Grover, Shweta; Sharma, Ajay

Working Paper

Education-occupation mismatch and dispersion in returns to education: Evidence from India

GLO Discussion Paper, No. 651

Provided in Cooperation with:
Global Labor Organization (GLO)

Education-occupation mismatch and dispersion in returns to education: Evidence from India

Shweta Grover
Indian Institute of Management Indore

Ajay Sharma
Indian Institute of Management Indore
Global Labor Organization

Abstract

Using a national level sample survey on labour market in India, we analyze the role of education-occupation (mis-)match (EOM) in explaining within-group dispersion in returns to education. Applying a double sample selection bias correction and Mincerian quantile wage regression estimation, the analysis reveals interesting findings. First, on average, overeducated workers suffer a wage penalty of seven percent and undereducated workers do not receive a wage reward as compared to their adequately educated counterparts. Second, the inclusion of match status reduces within-education group dispersion in returns. The finding highlights that ignoring EOM and thus, adopting a restrictive view of similarity across workers may lead to overestimation of the within-education group dispersion in returns. This study argues for focusing on EOM to increase both pecuniary and social benefits of education in terms of productivity gains and wages as well as to reduce wage dispersion.

Keywords: Education-occupation mismatch; Dispersion in returns to education; Wage dispersion; India; Quantile regression

JEL Code: I24, J24, J31

Corresponding Address:

Shweta Grover, Indian Institute of Management Indore, Rau-Pithampur Road, Prabandh Shikhar, Indore, (M.P) India-453556; e-mail: f15shwetag@iimidr.ac.in.

Ajay Sharma, Indian Institute of Managment Indore, Rau-Pithampur Road, Indore, (M.P.) India-453556; e-mail: ajays@iimidr.ac.in; ajaysharma87@gmail.com. Phone: +91-731-2439622.

1. Introduction

Understanding wage determination and resulting wage inequality/dispersion[1]remain one of the central themes of labour economics due to the interest of multiple agents *viz.*, workers, firms, and government. In wage dispersion literature, education is found to be one of the most consistent determinants of wage dispersion across and within economies (Alejos 2003; Dutta 2005; Kijima 2006; Krstic and Reilly 2003). Although it is understandable why workers with different education earn different returns, one unsolved puzzle in this literature relates to the heterogeneity in returns to the same level of education that is, within-education group dispersion in returns. One explanation for this phenomenon could be drawn from human capital theory (HCT) (Becker 1964). HCT asserts wages to be dependent on the combination of human capital characteristics and not on education alone. This implies that workers with a given education may experience dispersion in returns due to differences in their other human capital aspects such as experience in the labour market, on-the-job training, and so on. Another elucidation could be inferred from the job competition model (JCM) (Thurow 1975). JCM stresses on the job definition and its requirements as contributors to within-education group dispersion in returns. In other words, JCM argues that workers with the same level of education employed in different occupations would get different returns. Apart from worker and firm characteristics, researchers have examined various other channels which could partially explain this heterogeneity, for example, unemployment duration (Eckstein and Van den Berg 2007), the nature of job-search process and consequent labour market frictions (Eckstein and Van den Berg 2007; Postel–Vinay and Robin 2002), social networks and job-referrals (Montgomery 1991), institutional structure and policy changes (such as active labour market policies, trade and immigration policy, and so on) (Kierzenkowski and Koske 2012), and labour market discrimination (Becker 2010). In a similar spirit, this study contributes to this literature by providing an alternative explanation using the education-occupation (mis-)match (EOM) framework.

[1]Strictly speaking, the concepts of wage inequality and wage dispersion are slightly different (see, Salverda and Checchi (2014) for conceptual discussion). However, this study uses these terms interchangeably.

EOM exists when there is a misalignment between the attained years of education of an individual and required years of education by her occupation (Duncan and Hoffman 1981; Verdugo and Verdugo 1989). A person is said to be overeducated (undereducated) if her attained education is higher (lower) than the required education by her occupation. In contrast, a person is categorized as adequately educated when the attained and required education are aligned. A consistent finding in this literature is that overeducated workers endure significant wage penalties and undereducated workers receive considerable wage rewards as compared to their matched counterparts with the same level of education (Hartog 2000; Leuven and Oosterbeek 2011). This signals that match status (adequately educated, overeducated, or undereducated) in the labour market could explain a part of within-education group dispersion in returns by affecting returns to education (Martins and Pereira 2004). This study primarily explores whether and how much of the within-education group dispersion in returns couldbe attributed to EOM in the labour market. Thisis done by using the 2011-12 data available from India's National Sample Survey Office (NSSO) on the labour marketparticulars. The analysis is restricted to wage/salary employed individuals as income information is unavailable for self-employed individuals[2].

The main contributions of this study are to investigate (i) patterns of within-education group dispersion in returns and (ii) impact of EOM on within-education group dispersion in returns. The issue of within-education group dispersion in returns has received its due share of attention. Researchers document higher within-education dispersion in returns for tertiary-educated workers (Azam 2012) and a positive relationship between education and within-education dispersion in returns.[3] This relation is termed as inequality increasing effect of education (Budría and Telhado-Pereira 2011) and is found for several countries (see, Martins and Pereira (2004)for Europe and the United States; Tansel and Bodur (2012) for Turkey). On the other side, studies examining the relationship between EOM and within-education dispersion in returns are sparse. An exception to this is the notable study conducted by Green and Zhu (2010)in the case of Britain. The study focuses on

2 A detailed discussion related to the sample would follow in the data and descriptive statistics section of the paper.

3 One important point to note here is that in the relevant literature, researchers have used dispersion in returns as an indicator of dispersion in wages but this may not always be true. This is because wage dispersion considers the absolute wage differences while dispersion in returns consider the premium rate which market pays to get one higher level of education.

estimating the relationship between overtime changing levels of over-education and dispersion in returns to education. Green and Zhu (2010) find no significant dispersion in returns to education for matched male workers whereas there was notable fanning out of the average returns for the overall group (including both matched and mismatched workers). Apart from this, the study by Budría and Moro-Egido (2008) finds that EOM only when combined with skill mismatch leads to within-education dispersion. In this literature, the contribution of our study emerges from two fronts. On the one hand, our study sheds light on the incidence of within-education dispersion in returns using various measures of dispersion such as Kuznets ratio, range, Gini coefficient, coefficient of variation, and Lorenz curve. On the other hand, while focusing on the issue of heterogeneity in the returns to education (Henderson et al. 2011), this study also deals with the problem of sample selection bias (Heckman 1977) to estimate the unbiased and consistent estimates. In the empirical literature on EOM, only one of the problems is considered at a time. To address the heterogeneity issue, the study uses quantile regression (QR) estimation that examines the effect of education and EOM across the wage distribution. Further, to account for the sample selection bias, this study employs a double sample selection framework (Catsiapis and Robinson 1982) and considers two fundamental decisions – the decision to work (to be or not to be engaged in economic activity)[4] and the choice of economic activity status (wage/salaried employment or self-employment). In this process, this study contributes to the EOM literature by estimating returns to education and EOM at various percentiles.

The study could also aid policymakers in developing countries for understanding some of the key issues such as the role of human capital formation in the process of sustained economic growth and development. One major ongoing debate in developing countries is whetherthe available and limited resources should be channelised for universal primary education or should be targeted for tertiary education for a selected part of the population. Recent studies have argued for the latter case on account of the deeper role of tertiary education in improving the economic growth rate (Castelló-Climent and Mukhopadhyay 2013). The missing piece in this debate is the efficiency of the labour market in utilizing the workers' education. Thisis based on the argument that the productivity of an

[4] Economic activity is defined as an activity that results in the production of goods and services which, in turn, leads value addition to the national product (NSSO 2014).

individual is ascertained not only by her ability but also by the adequacy of her match to the job (Kieferand Devine 1991). Therefore, this study aims to shed light on the efficiency of higher education from the perspective of differential returns for matched and mismatched workers.

The salient findings from the study are as follows. First, returns to education increase with the level of education. Workers with graduation or above degrees earn the highest returns. Second, the analysis finds that, on average, overeducated workers suffer a wage penalty of around seven percent and undereducated workers do not receive a wage reward as compared to their adequately educated counterparts once we take account of human capital, job, and other personal and household characteristics. However, workers at the higher end of the wage distribution get rewarded for being undereducated. Further, the penalty for being overeducated decreases along with the wage distribution. This highlights that workers at the lower end of the wage distribution suffer higher wage losses if they are overeducated and earn negligible wage rewards if they are undereducated. Lastly, the study finds evidence that including EOM decreases the within-group dispersion in returns to education.

The rest of the study is structured as follows. Section 2 discusses the measures to estimate EOM, followed by the description of data in Section 3. Section 4 provides the estimation methodology. Section 5 presents the results and discussion. The final section presents the conclusion.

2. Approaches to measuringeducation-occupation (mis-)match

In the process of wage determination, the matching of attained education of a worker and required education by her occupation plays an important role (see, Leuven and Oosterbeek 2011). Thus, to identify whether a worker matches to her occupation, apart from attained education, the information on required education is also necessary. The relative importance of different agents in the matching process can lead to different approaches proposed in the literature to measure the required education.

When workers have more information to make well-informed decisions to choose their jobs and gauge their market wages, their self-assessment is considered appropriate. This approach is termed as workers' self-assessment (WA) in the EOM literature (Chevalier 2003; Duncan and Hoffman 1981). The inherent assumptions in the case of WA are that workers are well educated, and job-search mechanisms are well-defined, there is a prevalence of tight labour markets (number of jobs are more than number of workers)[5], and workers are well placedfor evaluating their jobs. In this method, individuals are asked about the educational requirements of their job and whether they are adequately educated for it. A fundamental criticism of WA approach is that workers may not be equipped with the expertise to give an unbiased opinion about their job requirements.

Keeping the limitations of WA in mind, the literature has proposed an alternative approach where job experts define the boundary of tasks to be performed at work and, in turn, provide a mechanism for wage determination. The proponents of this method assume that the tasks which need to be performed in an occupation are decisive factors for stating the job requirements and consequently the level of wages. This approach is called job analysis (JA) (Rumberger 1981). Given that experts are involved in the process of job descriptions, it is devoid of any individual biases and provides an objective measure of EOM. However, this method cannot be applied if jobs do not have well-defined requirements.

Apart from these two, a third approach tackles the question of matching workers to occupations using the perspective of the labour market and hence involves the interaction of demand and supply-side factors. This approach is called realized matches (RM) and hinges on the integral assumption that observable characteristics of workers and their revealed preferences through the choice of occupation are essential to EOM. In this method, a worker’s attained educationis compared against a statistical threshold, such as mean (Verdugo and Verdugo 1989) and mode (Kiker et al. 1997) for investigating whether she is adequately educated for a job. The statistical threshold acts as a proxy for required education due to the unavailability of an exogenous threshold. With the increase in the average level of education of the labour force, there is likely to be an upward bias in statistical threshold such as average education of workers in that occupation. This, however, highlights the advantage of this method that is, using the full educational distribution of workers in an occupation aids in estimating

[5] This assumption ensures that workers are working in a particular occupation by choice rather than due to shortage of jobs that suits their respective education.

the intrinsic requirements of an occupation existing at a point in time and thus, provides a 'realistic' picture of the incidence of EOM. This also helps in identifying not only the type of match status but also the extent of it. Further, given the ease of measurement and being less data demanding, it is suitable for cross-country and cross-sectional analysis of EOM.

Given the pros and cons of all three approaches, this study uses the RM method for the following reasons. First, the analysis in this study deals with a developing country where the informal sector is predominant (Ruppert Bulmer 2018), and thus jobs are either unstructured or lack specific requirements. Hence, WA and JA cannot serve as effective measures. Second, due to the less frequent revision of job requirements under JA method (Hartog 2000) as well as lack of availability of concordance of occupations between developed and developing countries (Balasubramanian 2016), it remains inapplicable for developing countries. Third, in the case of developing countries, WA surveys are scant and even when available cannot be used due to heterogeneity in the education distribution (such as quality) and micro-macro environment (such as slack labour markets). The only approach that remains applicable is RM. An advantage of using RM is that many studies in the literature have used this approach in the context of developed and developing countries (e.g., Blázquez and Rendon 2014; Haddad and Habibi 2017; Herrera-Idárraga et al. 2015; Sharma and Sharma 2017). This provides leverage to compare the findings of this study with the literature.

3. Data and descriptive statistics

This section discusses the data set and descriptive statistics to set the context for the study.

3.1. Source description

The data used in this study come from the latest employment and unemployment survey, 2011-12, collected by the National Sample Survey Office (NSSO). The survey is carried out in all the states and union territories of India, the exception being inaccessible places in parts of India and thus, is representative of India's labour market conditions. The survey covers 101,724 households (59,700 in rural areas and 42,024 in urban areas) comprising a total of 456,999 persons (280,763 in rural areas and 176,236 in urban areas). NSSO gives a sampling weight to each observation which makes the number of surveyed households and individuals equal to the number of households and the total population of India respectively. NSSO (2014) also provides the details of the sampling procedure for the survey. The NSSO survey contains detailed information about demographic variables such as age, gender, place of residence, and so on, together with the job-related characteristics such as the type of enterprise, number of workers, industry, and so on. The survey also collects data on wages for wage/salaried employees. The income information for self-employed workers, however, is not available.

The sample, in this study, is restricted to the working-age group that is, 15-59 years. This is to be consistent with the Government of India which considers age 15-59 years as economically active age group (NSSO 2014). Further, the main analysis in this study is conducted for wage/salaried employed workers for whom information on both education and occupation are available. Thus, the final sample comprises of 55,630 males and 14,467 females.

The next sub-section provides a detailed discussion on the measurement of EOM.

3.2. Education-occupation (mis-)match: definition and measurement

Based on the discussion in the previous section on measures to estimate EOM, this study adopts Realized Matches (RM) approach, particularly mean approach,[6]to measure the required education.

Definition: RM categorizes a person to be matched or mismatched by comparing her attained education with the mean years of education of workers in a given occupation (Verdugo and Verdugo 1989). Following Verdugo and Verdugo (1989), this study uses one standard deviation limit to ascertain the boundaries of the required years of education for an occupation. More precisely, suppose e_i represents years of attained education of an individual$i$, and $e_0$and s_0 are the mean and standard deviation of years of education respectively for her occupation, o. As per RM, she will be considered:

Overeducated if: $e_i > e_o + s_o$;

[6] The mode is not used to avoid the problem of multi-mode in case of some occupations.

Undereducated if: $e_i < e_0 - s_o$; and

Adequately educated if: $e_0 - s_o \leq e_i \leq e_o + s_o$

In particular, she is adequately educated, if her years of education lie between mean plus and minus one standard deviation threshold and overeducated (undereducated) if above (below) mean plus (minus) one standard deviation.

Measurement: As mentioned above, two inputs are required to use RM – occupation categories and years of education of the workers. Occupation of an individual is collected as per the National Classification of Occupation (NCO) 2004 three-digit codes using 111 occupation titles in NSSO survey data.[7] A full list of occupation titles is available in the report published by NSSO (2014).

The NCO code is available only for the employed (wage/salaried employed and self-employed) individuals. Further, this study does not include self-employed since defining required education is not only challenging but also lacks the theoretical justification for this group. This is because being self-employed in a particular occupation depends on a myriad of factors other than education such as available opportunities, risk-aversion, and so on. Therefore, the estimates of EOM are limited to wage/salaried employed workers. NSSO provides information on the general level of education. This study converts the levels of education into years of formal education following Duraisamy (2002)[8]. Next, the two-stage procedure is followed to capture EOM. In the first stage, required years of education are estimated. For that, the study calculates mean and standard deviation of years of education of individuals employed in a specific occupation category using sampling weights and then establishing the cut-offs of the mean plus and minus one standard deviation. In the second that is, the identification stage, individuals are categorized into adequately educated, under-educated, and overeducated workers using the required years of education as a threshold (Verdugo and Verdugo1989). The next sub-section provides estimates of EOM and within-education group wage dispersion.

[7] The number of observations in different occupation titles ranges from 14 to 36,041.

[8] No formal schooling corresponds to zero years of formal schooling, below primary corresponds to three years of formal education and completion of primary corresponds to five years of formal schooling. The corresponding years for middle, secondary, and higher secondary are eight, ten, and twelve respectively. Lastly, workers with graduate degree have been assigned 15 years of formal education and workers with postgraduate degree have been assigned 17 years of formal education.

3.3. Descriptive statistics

This sub-section provides the descriptive statistics on EOM and prevalence of within-education group dispersion in the Indian labour market.

In 2011-12, the total stock of matched workers was around 118 million. Despite having overall lower education levels as compared to other countries (Tilak 2018), the stock of overeducated workers (35 million) in India was significantly higher than the undereducated workers (26 million). Table 1 depicts the incidence of EOM at an overall level and by gender.

<INSERT TABLE 1 HERE>

Table 1 highlights that the incidence of over-education is higher than that of under-education in India. Further, over-education is a more common phenomenon among male rather than female workers. This is in contrast with earlier findings (Groot and Van Den Brink 2000). A plausible explanation could be that women being secondary earners choose not to work when they are not able to find an adequate job for their level of education. Hence, they are not a part of the employed population. For example, in India, around 73 percent of graduate and above females in the working-age group are either unemployed or out of the labour force. The corresponding proportion for males is only 12 percent. Besides, among graduates or above, the proportion of females is only 39 percent (that is, of the total graduates (males plus females), 39 percent are females). Thus, the other explanation could be attributed to the lower level of education among females as compared to males.[9]

Next, analysing average daily wages for the sample, the study finds that workers residing in urban areas (Rupees (Rs.) 238) fare better than workers in rural areas (Rs. 172).[10] The gender wage gap is also evident with male and female workers earning an average daily wage of Rs. 264 and Rs. 174 respectively. Table 2 shows average daily wages across different education groups by match status and reveals that undereducated workers earn the highest wages followed by adequately educated and

[9] In absolute terms, around 27 million women with graduation or above education are either unemployed or out-of-labour force. Further, 22 million women have graduate or above degree as compared to 61 million men.
[10] These are the nominal average daily wages that is, they are not adjusted for urban-rural price differences.

then overeducated workers (the exception being graduates or above).[11] The phenomenon represents a situation where workers with the same level of education earn different wages due to their employment in different occupations. Undereducated workers are employed in an occupation that requires more education as compared to their respective attained education and thus have a pay premium attached to it. In other words, they can be regarded as excellent performers (Verdugo and Verdugo 1989) or are lucky to get occupation premium. On the other hand, overeducated workers are generally employed in jobs which have lower mean education and consequently reflect low-paying jobs. Also, an increase in education does not necessarily lead to an increase in productivity. Therefore, lower wages earned by overeducated workers highlight that either they underutilize the skills or they possess lower skills (Mateos-Romero and del Mar Salinas-Jiménez 2015). In a nutshell, there are significant wage differences among adequately educated, overeducated, and undereducated workers at all levels of education groups. Hence, it indicates the presence of wage differential among workers similar in terms of education but different in match status.

<INSERT TABLE 2 HERE>

Table 3 exhibits wages at 90th, 50th, and 10th quantiles by education. The positive relationship between education and wages exists at all quantiles. However, a closer look indicates that moving up the ladder of education, the difference and ratio of wages at 90th and 10th percentile increases and hence confirms the 'inequality increasing effect of education' found by previous studies (Martins and Pereira 2004). One possible reason could be that with an increase in education, available spectrum of job opportunities widen. Hence, it leads to varied wage profiles.

<INSERT TABLE 3 HERE>

However, note that these figures do not take account of heterogeneity in the job and demographic characteristics. Therefore, this presents a partial picture. The empirical strategy used in this study and described in the following section addresses this concern.

[11] The plausible reason could be the lower representation of undereducated workers among graduates or above category. Undereducated, by definition, are workers working in an occupation that requires more education than their attained education. Therefore, graduates or above (which includes graduates, postgraduates or above) can be undereducated only when they are working in the occupations which requires postgraduate or above degree.

4. Estimation strategy

This section has two sub-parts. The first part explains the outcome and explanatory variables. The next part discusses the empirical model, first unveiling the econometric model followed by various empirical specifications to be used.

4.1. Variables used

The outcome variable is the logarithm of daily wages calculated as the sum of cash and in-kind emoluments. Since individuals may have a varied number of working days which can influence wages, all the comparisons in this study are made using daily wages. Further, the key interest variable is education category which is captured using the dummy for the attained level of education and categorized into no formal schooling, primary, middle, secondary, higher secondary, and graduates or above. Moreover, we also include three kinds of covariates.

First, comprises human capital variables, namely, age and age squared. The age of a worker is used as a proxy for her labour market experience. Also, the quadratic term for age allows for possible diminishing returns to experience. Most of the past studies use age minus years of education minus five as potential labour market experience. However, due to a significant proportion of workers with no formal education, the measure is rendered unsuitable for India (Goel 2017).

Second, this study includes indicators of job characteristics, namely, occupation category (legislators, senior officials and managers, professionals, associate professionals, clerks, service workers and market sales workers, skilled agricultural and fishery workers, craft and related trades workers, plant and machine operators and assemblers, and elementary occupations), broad industry groups (agriculture, manufacturing, construction, and services), the location of workplace (rural, urban, and no fixed location), enterprise type (proprietary, partnership, government, public/private, and other), firm size (less than 10 workers, 10-20, 20 and above, and unknown), and type of work contract (unwritten and written contract).

Lastly, the analysis includes a set of personal and household characteristics, namely, gender (male and female), marital status (unmarried, married, and others), the interaction of gender and marital status, social group (scheduled tribe, scheduled caste, other backward class and others), religion (Hindu, Muslim, Christian, and others), sector (rural and urban), and state of residence. Introducing dummy for state helps to control for state-level heterogeneity. These are the standard variables that have been used by previous studies as well (Agrawal 2012; Duraisamy 2002). Table 4 provides summary statistics for all the variables used in this study.

<INSERT TABLE 4 HERE>

4.2. Empirical model

Given that the purpose of this study is to ascertain whether EOM can explain the within-education group dispersion in returns, the following two-step methodology is employed. The first step measures the within-education group dispersion in returns taking account of human capital variables, job characteristics, and other personal and regional characteristics. The latter stage takes account of EOM to explore how the match status of workers affects the within-education group dispersion in returns.

Given the need of estimates at different percentiles to ascertain the covariates of dispersion in returns, the usual approach is to estimate a quantile regression (QR) model which is an extension of Mincerian (Mincer 1974)[12] wage equation in the context of distributional analysis. QR model was first introduced by Koenker and Basset (1978). The following equation represents QR:

$$y_{\theta i} = X_i \beta_\theta + \ e_{\theta i} with Quant_\theta(y_i|X_i) = X_i \beta_\theta \quad (1)$$

where β and X are the vectors of parameters and explanatory variables respectively. $Quant_\theta(y_i|X_i)$ represents the θ^{th} conditional quantile of y given X.[13] QR method involves minimizing the weighted

[12] See, Heckman et al. (2003) for detailed discussion on Mincerian wage equation.

[13]The θ^{th} regression quantile, $0 < \theta < 1$, is estimated by solving the following minimization problem: $\min_\beta \frac{1}{n} [\sum_{i:y_i \geq X_i\beta} \theta \ |y_i - X_i\beta| +$ $i{:}yi{<}Xi\beta(1{-}\theta)\ yi{-}Xi\beta$ (Buchinsky 1998). If we define 'check function,' $\rho\theta\varepsilon$ as: $\rho\theta\varepsilon = (\theta{-}1)\varepsilon, \ \ \varepsilon{<}0\theta\varepsilon, \qquad \varepsilon{\geq}0$; the minimization problem can be rewritten as $\min_\beta \frac{1}{n} \sum_{i=1}^{n} \rho_\theta(\varepsilon_{\theta i})$ (see Koenkar and Basset (1978)).

absolute value of the residuals. This can be done using linear programming methods where standard errors are obtained using bootstrap methods.

Using the above framework, Mincerian quantile wage equation to estimate the within-education group dispersion in returns is:

$$logw_{\theta i} = \beta_0 + \beta_{1\theta i,k} S_{i,k} + \xi_\theta X_i + e_{\theta i} \quad (2)$$

where the subscript θdenotes estimate at the θ^{th} conditional quantile. The dependent variable is the logarithm of daily wages.S_{ik} represent the dummy variables identifying the highest level of education (k) attained by an individual i. The vector of other covariates is denoted by X and e is the error term.

One of the critical issues in the Mincerian wage equation is that wages are not observed for all individuals. The truncated wage distribution, therefore, could lead to the problem of sample selection bias (Heckman 1977) which, in turn, results in biased and inconsistent estimates in the standard OLS as well as QR framework. Another issue which is specific to survey data in developing countries is the unavailability of earnings information for the self-employed individuals. However, the choice of self-employment versus wage employment is not random and thus, disregarding the possible sample selection into self-employment may again lead to sample selection bias (Dolton and Makepeace 1990).[14]

Given this, the study uses a double sample selection framework and considers two selection decisions – the decision to work (to be or not to be engaged in economic activity) and the choice of economic activity status (wage/salariedemployeeor self-employment). The two-stage estimation procedure suggested by Heckman (1977) is the widely accepted method to correct for sample selection bias. In the first stage, the participation equation is estimated to obtain the selection term captured by the inverse mills ratio[15]. In the later stage, the wage equation is estimated using a selection term as one of the covariates. This model works on the standard assumptions of ordinary least squares. However, in

[14] Standard models in the literature have considered only the employment selection, and thus ignored this aspect. A few notable exceptions are Agrawal and Agrawal (2018), Dolton and Makepeace (1990), among others.
[15] Inverse Mills ratio is defined as the ratio of the probability density function to the cumulative distribution function.

the case of QR, the bias term is of unknown form, and hence, the premise of normality of errors is usually rejected. This renders the Heckman method unsuitable for the analysis (Buchinsky 2002). Therefore, as suggested by Buchinsky (1998, 2002), the study employs the semi-non-parametric (SNP) correction method.[16] In the first step, the participation equations are estimated using SNP estimation:

$$Y_{1i}^{*} = z_{1i}\xi + u_{1i} \qquad (3.1)$$

$$Y_{2i}^{*} = z_{2i}\xi + u_{2i} \qquad (3.2)$$

where * indicates the unobserved variable. The dependent variable in equation (3.1) is a binary variable which takes a value of 1 when a person is employed and 0 otherwise, and the dependent variable in equation (3.2) is a binary variable which takes a value of 1 when a person is employed, and her wage is observed (wage/salaried employee) and 0 when a person is employed, but her wage is not observed (self-employed). Here,z_j, j =1,2 captures the observed variables and u_j, j =1,2 corresponds to the error term that captures the impact of unobserved variables. It is a prerequisite to identify at least one variable that does not affect the wages but influences the probability of participation, called as exclusion variable(s), to use this method. This implies that $z_{1i}$and z_{2i}should include at least one exclusion variable that influences the decision to work and choice of economic activity status respectively but not the wages. For equation (3.1), the study uses the number of dependent members (aged below six years or above 60 years) in a household[17] and household type and size as exclusion variables. The choice of these variables is consistent with the literature that suggests that family characteristics variables are appropriate exclusion variables for the choice of work (Agrawal 2012; Buchinsky 2002). Further, landholding by a household[18] and a dummy for having a hereditary vocational training are used as exclusion variables for equation (3.2). Hence, participation equations are estimated using human capital variables, other personal and regional characteristics, and respective exclusion variables. Job characteristics are not included while

[16] This is done by using snp command in STATA v15. See De Luca (2008) for its detailed application in STATA.
[17] The study considers absolute number of dependent members irrespective of the size of the household. This is because household size is controlled as a separate variable.
[18] Land holding is estimated by considering the maximum of land owned and land possessed.

estimating participation equations since these are only observed after a person is employed and thus, cannot be considered as determinants of choice of work or economic activity status.

In the second step, QR is estimated using inverse Mills ratios and its square as one of the independent variables.[19] Hence, equation (2) can be modified as follows:

$$logw_{\theta i} = \beta_0 + \beta_{1\theta i,k} S_{i,k} + \xi_\theta X_i + \delta_{\theta 1} Sel_{wc} + \delta_{\theta 2} Sel_{wc}^2 + \delta_{\theta 3} Sel_{we} + \delta_{\theta 4} Sel_{we}^2 + e_{\theta i} \qquad (4)$$

where Sel_{wc} and Sel_{we} refer to inverse mills ratio estimated respectively from participation equation (3.1) that is, the decision to work and participation equation (3.2) that is, the choice of economic activity status. Other variables are interpreted as before. This model is estimated for every decile from 10th to 90th percentile and the within-education dispersion in returns is represented by the differential returns spanning the 10th-90th percentiles.[20]

The next sub-section provides a detailed description of the empirical specifications used in the econometric analysis.

4.3. Empirical specifications

Given the near absence of a theoretical framework for examining the relationship between EOM and wages, the study adopts from theories of wage determination that is, human capital theory (HCT) (Becker 1964), job competition model (JCM) (Thurow 1975), and assignment model (Sattinger 1993) in this context. HCT affirms that individuals earn wages based on their respective productivity level, which is determined by the attained education, experience in the labour market, and so on, combined to be called as human capital. It contends that workers with similar human capital earn equal wages irrespective of their job characteristics. Contrary to HCT, JCM contends wages to be a function of job characteristics only. Hence, it argues that the job definition and its requirements are the main

[19] The study uses both mills ratio and its square term in the wage equation following Buchinsky (1998, 2002).
[20] We also estimated the results using returns at every fifth percentile from 5th to 95th percentile. The results were similar to the presented analysis.

contributors to the dispersion in returns among similar workers. Assignment model lies between HCT and JCM and argues wages to be dependent on both, human capital and job characteristics.

Keeping these in mind, the following empirical specifications are employed to answer the question in hand that is, what is the impact of education-occupation (mis-)match (EOM) on within-education group dispersion in returns?

The first empirical specification (Specf. 1) includes apart from education, other human capital variables, personal and regional characteristics and self-selection terms.

$$logw_{\theta i} = \beta_0 + \beta_{1\theta i,k} S_{i,k} + \xi_\theta Z_i + \delta_{\theta 1} Sel_{wc} + \delta_{\theta 2} Sel_{wc}^2 + \delta_{\theta 3} Sel_{we} + \delta_{\theta 4} Sel_{we}^2 + e_{\theta i} \quad (5)$$

Variables are interpreted as before. However, in this equation, to capture the sole effect of human capital characteristics on wages, job characteristics in are not included in Z_i.

The second empirical specification (Specf. 2) is consistent with the assignment model and includes both human capital and job characteristics.[21]

$$logw_{\theta i} = \beta_0 + \beta_{1\theta i,k} S_{i,k} + \xi_\theta X_i + \delta_{\theta 1} Sel_{wc} + \delta_{\theta 2} Sel_{wc}^2 + \delta_{\theta 3} Sel_{we} + \delta_{\theta 4} Sel_{we}^2 + e_{\theta i} \quad (6)$$

Variables are interpreted as before. Moreover, in this equation X_i includes all the three kinds of covariates that is, human capital, job, and other personal and regional characteristics. The study hypothesizes that equation (6) should exhibit lower dispersion in returns than the previous specification. This is because workers with the same education may witness dispersion in returns due to differences in their job characteristics. Hence, by taking account of both human capital and job characteristics, the above specification expands the similarity quotient of workers. Thus, it should reduce dispersion in returns among workers with similar education.

[21]To be consistent with job competition model, we also estimated the solitary impact of job characteristics on wages. For the sake of brevity, the JCM estimation and corresponding results are not provided, but are available on the request.

Subsequently, in the third empirical specification (Specf. 3), the study includes EOM variables to measure the impact of EOM on dispersion in returns to education. This is done using Verdugo and Verdugo (1989)'s procedure.[22]

$$logw_{\theta i} = \beta_0 + \beta_{1\theta i,k}S_{i,k} + \xi_{\theta}X_i + \beta_{3\theta i}D_i^o + \beta_{4\theta i}D_i^u + \delta_{\theta 1}Sel_{wc} + \delta_{\theta 2}Sel_{wc}^2 + \delta_{\theta 3}Sel_{we} + \delta_{\theta 4}Sel_{we}^2 + e_{\theta i} \quad (7)$$

where D_i^o and D_i^u are the dummies for overeducated and undereducated, respectively. Other variables are interpreted as before.

The study expects Specf. (3) to explain the within-education group dispersion in returns better. This is because categorizing the workers by match status further accentuates the criterion of similarity among workers. Workers have not only the same education and other characteristics but also the match status. Hence, by taking account of differential match status, the within-education group dispersion should be further explained.

The next section contains the detailed results.

5. Results and discussion

This section has three sub-parts. The first part discusses the ordinary least squares (OLS) results. The second part presents the quantile regression (QR) estimates. Finally, the last part answers the critical question raised in this study by measuring the differences in the within-education group dispersion in returns across empirical specifications.

For the sake of brevity, the estimates of only concerned variables are given. The estimates for the full set of variables are available in the annexure.

[22] Duncan and Hoffman (1981) provide another approach to estimate the impact of EOM on wages. The authors used continuous variable for education that is, years of education and divided this into two parts: required years of education and over/under years of education. However, the problem is including continuous years of education does not allow the consideration of level of education and hence fails to solve the purpose of this study which is to estimate the within-education group dispersion in returns.

5.1 Results from ordinary least square (OLS)

Table 5 presents OLS estimates of the wage equation for different specifications described in the previous section (Section 4.3). Specf. (1) depictsa positive association between level of education and wages (in line with Barrett et al. (1999)). In Specf. (2), even though the association between level of education and wages remains unchanged, the magnitude of coefficients reduces. A possible reason could be that the level of education and job characteristics are correlated and taking only one set of variables would lead to omitted variable bias and thus, may bias the estimates.

Analysing Specf. (3),which accounts for EOM, reveals that on average, overeducated workers suffer a wage penalty of around seven percent as compared to their adequately educated counterparts. The earlier studies have also found similar results for other countries (Hartog 2000; Leuven and Oosterbeek 2011). The finding could be due to the phenomenon that overeducated workers are generally employed in the jobs which have lower mean education and consequently reflect low-paying jobs. However, in contrast to conventional literature, the analysis reveals that undereducated workers do not receive any wage rewards. Further, Specf. (3) witnesses a considerable increase in returns to education as compared to Specf. (2). EOM being correlated to wages and also to education leads to omitted variable bias if ignored. As shown, there is a penalty for being overeducated. Therefore, the presence of overeducated workers brings down the average returns for a given level of education.

<INSERT TABLE 5 HERE>

Though informative, OLS does not unveil significant differences in the returns across the wage distribution. Therefore, recent literature makes use of quantile regression to uncover this dimension (e.g., Agrawal 2012; Azam 2012). The results for conditional quantile regression are presented in the next sub-section.

5.2. Results from quantile regression (QR)

The results for conditional quantile regression are provided in Table 6. The results are presented for specf. (1), (2),and (3) at 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentile. Also, to ease the comparison of estimates at different quantiles with the average returns, OLS estimates are reproduced in column 1.

<INSERT TABLE 6 HERE>

The following results are noticeable. First,consistent with the OLS estimates, there is a positive relationship between the level of education and associated returns, irrespective of the wage percentiles. Thus, it suggests that there are increasing additional returns to increase in education. Further, for any education level, there are increasing returns up to 75^{th} percentile. A plausible reason often quoted in the literature is that it could be indicative of an increase in ability and quality of education of workers inhigher-wage percentiles (Herrera-Idárraga et al. 2015).

Second, the reward associated with under-education exists for the workers in the higher percentiles. Further, the penalty associated with over-education decreases over the percentiles. This finding could be looked from the perspective of ability segments (McGuinness and Bennett 2007). McGuinness and Bennett (2007) categorize and associate the top end of wage distribution with the high-ability segments and vice-versa. Thus, workers tend to compensate for their over-education and complement their under-education with high ability. The opposite holds for workers in low-ability segments. The finding indicates that workers at the lower end of the wage distribution suffer the double penalty. On one side, they earn lower returns to education and on the other, they receive a higher penalty for being overeducated and negligible rewards for being undereducated.

The key takeaways from these results are that there is heterogeneity in returns to the same level of education, as well as in coefficients of over and under education indicators. The next sub-section does a more detailed analysis of this heterogeneity and its explanation. Also, wewill turn to answer the primary question of this study that is, whether EOM can explain within-education group dispersion in returns?

5.3. Education-occupation mismatch and within-education group dispersion in returns

The differences in the returns to education across the wage distribution indicate the presence of dispersion in the returns to the same level of education. This sub-section explores whether EOM affects this dispersion.

Table 7 presents the ratio of returns at 90th and 10th quantile (similar to Kuznets ratio: θ_{90}/θ_{10}) for different specifications. Note that it is not the absolute but excess dispersion in returns within the group compared to the reference group. For example, dispersion in middle level indicates an excess of dispersion in middle level over 'no formal schooling'. The study further divides θ_{90}/θ_{10} into θ_{90}/θ_{50} and θ_{50}/θ_{10} percentile to analyse the top-end and low-end disparities respectively. The study observes that relatively large disparities in returns take place mainly in the lower half of the distribution for all levels of education for Specf.1 and Specf. 2. Thisis indicated by higher θ_{50}/θ_{10} ratio as compared to θ_{90}/θ_{50} ratio. The flat returns at the top end of the wage distribution are an obvious reason for this finding. However, the relationship reverses once we take account of EOM. Specf. 3 observes higher dispersion in the upper half of the distribution. The reason could be as follows. At lower percentiles, the penalty for being overeducated is substantially higher than at higher percentiles. Therefore, not taking note of EOM leads to higher changes in returns to education at lower percentiles which in turn, leads to higher dispersion at the lower end of the distribution in case of Specf. 1 and Specf. 2. Further, comparing θ_{90}/θ_{10} across specifications reveals that Specf. 3 has the lowest dispersion in returns to the middle level, secondary, and higher secondary.

<INSERT TABLE 7 HERE>

The similar analysis is conducted using the difference between the returns at 90th and 10th quantile (Range: θ_{90} - θ_{10}) for different specifications (Table 8). The results are aligned with the earlier findings. In contrast to Specf. 1 andSpecf. 2, Specf. 3 observes higher dispersion at the top end of the wage distributions. Further, Specf. 3 exhibits the lowest dispersion in returns to secondary and higher secondary.

<INSERT TABLE 8 HERE>

Next, the study analyses within-education dispersion in returns using the coefficient of variation, and Gini coefficient and Lorenz curve (Cowell 2011). However, there are some caveats that need to be kept in mind while analysing the results. First, these methods require the whole distribution to summarize the dispersion. But this study uses returns across deciles and not the entire distribution and therefore, results can only be considered as indicative of true incidence of within-education dispersion in returns. Second, the use of these methods in estimating dispersion in returns is not very common in the education literature. Thus, it limits our ability to compare the results of the present study with the existing literature. Given these caveats, Table 9 presents the Gini coefficient for the returns to education across the wage distribution. Table 9 reveals that for the middle, secondary, and higher secondary, Specf. (3) witnesses lower dispersion in returns to education as compared to Specf. (2).

<INSERT TABLE 9 HERE>

Also, using the coefficient of variation as a measure of inequality, the study finds that there is a significant decline in the coefficient of variation for the middle, secondary, and higher secondary after taking account of EOM (Table 10). For primary, there is an increase in the coefficient of variation moving from specf. (2) to (3). However, for graduates or above, there is no change in the dispersion in returns.

<INSERT TABLE 10 HERE>

Figure 1 presents the Lorenz curve for returns to education as per Specf. (2) andSpecf.(3). The line at the 45° angle indicates perfect equality in returns to education, while the other lines show the actual distribution of returns to education as per both the specifications. The further away from the diagonal, the more unequal the returns to education or, in other words, there is higher dispersion in returns to education. The study finds that except for primary education, in all the other education groups, Specf. (3) exhibits either lower or equal dispersion in returns as compared to Specf. (2).

<INSERT FIGURE 1 HERE>

Therefore, contradictory to earlier studies that have found inequality increasing effect of education (e.g., Azam 2012; Martins and Pereira 2004), the present study does not find a positive relationship between education and within-education dispersion in returns across all the measures of inequality exception being range. Omitted variables such as information on occupation, and EOM in the previous studies may be responsible for the deviation in the results or it could also be due to the use of only one measure of dispersion. By the observation of higher wage inequality amongst higher educated workers, Azam (2012) warned that wage inequality in urban India would increase with the increase in the proportion of people getting a higher education. This study, on the other hand, claims that differences in the job characteristics and in the match status could be responsible for the within-education group dispersion in returns.

To summarize, as hypothesized, EOM does explain some part of the within-education group dispersion in returns. The finding could be owed to the following reasons. First, wages are not decided individually by education or occupation but by a combination of these two. Depending on the extent and magnitude of EOM for a particular education group, the degree of within-group dispersion in returns may vary. Hence, workers with the same level of education may receive higher or lower returns to education than average depending on their alignment with the respective occupation which leads to within-education group dispersion in returns. Therefore, taking account of EOM succinct match type in a particular education group and consequently leads to a better explanation of within-group dispersion. Second, returns to education comprises of returns to two segments: required education and over/under education. While the returns to former are always positive, the latter may lead to differential returns and thus can lead workers with the same education to command different returns. Thus, education in itself does not lead to within-group dispersion in returns, but the differences in the match status among workers with the same education result in the varied return profiles. This indicates that EOM is an important factor that aids in explaining the within-education group dispersion. EOM does not only lead to varied returns between education groups but also dispersion in returns for a particular education group. The results are crucial since they highlight that

EOM needs to be corrected if policymakers intend to reduce within-education group dispersion in returns, as well as need to estimate correct returns to education level as inputs in policies.

6. Conclusion

The study empirically examines the relevance of education-occupation (mis-)match in explaining heterogeneity in returns to the same level of education. Through EOM, the study captures the interplay between workers and job characteristics in understanding within-group dispersion in returns to education. The analysis is done using data from India's National Sample Survey Office (NSSO) on the labour market particulars. The analysis is restricted to wage/salary employed individuals as income information is unavailable for self-employed individuals.The main findings can be summarized as follows. First, on average, overeducated workers suffer a wage penalty of around seven percent and undereducated workers do not receive a wage reward as compared to their adequately educated counterparts. Second, the study provides estimates for returns to education while taking account of double sample selection bias. This is a novel contribution. Last and the most important finding is that the inclusion of match status affects within-education group dispersion in returns. The finding highlights that ignoring EOM and thus, adopting a restrictive view of similarity across workers may lead to overestimation of the within-education group dispersion in returns. Also, contrary to previous studies, this study does not find inequality increasingeffect of education. This indicates that capturing the impact of isolated labour market aspects presents the partial picture of dispersion in returns across similarly educated workers.

The study highlights that policymakers interested in reducing wage inequality should pay attention to a mismatch between workers and firm characteristics. This is crucial to harness the wage benefits of education. This is depicted by the higher returns to education after taking account of EOM. Hence, the varied match type will bring down the average returns to a particular education group. From the workers' perspective, there should be labour market institutions facilitating the proper matching of skills and education to the requirement of occupation. Such institutions should not just focus on providing employment but decent and well-matched. A failure to do so can lead to hampering of prospects in terms of lower wages, limited upward career mobility choices, and lower satisfaction. Besides, when firms hire mismatched workers especially overeducated workers, given the wage penalty and job dissatisfaction due to wage dispersion (Fleming andKler 2008), the attrition is likely to be higher. This leads to an increase in the cost in terms of training, hiring, and so on. Providing excessive education without considering the demand side can lead to potential productivity losses especially based on pecuniary returns and barring the social returns to education. This is a severe issue especially for developing countries where resources are even scarce as compared to the developed countries.

These results could also be used for policymaking in developing countries. One such issue is the long-standing debate of choice between universal primary education versus selected population with tertiary education in accelerating and sustaining the economic growth when resources are limited. There is supporting evidence towards the latter (Castelló-Climent and Mukhopadhyay2013). Hence, it supports the view that higher education is the key to growth. The results in this study further substantiate and complement these findings. The study argues that providing higher education and facilitating the adequate match between workers' education and required education by the occupation will lead to higher returns and therefore accentuates the pecuniary benefits of education. However, focusing only on one aspect that is, providing higher education or creating education-intensive jobs may not only lead to loss of productivity but also lower wages. In a nutshell, the study argues that providing higher education would fetch higher growth if the available job opportunities are well matched to the workers. Therefore, in developing countries, where resources are scarce, the adequate balance between providing higher education and creating education-intensive jobs should be maintained.

Although the study provides a rich description of within-education group dispersion in returns, the limitations are inevitable. The unavailability of information such as quality of schooling, the field of training, cognitive and non-cognitive skills, and so on limits the scope of the study to differentiate workers within the same education group. Another dimension which can further enrich the study is to include spatial aspects of the labour market. The reason is that the probability of employment and consequent match status is affected by the prevailing conditions in the local labour market. When the person chooses her job or occupation, the physical limits of the labour market plays a crucial role.

To capture the dynamic aspect of dispersion in returns in the labour market, the natural extension of this study would be to conduct the inter-temporal analysis. This would help in understanding the changing dynamics of educational inequality and its long- term relationship with wage inequality using the current EOM framework.

References

Agrawal, T. (2012). Returns to education in India: Some recent evidence. Journal of Quantitative Economics, 10, 131–151.

Agrawal, T., & Agrawal, A. (2018). Who gains more from education? A comparative analysis of business, farm and wage workers in India. The Journal of Development Studies, 1-18.

Alejos, L. A. (2003). Contribution of the determinants of income inequality in Guatemala.

Azam, M. (2012). Changes in wage structure in urban India, 1983–2004: A quantile regression decomposition. World Development, 40(6), 1135-1150.

Balasubramanian, I. (2016). Local skill concentrations and district employment growth: a simultaneous equation approach for India. The Annals of Regional Science, 56(2), 491-511.

Barrett, A., Callan, T., & Nolan, B. (1999). Rising wage inequality, returns to education and labour market institutions: evidence from Ireland. British Journal of Industrial Relations, 37(1), 77-100.

Becker, G. S. (1964). Human capital theory. Columbia, New York, 1964.

Becker, G. S. (2010). The economics of discrimination. University of Chicago press.

Blázquez, M., & Rendon, S. (2014). Over-education in multilingual economies: Evidence from Catalonia. International Migration, 52(6), 149-164.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. Journal of Applied Econometrics, 1-30.

Buchinsky, M. (2002). Quantile regression with sample selection: Estimating women's return to education in the US. In Economic Applications of Quantile Regression (pp. 87-113). Physica, Heidelberg.

Budría, S., & Moro-Egido, A. I. (2008). Education, educational mismatch, and wage inequality: Evidence for Spain. Economics of Education Review, 27(3), 332-341.

Budría, S., & Telhado-Pereira, P. (2011). Educational qualifications and wage inequality: evidence for Europe. Revista de Economía Aplicada, 19(56), 5.

Castelló-Climent, A., & Mukhopadhyay, A. (2013). Mass education or a minority well educated elite in the process of growth: the case of India. Journal of Development Economics, 105, 303-320.

Catsiapis, G., & Robinson, C. (1982). Sample selection bias with multiple selection rules: An application to student aid grants. Journal of Econometrics, 18(3), 351-368.

Chevalier, A. (2003). Measuring over-education. Economica, 70(279), 509-531.

Cowell, F. (2011). *Measuring inequality*. Oxford University Press.

De Luca, G. (2008). SNP and SML estimation of univariate and bivariate binary-choice models. Stata Journal, 8(2), 190.

Dolton, P. J., & Makepeace, G. H. (1990). Self-employment among graduates. Bulletin of Economic Research, 42(1), 35-54.

Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. Economics of Education Review, 1(1), 75-86.

Duraisamy, P. (2002). Changes in returns to education in India, 1983–94: by gender, age-cohort and location. Economics of Education Review, 21(6), 609-622.

Dutta, P. V. (2005). Accounting for wage inequality in India. Indian Journal of Labour Economics, 48(2), 273-295.

Eckstein, Z., & Van den Berg, G. J. (2007). Empirical labor search: A survey. Journal of Econometrics, 136(2), 531-564.

Fleming, C. M., & Kler, P. (2008). I'm too clever for this job: a bivariate probit analysis on overeducation and job satisfaction in Australia. Applied Economics, 40(9), 1123-1138.

Goel, M. (2017). Inequality between and within skill groups: The curious case of India. World Development, 93, 153-176.

Green, F., & Zhu, Y. (2010). Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education. Oxford Economic Papers, 62(4), 740-763.

Groot, W., & Van Den Brink, H. M. (2000). Overeducation in the labor market: a meta-analysis. Economics of Education Review, 19(2), 149-158.

Haddad, G. K., & Habibi, N. (2017). Vertical skill mismatch and wage consequences in low-skilled jobs: Evidence from Iran. International Labour Review, 156(1), 45-72.

Hartog, J. (2000). Over-education and earnings: where are we, where should we go?. Economics of Education Review, 19(2), 131-147.

Heckman, J. J. (1977). Sample selection bias as a specification error (with an application to the estimation of labor supply functions).

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2003). Fifty years of Mincer earnings regressions (No. w9732). National Bureau of Economic Research.

Henderson, D. J., Polachek, S. W., & Wang, L. (2011). Heterogeneity in schooling rates of return. Economics of Education Review, 30(6), 1202-1214.

Herrera-Idárraga, P., López-Bazo, E., & Motellón, E. (2015). Double penalty in returns to education: informality and educational mismatch in the Colombian labour market. The Journal of Development Studies, 51(12), 1683-1701.

Kiefer, N. M., & Devine, T. J. (1991). Empirical labor economics: the search approach. Oxford University Press on Demand.

Kierzenkowski, R., & Koske, I. (2012). Less income inequality and more growth–are they compatible? Part 8. The drivers of labour income inequality–a literature review.

Kijima, Y. (2006). Why did wage inequality increase? Evidence from urban India 1983–99. Journal of Development Economics, 81(1), 97-117.

Kiker, B. F., Santos, M. C., & De Oliveira, M. M. (1997). Overeducation and undereducation: evidence for Portugal. Economics of Education Review, 16(2), 111-125.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. Econometrica. 33-50.

Krstic, G., & Reilly, B. (2003). An Anatomy of Labour Market Earnings Inequality in Serbia-1995 to 2000. IMF Country Report, 5, 13.

Leuven, E., & Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. In Handbook of the Economics of Education (Vol. 4, pp. 283-326). Elsevier.

Martins, P. S., & Pereira, P. T. (2004). Does education reduce wage inequality? Quantile regression evidence from 16 countries. Labour Economics, 11(3), 355-371.

Mateos-Romero, L., & del Mar Salinas-Jiménez, M. (2015). Skills heterogeneity among graduate workers: real and apparent overeducation in the Spanish labor market. Social Indicators Research, 1-18.

McGuinness, S., & Bennett, J. (2007). Overeducation in the graduate labour market: A quantile regression approach. Economics of Education Review, 26(5), 521-531.

Mincer, J. (1974). Schooling, experience, and earnings. Human Behavior & Social Institutions No. 2.

Montgomery, J. D. (1991). Equilibrium wage dispersion and interindustry wage differentials. The Quarterly Journal of Economics, 106(1), 163-179.

National Sample Survey Office (NSSO) (2014). "Employment and unemployment situation in India". Report No. 554 (68/10/1). New Delhi: Government of India.

Postel–Vinay, F., & Robin, J. M. (2002). Equilibrium wage dispersion with worker and employer heterogeneity. Econometrica, 70(6), 2295-2350.

Rumberger, R. W. (1981). The rising incidence of overeducation in the US labor market. Economics of Education Review, 1(3), 293-314.

Ruppert Bulmer, E. (2018). Defining informality vs mitigating its negative effects. IZA World of Labor:442. doi: 10.15185/izawol.442

Salverda, W., & Checchi, D. (2015). Labor market institutions and the dispersion of wage earnings. In Handbook of Income Distribution (Vol. 2, pp. 1535-1727). Elsevier.

Sattinger, M. (1993). Assignment models of the distribution of earnings. Journal of Economic Literature, 31(2), 831-880.

Sharma, S., & Sharma, P. (2017). Educational mismatch and its impact on earnings: evidence from Indian labour market. International Journal of Social Economics, 44(12), 1778-1795.

Tansel, A., & Bodur, F. B. (2012). Wage inequality and returns to education in Turkey: A quantile regression analysis. Review of Development Economics, 16(1), 107-121.

Thurow, L. C. (1975). Generating inequality. Basic books.

Tilak, J. B. (2018). The political economy of education in India. In Education and Development in India (pp. 33-84). Palgrave Macmillan, Singapore.

Verdugo, R. R., & Verdugo, N. T. (1989). The impact of surplus schooling on earnings: Some additional findings. Journal of Human Resources, 629-643.

Tables

Table 1: Education and occupation (mis-)match for wage/salaried employed workers in the working-age group (15-59 years): overall and by gender (in percentage)

Match Type	Overall	Gender	
		Male	Female
Undereducated	14.86	15.15	13.84
Adequately educated	65.73	63.63	73.09
Overeducated	19.41	21.22	13.07

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.
Note: Sampling weights have been used.

Table 2: Average daily wages (in Indian rupees) for wage/salaried employed workers in the working-age group(15-59 years) across match status by education

Education level	Undereducated	Adequately Educated	Overeducated
No Formal Schooling	159	116	-
	(39.65)	(60.35)	(0.00)
Primary or Below	202	152	-
	(7.69)	(92.31)	(0.00)
Middle	285	208	147
	(3.91)	(53.21)	(42.88)
Secondary	456	285	178
	(7.08)	(43.71)	(49.21)
Higher Secondary	659	406	229
	(5.29)	(56.46)	(38.25)
Graduate and Above	589	744	549
	(0.61)	(67.75)	(31.64)

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.
Note: (i) Sampling weights have been used.
(ii) Numbers in parenthesis indicate proportion of workers in that category.

Table 3: Daily wages (in Indian rupees) for wage/salaried employed workers in the working-age group (15-59 years) across education

	Quantile			Ratio		
	10th	50th	90th	50/10	90/50	90/10
Education:						
No Formal Schooling	69	120	205	1.74	1.71	2.97
Primary or Below	71	134	250	1.89	1.87	3.52
Middle	80	150	350	1.88	2.33	4.38
Secondary	100	179	500	1.79	2.79	5.00
Higher Secondary	100	250	752	2.50	3.01	7.52
Graduate and Above	143	543	1314	3.80	2.42	9.19

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.
Note: Sampling weights have been used.

Table 4: Summary statistics for wage/salaried employed workers in the working-age group (15-59 years): by match status

Variables		Overall	Match Status		
			Undereducated	Adequately Educated	Overeducated
Education	No formal schooling	0.19	0.66	0.13	-
	Primary	0.20	0.13	0.28	-
	Middle	0.17	0.07	0.16	0.28
	Secondary	0.13	0.09	0.10	0.26
	Higher Secondary	0.12	0.04	0.12	0.18
	Graduate or above	0.19	0.01	0.21	0.28
Mean Age (in years)		36.05	38.40	36.36	32.90
Gender	Male	0.79	0.80	0.78	0.85
	Female	0.21	0.20	0.22	0.15
Marital status	Never married	0.20	0.14	0.20	0.28
	Currently married	0.75	0.78	0.76	0.70
	Others	0.05	0.08	0.05	0.02
Social group	Scheduled tribe	0.14	0.15	0.14	0.11
	Scheduled caste	0.21	0.25	0.19	0.20
	Other backward class	0.37	0.37	0.37	0.39
	Others	0.29	0.24	0.30	0.29
Religion	Hinduism	0.77	0.75	0.77	0.80
	Islam	0.12	0.17	0.11	0.10
	Christianity	0.07	0.04	0.08	0.07
	Others	0.04	0.04	0.04	0.04
Sector	Rural	0.55	0.55	0.54	0.58
	Urban	0.45	0.45	0.46	0.42
Occupation	Legislators, Senior Officials and Managers	0.02	0.02	0.03	0.00
	Professionals	0.08	0.07	0.09	0.02
	Associate Professionals	0.11	0.09	0.12	0.10

Variables		Overall	Match Status		
			Undereducated	Adequately Educated	Overeducated
	Clerks	0.06	0.05	0.07	0.04
	Service Workers and Shop and Market Sales Workers	0.10	0.11	0.11	0.08
	Skilled Agricultural and Fishery Workers	0.01	0.01	0.01	0.02
	Craft and Related Trades Workers	0.17	0.22	0.16	0.17
	Plant and Machine Operators and Assemblers	0.08	0.09	0.09	0.05
	Elementary Occupations	0.36	0.34	0.32	0.52
Location of Work	Rural	0.44	0.50	0.42	0.45
	Urban	0.53	0.46	0.55	0.52
	No-fixed location	0.03	0.04	0.03	0.02
Type of job contract	Unwritten	0.71	0.82	0.67	0.74
	Written	0.29	0.18	0.33	0.26
Enterprise Type	Proprietary	0.44	0.55	0.40	0.45
	Partnership	0.02	0.02	0.02	0.02
	Government/Public Sector	0.33	0.23	0.37	0.28
	Public/Private Limited Company	0.10	0.07	0.10	0.13
	Others	0.11	0.12	0.10	0.11
Number of Workers	Less than 10	0.53	0.59	0.51	0.53
	10-19	0.13	0.11	0.13	0.12
	Above 19	0.26	0.20	0.27	0.26
	Not Known	0.09	0.10	0.09	0.09

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.

Table 5: Returns to education and EOM for wage/salaried employed workers in the working-age group (15-59 years) – Ordinary Least Square (OLS) estimates

Explanatory Variables	Outcome Variable: Logarithm of Daily Wages		
	(1)	(2)	(3)
	Specf. 1	Specf. 2	Specf. 3
Education (Base Cat.: No Formal Schooling)			
Primary	0.118***	0.065***	0.072***
	(0.010)	(0.010)	(0.015)
Middle	0.216***	0.101***	0.129***
	(0.012)	(0.012)	(0.018)
Secondary	0.412***	0.186***	0.230***
	(0.014)	(0.013)	(0.020)
Higher Secondary	0.768***	0.358***	0.410***
	(0.015)	(0.015)	(0.023)
Graduate or Above	1.227***	0.586***	0.654***
	(0.015)	(0.017)	(0.027)
Match Status (Base Cat.: Adequately Educated)			
Undereducated			0.004
			(0.014)
Overeducated			-0.072***
			(0.011)
Number of Observations	65,792	56,168	56,142
R-Squared	0.390	0.475	0.476

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.

Note: (i) *** signals significant at 1% level, ** signals significant at 5% level, and * signals significant at 10% level.
(ii) Robust standard errors are given in parenthesis.
(iii) Specf. 1 includes human capital and self-selection terms, specf. 2 includes human capital, job characteristics and self-selection terms, and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.

Table 6: Returns to education and EOM for wage/salaried employed workers in the working-age group (15-59 years)- OLS and QR

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
Specification 1						
Education (Base Cat.: No Formal Schooling)						
Primary	0.118***	0.036***	0.054***	0.098***	0.125***	0.155***
	(0.010)	(0.012)	(0.008)	(0.007)	(0.008)	(0.010)
Middle	0.216***	0.073***	0.099***	0.176***	0.245***	0.313***
	(0.012)	(0.011)	(0.009)	(0.007)	(0.009)	(0.012)
Secondary	0.412***	0.183***	0.218***	0.345***	0.530***	0.625***
	(0.014)	(0.014)	(0.009)	(0.012)	(0.012)	(0.014)
Higher Secondary	0.768***	0.353***	0.454***	0.776***	0.980***	0.982***
	(0.015)	(0.013)	(0.013)	(0.012)	(0.011)	(0.012)
Graduate or Above	1.227***	0.719***	1.037***	1.324***	1.398***	1.405***
	(0.015)	(0.015)	(0.016)	(0.012)	(0.011)	(0.013)
Specification 2						
Education (Base Cat.: No Formal Schooling)						
Primary	0.065***	0.048***	0.048***	0.061***	0.069***	0.064***
	(0.010)	(0.018)	(0.010)	(0.010)	(0.009)	(0.014)
Middle	0.101***	0.062***	0.063***	0.096***	0.116***	0.115***
	(0.012)	(0.016)	(0.010)	(0.009)	(0.010)	(0.014)
Secondary	0.186***	0.134***	0.133***	0.169***	0.202***	0.206***
	(0.013)	(0.017)	(0.013)	(0.011)	(0.014)	(0.014)
Higher Secondary	0.358***	0.271***	0.289***	0.319***	0.332***	0.327***
	(0.015)	(0.019)	(0.014)	(0.014)	(0.014)	(0.015)
Graduate or Above	0.586***	0.524***	0.556***	0.555***	0.536***	0.527***
	(0.017)	(0.025)	(0.018)	(0.014)	(0.013)	(0.016)
Specification 3						
Education (Base Cat.: No Formal Schooling)						

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
Primary	0.072***	0.053**	0.046***	0.062***	0.093***	0.092***
	(0.015)	(0.024)	(0.017)	(0.014)	(0.013)	(0.015)
Middle	0.129***	0.093***	0.091***	0.112***	0.154***	0.158***
	(0.018)	(0.027)	(0.020)	(0.015)	(0.016)	(0.018)
Secondary	0.230***	0.193***	0.182***	0.200***	0.245***	0.260***
	(0.020)	(0.031)	(0.020)	(0.017)	(0.017)	(0.023)
Higher Secondary	0.410***	0.347***	0.345***	0.352***	0.389***	0.389***
	(0.023)	(0.036)	(0.023)	(0.016)	(0.021)	(0.026)
Graduate or Above	0.654***	0.620***	0.633***	0.596***	0.605***	0.608***
	(0.027)	(0.042)	(0.028)	(0.020)	(0.024)	(0.031)
Match Status (Base Cat.: Adequately Educated)						
Undereducated	0.004	-0.007	-0.010	-0.002	0.026**	0.031**
	(0.014)	(0.023)	(0.015)	(0.012)	(0.012)	(0.014)
Overeducated	-0.072***	-0.118***	-0.095***	-0.051***	-0.037***	-0.038***
	(0.011)	(0.016)	(0.013)	(0.010)	(0.008)	(0.012)

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.

Note: (i) *** signals significant at 1% level, ** signals significant at 5% level, and * signals significant at 10% level.
(ii) Robust standard errors are given in parenthesis.
(iii) Specf. 1 includes human capital and self-selection terms, specf. 2 includes human capital, job characteristics and self-selection terms, and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.

Table 7: Kuznets ratio of returns to education

Education (Base Cat.: No Formal Schooling)	Specification 1			Specification 2			Specification 3		
	θ_{90}/θ_{10}	θ_{90}/θ_{50}	θ_{50}/θ_{10}	θ_{90}/θ_{10}	θ_{90}/θ_{50}	θ_{50}/θ_{10}	θ_{90}/θ_{10}	θ_{90}/θ_{50}	θ_{50}/θ_{10}
Primary	4.342	1.585	2.739	1.314	1.041	1.262	1.737	1.492	1.164
Middle	4.276	1.778	2.404	1.849	1.202	1.539	1.706	1.411	1.210
Secondary	3.415	1.812	1.885	1.537	1.219	1.261	1.347	1.300	1.036
Higher Secondary	2.782	1.265	2.198	1.207	1.025	1.177	1.121	1.105	1.014
Graduate or Above	1.954	1.061	1.841	1.006	0.950	1.059	0.981	1.020	0.961

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.

Note: Specf. 1 includes human capital and self-selection terms, specf. 2 includes human capital, job characteristics and self-selection terms, and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.

Table 8: Range of returns to education

Education (Base Cat.: No Formal Schooling)	Specification 1			Specification 2			Specification 3		
	θ_{90}/θ_{10}	θ_{90}/θ_{50}	θ_{50}/θ_{10}	θ_{90}/θ_{10}	θ_{90}/θ_{50}	θ_{50}/θ_{10}	θ_{90}/θ_{10}	θ_{90}/θ_{50}	θ_{50}/θ_{10}
Primary	0.119	0.057	0.062	0.015	0.003	0.013	0.039	0.030	0.009
Middle	0.240	0.137	0.103	0.053	0.019	0.034	0.065	0.046	0.019
Secondary	0.442	0.280	0.162	0.072	0.037	0.035	0.067	0.060	0.007
Higher Secondary	0.629	0.206	0.423	0.056	0.008	0.048	0.042	0.037	0.005
Graduate or Above	0.686	0.081	0.605	0.003	-0.028	0.031	-0.012	0.012	-0.024

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.

Note: Specf. 1 includes human capital and self-selection terms, specf. 2 includes human capital, job characteristics and self-selection terms, and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.

Table 9: Gini coefficient – Returns to education

Education (Base Cat.: No Formal Schooling)	Specf. 1	Specf. 2	Specf. 3
Primary	0.23	0.07	0.14
Middle	0.24	0.12	0.11
Secondary	0.23	0.09	0.07
Higher Secondary	0.18	0.04	0.03
Graduate or Above	0.10	0.02	0.02

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.
Note: Specf. 1 includes human capital and self-selection terms, specf. 2 includes human capital, job characteristics and self-selection terms, and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.

Table 10: Coefficient of variation – Returns to education

Education (Base Cat.: No Formal Schooling)	Specf. 1	Specf. 2	Specf. 3
Primary	0.42	0.13	0.27
Middle	0.44	0.23	0.21
Secondary	0.43	0.17	0.13
Higher Secondary	0.35	0.08	0.05
Graduate or Above	0.20	0.03	0.03

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.
Note: Specf. 1 includes human capital and self-selection terms, specf. 2 includes human capital, job characteristics and self-selection terms, and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.

Figures

Figure 1: Lorenz Curves – Returns to education

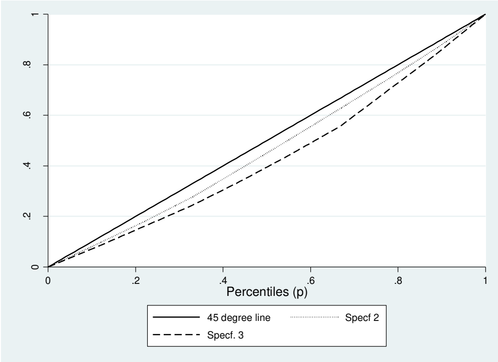

Figure 1a: Primary education

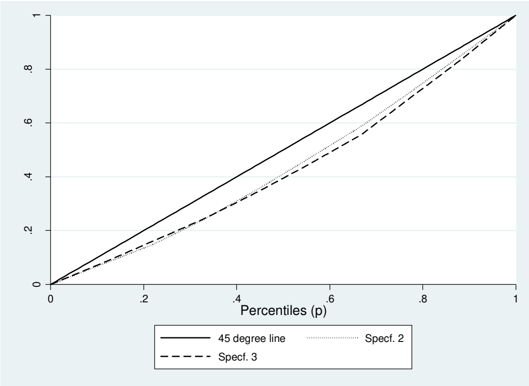

Figure 1b: Middle education

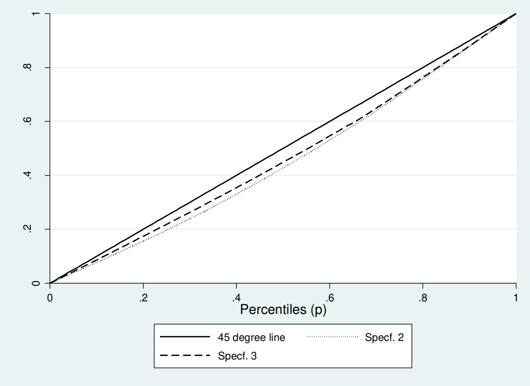

Figure 1c: Secondary education

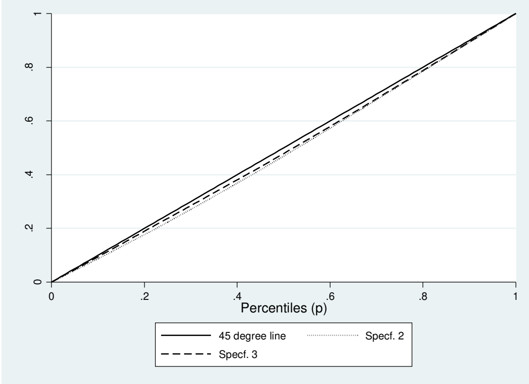

Figure 1d: Higher secondary education

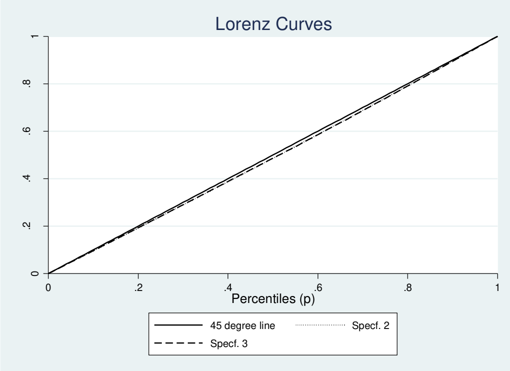

Figure 1e: Graduates or above education

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.
Note: (i) Reference category is 'no formal schooling'.
(ii) Specf. 2 includes human capital, job characteristics and self-selection terms and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.

Table A-1: Returns to wage-determining characteristics for wage/salaried employed workers in the working-age group (15-59 years)– Ordinary Least Square (OLS) estimates

Explanatory Variables	Outcome Variable: Logarithm of Daily Wages		
	(1)	(2)	(3)
	Specf. 1	Specf. 2	Specf. 3
Match Status (Base Cat.: Adequately Educated)			
Undereducated			0.004
			(0.014)
Overeducated			-0.072***
			(0.011)
Education (Base Cat.: No Formal Schooling)			
Primary	0.118***	0.065***	0.072***
	(0.010)	(0.010)	(0.015)
Middle	0.216***	0.101***	0.129***
	(0.012)	(0.012)	(0.018)
Secondary	0.412***	0.186***	0.230***
	(0.014)	(0.013)	(0.020)
Higher Secondary	0.768***	0.358***	0.410***
	(0.015)	(0.015)	(0.023)
Graduate or Above	1.227***	0.586***	0.654***
	(0.015)	(0.017)	(0.027)
Age	-0.003	-0.0002	-0.0002
	(0.003)	(0.003)	(0.003)
Occupation (Base Cat. Legislators, Senior Officials and Managers)			
Professionals		-0.147***	-0.147***
		(0.027)	(0.027)
Associate Professionals		-0.323***	-0.308***
		(0.027)	(0.027)
Clerks		-0.419***	-0.406***
		(0.026)	(0.026)
Service Workers and Shop & Market Sales Workers		-0.582***	-0.548***
		(0.027)	(0.028)
Skilled Agricultural and Fishery Workers		-0.570***	-0.520***
		(0.044)	(0.045)
Craft and Related Trades Workers		-0.531***	-0.489***
		(0.027)	(0.029)

Explanatory Variables	Outcome Variable: Logarithm of Daily Wages		
	(1)	(2)	(3)
	Specf. 1	Specf. 2	Specf. 3
Plant and Machine Operators and Assemblers		-0.436***	-0.399***
		(0.027)	(0.029)
Elementary Occupations		-0.637***	-0.581***
		(0.027)	(0.030)
Location of Workplace (Base Cat.: Rural)			
Urban		0.082***	0.081***
		(0.015)	(0.015)
No Fixed Location		0.039*	0.038
		(0.024)	(0.024)
Enterprise Type (Base Cat.: Proprietary)			
Partnership		-0.047**	-0.047**
		(0.019)	(0.019)
Government		0.340***	0.337***
		(0.013)	(0.013)
Public/Private		0.073***	0.072***
		(0.014)	(0.014)
Other		-0.138***	-0.135***
		(0.013)	(0.013)
Firm Size (Base Cat.: Less than ten)			
10-19		0.104***	0.102***
		(0.011)	(0.011)
20 or Above		0.210***	0.208***
		(0.010)	(0.010)
Unknown		0.140***	0.138***
		(0.016)	(0.016)
Contract (Base Cat.: Unwritten)			
Written		0.265***	0.264***
		(0.013)	(0.013)
Gender (Base Cat.: Male)			
Female	-0.361***	-0.361***	-0.362***
	(0.026)	(0.027)	(0.027)
Marital Status (Base Cat.: Unmarried)			
Married	0.144***	0.097***	0.097***
	(0.015)	(0.016)	(0.016)
Others	-0.015	-0.008	-0.008

Explanatory Variables	Outcome Variable: Logarithm of Daily Wages		
	(1)	(2)	(3)
	Specf. 1	Specf. 2	Specf. 3
	(0.032)	(0.028)	(0.028)
Social Group (Base Cat.: Scheduled Tribe)			
Scheduled Caste	-0.012	-0.013	-0.011
	(0.017)	(0.018)	(0.018)
OBC	-0.056***	-0.013	-0.012
	(0.016)	(0.016)	(0.016)
Others	0.020	0.049***	0.050***
	(0.018)	(0.018)	(0.018)
Religion (Base Cat.: Hindu)			
Muslim	-0.039***	-0.028**	-0.027**
	(0.012)	(0.012)	(0.012)
Christian	0.001	0.006	0.006
	(0.027)	(0.024)	(0.024)
Others	0.053**	0.048**	0.048**
	(0.025)	(0.022)	(0.022)
Sector (Base Cat.: Rural)			
Urban	0.160***	0.062***	0.063***
	(0.010)	(0.016)	(0.016)
Selection term			
Work	-0.058**	-0.075***	-0.075***
	(0.026)	(0.028)	(0.028)
Wage/salaried Employment	-0.138***	-0.061***	-0.061***
	(0.021)	(0.020)	(0.020)
Number of Observations	65,792	56,168	56,142
R-squared	0.390	0.475	0.476

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.

Note: (i) *** signals significant at 1% level, ** signals significant at 5% level,and * signals significant at 10% level.
(ii) Robust standard errors are in parenthesis.
(iii) Specf. 1 includes human capital and self-selection terms, specf. 2 includes human capital, job characteristics and self-selection terms, and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.
(iv) The analysis also controls for broad industry groups, age squared, interaction of gender and marital status, and 35 states and union territories.

Table A-2: Returns to wage-determining characteristics for wage/salaried employed workers in the working-age group (15-59 years)

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
Specification 1						
Education (Base Cat.: No Formal Schooling)						
Primary	0.118***	0.036***	0.054***	0.098***	0.125***	0.155***
	(0.010)	(0.012)	(0.008)	(0.007)	(0.008)	(0.010)
Middle	0.216***	0.073***	0.099***	0.176***	0.245***	0.313***
	(0.012)	(0.011)	(0.009)	(0.007)	(0.009)	(0.012)
Secondary	0.412***	0.183***	0.218***	0.345***	0.530***	0.625***
	(0.014)	(0.014)	(0.009)	(0.012)	(0.012)	(0.014)
Higher Secondary	0.768***	0.353***	0.454***	0.776***	0.980***	0.982***
	(0.015)	(0.013)	(0.013)	(0.012)	(0.011)	(0.012)
Graduate or Above	1.227***	0.719***	1.037***	1.324***	1.398***	1.405***
	(0.015)	(0.015)	(0.016)	(0.012)	(0.011)	(0.013)
Age	-0.003	0.0006	-0.007**	-0.014***	-0.002	0.018***
	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)
Gender (Base Cat.: Male)						
Female	-0.361***	-0.398***	-0.431***	-0.383***	-0.349***	-0.314***
	(0.026)	(0.036)	(0.023)	(0.022)	(0.024)	(0.029)
Marital Status (Base Cat.: Unmarried)						
Married	0.144***	0.140***	0.151***	0.132***	0.112***	0.111***
	(0.015)	(0.017)	(0.014)	(0.010)	(0.010)	(0.013)
Others	-0.015	-0.003	-0.015	-0.006	-0.003	0.038
	(0.032)	(0.053)	(0.031)	(0.025)	(0.025)	(0.050)
Social Group (Base Cat.: Scheduled Tribe)						

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
Scheduled Caste	-0.012	0.039**	0.039***	0.035***	0.017	-0.027*
	(0.017)	(0.016)	(0.013)	(0.010)	(0.011)	(0.016)
OBC	-0.056***	-0.009	-0.015	-0.019*	-0.037***	-0.073***
	(0.016)	(0.015)	(0.011)	(0.010)	(0.011)	(0.015)
Others	0.020	0.043***	0.037***	0.054***	0.041***	0.027**
	(0.018)	(0.015)	(0.014)	(0.010)	(0.010)	(0.013)
Religion (Base Cat.: Hindu)						
Muslim	-0.039***	-0.042***	-0.009	0.004	-0.009	-0.044***
	(0.012)	(0.014)	(0.009)	(0.007)	(0.010)	(0.013)
Christian	0.001	0.007	0.023	0.010	0.005	0.028
	(0.027)	(0.030)	(0.019)	(0.016)	(0.018)	(0.020)
Others	0.053**	0.001	0.028	0.069***	0.074***	0.062**
	(0.025)	(0.026)	(0.019)	(0.019)	(0.015)	(0.024)
Sector (Base Cat.: Rural)						
Urban	0.160***	0.085***	0.101***	0.130***	0.164***	0.195***
	(0.010)	(0.010)	(0.008)	(0.007)	(0.007)	(0.010)
Selection term						
Work	-0.058**	-0.498***	-0.335***	-0.266***	-0.019	0.227***
	(0.026)	(0.060)	(0.059)	(0.055)	(0.054)	(0.063)
Work2		0.154***	0.104***	0.116***	0.048	-0.046
		(0.042)	(0.037)	(0.038)	(0.032)	(0.036)
Wage/salaried Employment	-0.138***	0.094**	0.171***	0.156***	-0.064*	-0.263***
	(0.021)	(0.037)	(0.036)	(0.033)	(0.035)	(0.047)
Wage/salaried Employment2		-0.085***	-0.127***	-0.119***	-0.022	0.063***

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
		(0.014)	(0.014)	(0.013)	(0.014)	(0.019)
Specification 2						
Education (Base Cat.: No Formal Schooling)						
Primary	0.065***	0.048***	0.048***	0.061***	0.069***	0.064***
	(0.010)	(0.018)	(0.010)	(0.010)	(0.009)	(0.014)
Middle	0.101***	0.062***	0.063***	0.096***	0.116***	0.115***
	(0.012)	(0.016)	(0.010)	(0.009)	(0.010)	(0.014)
Secondary	0.186***	0.134***	0.133***	0.169***	0.202***	0.206***
	(0.013)	(0.017)	(0.013)	(0.011)	(0.014)	(0.014)
Higher Secondary	0.358***	0.271***	0.289***	0.319***	0.332***	0.327***
	(0.015)	(0.019)	(0.014)	(0.014)	(0.014)	(0.015)
Graduate or Above	0.586***	0.524***	0.556***	0.555***	0.536***	0.527***
	(0.017)	(0.025)	(0.018)	(0.014)	(0.013)	(0.016)
Age	-0.0002	0.008*	-0.003	-0.002	0.006***	0.009***
	(0.003)	(0.005)	(0.004)	(0.002)	(0.002)	(0.003)
Occupation (Base Cat. Legislators, Senior Officials and Managers)						
Professionals	-0.147***	-0.123**	-0.134***	-0.153***	-0.197***	-0.187***
	(0.027)	(0.049)	(0.032)	(0.024)	(0.020)	(0.028)
Associate Professionals	-0.323***	-0.262***	-0.323***	-0.345***	-0.399***	-0.423***
	(0.027)	(0.040)	(0.030)	(0.024)	(0.019)	(0.033)
Clerks	-0.419***	-0.255***	-0.358***	-0.444***	-0.552***	-0.594***
	(0.026)	(0.040)	(0.033)	(0.025)	(0.019)	(0.033)
Service Workers and Shop & Market Sales Workers	-0.582***	-0.433***	-0.550***	-0.630***	-0.707***	-0.730***

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
	(0.027)	(0.042)	(0.034)	(0.022)	(0.018)	(0.031)
Skilled Agricultural and Fishery Workers	-0.570***	-0.430***	-0.530***	-0.647***	-0.720***	-0.752***
	(0.044)	(0.071)	(0.053)	(0.037)	(0.032)	(0.050)
Craft and Related Trades Workers	-0.531***	-0.361***	-0.478***	-0.583***	-0.700***	-0.723***
	(0.027)	(0.039)	(0.033)	(0.024)	(0.019)	(0.030)
Plant and Machine Operators and Assemblers	-0.436***	-0.231***	-0.388***	-0.501***	-0.605***	-0.651***
	(0.027)	(0.037)	(0.031)	(0.022)	(0.018)	(0.031)
Elementary Occupations	-0.637***	-0.437***	-0.567***	-0.684***	-0.817***	-0.855***
	(0.027)	(0.039)	(0.032)	(0.023)	(0.019)	(0.030)
Location of Workplace (Base Cat.: Rural)						
Urban	0.082***	0.099***	0.096***	0.077***	0.048***	0.040***
	(0.015)	(0.019)	(0.011)	(0.009)	(0.010)	(0.013)
No Fixed Location	0.039*	0.020	0.065***	0.047***	0.037**	0.016
	(0.024)	(0.029)	(0.018)	(0.014)	(0.018)	(0.022)
Enterprise Type (Base Cat.: Proprietary)						
Partnership	-0.047**	-0.044	-0.058***	-0.055***	-0.046*	-0.011
	(0.019)	(0.031)	(0.021)	(0.016)	(0.025)	(0.030)
Government	0.340***	0.147***	0.296***	0.438***	0.455***	0.411***
	(0.013)	(0.014)	(0.014)	(0.013)	(0.011)	(0.014)
Public/Private	0.073***	-0.009	-0.001	0.040***	0.113***	0.210***
	(0.014)	(0.016)	(0.011)	(0.012)	(0.015)	(0.016)
Other	-0.138***	-0.152***	-0.121***	-0.108***	-0.104***	-0.092***
	(0.013)	(0.021)	(0.010)	(0.009)	(0.010)	(0.012)
Firm Size (Base Cat.: Less than ten)						

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
10-19	0.104***	0.118***	0.102***	0.087***	0.073***	0.058***
	(0.011)	(0.014)	(0.007)	(0.009)	(0.008)	(0.009)
20 or Above	0.210***	0.209***	0.199***	0.175***	0.170***	0.172***
	(0.010)	(0.014)	(0.009)	(0.007)	(0.007)	(0.009)
Unknown	0.140***	0.107***	0.110***	0.098***	0.113***	0.140***
	(0.016)	(0.016)	(0.011)	(0.010)	(0.010)	(0.012)
Contract (Base Cat.: Unwritten)						
Written	0.265***	0.265***	0.310***	0.283***	0.253***	0.232***
	(0.013)	(0.014)	(0.013)	(0.009)	(0.007)	(0.009)
Gender (Base Cat.: Male)						
Female	-0.361***	-0.458***	-0.461***	-0.383***	-0.307***	-0.248***
	(0.027)	(0.040)	(0.028)	(0.024)	(0.023)	(0.025)
Marital Status (Base Cat.: Unmarried)						
Married	0.097***	0.154***	0.130***	0.083***	0.063***	0.070***
	(0.016)	(0.020)	(0.012)	(0.012)	(0.011)	(0.013)
Others	-0.008	-0.031	-0.046	0.015	-0.004	-0.012
	(0.028)	(0.032)	(0.031)	(0.037)	(0.021)	(0.030)
Social Group (Base Cat.: Scheduled Tribe)						
Scheduled Caste	-0.013	-0.006	0.020*	0.039***	0.028**	-0.004
	(0.018)	(0.020)	(0.012)	(0.011)	(0.013)	(0.016)
OBC	-0.013	-0.005	0.0009	0.019**	0.004	-0.023*
	(0.016)	(0.020)	(0.011)	(0.009)	(0.011)	(0.013)
Others	0.049***	0.020	0.047***	0.076***	0.073***	0.068***
	(0.018)	(0.021)	(0.011)	(0.009)	(0.010)	(0.015)

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
Religion (Base Cat.: Hindu)						
Muslim	-0.028**	-0.034*	-0.014	-0.005	-0.013*	-0.038***
	(0.012)	(0.020)	(0.010)	(0.008)	(0.007)	(0.007)
Christian	0.006	0.0002	-0.011	0.011	0.008	0.016
	(0.024)	(0.028)	(0.018)	(0.015)	(0.016)	(0.018)
Others	0.048**	0.023	0.010	0.044***	0.039***	0.066***
	(0.022)	(0.027)	(0.018)	(0.013)	(0.013)	(0.016)
Sector (Base Cat.: Rural)						
Urban	0.062***	0.062***	0.043***	0.033***	0.050***	0.061***
	(0.016)	(0.019)	(0.011)	(0.008)	(0.010)	(0.014)
Selection term						
Work	-0.075***	-0.199**	-0.209***	-0.205***	-0.075*	-0.012
	(0.028)	(0.088)	(0.059)	(0.044)	(0.044)	(0.060)
Work2		0.027	0.062**	0.099***	0.056**	0.010
		(0.059)	(0.028)	(0.027)	(0.023)	(0.037)
Wage/salaried Employment	-0.061***	0.112**	0.147***	0.077***	-0.0009	-0.042
	(0.020)	(0.050)	(0.032)	(0.024)	(0.029)	(0.045)
Wage/salaried Employment2		-0.071***	-0.098***	-0.061***	-0.023**	-0.007
		(0.020)	(0.015)	(0.010)	(0.011)	(0.017)
Specification 3						
Match Status (Base Cat.: Adequately Educated)						
Undereducated	0.004	-0.007	-0.010	-0.002	0.026**	0.031**
	(0.014)	(0.023)	(0.015)	(0.012)	(0.012)	(0.014)
Overeducated	-0.072***	-0.118***	-0.095***	-0.051***	-0.037***	-0.038***

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
	(0.011)	(0.016)	(0.013)	(0.010)	(0.008)	(0.012)
Education (Base Cat.: No Formal Schooling)						
Primary	0.072***	0.053**	0.046***	0.062***	0.093***	0.092***
	(0.015)	(0.024)	(0.017)	(0.014)	(0.013)	(0.015)
Middle	0.129***	0.093***	0.091***	0.112***	0.154***	0.158***
	(0.018)	(0.027)	(0.020)	(0.015)	(0.016)	(0.018)
Secondary	0.230***	0.193***	0.182***	0.200***	0.245***	0.260***
	(0.020)	(0.031)	(0.020)	(0.017)	(0.017)	(0.023)
Higher Secondary	0.410***	0.347***	0.345***	0.352***	0.389***	0.389***
	(0.023)	(0.036)	(0.023)	(0.016)	(0.021)	(0.026)
Graduate or Above	0.654***	0.620***	0.633***	0.596***	0.605***	0.608***
	(0.027)	(0.042)	(0.028)	(0.020)	(0.024)	(0.031)
Age	-0.0002	0.008*	-0.003	-0.001	0.006***	0.008**
	(0.003)	(0.005)	(0.003)	(0.003)	(0.002)	(0.003)
Occupation (Base Cat. Legislators, Senior Officials and Managers)						
Professionals	-0.147***	-0.147***	-0.139***	-0.152***	-0.197***	-0.189***
	(0.027)	(0.041)	(0.031)	(0.021)	(0.021)	(0.021)
Associate Professionals	-0.308***	-0.270***	-0.314***	-0.334***	-0.389***	-0.407***
	(0.027)	(0.037)	(0.029)	(0.019)	(0.022)	(0.021)
Clerks	-0.406***	-0.260***	-0.354***	-0.433***	-0.545***	-0.580***
	(0.026)	(0.039)	(0.030)	(0.021)	(0.024)	(0.025)
Service Workers and Shop & Market Sales Workers	-0.548***	-0.405***	-0.516***	-0.608***	-0.685***	-0.704***
	(0.028)	(0.041)	(0.033)	(0.022)	(0.024)	(0.026)

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
Skilled Agricultural and Fishery Workers	-0.520***	-0.337***	-0.481***	-0.609***	-0.686***	-0.701***
	(0.045)	(0.088)	(0.050)	(0.037)	(0.043)	(0.054)
Craft and Related Trades Workers	-0.489***	-0.323***	-0.431***	-0.556***	-0.670***	-0.688***
	(0.029)	(0.044)	(0.030)	(0.019)	(0.024)	(0.026)
Plant and Machine Operators and Assemblers	-0.399***	-0.199***	-0.353***	-0.476***	-0.579***	-0.619***
	(0.029)	(0.048)	(0.032)	(0.022)	(0.023)	(0.027)
Elementary Occupations	-0.581***	-0.381***	-0.505***	-0.646***	-0.779***	-0.807***
	(0.030)	(0.046)	(0.031)	(0.020)	(0.025)	(0.029)
Location of Workplace (Base Cat.: Rural)						
Urban	0.081***	0.104***	0.010***	0.074***	0.046***	0.037***
	(0.015)	(0.016)	(0.009)	(0.009)	(0.007)	(0.011)
No Fixed Location	0.038	0.021	0.075***	0.042***	0.032***	0.0138
	(0.024)	(0.031)	(0.019)	(0.016)	(0.012)	(0.019)
Enterprise Type (Base Cat.: Proprietary)						
Partnership	-0.047**	-0.051	-0.059***	-0.059***	-0.046*	-0.004
	(0.019)	(0.035)	(0.021)	(0.018)	(0.024)	(0.031)
Government	0.337***	0.138***	0.295***	0.436***	0.452***	0.407***
	(0.013)	(0.016)	(0.011)	(0.014)	(0.013)	(0.015)
Public/Private	0.072***	-0.008	0.003	0.038***	0.116***	0.206***
	(0.014)	(0.014)	(0.012)	(0.012)	(0.016)	(0.018)
Other	-0.135***	-0.152***	-0.111***	-0.104***	-0.102***	-0.095***
	(0.013)	(0.019)	(0.010)	(0.007)	(0.008)	(0.013)
Firm Size (Base Cat.: Less than ten)						
10-19	0.102***	0.111***	0.104***	0.086***	0.072***	0.056**

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
	(0.011)	(0.014)	(0.009)	(0.008)	(0.007)	(0.010)
20 or Above	0.208***	0.203***	0.197***	0.173***	0.168***	0.171***
	(0.010)	(0.014)	(0.009)	(0.007)	(0.011)	(0.012)
Unknown	0.138***	0.101***	0.106***	0.096***	0.110***	0.140***
	(0.016)	(0.019)	(0.013)	(0.010)	(0.010)	(0.011)
Contract (Base Cat.: Unwritten)						
Written	0.264***	0.265***	0.304***	0.282***	0.254***	0.233***
	(0.013)	(0.014)	(0.010)	(0.011)	(0.008)	(0.010)
Gender (Base Cat.: Male)						
Female	-0.362***	-0.448***	-0.462***	-0.384***	-0.307***	-0.245***
	(0.027)	(0.034)	(0.024)	(0.025)	(0.025)	(0.029)
Marital Status (Base Cat.: Unmarried)						
Married	0.097***	0.150***	0.132***	0.082***	0.064***	0.071***
	(0.016)	(0.016)	(0.014)	(0.012)	(0.010)	(0.011)
Others	-0.008	-0.024	-0.045	0.005	-0.006	-0.009
	(0.028)	(0.036)	(0.030)	(0.034)	(0.026)	(0.030)
Social Group (Base Cat.: Scheduled Tribe)						
Scheduled Caste	-0.011	0.003	0.025*	0.036***	0.027***	0.005
	(0.018)	(0.020)	(0.014)	(0.010)	(0.010)	(0.014)
OBC	-0.012	0.0006	0.004	0.018*	0.003	-0.016
	(0.016)	(0.018)	(0.013)	(0.010)	(0.009)	(0.011)
Others	0.050***	0.031	0.051***	0.076***	0.071***	0.072***
	(0.018)	(0.019)	(0.013)	(0.010)	(0.009)	(0.015)
Religion (Base Cat.: Hindu)						

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	θ = 10	θ = 25	θ = 50	θ = 75	θ = 90
Muslim	-0.027**	-0.033**	-0.010	-0.005	-0.013*	-0.036***
	(0.012)	(0.013)	(0.010)	(0.008)	(0.008)	(0.011)
Christian	0.006	0.010	-0.003	0.009	0.009	0.013
	(0.024)	(0.032)	(0.020)	(0.016)	(0.013)	(0.016)
Others	0.048**	0.024	0.016	0.042***	0.044***	0.058***
	(0.022)	(0.026)	(0.021)	(0.014)	(0.013)	(0.015)
Sector (Base Cat.: Rural)						
Urban	0.063***	0.054***	0.040***	0.037***	0.053***	0.066***
	(0.016)	(0.015)	(0.010)	(0.008)	(0.008)	(0.013)
Selection term						
Work	-0.075***	-0.205**	-0.219***	-0.208***	-0.081	-0.016
	(0.028)	(0.089)	(0.058)	(0.065)	(0.053)	(0.067)
Work2		0.034	0.066**	0.100**	0.058**	0.012
		(0.056)	(0.032)	(0.041)	(0.029)	(0.036)
Wage/salaried Employment	-0.061***	0.106**	0.152***	0.082***	-0.003	-0.037
	(0.020)	(0.053)	(0.032)	(0.028)	(0.035)	(0.040)
Wage/salaried Employment2		-0.067***	-0.098***	-0.063***	-0.023*	-0.011
		(0.023)	(0.014)	(0.011)	(0.013)	(0.016)

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.

Note: (i) *** signals significant at 1% level, ** signals significant at 5% level,and * signals significant at 10% level.
(ii) Robust standard errors are in parenthesis.
(iii) Specf. 1 includes human capital and self-selection terms, specf. 2 includes human capital, job characteristics and self-selection terms, and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.
(iv) The analysis also controls for broad industry groups, age squared, interaction of gender and marital status, and 35 states and union territories.

Table A-3: Returns to wage determining characteristics for wage/salaried employed workers in the working-age group (15-59 years)–Job characteristics model

Explanatory Variables	Outcome Variable: Logarithm of Daily Wages
Education (Base Cat.: No Formal Schooling)	
Primary	0.021**
	(0.010)
Middle	0.033***
	(0.012)
Secondary	0.126***
	(0.013)
Higher Secondary	0.295***
	(0.015)
Graduate or Above	0.550***
	(0.017)
Occupation (Base Cat.: Legislators, Senior Officials and Managers)	
Professionals	-0.179***
	(0.027)
Associate Professionals	-0.361***
	(0.027)
Clerks	-0.450***
	(0.027)
Service Workers and Shop & Market Sales Workers	-0.635***
	(0.027)
Skilled Agricultural and Fishery Workers	-0.605***
	(0.045)
Craft and Related Trades Workers	-0.583***
	(0.027)
Plant and Machine Operators and Assemblers	-0.493***
	(0.027)
Elementary Occupations	-0.692***
	(0.028)
Location of Workplace (Base Cat.: Rural)	
Urban	0.085***
	(0.015)
No Fixed Location	0.038
	(0.024)
Enterprise Type (Base Cat.: Proprietary)	
Partnership	-0.047**
	(0.019)
Government	0.388***
	(0.014)
Public/Private	0.070***
	(0.014)
Other	-0.131***

Explanatory Variables	Outcome Variable: Logarithm of Daily Wages
	(0.013)
Contract (Base Cat.: Unwritten)	
Written	0.283***
	(0.013)
Gender (Base Cat.: Male)	
Female	-0.364***
	(0.025)
Marital Status (Base Cat.: Unmarried)	
Married	0.204***
	(0.015)
Others	0.172***
	(0.028)
Social Group (Base Cat.: Scheduled Tribe)	
Scheduled Caste	0.025
	(0.018)
OBC	-0.009
	(0.017)
Others	0.058***
	(0.018)
Religion (Base Cat.: Hindu)	
Muslim	-0.051***
	(0.012)
Christian	0.016
	(0.024)
Others	0.036*
	(0.022)
Sector (Base Cat.: Rural)	
Urban	0.092***
	(0.016)
Selection term	
Work	-0.099***
	(0.021)
Wage/salaried Employment	-0.180***
	(0.019)
Number of Observations	56,168
R-squared	0.465

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.

Note: (i) *** signals significant at 1% level, ** signals significant at 5% level,and * signals significant at 10% level.
(ii) Robust standard errors are in parenthesis.
(iii) The model includes education, job characteristics, and self-selection terms.
(iv) The analysis also controls for broad industry groups, age squared, interaction of gender and marital status, and 35 states and union territories.

Table A-4: Education and occupation (mis-)match for wage/salaried employed workers in the working-age group (15-59 years): by human capital variables

	Variables	Under	Adequate	Over
Education	Illiterate	39.65	60.35	-
	Primary or Below	7.69	92.31	-
	Middle	3.91	53.21	42.88
	Secondary	7.08	43.71	49.21
	Higher Secondary	5.29	56.46	38.25
	Graduate or Above	0.61	67.75	31.64
Age	15-24	10.87	62.54	26.59
	25-34	12.76	63.45	23.79
	35-44	15.76	66.8	17.44
	45-59	19.75	69.99	10.26

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.
Note: Sampling weights have been used.

Table A-5: Education and occupation (mis-)match for wage/salaried employed workers in the working-age group (15-59 years): by job characteristics

	Variables	Under	Adequate	Over
Job Contract	Unwritten	24.61	57.66	17.73
	Written	11.58	69.34	19.08
Occupation	Legislators, Senior Officials and Managers	12.76	84.76	2.48
	Professionals	15.16	78.27	6.57
	Associate Professionals	12.8	66.61	20.6
	Clerks	15.18	72.22	12.6
	Service Workers and Shop & Market Sales Workers	19.41	65.4	15.19
	Skilled Agricultural and Fishery Workers	7.83	78.06	14.11
	Craft and Related Trades Workers	24.08	58.31	17.61
	Plant and Machine Operators and Assemblers	19.06	69.32	11.62
	Elementary Occupations	11.16	64.79	24.05
Industry Group	Agriculture	1.31	76.45	22.24
	Manufacturing	19.83	62.02	18.15
	Construction	32.93	43.92	23.15
	Services	16.33	69.14	14.53
Firm Size	Less than 10	24.6	57.93	17.47
	10-20	20.51	61.79	17.7
	20 and above	16.46	64.77	18.77
	Unknown	23.17	57.48	19.36
Enterprise Type	Proprietary	25.72	56.9	17.38
	Partnership	19.31	64.4	16.29
	Government	15.61	67.22	17.17
	Public/Private	13.78	63.43	22.78
	Other	24.82	58.04	17.15
Location of Work	Rural	25.69	56.01	18.3
	Urban	18.23	64.1	17.67
	No Fixed Location	24.49	56.86	18.66
Job Contract	Unwritten	24.61	57.66	17.73
	Written	11.58	69.34	19.08

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.
Note: Sampling weights have been used.

Table A-6: Education and occupation (mis-)match for wage/salaried employed workers in the working-age group (15-59 years): by other characteristics

Variables		Under	Adequate	Over
Gender	Male	15.15	63.63	21.22
	Female	13.84	73.09	13.07
Marital Status	Unmarried	10.18	62.09	27.73
	Married	15.6	66.43	17.97
	Others	23.15	70.21	6.65
Sector	Rural	13.67	66.23	20.11
	Urban	17.19	64.75	18.06

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.
Note: Sampling weights have been used.

Table A-7: Returns to education and EOM for wage/salaried employed workers in the working-age group (15-59 years) – Ventiles (5^{th} – 50^{th})

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	θ = 5	θ = 10	θ = 15	θ = 20	θ = 25	θ = 30	θ = 35	θ = 40	θ = 45	θ = 50
Specification 1										
Education (Base Cat.: No Formal Schooling)										
Primary	0.019	0.036**	0.038***	0.047***	0.054***	0.073***	0.08***	0.082***	0.092***	0.098***
Middle	0.049***	0.073***	0.081***	0.094***	0.099***	0.128***	0.138***	0.152***	0.168***	0.176***
Secondary	0.146***	0.183***	0.183***	0.198***	0.218***	0.245***	0.275***	0.295***	0.324***	0.345***
Higher Secondary	0.310***	0.353***	0.373***	0.404***	0.454***	0.521***	0.582***	0.641***	0.711***	0.776***
Graduate or Above	0.575***	0.719***	0.828***	0.931***	1.037***	1.133***	1.208***	1.258***	1.293***	1.324***
Specification 2										
Education (Base Cat.: No Formal Schooling)										
Primary	0.037*	0.048***	0.056***	0.049***	0.048***	0.053***	0.053***	0.062***	0.059***	0.061***
Middle	0.047*	0.062***	0.071***	0.062***	0.063***	0.079***	0.081***	0.089***	0.095***	0.096***
Secondary	0.124***	0.134***	0.131***	0.130***	0.133***	0.143***	0.156***	0.160***	0.165***	0.169***
Higher Secondary	0.252***	0.271***	0.282***	0.280***	0.289***	0.297***	0.302***	0.307***	0.312***	0.319***
Graduate or Above	0.455***	0.524***	0.552***	0.563***	0.556***	0.573***	0.569***	0.566***	0.560***	0.555***
Specification 3										
Education (Base Cat.: No Formal Schooling)										
Primary	0.033	0.053**	0.058***	0.051***	0.046***	0.054***	0.054***	0.066***	0.063***	0.062***
Middle	0.076**	0.093***	0.104***	0.097***	0.091***	0.103***	0.109***	0.120***	0.118***	0.112***
Secondary	0.178***	0.193***	0.192***	0.187***	0.182***	0.191***	0.203***	0.208***	0.205***	0.200***
Higher Secondary	0.326***	0.347***	0.363***	0.352***	0.345***	0.355***	0.360***	0.368***	0.363***	0.352***
Graduate or Above	0.543***	0.620***	0.645***	0.649***	0.633***	0.644***	0.645***	0.640***	0.623***	0.596***
Match Status (Base Cat.: Adequately Educated)										
Undereducated	-0.010	-0.007	0.001	-0.003	-0.010	-0.006	-0.003	0.001	-0.001	-0.002
Overeducated	-0.116***	-0.118***	-0.113***	-0.104***	-0.095***	-0.095***	-0.089***	-0.082***	-0.070***	-0.051***

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.

Note: (i) *** signals significant at 1% level, ** signals significant at 5% level, and * signals significant at 10% level.

(ii) Robust standard errors are given in parenthesis.

(iii) Specf. 1 includes human capital and self-selection terms, specf. 2 includes human capital, job characteristics and self-selection terms, and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.

Table A-8: Returns to education and EOM for wage/salaried employed workers in the working-age group (15-59 years) – Ventiles (55^{th} – 90^{th})

Explanatory Variable	Outcome Variable: Logarithm of Daily Wages									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	θ = 55	θ = 60	θ = 65	θ = 70	θ = 75	θ = 80	θ = 85	θ = 90	θ = 95	θ = 55
Specification 1										
Education (Base Cat.: No Formal Schooling)										
Primary	0.102***	0.111***	0.119***	0.121***	0.125***	0.138***	0.143***	0.155***	0.150***	0.102***
Middle	0.187***	0.206***	0.224***	0.234***	0.245***	0.264***	0.287***	0.313***	0.333***	0.187***
Secondary	0.374***	0.413***	0.450***	0.484***	0.530***	0.573***	0.606***	0.625***	0.623***	0.374***
Higher Secondary	0.827***	0.882***	0.931***	0.966***	0.980***	0.985***	0.991***	0.982***	0.950***	0.827***
Graduate or Above	1.345***	1.364***	1.383***	1.394***	1.398***	1.397***	1.401***	1.405***	1.390***	1.345***
Specification 2										
Education (Base Cat.: No Formal Schooling)										
Primary	0.063***	0.063***	0.069***	0.072***	0.069***	0.066***	0.070***	0.064***	0.062***	0.063***
Middle	0.098***	0.097***	0.105***	0.116***	0.116***	0.111***	0.122***	0.115***	0.115***	0.098***
Secondary	0.175***	0.180***	0.185***	0.196***	0.202***	0.204***	0.216***	0.206***	0.200***	0.175***
Higher Secondary	0.322***	0.335***	0.335***	0.333***	0.332***	0.332***	0.350***	0.327***	0.303***	0.322***
Graduate or Above	0.548***	0.541***	0.535***	0.532***	0.536***	0.537***	0.551***	0.527***	0.502***	0.548***
Specification 3										
Education (Base Cat.: No Formal Schooling)										
Primary	0.071***	0.069***	0.088***	0.099***	0.093***	0.092***	0.081***	0.092***	0.091***	0.071***
Middle	0.124***	0.124***	0.143***	0.158***	0.154***	0.153***	0.142***	0.158***	0.156***	0.124***
Secondary	0.216***	0.214***	0.236***	0.245***	0.245***	0.252***	0.246***	0.260***	0.247***	0.216***
Higher Secondary	0.371***	0.377***	0.392***	0.393***	0.389***	0.388***	0.385***	0.389***	0.370***	0.371***
Graduate or Above	0.605***	0.596***	0.608***	0.605***	0.605***	0.604***	0.593***	0.608***	0.582***	0.605***
Match Status (Base Cat.: Adequately Educated)										
Undereducated	0.007	0.005	0.022**	0.031***	0.026**	0.030**	0.016	0.031**	0.039**	0.007
Overeducated	-0.053***	-0.051***	-0.047***	-0.041***	-0.037***	-0.035***	-0.027***	-0.038***	-0.034***	-0.053***

Source: Authors' calculation based on NSSO employment and unemployment survey, 2011-12.

Note: (i) *** signals significant at 1% level, ** signals significant at 5% level, and * signals significant at 10% level.

(ii) Robust standard errors are given in parenthesis.

(iii) Specf. 1 includes human capital and self-selection terms, specf. 2 includes human capital, job characteristics and self-selection terms, and specf. 3 includes human capital, job characteristics, self-selection terms and EOM.