Economic Modelling 100 (2021) 105525

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Educational mismatch and income inequality☆

Rongsheng Tang[a,b,*], Gaowang Wang[c]

[a] Institute for Advanced Research, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, 200433, China
[b] Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, 777 Guoding Road, Shanghai, 200433, China
[c] Center for Economic Research, Shandong University, Jinan, China

ARTICLE INFO

JEL classification:
I24
J24
J31
O15

Keywords:
Educational mismatch
Income inequality
Wage effect
Search friction
Promotion

ABSTRACT

We build a model to understand educational mismatch and income inequality among highly educated workers. For occupations and workers with college majors, educational mismatch negatively impacts wage but positively correlates with wage inequality. We examine different channels that contribute to wage inequality by identifying three underlying reasons behind the mismatch, namely, preference, promotion, and search friction and quantifying their impacts. Quantitatively, preference and promotion negatively contribute to an inequality increase from 1990 to 2000; match premium and search friction contribute to a 28.4% and 5.3% increase in inequality, respectively. We conclude that educational mismatch significantly affects income inequality and that this impact varies based on the underlying reasons. The study has important policy implications as it shows that policies for improving education match rate and educational signaling and less market friction can reduce wage inequality.

1. Introduction

Educational mismatch occurs when workers' skill type does not match their job requirements, for example, a chemistry major who works as a general manager. However, this mismatch differs from that in which college graduates work in low-skill jobs. Although the first mismatch has received much attention in the literature (e.g., Sloane, 2003; Guironnet and Peypoch, 2007; Lee et al., 2015), research on the type of mismatch is scarce. Accordingly, this study examines the impact of educational mismatch on residual wage inequality among the highly educated. Considering educational mismatch may be helpful because a substantial portion of inequality remains unexplained by observations (Violante, 2002; Tang et al., 2020). We analyze different channels that contribute to wage inequality using survey data to identify three fundamental reasons behind the mismatch and quantify their impacts. This study is meaningful considering the existence of several mismatched workers due to varying reasons. The reasons behind educational mismatch have not been investigated, though its effect on wage inequality has been discussed in the literature to a certain extent (Altonji et al., 2014).

We define educational mismatch based on the subjective responses from the National Survey of College Graduates (NSCG) in the United States. Survey participants were asked to select one of the three responses about the relatedness between their current occupation and field of study in which they have the highest qualification: "closely related," "somewhat related," and "not related." We identify a participant with a skill match when his/her response is "closely related" and as a mismatch otherwise.[1] A mismatch is not a mistake from the participants' perspective as the response reflects their optimal behavior. Nevertheless, the response may partially reflect their knowledge usage efficiency. The survey also gathered data on the main reasons behind the mismatch, indicating that 70% of the mismatches can be attributed to the three reasons: preference, promotion, and search friction.[2] This

☆ This paper has benefited by comments and questions from Liyan Yang, David Wiczer, Ping Wang, two anonymous referees, and the co-editor Angus Chu as well as participants of several conferences. Rongsheng Tang is grateful for the financial support from the National Natural Science Foundation of China (Grant No.71803112). Gaowang Wang is thankful to the Qilu Young Scholar Research Fund at Shandong University for its financial support.

* Corresponding author. Institute for Advanced Research, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai, 200433, China.
E-mail addresses: tang.rongsheng@sufe.edu.cn (R. Tang), gaowang.wang@sdu.edu.cn (G. Wang).

[1] We also define the mismatch only with the response of "not related" as a robustness check.

[2] For example, as documented in section 2, in 1990, 32%, 20%, and 16% of the mismatch was attributed to promotion, preference, and search friction, respectively.

https://doi.org/10.1016/j.econmod.2021.105525
Received 15 March 2020; Received in revised form 18 April 2021; Accepted 18 April 2021
Available online 28 April 2021

study will model the educational mismatch based on these reasons.

First, we document the wage effect of educational mismatch. Consistent with Ritter and West (2014), educational mismatch statistically negatively impacts wage. The average wage ratios of mismatched to matched workers were 0.91, 0.94, and 0.87 in 1990, 2000, and 2010, respectively, indicating approximately 10% mismatch penalty. Demographic variables were controlled for, changing the ratios to 17.1%, 22.9%, and 26.5%, respectively. Therefore, the penalty could be as high as a quarter for workers with similar characteristics. Moreover, mismatched workers have different wage types.

We group mismatched workers by the three reasons and computed the wage ratios for each group. The results show that the ratios in 1990 were 0.98, 0.85, and 0.66 for promotion, preference, and search friction, respectively. Promotion-induced mismatch has negligible penalty, whereas the penalty for search friction could be as high as one third.

Furthermore, educational mismatch and wage inequality among workers with college majors and occupations are positively correlated. Thus, occupations and workers with college majors with high educational mismatch have a high wage inequality. This finding may be counterintuitive at first glance as it may lead to the conclusion that a good match amplifies wage difference. However, inequality depends not only on the variety within the matched or mismatched workers' group but also on their labor component. Thus, inequality and the degree of mismatch have an inverse U-shaped relationship. However, the United States data shows that inequality is on the rising part of the curve. Although this finding does not imply causality, it suggests the potential importance of educational mismatch, particularly after controlling for the demographic characteristics. Section 2 presents a simple accounting exercise showing that educational mismatch contributes to 15% of the inequality. We obtain a sophisticated result by employing our model and conducting a quantitative analysis.

In the model, workers and firms vary by skill and productivity types, respectively. Based on a worker's skill type, jobs are categorized into two types: matched and mismatched. Workers need to possess related skill types that will draw a match premium to obtain a matched job. Here, joint output is the product of firm productivity, workers' skill, the match premium, and the workers' effort. A worker acquires a share of the output in terms of labor compensation, and an increase in this share is referred to as promotion. Job amenity on a matched job is random and affects workers' effort. Therefore, the match degree between the worker and job affects human capital level and workers' occupational choice. In the quantitative analysis, we calibrated the model to target the US economy in 1990 and 2000, during which wage inequality increased rapidly. We also allowed channel- and non-channel-specific parameters to change across years. Subsequently, we quantified each channel's contribution to the increase in wage inequality through counterfactual analysis.

Theoretically, the underlying channels affect wage inequality through occupational choice. A high promotional level in a mismatched job, poor job amenities, and low match premium in a matched job will reduce the likelihood of accepting the matched job. Search friction on a matched job will also affect employment distribution. This condition will affect wage inequality through labor component and human capital quality changes. Quantitatively, we find that preference and promotion contributes negatively to wage inequality increase, while match premium and search friction contribute 28.4% and 5.3%, respectively. We only focus on residual wage inequality, and the model does not include the workers' ability and heterogeneity in firm productivity. Hence, we consider the effect of the mismatch significant.

Our study contributes to the literature in the following ways. First, we explain residual wage inequality by introducing educational mismatch in a structural model. Second, we measure educational mismatch in a novel and direct way by employing survey data. Third, we identify the underlying reasons behind the mismatch to examine different mechanisms contributing to the inequality. Fourth, we find that educational mismatch significantly affects income inequality and that this impact varies based on the underlying reasons.

1.1. Related literature

The related literature primarily includes studies on wage inequality and education and skill mismatch. Studies on wage inequality for the highly educated are increasing. The literature documents a rapid increase in wage inequality since the 1980s (e.g., Autor et al., 2008; Piketty and Saez, 2014). Skill premium has also been studied intensively (e.g., Acemoglu and Autor, 2011; Burstein et al., 2015). Several studies on wage inequality among educational groups also exist (e.g., Violante, 2002; Lee et al., 2015; Tang et al., 2020).

Studies on within-group and residual wage inequality emphasize the impact of unobserved skills (e.g., Lemieux, 2006). Violante (2002) and Kambourov and Manovskii (2009) findings are closely related to our study. Violante (2002) emphasized the role of skill transferability across machines of different vintages in explaining wage differences among ex-ante workers and finds that this channel could explain one third of the residual wage inequality. Kambourov and Manovskii (2009) connected occupational mobility with wage inequality by emphasizing occupation-specific human capital. Although they concluded that occupational mobility would explain most of the residual wage inequality, they attribute occupational mobility to idiosyncratic productivity shocks. The aforementioned study does not identify the fundamental reasons behind the shocks. Thus, the present study complements it by providing certain specific reasons behind occupational choice.

Recent studies on high-skilled workers have focused on the match between skill type and jobs. Altonji et al. (2014) argued that the earnings difference between workers with college majors can be larger than the skill premium between workers with college and high school education levels. They also argued that substantial wage widening or income inequality between workers with college majors is related to their selected occupations' task composition. Ritter and West (2014) found that the changing distribution of college majors causes slight, if any, shift in earnings distribution. We build on the results of these two studies to determine the reasons behind educational mismatch, given the wage difference between occupations of people with the same major. We also determine the extent to which the mismatch accounts for the increase in wage inequality among the highly educated.

Educational mismatch studies investigate its effect on wage between workers with college majors and occupations. Robst (2007) was one of the first to employ the NSCG data for measuring educational mismatch. Lemieux (2014) showed that return to education significantly varies depending on occupation, the field of study, and their match. The college major match-related channel accounts for approximately half of the conventionally measured return to education. Other studies have also shown empirical results between field studies and earnings difference (e.g., Arcidiacono, 2004; Freeman and Hirsch, 2008; Nordin et al., 2010; Kirkeboen et al., 2016).

The present study also relates to skill mismatch literature (e.g., Guvenen et al., 2020; Lise and Postel-Vinay, 2020). These studies measure skill level using a test score, such as the ASVAB and occupational skill requirements. Skill mismatch is measured as the distance between a worker's skill acquirement and a job's skill requirement. Cooper and Liu (2019) recently measured the mismatch between skills and educational attainment. Skill mismatch studies in countries other than the United States, for example, Desjardins and Rubenson (2011), measured the mismatch using European data. In another study, Gil et al. (2020) examined the skill mismatch due to immigration.

The model setup is similar to Berliant et al. (2006) for illustrating knowledge exchange and its consequences on agglomerative activity in a general-equilibrium search-theoretic framework. The idea of occupational choice in this paper is built on Rosen (1986).

The remainder of this paper is organized as follows. Section 2 describes statistical and empirical facts. Section 3 presents a model on educational mismatch incorporating its underlying reasons. Section 4

describes the stationary equilibrium, and Section 5 presents the quantitative analysis. Section 6 presents the quantitative results under alternative calibration strategies. Lastly, Section 7 concludes the paper.

2. Facts

This section documents certain main features of educational mismatch. Data were collected from the NSCG, a census survey of individuals with college degrees. The NSCG is conducted every 10 years to gather information about the relatedness between each sampled individual's field of study and occupation. Survey participants selected from one of the following responses to the questions on relatedness: "closely related," "somewhat related," and "not related." If the response was "closely related," the participant was skill-matched and mismatched otherwise.[3]

The sample includes NSCG (1993, 2003, and 2013) data based on the 1990, 2000, and 2010 censuses, respectively. For data trimming, only full-time workers aged between 16 and 65 years were included. The top annual earning, that is, $4 million and earnings less than $2,800 were excluded. Regarding race, only white, black, and Hispanic races were considered. Following Altonji et al. (2014), the major code was regrouped under the 50 major categories listed in the Department of Education. Subsequently, occupation was regrouped based on Dorn (2009) that used a consistent three-digit occupation code. The data considered four school year levels: 16, 18, 19, and 21 years. These levels were regrouped into three categories: bachelors (16), masters (18, 19), and doctorate (21). In addition, tenure was calculated as a potential experience, that is, $max(age - schooling - 6, 0)$. Following Kambourov and Manovskii (2009), we obtained the residual wage based on the following regression:

$$ln(wage) = constant + \beta_1 edu + \beta_2 exp + i.gender + i.race + \epsilon \quad (1)$$

where $ln(wage)$ is the log value of annual earnings, edu is the education level, and exp is the potential labor market experience. Gender and race were also controlled. The residual wage is the exponential of the residual, $exp(\epsilon)$, and wage inequality is the residual wage inequality measured as the residual variance, $var(\epsilon)$. The analyses in this study focus on this residual wage inequality.

2.1. Statistical description

Tables A.1–A.3 present several data characteristics, including observations, average tenure, earnings, inequality, and employment share for different demographic groups from 1990 to 2010. In the sample, the wage inequality increases from 0.23 in 1990 to 0.39 in 2010. Men have higher earnings and wage inequality than women. Annual earnings increase with education level. However, master's-level workers have the lowest wage inequality. Moreover, white people earn higher wages and have higher inequality than other racial groups. The proportion of job relatedness does not substantially change, that is, the proportion of closely related jobs is approximately 0.56. However, even if the match premium increases, educational mismatch would continue to contribute to income inequality. In addition, if the main reason behind this mismatch changes, this change would also influence income inequality.

Table A.4 lists several inequalities, where Var_{raw} depicts wage inequality with raw data, that is, the variance of log of wages $var(ln(wage))$, Var_{res} is the residual wage inequality, var_1 is the residual wage inequality after further controlling the college major dummy, and var_2 is the residual wage inequality after further controlling the variables of major, occupation, and match status. Thus, the contribution of the college major-occupation match is $\frac{var_1 - var_2}{var_1}$. Subsequently, educational mismatch statistically contributes 15% to the inequality. Accordingly, we will build a model and conduct a quantitative analysis in the later sections.

Table A.5 lists the proportion of match degree in different demographic groups. We defined job relatedness as the percentage of individuals whose response is "closely related" and use it as the proxy of match degree. The match degree does not significantly differ between gender and racial groups. However, job relatedness increases with education level. In 1990, educational level increased from 0.46 for bachelor's degree to 0.88 for doctorate. Similar differences were observed for the other two years.

Table A.6 shows the reasons behind the mismatch. In the survey, people were asked about the most important reason behind working outside their fields. The table lists seven potential reasons and the calculation for the annual proportion for each reason. The data show three main reasons behind the mismatch: pay or promotion opportunities, career interests, and job unavailability in their highest degree field. These three factors constitute approximately 70% of the mismatch reported for each year.[4]

Table A.8 presents the wage ratios between matched and mismatched workers; these ratios are grouped by the aforementioned reasons. The raw data show that matched workers' average wage is higher than that of mismatched workers, regardless of the reason. However, for the residual wage, the mismatched group with the reason "pay or promotion opportunities" reports a higher or approximately the same wage as that of the matched workers. This finding implies wage variety in the mismatched workers' group. The data show that the ratios in 1990 were 0.98, 0.85, and 0.66 for promotion, preference, and search friction, respectively. Hence, distinguishing the reasons and examining the wage inequality would be meaningful.[5]

2.2. Wage effect

This subsection documents the wage effect of educational mismatch. Following Ritter and West (2014), we regress the log annual earnings on the demographic and match variables as follows:

$$ln(earnings)_{ijm} = \beta D_i + \alpha Z_j + \theta M_m + \delta_1 close_{jm} + \delta_2 some_{jm} + \gamma X_i + \epsilon_{ijm},$$

where $ln(earnings)_{ijm}$ is the log value of annual earning for individual i with occupation j and a college major m, D_i includes a vector of demographic variables (tenure, tenure^2, gender, education, and race) for individual i, Z_j denotes the occupation j, and M_m denotes the major m. $close_{jm}$ and $some_{jm}$ denote that occupation j and college major m are closely related and somewhat related, respectively. X_i includes all other factors for individual i, namely, parents' education, degree location, and work location. Finally, ϵ_{ijm} depicts the residual term. Therefore, δ_1 and δ_2 capture the wage effect of the mismatch. Particularly, $\delta_1(\delta_2)$ represents the percentage of change in earnings when mismatched workers become closely (somewhat) related.

Table 1 presents a part of the regression results and shows that $\delta_1 = 0.171$ and $\delta_2 = 0.118$; $\delta_1 = 0.229$ and $\delta_2 = 0.170$; and $\delta_1 = 0.265$ and $\delta_2 = 0.160$ in 1990,2000, and 2010, respectively. The result $\delta_1 > \delta_2 > 0$ suggests that a mismatch significantly negatively affects earnings (e.g., 17.1% in 1990). Moreover, the results show that the educational mismatch effect grows over time. Matched workers ("closely related")

[3] We also consider the response "not related" a mismatch as a robustness check.

[4] Table A.7 presents the reasons behind the mismatch for a worker with an experience of less than 10 years. The result is similar. Kambourov and Manovskii (2009) report that becoming an experienced worker typically requires 10 years. This rule is also applied in Ritter and West (2014). In our study, we will include the experienced and inexperienced worker and control for tenure.

[5] Table A.9 shows the wage ratio between the matched and mismatched groups for inexperienced workers. The result is similar.

Table 1
Wage effect of educational mismatch.

Variables	1990	2000	2010
closely related	0.171***	0.229***	0.265***
	(0.00450)	(0.00687)	(0.00681)
some related	0.118***	0.170***	0.160***
	(0.00450)	(0.00690)	(0.00688)
exp	0.0361***	0.0386***	0.0441***
	(0.000651)	(0.00101)	(0.000796)
male	0.158***	0.206***	0.165***
	(0.00332)	(0.00498)	(0.00479)
hgc	0.0681***	0.0666***	0.0858***
	(0.00134)	(0.00212)	(0.00203)
black	−0.0381***	−0.0519***	−0.107***
	(0.00620)	(0.00924)	(0.00900)
Constant	9.395***	9.711***	9.348***
	(0.0256)	(0.103)	(0.108)
Observations	92,802	55,039	62,452
R-squared	0.354	0.334	0.380

Standard errors are in parentheses ***p < 0.01, **p < 0.05, *p < 0.1
Note: This table presents the results of the following regression:

$$ln(earnings)_{ijm} = \beta D_i + \alpha Z_j + \theta M_m + \delta_1 close_{jm} + \delta_2 some_{jm} + \gamma X_i + \epsilon_{ijm},$$

where D_i includes a vector of demographic variables (tenure, tenure^2, gender, education, and race) for individual i. Z_j denotes job j, and M_m denotes college major m. The terms $close_{jm}$ and $some_{jm}$ denote that the job j and college major m are closely related and somewhat related, respectively. X_i includes all other factors for individual i, namely, parents' education, degree location, and work location. ϵ_{ijm} is the residual term. (Data source: NSCG [1993, 2003, and 2013]).

Table 2
Job relatedness (major) and wage inequality.

Variables	$var(\epsilon)$		
	1990	2000	2010
related	−0.146***	−0.278***	−0.157**
	(0.0341)	(0.0711)	(0.0673)
Constant	0.277***	0.439***	0.407***
	(0.0186)	(0.0382)	(0.0363)
Observations	45	44	44
R-squared	0.300	0.267	0.115

Standard errors are in parentheses ***p < 0.01, **p < 0.05, *p < 0.1
Note This table presents the result of the following regression:

$$var(\epsilon)_j = \beta * relatedness_j + \eta_j,$$

where $var(\epsilon)_j$ is the residual wage inequality in major j, $relatedness_j$ is the job relatedness in major j, and η_j is the residual term. Data source: NSCG.

Table 3
Job relatedness (occupation) and wage inequality.

Variables	$var(\epsilon)$		
	1990	2000	2010
relatedness	−0.0685**	−0.206***	−0.199***
	(0.0327)	(0.0601)	(0.0611)
Constant	0.208***	0.360***	0.379***
	(0.0195)	(0.0376)	(0.0388)
Observations	64	67	69
R-squared	0.066	0.153	0.137

Standard errors are in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1
Note This table presents the result of the following regression:

$$var(\epsilon)_j = \beta * relatedness_j + \eta_j,$$

where $var(\epsilon)_j$ is the residual wage inequality in occupation j, $relatedness_j$ is the job relatedness in occupation j, and η_j is the residual term. Data source: NSCG.

have 17.1%, 22.9%, and 26.5% higher annual earnings in 1990, 2000, and 2010, respectively, than mismatched workers ("not related").

2.3. Wage inequality effect

We plot the correlation between job relatedness and wage inequality among workers with college majors or occupations for different years to discern the relationship between educational mismatch and wage inequality.Figure A.1 Fig. A.1 displays the case of majors where each point represents one major. Relatedness is calculated as the proportion of matched workers (job closely related) in this major.[6] Moreover, wage inequality is calculated as the variance of the log value of residual annual earning, controlling for demographic characteristics ($var(\epsilon)$) in each major. Job relatedness and wage inequality across majors show a negative correlation. Moreover, a simple regression is performed to determine the correlation significance as follows:

$$var(\epsilon)_j = \beta * relatedness_j + \eta_j.$$

Table 2 presents the result, showing that the correlation is negative and significant, with values of −0.146, −0.278, and −0.157 in 1990, 2000, and 2010, respectively. Although implying the causality is not necessary, job relatedness suggests potential importance, including within majors. Similarly,Figure A.3 plots job relatedness and wage inequality across occupations,[7] showing a negative correlation. Furthermore, Table 3 shows that the correlation is significant at −0.0685, −0.206, and −0.199 in 1990, 2000, and 2010, respectively. This result also suggests that job relatedness is potentially important, including within occupations.

This finding may be counter intuitive at first glance as it may conclude that a match amplifies wage difference. However, inequality depends not only on variety within matched or mismatched workers but also on the component of these groups. Thus, inequality and mismatch degree have an inverse U-shaped relationship, though the US data show inequality on the rising part of the curve.

3. Model

This section presents a tractable model that incorporates the underlying reasons behind educational mismatch. The model includes discrete variables only, and a general model with continuous variables is presented in the online Appendix. A worker has skills in two dimensions (a, k), where a and k denote the skill level and type, respectively. A firm has productivity in two dimensions (A, k'), where A and k denote productivity level and firm type, respectively. Following Berliant et al. (2006), we measured the mismatch by assuming that k and k' are in a cycle of $[0, 1]$ with a total length of 1; the distance between 0 and 1 is 0; and the maximal distance for any two points on the cycle is $\frac{1}{2}$. This distance between skill type k and productivity type k' reflects the degree of educational mismatch. In the discussion below, we assume that productivity A and skill a are homogeneous. Then, we normalized them to 1 to make the model tractable and focus on mismatch type.

3.1. Production

A worker and firm has a linear joint output, indicating firm productivity A, worker's skill level a, work effort e, and the match premium

[6] Fig. A.2 presents the case of measuring job relatedness as the percentage of people who respond with "somewhat close" and "very close".
[7] Fig. A.2 presents job relatedness measured as the percentage of people who respond with "somewhat close" or "very close."

between the worker and firm $h(k,k')$, that is, $y(a,k,k') = Aah(k,k')e$. A worker takes α share of the joint output in the form of labor compensation, that is, $w(a,k,k') = \alpha y(a,k,k')$. Then, promotion is modeled as an increase in α. As promotion can be modeled in different ways though the survey it indicates that promotion and pay are categorized together by the reason behind the mismatch (see Table A.6). In addition, promotion level in a matched job is always α_0, and the promoted value is $\alpha_1(> \alpha_0)$. Given the assumption of homogeneous ability and productivity and the exogenous offer arrival and distribution, modeling a higher promotion α is the same as a higher output. Match premium decreases in distance between job and worker's skill type $d(k,k')$ and will be effective only if they are close or spread in the match-specific knowledge as in Berliant et al. (2006). Particularly, cutoff δ_k exists such that the match premium is h_L if $d(k,k') \geq \delta_k$, and $h_H(> h_L)$, if $d(k,k') < \delta_k$. We assume that skill has the same knowledge spread, that is, $\delta_k = \delta$. Thus, formally, the match premium is as follows:

$$h(k,k') \begin{cases} = h_L & d(k,k') \geq \delta \\ = h_H d(k,k') < \delta \end{cases}.$$

3.2. Worker

A worker's utility depends on consumption level c and work effort e as follows:

$$u(c,e) = \frac{c^{1-\theta}}{1-\theta} - \frac{1}{\tau}\frac{e^{1-\rho}}{1-\rho},$$

where θ and ρ capture risk aversion and the elasticity of effort, respectively, and τ denotes job preference. A large τ implies high job amenity or a low disutility of effort. The preference level distribution for a matched job is across three values $\{\tau_L, \tau_M, \tau_H\}$, where $\tau_L < \tau_M < \tau_H$, and is τ_M in a non-matched job. Hence, preference level in the matched job could be τ_L or τ_H. After τ is realized, worker's utility maximization implies the wage function $w(\alpha, A, a, h, \tau) = [(\alpha Aah)^{1-\rho}\tau]^{\frac{1}{\theta-\rho}}$ and effort function $e(\alpha, A, a, h, \tau) = [(\alpha Aah)^{-(\theta-1)}\tau]^{\frac{1}{\theta-\rho}}$. Subsequently, the indirect utility function is as follows:

$$U(\alpha, A, a, h, \tau) = -\frac{\theta-\rho}{(\theta-1)(1-\rho)}[(\alpha Aah)^{1-\rho}\tau]^{-\frac{\theta-1}{\theta-\rho}}.$$

Given the assumption that $\rho < 1 < \theta$, wages and utility increase in α, A, a, h, and τ, whereas effort increases in τ but decreases in α, A, a, and h.

We proceed with further assumptions. Workers have the same skill level a, which is normalized to 1, and firms are homogeneous in productivity ($A = 1$). These simplifications are established for two reasons. First, this study examines inequality only for the highly educated workers for whom the skill difference could be small. Second, this model focuses on type mismatch, omitting productivity and skill differences.

3.3. Value function

Instant utility is $U(\alpha_0, h_H, \tau)$ for a worker who is working in a matched job (S) with preference τ. Here, we simplify the notation by denoting $U(\alpha, h, \tau) = U(\alpha, 1, 1, h, \tau)$. The next period has a probability of P_s that the job and worker will be separated. Otherwise, a probability P_δ that the worker will be offered a matched job from outside exists. Job amenity in any matched job is τ', which is drawn from $\{\tau_L, \tau_M, \tau_H\}$. With a probability of $(1-P_\delta)$, the worker will receive an offer for a mismatched job (g). For the mismatched job with a probability of P_α, the promotion level is α_1. Probability $(1-P_\alpha)$ that the promotion is α_0 exists. In both cases, the worker opts between accepting and rejecting the job offer. Hence, the value function can be written as follows:

$$\begin{aligned} V_s(\tau) = U(\alpha_0, h_H, \tau) + \beta E_{\tau'} \{ & P_s V_U + (1-P_s)\left[P_\delta V_s(\tau') \right. \\ & + (1-P_\delta)\left\{P_\alpha max[V_g(\alpha_1), V_s(\tau')] \right. \\ & \left. \left. + (1-P_\alpha) max[V_g(\alpha_0), V_s(\tau')]\right\}\right]\}, \end{aligned}$$

where β is the time discount rate in the utility function.

The indirect utility function for the worker in a mismatched job with promotion α is $U(\alpha, h_L, \tau_M)$ because the match premium is h_L, and the preference is τ_M. The next period has a probability of P_s that the job and worker will be separated. Otherwise, a probability of P_δ exists indicating that the worker will be offered a matched job outside the firm with a promotion level of α_0, of which the preference is τ' drawn from $\{\tau_L, \tau_M, \tau_H\}$. A probability of $(1-P_\delta)$ exists indicating that the worker will receive an offer for a mismatched job with the same promotion level. Subsequently, the value function is as follows:

$$\begin{aligned} V_g(\alpha) = U(\alpha, h_L, \tau_M) + \beta E_{\tau'} \{ & P_s V_U + (1-P_s)\left[(1-P_\delta)V_g(\alpha) \right. \\ & \left. + P_\delta max[V_g(\alpha), V_s(\tau')]\right]\}. \end{aligned}$$

An unemployed worker will have an unemployment benefit $\overline{V}$ in the current period. In the next period, a probability of P_f exists indicating that this worker will receive an offer. A probability of P_δ exists indicating that the offer is a matched job with preference from a random draw and a probability of $(1-P_\delta)$ exists indicating that the offer is a mismatched job with a promotion level of α_0. Then, the value function can be written as follows:

$$V_U = \overline{V} + \beta E_{\tau'}\{(1-P_f)V_U + P_f[P_\delta V_s(\tau') + (1-P_\delta)V_g(\alpha_0)]\}.$$

The model shows that mismatch occurs when $V_g(\alpha) > V_s(\tau)$ for all (α, τ). Let $D(\alpha)$ be the set of preferences on a matched job where a mismatch occurs, that is, $D(\alpha) = \{\tau : V_g(\alpha) > V_s(\tau)\}$. Thus, the profile for mismatch is $\{(\alpha_1, \tau) : \tau \in D(\alpha_1)\}$ and $\{(\alpha_0, \tau) : \tau \in D(\alpha_0)\}$. A mismatch through occupational choice is differentiated into two parts based on two underlying reasons, namely, preference and promotion. Subsequently, the set of promotion is $PM = \{(\alpha_1, \tau) : \tau \in D(\alpha_1), \tau \notin D(\alpha_0)\}$, and others are due to preference.

In the model, firm behavior is passive. Firm entry is absent, and the offer's distribution is also exogenous. Although the firm effect is substantial, as documented in the literature (e.g. Song et al., 2019), we simplify it to focus on occupational choice. We consider the possibility wherein an endogenous distribution may amplify the sorting between a worker and firms, increasing wage inequality. Although we do not model knowledge spread explicitly, we allow its changes in the quantitative analysis to capture this effect to a certain extent.

4. Equilibrium definition

An equilibrium consists of employment allocation $\{N_U, N_s, N_{g0}, N_{g1}\}$, where N_U is the number of unemployment, N_s is the employment in matched job, N_{g0} is the employment in mismatched job with promotion level α_0, and N_{g1} is the employment in mismatched job with promotion level α_1. Workers make an occupational choice in every period based on the current status (α, τ, h) to maximize the expected utility $\{V_U, V_s(\tau), V_g(\alpha)\}$. In the stationary equilibrium, the employment distribution requires the following conditions be satisfied:

1. Unemployed workers include unlucky job seekers and unlucky employed workers as follows:

   $$N_U = N_U(1-P_f) + (N_{g0} + N_{g1} + N_s)P_s.$$

2. Workers in mismatched jobs with promotion α_0 comprise lucky job seekers, stayers, and switchers from matched jobs as follows:

   $$N_{g0} = N_U P_f(1-P_\delta) + N_{g0}Pr(g_0|g_0) + N_s Pr(g_0|s),$$

   where $Pr(g_0|g_0)$ is the probability of staying in a mismatched job with promotion α_0, and $Pr(g_0|s)$ is the probability of switching from a matched job.
3. Workers in a mismatched job with promotion α_1 include stayers and switchers from matched jobs as follows:

   $$N_{g1} = N_{g1}Pr(g_1|g_1) + N_s Pr(g_1|s),$$

Table 4
Multiple equilibria.

Equilibrium	$D(\alpha_0)$	$D(\alpha_1)$	*PM*	*PF*
Eq.(1)	$\{\tau_L\}$	$\{\tau_L, \tau_M, \tau_H\}$	$\{(\alpha_1, \tau_M), (\alpha_1, \tau_H)\}$	$\{(\alpha_0, \tau_L), (\alpha_1, \tau_L)\}$
Eq. (2)	$\{\tau_L\}$	$\{\tau_L, \tau_M\}$	$\{(\alpha_1, \tau_M)\}$	$\{(\alpha_0, \tau_L), (\alpha_1, \tau_L)\}$
Eq. (3)	$\{\tau_L\}$	$\{\tau_L\}$	$\emptyset$	$\{(\alpha_0, \tau_L), (\alpha_1, \tau_L)\}$
Eq. (4)	$\emptyset$	$\{\tau_L, \tau_M, \tau_H\}$	$\{(\alpha_1, \tau_L), (\alpha_1, \tau_M), (\alpha_1, \tau_H)\}$	$\emptyset$
Eq. (5)	$\emptyset$	$\{\tau_L, \tau_M\}$	$\{(\alpha_1, \tau_L), (\alpha_1, \tau_M)\}$	$\emptyset$
Eq. (6)	$\emptyset$	$\{\tau_L\}$	$\{(\alpha_1, \tau_L)\}$	$\emptyset$
Eq. (7)	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Note This table lists all seven equilibria. Columns $D(\alpha_0)$ and $D(\alpha_1)$ represent the set of job amenities of a job switcher when the promotion level is α_0 and α_1, respectively. "*PM*" and "*PF*" columns list the combination of promotion levels and amenities due to promotion and preference, respectively.

where $Pr(g_1|g_1)$ is the probability of staying in a mismatched job with promotion α_1, and $Pr(g_1|s)$ is the probability of switching from a matched job.

4. Workers in a matched job are lucky job seekers, switchers from mismatched jobs, and stayers as follows:

$$N_s = N_U P_f P_\delta + N_{g0} Pr(s|g_0) + N_s Pr(s|s),$$

where $Pr(s|s)$ is the probability of staying in a matched job; and $Pr(s|g_0)$ is the probability of switching from a mismatched job with promotion α_0.

5. The total number of labor force is normalized to 1, leading to the following

$$1 = N_U + N_{g0} + N_{g1} + N_s.$$

4.1. Equilibrium results

Theoretically, Table 4 lists potential multiple equilibria, though only two equilibria with non-empty sets (Eq (1), Eq (2)) are listed. As shown in the quantitative part, only Eq (1) match the data. Hence, we will focus on this equilibrium only.[8] Eq (1) is characterized as $D(\alpha_0) = \{\tau_L\}$ and $D(\alpha_1) = \{\tau_L, \tau_M, \tau_H\}$, and the profile of a mismatched worker is $\{(\alpha_0, \tau_L), (\alpha_1, \tau_L), (\alpha_1, \tau_M), (\alpha_1, \tau_H)\}$. Furthermore, a mismatch due to promotion is described. When given a promotion level α_1, workers will select a mismatched task. However, if the promotion level is downgraded to α_0, then the worker will select a matched task. Accordingly, the sets for a promotion-driven mismatched worker and preference are $(\alpha_1, \tau_M), (\alpha_1, \tau_H)$, and $\{(\alpha_0, \tau_L), (\alpha_1, \tau_L)\}$, respectively. Employment and wages are computed in Appendix B and summarized as follows.

Employment: First, N_U, N_s, N_{g0}, and N_{g1} are computed by solving the equilibrium conditions. Let N_{PF}, N_{PM},and N_{SF} be the number of total mismatched employment due to preference, promotion, and search friction, respectively, and N_{sL}, N_{sM}, and N_{sH} the number of workers in a matched job with preference τ_L, τ_M, and τ_H, respectively. Then, they are computed in Equations (2)–(7).

Wages: Let w_{PF}, w_{PM}, and w_{SF} be the wages of mismatched workers due to preference, promotion, and search friction, respectively, and w_{sL}, w_{sM}, and w_{sH} the wages of matched workers with the preferences τ_L, τ_M, τ_H, respectively. Furthermore, w_g and w_s are the average wages of mismatched and matched workers, respectively. Given the wage function $w(\alpha, h, \tau) = [(\alpha h)^{1-\rho}\tau]^{\frac{1}{\theta-\rho}}$, average wages are computed in Equations (8)–(15).

Income inequality: Total inequality can be decomposed into within-group inequality ($Var_j, j = g, s$) and between-group inequality $((\overline{lnw_j} - \overline{lnw})^2$ as follows:

$$Var(lnw) = \sum_{j=g,s} \frac{N_j}{N_g + N_s}[Var_j + (\overline{lnw_j} - \overline{lnw})^2],$$

where Var_g is the inequality within a mismatched job,

$$Var_g = \sum_{j=0,1} \frac{N_{gj}}{N_g}(lnw(\alpha_j, h_L, \tau_M) - \overline{lnw_g})^2,$$

and Var_s is the inequality within a matched job,

$$Var_s = \sum_{j=L,M,H} \frac{N_{sj}}{N_s}(lnw(\alpha_0, h_H, \tau_j) - \overline{lnw_s})^2.$$

We will quantify the impact of educational mismatch on income inequality in Section 5 below based on the functions in Appendix B.

5. Quantitative analysis

We calibrated the model with the data for 1990 as the benchmark in the quantitative analysis. Subsequently, we recalibrated the channel- and non-channel specific parameters in the model by targeting the economy in 2000. Finally, we conducted several counterfactual experiments, given the parameters in 1990 and 2000 to examine the impact of each factor on wage inequality. We performed the analysis only for these two years because wage inequality increases rapidly until 2000. Tables A.1 and A.2 show that wage inequality increases from 0.23 in 1990 to 0.34 in 2000 and turns to only 0.39 in 2010.

5.1. Calibration

The following parameters need to be calibrated: preference parameters τ_H, τ_M, τ_L, and P_H, P_M, and P_L; promotion parameters α_0, α_1, and P_α; search friction parameters P_δ, P_f, and P_s; skill premium parameters h_L, h_H; unemployment benefit $\overline{V}$; elasticity of effort ρ; risk aversion parameter θ; and time discount β. First, we normalized the parameters of match premium and preference in a mismatched job, that is ($h_L = 1$) and ($\tau_M = 1$), respectively. Following the literature, we set $\theta = 2$ and $\beta = 0.95$. Other parameters were calibrated by jointly targeting several main characteristics.

Table 5 presents the calibration results in 1990. The terms (α_0, α_1) denote promotion levels (low, high). P_α is the promotion probability; h_H is the match premium; (τ_H, τ_L) are job amenities (high, low); and (P_H, P_M, P_L) are the probability distributions on job amenity with $P_H + P_M + P_L = 1$. The term P_δ denotes knowledge spread, and the terms (P_f, P_s) indicate job finding and separation rates, respectively. The terms $(\overline{V}, \rho)$ denote the unemployment disutility and elasticity of effort, respectively.

Data are taken from NSCG (1993), and the targets include labor share (ls); matched-to-mismatched workers employment ratio $(\frac{N_s}{N_g})$

[8] In the quantitative analysis, we will allow the model to choose an equilibrium to match the data.

Table 5
Parameters in 1990.

Parameters	Descriptions	Value	Target	Data	Model
α_0	Promotion level (low)	0.34	ls	0.60	0.43
α_1	Promotion level (high)	0.99	Var_g	0.27	0.19
P_α	Promotion probability	0.09	$\frac{N_{PM}}{N_g}$	0.36	0.36
h_H	Match premium	2.05	$\frac{W_g}{W_s}$	0.88	0.88
τ_H	Job amenity(high)	13.87	Var_s	0.22	0.24
τ_L	Job amenity(low)	0.01	$\frac{W_{PF}}{W_s}$	0.83	0.73
P_M	Preference probability(M)	0.11	$\frac{W_{PM}}{W_s}$	1.02	1.12
P_L	Preference probability(L)	0.35	$\frac{N_{PF}}{N_g}$	0.46	0.47
P_δ	Knowledge spread	0.83	$\frac{N_s}{N_g}$	3.52	3.53
P_f	Job finding rate	0.75	N_u	0.05	0.05
P_s	Job separation rate	0.04	$\frac{W_{SF}}{W_s}$	0.72	0.81
$\overline{V}$	Unemployment disutility	−576.24	$\frac{N_{SF}}{N_g}$	0.19	0.19
ρ	Elasticity on effort	−4.09	Var	0.23	0.23

Note: The data for 1990 are from NSCG (1993). The targets include labor share (ls); matched-to-mismatched employment ratio ($\frac{N_s}{N_g}$) and wage ratio ($\frac{W_g}{W_s}$); matched-to-mismatched employment ratio due to promotion ($\frac{N_{PM}}{N_g}$), preference ($\frac{N_{PF}}{N_g}$), and search friction ($\frac{N_{SF}}{N_g}$); wage ratios between mismatched and matched employment due to promotion ($\frac{W_{PM}}{W_s}$), preference ($\frac{W_{PF}}{W_s}$), and search friction ($\frac{W_{SF}}{W_s}$); wage inequality within matched group (Var_s) and unmatched group (Var_g) and total wage inequality (Var); and unemployment rate (N_U).

Table 6
Parameters in 2000.

Parameters	Descriptions	Value	Target	Data	Model
α_0	Promotion level (low)	0.30	ls	0.60	0.41
α_1	Promotion level (high)	0.99	Var_g	0.40	0.19
P_α	Promotion probability	0.12	$\frac{N_{PM}}{N_g}$	0.35	0.35
h_H	Match premium	3.76	$\frac{W_g}{W_s}$	0.82	0.82
τ_H	Job amenity(high)	3.17	Var_s	0.31	0.38
τ_L	Job amenity(low)	0.01	$\frac{W_{PF}}{W_s}$	0.74	0.71
P_M	Preference probability(M)	0.22	$\frac{W_{PM}}{W_s}$	1.00	1.02
P_L	Preference probability(L)	0.42	$\frac{N_{PF}}{N_g}$	0.50	0.52
P_δ	Knowledge spread	0.87	$\frac{N_s}{N_g}$	3.23	3.23
P_f	Job finding rate	0.94	N_u	0.05	0.05
P_s	Job separation rate	0.05	$\frac{W_{SF}}{W_s}$	0.66	0.74
$\overline{V}$	Unemployment disutility	−343.73	$\frac{N_{SF}}{N_g}$	0.15	0.15
ρ	Elasticity on effort	−2.22	Var	0.34	0.34

Note: The data for 2000 are taken from NSCG (2003). The targets include labor share (ls); matched-to-mismatched employment ratio ($\frac{N_s}{N_g}$) and wage ratio ($\frac{W_g}{W_s}$); matched-to-mismatched employment ratio due to promotion ($\frac{N_{PM}}{N_g}$), preference ($\frac{N_{PF}}{N_g}$), and search friction ($\frac{N_{SF}}{N_g}$); wage ratios between mismatched and matched groups due to promotion ($\frac{W_{PM}}{W_s}$), preference ($\frac{W_{PF}}{W_s}$), and search friction ($\frac{W_{SF}}{W_s}$); wage inequality within the matched group (Var_s) and unmatched group (Var_g) and the total wage inequality (Var); and unemployment rate (N_U).

and corresponding wage ratio ($\frac{W_g}{W_s}$); employment components of mismatched workers by promotion ($\frac{N_{PM}}{N_g}$), preference ($\frac{N_{PF}}{N_g}$), and search friction ($\frac{N_{SF}}{N_g}$); wage ratios between mismatched and matched workers due to promotion ($\frac{W_{PM}}{W_s}$), preference ($\frac{W_{PF}}{W_s}$), and search friction ($\frac{W_{SF}}{W_s}$); wage inequality within matched and mismatched groups (Var_s, Var_g) and for all the highly educated (Var); and unemployment rate (N_u). Again, wage inequality is the residual term, that is, $var(\epsilon)$. Overall, the model significantly matches the data, particularly for inequalities Var, Var_s, and Var_g and employment ratios $\frac{N_s}{N_g}$, $\frac{N_{PM}}{N_g}$, $\frac{N_{PF}}{N_g}$, and $\frac{N_{SF}}{N_g}$. The fitness of these variables is important as the model examines inequality through occupational choice.

The following parameter values require explanation. First, $\alpha_1 = 0.99$ implies that the worker will take most of the output in the promoted job. This situation occurs because labor is the only input in the production function. By assumption, only two promotion levels exist. Second, the large job amenity difference $\tau_H/\tau_L = 13.87/0.01$ may be the result of the assumption that the preference is drawn from the discrete distribution. The next subsection discusses the other parameters.

5.2. Counterfactual analysis

We recalibrated channel- and non-channel-specific parameters to quantify the contribution of each channel by targeting the economy in 2000. Table 6 presents the results indicating that the model matches the data well, particularly for inequalities and employment ratios, similar to that in 1990. The decade from 1990 to 2000 witnessed many changes. The difference in promotion level (α_1/α_0) has increased from 0.99/0.34 to 0.99/0.30 as well as promotion probability from 0.09 to 0.12. Hence, the change in promotion channel will potentially enlarge wage inequality. Match premium (h_H) has also increased from 2.05 to 3.76, which will also potentially increase the wage gap between matched and mismatched workers. Job amenity difference (τ_H/τ_L) has decreased from 13.87/0.01 to 3.17/0.01, and probability distribution (P_H, P_M, P_L) has changed from $(0.54, 0.11, 0.35)$ to $(0.36, 0.22, 0.42)$. The decrease in the high amenity level and increase in its probability led the preference channel to contribute negatively to wage inequality increase. Knowledge spread (P_δ), job finding rate, and job separation rate have increased from 0.83 to 0.87, from 0.75 to 0.94, and from 0.04 to 0.05, respectively. These changes may increase wage inequality by expanding the wage gap between matched and mismatched workers. Moreover, we also recalibrated two non-channel-specific parameters to capture all the impact from the residuals, that is, those excluded from the model. The residuals could have large impacts because $\overline{V}$ (from −576.24 to −343.73) and ρ (from −4.09 to − 2.22) have significant changes.

Several counterfactual experiments are conducted to examine the impact of each factor on inequality, given the parameters in 1990 and 2000. First, the preference parameters ($\tau_L, \tau_H, P_L, P_M, P_H$) in 1990 are replaced with the values in 2000. The "Wage inequality" column of Table 7 lists the residual wage inequality based on the data for 1990 and 2000 and on the benchmark model calibrated with the data for 1990. The "Counterfactual analysis" column lists wage inequalities under different counterfactual cases. Column τ_L represents the result of replacing τ_L and retaining others with the benchmark values. Similar exercises are conducted in "τ_H", "(τ_L, τ_H)", and "(P_L, P_M, P_H)" columns. The "Preference(PF)" column presents the result derived after replacing all preference parameters ($\tau_L, \tau_H, P_L, P_M, P_H$). The first row in each column of the counterfactual analysis depicts the inequality level when replacing the parameter in 1990 with that in 2000. The second row presents the difference in inequality between the counterfactual case and benchmark, where the negative value indicates smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000. Table 7 shows that probability distribution plays an important role in explaining wage inequality increase. Particularly, if (P_L, P_M, P_H) are replaced, the wage inequality change will be −5.8% of the change in the data. We conclude that the amenity probability distribution contributes −5.8% to wage inequality increase. In addition, if job amenity level (τ_L, τ_H) is replaced, then the wage inequality change will be −55.8% of the change in the data. This result is consistent with the calibration result that the difference between τ_H

Table 7
Counterfactual analysis: preference.

Wage inequality			Counterfactual analysis				
Data	Data	Model	τ_L	τ_H	(τ_L, τ_H)	(P_L, P_M, P_H)	Preference(PF)
(1990)	(2000)	(1990)					
0.232	0.336	0.232	0.251	0.158	0.174	0.226	0.184
			0.019	−0.074	−0.058	−0.006	−0.048
			0.183	−0.712	−0.558	−0.058	−0.462

Note: The "Wage inequality" column lists the inequality in 1990 and 2000 and from the benchmark that is calibrated with the data from 1990. The "Counterfactual analysis" column lists wage inequality under different counterfactual cases. Column τ_L represents the result of replacing τ_L in 1990 with the value in 2000 and retaining others with the benchmark values. Similar exercises are conducted for the columns "τ_H", "(τ_L, τ_H)", and "(P_L, P_M, P_H)", and the column "Preference(PF)" presents the result derived after replacing (τ_L, τ_H) and (P_L, P_M, P_H). The first row in each column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row presents the difference in inequality between the counterfactual case and that of the benchmark, where a negative value implies smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000.

Table 8
Counterfactual analysis: match premium.

Wage inequality			Counterfactual analysis
Data	Data	Model	Match premium (PE)
(1990)	(2000)	(1990)	
0.232	0.336	0.232	0.296
			0.064
			0.615

Note: The "Wage inequality" column lists the inequality in 1990 and 2000 and from the benchmark that is calibrated with data from 1990. The "Match premium(PE)" column presents the result of replacing h_H in 1990 with the value in 2000 and retaining others with the benchmark values. The first row of the "Counterfactual analysis" column shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the difference in inequality between the counterfactual case and that of the benchmark, where the negative value indicates smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to wage inequality change from 1990 to 2000.

and τ_L in 1990 will be considerably larger than that in 2000. In addition, if (τ_L, τ_H), and (P_L, P_M, P_H) are replaced, then the inequality will be −46.2% of the change in the data. Therefore, overall, the preference channel contributes negatively to wage inequality increase.

A similar experiment is conducted for match premium (h_H). Table 8 shows that if h_H is replaced, the inequality change is 61.5% of the change in the data. The increase in match premium from 2.05 to 3.76 suggests 61.5% contribution to wage inequality increase. Three parameters are considered for search friction: job finding rate (P_f), separation rate (P_s), and knowledge spread (P_δ). Table 9 shows that the contribution of each channel is 5.8%, 0%, and 5.8%, respectively. The overall contribution of search friction is 11.5%.

The counterfactual exercise on the promotion channel involves the parameters of promotion offer probability (P_α) and promotion levels (α_0, α_1). Table 10 illustrates that promotion probability negatively contributes to inequality increase (−11.5%), whereas promotion level distribution contributes 9.6%. The overall contribution of the promotion channel is −1%. Thus, the effect amplifies when the promotion probability and level are changed simultaneously.

Wage inequality change may be caused by factors not included in the model. We check this idea by performing similar counterfactual exercises for $(\overline{V}, \rho)$ to account for residual contribution. Although $\overline{V}$ and ρ are the parameters of unemployment utility and elasticity on effort, respectively, they do not necessarily represent their channels but all factors not included in the model. Table 11 shows that the remaining

Table 9
Counterfactual analysis: search friction.

Wage inequality			Counterfactual analysis			
Data	Data	Model	P_δ	P_f	P_s	Search friction(SF)
(1990)	(2000)	(1990)				
0.232	0.336	0.232	0.238	0.232	0.238	0.244
			0.006	0	0.006	0.012
			0.058	0	0.058	0.115

Note: The "Wage inequality" column lists the inequality in 1990 and 2000 and from the benchmark that is calibrated with the data from 1990. The "P_δ" represents the result of replacing P_δ in 1990 with the value in 2000 and retaining others with the benchmark values. Similar exercises are conducted on the "P_f" and "P_s" columns. The "Search friction(SF)" column presents the result derived after replacing (P_δ, P_f, P_s). The first row in each column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the difference in inequality between the counterfactual case and that of the benchmark, where the negative value indicates smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to wage inequality change from 1990 to 2000.

Table 10
Counterfactual analysis: promotion.

Wage inequality			Counterfactual analysis				
Data (1990)	Data (2000)	Model (1990)	P_α	α_0	α_1	(α_0, α_1)	Promotion(PM)
0.232	0.336	0.232	0.220	0.241	0.232	0.242	0.231
			−0.012	0.009	0	0.01	−0.001
			−0.115	0.087	0	0.096	−0.01

Note: The "Wage inequality" column lists the inequality in 1990 and 2000 and from the benchmark that is calibrated with the data from 1990. The "P_α" column represents the result of replacing P_α in 1990 with the value in 2000 and retaining others with the benchmark values. Similar exercises are conducted on the "α_0", "α_1", and "(α_0, α_1)" columns. The "Promotion(PM)" column presents the result derived after replacing P_α and (α_0, α_1). The first row in each column of "Counterfactual analysis" shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the difference in inequality between the counterfactual case and that of the benchmark, where the negative value indicates smaller inequality than the benchmark value in counterfactual case. The third row shows the ratio of the value in the second row to wage inequality change from 1990 to 2000.

Table 11
Counterfactual analysis: residues.

Wage inequality			Counterfactual analysis		
Data (1990)	Data (2000)	Model (1990)	$\overline{V}$	ρ	Residue(RS)
0.232	0.336	0.232	0.232	0.43	0.43
			0	0.198	0.198
			0	1.904	1.904

Note The "Wage inequality" column lists the inequality in 1990 and 2000 and from the benchmark that is calibrated with the data from 1990. The "$\overline{V}$" column presents the result of replacing $\overline{V}$ in 1990 with the value in 2000 and retaining others with the benchmark values. Similar exercises are conducted on the "ρ" column. The "Residue(RS)" column presents the result derived after replacing $\overline{V}$ and ρ. The first row in each column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the difference in inequality between the counterfactual case and that of the benchmark, where the negative indicates smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000.

part contributes to inequality by 190.4%. Therefore, the model misses a significant portion of the explanation for the inequality increase. To emphasize the magnitude of each channel, we sum all the contributions from the five components: preference (PF), match premium (PE), SF, promotion (PM), and residual (RS). Then, we computed the ratio of each component to their sum. The first and fourth rows of Table 12 present the inequality level in the counterfactual cases, the inequality difference between the counterfactual case and the benchmark, the ratio of the value in the second row to the wage change from 1990 to 2000, and the standardized value, that is, the ratio of the value in the third row to their sum, respectively. Subsequently, this standardized measurement shows that the contribution of preference, match premium, SF, and promotion are −21.4%, 28.4%, 5.3%, and −0.5%, respectively. We consider the contribution significant, given that the wage inequality in this study controls for demographic characteristics and that the model does not include worker skill or heterogeneity in firm productivity.

6. Alternative calibration

The counterfactual analysis in Section 4 allows all parameter changes from 1990 to 2000. Nevertheless, certain parameters could remain constant across years. Detailed information, such as match quality and knowledge spread, is required to accurately identify which parameters will change. However, such detailed information is not available in the data. Hence, we considered about two alternative calibration strategies in this section. We retained $\overline{V}$ and ρ as constant and recalibrated other parameters in the first exercise. In the model, $\overline{V}$ and ρ are the parameters of unemployment utility and elasticity on effort, respectively, though they may capture all other factors not included in the model as argued in Section 5.

Thus, this alternative calibration may significantly change the results. In the second exercise, we recalibrated only five major parameters, namely h_H, P_α, P_L, P_M, and P_δ and retained others as constant. h_H captures match premium, which is subject to change as job match

Table 12
Counterfactual analysis: decomposition.

Wage inequality			Counterfactual analysis				
Data (1990)	Data (2000)	Model (1990)	Preference(PF)	Match premium (PE)	Search friction(SF)	Promotion(PM)	Residue(RS)
0.232	0.336	0.232	0.184	0.296	0.244	0.231	0.43
			−0.048	0.064	0.012	−0.001	0.198
			−0.462	0.615	0.115	−0.01	1.904
			−0.214	0.284	0.053	−0.005	0.88

Note The "Wage inequality" column lists the inequality in 1990 and 2000 and from the benchmark that is calibrated in 1990. The "Counterfactual analysis" column lists the wage inequality under different counterfactual cases. The first row is the wage inequality level when replacing the parameters in 1990 in the counterfactual cases with those in 2000. The second row presents the difference in inequality between the counterfactual case and that in the benchmark. The third row is the ratio of the value in the second row to the wage change from 1990 to 2000. The fourth row is the standardized value, that is, the ratio of the value in the third row to the total sum.

Table 13
Alternative 1: parameters in 2000.

Parameters	Descriptions	Value	Target	Data	Model
α_0	Promotion level (low)	0.32	ls	0.60	0.39
α_1	Promotion level (high)	0.74	Var	0.34	0.31
P_α	Promotion probability	0.11	$\frac{N_{PM}}{N_g}$	0.35	0.30
h_H	Match premium	3.57	$\frac{W_g}{W_s}$	0.82	0.55
τ_H	Job amenity(high)	2.90	Var_s	0.31	0.30
τ_L	Job amenity(low)	0.0005	$\frac{W_{PF}}{W_s}$	0.75	0.47
P_M	Preference probability(M)	0.03	$\frac{W_{PM}}{W_s}$	1.00	0.68
P_L	Preference probability(L)	0.26	$\frac{N_{PF}}{N_g}$	0.50	0.37
P_δ	Knowledge spread	0.75	$\frac{N_s}{N_g}$	3.23	2.90
P_f	Job finding rate	0.70	$\frac{N_{SF}}{N_g}$	0.15	0.29
P_s	Job separation rate	0.05	$\frac{W_{SF}}{W_s}$	0.67	0.52

Note: The data for 2000 are taken from NSCG (2003), and the targets include labor share (ls); matched-to-unmatched employment ratio ($\frac{N_s}{N_g}$) and wage ratio ($\frac{W_g}{W_s}$); matched-to-unmatched employment ratio due to promotion ($\frac{N_{PM}}{N_g}$), preference ($\frac{N_{PF}}{N_g}$), and SF ($\frac{N_{SF}}{N_g}$); wage ratios between the unmatched and matched groups due to promotion ($\frac{W_{PM}}{W_s}$), preference ($\frac{W_{PF}}{W_s}$), and SF ($\frac{W_{SF}}{W_s}$); wage inequality within matched group (Var_s); and total wage inequality (Var).

quality changes. For the other three channels, we recalibrated all probability parameters as they are likely to respond directly to educational mismatch-related policy. Particularly, P_α captures the promotion probability; P_L and P_M are preference probability distribution; and P_δ is the knowledge spread.

6.1. Alternative 1

In this exercise, we retained $\overline{V}$ and ρ as constant and recalibrated other parameters. Then, we performed similar counterfactual analysis as in section 5. Table 13 presents the calibration result where we do not target unemployment rate (N_u) and wage inequality for the unmatched group (Var_g). The result differs from that in Table 6 due to smaller τ_L and P_M, though the model matched the data well. Tables A.10–A.13 present the counterfactual results, and Table 14 presents the decomposition results. It shows that match premium explains a significant part of wage inequality increase (52.9%) and that promotion channel has a negative contribution (−19.2%). However, the preference channel in this case contributes positively (73%), whereas SF channel contributes

Table 15
Alternative 2: parameters in 2000.

Parameters	Descriptions	Value	Target	Data	Model
α_0	Promotion level (low)	0.34	ls	0.60	0.50
α_1	Promotion level (high)	0.99	Var	0.34	0.10
$\boldsymbol{P_\alpha}$	**Promotion probability**	**0.92**	$\frac{N_{PM}}{N_g}$	**0.35**	**0.86**
$\boldsymbol{h_H}$	**Match premium**	**2.26**	$\frac{W_g}{W_s}$	**0.82**	**0.86**
τ_H	Job amenity(high)	13.87	Var_s	0.31	0.12
τ_L	Job amenity(low)	0.01	$\frac{W_{PF}}{W_s}$	0.74	0.86
$\boldsymbol{P_M}$	**Preference probability(M)**	**0.014**	$\frac{W_{PM}}{W_s}$	**1.00**	**0.87**
$\boldsymbol{P_L}$	**Preference probability(L)**	**0.13**	$\frac{N_{PF}}{N_g}$	**0.50**	**0.13**
$\boldsymbol{P_\delta}$	**Knowledge spread**	**0.985**	$\frac{N_s}{N_g}$	**3.23**	**3.24**
P_f	Job finding rate	0.75	$\frac{N_{SF}}{N_g}$	0.15	0.02
P_s	Job separation rate	0.04	$\frac{W_{SF}}{W_s}$	0.66	0.79

Note: The data for 2000 are taken from NSCG (2003), and the targets include matched-to-unmatched employment ratio ($\frac{N_s}{N_g}$) and wage ratio ($\frac{W_g}{W_s}$); matched-to-unmatched employment ratio due to promotion ($\frac{N_{PM}}{N_g}$) and preference ($\frac{N_{PF}}{N_g}$); and wage ratios between the unmatched and matched groups due to promotion ($\frac{W_{PM}}{W_s}$).

Table 14
Alternative 1: decomposition.

Wage inequality			Counterfactual analysis			
Data (1990)	Data (2000)	Model (1990)	Preference(PF)	Match premium (PE)	Search friction(SF)	Promotion(PM)
0.232	0.336	0.232	0.308	0.287	0.229	0.212
			0.076	0.055	−0.003	−0.02
			0.73	0.529	−0.029	−0.192

Note The "Wage inequality" column lists the inequality in 1990 and 2000 and from the benchmark that is calibrated in 1990. The "Counterfactual analysis" column lists the wage inequality under different counterfactual cases. The first row lists the wage inequality level when replacing the parameters in 1990 in the counterfactual cases with those of 2000. The second row presents the inequality difference between the counterfactual case and the benchmark. The third row presents the ratio of the value in the second row to the wage change from 1990 to 2000.

Table 16
Alternative 2: decomposition.

Wage inequality			Counterfactual analysis			
Data (1990)	Data (2000)	Model (1990)	Preference(PF)	Match premium (PE)	Search friction(SF)	Promotion(PM)
0.232	0.336	0.232	0.265	0.236	0.265	0.04
			0.033	0.004	0.033	−0.228
			0.317	0.038	0.317	−1.85

Note The "Wage inequality" column lists the inequality in 1990 and 2000 and from the benchmark that is calibrated in 1990. The "Counterfactual analysis" column lists the wage inequality under different counterfactual cases. The first row is the wage inequality level when replacing the parameters in 1990 in the counterfactual cases with those in 2000. The second row presents the inequality difference between the counterfactual case and the benchmark. The third row presents the ratio of the value in the second row to the wage change from 1990 to 2000.

negatively (−2.9%). These results are contrary to our benchmark result.

6.2. Alternative 2

In this exercise, we recalibrate only five major parameters, namely, h_H, P_α, P_L, P_M, and P_δ and retained others unchanged. Thus, P_α; P_L, P_M; and P_δ represent the promotion, preference, and SF channels, respectively. Correspondingly, we targeted only five moments that are highlighted in Table 15. Here, although the model could not match two employment ratios well, namely, $\frac{N_{PM}}{N_g}$ and $\frac{N_{PF}}{N_g}$, it matches certain untargeted moments well, such as the employment ratio and wage ratio between the matched and unmatched groups ($\frac{w_g}{w_s}$, $\frac{N_s}{N_g}$). Tables A.14–A.17 and Table 16 present the counterfactual and decomposition results, respectively. Match premium explains a small part of wage inequality increase (3.8%), whereas preference has a positive contribution (31.7%). Moreover, the SF channel explains a large part of wage inequality increase (31.7%), whereas the promotion channel has a negative but much larger contribution (−185%).

7. Conclusion

This study explains residual wage inequality by introducing educational mismatch in a structural model. First, we measured educational mismatch in a novel and direct way by employing survey data. Subsequently, we identified the underlying reasons behind the mismatch to examine different mechanisms contributing to the inequality. Finally, we found that educational mismatch significantly affects income inequality and that the impact varies based on the underlying reasons.

The policy implications of this paper are as follows. First, an improvement in the education match rate will decrease wage inequality, given the negative correlation between inequality and job relatedness. Second, promotion, preference, and search friction are the three main reasons behind the mismatch. Thus, improving educational signaling and lowering market friction to help college graduates maximize their knowledge use could helpful in lowering wage inequality. Third, policies on improving match quality may automatically increase wage inequality considering that the match premium channel explains a significant part of wage inequality increase. Fourth, this study provides channels for understanding wage inequality in other countries, such as China (Piketty et al., 2019; Huang, 2019).

The model could be extended in various dimensions. First, it may include incorporate dynamics as workers may update their preference based on working experience and on-the-job learning that can increase their skill match. Second, the model could be extended to include skill and productivity heterogeneity. Under these two extensions, preference and search friction may have a higher quantitative importance. Third, the heterogeneity of preference and promotion level could be extended into a continuum where people have continuous attitudes and promotion levels.

Declaration of competing interest

We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix.

A. Tables and Figures

Table A.1
Statistical description: 1990

Groups		Observations	Tenure	Earnings	Inequality	Proportion
Gender	Female	34,467	18.17	53,814.12	0.20	0.39
	Male	59,893	19.73	76,401.97	0.26	0.61
Education	Bachelor	58,063	18.77	61,039.09	0.23	0.64
	Master	25,757	20.17	68,416.25	0.20	0.25
	PhD	10,540	18.68	104,475.18	0.32	0.11
Race	White	79,175	19.16	68,444.02	0.24	0.91
	Black	9478	19.20	55,785.14	0.19	0.06
	Hispanic	5707	17.27	62,247.22	0.20	0.03
Relatedness	Close	55,613	18.96	70,803.82	0.22	0.56
	Somewhat	24,066	19.04	67,695.83	0.23	0.26
	Not	14,681	19.69	57,021.46	0.28	0.18
All sample		94,360	19.11	67,514.19	0.23	1

Note: The data for 1990 are taken from the NSCG(1993). The "Observations," "Tenure," "Earnings," "Inequality," and "Proportion" columns present the number of observations in the sample, average tenure of each subgroup, average earning in USD in current year value, residual wage inequality of each subgroup, and employment share of each subgroup, respectively.

Table A.2
Statistical description: 2000

Groups		Observations	Tenure	Earnings	Inequality	Proportion
Gender	Female	21,180	20.13	60,732.11	0.28	0.43
	Male	34,285	21.51	90,929.36	0.38	0.57
Education	Bachelor	31,915	20.95	69,964.33	0.34	0.64
	Master	16,202	21.31	77,226.22	0.28	0.26
	PhD	7348	19.72	131,856.72	0.49	0.10
Race	White	47,212	21.07	79,947.82	0.35	0.89
	Black	4411	20.70	61,798.49	0.25	0.07
	Hispanic	3842	18.50	65,085.11	0.33	0.05
Relatedness	Close	33,377	20.45	83,487.92	0.30	0.56
	Somewhat	13,871	21.14	77,929.15	0.33	0.26
	Not	8217	22.05	61,807.33	0.40	0.19
All sample		55,465	20.92	78,042.82	0.34	1

Note: The data for 2000 are taken from the NSCG (2003). The "Observations," "Tenure," "Earnings," "Inequality," and "Proportion" columns present the number of observation in the sample, average tenure of each subgroup, average earnings in USD in current year value, residual wage inequality of each subgroup, and employment share of each subgroup, respectively.

Table A.3
Statistical description: 2010

Groups		Observations	Tenure	Earning	Inequality	Proportion
Gender	Female	26,871	19.21	74,392.51	0.34	0.47
	Male	36,619	21.22	111,361.32	0.43	0.53
Education	Bachelor	32,722	20.51	82,050.90	0.39	0.65
	Master	22,924	20.09	96,163.92	0.31	0.26
	PhD	7844	19.16	170,399.52	0.53	0.09
Race	White	51,949	20.44	96,302.70	0.38	0.86
	Black	4873	20.54	80,507.39	0.43	0.07
	Hispanic	6668	18.16	79,348.42	0.37	0.08
Relatedness	Close	39,174	19.92	104,512.62	0.35	0.56
	Somewhat	16,218	20.54	89,270.52	0.37	0.26
	Not	8098	21.01	67,644.66	0.44	0.18
All sample		63,490	20.28	93,939.57	0.39	1

Note: The data for 2010 are taken from the NSCG (2013). The "Observations," "Tenure," "Earnings," "Inequality," and "Proportion" columns present the number of observation in the sample, average tenure of each subgroup, average earnings in USD in current year value, residual wage inequality of each subgroup, and employment share of each subgroup.

Table A.4
Income inequality

Year	Var_{raw}	Var_{res}	Var_1	Var_2
1990	0.29	0.24	0.21	0.18
2000	0.40	0.34	0.31	0.26
2010	0.47	0.39	0.35	0.29

Note: The "Var_{raw}", "Var_{res}", "Var_1", and "Var_2" columns present the earnings inequality with raw data, residual wage inequality as in Equation (1), residual wage inequality after further controlling for the college major dummy, and residual wage inequality after further controlling for college major, occupation, and match status, respectively.

Table A.5
Proportion of match

Groups		1990			2000			2010		
		Close	Some	Not	Close	Some	Not	Close	Some	Not
Gender	Female	0.61	0.21	0.18	0.61	0.22	0.17	0.59	0.23	0.17
	Male	0.53	0.29	0.18	0.52	0.29	0.19	0.53	0.29	0.18
Education	Bachelor	0.46	0.30	0.24	0.46	0.30	0.24	0.46	0.31	0.23
	Master	0.68	0.22	0.10	0.68	0.22	0.10	0.69	0.22	0.09
	PhD	0.88	0.08	0.04	0.87	0.09	0.04	0.88	0.09	0.03
Race	White	0.59	0.26	0.15	0.60	0.25	0.15	0.56	0.27	0.17
	Black	0.60	0.22	0.18	0.59	0.24	0.17	0.54	0.26	0.20
	Hispanic	0.61	0.23	0.15	0.66	0.21	0.13	0.58	0.23	0.19
All sample		0.56	0.26	0.18	0.56	0.26	0.19	0.56	0.26	0.19

Note: The data are taken from the NSCG (1993, 2003, and 2013). The "Close," "Somewhat," and "Not" columns present the proportion of people who reported "closely related," "somewhat related," and "not related," respectively.

Table A.6
Reasons for mismatch

Reason	1990	2000	2010
Pay and promotion opportunities	0.32	0.32	0.29
Working conditions [hours, equipment, working environment]	0.08	0.10	0.09
Job location	0.04	0.06	0.07
Change in career or professional interests	0.20	0.20	0.19
Family-related reasons	0.08	0.10	0.10
Unavailability of jobs in highest degree field	0.16	0.14	0.18
Others	0.12	0.07	0.08

Note: The "1990," "2000," and "2010" columns present the percentage of workers reporting different mismatch reasons among those who reported "not related" for 1990, 2000, and 2010, respectively.

Table A.7
Reasons for mismatch: exp ≤ 10

Reasons	1990	2000	2010
Pay and promotion opportunities	0.30	0.35	0.27
Working conditions [hours, equipment, working environment]	0.07	0.09	0.09
Job location	0.04	0.05	0.05
Change in career or professional interests	0.18	0.22	0.15
Family-related reasons	0.07	0.08	0.07
Unavailability of jobs in highest degree field	0.21	0.17	0.28
Others	0.12	0.05	0.09

Note: The "1990," "2000," and "2010" columns present the percentage of workers reporting different mismatch reasons among those who reported "not related" for 1990, 2000, and 2010, respectively. This table includes only workers with an experience of 10 years and below.

Table A.8
Wage ratio to matched group

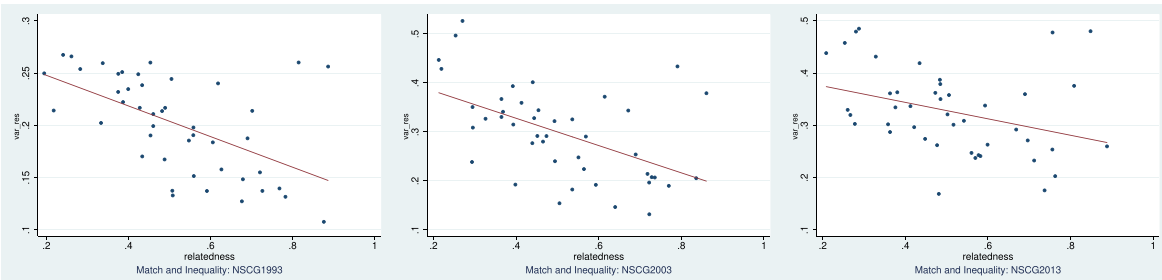
	Raw			Residual		
Reasons	1990	2000	2010	1990	2000	2010
Pay and promotion opportunities	0.98	0.94	0.89	1.02	1.00	0.99
Working conditions [hours, equipment, working environment]	0.74	0.66	0.59	0.78	0.73	0.66
Job location	0.68	0.64	0.57	0.72	0.69	0.65
Change in career or professional interests	0.85	0.76	0.74	0.89	0.81	0.81
Family-related reasons	0.70	0.59	0.52	0.78	0.66	0.59
Unavailability of jobs in highest degree field	0.66	0.60	0.49	0.72	0.67	0.59
Others	0.74	0.71	0.57	0.79	0.75	0.65

Note: The "Raw" and "Residual" columns present the raw wage ratio of the mismatched group (under different reasons) to the matched group for 1990, 2000, and 2010 and the residual wage ratio (after controlling for demographic characteristics) of the mismatched group to the matched group for 1990, 2000, and 2010, respectively.

Table A.9
Wage ratio to matched group: exp ≤ 10

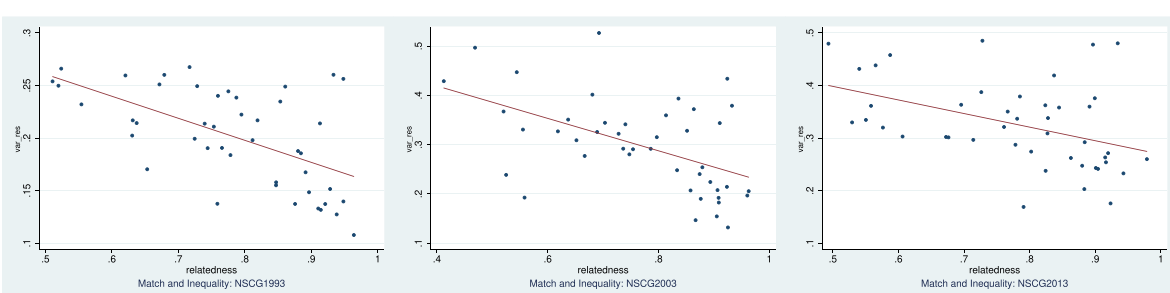
	Raw			Residual		
Reasons	1990	2000	2010	1990	2000	2010
Pay and promotion opportunities	0.92	0.94	0.82	0.98	1.01	0.92
Working conditions [hours, equipment, working environment]	0.74	0.80	0.61	0.80	0.88	0.67
Job location	0.70	0.72	0.59	0.73	0.81	0.72
Change in career or professional interests	0.78	0.77	0.69	0.84	0.83	0.80
Family-related reasons	0.76	0.56	0.70	0.81	0.64	0.81
Unavailability of jobs in highest degree field	0.64	0.58	0.52	0.71	0.66	0.63
Others	0.66	0.71	0.60	0.75	0.75	0.67

Note: The "Raw" and "Residual" columns present the raw wage ratio of the mismatched group (under different reasons) to the matched group for 1990, 2000, and 2010 and the residual wage ratio (after controlling for demographic characteristics) of the mismatched group to the matched group for 1990, 2000, and 2010, respectively. This table includes only workers with an experience of 10 years and below.



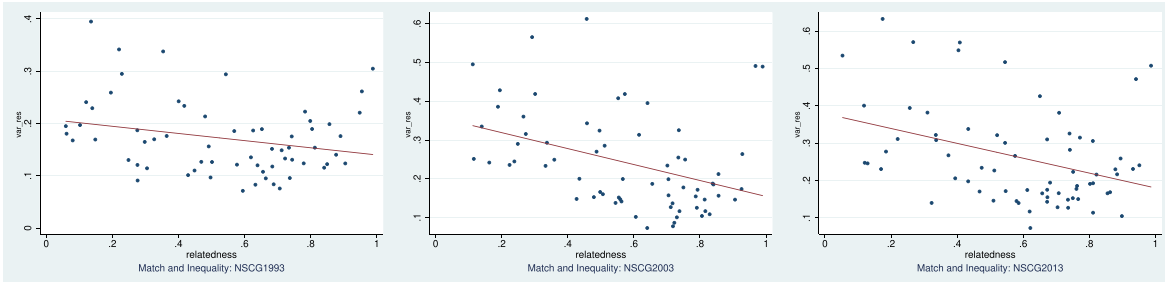
Note: This figure shows the correlation between major relatedness and within-major inequality. Relatedness is defined as the percentage of workers who respond "very close." In each panel, a dot, the x-axis, and the y-axis represent a college major, the relatedness in that major, and the residual wage inequality ($var(\epsilon)$) within that major, respectively. The left, middle, and right panels represent the results for 1990, 2000, and 2010, respectively.

Fig. A.1 Job relatedness (major) and wage inequality.



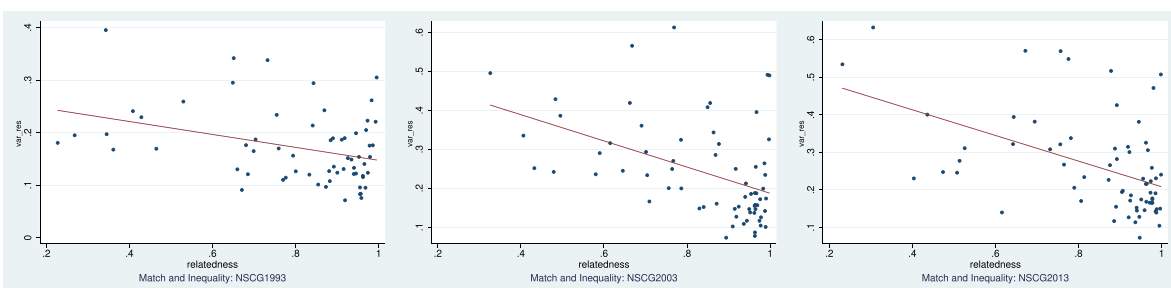
Note: This figure shows the correlation between major relatedness and within-major inequality. Relatedness here is defined as the percentage of workers who respond "somewhat close" or "very close." In each panel, a dot, the x-axis, and the y-axis represent a college major, the relatedness in that major, and the residual wage inequality ($var(\epsilon)$) within that major, respectively. The left, middle, and right panels represent the results for 1990, 2000, and 2010, respectively.

Fig. A.2 Job relatedness (major) and wage inequality.



Note: This figure shows the correlation between job (occupation) relatedness and within-job inequality. Relatedness is defined as the percentage of workers who respond "very close." In each panel, a dot, the x-axis, and the y-axis represent an occupation, the job relatedness in that occupation, and the residual wage inequality ($var(\epsilon)$) within that occupation, respectively. The left, middle, and right panels represent the results for 1990, 2000, and 2010, respectively.

Fig. A.3 Job relatedness (occupation) and wage inequality.



Note: This figure shows the correlation between job (occupation) relatedness and within-job inequality. Relatedness here is defined as the percentage of workers who respond "somewhat close" or "very close." In each panel, a dot, the x-axis, and the y-axis represent an occupation, the job relatedness in that occupation, and the residual wage inequality ($var(\epsilon)$) within that occupation, respectively. The left, middle, and right panels represent the results for 1990, 2000, and 2010, respectively.

Fig. A.4 Job relatedness (occupation) and wage inequality.

Table A.10
Alternative 1: preference

Wage inequality			Counterfactual analysis				
Data	Data	Model	τ_L	τ_H	(τ_L, τ_H)	(P_L, P_M, P_H)	Preference(PF)
(1990)	(2000)	(1990)					
0.232	0.336	0.232	0.46	0.154	0.353	0.211	0.308
			0.228	−0.078	0.121	−0.021	0.076
			2.19	−0.75	1.163	−0.202	0.73

Note: The "Wage inequality" column lists the inequality from 1990 to 2000 and from the benchmark that is calibrated with the data for 1990. The "Counterfactual analysis" column lists the wage inequality under different counterfactual cases. The τ_L column presents the result of replacing τ_L in 1990 with the value in 2000 and retaining others with the benchmark values. Similar exercises are conducted on the "τ_H", "(τ_L, τ_H)", and "(P_L, P_M, P_H)" columns. The "Preference(PF)" column is the result derived after replacing (τ_L, τ_H) and (P_L, P_M, P_H). The first row in each column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the inequality difference between the counterfactual case and the benchmark, where a negative value implies smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000.

Table A.11
Alternative 1: match premium

Wage inequality			Counterfactual analysis
Data	Data	Model	Match premium (PE)
(1990)	(2000)	(1990)	
0.232	0.336	0.232	0.287
			0.055
			0.529

Note: The "Wage inequality" column lists the inequality from 1990 to 2000 and from the benchmark that is calibrated with the data from 1990. The "Match premium(PE)" column presents the result of replacing h_H in 1990 with the value in 2000 and retaining others with the benchmark values. The first row in the column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the inequality difference between the counterfactual case and the benchmark, where the negative value indicates smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000.

Table A.12
Alternative 1: search friction

Wage inequality			Counterfactual analysis			
Data	Data	Model	P_δ	P_f	P_s	search_friction(SF)
(1990)	(2000)	(1990)				
0.232	0.336	0.232	0.221	0.232	0.238	0.229
			−0.011	0	0.006	−0.003
			−0.106	0	0.058	−0.029

Note: The "Wage inequality" column lists the inequality from 1990 to 2000 and from the benchmark that is calibrated with the data from 1990. The "P_δ" column presents the result of replacing P_δ in 1990 with the value in 2000 and retaining others with benchmark values. Similar exercises are conducted on the "P_f" and "P_s" columns. The "search_friction(SF)" column presents the result after replacing (P_δ, P_f, P_s). The first row in each column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the inequality difference between the counterfactual case and the benchmark, where the negative value indicates smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000.

Table A.13
Alternative 1: promotion

Wage inequality			Counterfactual analysis				
Data	Data	Model	P_α	α_0	α_1	(α_0, α_1)	promotion(PM)
(1990)	(2000)	(1990)					
0.232	0.336	0.232	0.222	0.236	0.222	0.223	0.212
			−0.01	0.004	−0.01	−0.009	−0.02
			−0.096	0.038	−0.096	−0.087	−0.192

Note: The "Wage inequality" column lists the inequality from 1990 to 2000 and from the benchmark that is calibrated with the data from 1990. The P_α column presents the result of replacing P_α in 1990 with the value in 2000 and retaining others with the benchmark values. Similar exercises are conducted on the "α_0", "α_1", and "(α_0, α_1)" columns. The "promotion(PM)" column presents the result derived after replacing P_α and (α_0, α_1). The first row in each column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the inequality difference between the counterfactual case and the benchmark, where the negative value indicates smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000.

Table A.14
Alternative 2: preference

Wage inequality			Counterfactual analysis	
Data	Data	Model	(P_L, P_M, P_H)	Preference(PF)
(1990)	(2000)	(1990)		
0.232	0.336	0.232	0.131	0.131
			−0.101	−0.101
			−0.971	−0.971

Note: The "Wage inequality" column lists the inequality from 1990 to 2000 and from the benchmark that is calibrated with the data for 1990. The "Counterfactual analysis" column lists the wage inequality under different counterfactual cases. The τ_L column represents the result after replacing τ_L in 1990 with the value in 2000 and retaining others with the benchmark values. Similar exercises are conducted on the "τ_H", "(τ_L, τ_H)", and "(P_L, P_M, P_H)" columns. The "Preference(PF)" column presents the result derived after replacing (τ_L, τ_H) and (P_L, P_M, P_H). The first row in each column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the inequality difference between the counterfactual case and the benchmark, where a negative value implies smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000.

Table A.15
Alternative 2: match premium

Wage inequality			Counterfactual analysis
Data	Data	Model	Match premium (PE)
(1990)	(2000)	(1990)	
0.232	0.336	0.232	0.236
			0.004
			0.038

Note: The "Wage inequality" column lists the inequality from 1990 to 2000 and from the benchmark that is calibrated with the data from 1990. The "Match premium(PE)" column presents the result of replacing h_H in 1990 with the value in 2000 and retaining others with the benchmark values. The first row in the column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the inequality difference between the counterfactual case and the benchmark, where the negative value indicates smaller inequality in the counterfactual case is than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000.

Table A.16
Alternative 2: search friction

Wage inequality			Counterfactual analysis	
Data	Data	Model	P_δ	search_friction(SF)
(1990)	(2000)	(1990)		
0.232	0.336	0.232	0.265	0.265
			0.033	0.033
			0.317	0.317

Note: The "Wage inequality" column lists the inequality from 1990 to 2000 and from the benchmark that is calibrated with the data from 1990. The "P_δ" column presents the result after replacing P_δ in 1990 with the value in 2000 and retaining others with benchmark values. Similar exercises are conducted on the "P_f" and "P_s" columns. The, "search_friction(SF)" column presents the result after replacing (P_δ, P_f, P_s). The first row in each column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the inequality difference between the counterfactual case and the benchmark, where the negative value indicates smaller inequality in the counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000.

Table A.17
Alternative 2: promotion

Wage inequality			Counterfactual analysis	
Data	Data	Model	P_α	promotion(PM)
(1990)	(2000)	(1990)		
0.232	0.336	0.232	0.04	0.04
			−0.228	−0.228
			−1.85	−1.85

Note: The "Wage inequality" column lists the inequality from 1990 to 2000 and from the benchmark that is calibrated with the data from 1990. The P_α column presents the result after replacing P_α in 1990 with the value in 2000 and retaining others with the benchmark values. Similar exercises are conducted on the "α_0", "α_1", and "(α_0, α_1)" columns. The "promotion(PM)" column presents the result derived after replacing P_α and (α_0, α_1). The first row in each column of the counterfactual analysis shows the inequality level when replacing the parameter in 1990 with that in 2000. The second row shows the inequality difference between the counterfactual case and the benchmark, where the negative value indicates smaller inequality in counterfactual case than the benchmark value. The third row shows the ratio of the value in the second row to the wage inequality change from 1990 to 2000.

B. *Appendix of benchmark model*

Stationary Equilibrium: An equilibrium consists of employment allocation $\{N_U, N_s, N_{g0}, N_{g1}\}$, where N_U, N_s, N_{g0}, and N_{g1} is the number of unemployment, employment in matched job, employment in mismatched job with promotion level α_0, and employment in mismatched job with promotion level α_1, respectively. In each period, workers make occupational choice based on the current status (α, τ, h) to maximize the expected utility $\{V_U, V_s(\tau), V_g(\alpha)\}$. In the stationary equilibrium (Eq (1)), employment distribution requires following conditions.

1. The unemployed workers include unlucky job seekers and unlucky employed workers as follows:

$$N_U = N_U(1 - P_f) + (N_{g0} + N_{g1} + N_s)P_s.$$

2. The workers in mismatched jobs with promotion a0 are lucky job seekers, stayers, and switchers from matched jobs due to low amenity as follows:

$$N_{g0} = N_U P_f(1 - P_\delta) + N_{g0}(1 - P_s)[(1 - P_\delta) + P_\delta P_L] + N_s(1 - P_s)(1 - P_\delta)(1 - P_\alpha)P_L.$$

3. The workers in mismatched job with promotion α_1 are stayers and switchers from matched jobs due to promotion as follows:

$$N_{g1} = N_{g1}(1 - P_s) + N_s(1 - P_s)(1 - P_\delta)P_\alpha.$$

4. The workers in matched job are lucky job seekers, switchers from mismatched jobs, and stayers as follows:

$$N_s = N_U P_f P_\delta + N_{g0}(1 - P_s)P_\delta(P_H + P_M) + N_s(1 - P_s) \times [P_\delta + (1 - P_\delta)(1 - P_\alpha)(P_H + P_M)].$$

5. The total number of labor force is normalized to 1, leading to the following

$$1 = N_U + N_{g0} + N_{g1} + N_s.$$

Employment: Let N_{PF}, N_{PM}, and N_{SF} be the mismatched employment due to preference, promotion, and SF, respectively. N_{sL}, N_{sM}, and N_{sH} are the number of workers in matched job with preference of τ_L, τ_M, and τ_H, respectively. Employment in an mismatched job of stayers with different promotion levels and switchers from matched jobs due to preference is as follows:

$$N_{PF} = N_{g0}(1 - P_s)P_\delta P_L + N_{g1}(1 - P_s)P_\delta P_L + N_s(1 - P_s)(1 - P_\delta)P_L. \tag{2}$$

Employment in a mismatched job of stayers and switchers from matched jobs due to promotion is as follows:

$$N_{PM} = N_{g1}(1 - P_s)P_\delta(P_M + P_H) + N_s(1 - P_s)(1 - P_\delta)P_\alpha(P_M + P_H). \tag{3}$$

Employment in an mismatched job of lucky job seekers receiving offers of mismatched jobs and stayers with different promotion levels who did not receive a matched job offer due to SF is as follows:

$$N_{SF} = N_U P_f(1 - P_\delta) + N_{g0}(1 - P_s)(1 - P_\delta) + N_{g1}(1 - P_s)(1 - P_\delta). \tag{4}$$

Employment in a matched job with low job amenity that comes from lucky job seekers and stayers is as follows:

$$N_{sL} = N_U P_f P_\delta P_L + N_s(1 - P_s)P_\delta P_L. \tag{5}$$

Employment in matched jobs with medium job amenity of lucky job seekers, switchers from mismatched jobs with promotions of α_0, and stayers is as follows:

$$N_{sM} = N_U P_f P_\delta P_M + N_{g0}(1-P_s)P_\delta P_M + N_s(1-P_s) \times (1-P_\delta)(1-P_\alpha)P_M + N_s(1-P_s)P_\delta P_M. \tag{6}$$

Employment in matched jobs with high job amenity of lucky job seekers, switchers from mismatched jobs, and stayers is as follows:

$$N_{sH} = N_U P_f P_\delta P_H + N_{g0}(1-P_s)P_\delta P_H + N_s(1-P_s) \times [(1-P_\delta)(1-P_\alpha) + P_\alpha]P_H. \tag{7}$$

Wages: Let w_{PF}, w_{PM}, and w_{SF} be the wage of mismatched workers due to preference, promotion, and SF, respectively; w_{sL}, w_{sM}, and w_{sH} are the wages of matched workers with the preference τ_L, τ_M, and τ_H, respectively; and w_g and w_s are the average wage of mismatched and matched workers, respectively. Given the wage function $w(\alpha, h, \tau) = [(\alpha h)^{1-\rho}\tau]^{\frac{1}{\theta-\rho}}$, the wage in mismatched job due to preference is the average wage of workers with different promotion levels as follows:

$$\begin{aligned} w_{PF} = {} & w(\alpha_0, h_L, \tau_M)\left[N_{g0}(1-P_s)P_\delta P_L + N_s(1-P_s) \times (1-P_\delta)(1-P_\alpha)P_L\right]/N_{PF} \\ & + w(\alpha_1, h_L, \tau_M)\left[N_s(1-P_s)(1-P_\delta)P_\alpha P_L + N_{g1}(1-P_s)P_\delta P_L\right]/N_{PF}. \end{aligned} \tag{8}$$

Similarly, the wage in mismatched jobs due to SF is the average wage of workers of different promotion levels as follows:

$$\begin{aligned} w_{SF} = {} & w(\alpha_0, h_L, \tau_M)[N_U P_f(1-P_\delta) + N_{g0}(1-P_s)(1-P_\delta)]/N_{SF} \\ & + w(\alpha_1, h_L, \tau_M)[N_{g1}(1-P_s)(1-P_\delta)]/N_{SF}. \end{aligned} \tag{9}$$

Given that other wages are the same for workers in the same group, the following holds:

$$w_{PM} = w(\alpha_1, h_L, \tau_M), \tag{10}$$

$$w_{sL} = w(\alpha_0, h_H, \tau_L), \tag{11}$$

$$w_{sM} = w(\alpha_0, h_H, \tau_M), \tag{12}$$

$$w_{sH} = w(\alpha_0, h_H, \tau_H), \tag{13}$$

Finally, the average wage in mismatched and matched jobs is as follows:

$$w_g = \sum_{j=0,1} \frac{N_{gj}}{N_g} w(\alpha_j, h_L, \tau_M), \tag{14}$$

$$w_s = \sum_{j=L,M,H} \frac{N_{sj}}{N_s} w(\alpha_0, h_H, \tau_j). \tag{15}$$

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.econmod.2021.105525.

References

Acemoglu, D., Autor, D., 2011. Skills, tasks and technologies: implications for employment and earnings. Handb. Labor Econ. 4, 1043–1171.

Altonji, J.G., Kahn, L.B., Speer, J.D., 2014. Trends in earnings differentials across college majors and the changing task composition of jobs. Am. Econ. Rev.: Papers & Proceedings 104 (5), 387–393.

Arcidiacono, P., 2004. Ability sorting and the returns to college major. J. Econom. 121 (1), 343–375.

Autor, D.H., Katz, L.F., Kearney, M.S., 2008. Trends in us wage inequality: revising the revisionists. Rev. Econ. Stat. 90 (2), 300–323.

Berliant, M., Reed, I.I.I.,R.R., Wang, P., 2006. Knowledge exchange, matching, and agglomeration. J. Urban Econ. 60 (1), 69–95.

Burstein, A., Morales, E., Vogel, J., 2015. Accounting for Changes in Between-Group Inequality. National Bureau of Economic Research(No. w20855).

Cooper, R., Liu, H., 2019. Mismatch in human capital accumulation. Int. Econ. Rev. 60 (3), 1291–1328.

Desjardins, R., Rubenson, K., 2011. An Analysis of Skill Mismatch Using Direct Measures of Skills. OECD Education Working Papers (63).

Dorn, D., 2009. Essays on Inequality, Spatial Interaction, and the Demand for Skills. (Ph.D. thesis).

Freeman, J.A., Hirsch, B.T., 2008. College majors and the knowledge content of jobs. Econ. Educ. Rev. 27 (5), 517–535.

Gil, P.M., Gabriel, S., Afonso, O., 2020. Is the skills mismatch important under skill-biased technological change and imperfect substitutability between immigrants and natives? Econ. Modell. 84, 38–54.

Guironnet, J.-P., Peypoch, N., 2007. Human capital allocation and overeducation: a measure of French productivity (1987, 1999). Econ. Modell. 24 (3), 398–410.

Guvenen, F., Kuruscu, B., Tanaka, S., Wiczer, D., 2020. Multidimensional skill mismatch. Am. Econ. J. Macroecon. 12 (1), 210–244.

Huang, K.X., 2019. Growth and cycles in China's unbalanced development: resource misallocation, debt overhang, economic inequality, and the importance of structural reforms. Front. Econ. China 14 (1), 53–71.

Kambourov, G., Manovskii, I., 2009. Occupational mobility and wage inequality. Rev. Econ. Stud. 76 (2), 731–759.

Kirkeboen, L.J., Leuven, E., Mogstad, M., 2016. Field of study, earnings, and self-selection. Q. J. Econ. 131 (3), 1057–1111.

Lee, S.Y., Shin, Y., Lee, D., 2015. The Option Value of Human Capital: Higher Education and Wage Inequality. National Bureau of Economic Research(No. w21724).

Lemieux, T., 2006. Increasing residual wage inequality: composition effects, noisy data, or rising demand for skill? Am. Econ. Rev. 461–498.

Lemieux, T., 2014. Occupations, fields of study and returns to education. Can. J. Econ. 47 (4), 1047–1077.
Lise, J., Postel-Vinay, F., 2020. Multidimensional skills, sorting, and human capital accumulation. Am. Econ. Rev. 110 (8), 2328–2376 110.
Nordin, M., Persson, I., Rooth, D.-O., 2010. Education–occupation mismatch: is there an income penalty? Econ. Educ. Rev. 29 (6), 1047–1059.
Piketty, T., Saez, E., 2014. Inequality in the long run. Science 344 (6186), 838–843.
Piketty, T., Yang, L., Zucman, G., 2019. Capital accumulation, private property and rising inequality in China, 1978–2015. Am. Econ. Rev. 109 (7), 2469–2496 7.
Ritter, J.A., West, K.L., 2014. Field of Study and Earnings Inequality Among the Highly Educated: 1993-2010. Minnesota Population Center.
Robst, J., 2007. Education and job match: the relatedness of college major and work. Econ. Educ. Rev. 26 (4), 397–407.
Rosen, S., 1986. The theory of equalizing differences. Handb. Labor Econ. 1, 641–692.
Sloane, P.J., 2003. Much ado about nothing? what does the overeducation literature really tell us. Overeducation in Europe 11–45.
Song, J., Price, D.J., Guvenen, F., Bloom, N., Von Wachter, T., 2019. Firming up inequality. Q. J. Econ. 134 (1), 1–50.
Tang, R., Tang, Y., Wang, P., June 2020. Within-job Wage Inequality: Performance Pay and Job Relatedness. Working Paper 27390. National Bureau of Economic Research.
Violante, G.L., 2002. Technological acceleration, skill transferability, and the rise in residual inequality. Q. J. Econ. 297–338.