Educational mismatch and earnings inequality

Tang, Rongsheng and Wang, Gaowang

Shanghai University of Finance and Economics, Shandong University

2 April 2021

Educational Mismatch and Earnings Inequality*

Rongsheng Tang

Shanghai University of Finance and Economics

Gaowang Wang

Shandong University

April 2021

Abstract

We build a model to understand educational mismatch and earnings inequality among highly educated workers. Educational mismatch has a negative wage effect and a positive correlation with wage inequality, for occupations and college majors. To disentangle different reasons or channels that contribute to wage inequality, we identity the three underlying reasons behind the mismatch—preference, promotion, and search friction—and quantify their impacts. Quantitatively, the preference and promotion channel negatively contribute to an inequality increase from 1990 to 2000; the match premium contributes to a 28.4% increase in inequality; and the contribution of search friction is 5.3%. We conclude that educational mismatch affects earnings inequality significantly and that the impact varies based on the underlying reasons. The study has important policy implications in that it shows that wage inequality can be reduced by policies for improving the education match rate and educational signaling and lowering market friction.

Keywords: educational mismatch; earnings inequality; wage effect; search friction; promotion

JEL Code: I24, J24, J31, O15

---

*Tang: Institute for Advanced Research, Shanghai University of Finance and Economics; Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, 777 Guoding Road, Shanghai, China 200433 (e-mail: tang.rongsheng@sufe.edu.cn). Wang: Center for Economic Research, Shandong University, Jinan, China. E-mail: gaowang.wang@sdu.edu.cn. 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.

1 Introduction

Educational mismatch occurs when a worker's skill type does not match the job requirements, such as in the case of a chemistry major who works as a general manager. This type of mismatch is different from that in which college graduates work in low-skill jobs. Although the latter has received much attention in the literature (e.g. Sloane (2003), Guironnet and Peypoch (2007), Lee et al. (2015)), research on type mismatch is scarce. This study examines the impact of educational mismatch on residual wage inequality among the highly educated. Since a substantial portion of inequality remains unexplained by observations (Violante (2002), Tang et al. (2020)), considering educational mismatch might be helpful. To disentangle the different channels contributing to wage inequality, we use survey data to identify the three fundamental reasons behind the mismatch and quantify their impacts. This is meaningful as there is a wide wage variety among mismatched workers due to different reasons. Although the effect of mismatch on wage inequality has been discussed in the literature to some extent (Altonji et al. (2014)), no study has examined the reasons behind the mismatch.

Specifically, we define the educational mismatch by employing the subjective responses from the National Survey of College Graduates (NSCG) in the United States. The survey participants were asked about the relatedness between their current occupation and the field of study in which they have the highest qualification; they were asked to select one of the following three possible responses: "closely related," "somewhat related," and "not at all related." We identify a participant with a skill match when the participant's response is "closely related;" otherwise, it is a mismatch.[1] For the participant, a mismatch does not necessarily mean a mistake, as the response reflects the participant's optimal behavior. However, it may, at least partially, reflect the knowledge usage efficiency. The survey also asked about the main reasons behind the mismatch; the data show that 70% of the mismatch can be attributed to the following three reasons: preference, promotion, and search friction.[2] This study will model the educational mismatch by considering the aforementioned reasons.

First, we document the wage effect of educational mismatch. Consistent with Ritter and West (2014), statistically, there is a negative wage effect of educational mismatch. The average wage ratios of mismatched to matched workers were 0.91, 0.94, and 0.87 in 1990, 2000, and 2010, respectively. In other words, these ratios depict a roughly 10% mismatch penalty. After controlling for demographic variables, the ratios become 17.1%, 22.9%, and 26.5%, respectively; it implies that, for workers with similar characteristics, the penalty could be as high as a quarter. Moreover, there is a wage variety among mismatched work-

[1] As a robustness check, we also define the mismatch only with the response of "not at all."

[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.

ers. We group mismatched workers by the three reasons and compute the wage ratios for each group. The results show that, for example, in 1990, the ratios were 0.98, 0.85, and 0.66 for promotion, preference, and search friction, respectively. There is nearly no penalty for promotion-induced mismatch, while the penalty for search friction could be as high as 1/3.

We also find a positive correlation between educational mismatch and wage inequality in the cases of both college majors and occupations. In other words, the occupations or college majors with high educational mismatch usually have a high wage inequality. This might be counter-intuitive at first glance, as it may lead to the conclusion that a better match will amplify the wage difference. However, the inequality depends not only on the variety within the matched or mismatched workers' group but also on the labor component of these two groups. In principal, there is an inverse U-shaped relationship between inequality and mismatch degree; however, as per the US data, the inequality is found on the rising part of the curve. Although this fact does not necessarily imply a causality, it suggests the potential importance of educational mismatch, especially when we have already controlled the demographic characteristics. As shown in section 2, in a simple accounting exercise, educational mismatch contributes toward 15% of the inequality. We obtain a more sophisticated result by employing our model and conducting a quantitative analysis.

In the model, workers and firms vary by skill type and productivity type, respectively. Given a worker's skill type, there are two types of jobs—matched and mismatched. In order to get a matched job, it is important for the worker to possess the related skill type, which will draw a match premium. In this case, the joint output depicts a product of the firm's productivity, the worker's skill, the match premium, and the worker's effort. A worker acquires a share of the output as the labor compensation, and promotion is defined as an increase in this share. The job amenity on a matched job is random and will affect the worker's effort. Therefore, the match degree between the worker and the job affects the human capital level and, eventually, the worker's occupational choice. In the quantitative analysis, we calibrate the model to target the US economy in both 1990 and 2000 because the wage inequality increased at a faster pace during these periods. We also allow both the channel-specific and non-channel-specific parameters to change across years. Subsequently, by counterfactual analysis, we quantify the contribution of each channel to the increase in wage inequality.

Theoretically, the underlying channels affect wage inequality through the occupational choice. In particular, a high level of promotion in a mismatched job, poor job amenities, or a low match premium in a matched job will lower the likelihood of accepting the matched job. Search friction on a matched job will also affect the employment distribution. This will affect the wage inequality through changes in the labor component and human capital quality. Quantitatively, we find that the preference and promotion channel contributes negatively to the wage inequality increase, and the match premium contributes 28.4% and search friction contributes 5.3%. Given that we only focus on the residual wage inequality, and the fact that the model does not include the worker's ability or the heterogeneity in firm productivity, we think that the effect of the mismatch is significant.

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

Related Literature

The related literature primarily includes studies on wage inequality and Educational and skill mismatch. There is a growing number of studies on wage inequality for the highly educated. The literature documents that there has been a rapid increase in wage inequality since the 1980s (e.g., Autor et al. (2008), Piketty and Saez (2014)). Skill premium has been studied intensively (e.g., Acemoglu and Autor (2011), Burstein et al. (2015)). There have been several studies on wage inequality among educational groups (e.g., Violante (2002), Lee et al. (2015), Tang et al. (2020)).

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

Recent studies on high-skilled workers have focused on the match between skill type and jobs. While Altonji et al. (2014) argue that the earnings difference across college majors can be larger than the skill premium between college and high school and that the substantial wage widening or earnings inequality between college majors is related to the task composition of occupations chosen by majors. Ritter and West (2014) found that the changing distribution of college majors causes slight, if any, shift in earnings distribution. Based on the results of these two studies, our study goes one step further and asks the reason behind the educational mismatch, given the wage difference between occupations of people with the same major. We also ask to what extent mismatch accounts for the increase in wage inequality among the highly educated.

Studies on educational mismatch usually investigate the wage effect of mismatch between college majors and occupations. Robst (2007) is one of the pioneers in employing NSCG data to measure educational mismatch. Lemieux (2014) shows that the return on education varies greatly depending on occupation, the field of study, and the match between these two factors. It is observed that the college major match-related channel accounts for close to 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 is also related to the literature on skill mismatch (e.g., Guvenen et al. (2020), Lise and Postel-Vinay (2015)). In these studies, the skill level is measured using a test score such as the ASVAB and the skill requirements in each occupation. Skill mismatch is measured as the distance between a worker's skill acquirement and the job's skill requirement. In a recent study, Cooper and Liu (2019) measured the mismatch between skills and educational attainment. Studies on skill mismatch in countries, other than the United States, for example, Desjardins and Rubenson (2011) measure it using the European data. In another study, Gil et al. (2020) examine the skill mismatch due to immigration.

The model setup is close to Berliant et al. (2006) who build a model to illustrate the exchange of knowledge as well as its consequences for agglomerative activity in a general-equilibrium search-theoretic framework. The idea of occupational choice is close to Rosen (1986).

Organization of the Paper The remainder of the present paper is organized as follows. Section 2 describes statistical and empirical facts. In Section 3, we build a model on educational mismatch, incorporating underlying reasons behind the mismatch. Section 4 describes the stationary equilibrium, and the quantitative analysis is presented in section 5. Section 6 presents quantitative results under alternative calibration strategies. Section 7 concludes the paper.

2 Facts

Some main features related to educational mismatch are documented in this section. Data is collected from the NSCG, which is a census survey of people with a college degree. Every 10 years, it provides information about the relatedness between each sampled individual's field of study and the individual's occupation. The survey participants choose from one of the following responses on the questions on relatedness: "closely related," "somewhat related," and "not at all related." We consider the participant to have a skill match if the response is "closely related," otherwise it is mismatched.[3]

The sample includes NSCG (1993, 2003, and 2013) data based on the censuses of 1990, 2000, and 2010, respectively. It adopts the following trimming strategy : it only includes full-time workers aged between 16 and 65 years. The top annual earnings are 4 million, and earnings less than 2,800 dollar are excluded. Concerning race, the study only considers white, black, and Hispanic races. Following the methodology in Altonji et al. (2014), the major code is regrouped under the 50 major categories listed in the Department of Education. Subsequently, occupation is regrouped under the method proposed by Dorn (2009), who develops a consistent three-digit occupation code. The data considers four levels of schooling year: 16, 18, 19, and 21 years. These levels are regrouped into three categories: Bachelor's (16), Master's (18, 19), and Doctorate (21). In addition, tenure is calculated as a potential experience, that is, $max(age - schooling - 6, 0)$. We obtain the wage residue based on the following regression, as in Kambourov and Manovskii (2009),

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

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

2.1 Statistical Description

Table A.1 through Table A.3 present several characteristics of the data, including the 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. The annual earning increases with the education level; however, the Master's level workers have the lowest wage inequality. White people earn higher wages and have higher inequality than other racial groups. The proportion of job relatedness does not change substantially—the proportion of closely related jobs is around 0.56. However, even if the match premium increases, educational mismatch would continue to contribute toward earnings inequality. In addition, if the main reason behind this mismatch changes, then this change would also influence earnings inequality.

Table A.4 lists several inequalities. In this table, Var_{raw} depicts the 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 major

[3] As a robustness check, we also define the mismatch only with the response of "not at all related."

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 major-occupation match is accounted as $\frac{var_1 - var_2}{var_1}$. Subsequently, it shows that, in statistical terms, the educational mismatch contributes toward 15% of the inequality. As it is a simply accounting exercise, 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 define the job relatedness as the percentage of people whose response is "closely related" and use it as the proxy of match degree. There is not much difference in match degree among the gender and racial groups. However, job relatedness increases with an increase in the educational level. For example, in 1990, it increased from 0.46 for Bachelor's degree to 0.88 for Doctorate; for the other two years, similar differences were observed.

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 all the seven potential reasons, and the proportion for each reason is calculated for each year. The data shows that there are three main reasons behind the mismatch: pay or promotion opportunities, career interests, and the unavailability of a job in their highest degree field. Roughly, these three factors constitute around 70% of the mismatch reported for each year.[4]

Table A.8 presents the wage ratios between the matched and the mismatched workers; these ratios are grouped by the aforementioned reasons. In the raw data, the average wage for matched workers is higher than that of the mismatched workers, regardless of the reason. However, for the residual wage, the mismatched group with the reason "Pay, promotion opportunity" reports a higher or almost the same wage as that of the matched workers. This implies wage variety in the mismatched workers' group. The data shows that, for example, in 1990, the ratios were 0.98, 0.85, and 0.66 for promotion, preference, and search friction, respectively. Hence, it would be meaningful to distinguish between the reasons and study the wage inequality.[5]

2.2 Wage Effect

This subsection documents the wage effect of educational mismatch. As in Ritter and West (2014), we regress the log annual earning on the demographic and match variables

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

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

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 in occupation j who graduated with 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. The terms $close_{jm}$ and $some_{jm}$ denote that the occupation j and the major m are closely related and somewhat related, respectively. X_i includes all the other factors for individual i—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. In particular, $\delta_1(\delta_2)$ represents the percentage of change in earnings when mismatched workers become closely (somewhat) related.

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 in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This is the result of the regression of

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

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

Table 1 presents a part of the regression results. It shows that $\delta_1 = 0.171$, $\delta_2 = 0.118$ in 1990; $\delta_1 = 0.229$, $\delta_2 = 0.170$ in 2000; and $\delta_1 = 0.265$, $\delta_2 = 0.160$ in 2010. The result

$\delta_1 > \delta_2 > 0$ suggests that a mismatch has a significant negative effect on earnings (e.g., 17.1% in 1990). Moreover, the results show that the effect of educational mismatch is becoming larger over time. The matched workers (closely related) have 17.1%, 22.9%, and 26.5% higher annual earnings in 1990, 2000, and 2010, respectively, than those of the mismatched workers (not related at all).

2.3 Wage Inequality Effect

To discern the relationship between the educational mismatch and wage inequality, we plot the correlation between job relatedness and wage inequality in the cases of college majors or occupations for different years. Figure A.1 displays the case of majors where each point represents one major. The relatedness is calculated as the proportion of matched workers (job closely related) in this major,[6] and wage inequality is the variance of log value of the residual annual earning, controlling for demographic characteristics ($var(\epsilon)$) in each major. It shows a negative correlation between job relatedness and wage inequality across majors. Moreover, to find out the correlation significance, we do a simple regression as

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

and Table 2 presents the result. It shows that the correlation is negative and significant, with values of -0.146, -0.278, and -0.157 in 1990, 2000, and 2010, respectively. Although it is not necessary to imply the causality, it suggests that job relatedness is potentially important even within majors. Similarly, Figure A.3 plots the job relatedness and wage

Table 2: Job relatedness (major) and wage inequality

VARIABLES	$var(\epsilon)$ 1990	$var(\epsilon)$ 2000	$var(\epsilon)$ 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 in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note This is the regression result of

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

where $var(\epsilon)_j$ is the residual wage inequality in major j, $relatedness_j$ is the job relatedness in major j, η_j is the residual term. Data source: National Survey of College Graduates.

[6] Figure A.2 presents the case of measuring job relatedness as the percentage of people who responds “some close” or “very close”.

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 in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note This is the regression result of

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

where $var(\epsilon)_j$ is the residual wage inequality in occupation j, $relatedness_j$ is the job relatedness in occupation j, η_j is the residual term. Data source: National Survey of College Graduates.

inequality across occupations,[7] and it also shows a negative correlation. Furthermore, Table 3 shows that the correlation is significant, with the value of -0.0685, -0.206, and -0.199 in 1990, 2000, and 2010, respectively. This also suggests that job relatedness is potentially important even within occupations.

This might be counter-intuitive at first glance as it may lead to the conclusion that a match will amplify the wage difference. However, the inequality depends on not only variety within matched or mismatched workers but also the component of these two groups. In principal, there is an inverse U-shaped relationship between inequality and mismatch degree; however, as per the US data, the inequality is found on the rising part of the curve.

3 The Model

In this section, we will present a tractable model incorporating the underlying reasons behind the educational mismatch. It only contains discrete type of variables, and a more general model with continuous type is presented in the online Appendix. A worker has skills in two dimensions (a, k), where a represents the skill level and k represents the skill type. A firm has productivity in two dimensions $(A, k^{'})$, where A represents the productivity level and $k^{'}$ represents the firm's type. To measure the mismatch, following Berliant et al. (2006), we assume 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 the skill type k and the productivity type $k^{'}$ reflects the degree of the educational mismatch. Later on, in order to make model tractable and focus

[7] Figure A.2 presents the case of measuring job relatedness as the percentage of people who responds "some close" or "very close".

on the type mismatch, we assume that productivity A and skill a are homogeneous and normalize them to 1.

Production

The joint output between a worker and a firm is linear; it depicts a firm's productivity A, worker's skill level a, work effort e, and the match premium between the worker and the firm $h(k,k^{'})$, that is, $y(a,k,k^{'}) = Aah(k,k^{'})e$. A worker takes α share of the joint output as labor compensation, that is, $w(a,k,k^{'}) = \alpha y(a,k,k^{'})$, and the promotion is modeled as an increase in α. There are different ways to model the promotion; however, as per the survey, given that the promotion and the pay are categorized together by the reason behind the mismatch (see Table A.6), it is straightforward to model it this way. In addition, the 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. The match premium decreases in the distance between job and worker's type $d(k,k^{'})$ and will be effective only if they are close enough or spread in the match-specific knowledge, as in Berliant et al. (2006). In particular, there is a cutoff δ_k such that the match premium is h_L if $d(k,k^{'}) \geq \delta_k$, and it is $h_H(> h_L)$ if $d(k,k^{'}) < \delta_k$. We assume that skill has the same knowledge spread, that is, $\delta_k = \delta$; then, formally the match premium is $h(k,k^{'}) \begin{cases} = h_L & d(k,k^{'}) \geq \delta \\ = h_H & d(k,k^{'}) < \delta \end{cases}$.

Worker

A worker's utility depends on the consumption level c and the work effort e, in particular,

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

where θ captures risk aversion, ρ captures the elasticity of effort, and τ denotes the job preference. A large τ implies high job amenity or a low disutility of effort. The distribution of preference level for a matched job is across three values $\{\tau_L, \tau_M, \tau_H\}$, where $\tau_L < \tau_M < \tau_H$, and, in a non-matched job, it is τ_M. Hence, in the matched job, the preference level could be τ_L or τ_H. After τ is realized, worker's utility maximization implies the following 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

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

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

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

Value function

For a worker who is working in a matched job (S) with preference τ, the instant utility is $U(\alpha_0, h_H, \tau)$; in this case, we simplify the notation by denoting $U(\alpha, h, \tau) = U(\alpha, 1, 1, h, \tau)$. For the next period, there is a probability of P_s that the job and worker will be separated. If they are not separated, there is a probability P_δ that the worker will be offered a matched job from outside. The job amenity in any matched job is $\tau^{'}$, 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. There is a probability $(1 - P_\alpha)$ that the promotion is α_0. In both cases, the worker chooses between accepting and rejecting the job offer. Hence, the value function can be written as

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

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

For the worker in a mismatched job with promotion α, since the match premium is h_L and the preference is τ_M the indirect utility function is $U(\alpha, h_L, \tau_M)$. For the next period, there is a probability of P_s that the job and worker will be separated. If they are not separated, there is a probability of P_δ that the worker will be offered a matched job outside the firm with a promotion level of α_0, of which the preference is $\tau^{'}$ drawn from $\{\tau_L, \tau_M, \tau_H\}$. There is a probability of $(1 - P_\delta)$ that the worker will receive an offer for a mismatched job with the same promotion level. Subsequently, the value function is as follows:

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

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

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

In the model, mismatch happens 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 happens, that is, $D(\alpha) = \{\tau : V_g(\alpha) > V_s(\tau)\}$. In this case, 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 the following two underlying reasons: preference and promotion. Subsequently, the set of promotion is $PM = \{(\alpha_1, \tau) : \tau \in D(\alpha_1), \text{and} \tau \notin D(\alpha_0)\}$ and others are due to preference.

In the model, firm's behavior is passive. There is no firm entry, 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 the occupational choice. We consider the possibility wherein an endogenous distribution might amplify the sorting between a worker and firms, which might eventually increase the wage inequality. Although we do not model it explicitly, we allow the knowledge spread change in the quantitative analysis, which will capture this effect to some extent.

4 Equilibrium

Definition

An equilibrium consists 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. In every period, workers make an 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, the employment distribution requires satisfying the following conditions:

1. The unemployed workers include unlucky job seekers and unlucky employed workers
$$N_U = N_U(1 - P_f) + (N_{g0} + N_{g1} + N_s)P_s$$

2. The workers in mismatched jobs with promotion α_0 comprise lucky job seekers, stayers, and switchers from matched jobs
$$N_{g0} = N_U P_f (1 - P_\delta) + N_{g0} Pr(g_0|g_0) + N_s Pr(g_0|s)$$
where $Pr(g_0|g_0)$ is the probability for staying in a mismatched job with promotion α_0; $Pr(g_0|s)$ is the probability for switching from a matched job

3. The workers in a mismatched job with promotion α_1 are stayers and switchers from

matched jobs

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

where $Pr(g_1|g_1)$ is the probability for staying in a mismatched job with promotion α_1; $Pr(g_1|s)$ is the probability for switching from a matched job

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

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

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

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

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

Table 4: Multiple equilibria

Equilibrium	$D(\alpha_0)$	$D(\alpha_1)$	PM	PF
Eq1	$\{\tau_L\}$	$\{\tau_L, \tau_M, \tau_H\}$	$\{(\alpha_1, \tau_M), (\alpha_1, \tau_H)\}$	$\{(\alpha_0, \tau_L), (\alpha_1, \tau_L)\}$
Eq2	$\{\tau_L\}$	$\{\tau_L, \tau_M\}$	$\{(\alpha_1, \tau_M)\}$	$\{(\alpha_0, \tau_L), (\alpha_1, \tau_L)\}$
Eq3	$\{\tau_L\}$	$\{\tau_L\}$	$\varnothing$	$\{(\alpha_0, \tau_L), (\alpha_1, \tau_L)\}$
Eq4	$\varnothing$	$\{\tau_L, \tau_M, \tau_H\}$	$\{(\alpha_1, \tau_L), (\alpha_1, \tau_M), (\alpha_1, \tau_H)\}$	$\varnothing$
Eq5	$\varnothing$	$\{\tau_L, \tau_M\}$	$\{(\alpha_1, \tau_L), (\alpha_1, \tau_M)\}$	$\varnothing$
Eq6	$\varnothing$	$\{\tau_L\}$	$\{(\alpha_1, \tau_L)\}$	$\varnothing$
Eq7	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

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. The columns "PM" and "PF" list the combination of promotion levels and amenities due to promotion and preference, respectively.

Equilibrium results

Theoretically, there are potential multiple equilibria listed in Table 4, but there are only two equilibria with non-empty sets (Eq1, Eq2). As shown in the quantitative part, only Eq1 can match the data. Hence, we will only focus on this equilibrium.[8] This equilibrium 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 as the following case. When given a promotion level of α_1, people will choose a mismatched task; however, if the promotion level is downgraded to α_0, the worker will choose a matched task. By this rule, the set for a promotion-driven mismatched worker is $(\alpha_1, \tau_M), (\alpha_1, \tau_H)$, and the set for preference is $\{(\alpha_0, \tau_L), (\alpha_1, \tau_L)\}$. The employment and wages are computed in Appendix B and are summarized as follows.

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

Employment First, N_U, N_s, N_{g0}, N_{g1} are computed by solving the equilibrium conditions. Let N_{PF}, N_{PM}, N_{SF} be the number of total employment mismatched owing to preference, promotion, and search friction, respectively; N_{sL}, N_{sM}, N_{sH} are the number of workers in a matched job with preference $\tau_L, \tau_M, and \tau_H$, respectively. They are then computed in Equation (2) through Equation (7).

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

Earnings inequality The total inequality can be decomposed into within-group inequality ($Var_j, j = g, s$) and between-group inequality ($(l\bar{n}w_j - l\bar{n}w)^2$, that is,

$$Var(lnw) = \sum_{j=g,s} \frac{N_j}{N_g + N_s}[Var_j + (l\bar{n}w_j - l\bar{n}w)^2]$$

where Var_g is the inequality within a mismatched job,

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

and Var_s is the inequality within a matched job,

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

Given all the functions in Appendix B, in the next section, we will quantify the impact of the educational mismatch on earnings inequality.

5 Quantitative analysis

In the quantitative analysis, we first calibrate the model with the data for 1990 as the benchmark. Subsequently, we re-calibrate the channel-specific and non-channel specific parameters in the model by targeting the economy in 2000. Finally, given the parameters in 1990 and 2000, we conduct several counterfactual experiments to examine the impact of each factor on wage inequality. We perform the analysis only for these two years because the wage inequality increases rapidly until 2000. As shown in Table A.1 and Table A.2, it increases from 0.23 in 1990 to 0.34 in 2000 and it turns to 0.39 in 2010.

5.1 Calibration

The following parameters need to be calibrated: the preference parameters τ_H, τ_M, τ_L, and P_H, P_M, P_L; the promotion parameters $\alpha_0, \alpha_1, P_\alpha$; the search friction parameters P_δ, P_f, P_s; the skill premium parameters h_L, h_H; the unemployment benefit $\bar{V}$; the elasticity of effort ρ; the risk aversion parameter θ; and the time discount β. First, we normalize the following parameters: match premium in a mismatched job ($h_L = 1$) and preference in a mismatched job ($\tau_M = 1$). Following the literature, we set $\theta = 2$ and $\beta = 0.95$. Other parameters are calibrated by jointly targeting several main characteristics.

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
$\bar{V}$	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 is from NSCG (1993), which collects the information for 1990; the targets are the following: 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 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), wage inequality within the unmatched group (Var_g) and total wage inequality (Var); and unemployment rate (N_U).

Table 5 presents the calibration results in 1990. The terms (α_0, α_1) are 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 the job amenity with $P_H + P_M + P_L = 1$. The term P_δ is the knowledge spread, and the terms (P_f, P_s) are the job finding and separation rates. The terms $(\bar{V}, \rho)$ are the unemployment disutility and the elasticity of effort.

The data is from NSCG (1993), and the targets are the following: the labor share (ls); the employment ratio of matched to mismatched workers ($\frac{N_s}{N_g}$) and the 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 by 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 and mismatched group (Var_s, Var_g) and the wage inequality for all the highly educated (Var); and unemployment rate (N_u). Again, the wage inequality is the residual term, that is, $var(\epsilon)$. Overall, the model matches the data quite well, particularly for inequalities Var, Var_s, 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 studies inequality through occupational choice.

The following parameter values need some explanation. First, $\alpha_1 = 0.99$ implies that the worker will take most of the output in the promoted job; this happens because, in the production function, labor is the only input. By assumption, there are only two promotion levels. Second, the large job amenity difference $\tau_H/\tau_L = 13.87/0.01$ might also be the result of the assumption that the preference is drawn from the discrete distribution. The other parameters are discussed in the next subsection.

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
$\bar{V}$	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 is taken from NSCG (2003), which collects the information for 2000. The targets are as follows: 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 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), wage inequality within the unmatched group (Var_g) and the total wage inequality (Var); and unemployment rate (N_U).

5.2 Counterfactual Analysis

To quantify the contribution of each channel, we re-calibrate both channel-specific and non-channel-specific parameters by targeting the economy in 2000. The result is shown in Table 6, and similar to the result for the period 1990, overall, the model matches the data well, particularly for inequalities and employment ratios. Several things have changed from 1990 to 2000. The difference in the promotion level (α_1/α_0) has increased from

0.99/0.34 to 0.99/0.30, and promotion probability has increased from 0.09 to0.12, and hence the change in promotion channel will potentially enlarge the wage inequality. The 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 an increase in its probability led the the preference channel to contribute negatively to the increase in wage inequality. The knowledge spread (P_δ) has slightly increased, from 0.83 to 0.87; the job finding rate has increased from 0.75 to 0.94, and the job separation rate has slightly increased from 0.04 to 0.05. These changes may increase the wage inequality by increasing the wage gap between matched and mismatched workers. Finally, we also re-calibrate two non-channel specific parameters to capture all the impact from residues, that is, anything else not being included in the model. Since there are significant changes in both $\bar{V}$ (from -576.24 to -343.73) and ρ (from -4.09 to -2.22), the impact from residues could be large.

Table 7: Counterfactual analysis: Preference

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

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with the data for 1990. The columns under "Counterfactual analysis" list the wage inequality under different counterfactual cases. The column τ_L represents the result that replaces τ_L in 1990 with the value in 2000 and retains others with the benchmark values. Similar exercises are conducted for the columns of "τ_H", "(τ_L, τ_H)", and "(P_L, P_M, P_H)", and the column of "Preference(PF)" is the result derived after replacing (τ_L, τ_H) and (P_L, P_M, P_H). In each column of the counterfactual analysis, the first row 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 a negative value implies that inequality in the counterfactual case is smaller than that of 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.

Given the parameters in 1990 and 2000, several counterfactual experiments are conducted to examine the impact of each factor on inequality. First, the parameters of preference $(\tau_L, \tau_H, P_L, P_M, P_H)$ in 1990 are replaced with the values in 2000. In Table 7, the columns under "Wage inequality" list the residual wage inequality, based on the data for 1990 and 2000 and on the benchmark model calibrated with the data for 1990. The columns under "Counterfactual analysis" list the wage inequalities under different counterfactual cases. The column τ_L represents the result of replacing τ_L and retaining others with the benchmark values. Similar exercises are conducted for the columns of "τ_H", "(τ_L, τ_H)", and "(P_L, P_M, P_H)", and the column of "Preference(PF)" is the result derived after replacing all the parameters of preference $(\tau_L, \tau_H, P_L, P_M, P_H)$. In each column of the counterfactual analysis, the first row depicts 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 the benchmark, where the negative value means that the inequality in the counterfactual case is smaller than that of 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.

As shown in Table 7, the probability distribution plays an important role in explaining the increase in wage inequality. In particular, 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 the increase in wage inequality. In addition, if the job amenity level (τ_L, τ_H) is replaced, the wage inequality change will be -55.8% of the change in the data. This result is consistent with the calibration result that, in 1990, the difference between τ_H and τ_L will be considerably larger than that in 2000. In addition, if both (τ_L, τ_H), and (P_L, P_M, P_H) are replaced, the inequality will be -46.2% of the change in the data. Therefore, overall, the preference channel contributes negatively to the increase in wage inequality.

Table 8: Counterfactual analysis: Match premium

Wage inequality			Counterfactual analysis
data (1990)	data (2000)	model (1990)	(h_H)match premium (PE)
			0.296
0.232	0.336	0.232	0.064
			0.615

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with data from 1990. The column of "(h_H)match premium(PE)" represents the result that replaces h_H in 1990 with the value in 2000 and retains others with the benchmark values. In the column of counterfactual analysis, the first row 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 means that the inequality in the counterfactual case is smaller than that of 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.

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

Table 9: Counterfactual analysis: Search friction

Wage inequality			Counterfactual analysis			
data (1990)	data (2000)	model (1990)	P_δ	P_f	P_s	search_friction(SF)
			0.238	0.232	0.238	0.244
0.232	0.336	0.232	0.006	0	0.006	0.012
			0.058	0	0.058	0.115

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with data from 1990. The column of "P_δ" represents the result that replaces P_δ in 1990 with the value in 2000 and retains others with benchmark values. Similar exercises are conducted for the columns of "P_f" and "P_s", and the column of "search_friction(SF)" is the result that is derived after replacing (P_δ, P_f, P_s). In each column of the counterfactual analysis, the first row 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 means that inequality the in counterfactual case is smaller than that of 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 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)
			0.220	0.241	0.232	0.242	0.231
0.232	0.336	0.232	-0.012	0.009	0	0.01	-0.001
			-0.115	0.087	0	0.096	-0.01

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with data from 1990. The column of P_α represents the result that replaces P_α in 1990 with the value in 2000 and retains others with the benchmark values. Similar exercises are conducted for the columns of "α_0", "α_1", and "(α_0,α_1)", and the column of "promotion(PM)" is the result derived after replacing $P_\alpha$and (α_0,α_1). In each column of counterfactual analysis, the first row 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 means that inequality in counterfactual case is smaller than that of 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 11: Counterfactual analysis: residues

Wage inequality			Counterfactual analysis		
data (1990)	data (2000)	model (1990)	$\bar{V}$	ρ	residue(RS)
			0.232	0.43	0.43
0.232	0.336	0.232	0	0.198	0.198
			0	1.904	1.904

Note The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with data from 1990. The column of $\bar{V}$ represents the result that replaces $\bar{V}$ in 1990 with the value in 2000 and retains others with the benchmark values. Similar exercises are conducted for the column of "ρ", and the column of "residue(RS)" is the result derived after replacing both $\bar{V}$ and ρ. In each column of counterfactual analysis, the first row 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 means that the inequality in the counterfactual case is smaller than that of 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.

The counterfactual exercise on the promotion channel involves the following parameters: the promotion offer probability (P_α) and the promotion levels (α_0, α_1). As shown in Table 10, the probability of promotion negatively contributes to the increase in inequality (-11.5%), the promotion level distribution contributes 9.6%, and the overall contribution of the promotion channel is -1%. In other words, there is an amplification effect when changing the probability and level at the same time.

The change in wage inequality may be caused by factors that were missed in the model. To check this, we performed similar counterfactual exercises for ($\bar{V}$,ρ) to account for the contribution from residues. Although $\bar{V}$ and ρ are the parameters of unemployment utility and elasticity on effort, respectively, they do not necessarily represent the channels of these two; instead, they represent all the factors being missed in the model. As shown in Table 11, the remaining part contributes to inequality by 190.4%; hence, the model still misses a significant portion of the explanation for the increase in inequality.

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 columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated in 1990. The columns under "Counterfactual analysis" list the wage inequality under different counterfactual cases. The first row is the wage inequality level when replacing the parameters in 1990 with those of 2000 in the counterfactual cases. The second row is 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.

To make more sense of the magnitude of each channel, we sum all the contributions from the five components—preference (PF), match premium (PE), search friction (SF), promotion (PM), and residue (RS)—and computed the ratio of each component to the sum of them. As reported in Table 12, the first row is the inequality level in the counterfactual cases. The second row is the difference in inequality between the counterfactual case and 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 sum of them. Subsequently, in this standardized measurement, the contribution of preference, match premium, search friction, and promotion are -21.4%, 28.4%, 5.3%, and -0.5%, respectively. We think the contribution is still significant, given that the wage inequality in this study controls for demographic characteristics, and the model does not include the worker skill or heterogeneity in firm productivity.

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 is from NSCG (2003), which collects the information of 2000, and the targets are the following: 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 search friction ($\frac{N_{SF}}{N_g}$); wage ratios between unmatched and matched 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 total wage inequality (Var).

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)
			0.308	0.287	0.229	0.212
0.232	0.336	0.232	0.076	0.055	-0.003	-0.02
			0.73	0.529	-0.029	-0.192

Note The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated in 1990. The columns under "Counterfactual analysis" list the wage inequality under different counterfactual cases. The first row is the wage inequality level when replacing the parameters in 1990 with those of 2000 in the counterfactual cases. The second row is 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.

6 Alternative calibration

While the counterfactual analysis in last section allows all the parameters change from 1990 to 2000, some parameters could be constant across years. To precisely identify which parameters will change, we need more detailed information, for example, the match quality, knowledge spread, etc. However, there is no such detailed information in the data. Hence, in this section, we consider about two alternative calibration strategies. In the first exercise, we keep $\bar{V}$ and ρ constant and re-calibrate other parameters. In the model, $\bar{V}$ and ρ are the parameters of unemployment utility and elasticity on effort, but as we argued in last section, it might capture all the other factors that are missed in the model, hence

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
P_α	**promotion probability**	**0.92**	$\frac{N_{PM}}{N_g}$	**0.35**	**0.86**
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
P_M	**preference probability(M)**	**0.014**	$\frac{W_{PM}}{W_s}$	**1.00**	**0.87**
P_L	**preference probability(L)**	**0.13**	$\frac{N_{PF}}{N_g}$	**0.50**	**0.13**
P_δ	**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 is from NSCG (2003), which collects the information of 2000, and the targets are the following: 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 ($\frac{N_{PF}}{N_g}$); wage ratios between unmatched and matched due to promotion ($\frac{W_{PM}}{W_s}$).

this alternative calibration might change the results significantly. In the second exercise, we only re-calibrate 5 major parameters– $h_H, P_\alpha, P_L, P_M, P_\delta$– and keep others unchanged. Note that h_H captures the match premium, which subjects to change as the job match quality change. For other three channels, we recalibrate all the parameters on probability as these are more likely to respond directly to the policy relating to educational mismatch, in particular, P_α captures the promotion probability, P_L, P_M are preference probability distribution, P_δ is the knowledge spread.

6.1 Alternative 1

In this exercise, we keep $\bar{V}$ and ρ constant and re-calibrate other parameters, and then we will do similar counterfactual analysis as in section 5. Table 13 presents the calibration result where we don't target unemployment rate (N_u) and wage inequality for unmatched group (Var_g). The main difference between the result and that in Table 6 are that τ_L and P_M are much smaller, but overall the model matched the data well. The counterfactual results are presented from Table A.10 to Table A.13. In Table 14, we present the decomposition results. It shows that match premium still explain a significant part of wage inequality increase (52.9%) and promotion channel still contribute negatively(-19.2%). However, in this case, the preference channel contributes it positively (73%) and search friction channel contributes it negatively(-2.9%), which are different from our bench mark result.

6.2 Alternative 2

In this exercise, we only re-calibrate 5 major parameters– $h_H, P_\alpha, P_L, P_M, P_\delta$– and keep others unchanged. In other words, P_α will represents all of promotion channel, P_L, P_M will represents all of preference channel, P_δ will represents all of search friction channel. Correspondingly, we only target 5 moments which are highlighted in Table 15. In this case, while the model couldn't match two employment ratio well: $\frac{N_{PM}}{N_g}$ and $\frac{N_{PF}}{N_g}$, it matched some untargeted moments well, for example, the employment ratio and wage ratio between matched and unmatched group ($\frac{w_g}{w_s}$, $\frac{N_s}{N_g}$) Counterfactual results are presented from Table A.14 to Table A.17, and the decomposition result is summarized in Table 16. It shows that match premium explains a smaller part of wage inequality increase (3.8%); and the preference also contributes it positively (31.7%); the search friction channel explain a larger part of wage inequality increase (31.7%); the promotion channel still contribute negatively but with much larger contribution (-185%).

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)
			0.265	0.236	0.265	0.04
0.232	0.336	0.232	0.033	0.004	0.033	-0.228
			0.317	0.038	0.317	-1.85

Note The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated in 1990. The columns under "Counterfactual analysis" list the wage inequality under different counterfactual cases. The first row is the wage inequality level when replacing the parameters in 1990 with those of 2000 in the counterfactual cases. The second row is 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.

7 Conclusion

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

The policy implications of this paper are as follow. First, an improvement in the education match rate will decrease wage inequality as there is a negative correlation between inequality and job relatedness. Second, as promotion, preference, and search friction are the three main reasons behind the mismatch, improving educational signaling and lowering market friction to help college graduates better utilize their knowledge could be helpful in lowering wage inequality. Third, since the match premium channel explains a significant part of the increase in wage inequality, the policy on improving match quality might automatically increase wage inequality. Fourth, this also provide channels to under the inequality in other countries, for example, China (Piketty et al. (2019),Huang (2019)).

The model could be extended to incorporate dynamics. A worker may update own preference based on the working experience, and on-the-job learning may increase the skill match. It could also be extended to include skill and productivity heterogeneity. Under these two extensions, both preference and search friction may have a higher quantitative importance. A third extension would be to turn the heterogeneity of preference and promotion level into a continuum wherein people have continuous attitudes or promotion levels.

References

Acemoglu, D., Autor, D., 2011. Skills, tasks and technologies: Implications for employment and earnings. Handbook of Labor Economics 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. The American Economic Review: Papers & Proceedings 104 (5), 387–393.

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

Autor, D. H., Katz, L. F., Kearney, M. S., 2008. Trends in us wage inequality: Revising the revisionists. The Review of Economics and Statistics 90 (2), 300–323.

Berliant, M., Reed III, R. R., Wang, P., 2006. Knowledge exchange, matching, and agglomeration. Journal of Urban Economics 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. International Economic Review 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. Economics of Education Review 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? Economic Modelling 84, 38–54.

Guironnet, J.-P., Peypoch, N., 2007. Human capital allocation and overeducation: A measure of french productivity (1987, 1999). Economic Modelling 24 (3), 398–410.

Guvenen, F., Kuruscu, B., Tanaka, S., Wiczer, D., 2020. Multidimensional skill mismatch. American Economic Journal: Macroeconomics 12 (1), 210–44.

Huang, K. X., 2019. Growth and cycles in china?s unbalanced development: Resource misallocation, debt overhang, economic inequality, and the importance of structural reforms. Frontiers of Economics in China 14 (1), 53–71.

Kambourov, G., Manovskii, I., 2009. Occupational mobility and wage inequality. The Review of Economic Studies 76 (2), 731–759.

Kirkeboen, L. J., Leuven, E., Mogstad, M., 2016. Field of study, earnings, and self-selection. The Quarterly Journal of Economics 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? The American Economic Review, 461–498.

Lemieux, T., 2014. Occupations, fields of study and returns to education. Canadian Journal of Economics 47 (4), 1047–1077.

Lise, J., Postel-Vinay, F., 2015. Multidimensional skills, sorting, and human capital accumulation. Working Paper.

Nordin, M., Persson, I., Rooth, D.-O., 2010. Education–occupation mismatch: Is there an income penalty? Economics of Education Review 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. American Economic Review.

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. Economics of Education Review 26 (4), 397–407.

Rosen, S., 1986. The theory of equalizing differences. Handbook of Labor Economics 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. The Quarterly Journal of Economics 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. The Quarterly Journal of Economics, 297–338.

Appendix

A Tables and Figures

Table A.1: Statistical description: 1990

groups		observations	tenure	earning	inequality	proportion
	Female	34467	18.17	53814.12	0.20	0.39
gender	male	59893	19.73	76401.97	0.26	0.61
	Bachelor	58063	18.77	61039.09	0.23	0.64
	Master	25757	20.17	68416.25	0.20	0.25
education	PhD	10540	18.68	104475.18	0.32	0.11
	White	79175	19.16	68444.02	0.24	0.91
	Black	9478	19.20	55785.14	0.19	0.06
race	Hispanic	5707	17.27	62247.22	0.20	0.03
	Close	55613	18.96	70803.82	0.22	0.56
	Some	24066	19.04	67695.83	0.23	0.26
relatedness	Not	14681	19.69	57021.46	0.28	0.18
all sample		94360	19.11	67514.19	0.23	1

Note: Data source is National Survey of College Graduates (1993) which has the information in 1990. Column "observations" is the number of observation in the sample; column "tenure" is the average tenure of each subgroup; "earning" is the average earning in USD in current year value; "inequality" is the residual wage inequality of each subgroup; "proportion" is the employment share of each subgroup.

Table A.2: Statistical description:2000

groups		observations	tenure	earning	inequality	proportion
	Female	21180	20.13	60732.11	0.28	0.43
gender	male	34285	21.51	90929.36	0.38	0.57
	Bachelor	31915	20.95	69964.33	0.34	0.64
	Master	16202	21.31	77226.22	0.28	0.26
education	PhD	7348	19.72	131856.72	0.49	0.10
	White	47212	21.07	79947.82	0.35	0.89
	Black	4411	20.70	61798.49	0.25	0.07
race	Hispanic	3842	18.50	65085.11	0.33	0.05
	Close	33377	20.45	83487.92	0.30	0.56
	Some	13871	21.14	77929.15	0.33	0.26
relatedness	Not	8217	22.05	61807.33	0.40	0.19
all sample		55465	20.92	78042.82	0.34	1

Note: Data source is National Survey of College Graduates (2003) which has the information in 2000. Column "observations" is the number of observation in the sample; column "tenure" is the average tenure of each subgroup; "earning" is the average earning in USD in current year value; "inequality" is the residual wage inequality of each subgroup; "proportion" is the employment share of each subgroup.

Table A.3: Statistical description: 2010

groups		observations	tenure	earning	inequality	proportion
gender	Female	26871	19.21	74392.51	0.34	0.47
	male	36619	21.22	111361.32	0.43	0.53
education	Bachelor	32722	20.51	82050.90	0.39	0.65
	Master	22924	20.09	96163.92	0.31	0.26
	PhD	7844	19.16	170399.52	0.53	0.09
race	White	51949	20.44	96302.70	0.38	0.86
	Black	4873	20.54	80507.39	0.43	0.07
	Hispanic	6668	18.16	79348.42	0.37	0.08
relatedness	Close	39174	19.92	104512.62	0.35	0.56
	Some	16218	20.54	89270.52	0.37	0.26
	Not	8098	21.01	67644.66	0.44	0.18
all sample		63490	20.28	93939.57	0.39	1

Note: Data source is National Survey of College Graduates (2013) which has the information in 2010. Column "observations" is the number of observation in the sample; column "tenure" is the average tenure of each subgroup; "earning" is the average earning in USD in current year value; "inequality" is the residual wage inequality of each subgroup; "proportion" is the employment share of each subgroup.

Table A.4: Earnings 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 second column "Var_{raw}" is the earnings inequality with raw data; the third column "Var_{res}" is the residual wage inequality as in equation (1); the fourth column "Var_1" is residual wage inequality after further controlling major dummy; and the fifth column "Var_2 " is residual wage inequality after further controlling major, occupation, and match status.

Table A.5: Proportion of match

		1990			2000			2010		
groups		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: Data source is National Survey of College Graduates (1993,2003,2013). Column "close" is the proportion of people who reported "closely related"; column "some" is the proportion of people who reported "somewhat related"; column "not" is the proportion of people who reported "not related at all".

Table A.6: Reasons for mismatch

Reason	1990	2000	2010
Pay, promotion opportunities	0.32	0.32	0.29
Working conditions [hours, equip., working envir.]	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
Job in highest degree field not available	0.16	0.14	0.18
Others	0.12	0.07	0.08

Note: Column "1990", "2000", "2010" is the percentage of people reporting different mismatch reasons among the people who reported "not related at all" for the year of 1990, 2000, 2010, respectively.

Table A.7: Reasons for mismatch: exp ≤ 10

Reasons	1990	2000	2010
Pay, promotion opportunities	0.30	0.35	0.27
Working conditions [hours, equip., working envir.]	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
Job in highest degree field not available	0.21	0.17	0.28
Others	0.12	0.05	0.09

Note: Column "1990", "2000", "2010" is the percentage of people reporting different mismatch reasons among the people who reported "not related at all" for the year of 1990, 2000, 2010, respectively. This table only includes workers with an experience of no more than 10 years.

Table A.8: Wage ratio to matched group

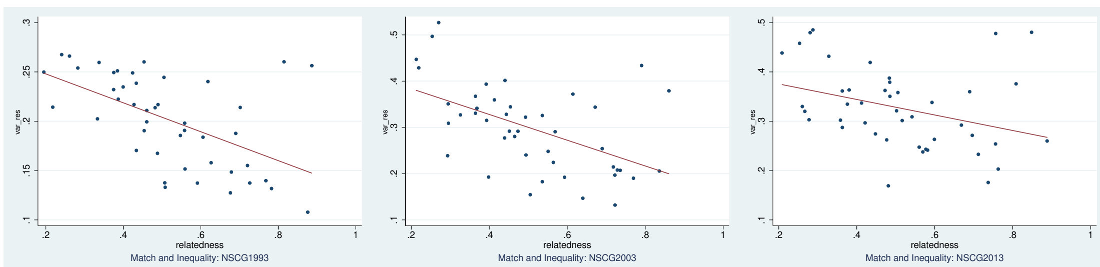
	raw data			residue		
Reasons	1990	2000	2010	1990	2000	2010
Pay, promotion opportunities	0.98	0.94	0.89	1.02	1.00	0.99
Working conditions [hours, equip., working envir.]	0.74	0.66	0.59	0.78	0.73	0.66
Job location	0.68	0.64	0.57	0.72	0.69	0.65
Change in career or professional interests	0.85	0.76	0.74	0.89	0.81	0.81
Family-related reasons	0.70	0.59	0.52	0.78	0.66	0.59
Job in highest degree field not available	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 column "raw data" presents the raw wage ratio of the mismatched group (under different reasons) to the matched group for 1990, 2000, and 2010. The column "residue" presents the residual wage ratio (after controlling for demographic characteristics) of the mismatched group to the matched group for 1990, 2000, and 2010.

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

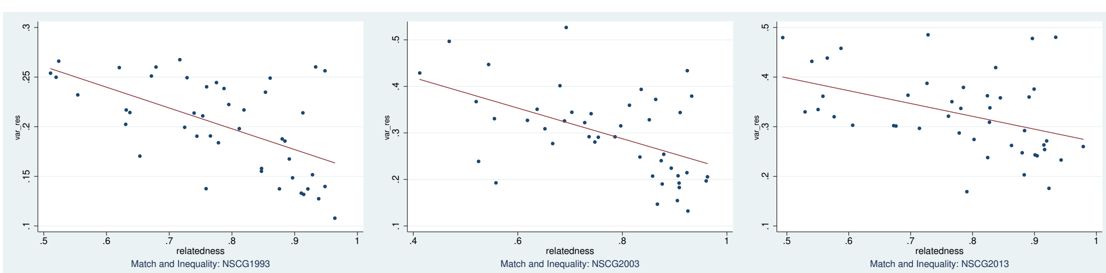
	raw data			residue		
Reasons	1990	2000	2010	1990	2000	2010
Pay, promotion opportunities	0.92	0.94	0.82	0.98	1.01	0.92
Working conditions [hours, equip., working envir.]	0.74	0.80	0.61	0.80	0.88	0.67
Job location	0.70	0.72	0.59	0.73	0.81	0.72
Change in career or professional interests	0.78	0.77	0.69	0.84	0.83	0.80
Family-related reasons	0.76	0.56	0.70	0.81	0.64	0.81
Job in highest degree field not available	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 column "raw data" presents the raw wage ratio of the mismatched group (under different reasons) to the matched group for 1990, 2000, and 2010. The column "residue" presents the residual wage ratio (after controlling for demographic characteristics) of the mismatched group to the matched group for 1990, 2000, and 2010. This table only includes workers with an experience of no more than 10 years.



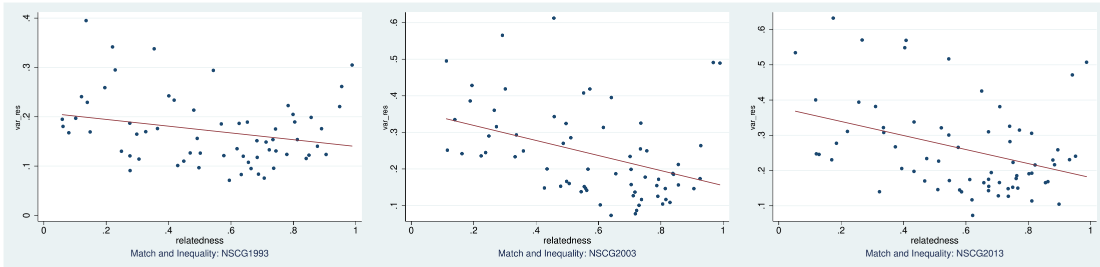
Note: This figure shows the correlation between major relatedness and within-major inequality. The relatedness is defined as the percentage of people who responds "very close". In each panel, a dot represents a major, the x-axis is the relatedness in that major, and the y-axis is the residual wage inequality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

Figure A.1: Job relatedness (major) and wage inequality



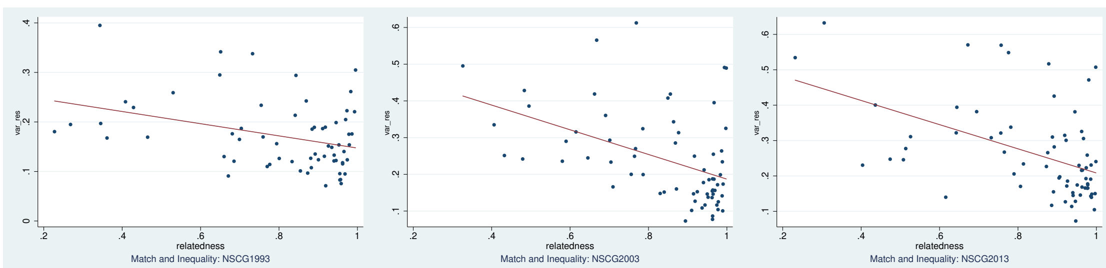
Note: This figure shows the correlation between major relatedness and within-major inequality. The relatedness here is defined as the percentage of people who responds "some close" or "very close". In each panel, a dot represents a major, the x-axis is the relatedness in that major, and the y-axis is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

Figure A.2: Job relatedness (major) and wage inequality



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

Figure A.3: Job relatedness (occupation) and wage inequality



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

Figure A.4: Job relatedness (occupation) and wage inequality

Table A.10: Alternative 1: Preference

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

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with the data for 1990. The columns under "Counterfactual analysis" list the wage inequality under different counterfactual cases. The column τ_L represents the result that replaces τ_L in 1990 with the value in 2000 and retains others with the benchmark values. Similar exercises are conducted for the columns of "τ_H", "(τ_L, τ_H)", and "(P_L, P_M, P_H)", and the column of "Preference(PF)" is the result derived after replacing (τ_L, τ_H) and (P_L, P_M, P_H). In each column of the counterfactual analysis, the first row 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 a negative value implies that inequality in the counterfactual case is smaller than that of 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 (1990)	data (2000)	model (1990)	(h_H)match premium (PE)
			0.287
0.232	0.336	0.232	0.055
			0.529

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with data from 1990. The column of "(h_H)match premium(PE)" represents the result that replaces h_H in 1990 with the value in 2000 and retains others with the benchmark values. In the column of counterfactual analysis, the first row 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 means that the inequality in the counterfactual case is smaller than that of 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 (1990)	data (2000)	model (1990)	P_δ	P_f	P_s	search_friction(SF)
			0.221	0.232	0.238	0.229
0.232	0.336	0.232	-0.011	0	0.006	-0.003
			-0.106	0	0.058	-0.029

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with data from 1990. The column of "P_δ" represents the result that replaces P_δ in 1990 with the value in 2000 and retains others with benchmark values. Similar exercises are conducted for the columns of "P_f" and "P_s", and the column of "search_friction(SF)" is the result that is derived after replacing (P_δ, P_f, P_s). In each column of the counterfactual analysis, the first row 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 means that inequality the in counterfactual case is smaller than that of 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 (1990)	data (2000)	model (1990)	P_α	α_0	α_1	(α_0, α_1)	promotion(PM)
			0.222	0.236	0.222	0.223	0.212
0.232	0.336	0.232	-0.01	0.004	-0.01	-0.009	-0.02
			-0.096	0.038	-0.096	-0.087	-0.192

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with data from 1990. The column of P_α represents the result that replaces P_α in 1990 with the value in 2000 and retains others with the benchmark values. Similar exercises are conducted for the columns of "α_0", "α_1", and "(α_0,α_1)", and the column of "promotion(PM)" is the result derived after replacing $P_\alpha$and (α_0,α_1). In each column of counterfactual analysis, the first row 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 means that inequality in counterfactual case is smaller than that of 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		
(1990)	(2000)	(1990)	(P_L, P_M, P_H)	Preference(PF)
			0.131	0.131
			-0.101	-0.101
0.232	0.336	0.232	-0.971	-0.971

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with the data for 1990. The columns under "Counterfactual analysis" list the wage inequality under different counterfactual cases. The column τ_L represents the result that replaces τ_L in 1990 with the value in 2000 and retains others with the benchmark values. Similar exercises are conducted for the columns of "τ_H", "(τ_L, τ_H)", and "(P_L, P_M, P_H)", and the column of "Preference(PF)" is the result derived after replacing (τ_L, τ_H) and (P_L, P_M, P_H). In each column of the counterfactual analysis, the first row 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 a negative value implies that inequality in the counterfactual case is smaller than that of 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	
(1990)	(2000)	(1990)	(h_H)match premium (PE)
			0.236
0.232	0.336	0.232	0.004
			0.038

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with data from 1990. The column of "(h_H)match premium(PE)" represents the result that replaces h_H in 1990 with the value in 2000 and retains others with the benchmark values. In the column of counterfactual analysis, the first row 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 means that the inequality in the counterfactual case is smaller than that of 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 (1990)	data (2000)	model (1990)	P_δ	search_friction(SF)
			0.265	0.265
0.232	0.336	0.232	0.033	0.033
			0.317	0.317

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with data from 1990. The column of "P_δ" represents the result that replaces P_δ in 1990 with the value in 2000 and retains others with benchmark values. Similar exercises are conducted for the columns of "P_f" and "P_s", and the column of "search_friction(SF)" is the result that is derived after replacing (P_δ, P_f, P_s). In each column of the counterfactual analysis, the first row 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 means that inequality the in counterfactual case is smaller than that of 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 (1990)	data (2000)	model (1990)	P_α	promotion(PM)
			0.04	0.04
0.232	0.336	0.232	-0.228	-0.228
			-1.85	-1.85

Note: The columns under "Wage inequality" list the inequality from 1990, 2000, and from the benchmark that is calibrated with data from 1990. The column of P_α represents the result that replaces P_α in 1990 with the value in 2000 and retains others with the benchmark values. Similar exercises are conducted for the columns of "α_0", "α_1", and "(α_0,α_1)", and the column of "promotion(PM)" is the result derived after replacing $P_\alpha$and (α_0,α_1). In each column of counterfactual analysis, the first row 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 means that inequality in counterfactual case is smaller than that of 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 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, N_{g1} is the employment in mismatched job with promotion level α_1. Every period workers make occupational choice based on current status (α, τ, h) to maximize the expected utility $\{V_U, V_s(\tau), V_g(\alpha)\}$, and in the stationary equilibrium(Eq1), the 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 α_0 come from lucky job seekers, stayers, and switchers from matched jobs due to low amenity.
$$N_{g0} = N_U P_f(1 - P_\delta) + N_{g0}(1 - P_s)[(1 - P_\delta) + P_\delta P_L] + N_s(1 - P_s)(1 - P_\delta)(1 - P_\alpha)P_L$$

3. The workers in mismatched job with promotion α_1 are stayers and switchers from matched jobs due to promotion.
$$N_{g1} = N_{g1}(1 - P_s) + N_s(1 - P_s)(1 - P_\delta)P_\alpha$$

4. The workers in matched job are lucky job seekers, switchers from mismatched jobs, and stayers.
$$N_s = N_U P_f P_\delta + N_{g0}(1 - P_s)P_\delta(P_H + P_M) + N_s(1 - P_s)[P_\delta + (1 - P_\delta)(1 - P_\alpha)(P_H + P_M)]$$

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}, N_{SF} be the mismatched employment due to preference, promotion, and search friction, respectively. N_{sL}, N_{sM}, N_{sH} are the number of workers in matched job with preference of τ_L, τ_M, τ_H, respectively. Employment in an mismatched job that comes from stayers with different promotion levels and switchers from matched jobs due to preference is

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

Employment in an mismatched job that comes from stayers and switchers from matched jobs due to promotion is

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

Employment in an mismatched job that comes from lucky job seekers receiving offers of mismatched jobs and stayers with different promotion levels who did not receive an offer of a matched job due to search friction is

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

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

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

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

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

Employment in matched jobs with high job amenity that comes from lucky job seekers, switchers from mismatched jobs, and stayers is

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

Wages Let w_{PF}, w_{PM}, w_{SF} be the wage of mismatched workers due to preference, promotion, and search friction, respectively; w_{sL}, w_{sM}, w_{sH} are the wages of matched workers with the preference τ_L, τ_M, τ_H, respectively; w_g, 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 among workers with different promotion levels.

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

Similarly, the wage in mismatched jobs due to search friction is the average wage among workers of different promotion levels.

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

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.

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

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