Misinformed, mismatched or misled? Explaining the gap between expected and realised graduate earnings in Mozambique

Sam Jones, Ricardo Santos, and Gimelgo Xirinda

UNU-WIDER, Mozambique

July 10, 2020

Abstract

Inaccurate expectations of future wages are almost ubiquitous. Yet, little is known about the sources of these errors, particularly outside high income countries. Based on a longitudinal survey of labour market transitions of graduates in Mozambique, this study provides a new decomposition of the gap between expected and realized earnings. We find this gap is extremely large (>100%), but is not driven by incorrect information about returns in different sectors or individual characteristics. Rather, job mismatches of various kinds account for over a third of the total error and the remaining error reflects a cognitive bias associated with misleading reference points (high-performing peers). Although this indicates a need for greater transparency regarding levels of remuneration, we find no evidence that optimistic expectations are associated with poorer labour market outcomes.

Acknowledgements: The study 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, as well as the active collaboration of the University of Eduardo Mondlane, Catholic University of Mozambique, Pedagogic University, Polytechnic University, St. Thomas Aquinas University and UniZambeze. We acknowledge the invaluable research assistance of Felix Mambo, Edson Mazive, Thomas Sohnesen, Vincenzo Salvucci and Yonesse Paris. We further thank conference and seminar participants at UEM (Maputo), IESE (Maputo), UNU-WIDER (Helsinki), and the UNU-WIDER annual conference in Bangkok 2019 for useful feedback. All errors and omissions are our own.

Funding: This study was undertaken 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.

1 Introduction

Since the seminal works of Schultz (1961) and Becker (1962), investments in education have been routinely conceptualized as forward-looking in nature, with expected future earnings being a cornerstone of choices. In this view, systematic biases in expected returns would typically lead to sub-optimal investment decisions (e.g., over-education). Expectations of future wages are also likely to inform decisions about job search (Becker, 1962), including whether to accept a particular job offer or remain in an existing job. So, unless wage expectations are correct, individuals may reject job offers they mistakenly consider to be low, or accept job positions for which they are overqualified.

Despite the critical role attributed to (wage) expectations across educational and labour market behaviours, much remains to be understood about their nature. Almost thirty years ago, Manski (1993) posed the need to deepen our understanding of expectations; but, looking over a wide range of studies, Behrman (2010) noted that research on expected labour market outcomes remains limited. That said, various studies have documented material differences between self-reported wage expectations and either econometric salary estimates (from market surveys) or later realizations. Focussing on university students or recent graduates, just a handful of papers find expectations are close to market rates (Webbink and Hartog, 2004; Van der Merwe, 2011) or below (Wolter, 2000; Klößner and Pfeifer, 2019). In contrast, the majority of studies encounter optimistic expectations (e.g., Jerrim, 2011, 2015; Wiswall and Zafar, 2015; Abbiati and Barone, 2017). Systematic errors in wage expectations also are not only encountered among graduates or school-leavers. For example, Hoxhaj (2015) finds that illegal migrants into Italy overestimate wages by over 80 percent, and this bias only increases with the size of their social network in other destinations.

At the same time, little is known about why these expectational errors persist. One explanation is that prospective workers are poorly informed about the distribution of wages across different occupations and, thus, mis-estimate differences in returns to specific occupations or individual characteristics (e.g., language skills; gender). This explanation is certainly plausible in low income countries, where labour market information and exposure tends to be scarce. Not only are such markets often thin, reflecting both their relative size and segmented nature (Hino and Ranis, 2014; Basu et al., 2019), but also many individuals simply do not have personal connections into the formal labour market (via family or friends) from which they might obtain credible information.

A separate literature indicates labour market frictions may create a gap between expected and realized wages. Evidence from high income countries, for instance, 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 characteristics of a realized job position do not match earlier expectations, this would imply realized wages fall below expectations. In developing countries, this kind of error also 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., Al-Samarrai and Bennell, 2007; Filmer and Fox, 2014). And while the specific issue of job mismatches *within* the formal sector has not received so much attention outside of high income economies, it stands to reason this phenomenon may be material elsewhere (for exceptions see Moleke, 2006; Sam et al., 2018).

Recognising the limited scope of research on wage expectations in developing countries, this paper contributes in three main ways. First, we quantify expectational errors in a new context. Namely, we run a longitudinal labour market transition survey for a large representative sample of graduates in Mozambique, allowing us to compare expected earnings before graduation to later outcomes. Second, we decompose the observed errors into four proximate sources – informational errors about returns to individual attributes; informational errors about differences in earnings across alternative jobs; labour market mismatches (or ‘assignment frictions’, as in Smith, 2010); and remaining optimism (pessimism), which has often been considered a main source of error. Third, based on a follow-up survey, we are able to probe the specific nature of this optimism, linking it to a common form of cognitive bias.

Our first finding is that expectational errors among Mozambican graduates are not just positive but also very large. On average, while around three quarters of the sample undertook some paid work within 18 months of finishing their university course, their starting salary was less than half of what they had expected. Decomposing this error, while specific informational errors (misinformation) do not appear to be so important, we observe that mismatches, both vertical and horizontal, translate into lower-than-expected realized wages. For instance, on beginning work, the majority of participants had not completed all formal study requirements and thus had not yet officially graduated. Furthermore, many were working as (paid) interns, on a part-time basis, without a contract, and/or were continuing to look for another job. Taken together, the wage penalties associated with these mismatches are large and account for around one-third of the overall (average) expectations gap. Expectation errors due to mismatch are, therefore non-negligible and attributing these to some form of ‘unrealistic optimism’ ignores the importance of assignment frictions in the transition from school to work.

A flip-side, however, is that most of the expectations gap cannot be attributed to misinformation or mismatch – that is, a large positive systematic bias remains to be explained. Drawing on the psychological literature on the role of reference points in expectations formation (e.g., Cruces et al., 2013), we demonstrate that forecasts of future earnings are heavily influenced by an unrepresentative reference group of upper-tier or superstar earners. To support this, we show that the distribution of expected wages closely draws from the highest deciles of the *ex post* wage distribution of the same cohort. Additionally, using a bespoke follow-up survey, we find that the highest known wage among their university colleagues represents the most robust and largest correlate of future wage expectations in comparison to other reference points, including the estimated average salary of their colleagues. However, we do not find that more optimistic wage expectations are associated with poorer job outcomes.

2 Expectations vs. reality

This section reviews the existing literature on errors in earnings expectations. The observation of systematic differences between the wages expected by students prior to entering the labour market and their eventual earnings 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 are well-informed. Since then, a range of published studies, summarised in Appendix Table A1, have examined the same issue. Typically, these focus on university and/or high school students – both of which are viewed 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 these previous studies. First, the majority 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 specific subgroups or study fields. 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 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 A1 show substantial variation in expectational errors, even within the same country. But, what accounts for the direction and magnitude of these errors is not clear.[1] While some studies suggest that younger students may incorrectly predict the final level of education at which they will enter the labour market (e.g., Jerrim, 2011), this would generally not account for expectational errors among university graduates. Rather, two different types of information frictions are likely to be relevant. One concerns knowledge about market returns to individual attributes, such as prior experience or gender. The other is knowledge about differences in earnings across alternative jobs or sectors, regardless of the particular worker in that position (earnings segmentation). In the USA, Carvajal et al. (2000) shows that both types of informational errors are present. Comparing the expectations of college seniors to the actual salaries of recent graduates, they find seniors under-estimate the gender wage gap but over-estimate both the minority wage gap and the premium associated with working in a large firm. Similarly, Wiswall and Zafar (2015) shows that college students are substantially misinformed about (population) earnings differences between different study majors. The literature also hints that students from more deprived backgrounds, as well as those exposed to more challenging labour market conditions, tend to make comparatively larger expectational errors (Rouse, 2004; de Paola et al., 2005; Van der Merwe, 2009; Vasilescu and Begu, 2019).

As noted in the Introduction, a second, potentially complementary, explanation for systematic

[1] This echoes a more general lack of attention to how expectations are actually formed. As Manski (1993) puts it: "Having chosen to make assumptions rather than to investigate expectations formation, economists do not know how youth infer the returns to schooling. ... Without an understanding of expectations, it is not possible to interpret schooling behavior nor to measure the objective returns to schooling. As a consequence, the economics of education is at an impasse." (p. 55)

gaps between expected and realized wage outcomes concerns difficulties in obtaining the *type* of job that was anticipated when wage expectations were elicited. Rather than staying unemployed, individuals may accept job offers in organizations or roles that they had not originally desired. Studies of these 'assignment frictions' (Smith, 2010), 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. An example of horizontal mismatch was found by Malamud (2010), with clear effects on initial wages and higher costs to those that specialized early. To these we might add completion or *certification mismatch*, which refers to cases where individuals begin work without having fully completed the final level of education they had earlier anticipated, meaning they cannot benefit from institutional wage-premia based on certified levels of formal educational attainment. Indeed, Jaeger and Page (1996) noted that failing to account for late degree completions may bias the estimated effects of a diploma. In this study, we quantify the contribution to expectational errors of these three types of mismatch.

Studies of various forms of mismatch and their implications also 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. Leuven and Oosterbeek (2011) survey over one hundred empirical studies of vertical mismatch; however, none of these refer to the African continent and just 18 to Asia. Nonetheless, 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 these authors, the average penalty associated with being over-educated for one's work position equals around half of the coefficient associated with the required or minimum level of schooling for that position (also Dolton and Silles, 2008; Li et al., 2018;

Caroleo and Pastore, 2018).

Existing literature related to certification mismatch has mostly focussed on the determinants and implications of dropping-out of college (e.g., Manski, 1989; Light and Strayer, 2000). However, a small group of studies consider the more specific problem of delayed completion, which occurs where individuals prolong the length of their studies beyond the minimum course duration and graduate late. As Aina et al. (2011) document, this is a serious problem in certain countries and appears closely associated with graduate 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 may be motivated by access to student funding but nonetheless can have consequences for later earnings – 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.

A third general explanation for expectational errors refers to cognitive biases. This goes beyond the specific tendency to over-estimate one's own ability or under-estimate the probability of negative events (the 'better-than-average effect'), some of which may be captured by including relevant variables in wage determination equations (see below). Rather, and as Jefferson et al. (2017) explain, 'unrealistic optimism' may be driven by a form of motivated cognition, in which individuals downplay or filter undesirable information. This can reflect the workings of a representativeness heuristic (Bar-Hillel, 1980; Shepperd et al., 2015), whereby information about specific individuals (e.g., known high earners) is perceived to be more relevant than generic salary information (e.g., minimum wages). We return to this issue in Section 6, but highlight for now that any such unrealistic optimism would emerge in the data as a systematic unexplained (residual) bias that remains after accounting for the contributions of either misinformation or mismatch on observed characteristics.

3 Analytical framework

The previous section distinguished between different proximate sources of expectational errors. We now set out these ideas formally, leading to a simple empirical decomposition procedure. In line with Dominitz (1998), we start with the assumption that subjective (point) estimates of expected wages are always of a conditional nature – i.e., they combine expectations of personal characteristics, being in a specific type of work, plus other relevant information available to the individual at the time of elicitation. Thus, the natural logarithm of the wage expected by individual i to be received at time $t + n$ is given by:

$$w^e_{i,t+n} = \mathrm{E}(w_i \mid O^e, \Omega^e, t + n) \tag{1}$$

where O^e represents a set of expected attributes deemed relevant to earnings, such as the individual's level of education and occupation; and Ω represents the current information set or beliefs regarding how these attributes are rewarded.[2] Focussing on the expected wage in the first job (after completing university), we place further empirical structure on this expression using a conventional Mincerian (hedonic) function:

$$\begin{aligned} w^e_i &= f^e(z^e_i, h^e_i, t^e) \\ &= z^{e\prime}_i \beta^e + h^{e\prime}_i \gamma^e + \delta^e t^e_i + (\mu^e + \varepsilon^e_i) \end{aligned} \tag{2}$$

Here, expected attributes are represented by z^e and h^e, which are individual and occupational characteristics respectively; and t^e is the expected time at which the first job is actually found. In relation to equation (1), the final term in parentheses (a constant plus residual) can be thought of as the individual-specific reference or base wage rate, while the other model parameters capture beliefs about how (expected) attributes are differentially rewarded – i.e., they capture variation

[2] Henceforth, superscript e denotes the expected future values; and superscript r denotes realized values.

around the reference wage rate.

A similar expression can be applied to the realized wage. Here a proportion of individuals accept employment offers and in turn report data on their wage income, as well as the characteristics of their job. Thus, conditional on finding work, the individual's realized wage at time t can be expressed as:

$$w_{it}^r = z_i^{r\prime}\beta^r + h_i^{r\prime}\gamma^r + \delta^r t^r + (\mu_t^r + \varepsilon_{it}^r) \tag{3}$$

In previous studies, expectational errors have often been modelled only as a function of baseline characteristics (e.g., Webbink and Hartog, 2004; Vasilescu and Begu, 2019). However, from the above it is evident not only that expected beliefs about rewards in the labour market may diverge from their later realizations, but also that the expected attributes of future job positions may not be realized. Taking this into account, a general expression for the gap between expected and realized earnings is just the simple difference:

$$w_i^e - w_{it}^r = (t_i^e\delta^e - t_i^r\delta^r) + (z_i^{e\prime}\beta^e - z_i^{r\prime}\beta^r) + (h_i^{e\prime}\gamma^e - h_i^{r\prime}\gamma^r) + (\mu^e - \mu_t^r) + (\varepsilon_i^e - \varepsilon_{it}^r) \tag{4}$$

From the perspective of empirical analysis, the above expression does not clearly identify the contribution of the different types of error discussed earlier. However, assuming individual characteristics are fixed over time ($z^e = z = z^r$) and using standard Blinder-Oaxaca methods (e.g., Blinder, 1973),[3] we algebraically transform the expression to distinguish between four

[3] For instance, $h_i^e\gamma^e - h_i^r\gamma^r \equiv h_i^e\Delta\gamma + \Delta h_i\gamma^r$, and where $\Delta h_i = h_i^e - h_i^r$.

distinct components:

$$w_i^e - w_{it}^r \equiv g_{it} = g_{I,i} + g_{J,i} + g_{M,i} + g_{R,it}, \tag{5}$$

$$\begin{aligned} \text{where:} \quad g_{I,i} &= t_i^e \Delta\delta + z_i' \Delta\beta \\ g_{J,i} &= h_i^{e\prime} \Delta\gamma \\ g_{M,i} &= \Delta t_i \delta^r + \Delta h_j' \gamma^r \\ g_{R,it} &= \Delta\mu_t + \Delta\varepsilon_{it} \end{aligned}$$

The first component, g_I, captures the contribution to the total expectations gap of private informational errors, namely differences between the expected and actual returns to fixed individual attributes, including time. In principle, to the extent that any self-enhancement bias varies systematically with personal characteristics (e.g., by gender), this component should capture the contribution of such biases.[4] The second component, g_J, captures the contribution of public informational errors about rewards to different observable job characteristics (e.g., type of employer). The third component, g_M, captures the net wage contribution of mismatches between expected and realized job outcomes (not returns), where the difference terms capture matching errors across different job dimensions. The final component, g_R, represents the systematic component of any remaining unexplained error and is associated with the reference category wage – i.e., this will capture whether wage expectations in the reference category are systematically biased.[5] By construction, this term is distinct from any contribution of errors associated with private information and mismatch, both of which can reflect self-enhancement

[4] For instance, imagine if only men were prone to self-enhancement bias, but in reality there is no gender discrimination in actual wages. If so, we would expect to find a positive difference between the expected return to being male and the actual parameter. For discussion of this phenomenon, see Risse et al. (2018).

[5] Conceptually, we can think of this as relating to the average or default wage rate, in relation to which individuals shift their own expectations upwards or downwards depending on their expected divergence from the reference profile. As such, we define the reference category (throughout) as the most frequent unique combination of study area, expected employer and gender. This group is: male students of Education who intend to work in the public sector.

bias. That is, the final component plausibly captures some kind of 'absolute unrealistic optimism' in the sense of Shepperd et al. (2015).

To estimate the parameters of the error decomposition given by equation (5) we use conventional regression techniques, including both linear (least squares) and non-linear (quantile) methods. In doing so, the objective is to identify systematic associations in the data. This primarily constitutes a diagnostic exercise, not a formal causal analysis. Even so, we recognise the presence of omitted variables could bias coefficient estimates and, thereby, confound the accurate quantification (comparison) of different sources of expectational errors. To address this concern, 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 family background and a wide range of job characteristics (see Appendix C for a complete list). In addition, we attempt to correct for any selection bias associated with who eventually gains employment. To do so, we evaluate the (*ex post*) probability of obtaining a job, based on initial characteristics and job preferences, using a probit model. We then use the generalized residual from this procedure, plus its interactions with a set of baseline characteristics, as a control function in the subsequent decomposition regressions (see Wooldridge, 2015). Further details regarding the data and methods are given below.

4 Mozambique tracer survey

4.1 Background

In 2017 we implemented a representative survey of over 2,000 students in their final year of studies across the six largest public and private universities in Mozambique. Starting in early 2018, we proceeded to re-contact the same individuals on a quarterly basis, via mobile phone, in order to follow their transition into the labour force. The design of this tracer survey, described in detail in Jones et al. (2018a), was motivated by three uncontroversial facts. First, and not unlike other (low-income) countries, Mozambique has witnessed rapid growth in access to education at all levels over recent decades (Jones et al., 2018b). In the tertiary sector, the number of students graduating each year (across the country) has risen dramatically, from under 700 in 2003 to over 18,000 in 2016 (Jones et al., 2018a), implying an annual growth rate of around 30%. However, educational expansion has occurred from a very low base and stocks of tertiary-educated workers remain some of the lowest in the world. Based on the comparative statistics compiled by Barro and Lee (2013), in 2010 Mozambicans aged 15 and over had completed only 1.93 years of schooling on average (versus 5.05 for the SSA region), while only 0.3 percent of the same group had completed tertiary education (versus 0.96 for the region). More recent statistics from the 2017 population census indicate that less than 2% of Mozambicans aged 15 and over have completed studies at the bachelor level or above.

Given their scarcity, one might think that university graduates are unlikely to encounter great difficulty in finding work. However, a second fact is that new graduates face what can only be described as a challenging jobs environment. The formal employment 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 are found largely in the (informal) services sector. Since around the mid-2000s, economic growth has become increasingly driven by extractive industries. While these sectors have seen significant investment, they are capital intensive and have often relied on foreign workers to fill key technical and managerial positions. As such, neither rapid nor sustained growth in demand for workers with a university education has been evident. This challenge is compounded by recent macroeconomic developments. The discovery of a series of government-backed commercial debts in 2013 and 2016 provoked a freezing of foreign aid and large cuts in government spending. As a result, real economic growth slowed to around 3% (barely above population growth) and, over the survey period, recruitment into the public sector was reduced dramatically.

Third, information systems in Mozambique are weak. The country has no regular labour market survey, no history of (thematic) panel data collection, and the last household budget survey was undertaken in 2014/15. While some limited follow-up of alumni has been attempted by certain universities, this has not been systematic and relevant samples are small and non-representative. In sum, public policy as regards the tertiary sector is not supported by an extensive evidence-base.

4.2 Survey data

As described in Jones et al. (2018a), the 2017 baseline survey was designed to be representative of the population of Mozambican university graduates by gender and study area (*viz.*, Education, Humanities, Social Sciences (including law), Natural Sciences, Engineering, Agriculture, and Health).[6] The baseline survey collected data on personal characteristics, educational and professional histories, cognitive abilities and labour market expectations. Starting from early 2018, *after* their studies should have been completed, we re-contacted the same individuals six times by telephone on a quarterly basis, 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.[7]

Of the 2,175 finalists surveyed in the baseline (1,024 women and 1,151 men), a total of 1,920 (88% of the baseline sample) both consented to participate in the follow-up telephone rounds and

[6] 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.

[7] Further details regarding the follow-up survey (and baseline) can be found in Jones et al. (2019).

provided valid wage expectations.[8] Of these, we were able to track 1,892 (98.5% of the eligible sample) at least once during the follow-up period. This constitutes our primary analytical sample. Appendix Figure B1 illustrates the sample dimensions, identifying the number of participants (by gender) reporting a first job in each of the follow-up rounds, plus the number reporting no first job (who remained unemployed or inactive). As shown, of the 1,415 who found a job during the survey period, around half reported to be working in the first telephone round, reflecting that many were already working or had a job lined-up. These early entrants are dominated by men, while women predominate among those who did not report any job during the period.

Table 1 reports baseline characteristics for the primary sample, split between those that did and did not obtain a paid position during the follow-up period. It summarizes important controls of labour market participation and wages, suggested by the literature. Individual characteristics that can be significant determinants are age (following Devereux and Fan, 2011; Black et al., 2011), gender (following Black et al., 2018; Malamud and Pop-Eleches, 2010), marital and parental statuses and intention to seek work. Other factors are type of university students graduated from (following Andrews et al., 2016, 2020), course of study (following Altonji et al., 2014; Delaney and Devereux, 2019), preferred sector and employer (following Hanna and Wang, 2017), comparative perceptions of cognitive abilities (following Carrell et al., 2018) and market-valued skills (following Deming and Kahn, 2018; Delaney and Devereux, 2020) or geographical origin (following Borjas et al., 1992; Meng and Zhang, 2001; Trejo, 1995). Among those who did not find work, 82% had originally expressed an interest in seeking work after their studies, implying this group is mostly not unemployed (inactive) by choice. However, survey participants who did find work tended to be significantly older (by 2 years), more likely to be male, married and with children. Students of (lower cost) public universities are comparatively over-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.

[8] Individuals who had no foreseeable intention to look for work, were not asked this question.

Table 1: Descriptive statistics from baseline survey (2017)

	Obtained work in follow-up period?					
	No		Yes		All	
Individual characteristics:						
Age	24.46	(0.20)	26.49	(0.17)	25.97	(0.14)
Female	0.69	(0.02)	0.39	(0.01)	0.46	(0.01)
Married	0.11	(0.01)	0.16	(0.01)	0.14	(0.01)
Has kids	0.21	(0.02)	0.34	(0.01)	0.30	(0.01)
Plans to seek work	0.82	(0.02)	0.76	(0.01)	0.78	(0.01)
University attended:						
Public university	0.70	(0.02)	0.82	(0.01)	0.79	(0.01)
Total cost USD/month	75.97	(2.95)	63.81	(1.40)	66.91	(1.29)
Course of study:						
Education	0.23	(0.02)	0.34	(0.01)	0.31	(0.01)
Humanities	0.01	(0.01)	0.02	(0.00)	0.02	(0.00)
Social Sciences	0.55	(0.02)	0.42	(0.01)	0.45	(0.01)
Natural Sciences	0.04	(0.01)	0.04	(0.00)	0.04	(0.00)
Engineering	0.07	(0.01)	0.07	(0.01)	0.07	(0.01)
Agriculture	0.05	(0.01)	0.05	(0.01)	0.05	(0.01)
Health	0.05	(0.01)	0.06	(0.01)	0.06	(0.01)
Job expectations:						
Plans to seek work	0.82	(0.02)	0.76	(0.01)	0.78	(0.01)
Private sector employee	0.34	(0.02)	0.33	(0.01)	0.33	(0.01)
Public sector employee	0.43	(0.02)	0.46	(0.01)	0.45	(0.01)
NGO employee	0.06	(0.01)	0.05	(0.01)	0.05	(0.01)
Self/family employed	0.16	(0.02)	0.16	(0.01)	0.16	(0.01)
Wage (USD/month)	413.89	(8.73)	437.15	(4.83)	431.22	(4.24)
Observations	477		1,415		1,892	

Notes: cells are variable means calculated applying survey weights, with standard errors in parentheses; costs and wages are in constant (November 2019) values.
Source: own estimates.

In terms of job expectations as reported at the baseline, employment in the private sector dominates (45%), followed by the public sector (33%) and then self-employment (16%). The average expected starting salary was just over US$ 450 per month (after tax), which compares to a minimum wage of just less than US$ 100 per month.[9] Comparing those who did and did not eventually find work, the expected salary distributions are statistically different (at the 5% levels). Combined with other differences in the profiles of these two groups, the possibility of (unobserved) selection bias in finding employment cannot be dismissed and we return to this below.

Employment outcomes for the first paid position reported in the follow-up period are summarised in Table 2. In terms of the type of employer, average outcomes would appear to bear a reasonable resemblance to expectations (e.g., 52% work in the private sector vs. 45% in the baseline expectation). However, in line with Section 3, a closer look at the individual level reveals mismatches are in fact common.[10] At the time they were observed in their first job, a large proportion of individuals stated they: had not yet formally completed their studies (76%); were working in positions outside their field of studies (57%); were working as interns (13%) or on a part-time basis (51%); did not have a fixed/permanent contract (70%); were actively looking for another job (63%); were not working for the type of organization stated in the baseline (69%); and were not working in the sector identified in the baseline (53%). Each of these eight types of mismatch, which cover vertical, horizontal and certification dimensions, are operationalized as dummy variables in the decomposition analysis. On average, the individual-specific sum of mismatches is close to four, which suggests first jobs generally do not match closely with original expectations.

[9] Minimum wages vary by sector so this is the sector-wide mean minimum wage as agreed in April 2019. For ease of interpretation, all monetary values are stated in constant prices (November 2019 = 1) and, where relevant, converted to US$ at an exchange rate of 60 Meticais : 1 US$.

[10] 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: Realized outcomes in first labour market position (N = 1,415)

	Private uni.		Public uni.		
	Male	Female	Male	Female	All
Private sector employee	0.57	0.64	0.44	0.47	0.48
Public sector employee	0.20	0.12	0.26	0.30	0.26
NGO employee	0.09	0.06	0.09	0.07	0.08
Self/family employed	0.14	0.18	0.21	0.16	0.19
Study unfinished	0.70	0.66	0.82	0.76	0.78
Job unlike course	0.50	0.58	0.52	0.59	0.55
Internship	0.15	0.16	0.12	0.12	0.13
Works part time	0.47	0.42	0.59	0.46	0.52
No fixed contract	0.67	0.64	0.73	0.70	0.70
Searching for work	0.64	0.59	0.68	0.59	0.64
Employee mismatch	0.61	0.66	0.68	0.63	0.66
Sector mismatch	0.40	0.43	0.55	0.45	0.49
Mismatches (count)	4.13	4.14	4.67	4.29	4.46
Realized wage (USD/month)	221.63	210.55	158.82	157.16	168.40
Expected - realized wage (USD)	252.83	223.14	296.91	237.88	268.76
Expectational error (log.)	0.87	0.84	1.19	1.04	1.09

Notes: unless otherwise indicated, cells report the proportion of individuals in each column subgroup with the indicated job characteristic; mismatches are all 'positive' – i.e., score a zero if there is no mismatch.
Source: own estimates.

The last part of Table 2 compares realised wages to their baseline expectations. The gap is positive and large – on average, individuals in their first paid position after university earn US$ 173 per month, which is US$ 289 less than what they had expected. Transformed into natural logarithms, the expectational error, defined as expected minus realized wage, equals 1.15 points on average. The expected and realized wage 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$). The latter shows that fewer than 10% of the respondents who obtained a job received a wage that equalled or exceeded their earlier expectations; and close to 80% reported to be receiving at least US$ 100 *less* than they had expected per month. Overall, this confirms that university graduates face a tough jobs market, at least compared to their expectations in their final year of studies. And while the presence of a positive expectational error is not so surprising, the magnitude of this error in this case is large in relation to earlier studies. This motivates the decomposition analysis, to which we now turn.

5 Results

5.1 Wage determination

To begin the formal analysis of expectational errors, 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.

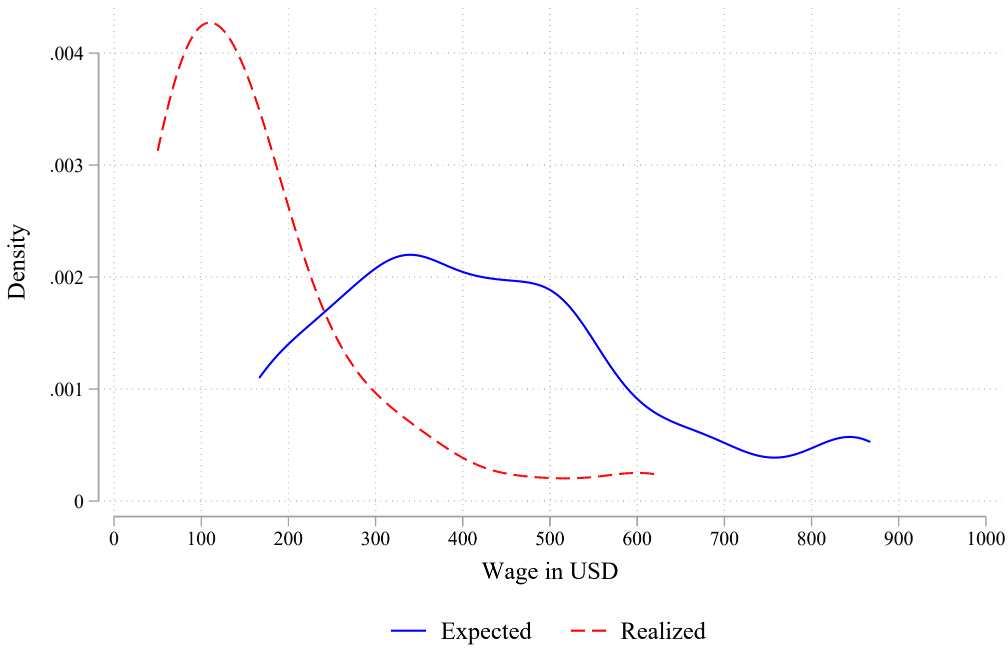
Column (I) of Table 3 summarises estimates from a linear probability model, where the dependent variable takes a value of one if the participant reported having a paid job post-baseline, using only baseline individual and future (desired) job characteristics as explanatory variables.[11] In this 'selection equation' we also include each participants' original stated interest in seeking work plus its interaction with gender and having children, which together are excluded from the subsequent outcome specifications (and thus operate as instrumental variables to address unobserved selection effects).[12] The model 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.

[11] See the variables under group I in Appendix C. Throughout, the (excluded) reference category is the largest group of students, namely men who attended courses in 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.

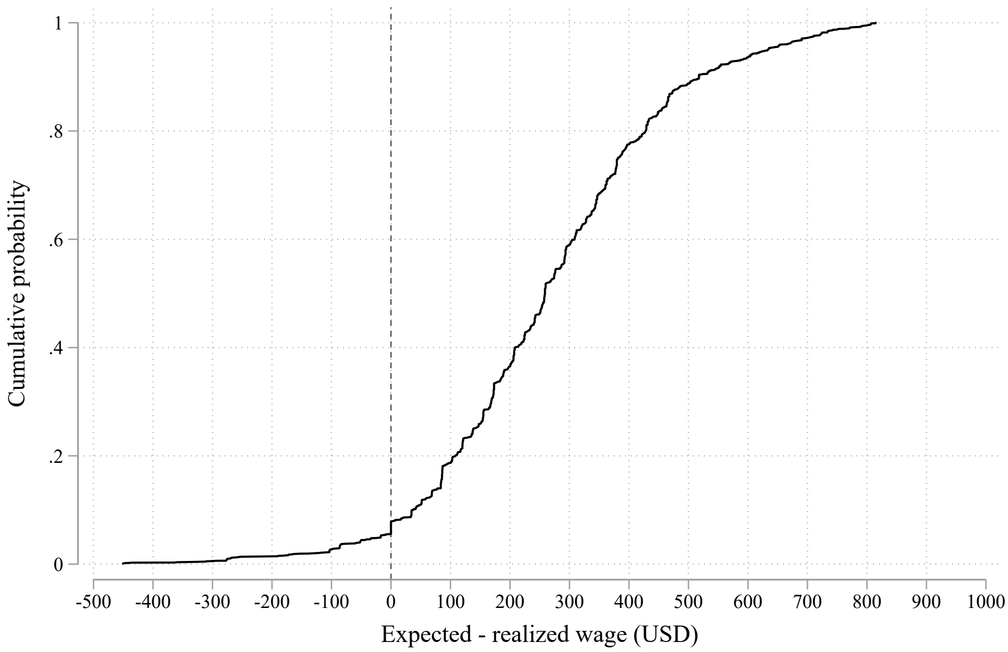
[12] Coefficient estimates for these variables are not shown; however, their joint significance is indicated in the 'control function' row in the footer.

Figure 1: Expected vs. realized wages

(a) Cross-sectional distributions



(b) Individual-level differences



Source: own calculations.

Table 3: Linear regression estimates of job expectations and outcomes

	(I) Job?	(II) Expected wage			(III) Realized wage		
	(.)	(a)	(b)	(c)	(a)	(b)	(c)
Constant	0.71***	3.09***	3.04***	3.05***	1.72***	2.51***	2.50***
	(0.07)	(0.11)	(0.12)	(0.12)	(0.11)	(0.17)	(0.17)
Age	0.00	-0.00	-0.00	-0.00	0.02***	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female	-0.15***	-0.13***	-0.12***	-0.09	0.10*	0.01	-0.03
	(0.02)	(0.03)	(0.03)	(0.07)	(0.05)	(0.05)	(0.07)
Married	-0.06**	0.04	0.03	0.04	0.11**	0.07	0.05
	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)
Private university	-0.09**	0.03	-0.00	0.01	0.24***	0.23***	0.23***
	(0.04)	(0.05)	(0.05)	(0.06)	(0.07)	(0.06)	(0.08)
Education	0.04	-0.03	-0.05	-0.06	-0.08	-0.16***	-0.15***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
Natural Sciences	-0.02	0.11*	0.13**	0.13**	0.10	0.13**	0.13**
	(0.04)	(0.05)	(0.05)	(0.05)	(0.07)	(0.06)	(0.06)
Engineering	-0.03	0.19**	0.15	0.15	0.25*	0.27**	0.27**
	(0.06)	(0.08)	(0.09)	(0.10)	(0.14)	(0.10)	(0.10)
Health	0.06	0.33***	0.29***	0.29***	0.22*	0.07	0.07
	(0.06)	(0.07)	(0.06)	(0.06)	(0.12)	(0.08)	(0.08)
English proficiency	0.07*	-0.03	-0.05	-0.05	0.12*	0.15**	0.17***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)	(0.06)	(0.06)
Academic level (self)	0.04**	0.02	-0.01	-0.02	0.14***	0.07**	0.08**
	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
Prev. internship	0.03	-0.01	-0.00	-0.00	0.09*	0.04	0.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.04)	(0.05)
Prev. work	0.10***	-0.01	0.01	0.00	-0.04	0.01	0.02
	(0.03)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
Prev. work exp.	0.01***	0.02***	0.02***	0.02**	-0.01	-0.02**	-0.01
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Self/family employed	-0.01	0.02	0.08	0.08*	0.14**	-0.27***	-0.27***
	(0.03)	(0.04)	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)
Study unfinished						-0.20***	-0.20***
						(0.05)	(0.05)
Works part time						-0.26***	-0.25***
						(0.05)	(0.05)
Internship						-0.33***	-0.33***
						(0.06)	(0.06)
Searching for work						-0.12***	-0.12***
						(0.03)	(0.03)
Job unlike course						-0.20***	-0.20***
						(0.05)	(0.05)
Obs.	1,892	1,892	1,415	1,415	1,415	1,415	1,415
R^2	0.16	0.14	0.15	0.15	0.23	0.35	0.35
Control func. (pr.)	0.03			0.50			0.51
Actual outcomes?	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(c) are only those that obtained a job; in column I(a) selection variables are included and their joint significance reported in 'control function'; in columns III(b) and III(c) all job outcomes are as realized, else they are as expected; only selected coefficients shown; control function terms to address selection bias are included in columns II(c) and III(c) (joint probability shown); robust standard errors clustered by baseline survey session are given in parentheses.

Source: own estimates.

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 = 1,892), while column (IIb) only contains the sub-sample for which we have a subsequent wage realization (N = 1,408). 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 to the sub-sample model the standardized generalized residual from a probit model on the form of column (I) plus its square and its interaction with gender. Following Wooldridge (2015) this represents a flexible control function to address unobserved selection bias. As shown in the footer of the table, these terms are jointly statistically significant at the 10% level; and, when included, they result in the shrinkage of the coefficient on being female toward zero, while other estimated coefficients remain largely unchanged. 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. In particular, a number of baseline factors that in practice are not material to obtaining a job nonetheless appear to be relevant determinants of expected wages. This is most clear for the area of study – e.g., students of engineering, health and natural sciences all expect higher starting salaries than those studying in the field of education (the base category); also participants expect to obtain lower salaries in the public sector relative to the private sector or self-employment. In line with our analytical framework, this supports the idea that wage expectations are conditional on realizing specific job outcomes and that participants expect the labour market to reward specific individuals and job types differently.

The remaining columns of Table 3 (IIIa-IIIc) shift the focus to realized wages in the first job observed in the follow-up period. Column (IIIa) replicates the specification of column (IIa), using only baseline characteristics as explanatory variables. Here some immediate differences are apparent. Women would appear to earn marginally more than men (*ceteris paribus*), as do graduates from private universities, while the discount associated with public sector work appears more severe than expected (-0.25 log points versus -0.05 points in column IIa). The problem with interpretations of this sort is that not all students who had expressed a desire to work in the public sector subsequently did so – i.e., the public sector dummy variable in column (IIIa) refers to the *desired* rather than the *actual* employer. To clarify the relevance of this point, columns (IIIb) and (IIIc) replace the expected labour market job characteristics (sector and employer) with their realized counterparts, now in accordance with equation (3). We also add controls for a range of job characteristics, which form the basis for identifying mismatches (see Section 4 and Appendix C).[13] Parameter estimates for the new specification shift substantially in magnitude relative to the (mis-specified) model including only baseline characteristics. Among these, the mismatch variables are not only statistically significant but are associated with large discounts to realized wages. For example, not having completed one's studies (a certificate mismatch) is associated with a discount of around 20% on realized wages; and having a job outside the field of study (horizontal mismatch) is associated with a 17% wage discount. Last, the control function variables included in column (IIIc) remain material; however, differences in parameters estimates are minor in comparison to those reported in column (IIIb).

5.2 Expectational errors

The simple difference between the models given in columns (IIc) and (IIIc) of Table 3 represents a basic model for the expectational error, as per equation (4). Applying the re-arrangement proposed in equation (5), Table 4 provides the preferred decomposition results. Columns (I) and (II) refer to alternative estimators, where the former is (sample weighted) OLS and the latter

[13] These additional variables are all assumed to take a value of zero in the (baseline) expected wage equation.

is the iteratively reweighted least squares (IRWLS) proposed by Huber (1973). Sub-columns (a) regress the expectation error on the set of baseline characteristics / expectations only, which is equivalent to assuming zero mismatches (as in Webbink and Hartog, 2004); sub-columns (b) relax this restriction, representing the complete specification; and sub-columns (c) add the control function terms, derived from the selection model (Table 3, column I).

Four principal findings merit note. First, as before, the complete specification adds significant explanatory value relative to its restricted counterpart. Accounting for labour market mismatches not only improves the overall goodness-of-fit of the model by around two thirds, increasing the R^2 from 0.15 to 0.25 (see columns IIa versus Ia), but also parameter estimates differ substantially between the two specifications. For instance, under the restricted model (columns Ia and IIa), the difference between expected and realized returns to self-employment are not different from zero. In contrast, under the complete model, our results suggest these same expectations are excessively optimistic (by around 0.20 log points).

Second, when mismatches are taken into account, the magnitude of the systematic unexplained error – the reference category error – falls considerably. While this is evident directly from the magnitude of the constant in the regression estimates, it can be seen more clearly from the contribution of each error term to the total error (at the average of the explanatory variables). To see this, for each error component of equation (5) we aggregate the relevant regression estimates using the following shrinkage formula:

$$c \in \mathcal{S}: \; g_{c,i} = \sum_{x \,\in\, c} \hat{\theta}_x x_i \times [1 - \Pr(\hat{\theta}_x = 0)] \tag{6}$$

where $\mathcal{S}$ is a collection of sets, the elements of which partition all explanatory variables (x) entering the decomposition regression according to the different error components: $\mathcal{S} = \{I, J, M, R\}$ (see Appendix C for a complete list of variables and their partitions). Thus, for $c = \{M\}$, we refer to the difference terms that capture the extent to which an individual is mismatched

Table 4: Regression estimates of expectational error (first job)

	(I) OLS			(II) Robust [M-estimator]		
	(a)	(b)	(c)	(a)	(b)	(c)
Constant	1.33***	0.76***	0.78***	1.42***	0.76***	0.79***
	(0.16)	(0.21)	(0.21)	(0.13)	(0.16)	(0.16)
Age	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***
	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Female	-0.21***	-0.14**	-0.07	-0.20***	-0.13***	-0.06
	(0.06)	(0.06)	(0.10)	(0.05)	(0.05)	(0.09)
Private university	-0.24***	-0.22***	-0.21**	-0.27***	-0.22***	-0.20***
	(0.07)	(0.07)	(0.09)	(0.06)	(0.06)	(0.07)
English proficiency	-0.17**	-0.18**	-0.20**	-0.19***	-0.18***	-0.21***
	(0.07)	(0.07)	(0.08)	(0.06)	(0.06)	(0.07)
Academic level (self)	-0.15***	-0.09**	-0.10**	-0.14***	-0.07*	-0.08*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Prev. internship	-0.09*	-0.06	-0.07	-0.13***	-0.09**	-0.10**
	(0.05)	(0.05)	(0.06)	(0.05)	(0.04)	(0.05)
Prev. work	0.05	0.01	-0.01	0.04	0.02	-0.01
	(0.05)	(0.05)	(0.06)	(0.05)	(0.04)	(0.06)
Prev. work exp.	0.02***	0.03***	0.02**	0.03***	0.03***	0.02*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Self/family employed	-0.06	0.21**	0.22**	-0.10	0.21**	0.21**
	(0.07)	(0.10)	(0.10)	(0.06)	(0.09)	(0.09)
Private services	0.09	-0.14	-0.13	-0.01	-0.18**	-0.18*
	(0.09)	(0.10)	(0.10)	(0.08)	(0.09)	(0.09)
Lives in Sofala (Δ)		-0.25**	-0.25**		-0.23*	-0.23*
		(0.12)	(0.12)		(0.12)	(0.12)
Study unfinished (Δ)		-0.17***	-0.17***		-0.17***	-0.17***
		(0.06)	(0.06)		(0.05)	(0.05)
Works part time (Δ)		-0.26***	-0.26***		-0.28***	-0.28***
		(0.05)	(0.05)		(0.05)	(0.05)
Internship (Δ)		-0.30***	-0.30***		-0.34***	-0.34***
		(0.08)	(0.09)		(0.07)	(0.07)
Searching for work (Δ)		-0.06	-0.06		-0.07*	-0.07*
		(0.04)	(0.04)		(0.04)	(0.04)
Job unlike course (Δ)		-0.15***	-0.15***		-0.20***	-0.20***
		(0.04)	(0.04)		(0.04)	(0.04)
NGO employee (Δ)		0.18**	0.18**		0.23***	0.22***
		(0.09)	(0.09)		(0.08)	(0.08)
Self/family employed (Δ)		-0.27***	-0.27***		-0.27***	-0.27***
		(0.08)	(0.08)		(0.07)	(0.07)
Private services (Δ)		0.19***	0.19***		0.18***	0.17***
		(0.07)	(0.07)		(0.07)	(0.07)
Obs.	1,415	1,415	1,415	1,415	1,415	1,415
R^2	0.14	0.25	0.25	0.16	0.30	0.30
Control func. (pr.)			0.41			0.62

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 5: Summary of expectational error components (first job)

	(I) OLS			(II) Robust [M-estimator]		
	(a)	(b)	(c)	(a)	(b)	(c)
Indiv. info.	-0.28	-0.13	-0.13	-0.26	-0.16	-0.17
	[-0.44,-0.13]	[-0.27,0.02]	[-0.27,0.01]	[-0.40,-0.12]	[-0.28,-0.04]	[-0.31,-0.04]
Job info.	0.07	0.08	0.08	-0.04	0.02	0.02
	[-0.07,0.21]	[-0.10,0.26]	[-0.10,0.26]	[-0.10,0.03]	[-0.12,0.16]	[-0.12,0.17]
Match quality	.	0.40	0.40	.	0.49	0.49
		[0.28,0.52]	[0.28,0.52]		[0.37,0.60]	[0.37,0.60]
Ref. point	1.31	0.72	0.74	1.39	0.72	0.75
	[0.99,1.63]	[0.31,1.12]	[0.32,1.15]	[1.14,1.65]	[0.42,1.03]	[0.44,1.07]
Total error	1.09	1.07	1.09	1.09	1.07	1.09
	[0.96,1.23]	[0.88,1.26]	[0.89,1.30]	[0.95,1.23]	[0.94,1.19]	[0.95,1.22]

Notes: cells report the point estimate and 95% confidence intervals associated with the overall contribution of different expectational error components, as derived from the models in the respective columns of Table 4; error contributions are shrunk, as per equation (6).
Source: own estimates.

in her first job; $\hat{\theta}_x$ are the coefficient estimates of this vector of variables; and the shrinkage factor is employed to downsize parameter estimates that are not statistically different from zero. Table 5 reports sample averages for these four predicted component errors (and 95% confidence intervals). Under both estimators, inclusion of the mismatch variables leads to an approximate 50% fall in the reference error; and the match quality error accounts for roughly 40% of the total expectational error. Notably, this is not driven by any single mismatch. As shown in Appendix Figure B2, which illustrates the magnitudes of the five largest contributors to the mismatch error (assessed at the sample mean), certification mismatch represents around a third of this error, followed by working part time (as opposed to full-time) and horizontal mismatch. Also, reflecting the earlier point that the restricted model is misspecified, the other component terms alter in magnitude when the mismatch terms are included.

Third, continuing to focus on the error component estimates from the complete model reported in Table 5 (columns b and c), job-related (public) informational errors are not different from zero on average, but individual (private) informational errors are negative and not immaterial. The former suggests that finalists are not so poorly informed about differences in returns to specific types of job (e.g., in public sector vs private sector). However, the latter suggests finalists generally *under*-estimated labour market returns to specific individual attributes. As shown in Appendix Figure B3 (also evident from the parameter estimates in Table 4; also Appendix Table A2), both finalists who had children and females expected to encounter a larger relative wage discount in the labour market than they actually encountered in practice. Also, individuals who rated their own academic performance as being above average under-estimated the premium associated with this characteristic, as did those who had attended private universities.

Last, even after accounting for job-related informational, individual informational and match quality errors, a large systematic positive residual error remains. Under the preferred OLS estimates of column I(c) (also the robust counterpart of IIc), which include control function terms, the reference error is very substantial at 0.80 log points (120%), representing more than two-thirds of the total error. Thus, observed characteristics account for under a third of the expectational error.

5.3 Validation

Before investigating what might explain the magnitude of the reference point error, we briefly validate the findings of the previous section. To do so, we run the decomposition regression (using the complete specification, including control function terms) across different percentiles of the expectational error distribution. These results, based on a conventional quantile regression estimator, are reported in Appendix Table A3 and summarised in Table A4. While the general pattern of estimates is fairly stable across percentiles, a few insights stand out. In particular, the contribution of job-related (public) informational errors appears to turn positive in the upper half of the distribution; and match quality is (perhaps unsurprisingly) smallest in the lower percentiles, implying at least a small share of the participants do find good job matches. However, the reference point error is always material and increases systematically across the percentiles, retaining a dominant relative contribution at all points in the distribution.

Second, we consider whether the magnitudes and proximate sources of expectational errors remain after individuals' have gained further experience in the labour market. The hypothesis is that labour market transitions may be not be smooth; even in the first 18 months, individuals may be able to move into better quality employment (e.g., from part-time to full-time, or from interns to permanent staff), and these later salaries may align more closer with earlier expectations. To examine this, we estimate the (linear) regression decomposition replacing the first realized wage observation provided by each participant with the last valid observation (in time). Table A5 summarises the results in the same fashion as before (see Appendix Table A6 for the full regression results). Overall, the total expectational error has diminished by around a third – to 0.74 log. points in column I(c) from 1.16 in the corresponding column of Table 5. This is only partly explained by a smaller match quality error (0.33 vs. 0.40 log points in column Ic); but since the job-related informational errors remain negligible and individual informational errors also have shrunken toward zero, the remaining change is in the systematic residual, which has fallen from 0.79 to 0.43 log. points. One interpretation is that the participants' subjective expectations of first wages did not account for the lower wages received in probationary or trial periods. But this may also reflect strong returns to experience amongst the more successful labour market entrants. In any case, the reference error is hardly trivial at over 50%, and continues to merit further investigation.

6 Optimism and its implications

Returning to the reference error, which to borrow from Abramovitz (1956) effectively just constitutes 'some measure of [our] ignorance', we have argued this term should not reflect self-enhancement bias to the extent that any enhancement varies by observed individual attributes (e.g., gender or self-assessed academic performance). Instead, as noted in Section 2, an alternative explanation for 'unrealistic absolute optimism' relates to the asymmetric or selective way in which information is processed and, in particular, how the representativeness heuristic can distort evaluations of the likelihood of (future) events (see Grether, 1992). While this heuristic can play out in various ways, one occurs when a statistic of interest is given a very high probability (weight) if it is deemed to come from a sample that is representative of a target population, regardless of the size or actual representativeness of the sample. For instance, Cruces et al. (2013) demonstrates how individuals treat information about incomes within their own narrow (similar-income) reference group as-if the group were representative of the general population, yielding systematically biased perceptions of the income distribution.

In the present context, a concern is that job seekers not only may have little concrete information about the relevant distribution of wages in the labour market, but also that any such information tends to come from more successful entrants (or those with more experience). Privacy norms around salaries, especially those of one's immediate peers or co-workers, have been documented in various contexts (Cullen and Perez-Truglia, 2020); and, in Mozambique, it is even the case that published job adverts almost never post information about the posts' salary range. In this light, we investigate whether finalists' salary expectations are distorted by placing excess weight on salaries in the upper tail of the wage distribution – i.e., whether they are referenced to a narrow group of higher earners.

To assess the plausibility of this argument, we begin by estimating where participants' expected first salaries (as elicited at baseline) are located on the distributions of the first and last salaries observed during the follow-up period.[14] That is, for each adjusted expected wage value, we identify its corresponding percentile (location) on the chosen outcome distribution. Figure 2

[14] To remove the bias in expectations that can be accounted for by the three observed errors (g_I, g_J, g_M), we use an adjusted measure of wage expectations, defined as: $\tilde{w}_i^e = w_i^e - g_{I,i} - g_{J,i} - g_{M,i}$, which places attention on the contribution of the reference error only.

plots the cumulative distribution of these effective percentiles. It shows that the median real (adjusted) expected wage corresponds to the 86th percentile of the distribution of realized real wages in the first job after university, or the 70th percentile of the distribution of wages in the last job. This indicates that baseline wage expectations were not completely unrealistic (unattainable), in the sense of being largely outside the support of the realized wage distribution. But expected wages would seem to have been drawn from a selective reference distribution of above-median earners, which also is consistent with the finding that initial salary expectations assumed a good quality job (matched to their preferences) would be obtained.

To test this proposition further, in November 2019 we invited the same group of students to participate in a short internet-based survey.[15] Within this, we not only asked their wage expectations for one year ahead, but we also elicited: (i) their reservation wage (lowest salary they would accept); (ii) their estimate of the current average earnings among their peers; and (iii) their estimate of the current highest earnings among their peers. To test the extent to which either one of these three quantities operate as reference points (anchors) for future expectations, we estimate regressions of the difference between the log. expected wage and the log. of each reference point:

$$w_i^e - \mu_{j,i} = a + x_i'\beta + \epsilon_i \tag{7}$$

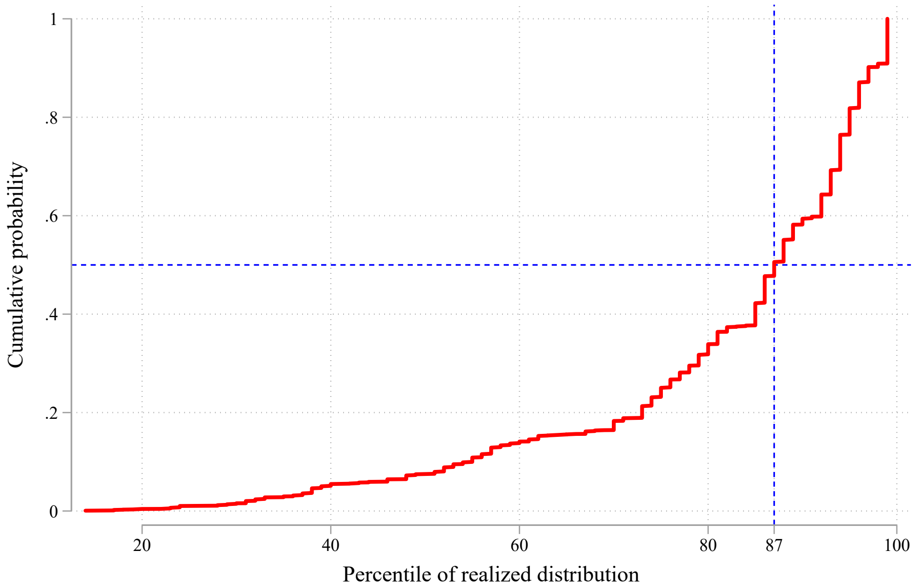
the idea being that the most salient point μ_j should yield an estimate for a that is closest to zero.[16] Results from this exercise are reported in Table 6, where columns I(a)-(c) refer to the full sample and columns II(a)-(c) refer to the matched sample, adding a series of baseline control variables, including study area, university and gender. In all specifications we account for the participants' current work situation, the number of peers in their reference group, as well as any bias that may be caused by the wording of the wage expectations question. In Portuguese, the language in

[15] This was implemented after the final round of the follow-up telephone interviews. We employed a lottery to incentivise responses and received 308 valid responses, of which we could match 275 to the baseline data.

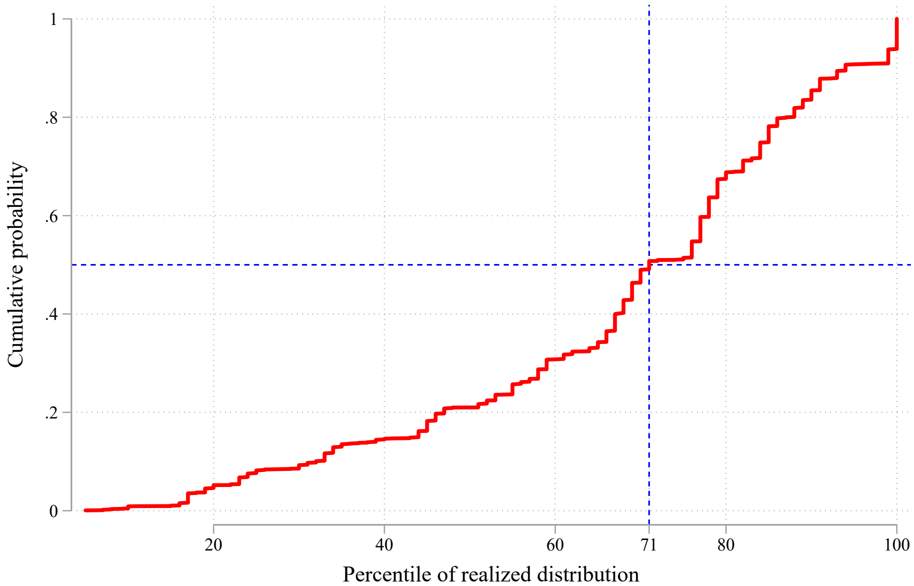
[16] Note, this specification is equivalent to: $w_i^e = \mu_{j,i} + a + x_i'\beta + \epsilon_i$, which clarifies that a represents any systematic difference between the expected wage and the candidate reference point.

Figure 2: Percentile location of (*ex ante*) expected wages on (*ex post*) realized wage distributions

(a) Location on the distribution of wages in first jobs



(b) Location on the distribution of wages in last jobs



Note: expected wages are adjusted to account for observed error components (g_I, g_J, g_M); all comparisons are made in constant prices.

Source: authors' calculations.

Table 6: Difference between expected wage and elicited reference point (internet survey)

	(I) Full sample			(II) Matched sample		
	(a)	(b)	(c)	(a)	(b)	(c)
Reference point →	Reserve	Mean	Highest	Reserve	Mean	Highest
Constant	0.88***	0.38*	0.04	0.96***	0.32	0.02
	(0.21)	(0.21)	(0.20)	(0.28)	(0.26)	(0.25)
Currently working	0.23	0.57***	0.14	0.42*	0.68***	0.22
	(0.21)	(0.19)	(0.18)	(0.23)	(0.21)	(0.20)
Years unemployed	-0.18	0.06	-0.19	-0.09	0.11	-0.18
	(0.19)	(0.18)	(0.16)	(0.19)	(0.17)	(0.15)
No. of peers (log.)	0.08	-0.07	0.06	0.02	-0.14**	0.02
	(0.07)	(0.05)	(0.05)	(0.07)	(0.07)	(0.06)
Wording (like to)	0.04	-0.06	0.10	0.04	-0.07	0.09
	(0.14)	(0.12)	(0.12)	(0.15)	(0.13)	(0.12)
Wording (realistic)	-0.23	-0.23*	-0.16	-0.28*	-0.26*	-0.16
	(0.14)	(0.13)	(0.13)	(0.14)	(0.14)	(0.14)
Female				-0.24**	0.05	0.05
				(0.12)	(0.12)	(0.12)
Baseline expected wage (log.)				0.16***	0.14***	0.13***
				(0.05)	(0.04)	(0.05)
Obs.	308	308	308	275	275	275
R^2	0.06	0.13	0.13	0.23	0.24	0.26
RMSE	1.00	0.90	0.87	0.94	0.88	0.85

Notes: each column reports summary results for models on the form of equation (7), where the dependent variable is indicated in the column sub-header and the 'Constant' coefficient gives the estimate for a; columns (I) refers to the full sample and columns (II) the matched sample allowing baseline characteristics to be included, such as study area and university (not shown).
Source: own estimates.

which our surveys have been administered, the word used to prompt for 'expected' wages can also mean 'hoped for' (*espera receber*). To control for differences in interpretation, we randomly allocated participants to one of three alternative future wage expectations wordings. These are: the same wording as in the baseline questionnaire (not shown); an alternative wording to refer to the salary they would 'like to' receive in one year (*gostaria de receber*); and a wording forcing them to reflect on what they could realistically obtain (*o salário que pensa, realisticamente, que estará a receber*).

The main finding is that the highest reported salary among the participants' peers is associated with the smallest constant term and lowest model RMSE. Indeed, in both columns I(c) and II(c) the constant is not significantly different from zero and the point estimate is almost precisely zero. In contrast, the reserve wage appears to be around 0.90 log points lower than the effective reference point underlying the expected wage, while the peers' average wage is around 0.35 log points lower. Findings for the control variables are generally rather imprecise. Nonetheless, the wording emphasising the realistic wage would appear to prompt participants to report somewhat lower expected wages (by around 0.20 log points), implying some default disposition toward optimism in the earlier responses. Also, the results are robust to including the (matched) baseline controls, including the (centered) expected wage reported at the baseline.

We recognise these results are not conclusive, particularly as they refer to a small convenience sample. Nonetheless, they are consistent with presence of a representativeness heuristic by which graduates place greater weight on information about (more desirable) salaries found at the upper-end of the salary distribution. This over-emphasis on superstar salaries thus appears to be misleading and does not provide an accurate representation of labour market realities. A final and perhaps more fundamental issue is whether optimistic expectations hold any implications for labour market outcomes, such as employment rates or attained salaries. Existing literature has not settled on whether excessive optimism has nefarious consequences. As summarised by Armor and Taylor (2002), on the one hand unrealistic optimism might generate disappointment and undermine motivation. On the other hand, high expectations could be motivational (i.e., operate as a kind of aspiration) and thus come to be self-fulfilling. Alternatively, expectations that refer to the distant future may only be weakly held and, even when expectations are unfulfilled, outcomes can be reinterpreted to minimize the gap between earlier expectations and subsequent reality.

We test this via a series of regressions of relevant (final) outcomes observed at time t, against

Table 7: Relationship between baseline wage expectations and later job outcomes

	Outcome	Obs.	Mean	Estimates $\hat{\theta}$	s.e.	Prob($\hat{\theta}=0 \mid x_i$) Raw	Adj.
(a)	Inactive (%)	1,892	0.07	0.01	(0.010)	0.48	0.60
	Unemployed (%)	1,892	0.27	-0.03	(0.019)	0.09	0.22
	Looking for work (%)	1,892	0.63	-0.04	(0.019)	0.05	0.16
	Working (%)	1,892	0.60	0.00	(0.020)	0.86	0.86
	Refused job offers (%)	1,892	0.14	0.01	(0.016)	0.36	0.52
	Number of different jobs	1,892	1.38	-0.10	(0.059)	0.10	0.20
(b)	Last job earnings (log.)	1,415	2.45	0.17	(0.048)	0.00	0.01
	Job quality score	1,415	0.59	0.03	(0.021)	0.16	0.26
(c)	Earnings meet expectations	1,165	0.55	0.01	(0.038)	0.77	0.85
	Choose same education	1,692	0.58	0.11	(0.037)	0.00	0.02

Notes: rows report results from a series of separate regressions as per equation (8); 'Outcome' refers to the dependent variable; in all models (rows) the independent variable of interest (attached to coefficient θ) is the baseline expected salary; baseline control variables included throughout; in panel (c) the realized salary in the last position is added to the vector of controls; $\hat{\theta}$ reports the estimated regression coefficient of interest, and 's.e.' its cluster-robust standard error; adjusted probability applies the Benjamini-Hochberg correction.
Source: own estimates.

initial wage expectations plus a full set of baseline controls (as per Tables 3 and 4):

$$y_{i,t} = a + \theta w_i^e + x_i'\beta + \varepsilon_{i,t} \tag{8}$$

Results for this exercise, focussing on the estimates for θ, are reported in Table 7. In terms of the chosen outcomes, panel (a) considers measures of labour market experience for the full sample (calculated across all follow-up rounds); panel (b) considers the salary and quality of the final job attained;[17] and in panel (c) we focus on subjective assessments, made in the final telephone survey round, as to whether their current salary was in line with earlier expectations, and whether they would choose to follow the same education (same university, course etc.) as before. In the latter panel we add the attained final salary to the vector of control variables.

[17] Quality is based on a jobs score constructed primarily from the mismatch variables previously discussed as well as indicators of formality. A positive value implies a higher quality job. Further details available on request.

Overall, the results suggest a fairly weak relationship between outcomes and initial expectations. After correcting for multiple hypothesis testing (via the Benjamini-Hochberg procedure, as per the final column), none of the labour market experience metrics show θ is likely to be different from zero. Among the remaining outcomes, the attained final salary appears positively related to initial expectations, as does the assessment of whether they would choose the same education again. As such, these results could be picking-up bias from omitted variables, such as having a positive mindset, in which case excessive optimism also may be symptomatic of certain personality traits that are valuable in the labour market. While further analysis goes beyond the scope of the present study, the main point is that we have no evidence that excessively optimistic wage expectations are associated with poorer labour market outcomes.

7 Conclusion

Based on detailed longitudinal data of a representative sample of university finalists in Mozambique, this study investigated the relationship between expected and realised salaries as the participants transitioned into the labour market. While most (3 in every 4) finalists found some work within 18 months of the end of their final year of studying, the gap between the expected and actual first wage was positive and an order of magnitude larger than encountered in studies elsewhere – on average, expected salaries were around US$ 430 per month; but observed first salaries were around US$ 170. To probe the sources of this gap, we proposed a simple decomposition procedure that distinguishes between: private informational errors (about returns to individual attributes); public information errors (about returns to observable job attributes); match quality errors; and reference error, which refers to the systematic unexplained component.

Results from the decomposition procedure revealed that the expectational error cannot be attributed to informational errors. In fact, private informational errors appeared to be negative, indicating participants tended to under-value the pecuniary returns to some personal attributes (e.g., women expected to receive less than they did). In contrast, individuals were generally not well matched in their first job (i.e., we found horizontal, vertical and certification mismatches), such that the match quality error accounted for around 40% of the total error. The counterpart to these findings was that around two thirds of the error cannot be explained from observed variables, leaving a large systematic residual (equal to around 0.80 log points). In other words, a large part of the expectational error would appear to reflect unrealistic absolute optimism.

Following the literature, we hypothesised this unrealistic optimism may reflect bias associated with a representativeness heuristic, namely where salary expectations are based on a narrow (unrepresentative) reference group of higher earners. We demonstrated this hypothesis is consistent with the observed data; and, using a bespoke internet survey, we found that the highest salary among the participants' peers represents the most salient reference point for future wage expectations, compared to both an estimate of the peers' mean wage and their own reservation wage. At the same time, we found no evidence that higher baseline wage expectation were associated with relatively worse labour market outcomes. If anything, the opposite may be the case.

What might this mean for policy? Certainly, access to information regarding starting salaries and typical career paths for university graduates is extremely scarce in Mozambique. While job entrants seem to have some notion of which positions are relatively better paid, they do not seem to be aware of the complete distribution of wage outcomes (for graduates), or of the extent of mismatch in (early) job positions. As such, and as in many countries, we recommend universities are required to systematically collect and disseminate data on alumni employment outcomes. Indeed, since graduate unemployment rates are hardly trivial despite the limited number of graduates in the country, this may be important to help both the government and individuals determine whether investments in higher education are worthwhile.

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.

Abramovitz, M. (1956). Resource and output trends in the United States since 1870. *The American Economic Review*, 46(2):5–23.

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).

Al-Samarrai, S. and Bennell, P. (2007). Where has all the education gone in sub-Saharan Africa? employment and other outcomes among secondary school and university leavers. *The Journal of Development Studies*, 43(7):1270–1300.

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

Altonji, J.G., Kahn, L.B. and Speer, J.D. (2014). Trends in earnings differentials across college majors and the changing task composition of jobs. *American Economic Review*, 104(5):387–393.

Andrews, R.J., Imberman, S.A. and Lovenheim, M.F. (2020). Recruiting and supporting low-income, high-achieving students at flagship universities. *Economics of Education Review*, 74.

Andrews, R.J., Li, J. and Lovenheim, M.F. (2016). Quantile treatment effects of college quality on earnings. *Journal of Human Resources*, 51(1):200–238.

Armor, D.A. and Taylor, S.E. (2002). When predictions fail: The dilemma of unrealistic optimism. In T. Gilovich, D. Griffin and D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*, p. 334–347. Cambridge University Press.

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.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44:211–233.

Barro, R.J. and Lee, J.W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104:184–198.

Basu, A.K., Chau, N.H., Fields, G.S. and Kanbur, R. (2019). Job creation in a multi-sector labour market model for developing economies. *Oxford Economic Papers*, 71(1):119–144.

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.

Behrman, J.R. (2010). Investment in education – inputs and incentives. In D. Rodrik and M. Rosenzweig (Eds.), *Handbook of Development Economics*, volume 5, chapter 73, pp. 4883–4975. Elsevier.

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

Black, S.E., Devereux, P.J., Lundborg, P. and Majlesi, K. (2018). Learning to take risks? The effect of education on risk-taking in financial markets*. *Review of Finance*, 22(3):951–975.

Black, S.E., Devereux, P.J. and Salvanes, K.G. (2011). Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics*, 93(2):455–467.

Blinder, A.S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8(4):436–455.

Borjas, G.J., Bronars, S.G. and Trejo, S.J. (1992). Assimilation and the earnings of young internal migrants. *The Review of Economics and Statistics*, 74(1):170.

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.

Carrell, S.E., Hoekstra, M. and Kuka, E. (2018). The long-run effects of disruptive peers. *American Economic Review*, 108(11):3377–3415.

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.

Cruces, G., Perez-Truglia, R. and Tetaz, M. (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*, 98(2013):100–112.

Cullen, Z.B. and Perez-Truglia, R. (2020). The salary taboo: Privacy norms and the diffusion of information. NBER Working Paper 25145, National Bureau of Economic Research. (Revised from October 2018).

de Paola, M., Lupi, C. and Ordine, P. (2005). Wage expectations in northern and southern Italian regions: An interpretation based on psychological and social factors. *International Review of Applied Economics*, 19(3):343–358.

Delaney, J.M. and Devereux, P.J. (2019). Understanding gender differences in STEM: Evidence from college applications. *Economics of Education Review*, 72:219–238.

Delaney, J. and Devereux, P. (2020). Math matters! The importance of mathematical and verbal skills for degree performance. *Economics Letters*, 186:108850.

Deming, D. and Kahn, L.B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36(S1):S337–S369.

Devereux, P.J. and Fan, W. (2011). Earnings returns to the British education expansion. *Economics of Education Review*, 30(6):1153–1166.

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: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.

Grether, D.M. (1992). Testing Bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization*, 17(1):31–57.

Hanna, R. and Wang, S.Y. (2017). Dishonesty and selection into public service: Evidence from India. *American Economic Journal: Economic Policy*, 9(3):262–290.

Hino, H. and Ranis, G. (Eds.) (2014). *Youth and employment in Sub-Saharan Africa: Working but poor*. Oxford: Routledge.

Hoxhaj, R. (2015). Wage expectations of illegal immigrants: The role of networks and previous migration experience. *International Economics*, 142:136–151.

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.

Jaeger, D.A. and Page, M.E. (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 78(4):733.

Jefferson, A., Bortolotti, L. and Kuzmanovic, B. (2017). What is unrealistic optimism? *Consciousness and Cognition*, 50:3–11.

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 www.wider.unu.edu/publication/baseline-survey-school-work-transitions-university-graduates-mozambique.

Jones, S., Santos, R. and Xirinda, G. (2019). Final report of the survey on the education-employment transitions of university finalists. Technical report, UNU-WIDER, Helsinki, Copenhagen, Maputo. URL igmozambique.wider.unu.edu/relat%C3%B3rio/final-report-survey-school-work-transition-university-students-portuguese.

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.

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.

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.

Malamud, O. (2010). Breadth versus depth: The timing of specialization in higher education: Timing of specialization in higher education. *Labour*, 24(4):359–390.

Malamud, O. and Pop-Eleches, C. (2010). General education versus vocational training: Evidence from an economy in transition. *Review of Economics and Statistics*, 92(1):43–60.

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

—— (1993). Adolescent econometricians: How do youth infer the returns to schooling? In C.T. Clotfelter and M. Rothschild (Eds.), *Studies of Supply and Demand in Higher Education*, chapter 2, pp. 43–60. University of Chicago Press.

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

Meng, X. and Zhang, J. (2001). The two-tier labor market in urban china. *Journal of Comparative Economics*, 29(3):485–504.

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.

Moleke, P. (2006). *Finding work: Employment experiences of South African graduates*. Pretoria: HSRC Press.

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.

Risse, L., Farrell, L. and Fry, T.R. (2018). Personality and pay: do gender gaps in confidence explain gender gaps in wages? *Oxford Economic Papers*, 70(4):919–949.

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.

Schultz, T.W. (1961). Investment in human capital. *American Economic Review*, 51(1):1–17. 06850.

Shepperd, J.A., Waters, E.A., Weinstein, N.D. and Klein, W.M. (2015). A primer on unrealistic optimism. *Current Directions in Psychological Science*, 24(3):232–237.

Smith, E. (2010). Sector-specific human capital and the distribution of earnings. *Journal of Human Capital*, 4(1):35–61.

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.

Trejo, J. (1995). The labor market skills of recent male immigrants: Evidence from the current population survey. *Industrial and Labor Relations Review*, p. 20.

Vasilescu, M.D. and Begu, L.S. (2019). Youth reservation wages in Romania: Are wage expectations realistic? *Applied Economics Letters*, 26(14):1157–1161.

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.

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

Wooldridge, J.M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2):420–445.

Supplementary material

Misinformed, mismatched or misled?
Explaining the gap between expected and realised graduate earnings in Mozambique

A Additional tables

Table A1: 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 (2019)	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 A1: 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.

Table A2: Error components (first job), by sub-groups

			Error components				
Group	Value	Obs.	Job info.	Ind. info.	Match q.	Ref. pnt	Total
Female	No	844	-0.17	0.06	0.42	0.83	1.14
	Yes	571	-0.21	0.07	0.39	0.75	0.99
Older	No	686	-0.12	0.04	0.43	0.79	1.14
	Yes	729	-0.24	0.08	0.38	0.81	1.04
Public uni.	No	279	-0.34	0.09	0.36	0.75	0.85
	Yes	1,136	-0.15	0.06	0.41	0.81	1.13
Mismatch	$\leq$1	38	-0.21	0.06	0.03	0.72	0.60
	2	146	-0.24	0.08	0.16	0.83	0.83
	3	244	-0.24	0.07	0.24	0.79	0.87
	4	277	-0.19	0.08	0.35	0.77	1.01
	5	281	-0.17	0.06	0.46	0.84	1.19
	6	429	-0.14	0.04	0.61	0.80	1.31
All		1,415	-0.18	0.06	0.40	0.80	1.08

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.

Table A3: Quantile regression estimates of expectational error (first job)

	Percentile				
	10	33	50	66	90
Constant	0.16	0.25	0.73***	1.07***	1.24***
	(0.36)	(0.26)	(0.25)	(0.25)	(0.26)
Age	-0.01*	-0.02***	-0.02***	-0.02***	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Female	-0.12	-0.14	-0.18	-0.10	-0.09
	(0.20)	(0.14)	(0.13)	(0.13)	(0.15)
Private university	-0.35**	-0.23**	-0.19*	-0.18*	-0.15
	(0.15)	(0.11)	(0.10)	(0.10)	(0.13)
English proficiency	-0.09	-0.12	-0.17*	-0.21**	-0.16
	(0.14)	(0.09)	(0.09)	(0.09)	(0.11)
Academic level (self)	-0.08	-0.03	-0.09	-0.10*	-0.06
	(0.09)	(0.07)	(0.06)	(0.06)	(0.07)
Prev. internship	-0.01	-0.10	-0.12*	-0.10	0.03
	(0.11)	(0.07)	(0.07)	(0.07)	(0.08)
Prev. work	0.05	-0.05	-0.06	-0.08	0.09
	(0.15)	(0.09)	(0.08)	(0.08)	(0.10)
Prev. work exp.	0.04	0.04**	0.03**	0.03	0.02
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Self/family employed	0.04	0.27*	0.33**	0.32**	0.35**
	(0.18)	(0.14)	(0.13)	(0.14)	(0.15)
Private services	-0.15	-0.03	-0.09	-0.15	-0.16
	(0.21)	(0.14)	(0.14)	(0.14)	(0.17)
Lives in Sofala (Δ)	-0.06	-0.33*	-0.19	-0.08	0.01
	(0.19)	(0.18)	(0.20)	(0.22)	(0.21)
Study unfinished (Δ)	-0.16	-0.14*	-0.16**	-0.12*	-0.25***
	(0.11)	(0.08)	(0.07)	(0.07)	(0.09)
Works part time (Δ)	-0.23***	-0.27***	-0.21***	-0.25***	-0.38***
	(0.08)	(0.07)	(0.07)	(0.08)	(0.09)
Internship (Δ)	-0.29**	-0.34***	-0.27***	-0.28***	-0.28**
	(0.14)	(0.10)	(0.10)	(0.10)	(0.13)
Searching for work (Δ)	-0.05	-0.10*	-0.08	-0.02	-0.02
	(0.08)	(0.06)	(0.06)	(0.06)	(0.06)
Job unlike course (Δ)	-0.12	-0.15***	-0.15***	-0.18***	-0.13**
	(0.08)	(0.06)	(0.06)	(0.06)	(0.07)
NGO employee (Δ)	0.17	0.17	0.14	0.27**	0.02
	(0.17)	(0.12)	(0.12)	(0.12)	(0.18)
Self/family employed (Δ)	-0.23	-0.34***	-0.38***	-0.28***	-0.33***
	(0.15)	(0.11)	(0.10)	(0.10)	(0.11)
Private services (Δ)	0.22	0.18*	0.17*	0.16	0.27**
	(0.16)	(0.11)	(0.10)	(0.10)	(0.12)
Obs.	1,415	1,415	1,415	1,415	1,415
Control func. (pr.)	0.60	0.92	0.64	0.68	0.92
Error at percentile	0.11	0.69	1.10	1.39	2.08

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.

Table A4: Summary of expectational error components (first job), by percentile

	Percentile				
	10	33	50	66	90
Indiv. info.	-0.00	-0.04	-0.21	-0.16	-0.05
	[-0.22,0.21]	[-0.23,0.14]	[-0.39,-0.03]	[-0.35,0.04]	[-0.24,0.15]
Job info.	-0.04	0.25	0.16	0.07	0.17
	[-0.16,0.09]	[0.07,0.42]	[-0.03,0.34]	[-0.12,0.27]	[-0.09,0.43]
Match quality	0.31	0.43	0.43	0.36	0.52
	[0.14,0.47]	[0.28,0.58]	[0.26,0.60]	[0.22,0.51]	[0.34,0.69]
Ref. point	-0.13	0.15	0.74	1.07	1.24
	[-0.39,0.14]	[-0.17,0.48]	[0.25,1.22]	[0.57,1.56]	[0.73,1.75]
Total error	0.14	0.79	1.12	1.35	1.88
	[-0.01,0.29]	[0.61,0.96]	[0.89,1.35]	[1.08,1.61]	[1.67,2.10]

Notes: cells report the point estimate and 95% confidence intervals associated with the overall contribution of different expectational error components, as derived from the models in the respective columns of Appendix Table A3; error contributions are shrunk, as per equation (6).
Source: own estimates.

Table A5: Summary of expectational error components (last job)

	(I) OLS			(II) Robust [M-estimator]		
	(a)	(b)	(c)	(a)	(b)	(c)
Indiv. info.	-0.22	-0.06	-0.09	-0.23	-0.04	-0.08
	[-0.33,-0.12]	[-0.18,0.06]	[-0.21,0.02]	[-0.37,-0.10]	[-0.15,0.08]	[-0.21,0.04]
Job info.	0.01	0.04	0.06	-0.02	0.01	0.01
	[-0.05,0.08]	[-0.12,0.20]	[-0.11,0.22]	[-0.08,0.05]	[-0.12,0.13]	[-0.12,0.14]
Match quality	.	0.34	0.34	.	0.43	0.43
		[0.27,0.42]	[0.26,0.42]		[0.34,0.53]	[0.34,0.53]
Ref. point	0.97	0.35	0.36	1.01	0.28	0.31
	[0.71,1.24]	[0.03,0.67]	[0.03,0.69]	[0.75,1.26]	[-0.01,0.58]	[0.01,0.62]
Total error	0.76	0.68	0.67	0.76	0.69	0.67
	[0.60,0.93]	[0.52,0.84]	[0.50,0.84]	[0.62,0.90]	[0.55,0.82]	[0.53,0.81]

Notes: cells report the point estimate and 95% confidence intervals associated with the overall contribution of different expectational error components, as derived from the models in the respective columns of Appendix Table A6; error contributions are shrunk, as per equation (6).
Source: own estimates.

Table A6: Regression estimates of expectational error (last job)

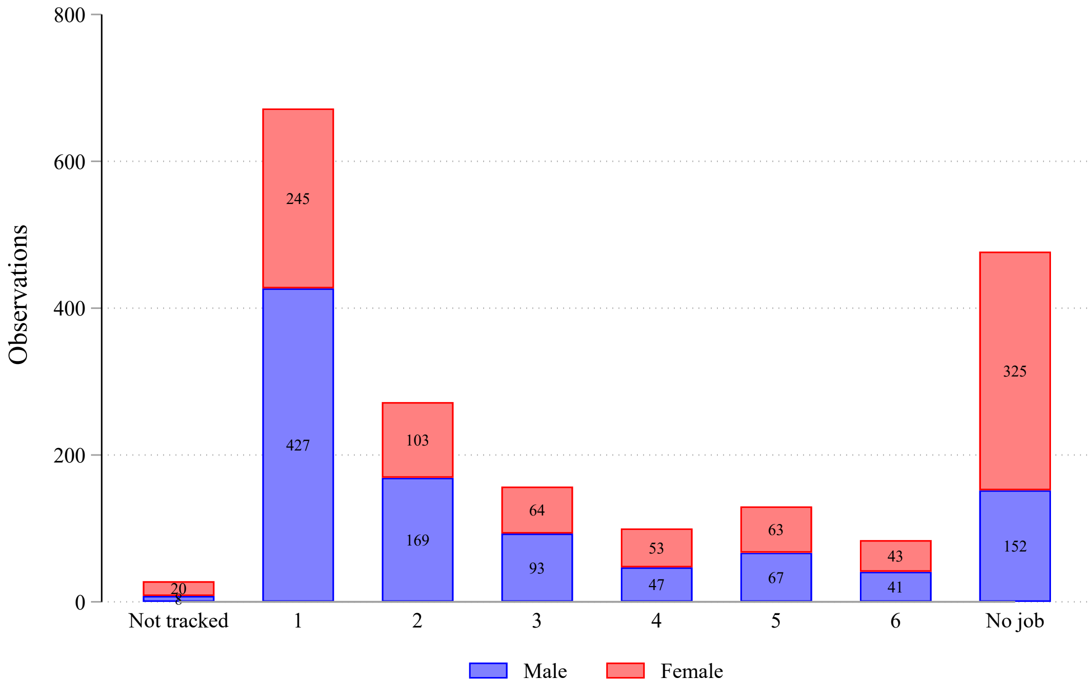
	(I) OLS			(II) Robust [M-estimator]		
	(a)	(b)	(c)	(a)	(b)	(c)
Constant	1.38***	0.63***	0.65***	1.41***	0.55***	0.58***
	(0.19)	(0.21)	(0.21)	(0.19)	(0.20)	(0.20)
Age	-0.01***	-0.01**	-0.01**	-0.01**	-0.01*	-0.01*
	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
Female	-0.14**	-0.07	-0.00	-0.12**	-0.04	0.05
	(0.06)	(0.06)	(0.09)	(0.05)	(0.05)	(0.09)
Private university	-0.09	-0.08	-0.09	-0.16***	-0.17***	-0.14*
	(0.07)	(0.06)	(0.08)	(0.06)	(0.06)	(0.07)
English proficiency	-0.24***	-0.23***	-0.23***	-0.26***	-0.23***	-0.26***
	(0.08)	(0.07)	(0.08)	(0.06)	(0.06)	(0.07)
Academic level (self)	-0.16***	-0.12***	-0.12***	-0.16***	-0.10**	-0.11**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Prev. internship	-0.06	-0.02	-0.02	-0.11**	-0.05	-0.06
	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)
Prev. work	0.05	0.03	0.03	0.03	0.02	-0.01
	(0.05)	(0.05)	(0.07)	(0.05)	(0.04)	(0.06)
Prev. work exp.	0.01	0.02***	0.02	0.01	0.02***	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Self/family employed	-0.02	0.18*	0.19*	-0.08	0.16*	0.17*
	(0.07)	(0.10)	(0.10)	(0.06)	(0.09)	(0.09)
Private services	-0.04	-0.20**	-0.19**	-0.04	-0.20**	-0.20**
	(0.08)	(0.09)	(0.09)	(0.08)	(0.09)	(0.09)
Lives in Sofala (Δ)		-0.36**	-0.37**		-0.23**	-0.24**
		(0.15)	(0.15)		(0.11)	(0.11)
Study unfinished (Δ)		-0.17***	-0.17***		-0.19***	-0.19***
		(0.05)	(0.05)		(0.04)	(0.04)
Works part time (Δ)		-0.21***	-0.20***		-0.22***	-0.22***
		(0.06)	(0.06)		(0.05)	(0.05)
Internship (Δ)		-0.43***	-0.41***		-0.52***	-0.52***
		(0.09)	(0.09)		(0.07)	(0.07)
Searching for work (Δ)		-0.22***	-0.23***		-0.22***	-0.23***
		(0.04)	(0.04)		(0.04)	(0.04)
Job unlike course (Δ)		-0.08*	-0.08*		-0.13***	-0.12***
		(0.05)	(0.05)		(0.04)	(0.04)
NGO employee (Δ)		0.22**	0.21**		0.25***	0.25***
		(0.09)	(0.09)		(0.08)	(0.08)
Self/family employed (Δ)		-0.16**	-0.17**		-0.17**	-0.17**
		(0.08)	(0.07)		(0.07)	(0.07)
Private services (Δ)		0.10	0.11		0.10	0.10
		(0.07)	(0.07)		(0.07)	(0.07)
Obs.	1,415	1,415	1,415	1,415	1,415	1,415
R^2	0.15	0.27	0.27	0.17	0.33	0.33
Control func. (pr.)			0.03			0.42

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 Additional figures

Figure B1: Observations, by round observed in first job



Source: own calculations.

Notes: bar values indicate the raw number of observations (unweighted), by follow-up survey round in which the participant first reports to have a job.

Figure B2: Main subcomponents of match quality error

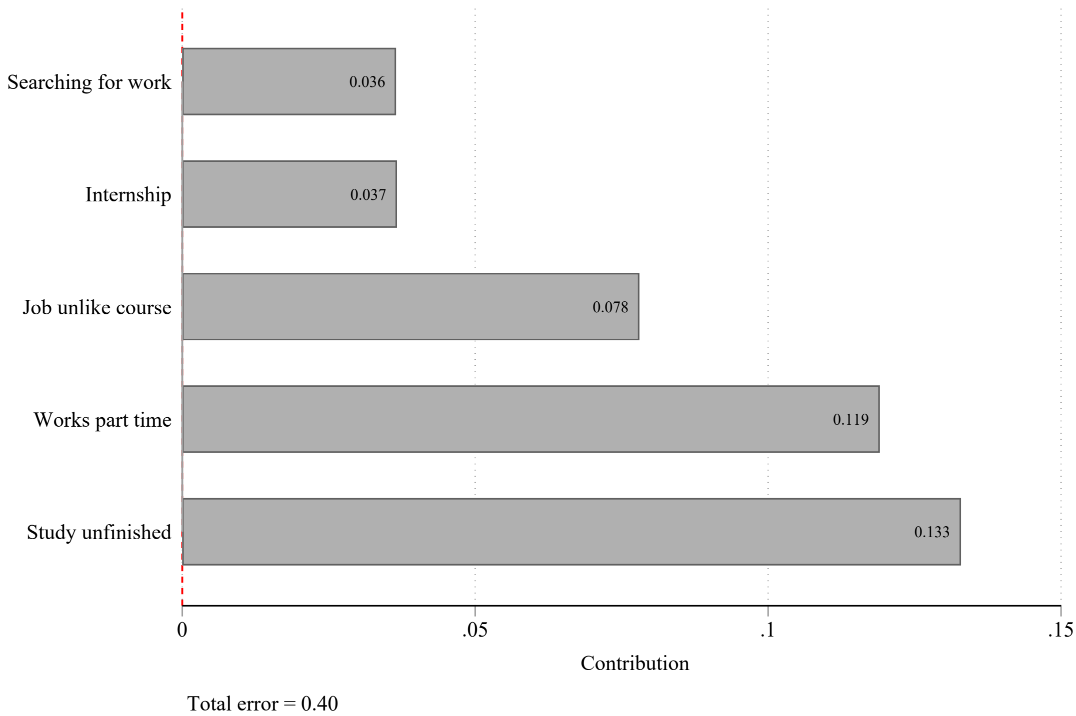
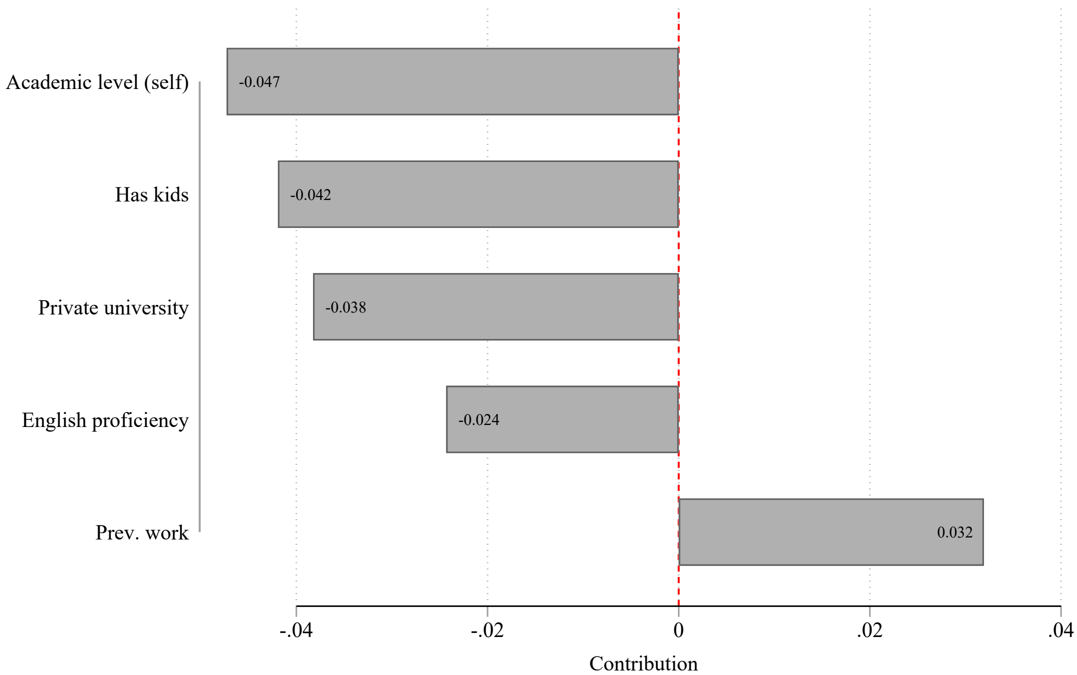
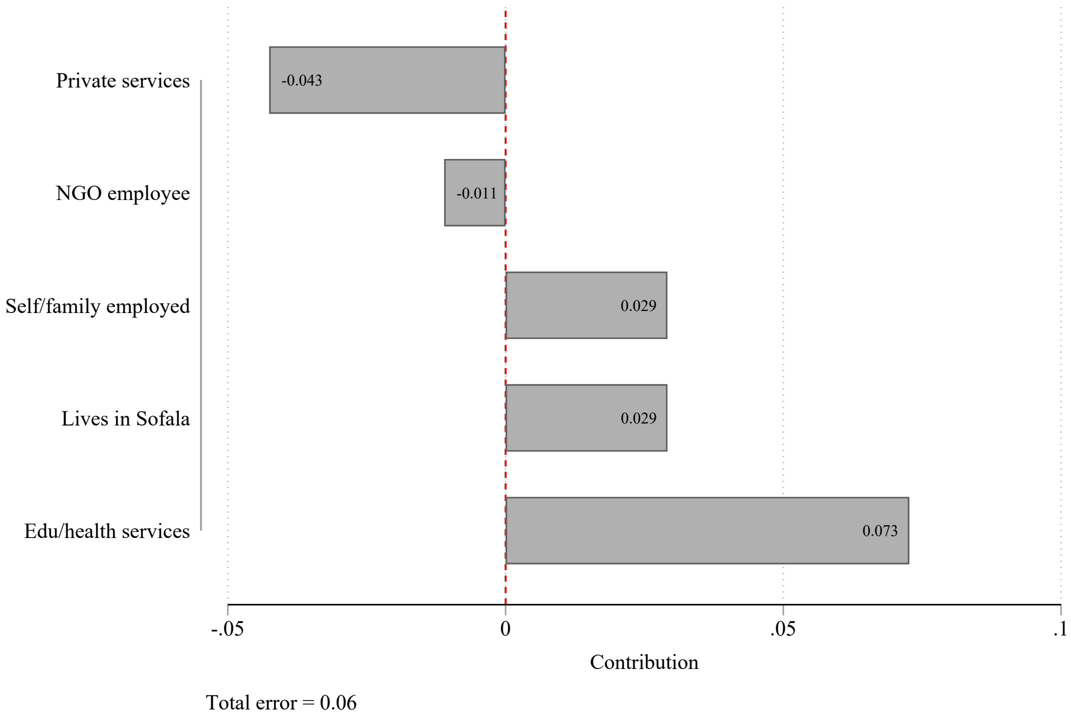


Figure B3: Main subcomponents of private information error



Total error = -0.18

Figure B4: Main subcomponents of public information error



C List of variables

Group		Variable
I	Individual attributes	Age in years
		Female
		Married
		Has kids
		Expected time to job
		English proficiency
		Family private sector job
		Family public sector job
		Family self-employed
		Female with kids
		First generation student
		Prev. work (dummy)
		Prev. work (length of time)
		Academic ability score
		Ravens score
		Academic level (self)
		Locus of control score
		Has adequate job info.
		Family job links
		Has job waiting
		Prev. internship
		Province of primary school (dummies)
		Relocated to uni.
		Received scholarship
		Education (study area)
		Humanities (study area)
		Natural Sciences (study area)
		Engineering (study area)
		Agriculture (study area)
		Health (study area)
		Private university
J	Job attributes	Lives in Sofala
		NGO employee
		Private sector employee
		Secondary sector
		Private services
		Edu/health services
		Self/family employed
M	Match quality	Time to job
		Internship
		Family job links

Group		Variable
		Lives outside Maputo/Sofala
		Lives in Sofala
		Study unfinished
		Temp. position
		Job unlike course
		Works part time
		Searching for work
		NGO employee
		Private sector employee
		Secondary sector
		Private services
		Edu/health services
		Self/family employed
R	Reference	Constant
		Round 2 (dummy)
		Round 2 (dummy)
		Round 3 (dummy)
		Round 4 (dummy)
		Round 5 (dummy)
		Round 6 (dummy)