mathematics

MDPI

Article

Changing Wage Effects of Educational Mismatch in China: Evidence from Threshold IV–Selection Models

Lulu Jiang [1], Woraphon Yamaka [2] and Paravee Maneejuk [2,*]

[1] Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand; lulu_jia@cmu.ac.th or miss_jianglu@163.com
[2] Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand; woraphon.yamaka@cmu.ac.th
* Correspondence: paravee.m@cmu.ac.th

Abstract

This study examines the wage effects of educational mismatch in China by jointly addressing sample selection, endogeneity, and nonlinear career-stage heterogeneity within a unified econometric framework. Although educational mismatch has been widely studied, existing evidence largely relies on linear models that overlook experience-dependent wage dynamics and potential selection and endogeneity biases. Using data from the 2020 wave of the China Family Panel Studies (CFPS), this study extends the Duncan–Hoffman model by integrating a sample-selection-corrected threshold regression estimated via instrumental variables. This approach allows the identification of experience thresholds at which the wage effects of overeducation and undereducation differ across regimes. The results reveal pronounced nonlinearities in mismatch-related wage differentials. Overeducation is associated with wage penalties at early career stages, but these penalties weaken and, in some cases, disappear once workers surpass the estimated experience threshold. In contrast, undereducation yields modest wage premiums early in the career but becomes increasingly penalized at higher experience levels. Substantial gender heterogeneity is also observed: male workers are better able to use accumulated experience to offset educational shortfalls, whereas female workers face more persistent penalties, particularly at later career stages.

Keywords: China; education mismatch; sample selection model; threshold effects; work experience

MSC: 66P20; 91B84; 62F10

Check for updates

Academic Editors: Alessandro Mazzoccoli, Raffaele Mattera and J.E. Trinidad-Segovia

Received: 29 January 2026
Revised: 6 March 2026
Accepted: 6 March 2026
Published: 9 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Labor constitutes a core component of human capital in modern economic activities, particularly in economies that rely on skilled and well-trained workers. Higher skill levels are generally associated with greater productivity and higher wages. Since 1999, China has experienced a rapid expansion of higher education. According to the National Education Development Statistical Bulletin, annual higher education enrollment increased from 1.6 million in 1999 to 10.145 million in 2022, while the gross enrollment rate rose from 10.5% to 59.6%. Although this expansion has substantially improved access to tertiary education, it has also contributed to growing educational mismatch, as acquired qualifications do not always align with labor market requirements [1].

Educational mismatch, encompassing both overeducation and undereducation, has become an increasingly prominent topic in labor market research [2–6]. Overeducation occurs

when workers possess more schooling than their jobs require, whereas undereducation arises when workers' schooling falls short of job requirements [7].

The impact of educational mismatch on earnings has been extensively examined, yet empirical findings remain inconclusive. Some studies document a wage penalty for overeducation and a wage premium for undereducation [8–11], while others report the opposite pattern [12,13]. For example, Xu and Chen [13], using CFPS data, find that overeducation is positively associated with income, whereas undereducation is negatively correlated with earnings. These conflicting results suggest that the wage effects of mismatch may depend on contextual and structural factors not fully captured in standard linear specifications. Moreover, recent structural transformations in labor markets may further complicate the mismatch–wage relationship. Rapid technological change, digital transformation, and the diffusion of artificial intelligence are reshaping task content and altering the relative importance of formal credentials and accumulated experience [14,15]. Such developments may increase heterogeneity in wage determination across career stages, reinforcing the possibility that the returns to educational mismatch vary over the life cycle rather than remaining constant.

Against this background, one potential explanation for the mixed evidence concerns the role of work experience. Although experience has been incorporated in the pioneering work of Duncan and Hoffman [16], it has typically been treated as an independent determinant of wages rather than as a moderating factor in the mismatch–wage relationship. This study argues that experience may fundamentally alter how educational mismatch affects earnings.

Although experience has been incorporated in the pioneering work of Duncan and Hoffman on the educational mismatchwage equation [16], it has traditionally been treated as an independent factor influencing wages rather than as a key determinant in the educational mismatchwage relationship. In general, experience enhances a worker's ability to apply acquired knowledge effectively, develop industry-specific skills, and adapt to job demands. As workers gain more experience, they accumulate tacit knowledge, improve their problem-solving abilities, and enhance their productivity, leading to higher wages. Conversely, less experienced workers may struggle to fully utilize their educational qualifications due to limited exposure to job-specific skills and fewer opportunities for workplace learning [17]. This gap in skill application not only affects individual wage growth but also weakens overall knowledge spillovers in the labor market, reducing the collective benefits of education. Consequently, educational mismatch may impose a more significant penalty on less experienced workers, who cannot compensate for skill gaps through accumulated work experience. In contrast, experienced workers may better navigate mismatch situations, mitigating potential wage penalties through acquired adaptability and workplace learning [18].

Another important issue that arises is the length of experience required to alter the educational mismatchwage relationship. While existing studies acknowledge the significance of experience, they do not explore how many years of experience are needed for this relationship to change. This gap remains inadequately examined, particularly in the context of China. To address this, this study introduces a threshold regression framework applied to the Duncan and Hoffman equation, allowing for a more precise identification of experience thresholds where wage effects shift significantly.

Traditionally, estimating returns to educational mismatch relies heavily on the Duncan and Hoffman equation, which links the logarithm of wages with educational mismatch, work experience, and other control variables [19]. However, this standard approach has significant limitations when applied to real-world labor market conditions. Specifically, the quadratic specification typically used in the traditional model assumes a smooth and

symmetrical wage-experience relationship, which may inadequately reflect actual labor market dynamics characterized by sudden shifts or structural breaks at critical experience thresholds. Moreover, quadratic specifications often yield distorted marginal effects at extreme values of experience, potentially leading to misleading policy implications [17]. By employing a threshold regression approach, this study overcomes these limitations, offering a more flexible and data-driven analysis of how experience influences the wage effects of educational mismatch. This innovation is especially relevant in labor markets characterized by distinct wage trajectories, such as significant differences between early-career growth and mid- or late-career stages, which are influenced by factors like job mobility, career transitions, and industry-specific wage structures [20]. Although threshold effects in human capital have been explored previously [17], their explicit application within the Duncan and Hoffman framework remains limited, underscoring the methodological novelty of this study.

Another significant concern when estimating returns to education is sample selection bias. The standard Duncan and Hoffman model implicitly assumes that observed wages represent the entire population. However, earnings often suffer from non-random selection, especially in economies with high rates of self-employment, informal employment, or labor market non-participation, leading to biased and inconsistent estimates [21]. To mitigate this bias, the study incorporates the Heckman [21] two-step selection model into the threshold Duncan and Hoffman regression. Initially, a selection equation employing a probit model estimates the likelihood of employment, generating an inverse Mills ratio (IMR). Subsequently, this ratio is incorporated into the second-stage threshold wage equation to correct for selection bias [22]. Furthermore, to address endogeneity concerns, we implement the heteroskedasticity-based IV estimator of Lewbel [23], which generates instruments internally from exogenous variables under testable heteroskedasticity conditions.

This study contributes to the literature in three main respects. First, although prior research documents wage penalties and premiums associated with overeducation and undereducation, existing studies typically impose linear specifications or model heterogeneity through quantile regression, interaction terms, or spline functions. This study instead introduces a threshold regression framework that endogenously determines an experience cutoff within the Duncan–Hoffman (ORU) wage model. By allowing the returns to required education, overeducation, and undereducation to shift discretely across experience regimes, the model uncovers nonlinear career-stage dynamics that are not captured by smooth functional forms or distributional approaches. The threshold structure identifies whether and when mismatch effects change across the work lifecycle, rather than assuming continuous variation. Second, while gender differences in returns to education are widely examined, little attention has been given to whether experience thresholds themselves differ across genders. By estimating gender-specific threshold models, this study shows that the timing and magnitude of mismatch effects vary substantially between men and women. This regime-dependent perspective offers a structural interpretation of gender wage disparities that complements existing linear decomposition approaches. Third, methodologically, the paper integrates threshold estimation with sample selection correction and heteroskedasticity-based instrumental variable identification. Standard mismatch studies typically rely on OLS, quantile regression, or IV methods in isolation. In contrast, this framework simultaneously addresses nonlinear regime shifts, potential endogeneity of education-related variables, and non-random labor force participation within a unified econometric structure.

The remainder of the study is structured as follows. Section 2 reviews the relevant literature. Section 3 describes the data and methodology. Section 4 presents empirical

results and Section 5 concludes the study with policy implications and recommendations for future research.

2. Literature

2.1. Theoretical Background

Educational mismatch, defined as the discrepancy between a worker's educational attainment and the requirements of their job, has attracted substantial attention in labor economics. Educational mismatch can manifest as either overeducation, where individuals possess more education than their job requires, or undereducation, where workers have less education than is required. There are two theories that might help to explain educational mismatch in China [24]. The first is signaling theory of Spence [25]. According to this theory, individuals may over-invest in education to signal their superior productivity, ability, or potential to prospective employers. In competitive labor markets, such signaling behavior can lead to a surplus of highly educated individuals whose qualifications exceed job requirements, resulting primarily in overeducation. This mechanism is particularly relevant in China, where the rapid expansion of higher education has substantially increased the supply of graduates. Between 1999 and 2022, annual higher-education enrollment increased from 1.6 million to over 10 million, and the gross enrollment rate rose from 10.5% to nearly 60%. Empirical studies document that educational mismatch and overeducation have become increasingly prevalent in China's labor market [1,24,26]. These trends suggest that intensified competition among graduates may reinforce incentives to pursue higher credentials as a differentiation strategy. The second relevant framework is job competition theory of Thurow [27]. This theory posited that individuals compete for available jobs based on their relative positions in the educational hierarchy. Employers rank applicants according to their education levels rather than matching education directly to specific job requirements. Consequently, intense job competition prompts individuals to acquire more education to improve their chances of employment, potentially exacerbating overeducation. However, undereducation can also occur, particularly in rapidly expanding or evolving industries where the labor force has not yet fully adapted to new educational requirements. These theories align closely with China's current labor market conditions, characterized by large population size, high competition, and significant educational mismatches.

The relationship between education and wages has been also primarily grounded in human capital theory [28,29], while the association between mismatched education (both overeducation and undereducation) and wages has been developed within the framework of assignment theory [30]. Both perspectives can be formalized through the Mincerian wage equation [31] and Duncan and Hoffman equation, respectively. According to human capital theory, additional investments in education generally lead to higher productivity and increased wages, but with diminishing marginal returns. This diminishing effect implies that the incremental wage gains from each additional year of education become smaller at higher education levels. Applying this theoretical framework to educational mismatch, it is plausible that wage returns to surplus education (overeducation) diminish when workers possess qualifications exceeding job requirements. In other words, the marginal wage benefit of being overeducated becomes smaller as the mismatch intensifies. Conversely, undereducated workers might initially face wage penalties due to their insufficient qualifications [16]. However, these penalties may diminish over time as individuals compensate through experience, on-the-job training, or informal learning [9]. Thus, theoretically, both overeducation and undereducation could exhibit diminishing marginal effects on wages, consistent with the principles outlined in human capital and assignment theories.

2.2. Theoretical Basis for Experience as Threshold

In modeling the relationship between educational mismatch and wages, Duncan and Hoffman [16] introduced a quadratic term for experience to capture diminishing returns. While this specification accounts for nonlinearity in experience, it treats the impact of mismatch as a smooth and continuous process. This assumption may be inconsistent with both human capital and assignment theories, which suggest that the role of experience in moderating mismatch can involve discrete regime shifts rather than gradual changes. From a human capital perspective, experience not only accumulates skills but also functions as a prerequisite for absorbing advanced knowledge. Zeng and Zhang [17] emphasize that human capital externalities depend on reaching a minimum threshold of initial endowments, such as literacy or basic work experience. Below this threshold, spillovers and on-the-job learning are limited; once it is crossed, individuals can effectively convert experience into productivity gains, thereby altering the wage impact of educational mismatch [18]. Assignment theory also supports threshold dynamics. Early in careers, employers rely heavily on formal education as a screening device [25,27], making undereducation particularly costly and overeducation less beneficial. Once workers accumulate sufficient experience, however, observable performance begins to outweigh credentials in job assignment and promotion decisions [30]. This shift implies that the wage effects of mismatch may change abruptly at certain experience levels, for instance, when workers transition from probationary status to permanent contracts, or from junior to senior positions.

Institutional features of labor markets further reinforce the plausibility of thresholds. Many firms structure pay and promotion along discrete seniority bands, where crossing a tenure threshold grants eligibility for supervisory roles, advanced training, or higher-level tasks [32]. These transitions create structural breaks in how surplus, or deficit education is rewarded. As a result, experience does not merely exert a smooth nonlinear effect but can generate threshold effects in the mismatch–wage relationship.

Therefore, this study proposes incorporating experience as a threshold variable influencing the impact of educational mismatch, specifically overeducation and undereducation, on wages. The theoretical intuition behind this approach is that undereducated workers can compensate for their educational deficit through extensive practical experience and on-the-job training, substituting formal qualifications and gradually offsetting initial wage penalties. Similarly, substantial experience may mitigate or neutralize the negative wage effects associated with educational mismatches, reflecting employers' valuation of productivity derived from practical skills and experience. Thus, this study aims to extend Duncan and Hoffman's model by explicitly integrating a threshold effect for experience, enhancing our understanding of the nuanced relationship between educational mismatch and wages.

Drawing on the above theoretical framework, this study tests the following hypotheses:

H1. *The wage effects of overeducation differ across career stages, with workers in different experience regimes exhibiting distinct returns to excess schooling.*

H2. *The wage effects of undereducation also vary with work experience, as accumulated job-specific skills can partially offset initial qualification deficits.*

H3. *Experience threshold segments workers into heterogeneous wage-determination regimes.*

H4. *The impacts of educational mismatch across experience regimes differ between men and women due to gendered labor-market dynamics.*

2.3. Empirical Evidence

The effect of overeducation on wages was first comprehensively examined by Duncan and Hoffman [16]. Their seminal study revealed that workers whose educational qualifications exceed the requirements of their jobs—commonly referred to as overeducated workers—earn lower wages compared to similarly educated workers whose qualifications match job requirements. This earnings difference is widely described in literature as the "wage penalty" associated with overeducation. However, subsequent research investigating the impact of educational mismatch on wages has yielded inconsistent results.

Commonly employed frameworks in analyzing educational mismatch and wage disparities include the Duncan and Hoffman approach, the Verdugo and Verdugo model [33,34], and the Over-Required and Undereducation (ORU) model [7,10,35]. These methodologies have been extensively used to understand and quantify the wage penalties associated with educational mismatch.

Numerous empirical analyses indicate that overeducated workers generally earn less than their adequately matched counterparts with similar educational qualifications, while undereducated workers often earn more than their less-educated peers. These phenomena are referred to as the penalty effect of overeducation and the premium effect of undereducation [12,36,37]. Recent evidence from Cameroon by Njifen and Smith [38] confirmed that overeducation results in a wage penalty, whereas undereducation leads to a wage premium, highlighting significant earnings disparities based on the alignment of education and job requirements.

However, these findings seem to contradict traditional labor economics theories, particularly human capital theory, which argue that education and training represent key forms of human capital investment. Generally, wage returns are expected to positively correlate with educational investments in the long term [39]. To explain the existence of educational mismatches, several alternative theories have been proposed in the literature. Some studies suggested that mismatches arise because certain essential components of human capital, such as work experience, have been overlooked [40,41]. According to Nordin et al. [42], education-specific skills and work experience may serve as substitutes to some extent. Büchel and Mertens [43] also suggested that work experience can mitigate wage penalties, especially in roles where skills acquisition and on-the-job training opportunities exist. Similarly, Green and McIntosh [40] and Green and Zhu [44] found that overeducated workers gradually close the wage gap with well-matched workers as they accumulate experience. These studies collectively indicate that work experience can serve as a compensating factor by enhancing productivity, thereby positively influencing income.

In the context of China, research on overeducation remains limited due to data constraints, with most studies published in Chinese journals being highly descriptive and primarily focused on documenting its incidence [24,26,45]. Zhu [45] found that mismatched graduates in China experience lower wage penalties than their Western counterparts, with some even achieving higher wages due to skill acquisition. Zheng et al. [24] reported that nearly half of online job seekers were overeducated, facing an estimated 5.1% wage penalty. Jiang and Guo [26] further examined different types of job mismatches, identifying skill-mismatch as imposing the most significant wage penalty. Shi [46] expanded this discussion by emphasizing the broader productivity losses resulting from skill mismatches in China's labor market.

However, these studies largely overlook the role of work experience in shaping the wage effects of educational mismatch. Experience can either mitigate or exacerbate wage penalties depending on a worker's ability to adapt and acquire job-specific skills over time. Existing empirical models fail to account for potential structural breaks in the educational mismatchwage relationship at different experience levels. To address this gap, this study

employs a threshold regression model, which provides a more precise framework for identifying how different levels of experience influence the wage effects of educational mismatch, offering deeper insights into China's labor market.

3. Methodology

3.1. Data and Variables

The existing literature identifies three primary approaches to measuring educational mismatch: the job-analysis method (objective), the worker self-assessment method (subjective), and the statistical method based on realized matches [47]. Owing to data availability, this study adopts the worker self-assessment method, which relies on individuals' evaluations of the educational requirements of their jobs. Compared with occupation-based or realized-match measures, the self-assessment approach captures job-specific skill requirements more directly and allows for within-occupation heterogeneity, which is particularly relevant in diverse and rapidly evolving labor markets. However, we acknowledge that subjective evaluations may be influenced by individual perceptions and reporting bias. Therefore, while this approach provides flexibility and job-level precision, potential measurement error cannot be entirely ruled out and constitutes a limitation of the analysis. A detailed description of all variables used in the analysis is provided in Table 1. This study uses data from the 2020 wave of the CFPS, the most recent survey year available. The CFPS is a nationally representative survey that collects detailed information on individuals' demographic characteristics, education, employment, and income through face-to-face interviews with eligible respondents. After data cleaning, the final analytical sample consists of 2926 observations. A key CFPS survey question asks respondents: "From the perspective of knowledge and skills, how much formal education do you think is needed to be competent for this job?" This question provides a self-reported measure of the education level required for the respondent's current job. Based on this self-reported job requirement, educational mismatch is defined as follows. A worker is classified as overeducated if their actual level of education exceeds the perceived educational requirement of the job, undereducated if their education falls short of the requirement, and well matched if actual and required education levels are equal. By relying on workers' own assessments, this approach captures perceived job requirements directly, rather than inferring mismatch from external occupational classifications or aggregate statistical benchmarks.

Table 1. Variable list.

Variables	Meaning	Variable Type
lninc	Log of real monthly income, deflated by CPI (2010 = 100) (Yuan/month)	numerical
D^o	D^o = 1 if overeducation; otherwise = 0	dummy
D^u	D^u = 1 if undereducation; otherwise = 0	dummy
A	Actual years of schooling attained	numerical
E^o	Absolute number of surplus years of schooling (overeducation intensity, actual years of schooling attained > number of years of schooling required to perform the current job)	numerical
E^u	Absolute number of missing years of schooling (undereducation intensity, (overeducation intensity, actual years of schooling attained < number of years of schooling required to perform the current job))	numerical

Table 1. *Cont.*

Variables	Meaning	Variable Type
exper	working experience	numerical
employ	employ = 1 if employed; otherwise = 0	dummy
gender	gender = 1 if male; otherwise = 0	dummy
age	Age (year)	
health_2	health_2 = 1, if unhealthy; otherwise = 0	dummy
health_3	health_3 = 1, if relatively healthy; otherwise = 0	dummy
health_4	health_4 = 1, if healthy; otherwise = 0	dummy
health_5	health_5 = 1, if super healthy; otherwise = 0	dummy
hukou	hukou = 1, if urban; otherwise = 0	dummy
marriage	marriage = 1, if married; otherwise = 0	dummy
party	party = 1, if he or she is member of the communist party; otherwise = 0	dummy

Note: Our dependent variable is the log of real monthly income (lninc). Nominal wages from CFPS are deflated using the China Consumer Price Index (CPI) published by the National Bureau of Statistics. We normalize 2010 as the base year (CPI = 100), so that 2020 wages are expressed in constant 2010 prices. This ensures comparability across waves by removing inflationary effects.

3.2. Baseline Model

The income equation proposed by Duncan and Hoffman [16] has become a cornerstone of applied labor economics. This empirical analysis adopts the educational mismatch framework introduced by Duncan and Hoffman [16], which relates wages to the degree of congruence between a worker's actual schooling and the level of schooling required for the job. Let A_{i} denote the actual years of schooling attained by individual i, and let R_{i} represent the number of years of schooling required to perform the current job, as reported by the respondent.

While the original Duncan–Hoffman model decomposes actual schooling into required, surplus, and deficit components, this study follows the absolute-value specification commonly used in the educational mismatch literature to obtain non-negative and directly interpretable measures of overeducation and undereducation. Specifically, the mismatch components are defined as:

$$\begin{aligned} E_i^o &= \left|A_i - R_i\right| \times D_i^o, \\ E_i^u &= \left|A_i - R_i\right| \times D_i^u, \end{aligned} \tag{1}$$

where the indicator variables $\mathrm{D_i^o}$ and $\mathrm{D_i^u}$ identify the direction of mismatch:

$$\begin{aligned} D_i^o &= \begin{cases} 1, & A_i > R_i \\ 0, & A_i \le R_i \end{cases}, \\ D_i^u &= \begin{cases} 1, & A_i < R_i \\ 0, & A_i \ge R_i \end{cases}, \end{aligned} \tag{2}$$

Thus, E_{i}^o measures the absolute number of surplus years of schooling for workers whose actual education exceeds job requirements, while $\mathrm{E_i^u}$ measures the absolute number of missing years of schooling for workers whose education falls short of job requirements. For matched workers $(A_i = R_i)$, both E_{i}^o and E_{i}^u equal zero. This absolute-value formulation avoids the sign-interpretation problem that arises when undereducation is defined using negative years of mismatch.

To estimate the Duncan and Hoffman income equation, we specify the following income equation:

$$\ln(inc_i) = \theta_0 + \theta_1 A_i + \theta_2 \mathrm{E}_i^u + \theta_3 \mathrm{E}_i^o + \theta_4 \mathrm{exper}_i + \varepsilon_i \tag{3}$$

However, to estimate this equation, selection bias must be addressed to account for both employed and unemployed individuals. A binary Probit model is used to determine an individual's employment status. Following Heckman [21], we estimate a selection equation to determine a worker's employment status. Consequently, a two-step sample selection model is employed. In the first stage, a Probit model is specified to model employment selection.

$$\begin{aligned} P(employ_i = 1) &= Z'_i \boldsymbol{\beta} \\ &= \Phi(\beta_0 + \beta_1 A_i + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 gender_i + \beta_5 health_2_i + \beta_6 health_3_i \\ &+ \beta_7 health_4_i + \beta_8 health_5_i + \beta_9 hukou_i + \beta_{10} mar_i + \beta_{11} party_i), \end{aligned} \tag{4}$$

where $P(employ_i = 1)$ is an individual's employment status. In the second stage, the Inverse Mills Ratio (IMR), estimated through the Probit model, is used to correct for sample selection bias.

$$IMR_i = \frac{\phi(Z'_i \boldsymbol{\beta})}{\Phi(Z'_i \boldsymbol{\beta})}, \tag{5}$$

where $\phi(\cdot)$ is the density function of the standard normal distribution. $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. It is included as an additional explanatory variable in the income equation.

$$\ln(inc_i) = \theta_0 + \theta_1 A_i + \theta_2 E_i^u + \theta_3 E_i^o + \theta_4 exper_i + \theta_5 IMR_i + \varepsilon_i \tag{6}$$

3.3. Threshold Regression Model

Considering the potential nonlinear relationship between educational mismatch and income, this study extends the Duncan and Hoffman framework by introducing a threshold structure based on work experience. The threshold specification allows the returns to required education and mismatch components to vary across career stages, thereby relaxing the restrictive assumption of constant coefficients imposed by the linear ORU model. In particular, work experience is treated as the threshold variable, capturing the idea that the economic consequences of overeducation and undereducation may differ between early and later career phases.

The sample is partitioned endogenously according to an estimated threshold value of work experience, γ, which determines regime membership. Individuals with $exper_i \leq \gamma$ belong to the first regime, while those with $exper_i > \gamma$ belong to the second regime. The resulting threshold Duncan–Hoffman income equation is specified as a piecewise function with regime-dependent coefficients, allowing the wage effects of educational mismatch to differ across experience intervals.

$$lninc_i = \begin{cases} \theta_0^{(1)} + \theta_1^{(1)} A_i + \theta_2^{(1)} \mathrm{E}_i^u + \theta_3^{(1)} \mathrm{E}_i^o + \theta_4^{(1)} exper_i + \theta_5^{(1)} IMR_i + \varepsilon_{1i}, & if \quad exper_i \leq \gamma \\ \theta_0^{(2)} + \theta_1^{(2)} A_i + \theta_2^{(2)} \mathrm{E}_i^u + \theta_3^{(2)} \mathrm{E}_i^o + \theta_4^{(2)} exper_i + \theta_5^{(2)} IMR_i + \varepsilon_{2i}, & if \quad exper_i > \gamma \end{cases} \tag{7}$$

where $\theta^{(1)}$ and $\theta^{(2)}$ are regime-dependent coefficients. γ represents the threshold value of work experience.

A well-recognized concern in cross-sectional wage equations is the potential endogeneity of educational mismatch variables ($\widehat{E_i^u}$ and $\widehat{E_i^o}$) as well as years of schooling (A_i) which may arise from omitted ability bias, measurement error, or simultaneity between education choices and earnings outcomes. If unaddressed, such endogeneity can lead to biased and inconsistent parameter estimates. Conventional instrumental variable (IV)

approaches require external instruments that satisfy both relevance and exclusion restrictions; however, identifying credible external instruments is often infeasible in practice. To overcome this challenge, we employ the heteroskedasticity-based IV estimator developed by Lewbel [23], which constructs instruments internally from the model's own exogenous variables. This method exploits the presence of heteroskedasticity in first-stage residuals to achieve identification without relying on traditional external instruments.

In our specification, the continuous exogenous variables used for instrument generation are age and a health index ranging from 1 ("unhealthy") to 5 ("super healthy"). While health status enters the selection equation as a set of categorical dummies, the continuous health index is treated as an exogenous regressor in the Lewbel procedure. Importantly, these variables are included directly in the wage equation as controls. Therefore, identification does not rely on excluding age or health from the structural equation. Instead, exclusion pertains to the interaction terms generated under heteroskedasticity, not to the level effects of these variables.

The Lewbel estimator proceeds in two steps. First, each endogenous variable G is regressed on all exogenous variables $X = \{age, Health\}$ in the structural equation to obtain the first-stage residuals.

$$G_i = \pi_0 + \pi' X + u_{G_i} \tag{8a}$$

Second, each continuous exogenous regressor is demeaned and interacted with these residuals to form the generated instruments.

$$IV_{G_j} = \left(X_{ij} - \overline{X}_j\right) u_{G_i} \tag{8b}$$

These instruments are valid under the condition that the covariance between the exogenous variables and the product of structural and first-stage errors satisfies $Cov(X_i, u_i \varepsilon_i) \neq 0$, while maintaining $E[X_i u_i] = 0$.

This procedure is repeated for all three endogenous variables A_i, $\widehat{E_i^u}$ and $\widehat{E_i^o}$, producing a total of six internally generated instruments. The predicted values of education and educational mismatch ($\widehat{A}_i$, $\widehat{E_i^u}$ and $\widehat{E_i^o}$) from the Lewbel-2SLS first stage are then substituted into the threshold regression model to estimate the effect of educational mismatch on log monthly income

$$lninc_i = \begin{cases} \theta_0^{(1)} + \theta_1^{(1)} \widehat{A}_i + \theta_2^{(1)} \widehat{E_i^u} + \theta_3^{(1)} \widehat{E_i^o} + \theta_4^{(1)} exper_i + \theta_5^{(1)} IMR_i + \varepsilon_{1i}, & if \quad exper_i \leq \gamma \\ \theta_0^{(2)} + \theta_1^{(2)} \widehat{A}_i + \theta_2^{(2)} \widehat{E_i^u} + \theta_3^{(2)} \widehat{E_i^o} + \theta_4^{(2)} exper_i + \theta_5^{(2)} IMR_i + \varepsilon_{2i}, & if \quad exper_i > \gamma \end{cases} \tag{9}$$

By integrating Lewbel's heteroskedasticity-based IV method with the threshold regression framework, this approach addresses potential endogeneity in educational mismatch measures when valid external instruments are unavailable. This integration enhances the consistency of parameter estimates and yields richer insights into how educational mismatch interacts with work experience to influence income across different regimes.

In the second step, the threshold parameter γ is treated as unknown and estimated using a grid-search procedure. To ensure reliable estimation and avoid regions with sparse observations, the search domain is restricted to the interior of the empirical distribution of labor market experience. Specifically, the candidate threshold values are selected from the interval

$$\Gamma = \{\gamma : q_{0.10}(exper) \leq \gamma \leq q_{0.90}(exper)\} \tag{10}$$

where $q_{0.10}(exper)$ and $q_{0.90}(exper)$ denote the 10th and 90th percentiles of the experience distribution, respectively.

For each candidate value $\gamma \in \Gamma$, regime-specific effects are introduced through an indicator function defined as

$$\mathbf{1}(exper_i \leq \gamma) = \begin{cases} 1 & \text{if } exper_i \leq \gamma, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Similarly, observations belonging to the higher-experience regime are identified by $\mathbf{1}(exper_i > \gamma)$. Let M_i denote the educational mismatch indicator, which takes the value

$$M_i = \begin{cases} 1 & \text{if individual } i \text{ is mismatched,} \\ 0 & \text{otherwise.} \end{cases}$$

The threshold wage equation can therefore be written as

$$\ln inc_i = X_{i'1}\boldsymbol{\theta}^{(1)}\mathbf{1}(exper_i \leq \gamma) + X_{i'}\boldsymbol{\theta}^{(2)}\mathbf{1}(exper_i > \gamma) + \varepsilon_i, \tag{12}$$

where X_i denotes the vector of explanatory variables, including education, educational mismatch measures, work experience, and the estimated inverse Mills ratio IMR_i obtained from the first-step selection equation. Because education and mismatch variables may be endogenous, the model is estimated using 2SLS. For each candidate value γ, the corresponding regression is estimated using the Lewbel-generated instruments together with the control variables. Let $\hat{\varepsilon}_i(\gamma)$ denote the residual obtained from the 2SLS estimation under threshold value γ. The associated sum of squared residuals is then computed as

$$SSR(\gamma) = \sum_{i=1}^{N} \hat{\varepsilon}_i(\gamma)^2 \tag{13}$$

The threshold estimator is defined as the value of γ that minimizes the sum of squared residuals over the candidate set Γ:

$$\hat{\gamma} = \text{argmin}_{\gamma \in \Gamma} SSR(\gamma) \tag{14}$$

This procedure follows the least-squares threshold estimation framework proposed by Hansen [48], adapted here to an instrumental-variable setting with selection correction through the inclusion of the estimated inverse Mills ratio.

To assess the statistical uncertainty surrounding $\hat{\gamma}$, we implement a bootstrap procedure applied to the threshold regression. Specifically, we resample observations from the outcome sample and re-estimate the threshold model in each replication while conditioning on the estimated inverse Mills ratio and first stage fitted values. The resulting empirical distribution of the bootstrap threshold estimates is used to construct confidence intervals and to conduct threshold tests based on bootstrap likelihood ratio (or Wald-type) statistics. Because the bootstrap conditions on the first-step estimates, inference should be interpreted as conditional on the estimated selection correction.

Despite these methodological features, several limitations should be acknowledged. First, the identification of Lewbel-generated instruments depends on the presence of heteroskedasticity in the first-stage residuals. If heteroskedasticity is weak, the resulting instruments may have limited explanatory power and may weaken the precision of the instrumental-variable estimates. Second, the Heckman [21] selection correction relies primarily on the nonlinear functional form of the inverse Mills ratio rather than on strong exclusion restrictions. Although the selection equation includes demographic, institutional, and health-related variables affecting labor market participation, these variables may plau-

sibly influence earnings directly and therefore are not treated as valid exclusion restrictions in the wage equation. Consequently, the selection correction is intended to mitigate bias arising from non-random employment selection rather than to provide strong causal identification. Third, the threshold regression framework assumes a single structural break in the relationship between educational mismatch and income, whereas in practice wage dynamics may evolve more gradually across the life cycle. Finally, because the analysis is based on cross-sectional data, the model cannot fully account for unobserved individual heterogeneity or dynamic career trajectories that may simultaneously influence educational attainment and earnings outcomes. Accordingly, the estimated coefficients should be interpreted primarily as conditional associations rather than strictly causal effects.

4. Empirical Results

4.1. Descriptive Statistics

Table 2 reports summary statistics for the 2926 individuals included in the analytical sample drawn from the 2020 wave of the CFPS. The sample consists of working-age individuals with complete information on education, employment status, income, and self-reported job-specific educational requirements. The average log of real monthly income (lninc) is 8.214 (SD = 0.605), indicating moderate dispersion in earnings across workers in 2020. Educational mismatch is widespread in the sample: 57.5% of individuals are classified as overeducated, 11.2% as undereducated, and 31.2% as well matched. In terms of human capital characteristics, the average years of schooling is 11.451 (SD = 3.467), corresponding approximately to upper secondary education. Average labor market experience is 13.220 years (SD = 10.432), providing substantial variation for identifying nonlinear effects in the threshold regression framework. The mean age of respondents is 32.553 years (SD = 10.803), indicating a relatively young working-age sample.

Table 2. Descriptive statistics.

Variable	Mean	Std. Dev.
lninc	8.214	0.605
age	32.553	10.803
exper	13.220	10.432
employ	0.866	0.341
gender (male = 1)	0.572	0.495
education (years)	11.451	3.467
overeducation	0.575	0.494
undereducation	0.112	0.316
well_matched	0.312	0.464
hukou (urban = 1)	0.570	0.495
marriage	0.612	0.487
party	0.023	0.151
health_2	0.214	0.410
health_3	0.247	0.431
health_4	0.017	0.131
health_5	0.031	0.173

Regarding demographic characteristics, males account for 57.2% of the sample. Urban hukou holders represent 57.0% of individuals, 61.2% are married, and 2.3% report Communist Party membership. The employment rate in the broader working-age sample is 86.6%, reflecting the inclusion of both employed and non-employed individuals in the selection equation of the Heckman framework.

The correlation matrix among the key explanatory variables is reported in Table A1. The pairwise correlation coefficients are generally moderate in magnitude, and no correla-

tion exceeds conventional thresholds commonly associated with severe multicollinearity. These results suggest that multicollinearity is unlikely to pose a serious concern in the baseline specification.

4.2. Results

As this study employs a two-step selection model integrated into the threshold Duncan and Hoffman regression framework, the empirical analysis consists of two equations: a selection equation estimated using a Probit model and an outcome equation estimated using a threshold regression with instrumental variables. Accordingly, the results are presented in two parts.

The first part reports the Probit estimates of employment participation in Table 3, with the corresponding marginal effects presented in Table 4. The results indicate a statistically significant and positive association between years of education and the likelihood of employment participation. This effect is particularly pronounced among females, with a marginal effect of 0.034, suggesting that education plays a crucial role in enhancing employment opportunities for women. This finding reflects ongoing transformations in China's labor market, where educational credentials have become increasingly important for labor market entry and attachment, especially among female workers [1,49].

Table 3. Probit model: Selection of employment participation.

	Whole	**Male**	**Female**
Constant	−1.616 ***	−0.383	−1.758 ***
	(0.298)	(0.431)	(0.416)
A	0.026 ***	0.040 ***	0.034 **
	(0.010)	(0.016)	(0.013)
age	0.110 **	0.092 ***	0.123 ***
	(0.016)	(0.022)	(0.024)
age^2	−0.001 ***	−0.001 ***	−0.001 ***
	(0.000)	(0.000)	(0.000)
gender	0.715 ***		
	(0.063)		
health_2	0.057 **	0.038 ***	−0.035 ***
	(0.030)	(0.012)	(0.010)
health_3	0.007 *	−0.003	0.051 ***
	(0.004)	(0.119)	(0.018)
health_4	0.039	−0.008 **	0.303
	(0.248)	(0.004)	(0.416)
health_5	−0.201	−0.236	−0.191
	(0.172)	(0.233)	(0.260)
hukou	0.009 ***	0.065 ***	−0.014 ***
	(0.003)	(0.005)	(0.007)
marriage	−0.055 *	0.539 ***	−0.435 ***
	(0.033)	(0.134)	(0.114)
party	0.255 **	0.101 *	0.303 *
	(0.131)	(0.060)	(0.174)
LR-test (*p*-value)	0.000	0.000	0.000

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4. Marginal effects.

	Whole	Male	Female
A	0.006 ***	0.003 *	0.008 **
	(0.002)	(0.002)	(0.004)
age	0.028 **	0.024 ***	0.030 ***
	(0.016)	(0.013)	(0.012)
gender	0.245 ***		
	(0.012)		
health_2	0.019 *	0.015 *	−0.015 **
	(0.011)	(0.009)	(0.008)
health_3	0.020 **	−0.011	0.027 *
	(0.011)	(0.015)	(0.011)
health_4	0.070	−0.113 ***	0.161
	(0.040)	(0.033)	(0.111)
health_5	−0.008	−0.030	−0.091
	(0.037)	(0.029)	(0.087)
hukou	0.010 *	0.006 ***	−0.019 ***
	(0.006)	(0.002)	(0.005)
marriage	−0.045 ***	0.070 ***	−0.010 ***
	(0.013)	(0.012)	(0.004)
party	0.054	0.014	0.093
	(0.047)	(0.039)	(0.094)

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Age exhibits a strong and positive effect on employment participation, while the coefficient on the squared age term is negative and statistically significant, indicating diminishing returns to age. This pattern suggests that employment likelihood increases with age and accumulated experience but weakens at later stages of the life cycle, consistent with prior evidence that employability eventually declines as individuals approach retirement age or face age-related constraints [50]. Gender has a statistically significant effect on employment participation, with males being more likely to be employed than females. This result is consistent with the existing literature documenting persistent gender disparities in China's labor market, which are often attributed to occupational segregation, differential labor market opportunities, and discrimination [51]. Health status is also an important determinant of employment participation. Individuals reporting better health are more likely to be employed, although the estimated effects vary across health categories. This finding aligns with previous studies showing that health conditions influence labor supply decisions and employment prospects by affecting productivity and work capacity [13].

Finally, marital status is negatively associated with employment participation, particularly among women. This result may reflect changing family roles, rising work–family conflicts, and evolving social norms that shape labor force attachment decisions. It is consistent with evidence suggesting that the employment advantage traditionally associated with marriage has weakened over time, especially for female workers [51].

In the second part of the results, we present the wage equation estimates obtained from the threshold regression model. Before interpreting the estimated coefficients, it is important to justify the use of both the threshold specification and the IV approach, which are employed to address potential nonlinearities in wage determination and endogeneity concerns related to education and educational mismatch. To this end, we conduct three diagnostic tests: the Hansen [48] threshold effect test, the Cragg–Donald Wald (CDW) F-statistic test for weak instruments, Hansen J test of overidentifying restrictions and the Durbin–Wu–Hausman (DWH) test for endogeneity, and Breusch–Pagan (BP) tests for heteroskedasticity. The results of these diagnostic tests are reported in the bottom panel of Table 5. The Hansen threshold test strongly rejects the null hypothesis of a linear

wage structure, confirming the presence of significant nonlinearities and supporting the use of a threshold regression framework for the full sample as well as for the male and female subsamples.

Table 5. Empirical results of threshold regression.

Threshold Model with IV						
	Whole		**Male**		**Female**	
	exper > γ	**exper ≤ γ**	**exper > γ**	**exper ≤ γ**	**exper > γ**	**exper ≤ γ**
Constant	8.295 ***	8.157 ***	8.410 ***	8.398 ***	8.103 ***	7.850 ***
	(0.013)	(0.015)	(0.021)	(0.022)	(0.020)	(0.024)
E_i^o	0.026 ***	−0.030 *	0.032 ***	−0.045 *	0.014 *	−0.044
	(0.013)	(0.017)	(0.004)	(0.025)	(0.019)	(0.030)
E_i^u	−0.066 **	0.019 *	−0.050 *	0.027 **	−0.075 ***	0.009
	(0.033)	(0.011)	(0.034)	(0.018)	(0.026)	(0.050)
A	0.041 ***	0.049 ***	0.041 ***	0.049 ***	0.050 ***	0.041 **
	(0.015)	(0.011)	(0.014)	(0.011)	(0.012)	(0.025)
exper	−0.003 ***	0.023 ***	−0.005 ***	0.033 ***	−0.016 ***	0.040 ***
	(0.001)	(0.003)	(0.002)	(0.004)	(0.002)	(0.005)
IMR	−0.662 ***	−0.177 ***	−0.272 ***	−0.660 ***	0.126	−0.386 ***
	(0.130)	(0.087)	(0.070)	(0.217)	(0.191)	(0.129)
Diagnostic						
Threshold (year)	13 (11–14)		14 (12–17)		9 (7–11)	
Threshold (F-test)	129.934 ***		23.020 ***		43.004 ***	
CDW test	15.034 ***		12.873 ***		20.302 **	
DWH test	3.201 **		6.108 ***		5.378 ***	
Hansen J *p*-value	0.527		0.344		0.602	
BP test	6.149 ***		5.401 ***		8.332 ***	

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

To address potential endogeneity of years of education, undereducation, and overeducation in the income equation, we employ six instruments generated using the heteroskedasticity-based identification strategy proposed by Lewbel [23]. The strength of the instruments is evaluated using the CDW F-statistic, following the critical values suggested by Stock and Yogo [52]. For the full sample, the CDW F-statistic equals 15.034, while the corresponding statistics are 12.873 for males and 20.302 for females. In all cases, the statistics exceed the conventional rule-of-thumb threshold of 10, indicating that the generated instruments are sufficiently strong and mitigating concerns regarding weak instrument bias. In addition, the DWH test rejects the null hypothesis of exogeneity. The DWH statistics are 3.201 for the full sample, 6.108 for males, and 5.378 for females, all statistically significant at conventional levels. These results provide empirical evidence that years of education, overeducation, and undereducation are endogenous in the wage equation, thereby justifying the use of IV estimation. Furthermore, the Hansen J test of overidentifying restrictions yields *p*-values of 0.527 (full sample), 0.344 (males), and 0.602 (females). In all cases, the null hypothesis of joint instrument validity cannot be rejected, supporting the exogeneity of the Lewbel-generated instruments under the maintained assumptions. Finally, the BP test in the first-stage residuals strongly reject the null of homoskedasticity. This confirms the presence of heteroskedasticity required for identification under the Lewbel [23] framework.

Table 5 reports the IV estimation results from the threshold regression model. Th The estimated thresholds of 13 years (full sample), 14 years (men), and 9 years (women) can be interpreted as career-stage transition points. In the Chinese labor market, early career years are often characterized by job search experimentation, higher mobility, and relatively limited bargaining power [53]. During this phase, educational mismatch may be more binding, as workers have not yet accumulated firm-specific or occupation-specific human capital [54]. Beyond the threshold, accumulated experience enhances adaptability, task-specific skills, and internal promotion opportunities, allowing workers to partially offset mismatch penalties through productivity gains and career advancement [32]. The higher threshold for men (14 years) suggests that male workers may require a longer accumulation of experience before mismatch effects change, possibly reflecting differences in occupational sorting and hierarchical promotion structures [55]. The lower threshold for women (9 years) may reflect institutional and life-cycle factors. In China, women's career trajectories are more likely to be influenced by family formation, maternity-related interruptions, and differential labor-force attachment patterns [51]. Empirical evidence suggests that gender wage gaps and promotion disparities in China are partly associated with career interruptions and structural labor market segmentation [56]. As a result, the wage effects of educational mismatch may shift earlier in the career cycle for women, reflecting earlier stabilization into long-term occupational tracks and constrained mobility.

Importantly, these thresholds should be interpreted as reduced-form transition points rather than strict causal breakpoints; they represent data-driven regime shifts in wage determination consistent with heterogeneous career dynamics [48] (see, Figure 1).

Considering overeducation, the wage effects of surplus schooling differ markedly across experience regimes. In the full sample, overeducation is associated with a statistically significant wage penalty in the low-experience regime (exper $\leq$ 13 years): each additional year of surplus schooling reduces income by approximately 3.0%. In contrast, in the high-experience regime (exper > 13 years), the coefficient on overeducation becomes positive and statistically significant (approximately +2.6%), indicating that surplus education is modestly rewarded once sufficient work experience has been accumulated. This regime shift suggests that accumulated experience can partially offset the initial productivity mismatch associated with overeducation. As workers gain job-specific skills, improve performance, and enhance internal mobility, surplus schooling becomes more effectively utilized. These findings are consistent with the literature emphasizing the compensatory role of experience in mitigating mismatch penalties [12,36,37,57].

Gender-disaggregated results reveal substantial heterogeneity. Among male workers, overeducation generates a statistically significant wage penalty in the low-experience regime (-4.5%), but becomes positively rewarded in the high-experience regime (+3.2%). This clear reversal implies that surplus education becomes economically valuable for men only after sufficient experience has been accumulated, possibly reflecting improved job matching or stronger productivity signaling at later career stages [50,58]. Among female workers, overeducation is negatively associated with wages in the low-experience regime (-4.4%), although the estimate is statistically insignificant, while the coefficient in the high-experience regime is positive but modest (+1.4%). Unlike men, women do not experience a pronounced transition from penalty to premium across experience regimes. Although the estimates do not indicate a strong gender-based dividend in returns to surplus schooling, they are consistent with broader evidence documenting narrowing gender gaps in educational attainment and labor market participation [49]. These structural changes may gradually reduce disparities in skill returns, even if full convergence is not yet observed in the 2020 data.

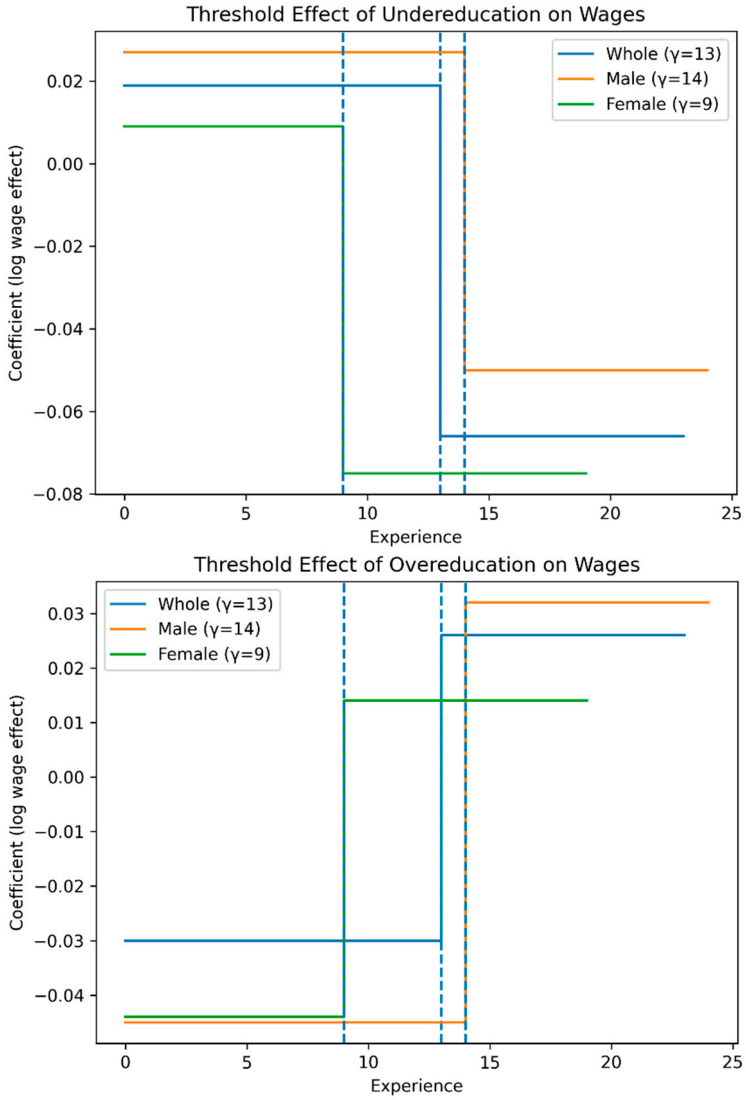


Figure 1. Threshold effects of educational mismatch on wages. The dashed blue vertical lines indicate the estimated threshold levels of experience (γ) at which the impact of undereducation on wages changes regime. Specifically, the thresholds are $\gamma = 13$ years for the full sample, $\gamma = 14$ years for males, and $\gamma = 9$ years for females.

Turning to undereducation, a clear regime-dependent reversal is also evident. In the low-experience regime (exper $\leq$ 13 years), undereducation is associated with a modest but statistically significant wage premium (+1.9%) in the full sample, suggesting that at early career stages, practical learning and on-the-job training can partially substitute for formal schooling. However, in the high-experience regime (exper > 13 years), undereducation becomes significantly penalized (-6.6%), indicating that the absence of formal educational credentials becomes increasingly binding at later career stages, potentially restricting promotion opportunities or access to higher-level occupations. This pattern aligns with the compensatory hypothesis of mismatch and with human capital theory, which emphasizes substitutability between experience and education early in the career but a growing importance of formal credentials over time [59,60].

Gender-specific estimates further highlight differences in how undereducation is rewarded. Among male workers, undereducation yields a statistically significant wage premium in the low-experience regime (+2.7%) but shifts to a penalty in the high-experience regime (-5.0%). Among female workers, the premium in the low-experience regime is small and statistically insignificant (+0.9%), whereas the penalty in the high-experience

regime is larger (−7.5%) and strongly significant. These results suggest that men are better able to leverage accumulated experience to compensate for insufficient schooling, whereas women face more persistent wage disadvantages at later career stages. This finding is consistent with evidence of structural gender inequalities in the Chinese labor market, including differential promotion opportunities, employer biases, and constraints related to family responsibilities [61–64].

4.3. Robustness Check

To assess the robustness of our findings, we re-estimate the threshold regression model using age as the threshold variable instead of work experience. This alternative specification allows us to examine whether the observed nonlinear wage effects of educational mismatch hold when the threshold reflects broader life-cycle stages. The results, presented in Table 6, are broadly consistent with our main findings.

Table 6. Empirical results of the threshold regression model with age as the threshold variable.

	Whole		**Male**		**Female**	
	age > γ	**age ≤ γ**	**age > γ**	**age ≤ γ**	**age > γ**	**age ≤ γ**
Constant	8.372 *** (0.164)	7.473 *** (0.079)	8.968 *** (0.182)	7.501 *** (0.101)	7.574 *** (0.192)	7.264 *** (0.122)
E_i^o	0.031 *** (0.008)	−0.005 (0.004)	0.027 *** (0.009)	−0.009 * (0.005)	0.040 *** (0.011)	−0.006 (0.007)
E_i^u	−0.001 (0.019)	0.006 (0.012)	−0.005 (0.021)	0.009 (0.015)	−0.024 (0.027)	0.003 (0.017)
A	0.057 *** (0.007)	0.029 *** (0.004)	0.039 *** (0.008)	0.024 *** (0.005)	0.069 *** (0.008)	0.030 *** (0.005)
exper	−0.016 *** (0.003)	0.017 *** (0.002)	−0.022 *** (0.003)	−0.024 *** (0.003)	−0.008 *** (0.004)	0.018 *** (0.004)
IMR	−0.648 *** (0.123)	−0.092 *** (0.301)	−0.383 ** (0.102)	−0.309 *** (0.134)	−0.224 *** (0.049)	−0.102 *** (0.034)
Threshold (year)	32.087		36.841		31.116	
Threshold (F-test)	103.945 ***		34.034 ***		9.309 ***	

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Moreover, to assess whether the main findings are sensitive to the Heckman selection correction, we re-estimate the threshold regression model on the employed subsample without including the IMR. The results are reported in Table 7.

The estimated experience thresholds remain identical to the baseline specification (13 years for the full sample, 14 for males, and 9 for females), and the threshold F-tests remain highly significant. Moreover, the qualitative pattern of educational mismatch effects across regimes is preserved. Overeducation exhibits positive wage effects above the threshold and negative or insignificant effects below, while experience maintains opposite-signed effects across regimes.

These findings indicate that the core nonlinear structure of the wage equation is not driven by the functional-form identification of the Heckman correction. The selection adjustment therefore refines the estimates but does not alter the substantive conclusions regarding regime-dependent mismatch effects.

Table 7. Empirical results of threshold regression without IMR.

	Whole		**Male**		**Female**	
	exper > γ	**exper ≤ γ**	**exper > γ**	**exper ≤ γ**	**exper > γ**	**exper ≤ γ**
Constant	8.297 ***	8.157 ***	8.404 ***	8.327 ***	8.100 ***	7.850 ***
	(0.013)	(0.015)	(0.019)	(0.019)	(0.020)	(0.024)
E_i^o	0.017 ***	−0.009 *	0.015 **	−0.011 *	0.019 *	−0.015 *
	(0.006)	(0.005)	(0.007)	(0.006)	(0.009)	(0.008)
E_i^u	−0.017	0.006	−0.017	0.015	−0.022	0.006
	(0.013)	(0.015)	(0.016)	(0.020)	(0.021)	(0.022)
A	0.025 ***	0.021 ***	0.018 ***	0.009	0.003	0.019 ***
	(0.003)	(0.005)	(0.004)	(0.006)	(0.045)	(0.007)
exper	−0.002 ***	0.023 ***	−0.003 **	0.031 ***	−0.014 ***	0.039 ***
	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)	(0.005)
Threshold (year)	13 [11–14]		14 [12–17]		9 [7–11]	
Threshold (F-test)	129.934 ***		23.020 ***		43.004 ***	

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5. Conclusions

This study provides new empirical evidence on the wage effects of educational mismatch in China by applying a threshold regression framework that jointly addresses endogeneity and sample selection. Using data from the 2020 wave of the CFPS, the analysis examines how work experience moderates the wage effects of educational mismatch across career stages and gender groups. By integrating a threshold instrumental-variable estimator into an augmented Duncan–Hoffman model, the study moves beyond conventional linear specifications and uncovers nonlinear and heterogeneous wage differentials associated with education and educational mismatch.

Several key findings emerge. First, the wage effects of educational mismatch depend strongly on workers' experience. Overeducated workers face wage penalties at early career stages, but these penalties diminish and, in some cases, become statistically insignificant once workers surpass the estimated experience threshold. This pattern suggests that accumulated work experience can partially compensate for surplus schooling, enabling workers to better utilize excess education through job-specific skills, improved performance, or more effective signaling. By contrast, undereducation is associated with a modest wage premium at early career stages, indicating that practical learning and on-the-job experience can substitute for formal schooling initially. However, this premium reverses into a wage penalty at higher experience levels, implying that the absence of formal educational credentials becomes increasingly binding as workers advance in their careers.

Second, pronounced gender differences are evident. Male workers appear better able to leverage accumulated experience to offset educational shortfalls. Undereducated men earn a meaningful premium below the experience threshold and face only a moderate penalty above it. For women, the premium associated with undereducation at early career stages is small and statistically insignificant, while the penalty at higher experience levels is larger and more robust. A similar asymmetry is observed for overeducation: experienced men face a substantially reduced mismatch penalty, whereas experienced women do not experience a comparable improvement. Although these findings do not imply convergence in mismatch-related wage outcomes, they are consistent with broader evidence on gender differences in labor-market opportunities, promotion prospects, and the returns to human capital investment.

Overall, the results highlight the critical role of work experience as a moderating factor in the wage effects of educational mismatch. Experience can compensate for both surplus and deficit schooling, but its effectiveness depends on the type of mismatch and gender. These nonlinear patterns underscore the importance of considering career-stage heterogeneity when assessing the economic consequences of educational mismatch.

Several policy implications follow from these findings. First, the presence of wage penalties for undereducated workers at later career stages suggests the need for targeted lifelong learning and skill-upgrading initiatives. Programs aimed at mid-career and senior workers without formal qualifications could help maintain productivity and reduce wage disadvantages associated with undereducation. Second, the reduced penalties for experienced overeducated workers point to the importance of expanding high-skilled employment opportunities. Policies that promote the development of knowledge-intensive industries, support innovation-driven firms, and facilitate job mobility across sectors can enhance the absorption of surplus education and mitigate mismatch. Third, the persistence of educational mismatch highlights the need to strengthen alignment between educational curricula and labor-market demand. Closer collaboration among educational institutions, employers, and government agencies can improve curriculum design, expand vocational and applied training pathways, and better equip students with skills valued in a rapidly evolving economy.

A potential limitation of this study concerns the measurement of educational mismatch using self-reported job requirements. Although the self-assessment approach is widely adopted in the literature and allows for job-specific heterogeneity, it may be subject to perception bias and measurement error. Workers' evaluations of the education required for their positions may reflect personal expectations, wage comparisons, or career satisfaction rather than purely objective job standards. Such reporting bias may generate attenuation effects or introduce endogeneity into the estimated mismatch coefficients. While our instrumental-variable strategy partially addresses concerns related to simultaneity and omitted-variable bias, measurement error arising from subjective reporting cannot be fully eliminated. In addition, although the threshold specification captures nonlinear heterogeneity in mismatch returns across experience regimes, it imposes a discrete regime shift at the estimated cutoff. If the true relationship between educational mismatch and income evolves smoothly over the career cycle, the threshold model should be interpreted as a parsimonious piecewise approximation rather than a literal structural break. Future research could extend the analysis using smooth-transition or spline-based approaches to explore more flexible nonlinear dynamics. Finally, because the threshold parameter is estimated rather than known a priori, statistical inference relies on nonstandard procedures [48], particularly when combined with instrumental-variable estimation and selection correction. Further methodological work may help refine inference strategies in such multi-stage nonlinear settings.

Author Contributions: Conceptualization, W.Y. and P.M.; methodology, W.Y.; software, P.M.; validation, P.M., W.Y. and L.J.; formal analysis, L.J.; investigation, L.J.; data curation, L.J.; writing—original draft preparation, L.J.; writing—review and editing, P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, and was conducted as part of the PhD program in Economics at the Faculty of Economics, Chiang Mai University, under the CMU Presidential Scholarship.

Informed Consent Statement: We declare that this manuscript is original, has not been published before, and is not currently being considered for publication elsewhere. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied

the criteria for authorship but are not listed. We further confirm that all have approved the order of authors listed in the manuscript.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Correlation matrix.

	Age	**exper**	**Employ**	**Gender**	***A***	***E^o^***	***E^u^***	**Hukou**	**Marriage**	**Party**
age	1.000	0.895	0.092	0.092	−0.302	0.095	−0.025	0.060	0.517	0.009
exper	0.895	1.000	0.060	0.087	−0.563	0.077	0.023	−0.056	0.517	−0.021
employ	0.092	0.060	1.000	0.216	0.030	0.001	−0.000	0.025	0.037	0.027
gender	0.092	0.087	0.216	1.000	−0.059	0.039	0.003	−0.052	−0.096	0.023
A	−0.302	−0.563	0.030	−0.059	1.000	0.058	−0.172	0.325	−0.201	0.105
E^o	0.095	0.077	0.001	0.039	0.058	1.000	−0.414	0.014	0.120	−0.001
E^u	−0.025	0.023	−0.000	0.003	−0.172	−0.414	1.000	−0.068	−0.064	−0.004
hukou	0.060	−0.056	0.025	−0.052	0.325	0.014	−0.068	1.000	0.022	0.001
marriage	0.517	0.517	0.037	−0.096	−0.201	0.120	−0.064	0.022	1.000	−0.012
party	0.009	−0.021	0.027	0.023	0.105	−0.001	−0.004	0.001	−0.012	1.000

Assumptions for Identification and Estimation

To ensure identification and consistent estimation of the model parameters, the following assumptions are imposed.

Assumption A1. *Exogeneity of control variables.*

$$E[X_i \varepsilon_i] = 0 \tag{A1}$$

where X_i denotes the vector of exogenous control variables included in the wage equation and ε_i represents the structural error term.

This assumption states that the exogenous covariates are uncorrelated with the unobserved determinants of income. Under this condition, the control variables do not introduce additional endogeneity into the structural equation, allowing them to serve as valid conditioning variables in both the first-stage and second-stage regressions.

Assumption A2. *Heteroskedasticity-based identification for Lewbel instruments.*

$$E[X_i u_i] = 0 \tag{A2}$$

$$Cov\left(X_i, u_i^2\right) \neq 0 \tag{A3}$$

where u_i denotes the residual from the first-stage regression of the endogenous variable on the exogenous controls. The first condition requires that the exogenous variables are orthogonal to the first-stage disturbance term. The second condition requires the presence of heteroskedasticity in the first-stage residuals that is systematically related to the exogenous covariates. Under these conditions, the heteroskedasticity-based instruments constructed following Lewbel [23] provide valid internal instruments for the endogenous regressors.

Assumption A3. *Threshold identification.*

$$P(exper_i = \gamma) = 0 \tag{A4}$$

where $exper_i$ denotes labor market experience and γ is the unknown threshold parameter. This condition ensures that the threshold value does not coincide with a mass point of the experience distribution. In other words, the probability that an observation lies exactly at the threshold is zero. This assumption guarantees that the threshold parameter is identifiable and that the objective function used in threshold estimation is well defined.

Assumption A4. *Joint normality in the selection model.*

$$(u_i, \varepsilon_i) \sim N(0, \Sigma) \tag{A5}$$

where u_i denotes the error term in the employment selection equation and ε_i denotes the structural error in the wage equation. This assumption implies that the disturbances of the selection equation and the wage equation follow a joint normal distribution with covariance matrix Σ. Under this assumption, the conditional expectation of the wage equation error term can be expressed as a function of the inverse Mills ratio, which forms the basis for the Heckman [21] two-step correction for sample selection bias.

References

1. Liu, S.; Ma, W.; Deng, N. The role of noncognitive skills in educational mismatch: Evidence from China. *Int. J. Manpow.* **2024**, *45*, 1365–1384. [CrossRef]
2. Kalfa, E.; Piracha, M. Immigrants' educational mismatch and the penalty of over-education. *Educ. Econ.* **2017**, *25*, 462–481. [CrossRef]
3. Veselinović, L.; Mangafić, J.; Turulja, L. The effect of education-job mismatch on net income: Evidence from a developing country. *Econ. Res.* **2020**, *33*, 2648–2669. [CrossRef]
4. Tang, R.; Wang, G. Educational mismatch and income inequality. *Econ. Model.* **2021**, *100*, 105525. [CrossRef]
5. Didier, N. Educational mismatch, labor market completeness, and gender: Evidence from Chile. *Int. J. Educ. Dev.* **2024**, *105*, 102990. [CrossRef]
6. Jiang, L.; Yamaka, W. Changing Wage Effects of Educational Mismatch in China: Evidence from Threshold IV–Selection Models. 2025. [CrossRef]
7. Muysken, J.; Kiiver, H.; Hoppe, M. The Impact of Education and Mismatch on Wages: Germany, 1984–2000. In *Overeducation in Europe: Current Issues in Theory and Policy*; Edward Elgar Publishing: Cheltenham, UK, 2003; pp. 109–132.
8. Zakariya, Z. Returns to education: What does over-education play. *J. Contemp. Issues Thought* **2013**, *3*, 138–155.
9. Zamfir, A.-M.; Matei, M.M.; Lungu, E.O. Influence of education-job mismatch on wages among higher education graduates. *Procedia Soc. Behav. Sci.* **2013**, *89*, 293–297. [CrossRef]
10. Wu, N.; Wang, Q. Wage penalty of overeducation: New micro-evidence from China. *China Econ. Rev.* **2018**, *50*, 206–217. [CrossRef]
11. Nan, J. Return to Education and Education Mismatch in China. Ph.D. Thesis, Cardiff University, Cardiff, UK, 2023.
12. Xiangrong, W. The impact of overeducation on earnings in China. *Front. Educ. China* **2008**, *3*, 123–136. [CrossRef]
13. Xu, Q.; Chen, Y. Educational mismatch and mental health: Evidence from China. *Soc. Sci. Med.* **2024**, *356*, 117140. [CrossRef] [PubMed]
14. Ajuzieogu, U. Human Capital Investment Under AI Transition: Educational Returns and Optimal Policy. *Econ. Educ. Rev.* **2025**. [CrossRef]
15. Poghosyan, S.; Poghosyan, M.; Nalbandyan, H.; Beglaryan, L.; Harutyunyan, A.; Adonts, N.; Azatyan, L.; Asatryan, H. Reskilling human capital in the AI era: A human-centric conceptual 'spider web' model of knowledge and technology integration. *Cogent Arts Humanit.* **2026**, *13*, 2617070. [CrossRef]
16. Duncan, G.J.; Hoffman, S.D. The incidence and wage effects of overeducation. *Econ. Educ. Rev.* **1981**, *1*, 75–86. [CrossRef]
17. Zeng, J.; Zhang, J. Education policies and development with threshold human capital externalities. *Econ. Model.* **2022**, *108*, 105744. [CrossRef]
18. Wantchekon, L.; Klašnja, M.; Novta, N. Education and human capital externalities: Evidence from colonial Benin. *Q. J. Econ.* **2015**, *130*, 703–757. [CrossRef]
19. Arrazola, M.; de Hevia, J. Three measures of returns to education: An illustration for the case of Spain. *Econ. Educ. Rev.* **2008**, *27*, 266–275. [CrossRef]
20. Heckman, J.J.; Lochner, L.J.; Todd, P.E. Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. *Handb. Econ. Educ.* **2006**, *1*, 307–458.
21. Heckman, J.J. Sample selection bias as a specification error. *Econometrica* **1979**, *47*, 153–161. [CrossRef]

22. Vella, F. Estimating models with sample selection bias: A survey. *J. Hum. Resour.* **1998**, *33*, 127–169. [CrossRef]
23. Lewbel, A. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *J. Bus. Econ. Stat.* **2012**, *30*, 67–80. [CrossRef]
24. Zheng, Y.; Zhang, X.; Zhu, Y. Overeducation, major mismatch, and return to higher education tiers: Evidence from novel data source of a major online recruitment platform in China. *China Econ. Rev.* **2021**, *66*, 101584. [CrossRef]
25. Spence, M. Job market signaling. *Q. J. Econ.* **1973**, *87*, 355–374. [CrossRef]
26. Jiang, S.; Guo, Y. Reasons for college major-job mismatch and labor market outcomes: Evidence from China. *China Econ. Rev.* **2022**, *74*, 101822. [CrossRef]
27. Thurow, L.C. The failure of education as an economic strategy. *Am. Econ. Rev.* **1982**, *72*, 72–76.
28. Rees, A. Review of Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education. *Am. Econ. Rev.* **1965**, *55*, 958–960.
29. Becker, G.S. Human Capital Revisited. In *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, 3rd ed.; University of Chicago Press: Chicago, IL, USA, 1994; pp. 15–28.
30. Sattinger, M. Assignment models of the distribution of earnings. *J. Econ. Lit.* **1993**, *31*, 831–880.
31. Mincer, J. Investment in Human Capital and Personal Income Distribution. *J. Political Econ.* **1958**, *66*, 281–302. [CrossRef]
32. Fanti, L.; Guarascio, D.; Tubiana, M. Skill mismatch and the dynamics of Italian companies' productivity. *Appl. Econ.* **2021**, *53*, 6790–6803. [CrossRef]
33. Iriondo, I.; Pérez-Amaral, T. The effect of educational mismatch on wages in Europe. *J. Policy Model.* **2016**, *38*, 304–323. [CrossRef]
34. Jacobs, V.; Rycx, F.; Volral, M. Wage effects of educational mismatch according to workers' origin: The role of demographics and firm characteristics. *Economist* **2022**, *170*, 459–501. [CrossRef]
35. Romero, L.M.; Huertas, I.P.M.; Jiménez, M.D.M.S. Wage effects of cognitive skills and educational mismatch in Europe. *J. Policy Model.* **2017**, *39*, 909–927. [CrossRef]
36. Hartog, J. Over-education and earnings: Where are we, where should we go? *Econ. Educ. Rev.* **2000**, *19*, 131–147. [CrossRef]
37. McGuinness, S. Overeducation in the labour market. *J. Econ. Surv.* **2006**, *20*, 387–418. [CrossRef]
38. Njifen, I.; Smith, P. Education-Job Mismatch and Heterogeneity in the Return to Schooling: Evidence from Cameroon. *J. Knowl. Econ.* **2024**, *15*, 860–887. [CrossRef]
39. Cutillo, A.; Di Pietro, G. The effects of overeducation on wages in Italy: A bivariate selectivity approach. *Int. J. Manpow.* **2006**, *27*, 143–168. [CrossRef]
40. Green, F.; McIntosh, S. Is there a genuine under-utilization of skills amongst the over-qualified? *Appl. Econ.* **2007**, *39*, 427–439. [CrossRef]
41. Groot, W.; Van Den Brink, H.M. Overeducation in the labor market: A meta-analysis. *Econ. Educ. Rev.* **2000**, *19*, 149–158. [CrossRef]
42. Nordin, M.; Persson, I.; Rooth, D.-O. Education–occupation mismatch: Is there an income penalty? *Econ. Educ. Rev.* **2010**, *29*, 1047–1059. [CrossRef]
43. Büchel, F.; Mertens, A. Overeducation, undereducation, and the theory of career mobility. *Appl. Econ.* **2008**, *36*, 803–816. [CrossRef]
44. Green, F.; Zhu, Y. Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education. *Oxf. Econ. Pap.* **2010**, *62*, 740–763. [CrossRef]
45. Zhu, R. The impact of major–job mismatch on college graduates' early career earnings: Evidence from China. *Educ. Econ.* **2014**, *22*, 511–528. [CrossRef]
46. Shi, T. The Economic Implications of Skill Mismatch in China's Labor Market: A Focus on Higher Education Graduates. *Law Econ.* **2024**, *3*, 30–38. [CrossRef]
47. Leuven, E.; Oosterbeek, H. Overeducation and Mismatch in the Labor Market. In *Handbook of the Economics of Education*; Elsevier: Amsterdam, The Netherlands, 2011; Volume 4, pp. 283–326. [CrossRef]
48. Hansen, B.E. Sample splitting and threshold estimation. *Econometrica* **2000**, *68*, 575–603. [CrossRef]
49. Chen, Y.; Jiang, S.; Zhou, L.A. Estimating returns to education in urban China: Evidence from a natural experiment in schooling reform. *J. Comp. Econ.* **2020**, *48*, 218–233. [CrossRef]
50. Lei, X.; Hu, Y.; McArdle, J.J.; Smith, J.P.; Zhao, Y. Gender differences in cognition among older adults in China. *J. Hum. Resour.* **2012**, *47*, 951–971. [PubMed]
51. Zhang, Y.; Hannum, E.; Wang, M. Gender-based employment and income differences in urban China: Considering the contributions of marriage and parenthood. *Soc. Forces* **2008**, *86*, 1529–1560. [CrossRef]
52. Stock, J.H.; Yogo, M. *Testing for Weak Instruments in Linear IV Regression*; NBER Technical Working Papers 0284; National Bureau of Economic Research, Inc.: Cambridge, MA, USA, 2002.
53. Quisumbing, A.R. From bargaining power to empowerment: Measuring the unmeasurable. *Agric. Econ.* **2025**, *56*, 419–430. [CrossRef]

54. Becker, G. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*; Columbia University Press: New York, NY, USA, 1964.
55. Blau, F.D.; Kahn, L.M. The gender wage gap: Extent, trends, and explanations. *J. Econ. Lit.* **2017**, *55*, 789–865. [CrossRef]
56. Liu, P.-W.; Meng, X.; Zhang, J. Sectoral Gender Wage Differentials and Discrimination in the Transitional Chinese Economy. *J. Popul. Econ.* **2000**, *13*, 331–352. Available online: http://www.jstor.org/stable/20007718 (accessed on 10 January 2020). [CrossRef]
57. Shi, Y.; Yu, H.; Huang, Y.; Shen, X.; Guo, W. Over-education and job satisfaction among new graduates in China: A gender perspective. *Soc. Indic. Res.* **2023**, *170*, 51–74. [CrossRef]
58. Kidd, M.P.; Meng, X. The Chinese state enterprise sector: Labour market reform and the impact on male–female wage structure. *Asian Econ. J.* **2001**, *15*, 405–423. [CrossRef]
59. Daw, J.; Hardie, J.H. Compensating differentials, labor market segmentation, and wage inequality. *Soc. Sci. Res.* **2012**, *41*, 1179–1197. [CrossRef] [PubMed]
60. Churchill, S.A.; Mishra, V. Returns to education in China: A meta-analysis. *Appl. Econ.* **2018**, *50*, 5903–5919. [CrossRef]
61. Iwasaki, I.; Ma, X. Gender wage gap in China: A large meta-analysis. *J. Labour Mark. Res.* **2020**, *54*, 17. [CrossRef]
62. Ma, X. Dual Labor Market: The Wage Gap Between Formal and Informal Workers. In *Labor Market Institutions in China*; Springer Nature: Singapore, 2024; pp. 171–198.
63. Chen, Y.; Zhang, H.; Zhou, L.A. Motherhood and gender wage differentials within a Chinese firm. *Econ. Dev. Cult. Change* **2021**, *70*, 283–320. [CrossRef]
64. Xing, C.; Yuan, X.; Zhang, J. City size, family migration, and gender wage gap: Evidence from rural–urban migrants in China. *Reg. Sci. Urban Econ.* **2022**, *97*, 103834. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.