0.08	-0.05
	(0.11)	(0.08)	(0.07)	(0.07)	(0.06)	(0.07)	(0.09)
Job unlike course (Δ)	-0.14	-0.11	-0.10	-0.13*	-0.19***	-0.18**	-0.12
	(0.10)	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.09)
NGO employee (Δ)	0.11	0.23	0.24*	0.26**	0.24*	0.19	0.21
	(0.18)	(0.15)	(0.13)	(0.13)	(0.13)	(0.16)	(0.17)
Self employed (Δ)	-0.22	-0.32**	-0.23**	-0.26**	-0.37***	-0.45***	-0.76***
	(0.20)	(0.14)	(0.11)	(0.10)	(0.11)	(0.15)	(0.16)
Nonselection hazard	-0.25	-0.21	-0.09	-0.01	0.06	0.05	0.12
	(0.15)	(0.13)	(0.11)	(0.11)	(0.12)	(0.13)	(0.17)
Obs.	1,187	1,187	1,187	1,187	1,187	1,187	1,187
Courses	0.03	0.19	0.73	0.64	0.78	0.69	0.03
Error at percentile	-0.16	0.49	0.87	1.15	1.43	1.85	2.65
Info. error	-0.21	0.09	0.42	0.63	0.70	1.03	1.67
	[-1.0,0.6]	[-0.5,0.7]	[-0.1,1.0]	[0.1,1.1]	[0.2,1.2]	[0.5,1.5]	[1.1,2.3]
Matching error	0.20	0.50	0.47	0.53	0.67	0.63	0.45
	[-0.1,0.5]	[0.2,0.8]	[0.2,0.7]	[0.3,0.8]	[0.4,1.0]	[0.4,0.9]	[0.2,0.7]

Notes: dependent variable is the log. difference between expected and real wages (reported in real terms); selected coefficients shown; columns I(a) and II(a) refer only to baseline characteristics, remaining column add differences (Δ) between expected and realized outcomes; non-selection hazard included in columns I(c) and II(c); robust standard errors in parentheses.
Source: own estimates.

received without having higher education, c_i is the cost and N is the length of the course. Thus, equation (9) is simply the ratio of the marginal benefit to the total cost. With the exception of the counterfactual wage, all the necessary elements to make this calculation were obtained from our survey. For the former, we use nationally-representative household survey data to estimate the expected wage for the survey participants under the assumption they had only completed high school (but not further education). This approach suggests an average gap relative to the (first) realised wage of: $\ln w_i - \ln \tilde{w}_i = 0.55$, which is almost identical to the 0.6 log point conditional premium to higher education obtained from the same household survey data.

Using these individual-level counterfactuals, Table 8 applies equation (9) and reports group-wise medians of three rates of return: (a) the return from the realized first salary; (b) the return from the realized first salary plus the estimated discount due to mismatch effects (i.e., what would be realized were the mismatches to disappear); and (c) the expected return, based on the anticipated wage elicited in the baseline survey. The distributions of these three returns are also illustrated in Figure 4. As might be anticipated, we find a substantial gap between actual and expected returns. While the midpoint of the latter is almost 50%, the former is just 4%. Nonetheless, correcting for job mismatches – e.g., under the assumption that they represent largely temporary phenomena – the actual return would increase to 24%, which is well in line with international levels including those in developing countries (ref??).

Perhaps of greater interest is the variation in returns. First, despite their generally higher cost, we find private universities offer higher returns (columns a and b). This is likely to explained by the different career choices of these graduates. In particular, many of those pursing education courses go on to be teachers, where salaries (in the public sector) are comparatively low inducing the lowest actual return, even after for mismatch adjustment (12%). And it is public universities that provide the bulk of education graduates. We also note that returns achieved by engineering graduates are substantially higher than all other courses of study (25% in column a, increasing to 70% in column b). Although these returns are still lower than expected at baseline, they are indicative of strong demand for these skills in the economy.

7 Conclusion

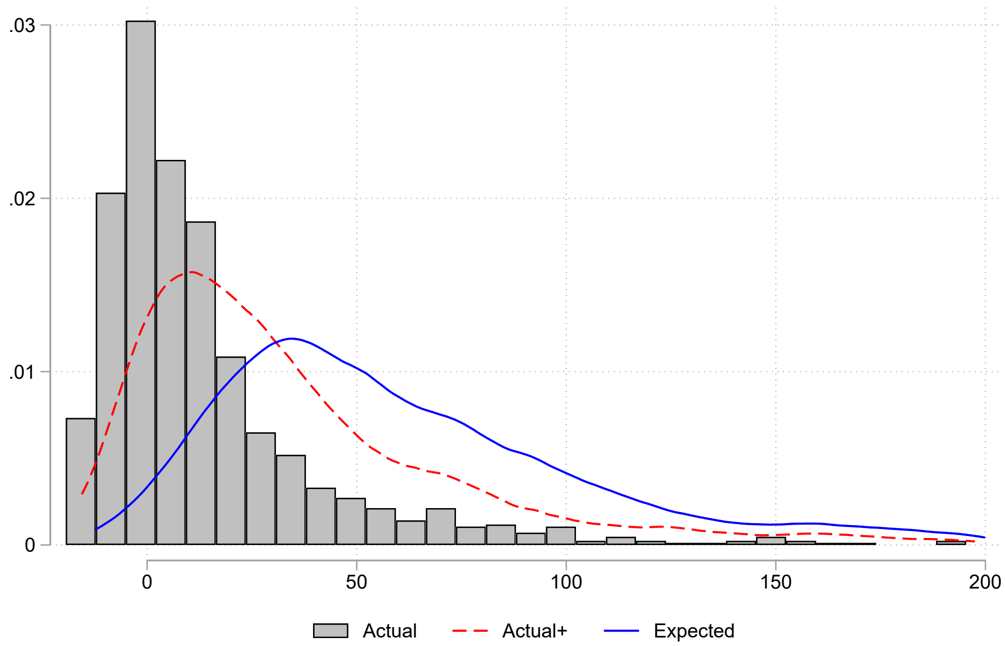
[To be written]

Table 8: Estimates of educational returns (in %), group medians

			Estimated returns			
			Actual	Actual+	Exp.	Δ
		Obs.	(a)	(b)	(c)	(c)-(b)
Public uni.	No	205	12.1	25.2	42.1	15.3
	Yes	982	1.9	20.5	52.9	28.7
Course	Agriculture	67	1.9	23.3	66.5	43.3
	Education	297	-1.8	10.3	38.9	27.7
	Engineering	115	24.7	67.0	99.2	37.9
	Health	75	14.6	31.5	76.1	35.7
	Humanities	63	6.9	28.5	57.7	19.4
	Natural Sciences	176	10.7	34.9	67.5	26.4
	Social Sciences	394	6.2	25.2	50.2	21.9
All		1,187	4.0	22.1	49.9	26.4

Notes: ... .
Source: own estimates.

Figure 4: Realized vs. expected returns



Source: own calculations.

References

Abbiati, G. and Barone, C. (2017). Is university education worth the investment? the expectations of upper secondary school seniors and the role of family background. *Rationality and Society*, 29(2):113–159.

Aina, C., Baici, E. and Casalone, G. (2011). Time to degree: students' abilities, university characteristics or something else? Evidence from Italy. *Education Economics*, 19(3):311–325.

Aina, C., Pastore, F. et al. (2012). Delayed graduation and overeducation: A test of the human capital model versus the screening hypothesis. Discussion Paper 6413, Institute for the Study of Labor (IZA).

Alonso-Borrego, C. and Romero-Medina, A. (2016). Wage Expectations for Higher Education Students in Spain. *Labour*, 30(1):1–17.

Avitabile, C. and De Hoyos, R. (2018). The heterogeneous effect of information on student performance: evidence from a randomized control trial in Mexico. *Journal of Development Economics*, 135:318–348.

Becker, G.S. (1960). Underinvestment in college education? *American Economic Review*, 50(2):346–354. URL http://www.jstor.org/stable/1815038.

—— (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5):9–49. URL http://www.jstor.org/stable/1829103.

Betts, J.R. (1996). What do students know about wages? evidence from a survey of undergraduates. *Journal of human resources*, pp. 27–56.

Böheim, R., Horvath, G.T. and Winter-Ebmer, R. (2011). Great expectations: past wages and unemployment durations. *Labour economics*, 18(6):778–785.

Caliendo, M., Lee, W.S. and Mahlstedt, R. (2017). The gender wage gap and the role of reservation wages: New evidence for unemployed workers. *Journal of Economic Behavior & Organization*, 136:161–173.

Caroleo, F.E. and Pastore, F. (2018). Overeducation at a glance. determinants and wage effects of the educational mismatch based on AlmaLaurea data. *Social Indicators Research*, 137(3):999–1032.

Carvajal, M.J., Bendana, D., Bozorgmanesh, A., Castillo, M.A., Pourmasiha, K., Rao, P. and Torres, J.A. (2000). Inter-gender differentials between college students' earnings expectations and the experience of recent graduates. *Economics of Education Review*, 19(3):229–243.

Clark, J. and Friesen, L. (2008). Overconfidence in forecasts of own performance: An experimental study. *The Economic Journal*, 119(534):229–251.

Dawson, C. (2017). The upside of pessimism: Biased beliefs and the paradox of the contented female worker. *Journal of Economic Behavior & Organization*, 135:215–228.

Dolton, P.J. and Silles, M.A. (2008). The effects of over-education on earnings in the graduate labour market. *Economics of Education Review*, 27(2):125–139.

Dominitz, J. (1998). Earnings expectations, revisions, and realizations. *Review of Economics and statistics*, 80(3):374–388.

Filmer, D. and Fox, L. (2014). Youth employment in Sub-Saharan Africa. Technical report, World Bank Group. doi:https://doi.org/10.1596/978-1-4648-0107-5.

Frick, B. and Maihaus, M. (2016). The structure and determinants of expected and actual starting salaries of higher education students in Germany: identical or different? *Education Economics*, 24(4):374–392.

Heckman, J.J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 1(5):799–821.

Huntington-Klein, N. (2015). Subjective and projected returns to education. *Journal of Economic Behavior & Organization*, 117:10–25.

INE (2019). IV Recenseamento geral da população e habitação 2017: Resultados definitivos – Moçambique. Technical report, Instituto Nacional de Estatistica.

Jerrim, J. (2011). Do UK higher education students overestimate their starting salary? *Fiscal Studies*, 32(4):483–509.

—— (2015). Do college students make better predictions of their future income than young adults in the labor force? *Education Economics*, 23(2):162–179.

Jones, S., Mambo, F., Mazive, E., Paris, Y., Santos, R. and Xirinda, G. (2018a). Baseline survey on university students' school-to-employment transition. Technical report, UNU-WIDER. URL https://www.wider.unu.edu/publication/baseline-survey-school-work-transitions-university-graduates-mozambique

Jones, S., Sohnesen, T.P. and Trifkovic, N. (2018b). The evolution of private returns to education during post-conflict transformation: Evidence from Mozambique. WIDER Working Paper 2018/143, UNU-WIDER.

Jones, S. and Tarp, F. (2016a). Mozambique: Growth experience through an employment lens. In F. Tarp and H. Bhorat (Eds.), *Africa's Lions: Growth Traps and Opportunities for Six African Economies*, chapter 5, pp. 145–180. Washington, DC: Brookings Institution Press.

—— (2016b). Mozambique: Jobs and welfare in an agrarian economy. In G. Betcherman and M. Rama (Eds.), *Jobs for Development: Challenges and Solutions in Different Country Settings*, chapter 2, pp. 20–61. Oxford University Press.

Klößner, S. and Pfeifer, G. (2019). The importance of tax adjustments when evaluating wage expectations. *The Scandinavian Journal of Economics*, 121(2):578–605.

Lachler, Ulrich; Walker, D.I. (2018). Mozambique – jobs diagnostic. Job Series 13, World Bank Group. URL http://documents.worldbank.org/curated/en/655951534181476346/Mozambique-jobs-diagnostic.

Leuven, E. and Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. In E.A. Hanushek, S. Machin and L. Woessmann (Eds.), *Handbook of the Economics of Education*, volume 4, chapter 3, pp. 283–326. Elsevier. URL https://EconPapers.repec.org/RePEc:eee:educhp:4-283.

Li, I., Harris, M. and Sloane, P.J. (2018). Vertical, horizontal and residual skills mismatch in the Australian graduate labour market. *Economic Record*, 94(306):301–315.

Light, A. and Strayer, W. (2000). Determinants of college completion: School quality or student ability? *Journal of Human Resources*, pp. 299–332.

Manski, C.F. (1989). Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon. *Economics of Education review*, 8(4):305–312.

McGuinness, S., Pouliakas, K. and Redmond, P. (2018). Skills mismatch: concepts, measurement and policy approaches. *Journal of Economic Surveys*, 32(4):985–1015.

Mendola, M. and Minale, L. (2018). Subjective expectations and occupational choices in low-income countries: Evidence from an information experiment in Mozambique. Working Paper F-36401MOZ-1, International Growth Center.

Menon, M.E., Pashourtidou, N., Polycarpou, A. and Pashardes, P. (2012). Students' expectations about earnings and employment and the experience of recent university graduates: Evidence from Cyprus. *International Journal of Educational Development*, 32(6):805–813.

Van der Merwe, A. (2009). A comparison of the labour market expectations and labour market experiences of new graduates. *South African Journal of Higher Education*, 23(2):398–417.

—— (2011). Earnings expectations of typical South African university of technology first-year students. *Education Economics*, 19(2):181–198.

Orazem, P.F., Werbel, J.D. and McElroy, J.C. (2003). Market expectations, job search, and gender differences in starting pay. *Journal of Labor Research*, 24(2):307–321.

Reuben, E., Wiswall, M. and Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, 127(604):2153–2186.

Rouse, C.E. (2004). Low-income students and college attendance: An exploration of income expectations. *Social Science Quarterly*, 85(5):1299–1317.

Sam, V. et al. (2018). Unemployment duration and educational mismatches: An empirical investigation among graduates in Cambodia. *Economics Bulletin*, 38(3):1554–1565.

Schweri, J. and Hartog, J. (2017). Do wage expectations predict college enrollment? evidence from healthcare. *Journal of Economic Behavior & Organization*, 141:135–150.

Smith, H.L. and Powell, B. (1990). Great expectations: Variations in income expectations among college seniors. *Sociology of Education*, pp. 194–207.

Somers, M.A., Cabus, S.J., Groot, W. and van den Brink, H.M. (2019). Horizontal mismatch between employment and field of education: evidence from a systematic literature review. *Journal of Economic Surveys*, 33(2):567–603.

Vasilescu, M.D. and Begu, L.S. (2018). Youth reservation wages in Romania: are wage expectations realistic? *Applied Economics Letters*, pp. 1–5.

Webbink, D. and Hartog, J. (2004). Can students predict starting salaries? Yes! *Economics of Education Review*, 23(2):103–113.

Wiswall, M. and Zafar, B. (2015). How do college students respond to public information about earnings? *Journal of Human Capital*, 9(2):117–169.

—— (2017). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507.

Wolter, S.C. (2000). Wage expectations: a comparison of Swiss and US students. *Kyklos*, 53(1):51–69.

A Appendix: Additional tables

Table A1: Regression estimates of expectational error, highest salary

	(I) OLS			(II) Robust [M-estimator]		
	(a)	(b)	(c)	(a)	(b)	(c)
Constant	1.30***	1.08***	1.13***	1.38***	1.05***	1.08***
	(0.21)	(0.21)	(0.23)	(0.25)	(0.27)	(0.28)
Age	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female	-0.08	-0.05	0.02	-0.06	-0.03	0.01
	(0.08)	(0.07)	(0.11)	(0.06)	(0.06)	(0.10)
Married	-0.09	-0.07	-0.06	-0.13	-0.08	-0.08
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Prev. work exp.	0.02**	0.03***	0.02	0.03***	0.03***	0.02
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Public university	0.26***	0.26***	0.21**	0.22***	0.23***	0.20**
	(0.08)	(0.08)	(0.09)	(0.08)	(0.07)	(0.10)
Health	-0.16	-0.03	-0.05	-0.15	0.00	-0.01
	(0.16)	(0.12)	(0.12)	(0.13)	(0.13)	(0.13)
English proficiency	-0.27***	-0.24***	-0.27***	-0.27***	-0.26***	-0.28***
	(0.09)	(0.08)	(0.10)	(0.07)	(0.07)	(0.08)
Future wage growth	-0.50***	-0.54***	-0.52***	-0.53***	-0.56***	-0.54***
	(0.10)	(0.09)	(0.10)	(0.08)	(0.08)	(0.08)
Self employed	0.10	0.39***	0.39***	0.01	0.21**	0.21**
	(0.08)	(0.11)	(0.11)	(0.07)	(0.09)	(0.09)
Study unfinished (Δ)		-0.22***	-0.22***		-0.25***	-0.25***
		(0.06)	(0.06)		(0.05)	(0.05)
Works part time (Δ)		-0.24***	-0.24***		-0.30***	-0.30***
		(0.07)	(0.07)		(0.06)	(0.06)
Intern position (Δ)		-0.40***	-0.40***		-0.38***	-0.38***
		(0.09)	(0.09)		(0.08)	(0.08)
Searching for work (Δ)		-0.18***	-0.18***		-0.16***	-0.16***
		(0.05)	(0.05)		(0.05)	(0.05)
Job unlike course (Δ)		-0.10	-0.10		-0.15***	-0.16***
		(0.06)	(0.06)		(0.06)	(0.06)
NGO employee (Δ)		0.17	0.17		0.23**	0.23**
		(0.10)	(0.10)		(0.09)	(0.09)
Self employed (Δ)		-0.31***	-0.31***		-0.22***	-0.22***
		(0.09)	(0.09)		(0.07)	(0.07)
Nonselection hazard			-0.08			-0.05
			(0.10)			(0.09)
Obs.	1,187	1,187	1,187	1,187	1,187	1,187
R^2	0.17	0.27	0.27	0.17	0.30	0.30
Courses	0.18	0.29	0.35	0.49	0.43	0.42
Info. error	1.04	0.78	0.75	1.04	0.75	0.73
	[0.9,1.2]	[0.5,1.1]	[0.5,1.0]	[0.8,1.3]	[0.4,1.1]	[0.4,1.0]
Matching error	.	0.30	0.30	.	0.34	0.33
	-	[0.1,0.5]	[0.1,0.5]	-	[0.1,0.5]	[0.1,0.5]

Notes: dependent variable is the log. difference between expected and real wages (reported in real terms); selected coefficients shown; columns I(a) and II(a) refer only to baseline characteristics, remaining column add differences (Δ) between expected and realized outcomes; non-selection hazard included in columns I(c) and II(c); cluster-robust standard errors in parentheses.

Source: own estimates.

Table A2: Regression estimates of expectational error, full time workers

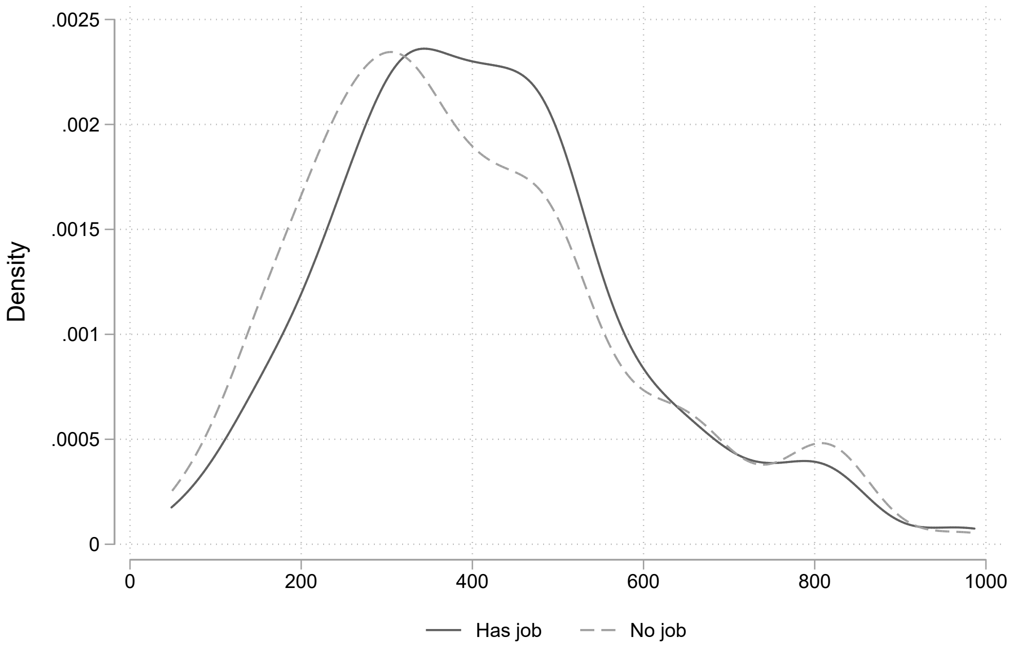
	(I) OLS			(II) Robust [M-estimator]		
	(a)	(b)	(c)	(a)	(b)	(c)
Constant	0.72***	0.65**	0.76**	0.91***	0.74**	0.82**
	(0.26)	(0.32)	(0.38)	(0.30)	(0.33)	(0.37)
Age	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female	-0.11	-0.11	-0.07	-0.08	-0.07	-0.05
	(0.08)	(0.08)	(0.09)	(0.07)	(0.07)	(0.08)
Married	-0.03	0.00	0.00	-0.06	-0.03	-0.03
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Prev. work exp.	0.03***	0.03***	0.02	0.03***	0.03***	0.02
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
Public university	0.07	0.10	0.08	0.06	0.06	0.05
	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.09)
Health	0.01	0.09	0.07	0.02	0.07	0.06
	(0.13)	(0.11)	(0.12)	(0.14)	(0.15)	(0.15)
English proficiency	-0.18*	-0.19**	-0.24*	-0.18**	-0.19**	-0.23**
	(0.10)	(0.09)	(0.13)	(0.08)	(0.08)	(0.11)
Future wage growth	-0.43***	-0.48***	-0.48***	-0.45***	-0.50***	-0.50***
	(0.10)	(0.10)	(0.10)	(0.09)	(0.09)	(0.09)
Self employed	0.03	0.41***	0.42***	0.02	0.37***	0.38***
	(0.08)	(0.15)	(0.15)	(0.08)	(0.12)	(0.12)
Study unfinished (Δ)		-0.27***	-0.27***		-0.24***	-0.24***
		(0.08)	(0.08)		(0.07)	(0.07)
Intern position (Δ)		-0.38***	-0.39***		-0.40***	-0.41***
		(0.09)	(0.09)		(0.07)	(0.07)
Searching for work (Δ)		-0.18***	-0.18***		-0.16***	-0.16***
		(0.05)	(0.05)		(0.05)	(0.05)
Job unlike course (Δ)		-0.09	-0.09		-0.14**	-0.14**
		(0.06)	(0.06)		(0.06)	(0.06)
NGO employee (Δ)		0.17*	0.17*		0.13	0.13
		(0.09)	(0.09)		(0.09)	(0.09)
Self employed (Δ)		-0.37**	-0.37**		-0.33***	-0.33***
		(0.15)	(0.15)		(0.10)	(0.10)
Nonselection hazard			-0.10			-0.06
			(0.14)			(0.13)
Obs.	794	794	794	794	794	794
R^2	0.14	0.23	0.23	0.15	0.24	0.24
Courses	0.55	0.20	0.20	0.85	0.61	0.58
Info. error	0.95	0.67	0.70	0.91	0.66	0.70
	[0.7,1.2]	[0.2,1.1]	[0.2,1.2]	[0.6,1.2]	[0.2,1.1]	[0.2,1.2]
Matching error	.	0.34	0.34	.	0.34	0.34
	-	[0.1,0.6]	[0.1,0.6]	-	[0.2,0.5]	[0.2,0.5]

Notes: dependent variable is the log. difference between expected and real wages (reported in real terms); selected coefficients shown; columns I(a) and II(a) refer only to baseline characteristics, remaining column add differences (Δ) between expected and realized outcomes; non-selection hazard included in columns I(c) and II(c); cluster-robust standard errors in parentheses.

Source: own estimates.

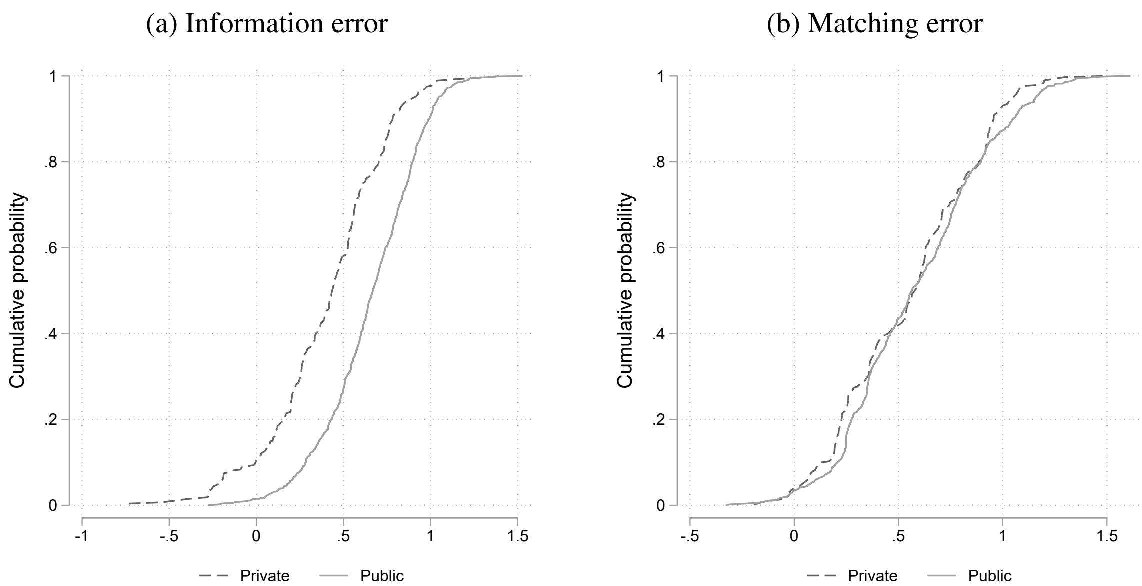
B Appendix: Additional figures

Figure B1: Expected wages, by later job outcomes



Source: own calculations.

Figure B2: Component errors, by university type



Source: own calculations.