Posted Compensation Inequality*

Xuanli Zhu†

December 7, 2022
(Preliminary and Incomplete)
Latest Version Here

Abstract

This paper studies the supply of non-wage compensations across different types of firms and jobs, as well as their impact on wage determination. Taking advantage of the data from a Chinese online job board where most firms document pecuniary and nonpecuniary amenities to attract applicants, we extract a large set of non-wage compensations that employers and jobseekers recognize as essential for their matching in the labor market. We find that different firms in different jobs also provide different non-wage compensations in a systematic way. In particular, high wage-premium firms sorted with high skill jobs also more likely provide advanced insurance packages, backloading wage and stock options, professional coworkers, and flexible work-time, and such amenities are positively correlated with posted wage. In contrast, low wage-premium firms sorted with low skill jobs more likely to offer weekend, holiday, and regular work-time, and such amenities are strongly subject to compensating differential. To account for these findings, we propose a new theory that combines the compensating differential with efficiency compensation and firm-worker sorting .

Keywords: compensating differential, wage inequality, firm-worker sorting

*I am very grateful to Tetsuji Okazaki for his guidance throughout the development of this project. I thank Daiji Kawaguchi and participants in the seminar of University of Tokyo and the “Labor, Firms, and Macro” Workshop for valuable feedback and comments. This paper is split from the previous version of the working paper "Posted Wage and Compensation Inequality", which is further previously circulated with the title "Job Task Variation, Firm Wage Premium, and Firm-Provided Compensation".

†University of Tokyo, Graduate School of Economics. Email: zhuxuanli46@gmail.com.

1 Introduction

Labor economists have long recognized the importance role that non-wage compensations play on the wage determination in the labor market. The term "compensating differential" (or "equalizing differences") refers to the straightforward idea that a firm that provides an amenity can offer a lower wage than an otherwise similar firm and still hire workers because those workers who benefit in utility from enjoying the amenity more than offsetting the pecuniary lose in their wage would prefer to work in that firm, and that the similar but inverse argument applies for a disamenity. In the classical model of compensating differential, variations in the provision of different types of compensation are generated by firms' heterogeneity in the compensation production function, and conditional on workers' heterogeneity, we should observe the effect of equalizing differences from the wage differences associated with the different levels of compensations provided. Despite this intuitive prediction, the empirical tests of the theory in the literature often result mixed results, with some coefficients of the hedonic regression in the direction supporting the theory but others being zero or in inverse direction, even under a specification with worker fixed effects to control for unobserved worker characteristics (Brown, 1980; Rosen, 1986). The culprit of this empirical failure in the early periods have been argued to be the identification problem of unobserved confounding variables. More recently, several studies (see, among others, Sorkin, 2018; Lamadon et al., 2022) use structural modelling to indirectly confirm the existence of compensating differential through a perspective of revealed preference and stress its importance in justifying on wage determination and worker mobility. However, it is fair to say that we still lack the knowledge about a large portion of the overall picture of the compensation provision and their impact on the labor market. The questions are: what are the primary non-wage compensations provided by firms that matter importantly for wage and worker's job choice? Do different firms provide different types of compensations on a systemic way? If so, then how and why? Is compensating differential the only channel that links non-wage compensation with wage differences or is there any other channels?

In this paper, we try to fill this gap first by using a new type of data to provide several new empirical facts on the provision of non-wage compensations and on their correlations with the wage differentials, and second by constructing a new theory to extend classic compensating differential theory so that we can reconcile our findings with the theory. On the empirical side, we utilize the online job vacancy data from a Chinese online job board to extract all the terms relevant to non-wage compensations that employers include in their job advertisements in order to attract their ideal workers.[1] The virtue of the combination of this data and approach is that we do not need to determine what compensations to investigate ex-ante, but rather let the data tell us what are the major non-wage compensations in the labor market that employers and jobseekers recognize as essential. The set of non-wage compensations we find in our data covers a large amount of different types of compensations, ranging from pecuniary ones like backloading payment, bonus, and stock options to nonpecuniary ones like insurance, worktime flexibility, fringe benefits, etc. Although for each vacancy or firm the set of non-wage compensations observed in our data is not necessarily a full list of the compensations that the job or

[1] In particular, we apply the same extraction procedure, which employs several machine learning algorithms, as the one in our companion paper Zhu (2022), where we use the same data to investigate the components of posted wage inequality.

the firm offer, these non-wage compensations mentioned in the job vacancies are arguably the most important ones because they are used by firms to attract potential workers or to justify their posted wage through compensating differential.[2]

Our empirical analysis has three main findings. First, while many non-wage compensation terms have strong predictive power on the posted wage, the share of the total wage variance accounted for by these terms as whole is rather small after controlling for all the skills and tasks documented in the job advertisement and firm fixed effect, and their explanatory power is primarily derived from their covariances with job controls and firm fixed effect. Put it differently, these non-wage compensation can predict the posted wage largely because they can indicate the job quality and firm wage premium, which are the main drivers of the wage inequality. This thus suggests that different firms in different jobs also systematically provide different non-wage compensations. In light of this result, we then directly examine the occurrence frequency of a set of different types of non-wage compensations by different types of firms and jobs, which leads to our second finding: non-wage compensations are provided differently in different firms and jobs in a systematic way. Specifically, we find that high wage-premium firms sorted with high skilled jobs are also more likely to provide advanced insurance package, backloading wage and stock options, and high qualified coworker and flexible work-time, but less likely to provide weekend, holiday, and fixed work-time. Whereas the inverse is true for low wage-premium firms sorted with low skilled jobs. Thirdly, we run a hedonic regression on these compensations with a full control of job characteristics and find that those amenities that high wage-premium firms are more likely to provide are significantly and positively related with the posted wage, whereas those amenities that low wage-premium firms are more likely to offer have a significantly negative correlation with the posted wage. We argue that these stylized facts reported in our data and in some other studies cannot be explained by the standard compensating differential model.

In light of these empirical evidences, we propose a new theory that combines the compensating differential with two other elements, efficiency compensation and firm-worker sorting, and show that it helps to reconcile these puzzling findings.[3] Essentially, as long as we accept the idea that many non-wage compensations can be efficient or inefficient in production or firm operation, there will be an additional efficiency channel along with the traditional equalizing differential channel when firms decide their levels of compensation. This new channel can work in either the inverse or the same direction to the compensating differential mechanism, and to what extent it matters depends on the level of the firm-worker sorting. For those compensations that are efficient like advanced insurance or backloading wages, and in those

[2]By construction, the non-wage compensations found in our data are those that firms recognize as attractable for the workers they are looking for. As a result, they are less likely to contain those compensations that the utilities vary hugely and can be either positive or negative across different workers due to personal preference. Example of such compensations include commuting time or some other personal preferences on the workplace which are rather random across workers. Although in some recent search models like Card et al. (2018) these idiosyncratic preference could be important for job moves and wage inequalities, to what extent that different workers vary their preferences on non-wage compensations in general is an empirical question worth future investigation. To simplify the analysis, for the empirical investigation here and for the theoretical model introduced later we will either implicitly or explicitly assume that workers have the similar preference on the non-wage compensations.

[3]Here by "efficiency compensation", we mean the efficiency aspects of compensations that is similar to the ones proposed in the efficiency wage theory: eliciting effort, reducing labor turnover costs, etc. In fact, we argue in **??** that this is a more natural property of non-wage compensations than monetary wage.

high-pay firms sorted with high-skilled workers and jobs, the efficiency effect will be large and can even dominate the compensating differential mechanism so that the firms that provide better compensations will not reduce but increase their workers' wage. Whereas in low-pay firms with low-skilled workers and jobs, such efficiency effect is low and if the compensation is mandated and its costs cannot be fully equalized from the efficiency benefits, it will be compensated from a reduction in the workers' wage. On the other hand, inefficient compensation like generous work-time or work-load will cause a large efficiency loss in those high pay-premium firms and high quality jobs, resulting that only low pay-premium firms and low quality jobs will bear the cost and provide such amenities. Our new theory can thus generate flexible patterns of compensation provision and wage impact, and thus provide important implications that help to better understand the labor market inequality. It explains why in some cases a hedonic wage regression, even well identified, could produce results that are inconsistent with the prediction of the classic compensating differential theory. It also gives some insights on how non-wage compensation provision can affect the overall compensation inequalities in the labor market, and what could be the key types of compensations that matter for certain job mobilities in the labor market.

Related Literature. Our paper contributes to the historical and recently resurgent literature of compensating differential. As shown by the classic paper of Rosen (1986), firms can provide different levels of amenities or disamenities to compensate their wage cost and workers will sort into different packages of wage and compensation to maximize their utility. In spite of the theory's intuitive idea and straightforward predictions, early empirical studies that run hedonic wage regressions on different types of compensations often lead to mixed results with both supporting and counterintuitive evidences, even with worker fixed effects (see the survey in Rosen, 1986). More recently, two different types of studies start to revive this topic. First, several recent empirical studies show evidences for compensating differential by using experimental or quasi-experimental methods to identify the wage effects of certain types of compensations in specific situations (see e.g. Mas and Pallais, 2017; Wissmann, 2022, among others). Second, through a perspective of revealed preferences, a few studies begin to model the labor market by using unobserved compensation as a wage wedge to justify job moves, especially for those moves to low wage-premium firms with wage loss observed in data (e.g. Card et al., 2018; Sorkin, 2018; Taber and Vejlin, 2020; Lamadon et al., 2022). We contribute to this literature through two aspects. Firstly we provide new empirical evidences on firms' non-wage compensation provision and their impact on wage determination by taking advantage of the online vacancy data, in which firms document their most important non-wage compensations to attract potential workers. In our data, we discover a large set of pecuniary and nonpecuniary compensations including insurance, backloading wage, stock option, coworker quality, training, weekend and holiday, and flexible work-time among many others, all of which hold predictive power on the posted wage. Moreover, we find high wage premium firms and low wage premium firms have distinguished patterns in the provision of non-wage compensations. In particular, those high wage premium firms are also more likely to offer advanced insurance, larger backloading wage, stock option, and better coworker quality, and that these amenities are not compensated by the posted wage. On the other hand, low pay premium firms are more likely to provide basic insurance and less work-time or more rest days to attract poten-

tial workers, and they equalize the costs of these amenities by reducing their posted wages. These findings are consistent with several recent studies that regress the provision of non-wage compensations on the firm fixed effect obtained from the AKM approach and find evidences of high-wage premium firms also providing better non-wage compensations (Sockin, 2022; Bana et al., 2022). Such a positive relationship between wage premium and non-wage compensation provision are, however, at odds with the prediction of compensating differentiation theory.[4] The second aspect of our contribution is to build a new theory that can reconcile these stylized facts. In particular, we combine two elements observed in our findings, efficiency wage and firm-worker sorting, with the traditional compensating differential mechanism and show that this new theory can generate flexible results on firms' compensation provision and impacts on the wage determination.[5]

The outline of this paper is following. First in Section 2 we use the vacancy data from China to show some empirical facts about the provision patterns of the non-wage compensations observed in our data as well as their relationship with the posted wage. We suggest that the classical theory of compensating differential cannot explain our empirical findings. Consequently, in Section 3 we construct a new theory which extends the canonical compensating differential mechanism with two new elements, namely efficiency compensations and firm-worker sorting, both of which have been observed in the data and long discussed in the literature. We show that this new theory can generate flexible provision patterns and wage impact of compensations and reconcile all the stylized facts that we find in our data. We conclude in Section 4.

2 Empirical Facts From Vacancy Data

2.1 Data Source and Processing

Our main data used for the empirical analysis is the online job vacancy data from a Chinese online job board, Lagou.com. It is the most popular national information technology (IT)-centered online job board in China and holds a large customer base of both IT-producing and IT-using firms. In total we collected from the site over 6 million job posts between 2013 and 2020, and compile them into a large cross-sectional sample. Although one third of the job

[4] In particular, the theory of compensating differential is largely based both the heterogeneous cost functions of compensation provision across and the heterogeneous preferences on these compensations across workers. The compensations that the economists often had in mind at the time when the theory was built are job injury, job mortality, or workplace pollutions that are no longer the major concerns in today's labor market. For the compensations found in our data, many are pecuniary and thus firms have exactly the same cost function. Also, for nonpecuniary compensation like health insurance, Dey and Flinn (2005) shows that the cost function of providing the insurance is likely to be similar across different employers. In a similar vein, workers are likely to have similar preference on these amenities although to what extent is an empirical question. Therefore, the theory of compensating differential will fail to generate systematic differences in non-wage compensation provision in the recent labor markets.

[5] The efficiency nature of alternative pay schemes have long been argued in the organizational literature, see for example Lemieux et al. (2009). And the efficiency of nonpecuniary compensation has been also argued in Dey and Flinn (2005), where the authors suggest that offering health insurance can be efficient for the employers through reductions in exogenous worker exit.

posts belong to IT professional occupations like IT engineers or programmers, there are also two thirds of the job vacancies ranging from other professional occupations like designers, writers, business operation specialists, and financial analysts to less professional or low-skilled occupations like sales and marketing clerks and custom service operators. Therefore, our data includes a large set of routine or non-routine cognitive jobs, and contains very little pure manual jobs. Given the secular trends of fast technological change like automation and persistent structural transformation from manufacturing to service all over the world, we think the labor market we focus here is the typical and major labor market in many countries in both the current era and future era.

Like many other online job vacancy data that have been studied in the literature, our collected data contains both job information like job name, wage range, job requirements and tasks, and etc., and firm information like firm name, firm size and industry, and etc. More importantly for our purpose is one often ignored information in the analysis of vacancy data—the information about non-wage compensations and amenities that firms claim in the job text to attract potential applicants. The major advantage of our data here is that most of the jobs in our data contains this information, partially due to the fact that there is a column of "job benefit" for firms to fill in when they post their job vacancies and partially due to the fact that most firms seem to do find it important to write down such information for their recruitment. Although the compensations claimed in the vacancy data do not constitute the full package of non-wage compensations and amenities provided by the firms, we suggest that they are likely the most important ones in the labor market perceived by firms and thus allow us to study the patterns of firm compensation provision and its impact on the wage determination and earning inequality in the labor market. Despite the richness of the information on job amenities in our data, the main difficulty of the empirical analysis, though, is to distill useful and tractable information from the raw texts that document those information in natural language so that we can conduct further econometric estimations.

Our information extraction procedure contains three steps. First, we use the least absolute shrinkage and selection operator (Lasso) regression to extract the words or terms that hold predictive power on posted wage. The selected terms mix those terms of non-wage compensations and amenities with a large set of skills and tasks that also documented in the job raw texts. To distinguish these different types of terms, in the second step we train a natural language processing (NLP) model—the word embedding model—on all the job vacancy texts to capture the potential relationships of all the terms within the context. We then use an unsupervised clustering algorithm—K-means clustering—to separate those compensation terms with other skill and task terms. Our final step is to first generate an indicator matrix of all the compensation terms that we select in previous steps and then use a dimensional reduction algorithm, partial least squares regression (PLS), to generates a low-dimensional representation of this indicator matrix. In the following analysis, we will utilize both the original indicator variables and the dimensional reduced proxy variables.

For the details of the data collection, data processing, and information extraction, we refer to our companion paper Zhu (2022).

2.2 Stylized Facts

Our first stylized fact is simply about the set of non-wage compensations discovered in our job vacancy text data. Given its large size, we illustrate this set by two ways. First, we show the terms of non-wage compensations that hold strong predictive power (high absolute value of coefficient) for the posted wages in our Lasso regression. In fact, in the results of our Lasso regression we find that a bunch of compensation terms appear in the top among all the terms including a huge amount of skills and tasks, and there are consistent patterns across the results using different samples of different occupations. For positive top tokens, we can find backloading compensation (e.g. "14th month pay"), fringe benefits (e.g. "three meals"), advanced insurance and fund (e.g. "six insurance & one pension"), coworker quality (e.g. "guru", "maestro"), and equity compensation (e.g. "stock", "options"). For negative top tokens, we can see compensation terms of mandated insurance (e.g. "five insurance", "social insurance"), leisure time (e.g. "two-day weekend", "holiday"), and also fringe benefits (e.g. "accommodation").[6] This result thus echos the confusing results in the compensation differential literature: the estimated coefficients from the hedonic regression are mixed and sometimes inconsistent with the theoretical prediction. However, one can notice that many compensations represented by the positive tokens are performance pay or fringe benefits that potentially encourage effort, long-hour or inflexible worktime, and learning, and prevent turnover cost. On the other hand the compensation in the negative tokens seem to indicate that the compensations related to work-life balance follows the classic compensation differential mechanism. Also, the case of insurance and fund provided by firm is quite interesting: the basic mandated level of insurance is negatively correlated with posted wage while the enhanced package of insurance and fund is positively correlated with posted wage, running exactly inverse to the theory of compensation differential. These features suggest that the efficiency of different non-wage compensations might be an important aspect when thinking about non-wage compensation provision in the labor market. One possible reason that these non-wage compensations hold such strong power on posted wage prediction is that they are actually correlated with firm effect. However, this argument will suggest that firms have very different strategies for compensation provision even when they are likely to have similar cost functions for providing these non-wage compensations, so that these differently provided compensations are correlated with the posted wage in a distinctive way. It thus raises the question of how and why firms decide the different packages of wage and non-wage compensations for their workers and the question that if there are some mechanisms other than compensation differential working in the labor market. We will try to examine and answer these questions using both some empirical analysis below and the theory in Section 3.

Our second way to illustrate the non-wage compensations extracted from our machine learning algorithm is to check the cluster that gather all the terms relevant to non-wage compensations or job characteristics through the K-Means clustering in the word embedding space.

[6]The most representative form in Chinese social insurance system are "five insurance and one fund". "Five insurance" means endowment insurance, medical insurance, employment insurance, employment injury insurance and maternity insurance and is mandated by law. "One fund" means housing provident fund, which is not compulsory by law but a large percent of formal firms, especially those large sized, will pay this fund for their workers. "Six insurance" means five basic insurance and one additional commercial supplementary medical insurance, which is only provided by a few well-paid firms. In some rare cases we can also observe "seven insurance" or "two fund" which basically indicates further advanced insurance or fund support.

In this cluster we also find many synonyms of the compensations in the top tokens that we have mentioned in the Lasso results, for example "business insurance", "five-day workday", and many similar compensations, for example "seven insurance & one fund", "two fund", "bonus", "tea time", "gym", "taxi", as well as some other typical compensations in the literature for example "flexible worktime", "overnight shift". The general picture of the compensation provided by firms are thus close to what we conclude from the top Lasso features: backloading payment and bonus, insurance & fund, worktime and leisure, fringe benefits, and learning or training environment.[7]

Our second stylized fact is about to what extent and through what channels the information of these non-wage compensations can account for posted wage variations. Put it differently, we want to study how the compensating differential of all extracted non-wage compensations as a whole impact the posted wage differentials after controlling for other important wage determinants like job or firm quality. To this end, we add the dimensional reduced proxy variables of the indicator matrix of the non-wage compensation cluster, Ξ_1, to an otherwise typical log wage regression with specification:

$$\ln w_i = X_i\beta + \psi_j + \delta_i + \iota_t + \epsilon_i \tag{1}$$

, where the wage w_i is the mean of the posted wage range for each vacancy i. X_i are the proxy variables of all skills and tasks extracted from the vacancy text in the same way as non-wage compensations, and we denote $\theta_i \equiv X_i\beta$ as the job quality or job effect. ψ_i is the firm fixed effects which indicate time-invariant firm pay premiums, and ι_t is the year effects. Our interest is $\delta_i \equiv \Xi_{1,i}b$, which is the product of the dimensional reduced proxy variables for all compensations Ξ_1 and its corresponding coefficients b at vacancy level. Under the logic of compensation differential, and by assuming similar preference on any specific non-wage compensation across different workers, the value of δ_i should represent the part of wage that is differential equalized in each job vacancy due to the compensation provision. More specifically, because in our case firms are generally more likely to document amenities rather than disamenities, the value of δ_i would indicate to what extent the posted wage of a vacancy is discounted due to the non-wage compensations provided by this job. Therefore, a low value of δ_i estimated in our case means that the amenities provided by the employer of this job are highly valued by the potential job-seekers and thus justify a large discount in the posted wage of this job. On the other hand, a high value of estimated δ_i means that the job amenities offered can bring only limited utility for the potential workers and thus cannot act as much compensation for the posted wage, or even that the set of compensations are in net disamenities so that wage should rise to equalize the potential worker's loss in utility.

[7] Given the nature of the job vacancy text, we would generally not have terms of disamenities in our vocabulary because firms will not voluntarily claim the cons of their job in the vacancy. As a result, part of the compensation that have been examined in the previous literature and are for sure disamenities like work injury and safety would not be taken into account here. However, most jobs in our sample, and also a majority of the jobs in the recent labor markets of middle-income or rich countries are office jobs, and thus typically not subject to those kinds of absolute disamenities in the traditional mining or construction industry. Also, many other amenities could have priorly undetermined level of benevolence due to the varying preference of workers or firms, and thus could be found in our data. Considering the general trend of technology advance and work environment improvement, we think our data illustrates a major part of non-wage compensations and amenities that are provided by firms in recent days and are recognized by both firms and workers as important factors of hiring in the labor market.

Table 1: Wage Variance Decomposition With Compensation

	Pooled		Computer		Design_Media		Admin	
	Comp.	Share	Comp.	Share	Comp.	Share	Comp.	Share
$\mathrm{Var}(\ln w)$	.362	-	.281	-	.254	-	.164	-
Panel A: $\delta_i \equiv \Xi_{1,i}\beta^c$								
$\mathrm{Var}(\theta_i)$	.158	.437	.079	.282	.082	.324	.063	.385
$\mathrm{Var}(\delta_i)$	.002	.004	.001	.003	.001	.002	.001	.006
$\mathrm{Var}(\epsilon_i)$	.097	.269	.074	.262	.070	.277	.057	.349
$\mathrm{Var}(\psi_j)$	.046	.128	.066	.234	.052	.207	.026	.161
$2\,\mathrm{Cov}(\theta_j, \psi_j)$	.049	.137	.051	.181	.041	.160	.011	.066
$2\,\mathrm{Cov}(\delta_i, \theta_i)$	.006	.017	.005	.018	.004	.015	.004	.027
$2\,\mathrm{Cov}(\delta_i, \psi_j)$	.003	.008	.006	.021	.004	.014	.001	.006
Panel B: Decompose $2\,\mathrm{Cov}(\delta_i, \theta_i)$								
$2\,\mathrm{Cov}(\delta_i, X_e)$	.002	.006	.002	.007	.002	.007	.002	.011
$2\,\mathrm{Cov}(\delta_i, \tilde{\Xi})$	.004	.011	.003	.011	.002	.009	.003	.016
$2\,\mathrm{Cov}(\delta_i, \Xi_g)$	.000	.001	.000	.001	.000	.001	.000	.001
$2\,\mathrm{Cov}(\delta_i, \Xi_m)$	.002	.004	.001	.003	.001	.004	.002	.012
$2\,\mathrm{Cov}(\delta_i, \Xi_s)$	.002	.006	.002	.007	.001	.005	.001	.003
Obs	3998840		1325260		548808		260364	
Firm	86165		62628		55664		41448	

Notes. Panel A shows the variance decomposition of the posted wage variance into job effect θ_i, firm effect ψ_j, compensation effect δ_i, and the interaction (i.e. sorting) of these effects along with an error part. Panel B further decomposes the covariance between job effect θ_i and compensation effect δ_i by splitting X_i into experience X_e and skills and tasks $\tilde{\Xi}$ and further splitting $\tilde{\Xi}$ into general skills Ξ_g, medium-specific skills Ξ_m, and most specific skills Ξ_s. We show all the results both for the pooled sample and for three subsamples with the typical high-, medium-, and low-skill occupations in our data.

The variance decomposition of the estimation results of Equation (1) are shown in Table 1. Panel A documents the components of variance decomposition according to

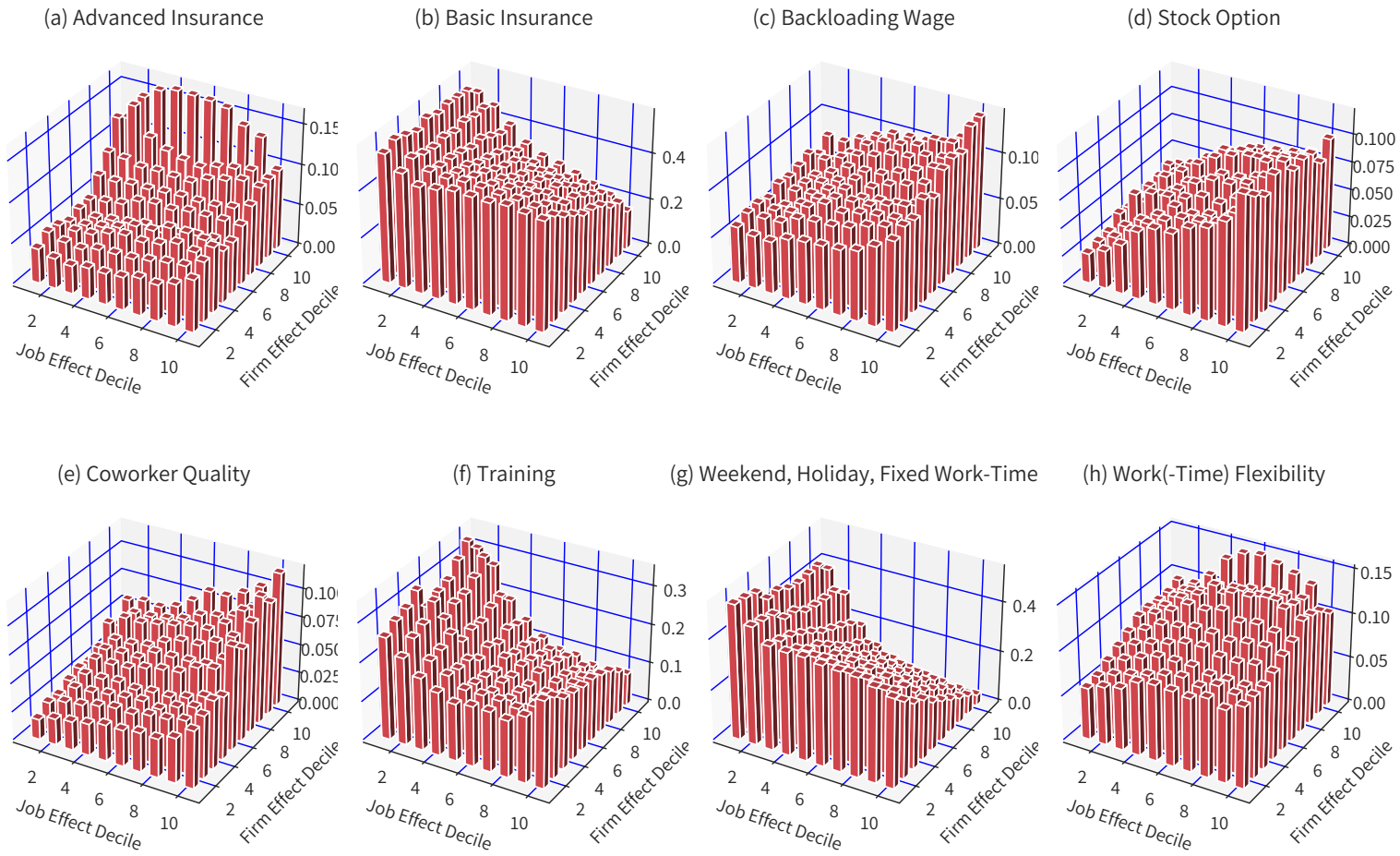
$$\text{var}(\ln w_i) = \text{var}(\theta_i) + \text{var}(\delta_i) + \text{var}(\psi_j) + 2\,\text{cov}(\theta_i, \psi_j) + 2\,\text{cov}(\theta_i, \delta_i) + 2\,\text{cov}(\psi_j, \delta_i) + \text{var}(\epsilon_i)$$

. And panel B further decomposes the covariance between the job effect and the compensation effect, $2\,\text{Cov}(\delta_i, \theta_i)$ into the covariance terms of different types of skills and tasks incorporated in X_i, including experience X_e and skills and tasks $\tilde{\Xi}$, where $\tilde{\Xi}$ can be further composed by general skills Ξ_g, medium-specific skills Ξ_m, and most specific skills Ξ_s. The results in panel A make it clear that the variation of non-wage compensation provision itself, i.e. $\text{Var}(\delta_i)$, holds very limited explanatory power for posted wage, accounting for only 0.2 to 0.6 percent of the for total wage variances. However, there are significant and positive relationship between compensation provision with both job effect and firm effect, as shown by the positive covariance terms $2\,\text{Cov}(\delta_i, \theta_i)$ and $2\,\text{Cov}(\delta_i, \psi_j)$ with shares ranging from 1 percent to 3 percent. These results indicate that our Lasso regression picks those compensation terms largely not because themselves have important impact on posted wage determination, but because that these compensation features can somehow indicate high quality jobs and high wage premium firms. In other words, high premium firms in high skill jobs somehow provide systematically different non-wage compensations. Also, if we again follow the logic of the compensation differential theory, these two positive correlations imply that high wage premium firms and high skill jobs are accompanied by amenities that have low values and thus are less compensated, while low wage premium firms with low skill jobs are more likely to provide amenities that worth more and thus are compensated more from posted wage.[8]

Our third stylized fact is about how different types of firms in different types of job provide systematically different set of non-wage compensations. This will give a more direct and clear picture of the compensation provision patterns across different types of firms and jobs, and help to inspect if the arguments following the compensating differential theory make intuitive sense in our data. To this end, we select a bunch of important compensation topics from our Lasso results and examine their occurrence ratios across different types of firms and jobs. In particular, we select eight genres of non-wage compensations that have terms show up with large absolute Lasso coefficients and/or are considered as important topics in the literature. These eight types of non-wage compensations are basic insurance, advanced insurance, backloading wage, stock and options, coworker quality, training, weekend, holiday and fixed work-time, and work-time flexibility. Then we find out all the synonymous in our vocabulary that indicate those genres by checking the group of terms with a small Gaussian distance in the embedding space constructed. Finally, we partition all the vacancies into the 10×10 job-firm joint decile cells and calculate the occurrence ratio of each compensation type for each cell by checking if its vacancy text contains any of the relevant terms. The generated occurrence distribution

[8] If we allow for large variations in idiosyncratic preference on non-wage compensations, we might also interpret the results as that high skill workers sorted with high wage premium firms value firm-provided amenities less whereas those low skill workers sorted with low premium firms value amenities more. Or if we allow for variations in firms' cost functions of non-wage compensation, we could interpret as that high premium firms are somehow more costly in providing high value amenities whereas low premium firms have a lower cost in providing amenities that are valued by their workers. In general, we can have both cases but if such substantial variations do exist is an empirical question with currently no strong evidences in the literature.

Figure 1: Compensation Occurrence in Pooled Sample



Notes. Job effects and firm effects here are the ones estimated using the specification in **??**. The occurrence ratio is calculated as the percentage of vacancies in each job-firm cell of which the vacancy text contains any of the terms related with a certain type of compensation. Basic insurance means five insurance and one fund, which is the most common compensation package in Chinese labor market. Advanced insurance means any other advanced package of insurance and fund which usually have additional business insurance or fund. Work flexibility relates to the work-time flexibility in most cases. See Figure A1 for the results of major occupation samples.

for the entire sample are shown in Figure 1, and we show in Figure A1 that the individual analysis for other single occupation sample depict largely similar patterns. For all eight types, we see compensation occurrence rate systematically changes along with either or both two axes. In particular, for advanced insurance, backloading wage, stock and option, coworker quality, and work-time flexibility, we observe that the occurrence increases in both the level of job effect and the level of the firm effect, although the extent to which effect matters more varies across compensation types. Conversely, for basic insurance and rest day and fixed work-time, their occurrence in job vacancy decrease significantly in both firm effect and job effect, and for training, the occurrence reduce strongly with job effect with ambiguous impact of firm effect.[9] In other words, our results suggest that high-pay firms with high-skill jobs are more likely to provide also better insurance and fund package, non-wage pecuniary compensations like backloading wage and stock option, and also nonpecuniary work place amenities of better coworkers and flexible worktime, whereas low-pay firms with low-skill jobs more often mention training and weekend, holiday, and fixed work-time as the amenities.[10] Our finding here thus largely contradicts our early interpretation of the positive relationship between compensating level and the levels of firm and job effects based on the theory of compensating differential. High wage premium firms also provide better non-wage compensation or amenities in many aspects, although they would less likely to offer training and leisure.

Our final stylized fact is about the wage impact of specific compensation types. To further examine the idea of compensating differential, we next follow the empirical literature of compensating differential and run a hedonic regression on the occurrence of those eight selected compensations. Similar to the specification in Equation (1), we now replace δ_i with the indicator matrix for the eight types of non-wage compensations that we have examined above. The identification is thus ensured by controlling for both the full set of proxy variables on heterogeneous skills and tasks and the firm fixed effects. In other words, our hedonic regressions control for almost all the information documented in the job vacancy about the potential candidate worker, and thus plausibly not subject to the problem of unobserved worker ability. The estimated coefficients for compensations in Table 2 show mixed evidences on compensating differential. For the compensations that are positively correlated with job effect and firm effect, i.e. advanced insurance, backloading wage, stock and option, coworker quality, and work-time flexibility, the coefficients are significantly positive in almost all cases. Whereas for

[9]The non-monotone relationship between job effect and training occurrence is because our method cannot distinguish that if the training terms mentioned in job text indicate receiving training or offering training. Actually after checking the raw data we find that the increase in training occurrence in the top deciles of job effect is completely due to these high-skill jobs require tasks of offering training to other workers in the firm. Although we can resolve this problem by applying more advanced NLP model to our text data, we argue that such case is relatively rare in our vacancy text data, and thus we stick with simpler method. The special pattern of training occurrence can be more clearly observed in the Computer occupation sample in Figure A1, where only the top two and bottom two deciles of job effect see a large increase while the middle deciles are generally flat.

[10]One note here is that our result does not necessarily mean that better firms with better jobs are less likely to provide basic insurance and fund package. This is because first obviously that such firms are more likely to offer advanced insurance package and thus correspondingly will not mention the basic package, and second that given that the basic insurance is compulsory for formal firms, high wage firms will generally not think it as an attractive compensation for their potential workers and thus not mention it even when they are actually providing it. We don't think a similar argument will go to the work-life balance because there are a large amounts of anecdotes on long working hours in many big and well-paid firms, and because income effect will make higher income workers prefer at least not less, if not more, leisure and so high pay premium firms can use it to attract workers if possible.

Table 2: Hedonic Regression on Selected Compensations with Full Controls

	Pooled	Computer	Design_ Media	Admin
	(1)	(2)	(3)	(4)
Advanced Insurance	.017**	.016**	.011**	.004
	(.001)	(.001)	(.002)	(.003)
Basic Insurance	-.026**	-.024**	-.018**	-.014**
	(.000)	(.001)	(.001)	(.001)
Backloading Wage	.009**	.012**	.022**	.011**
	(.001)	(.001)	(.002)	(.002)
Stock Option	.089**	.071**	.064**	.042**
	(.001)	(.001)	(.002)	(.004)
Commission	.029**	-.001	.003*	.032**
	(.001)	(.001)	(.002)	(.002)
Coworker Quality	.024**	.017**	.005*	.008*
	(.001)	(.001)	(.002)	(.004)
Training	-.001*	-.018**	-.002	.014**
	(.001)	(.001)	(.002)	(.002)
Work-Time	-.020**	-.019**	-.021**	-.022**
	(.000)	(.001)	(.001)	(.001)
Work Flexibility	.015**	.010**	.013**	.008**
	(.001)	(.001)	(.002)	(.002)
const	8.872**	9.155**	8.747**	8.336**
	(.002)	(.002)	(.004)	(.006)
Education FE	✓	✓	✓	✓
Experience FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
$\Xi_2, \ldots, \Xi_8$	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓
R^2	.743	.760	.757	.711
Adj. R^2	.738	.748	.730	.656
No. Obs	3998840	1325260	548808	260364

the compensations of basic insurance and work-time that are negatively correlated with job and firm effects, their coefficients are significantly negative.[11] Taking at the face value, these results indicate that those amenities provided by high wage premium firms in high skill jobs are not compensated at all but actually increase wage, while those amenities provided by low pay premium firms in low skill jobs are compensated from their posted wage. Therefore, our hedonic regression with detailed job controls produce results consistent with the findings in previous empirical studies: the mechanism of compensating different works in some cases, but in other cases we see exactly the inverse results that generates puzzles for the theory.

To sum up our empirical results, we find firms document a large amount of different types of non-wage compensations and amenities in their job advertisement to attract potential workers. Although the information of non-wage compensations itself explain little posted wage differential after controlling for all the job skills and tasks, we find that the provision of different types of compensations are correlated with job quality and firm wage premium. In other words, firms that have different levels of wage premium and post jobs with different levels of skills also distinguish in their non-wage compensations provision. In particular, those high wage premium firms sorted with high skilled jobs also provide many other pecuniary or nonpecuniary amenities including advanced insurance, additional payment, and high qualified work place, and these amenities are not compensated from posted wage but actually positively correlated with posted wage. In contrast, low wage premium firms sorted with low skilled jobs will provide basic insurance, training, and generous work-time as the job amenities, which are significantly compensated form posted wage. These empirical evidence provide hard challenges for the compensating differential theory, which claims that firms vary their provision on nonpecuniary compensation due to different cost functions of provision, and that workers select firms with different compensation through their heterogeneous preferences and allow their wage to be partially compensated for those compensations. For many pecuniary and nonpecuniary compensations provided by firms in our data, their cost functions are arguably similar for different firms, making it impossible to generate the strong linkages between their provision and firm and job effects. Similarly, it's a difficult empirical question that to what extent workers with different skills have different preference on these compensations. Actually if we believe that there is strong income effect on leisure, then it's hard to explain that why high wage premium firms sorted with high skill workers are substantially less likely to provide amenities of weekend, holiday, and less overtime, whereas low wage firms are more likely to provide such leisure to low income workers. And why for those amenities that they do generously provide unlike those low-pay firms, why do they not discount from their posted wages? In the next section, we suggest that compensation differential might not be the only force in the labor market for the provision of non-wage compensations and a new theory that takes efficient compensation and firm-worker sorting also into account can reconcile for all the empirical facts that we find here.

[11] The coefficients for training are misleading due to the reason that we talked earlier.

3 A New Theory

3.1 The Basic Idea

In this subsection, we suggest that the puzzle in the compensating differential literature, which is also occurred in our results, is not a problem of identification but a problem of incomplete theory. In particular, we argue that as long as we combine two additional elements, which are also observed in the literature and in our data, with the canonical mechanism of compensating differential, we can then generate patterns of compensation provision and different levels of compensating differential that are consistent with our empirical findings. The first new element is efficiency compensation, i.e. non-wage compensation can be efficient in production or in firm-operation.[12] The second new element is firm and worker sorting, or the firm and job sorting in our case. While the existence of the second element, sorting, have been confirmed by the recent literature on wage inequality and by the results here, the first element, the efficiency function of compensation, is often dismissed when empirically testing the impact of compensations on wage. Here we argue that the level of efficiency (or inefficiency) is a general and important feature of non-wage compensations.[13] First, it is not difficult to see the efficiency nature of those monetary compensations like backloading wage and stock option. In fact the literature have been long argued that alternative payment structure can help firm to improve efficiency through effort inducing, turnover reduction, and so on (see e.g. Lemieux et al., 2009). Similarly, it has been argued in the literature that health insurance and other insurance can reduce exogenous worker turnover (see e.g. Dey and Flinn, 2005), and that better coworker quality improves both production productivity and on-the-job learning efficiency in a complementary production setting (see e.g. Jarosch et al., 2021). In contrast, weekend, holiday, and less overtime or limited work duty are straightforward inefficient because they allow less work-time and effort. Other amenities like training or work-time flexibility are perhaps more unambiguous and if they are efficient or inefficient likely depends on the detailed cases.

A formal model setting and derivation of our new theory, which combines a simple framework of worker sorting with efficiency compensation, are documented in 3.2. For the rest of this subsection we briefly introduce the key ideas, intuitions, and implications of our new theory.[14]

[12] We call it "efficiency compensation" because it is analogous to the idea of efficiency wage theory, which suggests firms pay wages higher than market clearing level for various efficiency reasons such that it is optimal for the production or profit maximization. Actually we think efficiency compensation is even a more nature idea because one key critique on the efficiency wage theory is that firms should be able to take advantage of other non-wage compensations to achieve the same efficiency aim (see Katz, 1986).

[13] To be clear, in the canonical compensation differential the provision of a compensation can be also efficient or inefficient in production. However, the theory of compensation differential assumes that the sign of the impact of the compensation on the production be must inverse to the sign of its impact on the workers' utility. In other words, an amenity for the workers must cause a reduction in production productivity or a direct cost in production. Here we relax this restriction and allow an amenity to be either efficient or inefficient or having no impact on production at all.

[14] In fact, the setting in 3.2 is one of the simplest way, but not the only way, to generate the desired results, and there are many potential or further extensions that can be added to the basic framework. To distinguish with the traditional compensating differential model and to clarify our new mechanisms, in our model we assume workers are homogenous in their preference on all non-wage compensations and firms have the same direct cost functions on providing all compensations. However, both firms and workers are heterogeneous in their productivity, and they form pairs endogenously, and the joint production function is assumed to be supermodular—a necessary

The key idea is that when an amenity is allowed to be efficient in production, then in additional to the wage saving benefit, firm will also take the marginal product of efficiency improvement into account when considering the provision of a certain compensation. With firm-work sorting in the labor market, the level of this marginal production benefits from offering efficiency compensation will be larger in high wage premium firms that are sorted with high productivity workers or jobs. In other words, the better the firm or the job, more efficient will be the compensations. As a result, higher wage premium firms and higher wage jobs are more likely to provide those efficiency compensations, and because increase in productivity will often at least partially translate into increase in wage, this efficiency gain act in counter to the classical compensating differential mechanism. And if the level of the compensation has a large span and the marginal product does not decline too fast, it is also possible that the efficiency channel dominates the compensating differential above some threshold of firm and worker level, generating positive wage effect in net, i.e. firms providing better compensations now cause wage increase rather than wage decrease. In contrast, firms with lower wage premiums and sorted with low productive workers or jobs are less likely to provide efficient compensations because the marginal production benefits are small. And when firms do provide such compensations in some cases, say basic insurance that is mandated by the government, their net loss between the provision cost and the efficiency effect, if any, will be equalized through reduction in wage, and the lower is the rank and productivity of the firm and the work, the severe is the level of compensating differential. Therefore, our theory can generate the feature that while an amenity is significantly compensated from wage by low pay firms in low pay jobs, the same or even a superior amenity is not compensated from but actually positively correlated with wage in high wage firms and high skill jobs. The similar logic can be applied to compensations that are inefficient, say generous work-time or work-life balance.[15] Under complementarity, the higher the productivity and rank of the firm and worker, the larger the efficiency loss coming from the provision of such compensation.[16] Consequently, as long as the income effect on leisure is not too strong, high wage premium firms and jobs will not provide such compensations, but rather compensate workers for their utility loss with higher wage. On the other hand, such efficiency cost is small when the firm and the job have low rank and low productivity, and thus low wage premium firms with low skill jobs are more likely to document such inefficient compensations for attracting workers. In other words, now the efficiency channel is in the same direction as the compensating differential channel, and the impact of firm-worker sorting on the efficiency channel in fact act as an amplifier for compensating differential. Finally, when a compensation

condition to generate positively assortative matching between firms and workers in the economy. Compensations provided by firms are assumed to be either efficient or inefficient, i.e. they affect an efficiency terms of the firms' production which acts as another complementary input in the production function. We show that this simple and parsimonious setting that contains efficiency compensation and sorting is enough to generate rich features of compensations provision and different levels of compensating differential. More realistic models can be constructed by adding heterogeneous worker preference or search frictions so that the sorting becomes no longer monotone or perfect.

[15]It is arguable that in some cases generous rests like paid leave or maternity leave can be actually efficient if they help to retain workers and the turnover cost is very high. In fact Bana et al. (2022) find the in the U.S. high wage premium firms are more likely to participate in Paid Family Leave programs and have lower turnover rates. However, it could be a difficult empirical question to answer ex-ante that if an amenity like this is efficient or not.

[16]Note that in additional to the linkage with the firm-worker match productivity, such inefficient compensation also offsets the effect of other efficient compensations.

is neither efficient nor inefficient, the efficiency channel shuts down, and the model returns back to the traditional compensating differential model.

Our new theory have three implications that are important for understanding the labor market inequality in wage and non-wage compensation. First, the efficiency aspect of different pecuniary or nonpecuniary compensations could be the key to dissolve the puzzle that is brought by the mixed results found in the empirical tests for the theory of compensating differential. As our new theory shows, the efficiency effect can totally offset the effects of equalizing differential and generate results inverse to the predictions by the compensating differential theory. Our theory thus predicts that while it might be not difficult to find the clear evidences for compensating differential in the submarket with low-pay firms and low-skill workers, the similar evidences will be hard to find when targeting to the high-end labor market or the entire labor market. Also, directly adopting the estimation results and conclusions found from a particular compensation in a particular labor market to other compensations and other labor markets could be dangerous and misleading. Second, with firm-worker sorting and efficiency compensation, the labor market inequality could be underestimated by just looking at wage or monetary payments. The high-skill workers employed in high wage premium firms are likely to also enjoy the best non-wage compensations in many aspects, including both additional earnings from bonus and stock and nonpecuniary amenities like better insurance or fringe benefits, though at the expense of high effort. Perhaps more surprisingly, our theory suggest that the provision of compensations can not only generate inequality in non-wage compensation itself but also further enlarge the wage inequality. This is because efficiency compensations can simultaneously increase the workers' direct utility on non-wage compensations and increase workers' wage through a boost in their productivity. In other words, efficiency compensations work as an amplifier for the labor market inequality at both observed wage level and observed utility level. Third, our theory suggest that the set of the unobserved non-wage compensations that drive the large amount workers' moving to low-wage premium firms will be rather limited (see Sorkin, 2018; Bonhomme et al., 2019). In fact, our theory suggest that these compensations must be inefficient ones like less work-time because high-wage premium firms will also provide better efficient compensations. Moreover, a worker that goes down the firm ladder due to some changes in preferences for certain amenities like leisure will suffer not only a worse matching but also a downgrading on many other efficient compensations, both of which will negatively affect the wage that the worker receive.[17]

[17]In fact in the section 5.4 of Rosen (1986), Rosen suggests an application of the compensation differential theory as "hours of work (or work schedules more generally) may be formally treated as nonpecuniary aspects of jobs. Then the market transaction must be viewed as a tie-in in which a firm offers a fixed wage-hours package to workers, take it or leave it, with these package deals varying from firm to firm". He then suggests two sources for the equilibrium distribution of different packages generated in the labor market: coordination in production or set up costs. Our idea of labor market sorting as the source for heterogeneous provision of working hour and wage packages is close to the idea of coordinating production, but different from the classic compensation differential framework that Rosen suggest, in our argument the interpersonal differences in productivity affect the equilibrium allocation not only through the resulted heterogeneity in preference but also through firms' opportunity cost of offering such "inefficient" compensations. In addition, the nonpecuniary aspect of job we consider here can be more general and contains not only hours of work but also latent effort.

3.2 A Simple Model

We now build a simple model with firm-worker sorting and efficiency compensation and show that the results derived can be consistent with the results we find in our data and other similar results found in the recent literature. In comparison to the canonical compensation differential model, in our model workers and firms are not heterogeneous in their tastes of and cost functions of various peculiarity and nonpecuniary compensations, rather they are heterogeneous in their productivity like the typical assignment model.[18] The variation of the firms' provision of a certain compensation comes from different efficiency (or inefficiency) levels of the compensation in different firm-worker pairs. In fact, the traditional mechanism of compensation differential still exists but is now offset by a new efficiency mechanism when the compensation is efficient or magnified when the compensation is inefficient. The level of this new efficiency channel depends on the level of firm-worker sorting. We show that this model can thus generate flexible results on the wage impact of compensation provision that could be both consistent and inconsistent with the predictions of the traditional compensating differential model. We also show that this new theory can generate important implications for understanding labor market inequalities.

In the economy there is a continuum of workers with same utility function $U(C, a, h)$ but heterogeneous productivity $q \in [0, 1]$, where C is the monetary consumption, $a \in \{0, 1\}$ is the indicator of if the worker receiving a nonpecuniary amenity, e.g. the firm's provision of a certain insurance, and h is the level of a nonpecuniary disamenity, e.g. the requirement on additional working hour. To simplify the analysis, we assume the utility function takes an additively separable form,

$$U(C, a, h) = C + \phi_a a - \frac{h^{1+\phi_h}}{1+\phi_h} \tag{2}$$

, with $\phi_a, \phi_h > 0$. We assume that there is no income other than wage, thus $C = w(q)$, and the level of amenity a and disamenity h are decided by the firm that the worker matched in the equilibrium.

The economy also has a large continuum of potential firms that are ex-ante homogenous and want to hire workers. To facilitate sorting or separation, we assume that these firms face the same O-Ring type production technology, i.e. a production function with complementarity across all labor positions.[19] In addition, we extend the original O-Ring production function with efficiency compensation so that the labor productivity and the output Y_j of a firm j will also depend on the firm's choices on its supply of the compensations. In particular, the production

[18] A more general analysis with heterogeneous worker preferences (and heterogeneous firm cost functions of compensation provision) requires to extend the problem to multidimensional matching. However, this will significantly increase the analysis tractability, and thus we leave it for future research.

[19] An alternative way to establish sorting is to generate two-dimensional matching by assuming an exogenous distribution of firms with heterogeneous productivity and assuming a pair-wise production function with complementarity. One advantage of our one-dimensional matching setting is that we can easily generate endogenous firm size and have the firm size correlated with firm level productivity, as what we find the data and in the literature. The choice of this alternative setting and our setting does not affect our main results on compensation provision. More generally one can have both two-side heterogeneity and endogenous firm size, see Eeckhout and Kircher (2018).

function takes the form:

$$Y_j = AN_j^{1+\alpha} \prod_{i=1}^{N_j} q_i e(a,h) \tag{3}$$

, where A is a common productivity, N_j is the total number of worker the firm hires, and $e(\cdot)$ is an efficiency function which is increasing and strictly concave in both arguments. To simplify the exposition, we also assume the function $e(\cdot)$ takes an additively separable form:

$$e(a,h) = 1 + \gamma_a a + \frac{h^{\gamma_h}}{\gamma_h}$$

, where $\gamma_a, \gamma_h \in [0,1)$ control the decreasing return for these two efficiency compensation. This also helps us to illustrate that even when the decisions on the level of different amenities and disamenities are irrelevant by themselves, their occurrence could be still correlated through their relationships with the firm productivity. We further assume that for the amenity a, firm will pay the cost, which is a per-worker cost κ multiplied by the total number of worker N if $a = 1$. In comparison, for the disamenity h, a firm does not pay any direct cost but need to pay a higher wage w to compensate the loss in worker's utility. For analytical tractability, we also assume a fixed N for all firms but will show later that relaxing N to be another endogenous firm choice does not change our results.

The competitive equilibrium in this economy is defined as an assignment of worker types to firms and a utility schedule, $u(q)$, such that (i) given the utility schedule, all active firms maximize their profits by employing their workers in a way consistent with the assignment and by choosing the wages for their workers, $w(q)$, and the levels of two (dis)amenities, a and h, and that (ii) the labor market clears for workers of all productivity levels. The competitive equilibrium here coincides with the stable matching of the assignment problem: the workers' utility schedule and the firms' profit schedule is on the possibility frontiers and there does not exist other assignments that can generate larger payoffs.

We then characterize the competitive equilibrium by analyzing the firm's profit maximization problem:

$$\begin{aligned} \max_{\{q_i\}_{i=1}^N, a, h, w(q)} \; & AN^{1+\alpha} \prod_{i=1}^{N} q_i e(a,h) - \sum_{i=1}^{N} w(q_i) - a\kappa N \\ \text{s.t.} \quad & w(q) + \phi_a a - \frac{h^{1+\phi_h}}{1+\phi_h} \geq u(q) \quad \forall q \in \{q_i\}_{i=1}^N \end{aligned} \tag{4}$$

. The profit possibility frontier for the firm is $v(q_1, \ldots, q_N, u)$ with the utility compatibility constraint holding in equality and $a, h, w(q)$ chosen optimally. The derivative of v with respective to each single q_i is $\frac{dv}{dq_i} = AN^{1+\alpha} e(a,h) \prod_{i' \neq i}^{N} q_{i'} e(a,h) > 0$, and thus v is type increasing. It is then easy to see that the cross partial derivatives $\frac{d^2 v}{dq_i dq_{i'}} > 0$ and $\frac{d^2 v}{dq_i du} = 0$, and thus the equilibrium allocation in the economy satisfies positive assortative matching (PAM) and in our case this means all workers employed by any single firm will have the same type, q.[20] Under perfect

[20] In the original O-Ring mode there is perfect transferable utility, and thus the complementarity across different labor inputs in the O-Ring production function, i.e. $\frac{d^2 v}{dq_i dq_{i'}} > 0$, will be enough ensure that in the equilibrium assignment is PAM. However, the amenity and disamenity terms in the worker's utility function generate imper-

segregation, the firm's problem in Equation (4) now can be written as

$$\max_{q,a,h} AN^{1+\alpha}q^N(1+\gamma_a a+\frac{h^{\gamma_h}}{\gamma_h})-N\left(u(q)-\phi_a a+\frac{h^{1+\phi_h}}{1+\phi_h}\right)-a\kappa N \tag{5}$$

. Because the amenity a is a discrete choice, there is a productivity threshold q_a such that the firm that hires workers with q_a will be indifferent between providing or not providing amenity a. In particular the optimal provision strategy will be

$$a=\begin{cases}1, & \text{if } q\geq q_a\\ 0, & \text{if } q<q_a\end{cases}, \text{ and } AN^{\alpha}q_a^N\gamma_a+\phi_a=\kappa \tag{6}$$

This threshold is decided from the equation $AN^{\alpha}q_a^N\gamma_a+\phi_a=\kappa$, i.e. the marginal benefit of providing a equals the marginal cost of providing a. Note that if the amenity a is not efficient at all, i.e. $\gamma_a=0$, then it can return back to the canonical compensation differential where the dispersion of preference ϕ_a and of the cost κ generate sorting between workers and firms. As a result, more productive firm with more productive workers are more likely to offer these efficient compensations. The importance of this channel can be seen more clearly when the cost of the amenity is increasing in the level of the worker, as the case of many insurance and fund. If we assume the per-worker cost of a is actually $p\kappa$, then the cost of providing such amenity increases in firms which employ high q workers and pay high wage w. As a result with the traditional compensating differential mechanism alone, high rank firms are less likely to pay for such amenity, which is inconsistent with the empirical facts found in the literature. However, when a non-wage compensation is efficient and there is enough productivity dispersion across firms and workers, the differences in the efficiency effect could dominate and derive the discrepancy in compensation provision. However, with the efficiency channel, the increase in q has increased marginal benefit from efficiency effect more than covering the increased provision cost, generating positive relationship between firm productivity and compensation provision.

The first order conditions for the rest of two maximization choices q and h are

$$AN^{1+\alpha}q^{N-1}e(a,h)=u'(q), \tag{7}$$

$$AN^{\alpha}q^N h^{\gamma_h-1}=h^{\phi_h} \tag{8}$$

respectively. It's clear from Equation (8) that the optimal level of disamenity $h=\left(AN^{\alpha}q^N\right)^{\frac{1}{1+\phi_h-\gamma_h}}$ is also increasing in productivity q due to the same efficiency reason as amenity a. Therefore, more productive firm will also require high level of disamenity h. However, in this case, this disamenity will be fully compensated by the increase in the wage. To obtain the market wage, we first derive the market utility profile by replacing the optimal efficiency level $e(a,h)$ and

fectly transferable utility and as a result an additional condition $\frac{d^2v}{dq_i du}\geq 0$ is required for segregation assignment to be achieved in the equilibrium. For more details about the sufficient conditions for monotone matching in an economy with assignment problem see Legros and Newman (2007)

then integrate Equation (7) over the entire distribution of worker productivity:

$$u(q) = \begin{cases} \frac{\left(\bar{A}q^N\right)^{1+\omega}}{(1+\omega)(1+\gamma_h)} + (1+\gamma_a)\bar{A}q^N + u_a, & \text{if } q \geq q_a \\ \frac{\left(\bar{A}q^N\right)^{1+\omega}}{(1+\omega)(1+\gamma_h)} + \bar{A}q^N + u_0, & \text{if } q < q_a \end{cases} \tag{9}$$

, where $\bar{A} \equiv AN^{\alpha}$, $\omega = \frac{1+\gamma_h}{1+\phi_h-\gamma_h}$, and u_0 and u_a is the constant of integration. In fact u_0 will be the utility or wage that workers of $p = 0$ obtain, and thus is pinned down by free-entry condition such that $u_0 = 0$. Similarly, u_a is pinned down by the firm indifference at p_a such that $u_a = \phi_a - \kappa$. Finally, the market wage profile can thus be derived from Equation (9) as

$$w(q) = \begin{cases} \bar{A}q^N + \gamma_a\bar{A}q^N - \kappa + \frac{\left(\bar{A}q^N\right)^{1+\omega}}{(1+\omega)(1+\gamma_h)} + \frac{\left(\bar{A}q^N\right)^{\omega}}{1+\gamma_h}, & \text{if } q \geq q_a \\ \bar{A}q^N + \frac{\left(\bar{A}q^N\right)^{1+\omega}}{(1+\omega)(1+\gamma_h)} + \frac{\left(\bar{A}q^N\right)^{\omega}}{1+\gamma_h}, & \text{if } q < q_a \end{cases} \tag{10}$$

. The first term $\bar{A}q^N$ is the wage function when there is no non-wage compensation, which, perhaps not surprisingly, increases in productivity p, indicating that firms with high q workers also provide higher wage. The second part $\gamma_a\bar{A}q^N - \kappa$ in the case $q \geq q_a$ is a combination of the increase in wage due to efficiency effect and the compensation differential on a. Note when $q = q_a$ this term is $-\phi_a$, i.e. the workers' utility benefits from firms' provision of a is fully compensated from the reduction in wage, and thus for some range of the productivity $p > p_a$, it will generate a wage plunge comparing to firms that have a close productivity of workers but do no provide amenity a. However, going up the firm rank, the efficiency effect will increase in productivity q, and it's possible that the provision of the amenity a are not compensated from wage reduction at all but actually generate wage gain for the workers. The last two terms, $\frac{\left(\bar{A}q^N\right)^{1+\omega}}{(1+\omega)(1+\phi_h)}$ and $\frac{\left(\bar{A}q^N\right)^{\omega}}{1+\phi_h}$, are the increase in wage due to efficiency effect and compensation differential on h. In this case because h is an amenity and high rank firms demand more h, this generates a positive compensating differential on the wage. Therefore, our model shows that in a standard setting of firm-worker sorting or segregation, an efficient non-wage compensation, whether amenity or disamenity, could further enlarge market wage inequality, especially when the dispersion of productivity is large across firms. In other words, non-wage compensation can be not only an unobserved and overlooked labor market inequality, but also potential drivers behind the increasing in observed wage or earning inequality.

Next we discuss two more implications of our model on the empirical estimation of compensating differential. First, note for q close to q_a, one can find clear evidence of compensation differential by examining wage difference and controlling for worker characteristics. However, if a is a multiple discrete choice or close to an continuous choice, higher level of a will be correlated with higher wage, counteracting to the force of equalizing differential and thus confounding the estimation results. Also, if we assume that at some point in time the government mandate the firms' provision of amenity a (again for example the enforcement of a certain insurance) but the enforcement is not perfect. This will not change anything for all firms with worker productivity $q \geq q_a$, but it now requires all firms with $q < q_a$ also provide a. As a result, the compensation in wage reduction for these firms will be $\gamma_a\bar{A}q^N - \kappa$ and this reduction

is larger for firms with lower q, providing larger incentive for these firms to circumvent the provision of a. The empirical estimation will thus find strong evidence of compensation differential for this firms, but again the higher the labor market with firms and workers with high productivity, the smaller will be the equalizing effect and in some cases be even negative. Second, given that firms gathered with high productivity workers are likely to both provide high levels of efficiency compensations and require high levels of efficiency disamenities, a worker deviated from in the common utility structure for certain amenities will have to also deviate from the current optimal matching and be subject to changes in receiving other compensations. For example, if a middle-age female worker with a high productivity q gives a birth and thus has a large increase in disamenity in working hour (i.e. an increase ϕ_h), she has to go down the job ladder and match to a firm in which workers have productivity lower than q. Moreover, the level of other non-wage compensations might be also downgraded if these compensations are efficient. Therefore, in additional to the part of usual equalizing differential, the cost of such compensation differential will also incorporate the wage decline due to a worse matching and the utility decline due to a less generous package of other compensations.

Finally, we consider the case when N is also a choice of the firm. The additional first order condition with respect to N in this case is

$$AN^{\alpha}q^{N}e(a,h)\left(1+\alpha+N\ln(q)\right)=w+ac \tag{11}$$

. Further differentiating Equation (11) with respect to q and evaluated at the optimal level, we obtain the optimal choice on firm size:

$$N(p)=\frac{1+\alpha}{-\ln(p)} \tag{12}$$

. This result shows that the firm size increases in p and is irrelevant to the choices of amenities. Therefore, all the relationships we have found between productivity and amenity provision can be now directly translate to the firm size.

4 Conclusion

In this paper, we take advantage of online job advertisement data, where firms document their non-wage compensations and amenities to attract workers, to document several new empirical findings about firms' compensation provision behavior. Most importantly, we find while high wage premium firms sorted with high quality works or jobs are more likely to also provide many other non-wage compensations like advanced insurance or stock option, low wage premium firms sorted with low quality worker or jobs are more likely to provide weekend, holiday, and fixed work time. The classic compensating differential theory does not provide explanations for such distinguished behavior in compensation provision. We also find the puzzling results that the compensations that high wage premium firms provide are positively correlated with the posted wage, which is at odds with the prediction of compensating differential theory, although the inverse is true in for low wage premium firms, which supports the existence of equalizing differential. To reconcile these stylized facts, we suggest a new theory which extends

the classic mechanism of compensating differential with an additional channel of efficiency compensation, of which the extent depends on the level of firm-worker sorting. We use a simple model to show that our new theory can not only reconcile all the empirical findings we find, but also have important implications on the labor market inequalities in terms of both wage and non-wage compensations. For the purpose of tractability, our model is rather stylized and perhaps over-simplified in that we assume homogenous worker preference, homogenous firm production function, and perfect assortative matching. One potential future work is to further generalize the model so that we can bring the model to the data.

References

Bana, S., K. Bedard, M. Rossin-Slater, and J. Stearns (2022). Unequal use of social insurance benefits: The role of employers. *Journal of Econometrics*.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework for matched employer employee data. *Econometrica 87*(3), 699–739.

Brown, C. (1980). Equalizing differences in the labor market. *The Quarterly Journal of Economics 94*(1), 113–134.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(S1), S13–S70.

Dey, M. S. and C. J. Flinn (2005). An equilibrium model of health insurance provision and wage determination. *Econometrica 73*(2), 571–627.

Eeckhout, J. and P. Kircher (2018). Assortative matching with large firms. *Econometrica 86*(1), 85–132.

Jarosch, G., E. Oberfield, and E. Rossi-Hansberg (2021). Learning from coworkers. *Econometrica 89*(2), 647–676.

Katz, L. F. (1986). Efficiency wage theories: A partial evaluation. *NBER macroeconomics annual 1*, 235–276.

Lamadon, T., M. Mogstad, and B. Setzler (2022). Imperfect competition, compensating differentials, and rent sharing in the us labor market. *American Economic Review 112*(1), 169–212.

Legros, P. and A. F. Newman (2007). Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities. *Econometrica 75*(4), 1073–1102.

Lemieux, T., W. B. MacLeod, and D. Parent (2009). Performance pay and wage inequality. *The Quarterly Journal of Economics 124*(1), 1–49.

Mas, A. and A. Pallais (2017). Valuing alternative work arrangements. *American Economic Review 107*(12), 3722–59.

Rosen, S. (1986). The theory of equalizing differences. *Handbook of labor economics 1*, 641–692.

Sockin, J. (2022). Show me the amenity: Are higher-paying firms better all around? *CESifo Working Paper*.

Sorkin, I. (2018). Ranking firms using revealed preference. *The quarterly journal of economics 133*(3), 1331–1393.

Taber, C. and R. Vejlin (2020). Estimation of a roy/search/compensating differential model of the labor market. *Econometrica 88*(3), 1031–1069.

Wissmann, D. (2022). Finally a smoking gun? compensating differentials and the introduction of smoking bans. *American Economic Journal: Applied Economics 14*(1), 75–106.

Zhu, X. (2022). Posted wage inequality. *Working Paper*.

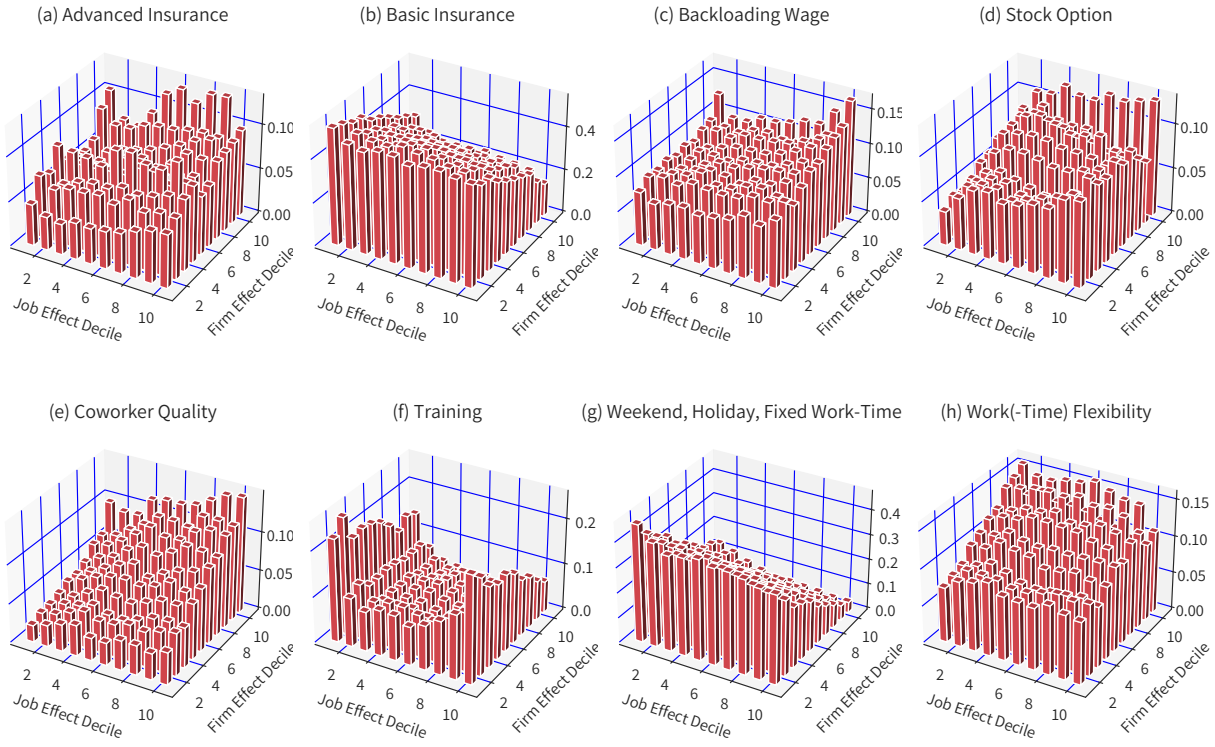
Appendices

Appendix A. Additional Tables And Figures

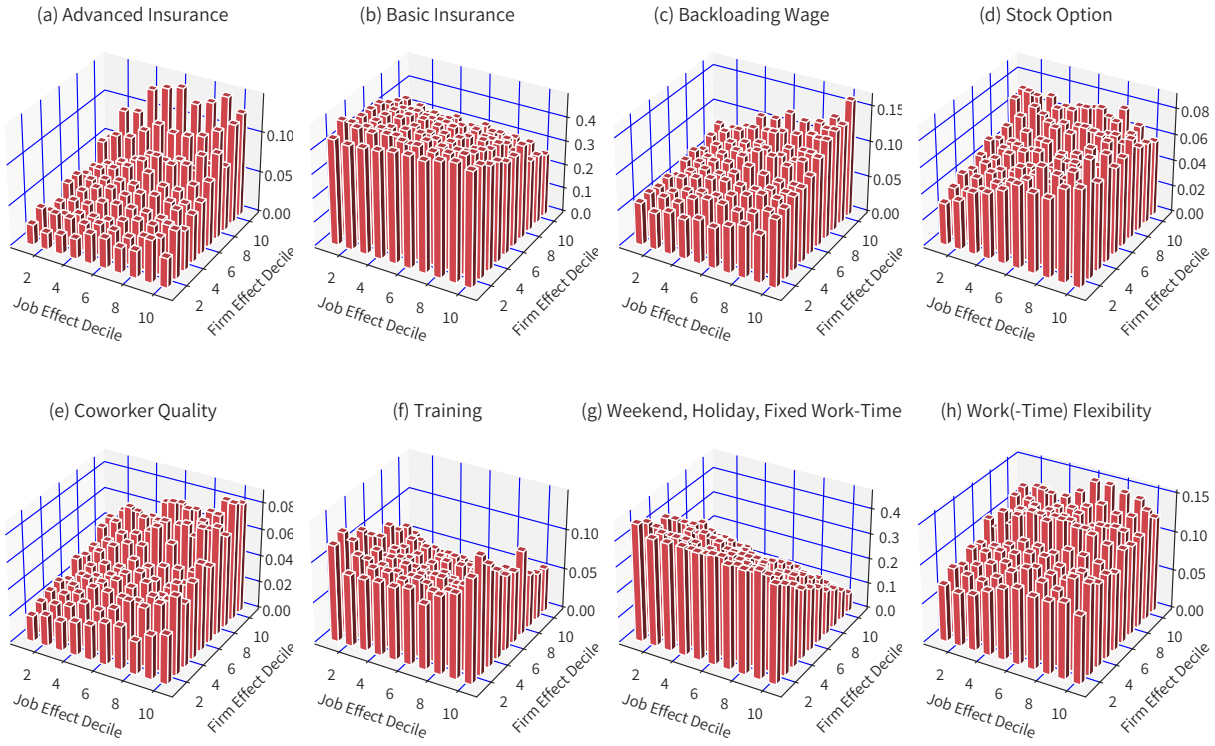
A.1 Compensation Occurrence

Figure A1: Compensation Occurrence

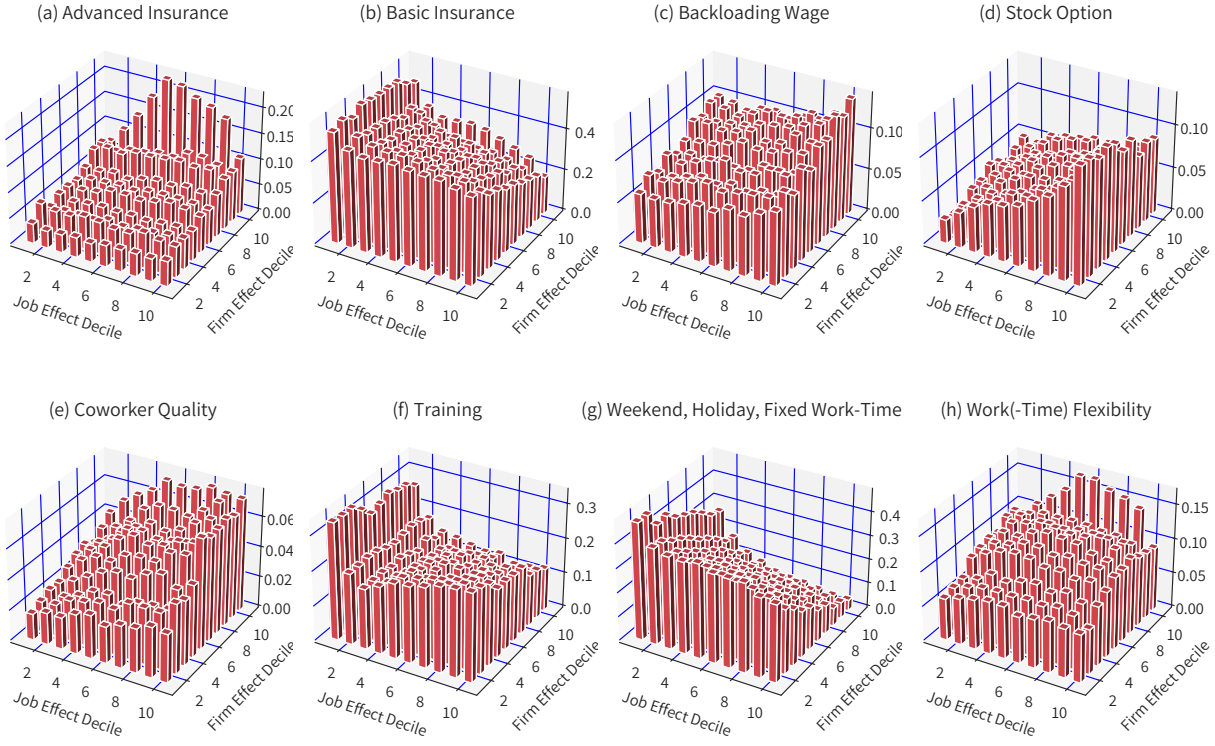
(a) Computer



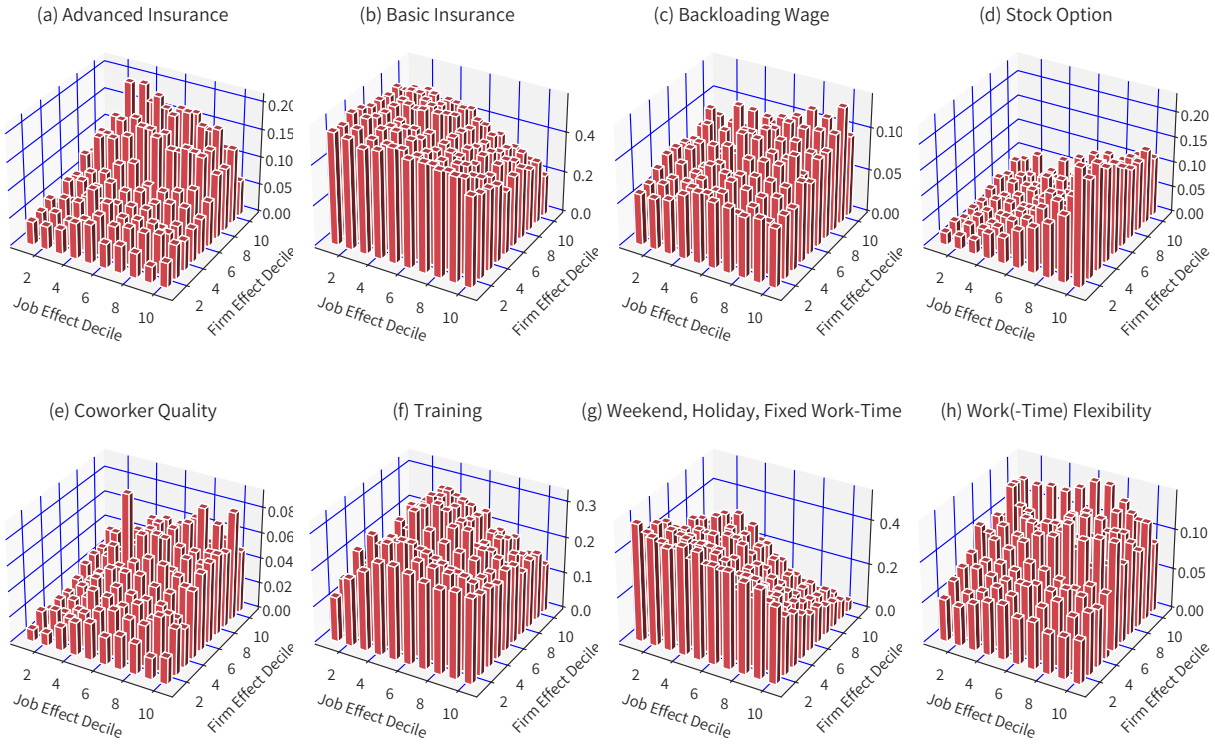
(b) Design_Media



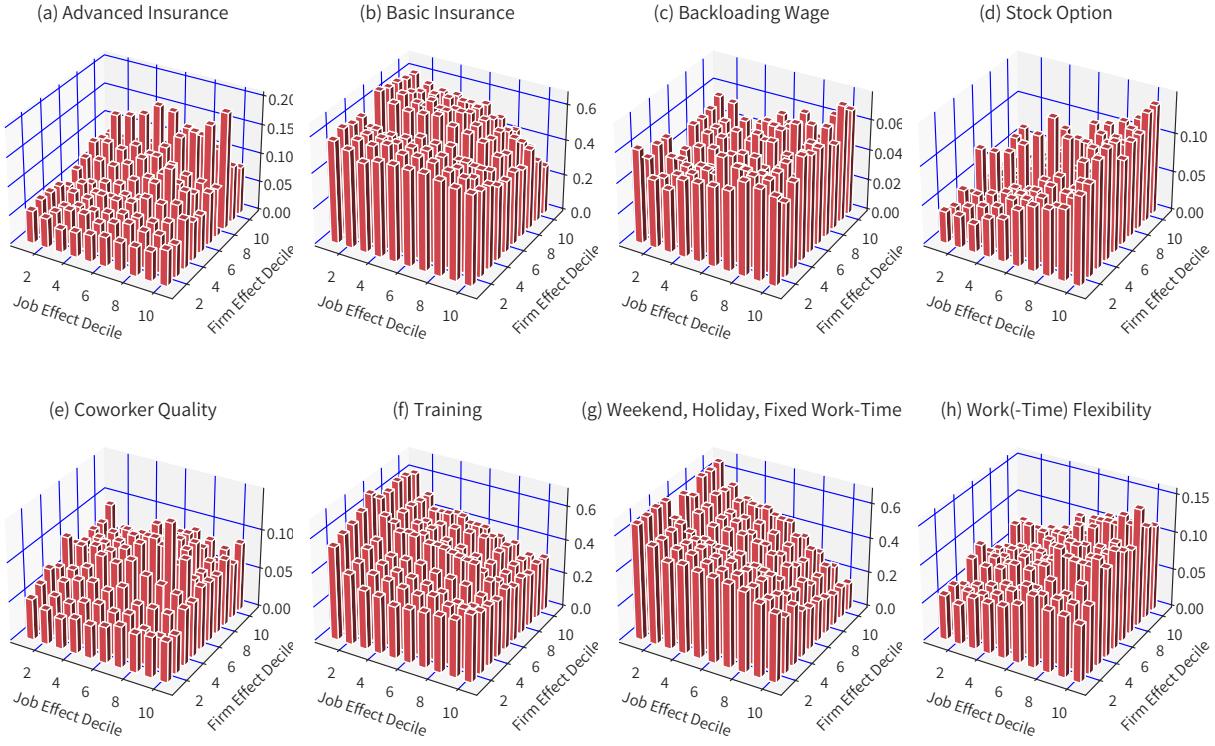
(c) Business_Operations



(d) Financial_Legal



(e) Sales



(f) Administrative