Macroeconomic Dynamics (2025), **29**, e158, pp. 1–34
doi:10.1017/S1365100525100631

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Within-job wage inequality: the nexus between performance pay and match quality

Rongsheng Tang[1], Yang Tang[2], and Ping Wang[3,4]

[1]Li Anmin Institute of Economic Research, Liaoning University, Shenyang, China
[2]Nanyang Technological University, Singapore
[3]Washington University in St. Louis, St. Louis, USA
[4]NBER, Cambridge, MA, USA
Corresponding author: Rongsheng Tang; Email: rstang@outlook.com

Abstract

Over recent decades, we find about 80% of widening residual wage inequality to be within jobs (industry-occupation pairs). To explore the underlying drivers, we incorporate into a sorting equilibrium framework with two extensive margin channels (across-job sorting and within-job selection of a performance-pay position) and an intensive margin channel (quality of skill match), in addition to residual job productivity. We show that equilibrium sorting is positively assortative both within and across jobs. By calibrating the model to the United States in 1990 and 2000, we find the improved match quality and rising performance-pay incidence amplify each other, jointly accounting for about 90% of the widening within-job wage inequality. Match quality and performance pay are particularly important in jobs with rising average wages and expansionary employment. Once performance pay and match quality channels are incorporated, job sorting becomes less important and residual job productivity becomes inconsequential throughout.

Keywords: Within-job wage inequality; performance pay; match quality; positive assortative job sorting

JEL classifications: E24; O15; O40

1. Introduction

Income inequality has been at the center of the macroeconomic platform over the past few decades. It has been extensively documented that since the 1980s the US residual wage inequality accounts for a major proportion of its overall wage inequality and tends to increase faster among individuals with higher education (having some college or above).[1] In this paper, we highlight two key stylized facts:

[Fact 1] Over 70% of overall inequality among the highly educated has been due to the residual inequality and about four-fifths of the residual wage inequality driven by wage dispersion within jobs, defined by industry–occupation pairs.[2]

[Fact 2] Over the period of 1990–2018, the rise in the residual wage inequality of the highly educated has been particularly prominent during the early phase of 1993–2001.[3]

Despite its important roles played in rising income inequality, little is known toward understanding the underlying drivers of within-job residual wage inequality, particularly so among the highly educated since the mid-1990s. To address this issue, we construct a novel sorting equilibrium model to explore theoretically several potentially important channels, both the intensive and the extensive margins, and then to quantify their contributions over the decades of 1990 and 2000 with the sharpest rising inequality.

© The Author(s), 2025. Published by Cambridge University Press.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

In the higher education group, even if we control for more job characteristics including industry and occupation, about 90% of the residual wage inequality remains. By decomposing the residual wage inequality into the between-job and within-job components over all the industry—occupation pairs, we find that the within-job wage inequality accounts for about 80% of the residual wage inequality between 1990 and 2000. This suggests that the wage inequality among individuals with higher education is primarily driven by the within-industry/occupation inequality. To explain the aforementioned facts, we construct a sorting equilibrium model where, in addition to different job productivity levels, we consider (i) two extensive margin channels—workers *optimally self-select into different jobs and pay positions* (performance-pay versus fixed-pay positions) and (ii) an intensive channel—the *quality of skill match*. The chief purpose of the model is to investigate whether these margins may be important to explain the two facts on within-job inequality we established.

Notably, the performance-pay incidence and match quality margins have direct data measurements. Yet, the sorting margin needs to be backed out upon calibrating the model—because this margin is between jobs, the calibration in general equilibrium is nontrivial due to computation of all job sorting wage cutoffs with the range of wages of each job all endogenously determined.

Workers in the performance-pay position are paid according to how much they contribute, and such payments usually include bonuses, commission, piece rates, and tips. The counterpart to this is the payment of a fixed hourly wage. While the literature has identified a positive wage effect of performance-pay, for example, by Lemieux et al. (2009), we further examine the relation between the within-job wage inequality and the performance-pay incidence and find a significant positive relationship: *jobs with higher performance-pay incidence usually have higher wage inequality*. This hints at the potential role of the rising performance-pay incidence for widening wage dispersion. We measure job match quality based on the inverse relationship with skill mismatch. We compute the skill mismatch index for each job using workers' occupation-relevant skill and ability measures from the 1979 National Longitudinal Survey of Youth (NLSY79) and occupation skill requirements from O*NET. Previous work such as Guvenen et al. (2020) identify a negative wage effect of skill mismatch whereby better job match quality is associated with higher compensation, we further show a negative relationship between the skill mismatch index and within-job wage inequality: *jobs with lower match quality are paid more equally*. Hence, improved match quality may also serve to explain the within-job wage inequality. By conducting regression analysis, we find that increases in performance-pay incidence or match quality raise wage inequality significantly. While the empirical analysis suggests a plausible relationship between the wage inequality of each job and the job characteristics of interests in partial equilibrium, the result might be different in general equilibrium and when interactions between different channels are allowed. In addition, the regression results are silent on the counterfactual scenario, thus limiting the ability to conduct rich policy analysis. This thereby requires a deep structure to discipline the interactions among them.

In our sorting equilibrium framework, workers are heterogeneous in their innate abilities. Each job is associated with different productivity and two payment schemes: performance pay and fixed pay. In a fixed-pay position, a worker earns a pooled wage independent of his/her ability, effort or match quality. In the performance-pay position, a worker's pay positively depends on his/her contribution to production. A worker in a better-matched job is captured by drawing a higher productivity premium from matching, which leads to a higher wage payment. Higher expected match quality thus plays an important role in position selection and job sorting.

Job-specific disutility is incurred for workers in a performance-pay position to capture monitoring costs. Match quality and the resulting productivity-premium also vary by job. In the sorting equilibrium, workers optimally choose their jobs and positions. Under proper assumptions, we show that equilibrium sorting is positive assortative, with the less talented workers choosing

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

the fixed-pay position and the more talented workers selecting the performance-pay position at different jobs based on job productivity and the expected returns dependent on matching premium.

The within-job wage inequality can then be decomposed into that within the performance-pay position and the differences in the average wage between the performance-pay and the fixed-pay positions. In general, performance-pay income is more dispersed than fixed-pay income. Analytically, in a one-job case, we can establish a positive relationship between the performance-pay incidence and within-job wage inequality when the share of the educated workers in the performance-pay position is not too large. This case is equivalent to a multi-job one under a given sorting outcome. A better technology or higher match quality is shown to amplify the effect of the performance-pay incidence on the within-job and the within-performance-pay inequality but dampens its effect on the within-job and between-position inequality, thereby leading to an ambiguous net impact overall. The relative magnitudes of these effects depend crucially on the degree of complementarity between workers in different positions and their relative employment size. With endogenous sorting across jobs, the resulting cutoff abilities further interact with performance pay and match quality in a complex manner. As a consequence, how various channels affect the within-job wage inequality becomes a quantitative question.

To quantify the roles of the performance-pay incidence, match quality, job productivity, and job sorting played in rising wage inequality, we calibrate the model by matching several of the key job-specific features of wage inequality and employment in the US economy in 2000. To conduct counterfactual–based decomposition exercises, we restore the value of each job-specific factor in 2000 to the corresponding value in 1990 by perturbing the distribution of each variation, while maintaining all the other parameters at their benchmark values. The results indicate that our model can capture most of the changes in within-job wage inequality from 1990 to 2000, with an essentially zero residual component. Our decomposition analysis suggests that changes in performance pay and match quality alone account for around 90% of widening within-job wage inequality. Moreover, we find amplifying interactive effect compared with the situation if we either let the performance-pay incidence or match quality take their values in 1990. While sorting makes a modest contribution to the widening inequality, the role of residual job productivity is inconsequential. Since performance pay and match quality are the main drivers, pro-active policy intervention aiming to reduce inequality may be at the expense of less prevalence of performance pay or skill mismatch.

Intuitively, once match quality is assured and workers are offered with correct incentives by performance pay with better ones selecting performance pay positions, the role played by positive assortative sorting across jobs must diminish as a consequence. Ex ante one cannot tell which type of sorting is strong and the existing literature is silent about this. In this paper, we provide a thorough quantitative contrast between them. Moreover, job-specific productivity alone does not convert to higher output—it must be combined with workers' ability and the quality of match. Thus, to some degree, it resembles the Solow residual in neoclassical aggregate production—departing from the neoclassical canonical framework; however, within- and cross-job sorting is in play in our paper. The results indicate that upon incorporating performance pay and match quality, residual productivity adds little explanatory power to underlying drivers in the model—a finding granting the paper as an endogenous TFP story.

We also perform similar decomposition exercises based upon, respectively, changes in the rankings of average wages and employment shares from 1990 to 2000. The lion's share of the contribution (more than 85%) to widening within-job inequality comes from jobs whose average earnings rankings change moderately (by 1 or 2) or remain unchanged, where both performance pay and match quality play comparably important roles. For jobs whose employment share rankings remain unchanged, which make up more than half of the changes in the overall within-job inequality, match quality is found to play a greater role than the performance-pay incidence.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

The main takeaway of our paper is that while the within-job wage inequality accounts for about 80% of the residual wage inequality between 1990 and 2000, approximately 90% of such rising dispersion is a consequence of higher performance-pay incidence and improved match quality. Once the extensive margin via performance pay and the intensive margin via match quality are incorporated, job sorting becomes much less important, while residual job productivity is inconsequential throughout for the within-job wage inequality. Policymakers concerning earned income inequality should thus be careful to account for the equity-efficiency tradeoff because the two primary sources of wage dispersions within jobs—performance pay and match quality—are natural in the eyes of production efficiency.

1.1 Related literature

Several studies have documented the trend of rising wage inequality since the 1970s (e.g., Piketty and Saez, 2003; Acemoglu and Autor, 2011). Although one convincing theory claims that skill–biased technology change can account for the rising skill premium (e.g., Juhn et al., 1993; Krusell et al., 2000; Acemoglu, 2003), Altonji et al. (2016) argue that earnings differences among college students across majors can be larger than the skill premium between college and high school. In general, the residual wage inequality accounts for most of the overall increased wage inequality (e.g., Katz and Autor, 1999; Lemieux, 2006).

Recent studies focus on within-industry or within-firm wage inequality. For example, Barth et al. (2011) emphasize the role of plant differences within industries and argue that this could explain two-thirds of wage inequality in the United States, whereas Card et al. (2013) find that plant heterogeneity and assortativeness between plants and workers explain a large part of the increase in wage inequality in West Germany. Additionally, Mueller et al. (2017) study the skill premium within firms and find that firm growth has increased wage inequality. While Papageorgiou (2011) highlights the labor markets within firms and concludes that the within-firm part might explain one-eighth to one-third of the increase in wage inequality, Song et al. (2019) argue that the between-firm component can explain two-third of the rise in wage inequality. To compare our results with theirs, we note that "within industry" includes both within-firm and between-firm, while "within firm" includes both within-occupation and between-occupation. Thus, there are intersections between within-firm and within industry-occupation, and the between-firm part appears to be the main reason for rising within industry-occupation inequality in our exercise.

There are also a growing number of studies that examine wage inequality between and within occupation (see, e.g., Erosa et al. 2024). While Kambourov and Manovskii (2009) argue that the variability of productivity shocks in occupations coupled with endogenous occupational mobility could account for most of the increase in within-group wage inequality between the 1970s and mid-1990s, Scotese (2012) shows that changes in wage dispersion within occupations are quantitatively as important as wage changes between occupations for explaining wage inequality between 1980 and 2000. In a more related study, Akerman et al. (2013) use Sweden employee–employer data to decompose the residual wage inequality and find that within the sector-occupation component (which we call within jobs) it accounts for a large proportion for both the level and change from 2001 to 2007. We confirm that a similar trend is found using the US data.

Performance pay The literature on performance pay either studies incentives and productivity (e.g., Jensen and Murphy, 1990; Lazear, 2000) or explains the racial or gender wage gap through the different performance-pay rates by race or gender (e.g., Heywood and Parent, 2012; Heywood and Parent, 2017). Makridis (2018) finds a positive relationship between the performance-pay incidence and wage premium and shows that longer working hours lead to more on-the-job human capital investment. While his study focuses on the dynamic effects of performance pay on human capital accumulation, our study emphasizes the *interactions among performance pay,*

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

match quality and sorting within and between jobs. In addition, Makridis and Gittleman (2022) find that performance-pay jobs adjust compensation more on the intensive margin, for which we concur but our incorporation of the aforementioned interactions produces richer results. Another relevant study is Lemieux et al. (2009), who suggest performance pay as a channel through which the underlying changes in returns to skills are translated into higher wage inequality. They conclude that wages in performance-pay jobs are more sensitive to abilities and that inequality is much higher. While the within-job selection in their work is a partial equilibrium outcome, our *general equilibrium* framework adds two important features: (i) the wages in the non-performance-pay position affect selections in all the jobs. (ii) sorting across jobs affects within-job sorting. In addition, while they emphasize the monitoring cost on selection, we highlight *match quality* as another driver affecting sorting both *within and between* jobs. Furthermore, Wallskog et al. (2024) identifies empirically that the increase of within-firm inequality is mainly driven by adoption of aggressive performance-pay bonus and management schemes, thereby lending support to the channels proposed in our theory.

Quality of skill match The general idea of skill mismatch is that workers that share the same characteristics might have different productivity from the job or machine in or on which they are working. Violante (2002) provides a channel through vintage capital to decompose the residual wage inequality into a worker's ability dispersion, a machine's productivity dispersion, and the correlation between them. He argues that this channel could explain most transitory wage inequality and 30% of residual wage inequality. Jovanovic (2014) builds a learning–by–doing model to emphasize the role of matching between employees and employers. Under such a framework, he discusses the roles of improving signal quality and assignment efficiency. Both studies are remotely related due to very different model settings and underlying issues of investigation. Concerning the measurement of match quality, there are two general two approaches: (i) the first is to measure the distance between skill requirement and acquirement based on the scores of skills from NLSY79 and O*NET (e.g., Sanders, 2014; Guvenen et al., 2020; Lise and Postel-Vinay, 2020); (ii) the second is to measure job relatedness between the field of study in the highest degree and current occupation (e.g., Robst, 2007; Arcidiacono, 2004; Ritter and West, 2014; Kirkeboen et al., 2016). Since the second measurement relies on subjective responses that might be biased depending on unobservable individual characteristics, we elect to use the first approach to construct a skill mismatch index following Guvenen et al. (2020).[4] Finally, Bandiera et al. (2024) study mismatch using internationally comparable microdata. Although the main focus is on meritocracy and income difference across countries, one result in this paper that increase technology, which increases the match return, will increase the wage inequality is consistent with our results that high match probability or quality will increase inequality.

2. Stylized facts

Here, we document several stylized facts on performance-pay incidence, skill mismatch, and wage inequality. First, we show how the performance-pay incidence and skill mismatch index are estimated. Second, we compute wage inequality using different measurements, and decompose it into the between- and within-job components. Third, we examine the relationships between the performance-pay incidence, skill mismatch, and wage inequality across jobs.

Our data are collected from several sources, including the March Current Population Survey (March CPS), Panel Study of Income Dynamics (PSID), 1979 National Longitudinal Survey NLSY79, and O*NET. The March CPS includes the most extended high-frequency data series enumerating labor force participation and earnings in the US economy. The PSID contains detailed information on earnings, including commission, bonuses, piece rates, and tips. In NLSY79, respondents were given an occupational placement test—the Armed Services Vocational Aptitude

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Battery (ASVAB)—that provides detailed measures of occupation-relevant skills and abilities. In addition, respondents reported various measures of non-cognitive skills, which can reflect a worker's ability for socially interactive work. Finally, O*NET provides the skill requirements for each occupation.

2.1 Estimation of the performance-pay incidence and skill mismatch index

Industry/occupation classification The Center for Economic Policy Research (CEPR) provides two- and three-digit occupation and industry codes. However, the classification is inconsistent from 1983 to 2018. To address this, we build consistent one-digit industry and occupation codes as in Lemieux et al. (2009). Meanwhile, we build a consistent three-digit code following Dorn (2009). Finally, the same method is used to group a consistent two-digit code.[5] To circumvent the mis-classification problem over different years due to changes in the content of either occupation or industry, and the problem of empty cells, we use the one-digit industry and occupation codes and check the inclusion of the two-digit occupation code when there is a potential concern.

Performance-pay incidence Performance pay includes bonuses, commission, and piece-rate payment. A significant challenge is to identify workers in a performance-pay position. The PSID, which has been extensively discussed in the literature, reports the format of the payment that a worker has received in a given year among the three performance pay components. However, for workers without those payments, we cannot distinguish whether they work in fixed-pay positions or do not merit, say, bonuses in that given period. Fortunately, the longitudinal nature of PSID data allows us to track the payment history to examine whether a worker has received any form of performance pay in their current job, providing a much more accurate measure (see Lemieux et al., 2009).[6]

Following Lemieux et al. (2009), we estimate performance-pay incidence over time using a linear probability model where we regress the dummy of the performance pay on job (industry by occupation) dummy. Specifically, we construct a performance-pay indicator variable looking at whether a part of a worker's total compensation includes a variable pay component (bonus, commission, or piece-rate) during the sample period. Naturally, conditional on the job duration, we tend to observe a given job match fewer times at the two ends of the sample period than in the middle of the sample. To solve this end-point problem, we follow their solution involving re-balancing the sample using regression or other methods. In practice, we adjust the measures of the performance-pay incidence over time by estimating a linear probability model in which dummies for calendar years and the number of times the job match is observed are included as regressor. We then compute an adjusted measure of the performance pay incidence by using the distribution of the number of times the job match is observed to its average value for the years 1982 to 2000, which are relatively unaffected by the end-point problem. Finally, we keep both less and highly educated workers in PSID for two reasons. First, the sample size would become too small to run the regression if we only retain highly educated workers. Second, we treat the performance-pay incidence as one of the job characteristics so that it would be better to retain the full sample.[7]

We estimate the performance-pay incidence for 80 jobs separately for the years up to 1990 and 2000. To have a consistent code with skill mismatch, we also group them into 25 jobs. Tables 1 and 2 present the results for 25 jobs and show that performance-pay incidence has increased from 1990 to 2000 in general. The results are similar for 80 jobs, as shown in Tables A.9 and A.10.

Skill mismatch index We approximate the quality of skill match by the inverse of the skill mismatch index. We compute the skill mismatch index using NLSY79 and O*NET following Guvenen et al. (2020). NLSY79 tracks a nationally representative sample of individuals aged between 14 and 22 years on January 1, 1979. It contains detailed information on the industry and occupation in which each individual worked. All respondents took the ASVAB test at the

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 1. Incidence of performance-pay for 25 jobs: 1990

	Professional	Manager	Sales	Production	Clerical
Goods	0.40	0.55	0.75	0.23	0.23
Transport	0.11	0.31	0.75	0.23	0.21
FIRE	0.65	0.66	0.85	0.13	0.32
Business	0.36	0.50	0.84	0.38	0.25
Trade	0.41	0.58	0.75	0.34	0.41

Notes: This table presents the estimated performance-pay incidence for the highly educated for 25 jobs for the year 1990, following the approach in Lemieux et al. (2009). Data source: PSID.

Table 2. Incidence of performance-pay for 25 jobs: 2000

	Professional	Manager	Sales	Production	Clerical
Goods	0.49	0.58	0.77	0.28	0.31
Transport	0.23	0.53	0.82	0.32	0.27
FIRE	0.74	0.72	0.87	0.24	0.37
Business	0.41	0.57	0.72	0.43	0.30
Trade	0.41	0.60	0.75	0.38	0.41

Notes: This table presents the estimated performance pay incidence for the highly educated for 25 jobs for the year 2000, following the way in Lemieux et al. (2009). Data source: PSID.

beginning of the survey. The respondents were also given a behavioral test to elicit their social attitudes (e.g., self-esteem and willingness to engage with others). The ASVAB version taken by NLSY79 respondents had 10 component tests. We focus on the following 4 components of verbal and math abilities, which can be linked to skill counterparts: Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, and Mathematics Knowledge. The NLSY79 included three attitudinal scales that measure a respondent's non-cognitive abilities. We use two of these measures as social abilities: the Rotter locus of control and Rosenberg self-esteem scales. Because age differences can affect scores, we equalized the mean and variance of each test score across ages. Similar to PSID, we retain the full sample to ensure that we have a sufficient sample size and capture the job's characteristics.

O*NET includes information on 974 occupations, which can be mapped to the 292 occupation categories included in NLSY79. For each occupation, O*NET analysts score the importance of the 277 descriptors. We use the 26 descriptors most related to the ASVAB component tests and another 6 descriptors related to social skills.

We aggregate information about workers' abilities and occupational skill requirements into verbal, math, and social ability following Guvenen et al. (2020). We restate them as follows. First, we convert the O*NET skills into four ASVAB test categories using the relatedness score created by the Defense Manpower Data Center. For each ASVAB category test, we create an O*NET analog by summing the 26 descriptors and weighting them by this relatedness score. Therefore, each occupation yields a set of scores comparable to the ASVAB categories, each a weighted average of the 26 original O*NET descriptors.

Second, we normalize each dimension's standard deviation to be 1 and reduce these 4 ASVAB categories into 2 composite dimensions—verbal and math—by applying principal component analysis. Similarly, we normalize each dimension of the 6 O*NET descriptors related to social skills standard deviation to be 1 and reduce them to a single dimension.[8] We convert all six scores (verbal, math, and social worker abilities, and verbal, math, and social occupation requirements) into percentile ranks among individuals or occupations.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 3. Skill mismatch index in 1990

	Professional	Manager	Sales	Production	Clerical	mm_ind
Goods	1.549	0.918	1.557	1.853	2.062	1.570
Transport	1.747	1.576	1.139	1.564	1.522	1.557
FIRE	1.255	1.380	1.355	1.215	1.242	1.329
Business	1.504	1.240	1.475	1.727	2.095	1.523
Trade	0.807	1.324	1.373	1.847	1.987	1.532
mm_occ	1.504	1.244	1.414	1.782	1.875	

Notes: This table presents the average skill mismatch for 25 jobs for the year 1990. "mm_occ" and "mm_ind" represent the average skill mismatch within occupation and industry, respectively. Data source: NLSY79 and ONET, and authors' calculation.

Table 4. Skill mismat3h index in 2000

	Professional	Manager	Sales	Production	Clerical	mm_ind
Goods	1.532	1.079	1.107	1.902	2.043	1.511
Transport	1.629	1.277	1.984	1.626	1.363	1.516
FIRE	1.258	1.055	1.504	2.533	1.508	1.280
Business	1.466	1.191	2.118	1.882	1.790	1.431
Trade	1.565	1.278	1.085	1.671	2.041	1.408
mm_occ	1.499	1.173	1.268	1.812	1.807	

Notes: This table presents the average skill mismatch for 25 jobs for the year 1990. "mm_occ" and "mm_ind" represent the average skill mismatch within occupation and industry, respectively. Data source: NLSY79 and ONET, and authors' calculation.

Third, we compute the individual-level mismatch index and then aggregate them into 25 jobs in 1990 and 2000. Tables 3 and 4 show that skill mismatch has decreased in general from 1990 to 2000, that is, the match quality has increased.

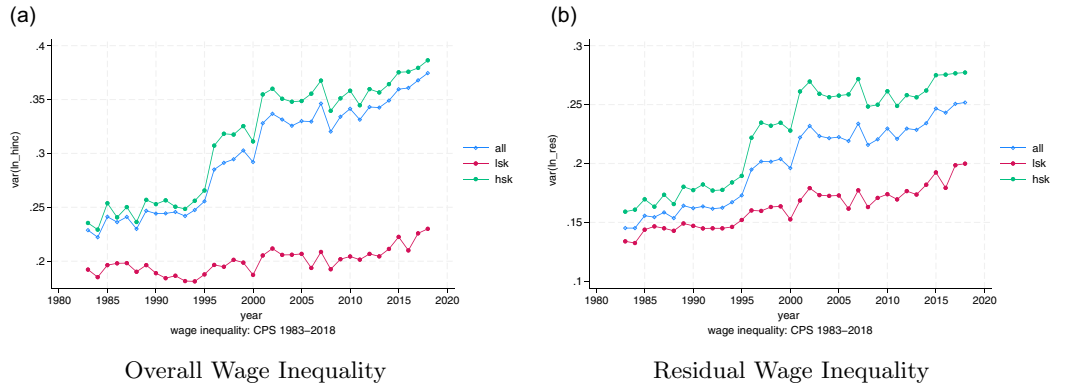
For robustness check, we also measure the skill mismatch in two alternative ways. First, the current method using total 26 skills in cognitive skills, while as an alternative, one use 6 skills as cognitive skills, representing the minimal level in this category. Figure A.1 presents the comparison in 1990 (left) and 2000 (right) for 2-digit code occupation. It shows that these two measurements are closely related, despite that the values in alternative are slightly higher than the benchmark ones. Second, we measure mismatch by a subjective response on the relatedness between college major and occupation using data from National Survey of College Graduate, then a high relatedness should reflect low skill mismatch.[9] Figure A.2 compares skill mismatch with the job relatedness and shows a clear-cut negative correlation as expected.

2.2 Measuring wage inequality

We compute the wage inequality using the March CPS where only full-time and full-year workers, defined as those working for at least 40 weeks in a year and 35 hours in a week, and aged between 16 and 65 years. Wages are defined as real hourly earnings. We drop hourly earnings below half the minimum wage in 1982 dollars or higher than USD 1000. The top-coding wage is low in some studies; for example, it is 100 (in 1979 dollars) in Lemieux et al. (2009) and around 180 in Acemoglu and Autor (2011). Because we focus on highly educated individuals, we do not trim the top incomes by as much.

In the March CPS, the education level is grouped into six categories: primary, high school dropout, high school graduate, some college, college graduate, and post-college. The implied years of schooling are 6, 9, 12, 14, 16, and 18, respectively. Potential experience is then computed as the difference between the year after graduation and age.[10] The highly educated group includes

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press



Overall Wage Inequality Residual Wage Inequality

Figure 1. Wage inequality by education group.
Notes: In the left (right) panel, the inequality is measured as the variance in the log value of hourly wages (residual wages). In both panels, the blue line represents the inequality for the whole sample (all) and the green line only includes those highly educated or high skilled (hsk). The red line is for the low education group or low skilled (lsk). Data source: March CPS from the CEPR.

workers who have some college education and above; the proportion of this group increased from 45% in 1983 to 66% in 2018.

We measure wage inequality as the variance in the logarithm of wage. The left panel of Figure 1 presents the pattern of wage inequality by education group from 1983 to 2018. In general, wage inequality has been increasing since the 1980s among all groups. However, patterns also vary between education groups. Compared with the less educated group, the highly educated group has higher wage inequality and increases faster, especially since the late 1990s.

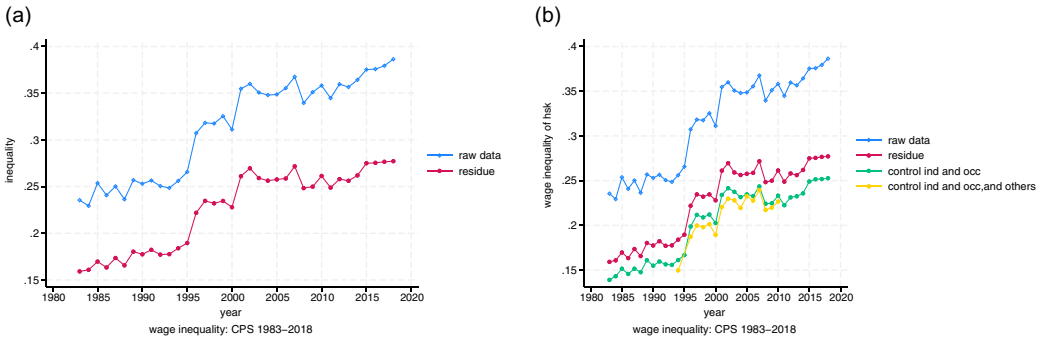
Following the convention in the literature (e.g., Kambourov and Manovskii, 2009), we obtain residual wages after controlling for sex, race, experience, and education. The residual wage inequality is then measured as the variance in the residual wages. As shown in the right panel of Figure 1, the highly educated group has higher residual wage inequality which increases faster, especially during 1993–2001. As a robustness check, we calculate the Gini coefficient and 90–10 ratio as well. Figure A.3 shows that both the Gini coefficient and 90–10 ratio exhibit a similar pattern in the highly educated group. In summary, both facts confirm the substantial level of within-group inequality. In addition, the residual wage inequality within the highly educated group exceeds the overall average and increases faster, especially between 1990 and 2000.

We now examine the inequality within the highly educated group. The left panel of Figure 2 presents the evolution of both the overall and residual wage inequalities. The residual wage inequality accounts for around 70% of the overall wage inequality among the highly educated. This proportion is higher than that commonly documented in the literature (e.g., Lemieux, 2006) for the overall sample comprising individuals of different education levels. The right panel shows the trend of the residual wage inequality when controlling for additional job characteristics, including industry, occupation, location, firm size, and citizenship. Controlling for industry and occupation can only explain 10% more of inequality. The result changes little when additional variables, such as location and firm size, are controlled. Thus, the wage inequality within industry and occupation significantly contributes to the residual wage inequality. To consolidate this finding, we decompose the residual wage inequality.

2.3 Decomposition of the wage inequality

Here, we decompose both the level and change in residual wage inequality into within- and between-job wage inequalities. A job is defined as an industry-occupation pair. Examples of jobs

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press



Overall and Residual Wage Inequality

Overall, Residual and Within-job Wage Inequality

Figure 2. Residual wage inequality of the highly educated.
Notes: In both panels, the inequality is measured among the highly educated. The blue line represents the overall wage inequality. The red line is the residual wage inequality after controlling for only demographic characteristics. The green line is the residual wage inequality after further controlling for occupation and industry. The yellow line is the residual wage inequality after further controlling for location, firm size, and citizenship. Data source: March CPS from the CEPR.

include sales in the FIRE industry, managers in business, clerical workers in retail/wholesale trade, and production workers in durable/nondurable goods, among others.

Decomposition of the level Suppose there are J jobs indexed as $j = 1, \ldots J$. In job j, we denote P_j as the employment share, V_j as the within-job wage inequality, and E_j as the average earnings. Then, $\sum_j P_j V_j$ is the average within-job wage inequality weighted by the employment share. Further, $\sum_j P_j(\ln E_j - \sum_{j'} P_{j'} \ln E_{j'})^2$ is the weighted average of between-job wage inequality, where $\sum_{j'} P_{j'} \ln E_{j'}$ is the weighted average of the logarithm of earnings in the economy. Finally, the overall wage inequality $var(\ln E)$ can be decomposed into the between- and within-job components as follows:

$$var(\ln E) = \sum_j P_j V_j + \sum_j P_j(\ln E_j - \sum_{j'} P_{j'} \ln E_{j'})^2. \quad (1)$$

The contribution of the within-job wage inequality to the total wage inequality then becomes the ratio of $\sum_j P_j V_j$ to $var(\ln E)$.

Decomposition of the change We also decompose the changes in the residual wage inequality over time into the various components of interest. In job j in year t, let $V_{j,t}$ be the wage inequality, $\ln E_{j,t}$ be the average log earnings, $P_{j,t}$ be the employment share, and $\ln E_t$ be the average log earnings among all jobs. Then, the change in the within-job wage inequality from t to $t+1$ is $V_{j,t+1} - V_{j,t}$, change in the between-job wage inequality is $(\ln E_{t+1} - \ln E_{j,t+1})^2 - (\ln E_t - \ln E_{j,t})^2$, and change in the employment share is $P_{j,t+1} - P_{j,t}$. Therefore, the change in the wage inequality between years $t+1$ and t, $V_{t+1} - V_t$, can be decomposed into four components: the weighted average change in the within-job wage inequality $\sum_{j=1}^{J} P_{j,t}[V_{j,t+1} - V_{j,t}]$, the weighted average change in the between-job wage inequality $\sum_{j=1}^{J} P_{j,t}[(\ln E_{t+1} - \ln E_{j,t+1})^2 - (\ln E_t - \ln E_{j,t})^2]$, the weighted average change in the employment share $\sum_{j=1}^{J} (P_{j,t+1} - P_{j,t})[V_{j,t} + (\ln E_t - \ln E_{j,t})^2]$, and an interaction term which is the product of the changes in the employment share, and sum of within- and between-job wage inequalities (simply termed as the "interaction" in the

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 5. How much residual wage inequality accounts for overall inequality

	level(%)				change(%)		
	1990	2000	2010	2018	1990–2000	2000–2010	2010–2018
Residual wage inequality	70.0	73.3	72.9	71.8	87.9	70.2	57.1
–within-job (1-d)	89.4	90.8	90.6	92.5	95.6	89.1	123.7
–within-job (2-d)	84.4	85.7	84.8	86.9	91.4	80.4	120.3

Notes: This table reports how much residual wage inequality accounts for overall inequality by levels or by changes over the previous decade. The first row reports the ratio of residual wage inequality to overall wage inequality. The second row reports the within-job component of residual wage inequality, where the jobs are classified by 1-digit occupations and industries. The third row reports within-job component of residual wage inequality, where the jobs are classified by 1-digit (1-d) industry and 2-digit (2-d) occupation codes. Data source: March CPS.

decomposition exercise):

$$\sum_{j=1}^{J}(P_{j,t+1}-P_{j,t})\{(V_{j,t+1}-V_{j,t})+[(\ln E_{t+1}-\ln E_{j,t+1})^2-(\ln E_t-\ln E_{j,t})^2]\}.$$

Formally, we decompose the change in wage inequality as follows:

$$\begin{aligned}V_{t+1}-V_t=&\sum_{j=1}^{J}P_{j,t}[V_{j,t+1}-V_{j,t}]\\&+\sum_{j=1}^{J}P_{j,t}[(\ln E_{t+1}-\ln E_{j,t+1})^2-(\ln E_t-\ln E_{j,t})^2]\\&+\sum_{j=1}^{J}(P_{j,t+1}-P_{j,t})[V_{j,t}+(\ln E_t-\ln E_{j,t})^2]\\&+\sum_{j=1}^{J}(P_{j,t+1}-P_{j,t})\{(V_{j,t+1}-V_{j,t})+[(\ln E_{t+1}-\ln E_{j,t+1})^2-(\ln E_t-\ln E_{j,t})^2]\}.\end{aligned}\tag{2}$$

Similarly, each component's contribution is defined as the ratio of its change to the total change in the residual wage inequality. Table 5 presents the composition of wage inequality from 1990 to 2018. Specifically, the first row shows that the residual wage inequality accounts for a large proportion of the overall wage inequality, which usually exceeds 70% for the level and 50% for the change. More importantly, among the past four decades, the share of changes in residual wage inequality to changes in overall wage inequality is the highest from 1990 to 2000 (87.9%). Hence, we focus on 1990 and 2000 in the quantitative analysis. The second row shows that the contribution of the within-job wage inequality is around 90% for both the level and change under the 1-digit industry and occupation codes. There may be a concern that the significant contribution of the within-job wage inequality is a result of the broad occupation categorization. Hence, we perform a robustness check using the 1-digit (1-d) industry and 2-digit (2-d) occupation codes. As shown in the third row, the contribution of the within-job wage inequality remains sizable, at about 85% for level and above 80% for change.

We also perform several other robustness checks. First, Table 6 presents the decomposition results using Census data under the 1-digit codes for the years of 1990 and 2000. The results indicate that the contribution of the within-job component is consistently above 80%. In addition, Table A.11 presents the decomposition results using Census data under the 3-digit codes for the years of 1990 and 2000, showing that the contribution of the within-job component exceeds 70%. Moreover, Figure A.4 shows that the contribution of the within-job wage inequality is quite

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 6. Decomposition of residual wage inequality under 1-digit code

	inequality		contribution		
year	overall	within	between	within	between
1990	.406	.330	.077	**81%**	19%
2000	.429	.351	.078	**82%**	18%

Notes: This table computes the contribution of the within-job and between-job inequality to the overall inequality in both 1990 and 2000 using 1-digit industry and occupation code. Data source: CENSUS from CEPR.

Table 7. Decomposition of the changes in residual wage inequality: March CPS

	Within-job	Between-job	Employment	Interaction
1-d code	**82.6%**	9.2%	2.7%	3.5%
1-d ind, 2-d occ	**79.5%**	8.1%	8.5%	0.3%

Notes: This table computes the contribution of within-job, between-job wage inequality, and the employment share interactions to the changes in the residual inequality from 1990 to 2000. The row "1-d code" indicates the one-digit industry and occupation code, "1-d ind, 2-d occ" indicates the one-digit industry and two-digit occupation codes. Data source: March CPS from the CEPR.

Table 8. Decomposition of the changes in residual wage inequality: CENSUS and CPSORG

	Within-job	Between-job	Employment	Interaction
CENSUS	**84.4%**	2.1%	6.9%	6.4%
CPSORG	**110%**	−31.8%	17.7%	4.6

Notes: This table computes the contribution of the within-job, between-job inequality, employment share and interactions to the changes in the residual wage inequality from 1990 to 2000 under 1-digit industry and occupation code. The row "CENSUS" indicates using data from CENSUS, "CPSORG" indicates using data from CPSORG. Data source: CENSUS and CPSORG from CEPR.

large throughout all years. Furthermore, the left panel of Figure A.5 decomposes the overall wage inequality, indicating that the contribution of the within-job wage inequality still exceeds 75%; and, the right panel that presents the decomposition result using the CPSORG dataset obtains similar findings as well.

Second, Table 7 reports the contribution of all components between 1990 and 2000 using March CPS, which shows that the within-job significantly dominates other components. Table 8 presents the results using the CENSUS and CPSORG dataset, again showing that the within-job component accounts for the lion's share. Additionally, Table A.12 summarizes the decomposition results using Census data under the 3-digit industry and occupation codes, where the contributions of the within-job component to the change in the residual wage inequality between 1990 and 2000 also exceed 80%.

In summary, based on the results in Sections 2.2 and 2.3, we establish two robust stylized facts:
[Fact 1] Over 70% of overall inequality among the highly educated has been due to the residual inequality, and about four-fifths of the residual wage inequality is driven by wage dispersion within jobs, defined by industry–occupation pairs.
[Fact 2] Over the 1990–2018 period, the rise in the residual wage inequality of the highly educated has been particularly prominent during the early phase of 1993–2001.

2.4 Performance-pay incidence, skill mismatch, and wage inequality

Next, we first explore the correlations between performance-pay incidence, skill mismatch, and within-job wage inequality. Then, we empirically examine how changes in the performance-pay incidence and skill mismatch over time may contribute to the increasing wage inequality. As

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

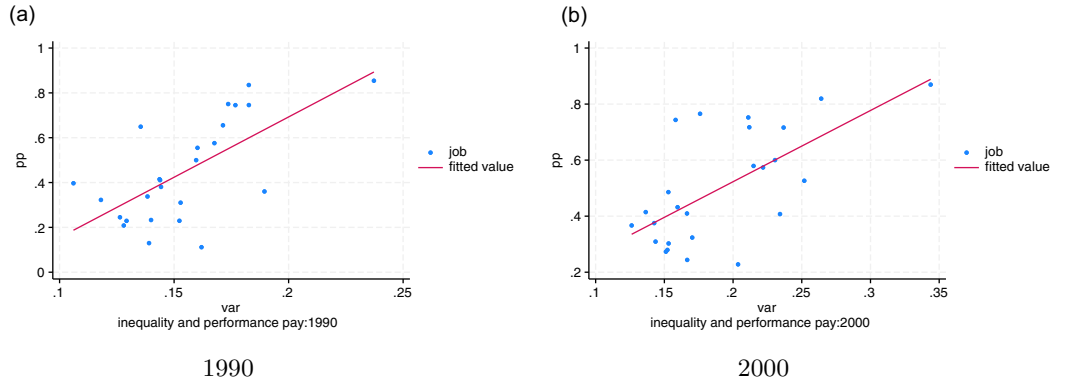


Figure 3. Performance-pay incidence and wage inequality.
Notes: In both panels, each dot represents a job(25 jobs in total). The *x*-axis is the performance-pay incidence. The *y*-axis is the within-job wage inequality. The left panel is based on 1990 data and right panel is based on 2000 data. The coefficients in two panels are 5.38 and 2.54, respectively. Data source: PSID and March CPS.

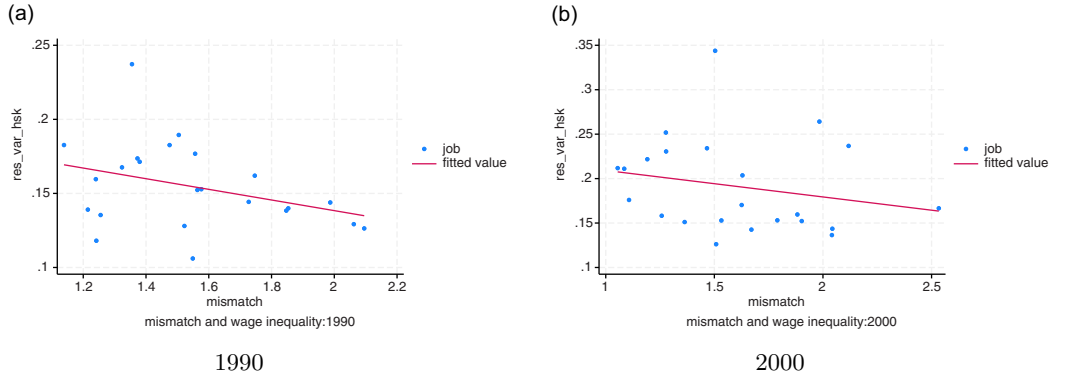


Figure 4. Wage inequality and skill mismatch index.
Notes: In both panels, each dot represents a job(25 jobs in total). The *y*-axis is the wage inequality. The *x*-axis is the mismatch index. The left panel is based on 1990 data, and the right panel is based on 2000 data. The coefficients in both panels are −0.03. Data source: NLSY79 and March CPS.

the datasets are from various sources, throughout the remainder of the study, we classify industries and occupations each into five categories, consistent with the classification in the PSID and NLSY79.[11]

Figure 3 plots the performance-pay incidence against the within-job wage inequality for each of the 25 jobs in 1990 and 2000.[12] We find a significant positive relationship; that is, jobs with a higher performance-pay incidence usually have higher within-job wage inequality. In Appendix A.2, we also show that the relation is robust to using the same job classification as in Lemieux et al. (2009), containing 52 jobs (Figure A.6). In another robustness check, we plot the performance-pay incidence using the PSID data against wage inequality from the CPSORG dataset in Figure A.7. Furthermore, we try an alternative measurement of performance-pay ratio, defined as the ratio of earnings from performance-pay position to total earnings. Figure A.8 reveals a positive correlation between performance pay and wage inequality under this alternative measurement.

Conversely, we find a negative relationship between the wage inequality and skill mismatch index, as shown in Figure 4. Thus, jobs with better matching quality tend to experience higher wage inequality.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 9. Performance-pay incidence, skill mismatch and wage inequality

Variables	(1) Δvar_{ij}	(2) Δvar_{ij}	(3) Δvar_{ij}
Δ_{pp}	0.04***	0.05***	
	(0.00)	(0.00)	
Δ_{mm}	−0.02**		−0.03***
	(0.01)		(0.01)
Constant	0.00	0.00	0.00**
	(0.00)	(0.00)	(0.00)
Observations	389	389	389
R-squared	0.212	0.203	0.027
Industry FE	YES	YES	YES
Occupation FE	YES	YES	YES

Note: Δvar_{ij} denotes the changes in residual wage inequality between two consecutive years. Δ_{pp} and Δ_{mm} is the change in performance-pay incidence and mismatch index between two consecutive years.
Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

To further explore how performance-pay incidence and match quality increase the rising wage inequality, we first compute the changes in the level of wage inequality, performance-pay incidence, and mismatch index between the current and previous years. We then regress the changes in the wage inequality against the changes in performance-pay incidence and skill mismatch index while controlling for industry and occupation fixed effects:[13]

$$\Delta var_{ij} = \beta_1 \Delta_{pp,ij} + \beta_2 \Delta_{mm,ij} + \beta_3 i + \beta_4 j + \epsilon_{ij}.$$

Table 9 presents the results. The positive coefficient on the change in performance-pay incidence (Δ_{pp}) implies that rising performance-pay incidence increases the wage inequality. Specifically, if the performance-pay incidence rises by 10 percentage points, the wage inequality rise by 0.4 percentage points. Similarly, the negative coefficient on the change of skill mismatch index (Δ_{mm}) implies that improving matching quality (lowering mismatch index) increases the wage inequality. The results are robust and significant if we let one or both variables be the regressor. Furthermore, Table A.14 reports significant results when we replace performance-pay incidence with performance-pay ratio.

Additionally, we conduct the Firpo–Fortin–Lemieux (FFL) decomposition as in Firpo et al. (2009). Results in Table A.15 and A.16 confirm that performance-pay incidence (mismatch index) is positively (negatively) associated with wage inequality.[14] Moreover, among other studies, Galindo da Fonseca et al. (2020) build a measurement of match quality using vacancy and unemployment information, and find that a high match quality can motivate greater performance pay. We plot performance-pay incidence and match quality using our measurement in Figure A.9 and find they are positively correlated.

Lastly, as the model predicts that the second moment of match quality motivates more performance pay (specified in Section 4), it would be useful to validify this theoretical property. Accordingly, we plot the correlation between the second moment and standard deviation of match quality with performance pay incidence in Figure A.10 and A.11, respectively. The results show that both the second moments and standard deviation of match quality (mismatch) are positively (negatively) correlated with performance pay incidence, conforming with the theoretical prediction.

The empirical analysis presents partial equilibrium results omitting the interactions between different channels and feedback effects via endogenous wage adjustments. Moreover, the results are silent on the counterfactual scenario, thus limiting the ability to conduct rich policy analysis.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Hence, this requires a comprehensive structure to discipline the interactions among performance-pay incidence, match quality, and wage inequality, which we explore next.

3. The model

Environment There are J jobs available in the economy indexed by $j = 1, 2, \ldots J$. In each job, two types of positions are offered: fixed- (FP) and performance-pay (PP) positions. Workers are heterogeneous in their innate ability a, and choose jobs, positions, and efforts to maximize their utilities. Job characteristics and workers' abilities are both public information.

Performance-pay position A worker's effective labor supply in the performance-pay position depends on their ability a, the job-specific productivity A_j, an idiosyncratic matching premium η, and their effort level e. Specifically, the effective labor supply is given by

$$h = A_j a \eta e. \tag{3}$$

The ability a follows a Pareto distribution $a \sim G(a) = 1 - (\frac{1}{a})^{\theta_a}$, $a \geq 1, \theta_a > 2$, with the minimum ability normalized to one. We assume $\theta_a > 2$ to guarantee finite variance, which is essential for studying wage inequality measured by the variance in log wages. To capture the idea of the quality of skill match, we assume that each worker draws an idiosyncratic matching premium from a job-specific Pareto distribution $\eta \sim F(\eta) = 1 - (\frac{\eta_j}{\eta})^{\theta_{sj}}$ defined over $[\underline{\eta_j}, \infty)$, where $\theta_{sj} > 2$ is the inverse of the tail parameter that captures the dispersion of the matching premium in job j.

We assume that abilities and the matching premium are distributed independently. Thus, the total effective labor supply in the performance-pay position of job j is:

$$H_{jP} = \int_{a \in D_{jP}} \int_{\eta} h_{jp}(a, \eta) dF_j(\eta) dG(a), \tag{4}$$

where D_{jP} is the ability domain for workers, and $h_{jp}(a, \eta)$ is the effective labor supply from a worker of ability a and matching premium η.

Fixed-pay position A worker's effective labor supply in the fixed-pay position depends neither on the innate ability nor effort level. It is assumed to be only linear in the job-specific productivity. Precisely, in job j, the effective labor supply from a worker in the fixed-pay position is A_j. Thus, the total effective labor supply in the fixed-pay position of job j is $H_{jF} = A_j N_{jF}$, where N_{jF} is the employment level in the fixed-pay position of job j.

Production and payment The production function in job j is a constant elasticity of substitution (CES) aggregator of the effective labor supply from the performance- and fixed-pay positions. Specifically, the output of job j is given by the following:

$$Y_j = \left[\alpha_j H_{jF}^{\gamma} + (1 - \alpha_j) H_{jP}^{\gamma}\right]^{\frac{1}{\gamma}}, \ \gamma < 1, \tag{5}$$

where $\frac{1}{1-\gamma}$ is the elasticity of substitution between the labor supply from the two positions—they are substitutes (complementary) if $\gamma > 0$ ($\gamma < 0$); α_j is job specific to reflect the intensity of the fixed-pay position in each job.

We denote w as the wages that each worker in the fixed-pay position can earn. In each job j, given wages w, the representative firm decides the employment level in the fixed-pay position to maximize the revenue net of the payment to workers in the fixed-pay position. That is,

$$\max_{N_{jF}} Y_j - w N_{jF}. \tag{6}$$

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

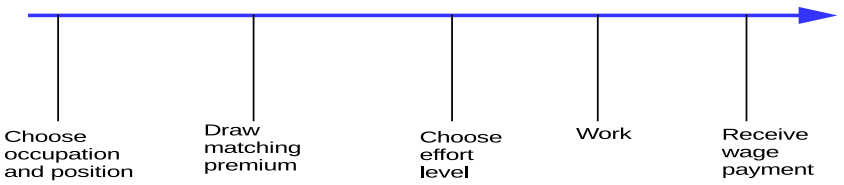


Figure 5. Timeline.
Note: This figure summarizes the timeline for a representative worker in the model economy.

It is straightforward to show that N_{jF} satisfies $N_{jF} = \frac{\chi_j(w)H_{jP}}{A_j}$, where $\chi_j(w) \equiv \left[\frac{(\frac{w}{\alpha_j A_j})^{\frac{\gamma}{1-\gamma}} - \alpha_j}{1-\alpha_j}\right]^{-\frac{1}{\gamma}}$. We show in Appendix D that the total payment to workers at the fixed-pay position, E_{jF}, is equivalent to $E_{jF} = \tilde{\alpha}_j(w)\tilde{A}_j(w)H_{jP}$, where $\tilde{A}_j(w) = \left[\alpha_j \chi_j(w)^{\gamma} + (1-\alpha_j)\right]^{\frac{1}{\gamma}}$, and $\tilde{\alpha}_j(w) = \frac{\alpha_j \chi_j(w)^{\gamma}}{\alpha_j \chi_j(w)^{\gamma} + (1-\alpha_j)}$. We also show in Appendix D that the total output can be expressed as $Y_j = \tilde{A}_j(w)H_{jP}$. Therefore, the residual profit is $R_{jP} = (1 - \tilde{\alpha}_j(w))\tilde{A}_j(w)H_{jP}$.

While the fixed-pay workers are paid at a pooled wage, performance-pay workers' payments crucially depend on their contribution to production. Given fixed-pay workers, the production function is linear in human capital in the performance-pay position. Hence, a worker's contribution to the firm is linear in her human capital. Then, the simplest way to model the payment scheme is that the firm and workers in the performance-pay position share the residual profit. We assume that workers' bargaining power is μ. Therefore, the total payment to workers in the performance-pay position is $E_{jP} = \mu(1 - \tilde{\alpha}_j)\tilde{A}_j(w)H_{jP}$, while the payment to a worker of ability a and matching premium η in the performance-pay position is $\mu(1 - \tilde{\alpha}_j)\tilde{A}_j(w)A_j a\eta e_j(a, \eta)$.

Workers A worker's utility is assumed to positively depend on their consumption c and negatively depend on their effort level e. In addition, workers in the performance-pay position suffer from the job-specific dis-utility of being monitored.[15] We further assume that a worker's utility function is linear in consumption and quadratic in the effort level. Specifically, a worker's utility function from working in the performance-pay position of job j takes the following form:

$$U_j^P = \max_e \quad \mu(1 - \tilde{\alpha}_j)\tilde{A}_j A_j a\eta e - \frac{1}{2}be^2 - M_j, \tag{7}$$

where b measures the degree of dis-utility on effort. M_j is the dis-utility incurred from working in job j.

Workers make decisions on jobs, positions, and efforts. Figure 5 summarizes the timeline of the worker's decision. Workers choose the job and position before the realization of the matching premium. Conversely, the optimal effort level is chosen after workers observe the outcome of the matching premium.

Thus, the ex ante expected utility from working in the performance-pay position of job j for a worker of ability a is $EU_j^P(a) = E_\eta[U_j^P(a, \eta)]$.

Workers in the fixed-pay position always exert the minimum effort since their pay is independent of their effort level. Thus, a worker's payoff from the fixed-pay position can be expressed as follows:

$$U^F = w - \frac{b}{2}\underline{e}^2. \tag{8}$$

Finally, a worker chooses the job and position that delivers the highest expected utility:

$$EV(a) = \max_j\{\max\{U^F, EU_j^P(a)\}\}. \tag{9}$$

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Thus, the fixed-pay position is chosen by a worker of ability a if $U^F > \max_j\{EU_j^P(a)\}$. Otherwise, a worker of ability a chooses the performance-pay position in a job given by $j^*(a) = \arg\max_j\{EU_j^P(a)\}$.

Two remarks are in order before turning to theoretical results. First, by focusing on the production side, we assume that all goods are perfect substitutes. While this implicitly assumes away the effects of demand shifts and relative price changes on labor and wage inequality, such simplification is not as vulnerable because we focus on highly education labor only. Second, to guarantee positive demand for all jobs or no empty job, we assume a supermodularity condition (as specified in Proposition 10 below). Therefore, all jobs will be perfectly sorted according to workers' abilities.

4. Theoretical analysis

4.1 Equilibrium

Given the job-specific characteristics $\{A_j, \alpha_j, M_j, \theta_{sj}, \underline{\eta_j}\}$, a *sorting equilibrium* is described by the wage rate w, effort e, and job choice ($j^*(a)$) and position (F or P), employment in the fixed-pay position N_{jF}, and effective labor supply in performance-pay position H_{jP}, such that:

1. Given the wage rate w, workers optimally choose jobs ($j^*(a)$) and positions (F or p) as in Equation (9).
2. Given the wage rate w, workers' effort e solve problems in Equation (7).
3. Given the wage rate w and job and position choices, the effective labor supply from the performance-pay position H_{jP} is determined in Equation (4).
4. Given the wage rate w and effective labor supply from the performance-pay position H_{jP}, firms decide employment in the fixed-pay position N_{jF} to maximize profits in Equation (6)
5. The wage rate w clears the labor market, such that:

$$\sum_j \int_{a\in\{D_{jF}\cup D_{jP}\}} dG(a) = 1,$$

where domain D_{jP} and D_{jF} are the ability domains for workers in performance and fixed-pay positions, respectively, and the total labor force is normalized to 1.

4.2 Analytical results

Next, we solve the worker's decision in a backward fashion:

- (Stage 2) We first solve the effort exerted by a worker of ability a given that the performance-pay position has been chosen and matching premium draw has been realized.
- (Stage 1) We then move backward to determine whether this worker would select a performance-pay position and, if so, which job they would take.

Specifically, in Stage 2, a worker of ability a in the performance-pay position of job j, after realizing the productivity premium η, chooses the effort to maximize their utility shown in equation (7). Consumption c_j equals the wage income: $c_j = \mu(1-\tilde{\alpha}_j)\tilde{A}_j A_j a\eta e$. Thus, the resulting effort level is $e_j(a,\eta) = \mu(1-\tilde{\alpha}_j)\tilde{A}_j A_j a\eta/b$.

Substituting in the aforementioned effort level, a worker's effective labor supply is:

$$h_{jp}(a,\eta) = \frac{\mu(1-\tilde{\alpha}_j)\tilde{A}_j[A_j a\eta]^2}{b}$$

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

and the ex post utility from working in job j can be derived as

$$U_j^P(a;\eta) = \frac{1}{2b}((1-\tilde{\alpha}_j)\tilde{A}_j A_j \eta \mu a)^2 - M_j.$$

In stage 1, before the realization of the matching premium draw, the decision to select the performance-pay position depends on the expected utility:

$$EU_j^P(a) \equiv E_\eta[U_j^P(a,\eta)] = \tilde{C}_j a^2 - M_j, \tag{10}$$

where $\tilde{C}_j = \frac{1}{2b}((1-\tilde{\alpha}_j)\tilde{A}_j A_j \mu)^2 E_j(\eta^2)$ and

$$E_j(\eta^2) = \frac{\theta_{sj}}{\theta_{sj}-2}(\underline{\eta_j})^2. \tag{11}$$

Because this second moment $E_j(\eta^2)$ term enters the expected utility from the performance-pay position, we can conveniently use this to measure the match quality in job j. Our result shows that match quality positively depends on both the job-specific minimum matching premium $\underline{\eta_j}$ and its dispersion measured by the tail parameter $1/\theta_{sj}$. Recall that $j^*(a)$ denotes the job in which a worker of ability a gains the highest expected utility from the performance-pay position. Thus, the performance-pay position in job $j^*(a)$ is selected by a worker of ability a if and only if $EU^P_{j^*(a)}(a) > U^F$.

In summary, the two extensive margins—selection into a performance-pay position and job sorting captured by $j^*(a)$—are interconnected, depending crucially on the worker's ability a and an intensive margin of the match quality measure $E_j(\eta^2)$. In the sorting equilibrium, all margins, in addition to general job productivity, affect the wage dispersion within each job.

4.3 The one-job case

To gain further insight, consider a scenario of only one job with two available positions: fixed pay and performance pay. Here, the sorting channel is eliminated by construction, thereby enabling us to focus on the effect of performance-pay incidence on wage inequality. It is also an interesting benchmark because it is is equivalent to a multi-job one *under a given sorting outcome*.

The income earned from each position positively depends on the individual's ability a. We denote $y(a)$ and $z(a)$ as the income of an agent of ability a in the performance- and fixed-pay positions, respectively.[16] As long as $y''(a) > z''(a)$ and $y(\underline{a}) < z(\underline{a})$, workers of higher ability work in the performance-pay position. Let a^* be the cutoff. The income inequality expressed as the variance in income can be obtained as follows:

$$VI = \int_{a^*}^{\bar{a}} y(a)^2 dG(a) + \int_{\underline{a}}^{a^*} z(a)^2 dG(a) - \left[\int_{a^*}^{\bar{a}} y(a) dG(a) + \int_{\underline{a}}^{a^*} z(a) dG(a)\right]^2 \tag{12}$$

Denote $n_p = \int_{a^*}^{\bar{a}} dG(a)$; then, $1 - n_p = \int_{\underline{a}}^{a^*} dG(a)$. In Appendix C, we show that:

$$\frac{d(VI)}{dn_p} = Var(y(a)) - Var(z(a)) + (1-2n_p)\big[E[y(a)] - E[z(a)]\big]^2. \tag{13}$$

Thus, under normal circumstances, where performance-pay incomes are more dispersed than fixed-pay income ($Var(y(a)) > Var(z(a))$), we have the following proposition.

Proposition 1. (Performance-pay Incidence and Wage Inequality) *Under* $Var(y(a)) > Var(z(a))$, *the overall inequality in the job increases with* n_p *if* n_p *is not too large.*

Intuitively, how much the performance-pay incidence affects inequality depends on the inequality within the performance-pay position relative to the fixed-pay position, and the gap

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

between the average wages in the two positions. The former effect is expected to be positive, whereas the latter is positive only if $n_p < 1/2$. If n_p is not too large, then the former plays a dominant role. We do return to this discussion later under a simplified production function.

Since wage inequality increases with the performance-pay incidence, we next focus on how n_p is determined within a job. For illustrative purposes, we further assume that the production function takes a Cobb–Douglas form. The firm's profit maximization problem then becomes

$$\max \quad (AN_F)^{\alpha} H_p^{1-\alpha} - wN_F.$$

The relationship between wages and employment in the fixed-pay position is $N_F = \left[\frac{w}{\alpha A}\right]^{\frac{1}{\alpha-1}} \frac{H_p}{A}$.

Defining $\chi(w) \equiv \left[\frac{w}{\alpha A}\right]^{\frac{1}{\alpha-1}}$, the total payment to the fixed-pay position can be summarized as $\chi(w)^{\alpha} \alpha H_p$. Following the same notation as in the full-fledged model, we let $\widetilde{A} = \chi(w)^{\alpha}$ and $\widetilde{\alpha} = \alpha$. The expected utility from working in the performance-pay position is $\frac{1}{2b}\left[(1-\widetilde{\alpha})\widetilde{A}A\mu\right]^2 E\left[\eta^2\right] a^2$, and in the fixed-pay position is w if we normalize the minimum effort level to zero. Thus, a worker chooses the performance-pay position if and only if $\frac{1}{2b}\left[(1-\widetilde{\alpha})\widetilde{A}A\mu\right]^2 E\left[\eta^2\right] a^2 > w$.

In this simple illustration, we also drop the monitoring cost because it is only useful for sorting workers across jobs. Moreover, since $(1-\widetilde{\alpha})\widetilde{A}A = (1-\alpha)\left[\frac{w}{\alpha A}\right]^{\frac{\alpha}{\alpha-1}} A$, the cutoff ability a can thus be pinned down from

$$\frac{\mu^2}{2b}\left[(1-\alpha)\left[\frac{w}{\alpha A}\right]^{\frac{\alpha}{\alpha-1}} A\right]^2 E\left[\eta^2\right] a^2 = w.$$

In addition, the labor market clearing condition can be expressed as follows:

$$\left[\frac{w}{\alpha A}\right]^{\frac{1}{\alpha-1}} \frac{H_p}{A} = 1 - a^{-\theta_a},$$

where $H_p = \frac{\mu}{b}(1-\alpha)^2 \left[\frac{w}{\alpha A}\right]^{\frac{2\alpha}{\alpha-1}} A^2 E\left[\eta^2\right] \frac{\theta_a}{\theta_a - 2} a^{2-\theta_a}$.

We can now prove the existence and uniqueness of the solution (a, w) and characterize the impacts of A and $E[\eta^2]$ on the equilibrium outcome. To do this, we substitute the cutoff condition into the labor market equilibrium condition to eliminate $E[\eta^2]$ and manipulate it to arrive at the following equation:

$$\frac{\mu}{2}\frac{\theta_a - 2}{\theta_a}\left(a^{\theta_a} - 1\right)\left[\frac{w}{A}\right]^{\frac{\alpha}{1-\alpha}} = \alpha^{\frac{1}{1-\alpha}}, \tag{14}$$

which is referred to as the Labor market Equilibrium (LE) locus. Manipulating the cutoff condition, we obtain the following:

$$\frac{\mu^2}{2b}(1-\alpha)^2 \left[\frac{1}{\alpha}\right]^{\frac{2\alpha}{1-\alpha}} AE\left[\eta^2\right] a^2 \left[\frac{w}{A}\right]^{\frac{1+\alpha}{\alpha-1}} = 1, \tag{15}$$

which is referred to as the cutoff ability (CA) locus.

Clearly, LE is increasing in $\frac{w}{A}$ and a but independent of A and $E[\eta^2]$. Moreover, CA is decreasing in w/A but increasing in a and A. Thus, the LE locus is downward sloping in the $(a, \frac{w}{A})$ space, whereas the CA locus is upward sloping, as shown in Figure 6. Intuitively, a higher cutoff ability means more workers in the fixed-pay position, thereby leading to lower fixed-pay wages under the labor market equilibrium. Conversely, for the marginal worker to be indifferent, higher fixed-pay wages raise one's outside options, thus requiring a higher cutoff ability to maintain indifference.

We first study the boundary properties. Recall that $a \geq 1$ and $\theta_a > 2$. Consider the LE locus: as $a \to 1$, $\frac{w}{A} \to \infty$; as $a \to \infty$, $\frac{w}{A} \to 0$. Consider the CA locus: as $a \to 1$, $\frac{w}{A} \to \left[\frac{w}{A}\right]_{\min} \equiv$

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

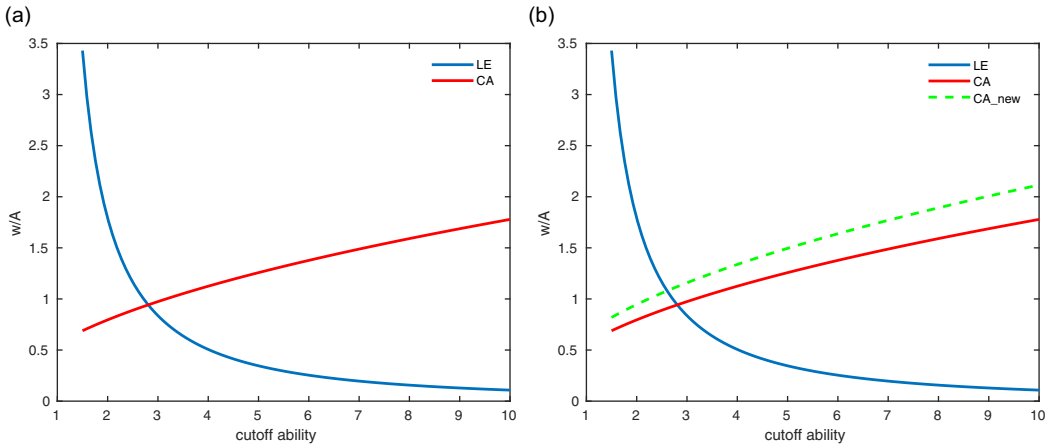


Figure 6. Comparative statics in the one-job case.
Notes: The blue curve denotes equation (14) and the red curve denotes equation (15). The left panel shows the existence and uniqueness of the equilibrium outcome in the simple one-job setting. In the right panel, we either double productivity A or match quality summarized by $E[\eta^2]$ to evaluate the changes in w/A.

$\left\{\frac{\mu^2}{2b}(1-\alpha)^2\left[\frac{1}{\alpha}\right]^{\frac{2\alpha}{1-\alpha}}AE\left[\eta^2\right]\right\}^{\frac{1-\alpha}{1+\alpha}}$; as $a\to\infty$, $\frac{w}{A}\to\infty$. These, together with the slope properties of the two loci, yield the following proposition.

Proposition 2. (Existence and Uniqueness) *For any job, the position cutoff ability and fixed-pay wages are uniquely determined.*

Moreover, in the $(a, \frac{w}{A})$ space, an increase in either A or $E[\eta^2]$ only shifts the CA locus up, without affecting the LE locus. This leads to a lower a, higher $\frac{w}{A}$, and hence, higher w.

Proposition 3. (Comparative Statics: Wage and Position Selection) *For any job, an increase in technology or match quality reduces the position cutoff ability and increases the fixed-pay wages.*

Figure 6 illustrates the comparative statics from doubling either productivity or match quality (summarized as $E(\eta^2)$). Intuitively, technological improvements lead to higher overall productivity and fixed-pay wages. However, these advancements also result in a lower cutoff ability due to the corresponding increase in performance-pay wages. Consequently, the labor supply to the fixed-pay position is lower, leading to a more-than-proportional increase in the fixed-pay wages. Conversely, an increase in match quality raises the reward in the performance-pay position. By complementarity, this also raises the productivity of the fixed-pay workers, and hence, fixed-pay wages.

We now return to the characterization of the within-job wage inequality in Equation (13). A natural question arises: How do technology and match quality influence the impacts of performance-pay incidence on the overall wage inequality within the job? As shown in Proposition 3, either shift reduces the cutoff ability and raises the fixed-pay wages. While the latter tends to reduce $\left[E[y(a)]-E[z(a)]\right]^2$, the former expands the range of workers in the performance-pay position but suppresses that in the fixed-pay position. Together, these lead to higher $Var(y(a))$, lower $Var(z(a))$, and hence, larger $Var(y(a))-Var(z(a))$. The presence of these channels of endogenous cross-position spillovers is novel. Thus, one may conclude that rising fixed-pay technology amplifies the effect of the performance-pay incidence in the overall within-job wage inequality if n_p is not too large. Regarding higher match quality, an additional direct

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

effect widens the performance-pay dispersion $Var(y(a))$ resulting from the heterogeneous premium draws. Thus, while its net impact on the effect of the performance-pay incidence on the overall within-job wage inequality is generally ambiguous, the likelihood of amplifying the effect of the performance-pay incidence on the overall within-job wage inequality is greater than the technology effect. In summary, we have the following proposition.

Proposition 4. (Comparative Statics: Inequality) *Suppose that* $Var(y(a)) > Var(z(a))$ *holds and* n_p *is not too large. Then, a better technology or higher match quality amplifies the effect of the performance-pay incidence* n_p *on the within-job and within-performance-pay inequality* ($Var(y(a))$), *but dampens the effect of the performance-pay incidence* n_p *on the within-job and between-position wage inequality* ($\left[E[y(a)] - E[z(a)]\right]^2$), *leading to an overall ambiguous net impact.*

We summarize the rich mechanisms underlying the comparative static results presented in Propositions 3 and 4. First, as in standard neoclassical theory, contributions from fixed- and performance-pay workers to production are Pareto complements. Thus, the fixed-pay technology raises the productivity of both fixed- and performance-pay workers. Likewise, match quality enhances the productivity of workers in both positions. Here, the direct effect on its own position should dominate the indirect effect on the other position. Thus, their impacts on the within-job and between-position wage inequality are unambiguous. Second, there is a labor composition effect via changes in the cutoff ability, as in occupational choice models. When the cutoff is lower, more workers are there in the performance-pay position but fewer in the fixed-pay position. This labor composition effect feeds back to affect income in each job via the labor market equilibrium, and subsequently, changes the between-position inequality. Third, because of the changes in the cutoff ability, the range of ability changes, thereby affecting the dispersion of income in both positions. When the cutoff is lower, inequality in the fixed-pay position narrows, whereas that in the performance-pay position widens. Fourth, changes in the cutoff ability also affect the selection of performance-pay workers. When the cutoff is lower, the performance-pay position is less selective, thus reducing the average productivity of performance-pay workers as a whole. In summary, even under this simplified setup, we have four interactive mechanisms: *complementarity*, *labor composition*, *relative ability dispersion*, and *performance-pay selectivity*.

If we consider a CES production or sorting mechanism across jobs, the comparative static properties regarding the effect of the performance-pay incidence on the overall within-job wage inequality would become even more complicated. While we examine sorting in greater depth in the next subsection, note that CES production enriches the complementarity mechanism, whereas sorting affects the labor composition, relative ability dispersion, and performance-pay selectivity mechanisms. As shown in Figure 7, when the performance- and fixed-pay positions become more complementary ($\gamma = -2$), the elasticity of substitution is $1/(1-\gamma) = 1/3$. The value of the fixed-pay position tends to decrease with both productivity and match quality, as opposed to increasing in the case of the Cobb–Douglas production function. This is because stronger complementarity dampens the negative labor composition effect in the performance-pay positions. This strengthens the positive interactions between match quality and performance-pay incidence, as indicated by the negative relationship between match quality and cutoff ability. Due to all these rich interactions, understanding such impacts thus becomes a quantitative question, which we investigate in Section 5.

4.4 Positive assortative sorting equilibrium

Here, we return to the full-fledged model. To circumvent the potential issues associated with multiple equilibria in a general setting, we restrict our attention to the case of a monotone job ranking, under which we order n such that $\{M_n\}$ and $\{\tilde{C}_n\}$ are increasing sequences of n.[17] In this case, we can characterize the sorting equilibrium.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

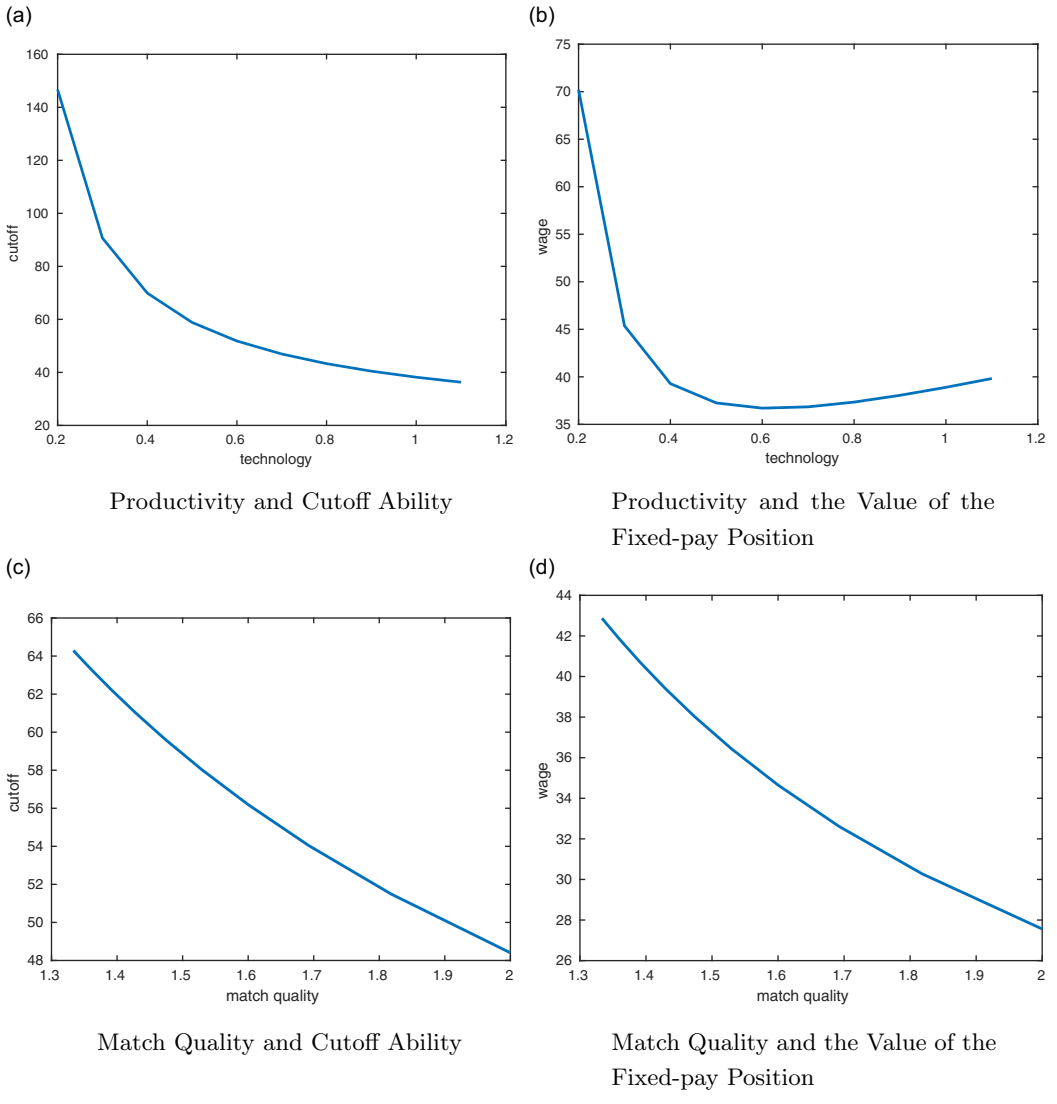


Figure 7. Simulation of the one-job case with the CES production function.
Notes: When the performance-pay and fixed-pay positions become more complementary ($\gamma = -2$) as opposed to in the benchmark Cobb–Douglas production function, we re-evaluate how productivity and match quality affect cutoff abilities and wages.

We start with the following lemma to show that in the performance-pay position, higher ability workers choose the job of a larger index.

Lemma 5. *If a worker of ability a chooses to work in the performance-pay position of job k , then the workers of ability $a' > a$ prefer the performance-pay position in any job $n \geq k$ than in job k.*

We further characterize workers' preferences over jobs within the performance-pay position in the following lemmas.

Lemma 6. *If* $\frac{M_n - M_{n-1}}{\tilde{C}_n - \tilde{C}_{n-1}}$ *is increasing in n for all* $n > 1$, *the workers of ability* $a \in \left(\sqrt{\frac{M_n - M_{n-1}}{\tilde{C}_n - \tilde{C}_{n-1}}}, \sqrt{\frac{M_{n+1} - M_n}{\tilde{C}_{n+1} - \tilde{C}_n}}\right)$ *prefer the performance-pay position at job n to that at* $n-1$ *and* $n+1$.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Lemma 7. *If* $\frac{M_n - M_{n-1}}{\tilde{C}_n - \tilde{C}_{n-1}}$ *is increasing in n for all* $n > 1$*, when a worker prefers the performance-pay position in job n to* $n-1$*, they also prefer the performance-pay position in job n to any job* $k < n-1$*.*

Lemma 8. *If* $\frac{M_n - M_{n-1}}{\tilde{C}_n - \tilde{C}_{n-1}}$ *is increasing in n for all* $n > 1$*, when a worker prefers the performance-pay position in job n to* $n+1$*, they also prefer the performance-pay position in job n to any job* $k > n+1$*.*

The proof of the above three lemmas can be found in Appendix C. These lemmas together imply that conditional on the performance-pay position, workers of ability $a \in \left(\sqrt{\frac{M_n - M_{n-1}}{\tilde{C}_n - \tilde{C}_{n-1}}}, \sqrt{\frac{M_{n+1} - M_n}{\tilde{C}_{n+1} - \tilde{C}_n}}\right)$ find it optimal to work in job n. We can denote $a_n^* := \sqrt{\frac{M_n - M_{n-1}}{\tilde{C}_n - \tilde{C}_{n-1}}}$ as the cutoff ability. Thus, we can conclude that workers of ability $a \in [a_n^*, a_{n+1}^*)$ prefer the performance-pay position in job n than in any other jobs.

The previous analysis is restricted to workers in the performance-pay position. Next, we explore how workers choose between the performance- and fixed-pay positions.

Lemma 9. *Denote* $\underline{U}$ *as the lowest level of utility, when* $\left(\frac{\underline{U}+M_1}{\tilde{C}_1}\right)^{1/2} > \underline{a}$*, workers of ability* $a \in \left[\underline{a}, \left(\frac{\underline{U}+M_1}{\tilde{C}_1}\right)^{1/2}\right)$ *choose to work in the fixed-pay position and workers of ability* $a \in \left[\left(\frac{\underline{U}+M_1}{\tilde{C}_1}\right)^{1/2}, a_2^*\right)$ *choose to work in the performance-pay position of job 1.*

Combining all lemmas above yields the following proposition.

Proposition 10. (Positive Assortative Sorting) *If* $\frac{M_n - M_{n-1}}{\tilde{C}_n - \tilde{C}_{n-1}}$ *is increasing in n for all* $n > 1$*, a sorting equilibrium is positive assortative both within and across jobs when* $\left(\frac{\underline{U}+M_1}{\tilde{C}_1}\right)^{1/2} > \underline{a}$*.*

(i) *Sorting across jobs: Workers of ability* $a \in \left(\sqrt{\frac{M_n - M_{n-1}}{\tilde{C}_n - \tilde{C}_{n-1}}}, \sqrt{\frac{M_{n+1} - M_n}{\tilde{C}_{n+1} - \tilde{C}_n}}\right)$ *find it optimal to work in the performance-pay position of job* $n > 2$*, whereas the workers of ability* $a \in \left[\left(\frac{\underline{U}+M_1}{\tilde{C}_1}\right)^{1/2}, a_2^*\right)$ *choose to work in the performance-pay position of job 1.*

(ii) *Sorting within jobs: Workers of ability* $a \in \left[\underline{a}, \left(\frac{\underline{U}+M_1}{\tilde{C}_1}\right)^{1/2}\right)$ *choose to work in the fixed-pay position of any job.*

An illustration of job and position choices To illustrate the sorting equilibrium defined above, Figure 8 provides a simple example. Suppose the light blue line is the ability distribution and there are four jobs: 1 (black), 2 (blue), 3 (red), and 4 (green). Hence, there are four cutoff abilities which, shown in the figure, are $a_1^* = 2$, $a_2^* = 4$, $a_3^* = 6$, and $a_4^* = 8$, respectively. The workers work in the fixed-pay position if $a < a_1^*$ and choose the performance-pay position of job n if $a_n^* \le a < a_{n+1}^*$, $n = 1, 2, 3$. Finally, workers work in the performance-pay position of job 4 if $a \ge a_4^*$. Therefore, in each job, some workers are there in the performance-pay position and others in the fixed-pay position.

4.5 Further discussion

While the model delineated above yields useful implications for wage dynamics and income inequality, some model limitations deserve further comments, particularly along the lines of dynamics, other dimensions of firm heterogeneities and the considerations of low-skilled labor, which are now discussed in order.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

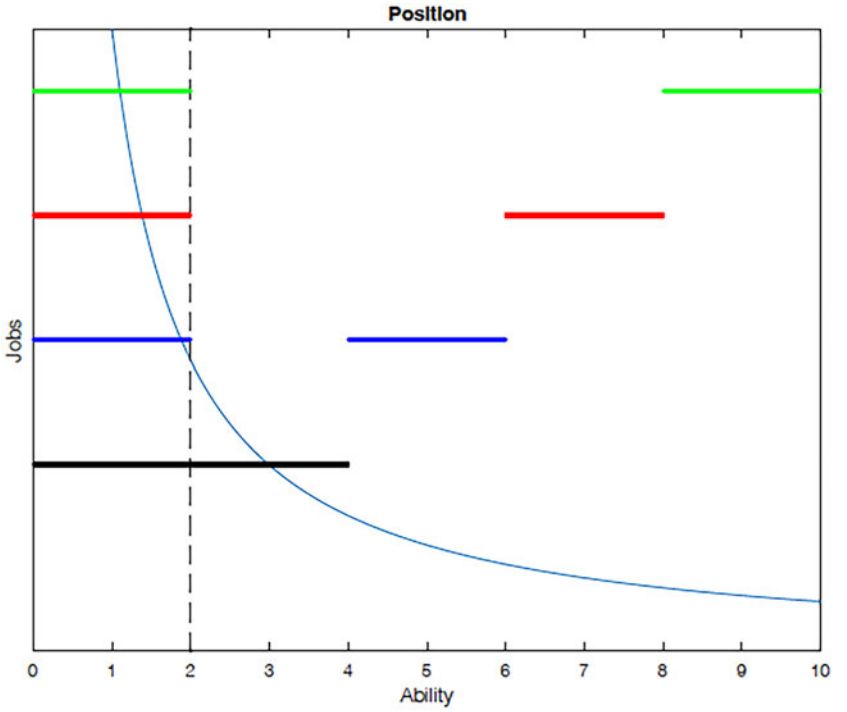


Figure 8. Job and position choice.
Notes: This figure illustrates the distribution of workers over different jobs' performance-pay and fixed-pay positions based upon their innate abilities. Workers of the lowest abilities work at fixed-pay positions of each job, followed by workers working at performance-pay position of job 1(black), 2(blue), 3(red), 4(green) in an ascending order.

Dynamics For the sake of simplicity, our equilibrium sorting framework is basically static. In the quantitative exercises to be conducted below, dynamics are, by construction, captured by changes in exogenous variables or parameters. Thus, potential dynamic channels through human capital accumulation and worker relocation are ignored. Should workers be allowed to accumulate human capital, the advancers can enjoy better performance pay and hence wage dispersion is likely to rise. Similar effects occur when workers may relocate to better paid jobs and improve match quality. While the latter case can be largely captured by our job sorting channel, the former may induce greater within-performance-pay inequality, yielding an even stronger role of performance-pay. That is, while the qualitative findings remain unchanged, one may view our computed contribution by performance-pay as a conservative one.

Other Dimensions of Firm Heterogeneity In our model, two key dimensions of firm heterogeneities under considerations are job-specific technologies and monitoring costs. A common dimension considered in the spatial sorting literature is amenity (e.g., see Diamond and Serrato, 2025, and papers cited therein). In addition to location-dependent amenity that is beyond the scope of our paper, it may include firm or job-specific amenities. Sockin (2022) identifies many such amenities and find higher-paying firms offer better amenities and amenities widen cross-firm wage dispersion. In the content of our paper, one may think of such amenities as being related to firm practices and management quality that may affect the design and prevalence of performance-pay contracts. These are qualitatively equivalent to reductions in the monitoring cost within our framework, which would induce higher performance-pay incidence and greater inequality. Of course, some amenities such as workplace environment may directly yield a positive utility. In this case, workers may be satisfied with a lower-paying job in exchange for better job amenity.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Because Sockin's empirical evidence suggests higher-paying firms offer better amenities, this latter channel is unlikely to be quantitatively dominant. As a consequence, our main findings are most likely to prevail.

On Low-Skilled Labor Our paper focuses exclusively on highly educated workers. This is motivated by the fact that the highly educated are the main contributors to wage inequality (Figure 1) and that both performance pay and match quality matters for wage inequality for this group of workers (Figures 3 and 4). On the contrary, these two channels may not play much a role for low-skilled labor. To see this, we present the correlation between performance pay and wage inequality and that between match quality and wage inequality for this group of workers in Figure A.12 and A.13 in the Appendix A.2. Both correlations turn out to be insignificant. That is, wage dispersion within the low-skilled is unlikely to be driven by performance pay or match quality.

5. Quantitative analysis

5.1 Calibration

We calibrate the benchmark model to the US economy in 2000, similar to Section 2.4, focusing on 25 jobs, as observed in the March CPS, PSID, and NLSY79 datasets. In our model, there are four common parameters $\{\mu, \theta_a, \gamma, b\}$ and five job-specific parameter series $\{A_j, M_j, \theta_{sj}, \underline{\eta}_j, \alpha_j\}_{j=1}^{J}$. Two parameters can be preset based on the literature or simple targets. One is μ, which governs workers' bargaining power in the performance-pay position. This is set to 0.54 by Hall and Milgrom (2008) and 0.72 by Shimer (2005), with others choosing lower values such as 0.4 and 0.5. Because we consider only highly educated workers whose bargaining strength is above average, we set $\mu = 0.6$. Another is θ_a, the shape parameter of the Pareto distribution for workers' innate abilities. We set $\theta_a = 8.0$ so that the 90–10 earnings ratio in our calibrated economy is about 5.

All other parameters are jointly calibrated based on the model. However, each is more closely connected to a primary target. For the crucial parameter of the elasticity of substitution between the efficient labor supply in the production function γ, as discussed in Section 4.3, we calibrate it to match the overall between-job wage inequality. The calibrated elasticity indicates that workers at different job positions are complementary, and the value of the fixed-pay position and cutoff ability are decreasing in the match quality. Thus, the higher the match quality is, the stronger are sorting incentives for the performance-pay position and job sorting. For the dis-utility parameter from exerting effort b, we set it such that the least talented individual in performance-pay position chooses to exert the same effort level as those workers in the fixed-pay position (i.e., the minimum effort: $e_{\min}$). Regarding job-specific monitoring costs $\{M_j\}$, we normalize M_1 to zero and calibrate the remaining M_j to match the number of workers in the performance-pay position in each job. As $\{\theta_{sj}\}$ captures the skewness of the matching premium of each job, they are calibrated to minimize the distance between the model-predicted within-job standard deviation of log earnings and their data counterparts. Regarding whether $\underline{\eta}_j$ governs the minimum scale of the Pareto distribution of the matching premium, we first establish the following one-to-one mapping between the mismatch index (m_j) measured in Section 2.1 and match quality(η_j) in job j:

$$m_j = m_{j0} + \frac{1}{\eta_j}.$$

If η_j follows a Pareto distribution defined over $[\underline{\eta}_j, \infty)$ with the shape parameter θ_{sj}, the distribution function of m_j is then

$$\tilde{F}(m_j) = [\underline{\eta}_j \cdot (m_j - m_{j0})]^{\theta_{sj}}.$$

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 10. Benchmark parameterizations

Description	Para.	Value.	Source/Targets
	External Parameters		
Number of jobs	N	25	Section 2.4
Bargaining power	μ	0.6	Shimer (2005) and Hall and Milgrom (2008)
Shape of the ability distribution	θ_a	8.0	90–10 earning ratio
	Jointly Determined Parameters		
Elas. of subs. between two pos	$\frac{1}{1-\gamma}$	0.357	between-job wage inequality
Coefficient on dis-utility from working	b	212.3	minimum effort is 1
Lowest job-specific productivity	A_1	0.86	fraction of workers at fixed-pay
Job-specific productivity	$\{A_j\}_{j=2}^N$	Figure B.1	relative wages of each job
Coefficient on fixed-pay in production	$\{\alpha_j\}$	Figure B.1	Performance-pay incidence
Monitoring cost	$\{M_j\}_{j=2}^N$	Figure B.1	dist. of workers at performance-pay
Shape of the matching premium distribution	$\{\theta_{sj}\}$	Figure B.1	Within-job wage inequality
Scale of the matching premium distribution	$\{\underline{\eta}_j\}$	Figure B.1	mean of the mismatch index

Notes: This table summarizes parameter value in the benchmark economy. The values for some standard parameters are directly obtained from the literature. The remaining parameters are calibrated to match stylized facts in the data.

We thus jointly calibrate the job-specific $\{\eta_j, m_{j0}\}$ to minimize the distance between the model-predicted mean and standard deviation of the job-specific mismatch index, and their data counterparts.[18] This enables us to obtain a match quality series using Equation (11). Recall that α_j is the coefficient on the labor supply from the fixed-pay position in the production function. Thus, it is calibrated to match the performance-pay incidence in each job, which is equivalent to the employment share of workers in the performance-pay position in any job. Finally, turning to job-specific productivity $\{A_j\}$, we calibrate A_1 to match the number of workers in the fixed-pay position and then compute $\{A_j\}_{j=2}^N$ according to the ratio of average pay in job j to that in job 1. Because these parameters are interconnected in the model, they are jointly calibrated to meet the targets by minimizing distance.

Table 10 summarizes the parameter values and their targets. While all other target moments are easily understood, we should note that the "performance-pay incidence" target is the performance-pay ratio within each job, whereas the "dist. of workers at performance-pay" target is the number of workers in the performance-pay position across jobs.

To facilitate decomposition analysis in the next subsection, we must obtain the job-specific series in 1990 by recalibrating the model to the 1990 data following an identical strategy to the exercise in 2000. Therefore, we re-calibrate the jobs specific parameters $(\{A_j, \alpha_j, M_j, \theta_{sj}, \underline{\eta}_j\})$ in 1990 using job-by-job relative wage, performance-pay incidence, distribution of worker at performance-pay, match quality, and within-job inequality measures, while maintaining other common parameters (e.g., b or γ). Figure B.1 presents the parameters in 2000 (benchmark) and Figure B.2 compares the parameters in 1990 and 2000. We find that α and M have no clear trend, whereas job-specific productivity A has increased and θ has decreased over time.

Owing to the strict parameter restriction,[19] it is not guaranteed that we can exactly match the data counterparts of each job. Nonetheless, as shown in Figure 9, our calibration is quite precise: the model perfectly matches the within-job inequality of each job in 2000; it perfectly matches the inequality in 1990 in all but one job, (Goods, professional).[20]

Notably, in our calibration, we have never targeted any aggregate inequality moments. Such untargeted moments can thus be used to test model fit. In Table 11, we report the model-predicted

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 11. Benchmark inequality: model versus data

	1990		2000	
	Data	Model	Data	Model
Within Job	0.150	0.151	0.183	0.183
Between Job	0.019	0.018	0.021	0.014
Overall	0.169	0.169	0.204	0.197

Note: The data on each type of wage inequality are computed from the March CPS dataset. Model-implied wage inequality is based on the calibrated parameters summarized in Table 10.

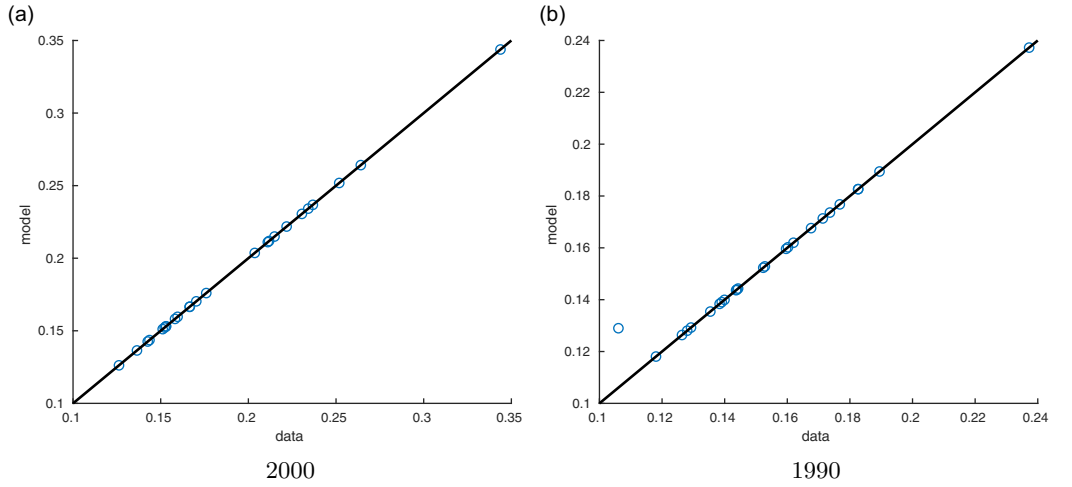


Figure 9. Model fitness in 2000 and 1990.
Note: The *x*-axis represents job-specific within-job wage inequality computed from the data and the *y*-axis is the wage inequality within each job computed from the model. Each dot represents a job.

within-job, between-job, and overall wage inequalities as opposed to the data counterpart in 1990 and 2000. As it can be seen, the model matches these untargeted between-job wage and overall wage inequalities reasonably well.

5.2 Decompose the changes in within-job inequality

Next, we further disentangle how the performance-pay incidence, match quality, sorting, and job-specific productivity affect the pattern of within-job wage inequality. Specifically, we conduct a decomposition analysis that changes the values of each job-specific series in 2000 into their respective values in 1990. All other parameters remain their benchmark values. Performance pay channel is governed by monitoring cost $\{M_j\}$, match quality channel by $\{\theta_{sj}\}$, and productivity by $\{A_j\}$. As sorting is endogenous, in the counterfactual exercise we shut it down by directly setting the sorting result in 1990 to the economy of 2000 (benchmark).

To evaluate the contribution of factor k to the changes in inequality from 1990 to 2000 in job j, denoted as π_k^j, we apply the following formula:

$$\pi_k^j = \frac{(model_{2000}^j - counter_k^j)}{\sum_{k'}(model_{2000}^j - counter_{k'}^j)}\left(\frac{model_{2000}^j - model_{1990}^j}{data_{2000}^j - data_{1990}^j}\right),$$

where $model_{1990}^j$ and $model_{2000}^j$ denote the model-predicted inequalities in 1990 and 2000 in job j, respectively. $data_{1990}^j$ and $data_{2000}^j$ denote the inequality in the data in these two years. $counter_k^j$

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 12. Contribution of each channel

Channel	Performance(%)	Match Quality(%)	Productivity(%)	Sorting(%)	Residual(%)	Total(%)
Contribution	37.6	51.9	−0.8	9.2	2.1	100.00

Notes: The contribution of factor k in any job is summarized in equation (16). The overall contribution of factor k across all the jobs is computed as the average of its contribution in each job weighted by the employment share.

Table 13. Contribution of performance pay and match quality

Channel	Performance(%)	Match Quality(%)	Performance+Match Quality(%)
Contribution	37.6	51.9	96.2

Note: We either set one channel or both channels to their values in 1990 and compute the contribution to changes in the wage inequality in each case.

is the inequality obtained from the counterfactual economy, wherein we restore the job-specific factor k from their values in 2000 to those in 1990.

Overall, we evaluate four contributing factors from the model: the performance-pay incidence, match quality, job-specific productivity, and sorting-induced changes in abilities and employment shares. The fraction that cannot be accounted for in the model is due to "residuals" arising from missing factors, which is small in the benchmark economy.

Finally, the overall contribution of factor k, denoted as $\bar{\pi}_k$, is an average of its contribution in each job weighted by the employment share of each job

$$\bar{\pi}_k = \frac{\sum_j \pi_k^j N_j}{\sum_{k'} \sum_j \pi_{k'}^j N_j}, \tag{16}$$

where N_j is the employment share in job j.

Table 12 reports the overall contribution of each channel. The contributions of the performance-pay incidence, match quality, job-specific productivity, and sorting are 37.6%, 51.9%, −0.8%, and 9.2%, respectively. Overall, the model mimics the data counterpart fairly well with a nearly zero residual component. The changes in the performance-pay incidence and match quality can capture almost 90% of the changes in within-job wage inequality from 1990 to 2000. Finally, the role of general job productivity is inconsequential.

One may inquire why job sorting or job-specific productivity does not widen within-job inequality by much. On the one hand, once match quality is assured and workers are offered correct incentives by performance pay with better ones sorted into performance-pay positions, the role played by positive assortative sorting across jobs substantially decreases. On the other hand, job-specific productivity resembles Solow residuals in neoclassical aggregate production. Thus, upon incorporating match quality and performance pay, this residual productivity adds little explanatory power to the underlying drivers in the model.

Interactions between performance-pay incidence and match quality One may wonder when both the channel of performance-pay incidence and match quality are incorporated, and whether they tend to reinforce or offset each other. Here, we address this question by either setting both the performance-pay incidence and match quality, or setting one of them to their levels in 1990. As shown in Table 13, where both are set to their levels in 1990, the joint contribution to changes in the overall wage inequality is 96.2%. This implies a positive amplification effect between the two channels by 6.7%,[21] which also echoes our findings in Proposition 4.

In summary, once the extensive margin via performance pay and intensive margin via match quality are incorporated, the contribution of the remaining extensive margin via job sorting becomes modest, accounting for only less than 10% of the rising within-job wage inequality from 1990 to 2000. Moreover, general job productivity—which has been emphasized in most studies—is inconsequential throughout when performance-pay and match quality are considered. Thus,

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 14. Decomposition by industry

Industry	Performance(%)	Match quality(%)	Productivity(%)	Sorting(%)	Residual(%)
Goods	30.6	52.5	0.5	10.5	6.0
Transport	24.4	65.0	0.3	10.3	0.0
FIRE	30.3	64.4	−2.3	7.7	0.0
Business	40.8	52.0	−1.4	8.6	−0.0
Trade	52.9	40.7	−1.8	8.2	0.0
Aggregate	37.6	51.9	−0.8	9.2	2.1

Notes: The five industries are (1) Goods (containing durables/non-durables, construction and mining), (2) Transport (transportation and utility), (3) FIRE, (4) Business (business and professional services), and (5) Trade (whole/retail trade and personal services). The contribution of each factor in an industry is the average of its contribution in each job belonging to that industry weighted by the employment share.

while widening wage inequality may be a concern, it is less warranted if improved matching quality and better incentivized performance-pay contracts are the underlying drivers. It is particularly so because the modest role played by job sorting suggests that typical labor market search frictions are less likely to be the main source of the widening within-job wage inequality. As such, policymakers concerned about wage inequality should recognize the equity-efficiency tradeoffs because these two primary sources of wage dispersions within jobs are natural in the eyes of production efficiency. Our results also offer, on a more micro-structured basis, insights for a better understanding of the main findings in Jovanovic (2014); in an incomplete information world, this improved signaling quality about workers' true abilities may widen the within-the-group inequality. While improved match quality may be thought of as due partly to more informed signaling, performance-pay contracts may induce better incentives for workers, mitigating informational frictions.

5.2.1 Contribution by industry and occupation

To further examine the role of each channel, we have conducted similar decomposition exercises at the industry/occupation level. The contribution of each channel in a given group is computed as the employment-weighted average of the job-level contributions:

$$\bar{\pi}_k^g = \frac{\sum_{j\in g} \pi_k^j N_j}{\sum_{j\in g} N_j}$$

where j denotes the job index that belongs to the group g. For brevity, we focus on those with significant deviations in group outcomes from the overall aggregate outcomes. To evaluate the relative importance of each industry/occupation in driving the widening wage inequality, we have also computed the contribution of each industry/occupation to the increase in the overall within-job inequality. The brief algorithm is to first compute the changes in the wage inequality of each job from 1990 to 2000. The changes in wage inequality at the industry (occupation) level are then an employment-weighted average of job-level changes. The contribution of each industry (occupation) is then the fraction of the changes in each industry or occupation for the changes in the overall within-job inequality. The results are reported in Table B.2 and B.3.

Table 14 presents the results of grouping by industry. Match quality is more important than the performance-pay incidence in almost all industries except for the Trade industry, which accounts for approximately 21% of the increase in the overall within-job inequality, as shown in Table B.2. The role of match quality is dominant and higher than the average contributions among all industries in the Transport and FIRE industries, which jointly account for another 19.3% of the increase in the overall within-job inequality. Conversely, performance-pay incidence in the Trade and Business industries contributes more than the average level among all industries, accounting

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 15. Decomposition by occupation

Occupation	Performance (%)	Match quality (%)	Productivity (%)	Sorting (%)	Residual (%)
Professional	32.8	46.3	−0.8	9.2	12.5
Manager	32.6	61.1	−1.8	8.2	0.0
Sales	40.6	62.5	−6.6	3.4	0.0
Production Labor	35.5	52.7	0.9	10.9	−0.0
Clerical	46.3	42.4	0.6	10.6	0.0
Aggregate	37.6	51.9	−0.8	9.2	2.1

Notes: The five occupations are (1) Professional, (2) Manager, (3) Sales, (4) Production/Operative Labor (including craftsmen), and (5) Clerical (including other services). The contribution of each factor in an occupation is the average of its contribution in each job belonging to that occupation weighted by the employment share.

for approximately 44.7% of the changes in the overall within-job inequality. That is, incentive provision is more crucial for explaining the widening inequality in the Trade and Business industries. The contribution of sorting is relatively larger in the Goods and Transport industries, exceeding 10%.

Table 15 presents the results of grouping by occupations. Specifically, in the Clerical occupation, which accounts for 21.8% changes in the overall within-job inequality as shown in Table B.3, performance pay is more important than match quality in driving up the wage inequality, Meanwhile, in all remaining occupations, the contribution of match quality is the largest among the different channels. In Manager, Sales, and Production Labor occupations, which together can explain 61.2% of the changes in the overall within-job inequality, the contribution of match quality is higher than the average level. In only two occupations, Production Labor and Clerical, the contribution of job sorting exceeds 10%.

Tables 14 and 15 show that, among all industries and occupations, the Professional occupation is the only category where the model cannot capture a noticeable portion (12.5%) of the widening inequality. In the Goods industry, the residual accounts for a mild 6%. Meanwhile, in all other categories, the residual shares are all negligible.

5.2.2 Contribution by wage, employment, or inequality rankings

In the benchmark economy, jobs are sorted by average wages. The rankings of jobs may change from 1990 to 2000. In our exercise, being ranked first implies that the job has the lowest average wage among all jobs. Next, we group jobs with similar changes in ranking to examine how the role of each channel varies by category. If the job's wage bin rises, this job stands at a higher ranking in 2000 than in 1990. That is, there are more jobs with a lower average wage than in 1990. Table 16 presents the decomposition results grouped by wage bin changes.[22] For jobs whose average earnings rankings rise/drop moderately (by one or two) or remain unchanged, which account for more than 85% of changes in the overall within-job inequality as shown in Table B.4, match quality and performance pay jointly contribute to the widening wage inequality. For jobs whose average earnings rankings rise/drop sharply (by three or more), which contribute to approximately 12% of changes in the within-job inequality, match quality plays a more dominant role with contributions higher than the overall average; meanwhile, performance pay tends to be more important than in other categories.

Similarly, we can group jobs by the rankings of their employment share; this also changes from 1990 to 2000, with a first-place ranking representing the lowest employment share. We again group jobs with similar changes in ranking. If the job's employment bin rises, there are more jobs with a lower employment share than the specific job in 2000 than in 1990. Table 17 presents the results grouped by employment bin changes.[23] For jobs whose employment share rankings rise, which accounts for about 15.3% changes in the overall within-job inequality as shown in Table

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Table 16. Decomposition by average wage bin change

	Performance(%)	Match quality(%)	Productivity(%)	Sorting(%)	Residual(%)
Rise by 3 or more	19.1	91.0	−10.1	−0.1	0.0
Rise by 1 or 2	43.0	51.5	−2.3	7.7	0.0
Stay	40.5	48.8	0.4	10.4	0.0
Drop by 1 or 2	25.0	49.1	−0.3	9.7	16.5
Drop by 3 or more	28.0	59.8	1.1	11.1	0.0
Aggregate	37.6	51.9	−0.8	9.2	2.1

Note: We group jobs with similar changes in ranking. If the job's average wage bin rises, then there are more jobs with a lower average wage than this job in 2000 than in 1990. The contribution of each factor in a category is again the weighted average of its contribution in each job within that category.

Table 17. Decomposition by employment bin change

	Performance(%)	Match quality(%)	Productivity(%)	Sorting(%)	Residual(%)
Rise	34.1	61.0	−2.6	7.4	0.0
Stay	37.7	54.1	−0.9	9.1	0.0
Drop	39.1	43.3	0.4	10.4	6.8
Aggregate	37.6	51.9	−0.8	9.2	2.1

Note: We group jobs with similar changes in the ranking of the employment share. If the job's employment bin rises, then there are more jobs with a lower employment share than this job in 2000 than in 1990. The contribution of each factor in a category is again the weighted average of its contribution in each job within that category.

B.5, match quality plays a dominant role. For jobs whose employment share rankings remain unchanged, which comprise more than 50% of contributions to changes in the overall within-job inequality, the dominant role of match quality diminishes. Finally, its role continues to decline for jobs whose employment share rankings fall, where sorting contributes more than 10% of changes in the within-job inequality.

5.2.3 Contribution in selected jobs

Finally, we investigate six selected jobs based on their large earnings and employment shares (see Table B.6 in the Appendix). Match quality contributes more than performance-pay incidence in almost all industries except for the Trade industry, which accounts for about 20% of the changes in the overall within-job inequality. In Transport and FIRE industries, which jointly account for another 20% of the changes in the overall within-job inequality, match quality plays a dominant role. In the Business industry, the contribution of performance-pay incidence is higher than the average level among all industries. At the occupation level, performance-pay incidence is more important than match quality only in the Clerical occupation, which accounts for approximately 22% of the changes in the overall within-job inequality. In Manager, Sales, and Production Labor occupations, which together account for more than 60% of the changes in the overall within-job inequality, the contribution of match quality is higher than the average level.

6. Conclusion

We document that the within-job component accounts for the lion's share of residual wage inequality, which itself has been the primary source of overall wage inequality. We examine four plausible channels widening within-job wage inequality, and identify match quality and performance pay as the two key drivers. We shed light on the primary source of residual wage inequality within jobs. While widening wage inequality may be a concern, it is less warranted if improved matching and more prevalent performance pay are the underlying drivers.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Some interesting future research avenues arise along these lines. First, the model can be extended to discuss the underlying reasons for skill mismatch. In particular, a worker may be employed in a mismatched job because of search frictions, or the trade-off between personal interests and earnings. Incorporating these underlying reasons, the model can generate richer policy implications. Second, the model can also be extended to include multiple dimension skills to study the effects of multi-tasking on wage inequality. With rising demand for multi-tasking, the dimension of skill requirements expands and that of job specialization declines. While a worker who fits the job or a special task well may have a higher wage, such a highly specialized skill may turn them into a loser both in multi-tasking requirements and in subsequent re-sorting. Thus, within-job wage inequality may change as the requirements for multi-tasking change in different types of jobs.

Acknowledgements. An earlier version of this paper was circulated under the title "Within-job Wage Inequality: Performance Pay and Job Relatedness", NBER Working Paper 27390. We thank Daniel Parent, Carl Sanders and David Wiczer for sharing their codes and Gaetano Antinolfi, Sangmin Aum, Michele Boldrin, Fatih Guvenen, Tim Lee, Rodolfo Manuelli, B. Ravikumar, Raul Santaeulalia-Llopis, Yongseok Shin, Guillaume Vandenbroucke, and David Wiczer, as well as an editor and two anonymous referees for helpful comments. We have also benefited from comments by participants at the Asian Meeting of the Econometric Society, China Meeting of Econometric Society, Midwest Macroeconomics Meeting, North American Meeting of Econometric Society, QMUL-SUFE Workshop in Economics, and Taipei International Conference on Growth, Trade and Dynamics. Rongsheng Tang is grateful for the financial support from the National Natural Science Foundation of China (Grant No. 72373088).

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1365100525100631.

Notes

1 See pivotal work by Violante (2002) and Jovanovic (2014) and follow-up studies.
2 See Figures 1 and 2 and the discussion in Section 2.2 below. We call a cell of industry-by-occupation as a job for two reasons. First, workers in practice choose both the industry and occupation a prior. Second, there are papers studying within-industry and within-occupation inequality independently, and both groups claim the significance. Thus it seems important to better-understand the interaction of industry and occupation.
3 See Figure 3 and the discussion in Section 2.3.
4 The reader is referred to our NBER working paper where we show the robustness of main findings under the second approach.
5 The original two-digit code is consistent in 1983–2002 and 2003–2018.
6 For reference, we present statistics for all workers in Appendix A.1. In particular, Table A.1 and Table A.2 present mean and standard deviation of wages and the number of observations for 80 jobs up to 1990 and 2000, respectively, and Table A.3 and Table A.4 present the results for 25 jobs up to 1990 and 2000, respectively.
7 For reference, we present statistic description for high-skill workers only in Appendix A.1. In particular, Table A.5 and A.6 present mean and standard deviation of wages and the number of observations for 80 jobs up to 1990 and 2000, respectively, and Table A.7 and Table A.8 present for 25 jobs up to 1990 and 2000, respectively.
8 We take the negative of the Rotter scale because a lower score implies more feeling of self-determination.
9 An earlier version of this paper (Tang et al., 2020) presents details on how to construct the job relatedness measure.
10 Specifically, the formula to compute the years of experience is max (age year of schooling-6,0).
11 In total, there are 25 jobs. The five industries are: (1) Goods (containing durables/non-durables, construction, and mining), (2) Transport (transportation and utility), (3) FIRE, (4) Business (business and professional service), and (5) Trade (whole/retail trade and personal services). The five occupations are: (1) Professional, (2) Manager, (3) Sales, (4) Production/Operative Labor (including craftsmen), and (5) Clerical (including other services).
12 While we only report the results in 1990 and 2000 as they are used in quantitative analysis, the pattern is true for all years with available data.
13 We execute the analysis as a pooled panel regression by ignoring the time dimension since our sample periods are not continuous and relatively short. Table A.13 presents the regression result while sequentially controlling for fixed effects.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

14 Specifically, we combine the performance-pay index (pp) and mismatch index (mm) in industry-occupation level with the CPS data and compute "Recentered Influence Functions" (rif) on the residual log value of wage after controlling for the observed variables, denoted as w:

$$rif(var) = (w - mean(w))^2 - var(w).$$

We then regress rif(var) on pp and mm on individual-level data in 1990 and 2000 in Table A.15. The results suggest that a one percentage point increase in performance-pay incidence raises the wage inequality 8 percentage points, whereas a one percentage point increase in the mismatch index suppresses the wage inequality by 2.7 percentage points. In Table A.16, we have done this exercise with all years data and obtain similar results.

15 Monitoring aims to prevent workers in the performance-pay position from shirking. To simplify the analysis, we make the following two assumptions. First, with shirking, the worker's effort decreases to $(1 - \delta)$ of the optimal effort level. Second, we assume the following condition, which assures incentive compatibility for any job j: $1 > \delta^2 > \frac{2\mu b M_j}{[\mu(1-\widetilde{\alpha}_j)\widetilde{A}_j A_j]^2}$. The condition is met in our quantitative analysis.

16 We use this general notation for proof although $z(a)$ is independent of a.

17 Recall that $\{\widetilde{C}_n\}$ are defined in Equation (10). In the remainder of this section, we relabel the job index as $n = 1, 2..J$ to denote the ranking of jobs in the sorting equilibrium.

18 We may not exactly match both moments since the estimation is subject to the constraint $\theta_{sj} > 2$.

19 In particular, $\{M_j\}$ must be sorted in ascending order.

20 [20]It overpredicts by about 30%. Table B.1 provides further details.

21 It is computed as $96.2 - (37.6 + 51.9)$.

22 To facilitate the comparison, we also compute the contribution of each category to the changes in the total within-job inequality similar to the approach in Section 5.2.1. The results are reported in Tables B.4 and B.5.

23 Each category's contribution to the changes in the overall within-job inequality can be found in Table B.5

References

Acemoglu, D. (2003). Patterns of skill premia. *The Review of Economic Studies* 70, 199–230.

Acemoglu, D. and D. Autor. (2011). Skills, tasks and technologies: implications for employment and earnings. *Handbook of Labor Economics* 4, 1043–1171.

Akerman, A., E. Helpman, O. Itskhoki, M.-A. Muendler and S. Redding. (2013). Sources of wage inequality. *American Economic Review, Papers and Proceedings* 103, 214–219.

Altonji, J.G., L.B. Kahn and J.D. Speer. (2016). Cashier or consultant? Entry labor market conditions, field of study, and career success. *Journal of Labor Economics* 34, S361–S401.

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

Bandiera, O., A. Kotia, I. Lindenlaub, C. Moser and A. Prat. (2024). Meritocracy across countries. NBER Working Paper 32375. https://doi.org/10.3386/w32375.

Barth, E., A. Bryson, J.C. Davis and R. Freeman. (2011). The contribution of dispersion across plants to the increase in US earnings dispersion. Working paper.

Card, D., J. Heining and P. Kline. (2013). Workplace heterogeneity and the rise of west German wage inequality. *The Quarterly Journal of Economics* 128, 967–1015.

Diamond, R. and J.C.S. Serrato. (2025). Spatial sorting and inequality. In *Handbook of Regional and Urban Economics*, Vol. 6 of Handbook of Regional and Urban Economics, Elsevier, pp. 463–513.

Dorn, D. (2009). Essays on inequality, spatial interaction, and the demand for skills, Ph.D. thesis, Verlag nicht ermittelbar.

Erosa, A., L. Fuster, G. Kambourov and R. Rogerson. (2024). Wage and earnings inequality between and within occupations: the role of labor supply. *Journal of Monetary Economics* 145, 103553.

Firpo, S., N.M. Fortin and T. Lemieux. (2009). Unconditional quantile regressions. *Econometrica* 77, 953–973.

Galindo da Fonseca, J.A., G. Gallipoli and Y. Yedid-Levi. (2020). Match quality and contractual sorting. *Labour Economics* 66, 101899.

Guvenen, F., B. Kuruscu, S. Tanaka and D. Wiczer. (2020). Multidimensional skill mismatch. *American Economic Journal: Macroeconomics* 12, 210–244.

Hall, R.E. and P.R. Milgrom. (2008). The limited influence of unemployment on the wage bargain. *American Economic Review* 98, 1653–1674.

Heywood, J.S. and D. Parent. (2012). Performance pay and the White-Black wage gap. *Journal of Labor Economics* 30, 249–290.

Heywood, J.S. and D. Parent. (2017). Performance pay, the gender gap, and specialization within marriage. *Journal of Labor Research* 38, 387–427.

Jensen, M.C. and K.J. Murphy. (1990). Performance pay and top-management incentives. *Journal of Political Economy* 98, 225–264.

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press

Jovanovic, B. (2014). Misallocation and growth. *The American Economic Review* 104, 1149–1171.
Juhn, C., K.M. Murphy and B. Pierce. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy* 101, 410–442.
Kambourov, G. and I. Manovskii. (2009). Occupational mobility and wage inequality. *The Review of Economic Studies* 76, 731–759.
Katz, L. F. and D.H. Autor. (1999). Changes in the wage structure and earnings inequality. *Handbook of Labor Economics* 3, 1463–1555.
Kirkeboen, L.J., E. Leuven and M. Mogstad. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics* 131, 1057–1111.
Krusell, P., L.E. Ohanian, J.-V. Ríos-Rull and G.L. Violante. (2000). Capital-skill complementarity and inequality: a macroeconomic analysis. *Econometrica* 68, 1029–1053.
Lazear, E.P. (2000). Performance pay and productivity. *The American Economic Review* 90, 1346–1361.
Lemieux, T. (2006). Increasing residual wage inequality: composition effects, noisy data, or rising demand for skill? *The American Economic Review* 96, 461–498.
Lemieux, T., W.B. MacLeod and D. Parent. (2009). Performance pay and wage inequality. *The Quarterly Journal of Economics* 124, 1–49.
Lise, J. and F. Postel-Vinay. (2020). Multidimensional skills, sorting, and human capital accumulation. *The American Economic Revew* 110, 2328–2376.
Makridis, C. (2018). *Work and Grow Rich: The Dynamic Effects of Performance Pay Contracts*. Available at SSRN 3308540.
Makridis, C.A. and M. Gittleman. (2022). On the cyclicality of real wages and employment: new evidence and stylized facts from performance pay and fixed wage jobs. *The Journal of Law, Economics, and Organization* 38, 889–920.
Mueller, H.M., P.P. Ouimet and E. Simintzi. (2017). Wage inequality and firm growth. *American Economic Review* 107, 379–383.
Papageorgiou, T. (2011). *Large Firms and Internal Labor Markets*. Pennsylvania State University, mimeo.
Piketty, T. and E. Saez. (2003). Income inequality in the United States, 1913–1998. *The Quarterly Journal of Economics* 118, 1–39.
Ritter, J.A. and K.L. West. (2014). Field of study and earnings inequality among the highly educated: 1993–2010. Working paper.
Robst, J. (2007). Education and job match: the relatedness of college major and work. *Economics of Education Review* 26, 397–407.
Sanders, C. (2014). Skill Uncertainty, Skill Accumulation, and Occupational Choice. Working Paper.
Scotese, C. A. (2012). Wage inequality, tasks and occupations. Working Paper.
Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review* 95, 25–49.
Sockin, J. (2022). Show Me the Amenity: Are Higher-Paying Firms Better All Around? CESifo Working Paper, 9842.
Song, J., D.J. Price, F. Guvenen, N. Bloom and T. Von Wachter. (2019). Firming up inequality. *The Quarterly Journal of Economics* 134, 1–50.
Tang, R., Y. Tang and P. Wang. (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* 117, 297–338.
Wallskog, M., N. Bloom, S.W. Ohlmacher and C. Tello-Trillo. (2024). Within-Firm Pay Inequality and Productivity. Working Paper 32240, National Bureau of Economic Research.

Cite this article: Tang R, Tang Y and Wang P (2025). "Within-job wage inequality: the nexus between performance pay and match quality." *Macroeconomic Dynamics* **29**(e158), 1–34. https://doi.org/10.1017/S1365100525100631

https://doi.org/10.1017/S1365100525100631 Published online by Cambridge University Press