Asian Journal of Applied Economics, 2026, Vol. 33, No. 1, Article No. 330105
ISSN: 2985-1610 (Online)

Education-Occupation Mismatch and the Role of Internal Migration Streams in Shaping Returns to Education

Farrukh Mahmood*, Abdul Wahab Rana
Faculty of Humanities and Social Sciences, Information Technology University, Lahore 54600, Pakistan.

*Corresponding author E-mail: farrukh.mahmood@itu.edu.pk
*ORCID iD: 0000-0002-9100-2786

Received: 25 December 2025; Received in revised form: 27 January 2026; Accepted: 2 March 2026

Abstract

Background and Objectives: Education is a primary pillar of human capital development, directly impacting labor market performance and economic growth. While global returns to education are increasing, the Pakistani labor market continues to face significant inefficiencies in the form of education-occupation mismatch (EOM), where individual skills fail to align with job demands. Literature suggests that nearly 45% of the Pakistani workforce experiences such mismatches, resulting in wage penalties, skill underutilization, and reduced productivity. This study examines the returns to education and EOM through the lens of internal migration. By integrating job-matching theory with migration economics, this paper hypothesizes that migration to diverse labor markets reduces the negative outcomes of employment mismatch and improves the distribution of match quality.

Methodology: The study utilizes the Pakistan Social and Living Standards Measurement (PSLM) 2019-20 survey, focusing on a refined sample of 114,586 household heads aged 15 to 65. The education-occupation fit is determined via the realized-matches approach, calculating the mean and standard deviation of education levels within four-digit occupational groups to identify overeducated, undereducated, and matched individuals. The empirical analysis employs an augmented Mincer equation to estimate returns to education. To address inflated variance and measurement errors in logarithmic monthly wages, the study utilizes advanced statistical learning models, including Ridge, Lasso, and Elastic Net regressions, with 10-fold cross-validation and 100 bootstrap replications. To mitigate endogeneity and omitted-variable bias, an instrumental-variables (IV) framework with a two-step least-squares estimator is applied, using family educational background as a valid instrument.

Key Findings: The empirical results yield three primary findings regarding the Pakistani labor market. First, educational returns are significant, with earnings increasing at a rate of 14.9% for each additional year of schooling after controlling for endogeneity. Second, the analysis reveals a concave relationship between education-occupation mismatch and earnings, wherein under-educated individuals face a substantial wage penalty of 18.2% compared to their adequately educated counterparts, while over-educated individuals experience diminishing marginal returns with income increases of only 7.68%. Third, internal migration exerts a transformational effect, as migrants on average earn more than non-migrants. Specifically, migration streams—particularly rural-to-urban and urban-to-urban flows—redefine educational returns and flatten the mismatch returns curve, thereby reducing wage inefficiencies caused by skill imbalances.

Policy Implications: Maximizing economic returns on human investment in Pakistan requires a multi-layered intervention. It is necessary to enhance vocational guidance and align educational curricula with market demands to reduce early-career mismatch penalties. Furthermore, policies should facilitate internal mobility to various urban labor markets to mitigate the wage inefficiencies identified in the study. Targeted measures are also required to address the severe wage penalties faced by married women, including promoting female entrepreneurship within independent contracting and agriculture, ensuring pay equity, and enhancing workplace inclusivity. Ultimately, achieving equitable labor market outcomes requires dismantling the structural and normative barriers that currently limit the career progression of the female workforce.

Keywords: Education-Occupation Mismatch; Migration; Returns to Education; Nonlinear Returns; Pakistan

JEL Classification Codes: I21; I26; J24; J61; R23

1. Introduction

Education serves as a fundamental pillar of human capital development, directly influencing labor market outcomes and, consequently, economic growth (Becker, 1975; Becker, 1993). Globally, returns to education have been rising, emphasizing the critical role of schooling in driving economic development (Colclough et al., 2009). However, despite notable investments in education, labor markets continue to experience substantial inefficiencies due to education-occupation mismatch (EOM)—a condition where an individual's educational attainment does not align with the specific requirements of their job. While wages ideally reflect productivity based on schooling, job-matching and assignment theories suggest this link is often weakened by search frictions, information asymmetries, and institutional rigidities (Leuven & Oosterbeek, 2011; Sattinger, 1993). In this context, productivity shortfalls in undereducated workers generate significant wage penalties, whereas overeducated workers may receive only limited premiums because surplus skills are not fully rewarded in their positions (Verdugo & Verdugo, 1989). Such asymmetric behavior implies a non-linear, often concave, relationship between EOM and earnings. This mismatch ultimately dampens both productivity and broader economic growth (Akhtar et al., 2018; Tilak, 2018). Literature regarding Pakistan suggests that nearly 45% of the labor force faces such mismatches; while undereducation is more prevalent among men, women disproportionately bear the adverse effects of these misalignments (Ahsan, 2024).

The presence of EOM weakens the incentive for workers to seek employment that fully utilizes their education, as overeducated workers may accept lower-paying positions due to limited opportunities or suppressed wages (Feliciano, 2021; Pakistan Institute of Development Economics, 2024; Sitorus & Wicaksono, 2022;). These wage penalties often motivate workers to migrate in search of better employment opportunities or improved living standards (Bahl & Sharma, 2023; Budría & Moro-Egido, 2014; Quinn & Rubb, 2005). While migration can increase the labor supply in destination areas and potentially distort wage structures, it also grants workers access to more diverse labor markets. In these diversified environments, superior education-occupation matches, and wage premiums are more likely to exist, thereby partially alleviating the adverse effects of the initial mismatch (Greenwood, 1997; Khan & Said, 2022). Consequently, migration can mitigate the wage penalties associated with EOM not only by raising overall wage levels but by fundamentally altering the distribution of match quality (Dustmann, 2023).

Combining job-matching theory and migration economics provides clear empirical prospects: when migration allows workers to move to more diversified labor markets with higher demand for skilled jobs, job matching improves. In such cases, returns to education are expected to be less concave or closer to linear. Furthermore, when migration mainly increases the labor supply in low-skill or segmented labor markets, EOM is likely to persist. As a result, marginal returns to additional education remain concave. These effects are expected to vary by gender and marital status, since mobility constraints limit women's ability to access better job matches. Our empirical analysis tests these expectations by comparing the structure of returns to education across migrants and non-migrants and across different migration motives.

Conversely, the potential of migration to improve labor-market outcomes is often offset by persistent barriers that limit the ability of married women to benefit fully. Household and caregiving responsibilities constrain their geographic and occupational mobility; these constraints function as search frictions that extend job searches, reduce match quality, and restrict access to positions that fully utilize their education (Awan, 2020). Additional obstacles, such as occupational segregation, discriminatory hiring practices, and insufficient workplace support, including limited maternity leave and inflexible arrangements, may further compress their returns to education (World Bank, 2021). As a result, even when they migrate, married women often experience a more concave education–earnings profile because additional years of education do not translate proportionally into better job matches. Therefore, migration alone may not fully eliminate the wage disadvantages arising from EOM for this group. The empirical analysis explicitly tests whether migration affects the curvature of returns to education differently for men and women, with a particular focus on married women.

Although the contribution of this study is focused, it adds to a growing body of literature that integrates the role of internal migration in shaping returns to education with EOM (Akram et al., 2024; Ashraf, 2011; Nasir & Nazli, 2000;). By incorporating migration streams and the reasons behind migration decisions, this research provides a nuanced understanding of labor-market dynamics that are often overlooked in conventional analyses. Using data from the Pakistan Social and Living Standards Measurement (PSLM) 2019-20 survey, this research estimates returns to education both with and without the migration stream. Furthermore, we employed several estimation methods to assess the consistency of estimates, offering policymakers novel insights to optimize the utilization of human capital

The rest of the study is structured as follows: Section 2 discusses the data and methodology. Section 3 presents the empirical results. Section 4 discusses the results and their economic significance. Section 5 concludes, and Section 6 provides policy implications.

2. Data and Methodology

2.1. The Data

This research utilizes the latest Pakistan Social and Living Standards Measurement (PSLM) survey (2019-20), published by the Pakistan Bureau of Statistics (PBS). The analysis focuses on the economically active population, specifically working-age individuals aged 15 to 65. The sample includes four employment categories: employers, paid employees, self-employed individuals, and agricultural workers. For self-employed and agricultural workers whose monthly income was not directly reported, monthly figures were derived by dividing their annual income by 12.

The unit of monthly wages is Pakistani rupees, and a logarithmic transformation was applied to this variable to mitigate the impact of extreme values and stabilize the data. This transformation also helps address potential heteroskedasticity, ensuring the reliability of the empirical results. After removing observations with missing wage data and performing data cleaning, the final sample size consists of 114,586 household heads, which includes a subset of 8,185 migrant heads.

Education is measured based on the highest degree attained by the individual. The PSLM also records reasons for migration. To focus on shifts in returns driven by job preferences, this study isolates "better economic opportunities" and "confirmed jobs" as specific categories, while using all other reasons (e.g., marriage, law and order, etc.) as the reference group. The reason for migration is treated as a categorical variable, coded as 1 for other reasons, 2 for better economic opportunities, and 3 for confirmed jobs.

Other control variables included in the model are the gender of the household head, marital status, and an interaction term between gender and marital status. Age is measured in years, and its quadratic form (age squared) is included to capture the potential nonlinear relationship between experience and earnings.

Consistent with the study's objectives, the unit of analysis is the head of the household. The PSLM identifies migration by checking if the head was born in the current district; if not, the original district is noted. For this study, an individual is defined as a migrant if they were born in a different district and have resided in their current district for more than six months. Furthermore, a "stream" variable was constructed to categorize migration flows based on the rural or urban status of the birth region and the current residence. This variable is coded as: (1) rural-to-rural, (2) rural-to-urban, (3) urban-to-rural, and (4) urban-to-urban.

Table 1 presents the frequency distribution for migrants versus non-migrants. Based on PSLM 2019–2020 data, approximately 7.14% of the individuals in the sample were identified as internal migrants. Among the migrant population, 74.64% are male and 25.36% are female. Regarding regional distribution, 67.76% of migrants reside in urban areas, while 32.24% live in rural areas.

An analysis of migration streams reveals that 25.30% of migrants moved between rural areas, whereas rural-to-urban migration accounted for the largest share at 41.85%. Urban-to-urban and urban-to-rural flows represented 6.94% and 25.91% of the migrant population, respectively. At the provincial level, Punjab holds the highest share of migrants at 62.33%, followed by Sindh (21.72%), Khyber Pakhtunkhwa (13.67%), and Balochistan (3.27%).

The motivations for migration are diverse: 44.36% of individuals moved in search of better economic opportunities, and 13.81% migrated for a confirmed job offer. The remaining 41.83% migrated for other reasons, including marriage, family considerations, law enforcement issues, or environmental factors. These figures demonstrate that internal migration in Pakistan is a widespread and multifaceted phenomenon, driven by varied objectives and directed toward diverse destinations.

Table 1. Descriptive Statistics by Migration Status

Variables	Migrants	Non-Migrants
Total Household Heads	8,185 (7.14%)	106,401 (92.00%)
Gender		
Male	6,109 (5.74%)	100,292 (94.26%)
Female	907 (12.93%)	6,109 (87.07%)
Region		
Urban	5,546 (12.31%)	39,519 (87.69%)
Rural	2,639 (3.80%)	66,882 (96.20%)
Province		
KPK	1,037 (5.17%)	19,019 (94.83%)
Punjab	5,102 (8.24%)	56,785 (91.76%)
Sindh	1,778 (7.46%)	22,064 (92.54%)
Balochistan	268 (3.06%)	8,533 (96.96%)
Migration Stream		
Rural-Rural	2,071	
Rural-Urban	3,425	
Urban-Rural	568	
Urban-Urban	2,121	
Reason for Migration		
Better Economic Opportunity	3,631	
Job Offered	1,130	
Other reasons	3,424	

Source: Author's calculations based on PSLM 2019-20.

Figure 1 illustrates the distribution of wages for migrants (blue) and non-migrants (black). The average wage for migrant household heads is 10.12 (blue solid line), which is higher than the average of 9.91 for non-migrants (black solid line). Additionally, the wage distribution for migrants exhibits a relatively lower variance. A two-sample t-test was conducted to evaluate the significance of these mean differences; with a t-statistic of 23.65, the null hypothesis of equal means is rejected at the 1% significance level. This confirms that migrant heads earn significantly more on average than their non-migrant counterparts.

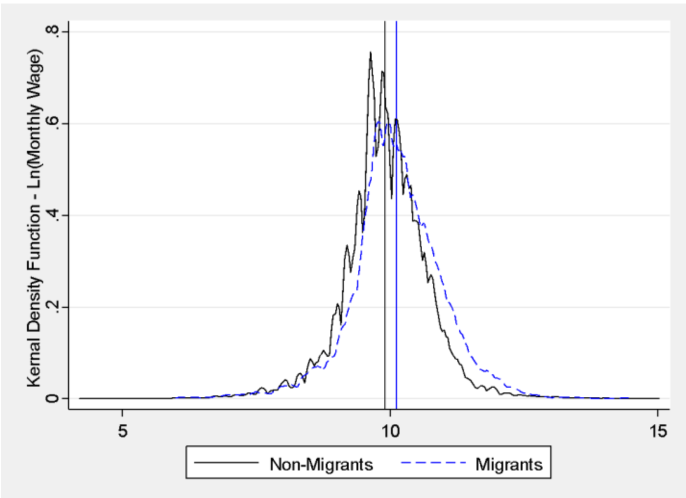
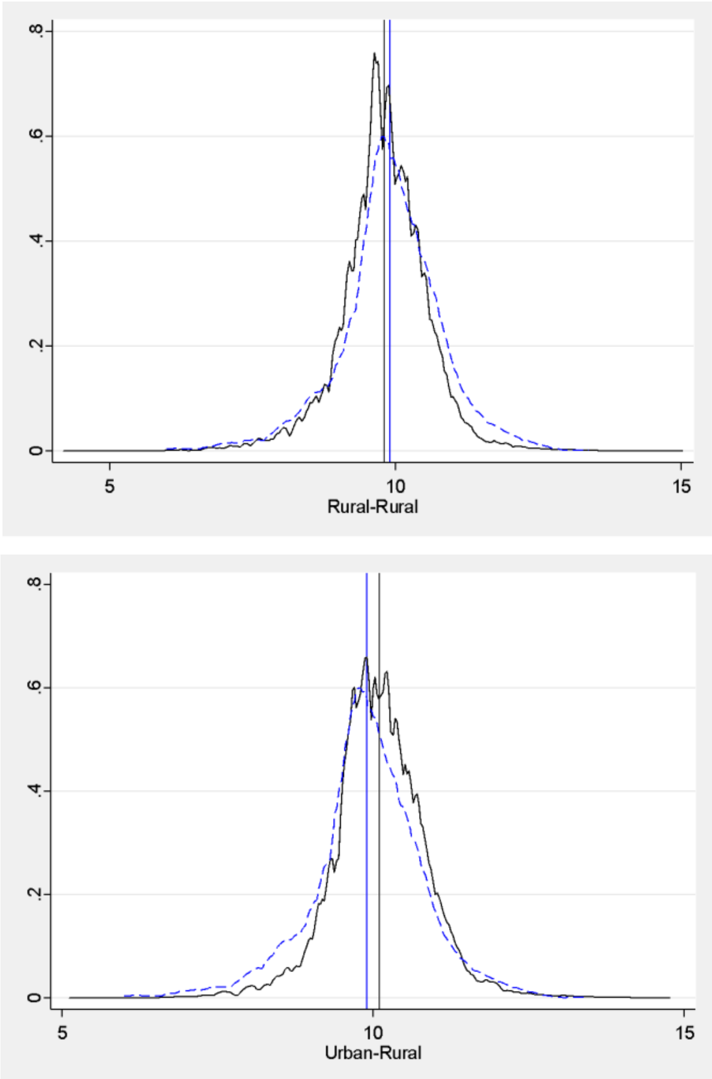


Figure 1. Kernel Density Function of Wage Across Migrants and Non-Migrants.
Source: Author's calculations based on PSLM 2019-20.
Note. The black line represents non-migrants, while the blue line represents migrants.

Figure 2 presents the wage distribution across migrants and non-migrants disaggregated by migration stream. Across most streams, migrants consistently earn more on average than non-migrants. The exception is the urban-to-rural stream, where the lower average wage reflects the reduced economic activity typical of rural areas, even when controlling for other socioeconomic factors.



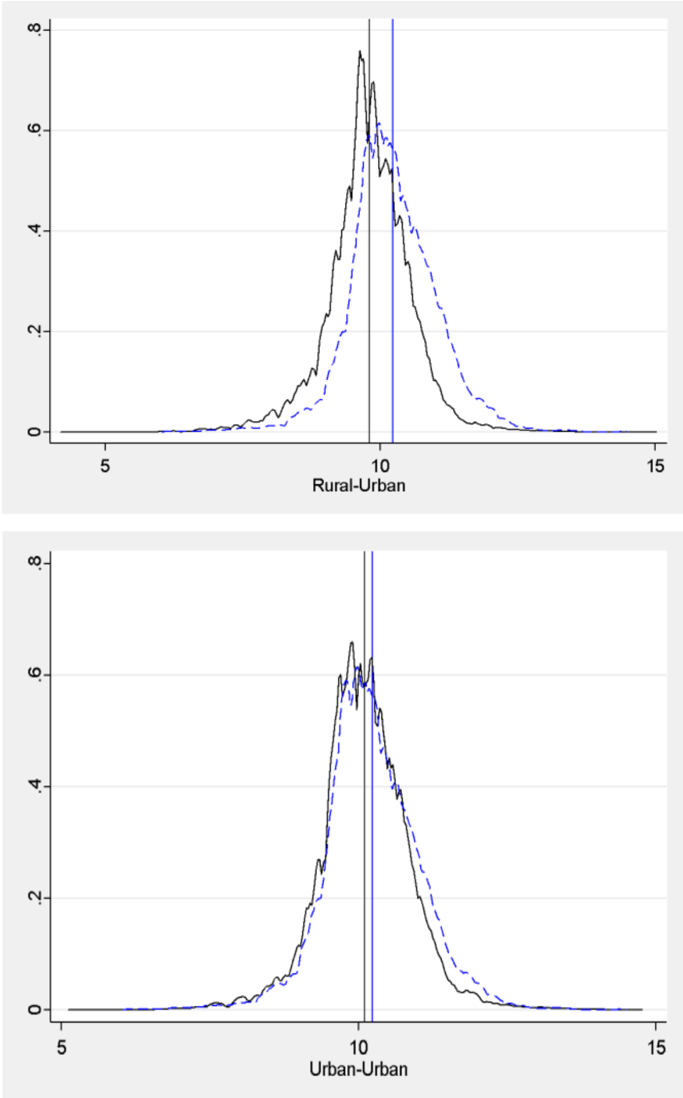


Figure 2. Kernel Density Function of Wage across Migrants and Non-Migrants at the Rural and Urban Levels.
Source: Author's calculations based on PSLM 2019-20.
Note. The black line represents non-migrants, while the blue line represents migrants.

2.2. Education-Occupation Mismatch (EOM) Definition and Measurement

Education-occupation mismatch is defined as the discrepancy between an individual's attained education and the level required by their job. Individuals with more education than required are classified as overeducated, while those with less are considered undereducated. When an individual's educational attainment aligns with job requirements, they are categorized as adequately educated

To measure required education, the literature identifies three primary methods: (1) worker self-assessment (subjective), (2) job analysis by experts, and (3) realized matches (objective) (Leuven & Oosterbeek, 2011). This study adopts the realized matches approach using occupation-specific mean education ±1 standard deviation (SD) (Duncan & Hoffman, 1981). Under this convention:

- **Overeducation** occurs when an individual's education exceeds the mean +1SD.
- **Undereducation** occurs when it falls below the mean −1SD.

- **Matching** occurs when the education level falls within the ±1SD range.

The realized matches approach is the most suitable for this research for two primary reasons. First, our objective is to analyze migrant outcomes across the entire labor market, which necessitates a method that reflects national-level demand and supply patterns. Second, in developing countries, the detailed survey data required for self-assessment or expert job analysis are often unavailable, making realized matches the only viable option. This method is widely utilized in research concerning spatial mobility and education mismatch (Poot & Stillman, 2016) and remains consistent with the empirical economic literature on survey data (Farooq, 2011). Consequently, this choice facilitates a more effective comparison of our findings with established research

To identify Education-Occupation Mismatch (EOM), this study utilizes two key data sources: the highest level of degree attained, and the four-digit occupational categories listed in the PSLM 2019–20. Following the methodology of Verdugo and Verdugo (1989), we calculate the mean degree level attained within each four-digit occupational category.

Let Edu_{ho} represent the highest level of degree attained (h) within a specific occupational category (o). Similarly, the standard deviation of the highest degree attained for each four-digit category is represented as sd_{ho}. The range for the required level of education is then calculated using the formula $Edu_{ho} \pm sd_{ho}$.

Under this framework:

- An individual is considered **overeducated** if their level of education is higher than $Edu_{ho} + sd_{ho}$.
- An individual is considered **undereducated** if their level of education is lower than $Edu_{ho} - sd_{ho}$.

While this ±SD convention is standard in the literature (Hartog, 2006), it exhibits classification sensitivity, meaning results can vary based on distributional assumptions and the specific threshold chosen. To test the robustness of our findings, we examined alternative thresholds within the PSLM data: Mean ± 0.75SD, Mean ± 1SD, and Mean ± 1.25SD. These alternatives yielded similar mismatch patterns or symmetric distributions, as summarized in Table 2.

Table 2. Mismatch Incidence and Returns by Threshold

	Mean $\pm 0.75SD$	**Mean $\pm 1SD$**	**Mean $\pm 1.25SD$**
Undereducation	27.16%	13.76%	9.99%
Overeducation	49.63%	72.31%	81.02%
Required education	23.21%	13.93%	8.99%

Source: Author's calculations based on PSLM, 2019-20.

Furthermore, in the context of the Pakistani labor market, informal and self-employment sectors often lack structured skill ladders, which may potentially inflate the prevalence of undereducation and obscure genuine skill-education mismatches. While the realized matches method remains empirically grounded for survey-based analysis, job analysis methods could provide a complementary approach for future research focused on formal sectors.

For the empirical analysis, the EOM variable was converted into a categorical variable following the methodology of Duncan and Hoffman (1981). Required education serves as the base category (1), with undereducation (2) and overeducation (3) acting as the comparison groups. The frequency distribution based on this EOM classification is detailed in Table 3.

Table 3. Frequency Distribution of Education-Occupation Mismatch.

	Undereducation	**Required Education**	**Overeducation**
Province			
KPK	3,159 (15.75%)	13,433 (66.98%)	3,464 (17.27%)
Punjab	11,256 (18.19%)	43,572 (70.41%)	7,059 (11.40%)
Sindh	3,463 (14.00%)	15,690 (63.41%)	4,689 (22.59%)
Balochistan	1,641 (18.65)	5,857 (66.55%)	1,303 (14.80%)
Region			
Urban	6,227 (13.82%)	31,572 (70.06%)	7,266 (16.12%)
Rural	13,292 (19.12%)	46,980 (67.58%)	9,249 (13.3%)
Gender			
Male	18,596 (17.29%)	73,118 (67.97%)	15,856 (14.74%)
Female	923 (13.16%)	5,434 (77.45%)	659 (9.39%)
Employment Status			
Employer	454 (14.51%)	2,048 (65.49%)	625 (20.00%)
Self-Employed	5,391 (17.86%)	21,153 (70.06%)	3,647 (12.08%)
Paid-Employed	10,200 (15.18%)	46,755 (65.59%)	10,232 (19.23%)
Agriculture-Employed	3,474 (24.67%)	8,596 (61.05%)	2,011 (14.28%)

Source: Author's calculations based on PSLM 2019–20.

2.3. The Econometric Model

This research utilizes an augmented Mincer equation to analyze the returns to education as influenced by migration and education-occupation mismatch (EOM) (Mincer, 1974). The baseline specification is defined as follows:

$$w_i = \alpha_0 + \alpha_1 edu_i + \beta_1 M_i + \beta_2 R_i + \beta_3 I_i + \varepsilon_i \tag{1}$$

In this model, the dependent variable w_i represents the logarithm of individual i's monthly wages. The primary variable of interest, edu_i, denotes the years of formal education attained by the individual.

The model incorporates three sets of control variables. The first set of controls is I_i, which represents the individual-level variables: year of formal education, age and its square, gender, marital status, and the interaction of gender and marital status. The second set of controls is R_i, which represents the regional level and focuses on local labor-market characteristics. The third set of controls is M_i, implemented in the empirical model to control for migrant heterogeneity by reason of migration, and migration stream.

To analyze the differential returns associated with EOM, Equation 1 is modified to include a categorical variable:

$$w_i = \alpha_0 ta + \sum_{j=1}^{3} \alpha_j * edu_i^j + \beta_1 M_i + \beta_2 R_i + \beta_3 I_i + \varepsilon_i \tag{2}$$

Where edu_i^j is a categorical variable for the EOM for an individual i. The education is at the required level for an occupation if $j = 1$ and serves as a base category for the comparison with undereducated and overeducated individuals. An individual will be considered as undereducated if their education level for an occupation is $j = 2$. The empirical value for an undereducated individual is captured by α_2, and the expected sign for $\alpha_2 < 0$, i.e., compared to the required level of education, undereducated individuals earn less. On the contrary, the empirical value of the overeducated individuals is captured by the α_3, and the expected sign for the $\alpha_3 > 0$ implies that the overeducated earn relatively more than those individuals who have the required level of education in the same occupation.

Furthermore, comparing $|\alpha_2|$ and $|\alpha_3|$ gives an important clue about the curvature of the relationship between returns and education-occupational mismatch. If

$|\alpha_2| = |\alpha_3|$, then the linear relationship between returns to education and EOM holds. If $|\alpha_2| > |\alpha_3|$, then it implies that returns are increasing at a decreasing rate across EOM – convex; conversely, $|\alpha_2| < |\alpha_3|$ implies that returns increase at an increasing rate across EOM – concave.

In addition, our second important variable is migration, and its impact is captured by β_1. The expected sign for $\beta_1 > 0$, which implies migration supports an increase in returns to education. However, it can be misleading, and the explanation can be inferred from the second panel of Figure 2, i.e., comparisons of average wages across migrants and non-migrants in the Urban-Rural case. It is more likely that, if an individual migrates from an urban to a rural area, his average wage might decrease, and hence the estimated parameter β_1 could be negative – $\beta_1 < 0$. It is therefore important to add a stream variable to overcome such issues. This stream variable will give us a clear picture of the migration direction.

2.4. Estimation Method

Empirically, the study begins with Ordinary Least Squares (OLS) estimation to establish baseline results. While OLS estimates are fully efficient under random sampling conditions, they often become biased when the Mincer equation suffers from sample selection bias, omitted-variable bias, or measurement error (Aslam, 2009). Although researchers typically mitigate these issues by focusing only on waged workers in the paid labor force, males in developing countries are frequently self-employed rather than wage workers.

For self-employed and agricultural workers, monthly income is calculated by dividing annual revenue by 12. This introduces a risk of measurement error in the dependent variable, as these groups may report income less precisely than paid employees due to irregular earnings and the difficulty of separating labor from capital income. To address this and improve prediction precision, the research employs 10-fold cross-validation—dividing the data into multiple sub-samples to compare inferential statistics—and the bootstrap method with 100 repetitions to stabilize standard error estimates (James et al., 2013; Kohavi, 1995).

Furthermore, advanced statistical shrinkage methods (Ridge, Lasso, and Elastic Net regression) are utilized with cross-validation (Hastie et al., 2009; Hilbert et al., 2021; Tibshirani, 1996; Zou & Hastie, 2005;). These methods offer substantial advantages over conventional OLS, particularly in their ability to handle multicollinearity among explanatory variables. This is especially relevant in educational datasets where factors such as parental education and income are often highly correlated. By introducing a penalty term into the loss function, ridge regression stabilizes coefficient estimates and mitigates the risk of overfitting. Similarly, lasso regression performs variable selection by shrinking non-essential coefficients to zero, thereby enhancing model interpretability and allowing for the identification of key factors influencing educational returns. Elastic net regression subsequently combines the strengths of both methods, making it particularly effective for high-dimensional data with correlated predictors. To assess model performance, Mean Squared Error (MSE) is utilized to quantify the alignment between predicted and actual values.

The classical linear regression assumption that explanatory variables are exogenous is often violated in Mincer equations, where the primary explanatory variable (education) is endogenous. To correct for unobserved ability and measurement error, the study employs an Instrumental Variables (IV) regression framework. The challenge lies in finding valid instruments; this research constructs a family educational background instrument by combining the education levels of parents, siblings, and spouses.

After developing the IV, the instrument's validity is tested empirically through under-identification and weak identification tests. For under-identification, the Kleibergen-Paap rk LM statistic is employed; the null hypothesis is rejected if the p-value is less than 10%, concluding that the instrument is identified. To ensure the instrument is strong, the study also employs the Cragg-Donald Wald F statistic to test for weak identification. The null hypothesis of weak identification is rejected by comparing the calculated values against the Stock-Yogo critical values at the 10%, 15%, 20%, and 25% levels, which are 16.38, 8.96, 6.66, and 5.53, respectively (Stock & Yogo, 2005).

3. Empirical Results

Empirical results are divided into three different sections. The first section presents the empirical results from OLS estimation for both years of education, and the EOM categorical variable. These results serve as a base for the comparisons. The second section presents empirical results based on shrinking methods, with bootstrapping and cross-validation, to ensure inference consistency across OLS and shrinking methods. Lastly, the estimates of years of education in the first section are likely endogenous; therefore, we employed an IV regression. Subsequently, this method will yield the maximum returns after correcting for endogeneity. Hence, our empirical results give us a range for the returns to education.

Table 4 presents the empirical results from OLS estimation. The table is divided into two main parts based on the measurement of education, i.e., years of education and EOM. Each part of the table is divided into four columns, each containing a specification. Each specification estimates returns to education using OLS to examine heterogeneity across EOM, migrants, and non-migrants. Other variables are controls, such as employment categories, regions, and provinces. This systematic forward selection method for the empirical model provides insights into the consistency and reliability of estimates, enhances the model's explanatory power, and mitigates omitted-variable bias, leading to a more comprehensive understanding of wage determinants in the Pakistani labor market.

Table 4. Empirical Results of the OLS Estimation Method.

	Dependent variable: ln (Monthly Wages)							
	Year of Education				Education-Occupation Mismatch			
	Overall	Migrants vs Non-Migrants	Migration Stream	Migration Stream with Reason	Overall	Migrants vs Non-Migrants	Migration Stream	Migration Stream with Reason
Constant	8.237***	8.244***	8.083***	8.148***	8.633***	8.641***	8.621***	8.702***
Years of Education	0.0770***	0.0769***	0.0983***	0.0961***				
EOM Base Category: Required Education								
Overeducation					0.0768***	0.0764***	0.202***	0.192***
Undereducation					-0.182***	-0.182***	-0.241***	-0.235***
Migration		0.0634***				0.0820***		
Urban Region	0.174***	0.169***			0.255***	0.249***		
Migration Stream Base Category: Rural-Rural								
Rural-Urban			0.105***	0.0958***			0.215***	0.202***
Urban-Rural			0.0852***	0.0867***			0.239***	0.231***
Urban-Urban			0.207***	0.204***			0.395***	0.381***
Migration Reason Base Category: Others								
Better Economic Opportunity				0.0609***				0.0360*
Confirmed Job				0.211***				0.348***
Female	-0.658***	-0.652***	-0.135	-0.130	-0.471***	-0.465***	0.0570	0.0552
Married	0.0861***	0.0854***	0.0794***	0.0619**	0.0529***	0.0520***	0.0264	0.00653
Female*Married	0.0218	0.0167	-0.324***	-0.297***	-0.00243	-0.00901	-0.366***	-0.333***
Age	0.048***	0.0482***	0.0571***	0.0547***	0.0671***	0.0670***	0.0827***	0.0789***
Age Squared	-0.0004***	-0.0004***	-0.0005***	-0.0005***	-0.001***	-0.0007***	-0.0008***	-0.0008***
Province Base Category: Khyber Pakhtunkhwa								
Punjab	0.0580***	0.0572***	0.0607***	0.0534**	-0.00200	-0.00301	-0.0246	-0.0252
Sindh	-0.041***	-0.040***	0.0133	0.0249	-0.064***	-0.062***	-0.0687**	-0.0389
Balochistan	0.119***	0.121***	0.0815*	0.0610	0.104***	0.106***	0.0788	0.0484
Employment Status Base Category: Employer								
Self-Employed	-0.409***	-0.408***	-0.415***	-0.402***	-0.515***	-0.513***	-0.550***	-0.526***

	Dependent variable: ln (Monthly Wages)							
	Year of Education				**Education-Occupation Mismatch**			
	Overall	**Migrants vs Non-Migrants**	**Migration Stream**	**Migration Stream with Reason**	**Overall**	**Migrants vs Non-Migrants**	**Migration Stream**	**Migration Stream with Reason**
Paid-Employed	-0.544***	-0.544***	-0.516***	-0.527***	-0.590***	-0.590***	-0.530***	-0.557***
Agriculture-Employed	-0.669***	-0.666***	-0.676***	-0.649***	-0.784***	-0.781***	-0.831***	-0.797***
N	114586	114586	8185	8185	114586	114586	8185	8185
R^2	0.296	0.296	0.390	0.396	0.196	0.197	0.225	0.241
adj. R^2	0.296	0.296	0.389	0.394	0.196	0.197	0.223	0.239

Source: Author's calculations based on PSLM, 2019-20.
*Note: ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively. The N varies across models due to the inclusion of the migration stream and variable. EOM represents the education-occupation mismatch*

The Model 1 (overall) in Table 4 serves as a baseline for comparison, suggesting that returns to education increase by about 7.7% per year of education. Furthermore, individuals who live in an urban area earn more than individuals who live in a rural area. On average, female earnings are 65.8% less than males and being married is associated with an 8.61% shift in returns. Subsequently, returns decrease at an increasing rate with age and are consistent with economic theory, i.e., concave. Lastly, there is heterogeneity in returns across employment status and provinces. In the second model, we introduced a dummy variable for migration, which is statistically significant and positive, implying that migration shifts the returns by 6.34%. However, the R^2 value of this model does not improve. In the third model, we replace the migration and urban variables with the stream variable. The inclusion of the stream variable not only improved the value of R^2 but also indicates that, compared to rural-rural migration, all other categories of migration earn significantly higher returns. However, the magnitude varies across each stream. It is important to note that migration from rural to urban or from urban to urban improves returns at higher rates, i.e., 10.5% and 20.7%, respectively. In the fourth model, we also included the reason for migration as a predictor variable, alongside the migration stream variable. This model has four important results. First, it implies that, compared to the base category, better economic opportunities and a confirmed job are significant sources of increased earnings. Second, due to the inclusion of migration stream and reason variables, the province variable becomes insignificant for Sindh and Balochistan, but significant for Punjab. It implies that, compared to other provinces, Punjab is a relatively preferred destination for increased earnings. Third, compared to the first two models, the average earnings difference between genders becomes insignificant, i.e., males and females now have the same earnings level. Fourth, if a woman is married, she earns significantly less (29.7%) than an unmarried woman, suggesting that married women's earnings may be influenced by labor-market penalties.

The second part of Table 4 presents the empirical results for the EOM, where we have replaced years of education with a categorical variable for the mismatch, while keeping the exact specification as in the first part. In the case of the Overall model, over-education is statistically positive and significant, implying that, compared to the base category (required education), an individual with over-education earns 7.68% more. In comparison, an individual with undereducation earns 18.2% less, indicating a concave relationship between returns and mismatch. All other results are approximately consistent, with only slight variation in values, except for Punjab, which shows an insignificant difference. It is important to note that in the last model (migration stream with reason), including the migration stream and reason led to drastic changes in the estimated parameters for overeducation and undereducation. In other words, the relationship changes from concave to slightly or no concave. Moreover, none of the provincial parameters is significant.

The empirical results in Table 4 can be summarized as follows: first, the returns to education increase at a rate of 9.61% per year of education; second, if an individual is overeducated, they earn significantly more than an individual with the required level of education, while an undereducated individual earns significantly less. Second, migration (stream) is associated with higher earnings than rural-rural migration. Third, on average, females earn less in the non-migrant data but earn approximately the same in the migrant data. Fourth, migration (stream) is a source of change, moving from a concave to a slightly or no concave relationship.

Table 5 presents the empirical results of the shrinking methods in comparison with OLS. To remain consistent with our empirical analysis, we use the specification from the last model in Table 4, with one modification, i.e., we have dropped the province variable because it is insignificant. Table 5 is divided into two sections: the first presents the results of conventional OLS, while the second presents result from shrinkage regression methods. In addition to shrinkage methods, we used bootstrapping and a cross-validation approach to analyze measurement error in the dependent variable and estimation precision. The first and second columns present the results of the OLS estimation method, but the second column uses bootstrapping. As the estimation method is the same, there is no difference in the estimates, but a slight difference in the standard errors. In the case of Bootstrapped OLS, the standard errors are smaller than those of conventional OLS; therefore, it yields precision in estimates. Furthermore, in the case of the shrinking methods, the estimates are slightly smaller due to the penalty that shrinks towards zero. Even after shrinking towards zero, the estimated parameter shows no drastic change, confirming that there is no overfitting or multicollinearity. Compared to OLS, Lasso estimates differ slightly due to the penalty. Overall, these results imply that there is no significant measurement error in the dependent variable, given the consistency of the results.

Table 5. Empirical Results of Shrinkage Regression Methods.

	Dependent variable: ln (Monthly Wages)				
	OLS	Bootstrapping			
		OLS	Lasso	Ridge	Elastic
Constant	8.687***	8.687***	9.456***	8.717***	8.725***
	(0.0976)	(0.102)	(0.0362)	(0.131)	(0.0978)
EOM Base Category: Required Education					
Overeducation	0.193***	0.193***	0.182***	0.193***	0.193***
	(0.0234)	(0.0232)	(0.0185)	(0.0237)	(0.0239)
Undereducation	-0.235***	-0.235***	-0.220***	-0.235***	-0.235***
	(0.0228)	(0.0202)	(0.0169)	(0.0209)	(0.0202)
Migration Stream Base Category: Rural-Rural					
Rural-Urban	0.196***	0.196***	0.118***	0.195***	0.194***
	(0.0212)	(0.0216)	(0.0136)	(0.0217)	(0.0198)
Urban-Rural	0.228***	0.228***	0.130***	0.227***	0.226***
	(0.0353)	(0.0348)	(0.0256)	(0.0390)	(0.0509)
Urban-Urban	0.372***	0.372***	0.275***	0.371***	0.371***
	(0.0234)	(0.0232)	(0.0153)	(0.0257)	(0.0252)
Migration Reason Base Category: Others					
Better Economic Opportunity	0.0350*	0.0350	0.0534***	0.0353**	0.0354*
	(0.0197)	(0.0216)	(0.0139)	(0.0175)	(0.0207)
Confirmed Job	0.352***	0.352***	0.317***	0.353***	0.352***
	(0.0274)	(0.0253)	(0.0189)	(0.0249)	(0.0269)
Female	0.0515	0.0515	-0.200***	0.0255	0.0169
	(0.118)	(0.144)	(0.0225)	(0.126)	(0.130)
Married	0.00774	0.00774	0.127***	0.00983	0.0103
	(0.0293)	(0.0269)	(0.0176)	(0.0200)	(0.0198)
Female*Married	-0.331***	-0.331***	-0.155***	-0.317***	-0.312***

	Dependent variable: ln (Monthly Wages)				
	OLS	Bootstrapping			
		OLS	Lasso	Ridge	Elastic
	(0.0648)	(0.0869)	(0.0136)	(0.0710)	(0.0721)
Age	0.0787***	0.0787***	0.00964***	0.0766***	0.0760***
	(0.00530)	(0.00588)	(0.000312)	(0.00554)	(0.00391)
Age Squared	-0.001***	-0.001***	0.0001***	-0.001***	-0.001***
	(0.00006)	(0.00007)	(0.0000066)	(0.00006)	(0.00005)
Employment Status Base Category: Employer					
Self-Employed	-0.525***	-0.525***	-0.112***	-0.518***	-0.515***
	(0.0443)	(0.0509)	(0.0179)	(0.122)	(0.0846)
Paid-Employed	-0.557***	-0.557***	-0.150***	-0.550***	-0.548***
	(0.0425)	(0.0466)	(0.0166)	(0.107)	(0.0771)
Agriculture-Employed	-0.803***	-0.803***	-0.362***	-0.796***	-0.793***
	(0.0582)	(0.0629)	(0.0354)	(0.115)	(0.0998)
N	8185	8185	8185	8185	8185
R^2	0.241	-	-	-	-
adj. R^2	0.239	-	-	-	-
Bootstrapped Replications	-	100	100	100	100
K-Folds	-	10	10	10	10

Source: Author's calculations based on PSLM, 2019-20.
*Note: ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively. Standard errors in parentheses. EOM represents EOM.*

Table 6 presents the empirical results of the employment-status-wise analysis to assess the consistency of EOM estimates. To remain consistent with our empirical analysis, we use the specification from the last model in Table 4. The estimates for overeducation and undereducation vary across employment statuses; however, there is less variation among the Paid-Employed and Self-Employed. In the Employer's model, the estimate for overeducation is 23.4%, while the estimate for undereducation is negative but statistically insignificant. Similar results hold for the Agriculture-Employed, with one additional finding: a positive sign for undereducation. In both models, it implies that undereducated individuals earn approximately the same amount as those with the required level of education. For Self-Employed and Paid-Employed, estimates are significant and in line with expectations. In relative comparisons across models, Paid-Employed faces a penalty for both undereducation and overeducation.

Table 6. Employment Status-based Empirical Results.

	Dependent variable: ln (Monthly Wages)			
	Employer	Self-Employed	Paid-Employed	Agriculture-Employed
Constant	7.289***	8.974***	7.994***	7.899***
EOM Base Category: Required Education				
Overeducation	0.234**	0.247***	0.142***	0.351**
Undereducation	-0.182	-0.206***	-0.265***	0.0151
Migration Stream Base Category: Rural-Rural				

	Dependent variable: ln (Monthly Wages)			
	Employer	Self-Employed	Paid-Employed	Agriculture-Employed
Rural-Urban	0.0111	0.0484	0.225***	0.742***
Urban-Rural	-0.0811	0.159**	0.248***	0.294
Urban-Urban	0.353**	0.263***	0.364***	0.726***
Migration Reason Base Category: Others				
Better Economic Opportunity	0.0337	0.172***	-0.0301	0.0483
Confirmed Job	-0.474***	0.151*	0.401***	0.362
Female	0.0214	-0.0715	-0.0997	1.428
Married	0.0237	0.0553	-0.0351	-0.0732
Female*Married	-0.240	-0.675***	-0.122*	-1.186***
Age	0.142***	0.0424***	0.0885***	0.0846***
Age Squared	-0.0014***	-0.00042***	-0.00089***	-0.000940**
N	327	2092	5417	349
R^2	0.321	0.306	0.236	0.288
adj. R^2	0.295	0.302	0.235	0.262

Source: Author's calculations based on PSLM, 2019-20.
*Note: ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively. EOM represents EOM.*

Based on the migration stream, urban-urban migration is significant in all models. Urban-rural is significant for Paid-Employed and Self-Employed. Furthermore, rural-urban migration is important for the Paid-Employed and Agricultural-Employed. Better economic opportunity is only valid for the Paid-Employed. Confirmed job as a reason for migration is beneficial for the Self-Employed and Paid-Employed, while it adversely affects the Employer. Lastly, except for the Employer model, all models show a significant difference in the labor market for married women, suggesting potential barriers in employment outcomes, which could be influenced by factors such as family roles and mobility constraints.

Table 7. Empirical Results of the Instrumental Regression Methods.

	Dependent variable: ln (Monthly Wages)			
	Overall	Migrants vs Non-Migrants	Migration Stream	Migration Stream with Reason
Constant	7.933***	7.938***	7.691***	7.728***
Years of Education	0.138***	0.138***	0.149***	0.149***
Migration		0.0376***		
Urban Region	0.116***	0.113***		
Migration Stream Base Category: Rural-Rural				
Rural-Urban			0.0110	0.00544
Urban-Rural			-0.00169	0.000943
Urban-Urban			0.0677**	0.0686**
Migration Reason Base Category: Others				
Better Economic Opportunity				0.0759***
Confirmed Job				0.144***
Female	-0.744***	-0.741***	0.0401	0.0269
Married	0.147***	0.147***	0.158***	0.133***

Female*Married	0.0138	0.0109	-0.441***	-0.406***
Age	0.0278***	0.0278***	0.0433***	0.0417***
Age Square	-0.0001***	-0.00018***	-0.0004***	-0.0004***
Province Base Category: Khyber Pakhtunkhwa				
Punjab	0.108***	0.107***	0.0999***	0.0945**
Sindh	-0.0189	-0.0182	0.0325	0.0413
Balochistan	0.0666***	0.0672***	0.0856	0.0696
Employment Status Base Category: Employer				
Self-Employed	-0.329***	-0.328***	-0.339***	-0.331***
Paid-Employed	-0.506***	-0.506***	-0.496***	-0.502***
Agriculture-Employed	-0.509***	-0.508***	-0.545***	-0.517***
N	49674	49674	4568	4568
R^2	0.285	0.285	0.396	0.399
adj. R^2	0.285	0.285	0.394	0.396
$\boldsymbol{H_0}$: **Under-indentification**				
Kleibergen-Paap rk LM statistic	6409.350 (p>0.000)	6397.608 (p>0.000)	853.566 (p>0.000)	854.932 (p>0.000)
$\boldsymbol{H_0}$: **Weak Identification**				
Cragg-Donald Wald F statistic	7356.775	7341.158	1046.036	1047.636

Source: Author's calculation based on PSLM, 2019-20.
*Note: ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively. Fixed effects are occupations, districts, and Provinces.*

The empirical results of IV regressions are presented in Table 7 to address the bias in the estimates of education due to endogeneity. They are verified by rejecting the null of exogeneity in the Durbin and Wu-Hausman tests at 10.275 $(p = 0.0013)$ and 9.21 $(p = 0.0042)$, respectively. We used family education as an instrument in our IV regression models. The null hypothesis of under-identification is rejected at the 1% level in all models; consequently, the null hypothesis of weak identification is also rejected, as the test statistics are well above the Stock-Yogo 10% critical values (16.38). Therefore, we can conclude that our family education as an instrument is valid.

Although the family education satisfies relevance and passes overidentification, even though the exclusion restriction family education affects wages only through own schooling remains untestable directly. In this case, there are potential violations such as (1) family networks facilitating job access, (2) social capital or cultural transmission affecting productivity conditional on schooling, and (3) inherited ability confounding human capital returns. Our estimates of the IV remain stable across specific ions controlling for occupations, employment status, regional fixed effects, which implies limited direct channels beyond education. Finally, we interpret IV estimates cautiously as local average treatment effects (LATE) for compliers influence by family education, not populations' average effects.

Compared to Table 4, the returns-to-education parameter has improved significantly in Table 7, from 9.6% to 13.8% per month. Furthermore, the estimates of migration and region are statistically significant and positive; suggesting that migration significantly improves earnings. While looking at the stream variable for the same table seems contradictory, based on the migration stream, only urban-urban migration is statistically significant and positive; compared to other migration categories, only urban-urban migration increases earnings. Moreover, both categories of migration reasons, compared to the base category, statistically improve the earnings. Based on age, earnings increase at a decreasing rate. Finally, all models confirm a significant labor market disparity faced by married women.

Given the cross-sectional nature of the data, the observed relationships between marital status, gender, and wage penalties should be viewed as associations. While these findings contribute to the understanding of gender wage disparities, they cannot confirm causal mechanisms, and further research using longitudinal data would be necessary to explore these relationships over time.

Lastly, Table 5 applies advanced statistical methods to high-dimensional controls, these methods mitigating overfitting and acting as a check for robustness and estimation stability, while the IV estimation (Table 7) provides insights into causal patterns. Ridge stabilizes EOM effects (0.017-0.019); Lasso validates key predictors; elastic net schooling returns (0.15-0.17) are aligned with the IV estimates range. The consistency of the estimates suggests the robustness of the model.

4. Discussion and Economic Significance

According to human capital theory, education is an important factor in enhancing an individual's productive skills and knowledge, thereby increasing their earning potential and employability. Our empirical results are consistent with this theory: the highest degree attained positively contributed to wages, with clear evidence of significant returns to higher levels of education.

On the contrary, the highest attained degree is only associated with higher earnings when there is alignment between the individual's education and the occupational role. Pakistan's labor market is inefficient due to EOM, leading to persistent misallocation of human capital. It leads to a wage penalty for the undereducated and limits wage premiums for the overeducated, reflecting a concave relationship between mismatch and earnings. Furthermore, it also suggests that higher education does not always translate into higher wages. These inefficiencies constrain not only individual earning potential but also broader economic growth by underutilizing the country's educational investments and human capital resources. Improving vocational guidance, implementing targeted labor-market policies, and aligning educational output with market needs are essential to unlocking the full economic benefits of human capital development in Pakistan.

Throughout our statistical and empirical results, migration consistently shows a positive shift with earnings, indicating that migrants tend to earn higher wages than non-migrants, which may reflect the selective migration of more skilled workers. The returns to education remain robust even in the presence of migration and EOM, confirming that years of education enhance earnings potential. However, the EOM continues to adversely affect earnings, especially for undereducated workers, who experience significant wage penalties. Interestingly, the coefficient for overeducation shows relatively modest positive effects on wages, indicating diminishing returns when education exceeds education-occupation. Furthermore, migration streams shows that those moving to urban centers typically earn higher incomes than those migrating across rural areas. Migration streams motivated by better economic opportunities or confirmed jobs also correspond with higher earnings, emphasizing how the reasons and destinations for migration shape labor market outcomes.

More importantly, the favorable earning premiums associated with both migration and specific migration streams, alongside the concave earning profile of mismatch, suggest that migration can partially mitigate these inefficiencies: the concavity of the mismatch-earnings relationship becomes less pronounced, and in some migration streams, it may even approach a linear or non-concave pattern. These results are consistent with the predictions of job-matching and spatial labor-market theories, which link better match quality to less concave returns. These results highlight the importance of considering both migration and migration stream in policies aimed at improving returns to education and reducing the negative impacts of the EOM, thereby maximizing labor market efficiency and income growth in Pakistan.

On the contrary, empirical results show that married women in Pakistan's labor market face persistently lower wages. Even though the estimates for females and married women show mixed results across employment categories, the interaction term for female and married is strongly negative, especially for self-employed and agriculture-employed groups (vulnerable or informal segments). This suggests that being a married woman may be associated with a greater wage penalty in these sectors relative to men and unmarried women. These disadvantages are likely influenced by structural and cultural barriers that limit married women's opportunities and bargaining power.

The varying degrees of earnings penalties across employment types highlight the need for targeted interventions: improving workplace inclusion, ensuring fair pay, and supporting female entrepreneurship and agricultural work are essential steps toward more equitable labor market outcomes and the full realization of Pakistan's human capital potential.

5. Conclusions

This study examines how returns to education in Pakistan vary with education–occupation mismatch (EOM), internal migration, and gender, using PSLM 2019–2020 data. The results confirm that schooling is a strong predictor of earnings, but that returns are shaped by labor-market frictions. EOM remains a persistent inefficiency: undereducated workers face substantial wage penalties relative to adequately matched workers, while overeducated workers receive smaller and diminishing gains, implying a concave mismatch–earnings profile. Internal migration is associated with higher earnings, especially for rural–urban and urban–urban movers, and appears to reduce mismatch-related inefficiencies by flattening the curvature of the mismatch–earnings relationship.

The findings also highlight pronounced distributional disparities. Women—and particularly married women—consistently experience wage penalties, with the largest disadvantages observed among self-employed and agricultural workers. This pattern is consistent with constraints that limit women's mobility and job-matching opportunities, including household responsibilities, restricted access to higher-quality jobs, and segmentation between formal and informal work. Overall, the evidence suggests that improving the economic payoff to human capital in Pakistan requires not only reducing EOM and facilitating mobility but also tackling structural barriers that prevent women—especially married women—from translating education into commensurate labor-market outcomes.

6. Policy Implications

To maximize economic returns on human capital investment in Pakistan, a multi-layered policy intervention is required to strengthen the alignment between the education system and the labor market. Policies should prioritize reducing education-occupation mismatch (EOM) by enhancing career guidance, improving labor-market information systems, and fostering stronger coordination between education providers and employers. Expanding demand-driven technical and vocational education and training (TVET), developing employer-linked curricula, and establishing robust certification frameworks are essential steps to ensure that the skills supplied by the education system better match occupational requirements, thereby lowering the wage penalties borne by undereducated workers and reducing inefficient overeducation.

Furthermore, supporting internal mobility is crucial for improving match quality and earnings, particularly given the wage premiums associated with rural-to-urban and urban-to-urban migration streams. Reducing mobility frictions requires a suite of priority measures, including the establishment of transparent job-matching platforms and placement services across cities, the provision of affordable and safe worker housing, and improved transport access in major labor markets. Ensuring the portability of documentation and access to basic services for migrants, alongside complementary urban labor-market policies that promote diversified formal employment, will help ensure that migration translates into higher-quality job matches rather than increased congestion in low-skill segments.

Addressing the persistent wage penalties faced by married women, particularly in vulnerable and informal sectors, remains a critical challenge for inclusive growth. Targeted interventions are necessary, such as expanding childcare availability, promoting flexible and safe work arrangements, and strengthening the enforcement of anti-discrimination laws. Additionally, improving access to finance, training, and market linkages for female entrepreneurs—especially those in self-employed and agricultural roles—is vital. Policies that enhance women's effective geographic and occupational mobility, as well as their bargaining power, are essential for narrowing gender wage gaps and ensuring that educational investments generate equitable returns across the workforce.

Finally, the success of these interventions depends on effective implementation and rigorous monitoring. All policy measures should be accompanied by measurable targets, such as specific goals for reducing mismatch incidence by occupation, narrowing gender wage gaps by sector, and improving migrant placement rates. Regular monitoring using household and labor-force data will allow for timely adjustments as labor demand, migration patterns, and economic conditions evolve, ensuring that Pakistan’s human capital potential is fully realized.

7. References

Ahsan, H. (2024). *Impact of education mismatch on earnings: Evidence from Pakistan’s labor market* (PIDE Working Paper No. 2024:1). Pakistan Institute of Development Economics.

Akhtar, M., Javed, M., & Noreen, S. (2018). Analysis of education occupation mismatch at Pakistani educational institutions. *Journal of Educational Research*, *21*(2), 175–189.

Akram, N., Hamid, A., & Gulza, A. (2024). Measuring educational returns and productivity in Pakistan. *Journal of Asian Development Studies*, *13*(4), 1121–1130. https://doi.org/10.62345/jads.2024.13.4.91

Ashraf, J. (2011). New evidence on rates of return to education in Pakistan. *Global Journal of Business Research*, *5*(3), 113–120.

Aslam, M. (2009). Education gender gaps in Pakistan: Is the labor market to blame? *Economic Development and Cultural Change*, *57*(4), 747–784. https://doi.org/10.1086/598761

Awan, S. Z. (2020). Financial misgivings of married working women in Lahore. *Pakistan Perspectives, 25*(1), 49–68.

Bahl, S., & Sharma, A. (2023). Heterogeneity among migrants, education–occupation mismatch and returns to education. *Regional Studies*, *57*(9), 1851–1865. https://doi.org/10.1080/00343404.2022.2154448

Becker, G. S. (1975). *Human capital: A theoretical and empirical analysis*. National Bureau of Economic Research.

Becker, G. S. (1993). *Human capital: A theoretical and empirical analysis, with special reference to education* (3rd ed.). University of Chicago Press.

Budría, S., & Moro-Egido, A. (2014). Overqualification, skill mismatches and wages in private sector employment in Europe. *Technological and Economic Development of Economy, 20*(3), 457–483. https://doi.org/10.3846/20294913.2014.883341

Colclough, C., Kingdon, G., & Patrinos, H. (2009). *The pattern of returns to education and its implications*. Research Consortium on Educational Outcomes & Poverty.

Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review*, *1*(1), 75–86. https://doi.org/10.1016/0272-7757(81)90028-5

Dustmann, C. (2023). The economic impacts of migration: A review. *Journal of Economic Perspectives*, *37*(4), 45–68. https://doi.org/10.1257/jep.37.4.45

Farooq, S. (2011). Mismatch between education and occupation: A case study of Pakistani graduates. *The Pakistan Development Review*, *50*(4), 531–552. https://doi.org/10.30541/v50i4pp.531-552

Feliciano, C. (2021). Educational mismatch and the labor market outcomes of migrants. *International Migration Review*, *55*(2), 494–519. https://doi.org/10.1177/0197918320942446

Greenwood, M. J. (1997). Internal migration in developed countries. In M. R. Rosenzweig & O. Stark (Eds.), *Handbook of population and family economics* (Vol. 1B, pp. 647–720). Elsevier.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed., pp. 485–585). Springer.

Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., Wild, J., Krauss, S., Goretzko, D., & Stachl, C. (2021). Machine learning for the educational sciences. *Review of Education*, *9*(3), e3310. https://doi.org/10.1002/rev3.3310

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.

Khan, M. Z., & Said, R. (2022). Incidence of educational mismatch and earning in Pakistan. *PLOS ONE*, *17*(5), e0268008. https://doi.org/10.1371/journal.pone.0268008

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, *2*, 1137–1145.

Leuven, E., & Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 4, pp. 283–326). North-Holland.

Mincer, J. A. (1974). *Schooling, experience, and earnings*. National Bureau of Economic Research.

Nasir, Z. M., & Nazli, H. (2000). *Education and earnings in Pakistan* (Research Report No. 177). Pakistan Institute of Development Economics.

Pakistan Institute of Development Economics. (2024). *Labour market issues in Pakistan: Unemployment, working conditions, and wage disparities*. https://pide.org.pk/research/

Poot, J., & Stillman, S. (2016). Skill composition of immigration flows and the measurement of education–occupation mismatch. *IZA Journal of Migration*, *5*(1), 18. https://doi.org/10.1186/s40176-016-0066-y

Quinn, M. A., & Rubb, S. (2005). The importance of education–occupation matching in migration decisions. *Demography*, *42*(1), 153–167. https://doi.org/10.1353/dem.2005.0008

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

Sitorus, D. A., & Wicaksono, F. (2022). The effect of educational mismatch on wages: Migrant vs. native workers in the Indonesian labor market. *Bulletin of Indonesian Economic Studies*, *58*(1), 1–22. https://doi.org/10.1080/00074918.2020.1837965

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference in econometric models: Essays in honor of Thomas J. Rothenberg* (pp. 80–108). Cambridge University Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tilak, J. B. (2018). The political economy of education in India. In J. B. Tilak (Ed.), *Education and development in India* (pp. 33–84). Palgrave Macmillan.

Verdugo, R. R., & Verdugo, N. T. (1989). The impact of surplus schooling on earnings: Some additional findings. *Journal of Human Resources*, *24*(4), 629–643. https://doi.org/10.2307/145998

World Bank. (2021). *Women, business and the law: Pakistan country profile*. https://wbl.worldbank.org/en/data/exploreeconomies/pakistan/2021

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x