10-2019

On analysing supply and demand in labor markets: Framework, model and system

Hendrik Santoso SUGIARTO
Singapore Management University, hendriks@smu.edu.sg

Ee-peng LIM
Singapore Management University, eplim@smu.edu.sg

Ngak Leng SIM
Singapore Management University, nlsim@smu.edu.sg

Citation

SUGIARTO, Hendrik Santoso; LIM, Ee-peng; and SIM, Ngak Leng. On analysing supply and demand in labor markets: Framework, model and system. (2019). *2019 IEEE International Conference on Data Science and Advanced Analytics 6th DSAA: Washington, October 5-8: Proceedings*. 511-520.

Available at: https://ink.library.smu.edu.sg/sis_research/4531

On Analysing Supply and Demand in Labor Markets: Framework, Model and System

Hendrik Santoso Sugiarto
Living Analytics Research Centre
Singapore Management University
Email: hendriks@smu.edu.sg

Ee-Peng Lim
Living Analytics Research Centre
Singapore Management University
Email: eplim@smu.edu.sg

Ngak-Leng Sim
Living Analytics Research Centre
Singapore Management University
Email: nlsim@smu.edu.sg

***Abstract*—The labor market refers to the market between job seekers and employers. As much of job seeking and talent hiring activities are now performed online, a large amount of job posting and application data have been collected and can be re-purposed for labor market analysis. In the labor market, both supply and demand are the key factors in determining an appropriate salary for both job applicants and employers in the market. However, it is challenging to discover the supply and demand for any labor market. In this paper, we propose a novel framework to built a labor market model using a large amount of job post and applicant data. For each labor market, the supply and demand of the labor market are constructed by using offer salaries of job posts and the response of applicants. The equilibrium salary and the equilibrium job quantity are calculated by considering the supply and demand. This labor market modeling framework is then applied to a large job repository dataset containing job post and applicant data of Singapore, a developed economy in Southeast Asia. Several issues are discussed thoroughly in the paper including developing and evaluate salary prediction models to predict missing offer salaries and estimate reserved salaries. Moreover, we propose a way to empirically evaluate of equilibrium salary of the proposed model. The constructed labor market models are then used to explain the job seeker and employer specific challenges in various market segments. We also report gender and age biases that exist in labor markets. Finally, we present a wage dashboard system that yields interesting salary insights using the model.**

I. Introduction

A. Motivation

With globalization and technology advancement, labor markets in different parts of the world are now experiencing fundamental changes which directly affect labor supply and demand. Companies embracing globalization tap on cross-border supply chain network to obtain raw materials, to manufacture products, to package and move them from factories to distributors, retailers and consumers. In this huge network, jobs are distributed among different world locations. The competition among countries and cities for a pie in this network thus affects the labor markets. Technology advancements in areas such as artificial intelligence and robotization are also creating new jobs at the same time replacing others. This triggers waves of rapid changes to labor markets which may matter a lot to job seekers (who may need up-skilling), employers (who may start hiring talents with new skills) and governments (that may have to introduce new training schemes). To keep a close tap on the pulses of labor markets, we need to model and analyze these markets accurately and quickly.

Traditionally, labor markets are analyzed by surveys on employers and job seekers. Two prominent examples of these surveys are EU's Labour market and Labour Force Survey (LFS) and U.S. Bureau of Labor Statistics (BLS) Survey. These surveys are usually conducted annually as they require significant costs and human efforts. This limits their applicability to fast-changing labor economy and analyses that requires fine-grained results which can directly be used to guide individual employers and job seekers in their recruitment and job application efforts respectively.

Beyond surveys, labor economists usually model competitive labor market interaction in terms of supply of labors (workers) and demand of labors (jobs) to determine a market's equilibrium point where both applicants and employers find an ideal salary point which has the number of applicants matching the number of available jobs [1]. This salary point, also known as **equilibrium salary**, has been used to explain behaviors of employers and job seekers having to adjust their expected salaries according to the labor supply and demand respectively. Equilibrium salary and equilibrium quantity (the quantity where supply is equal with demand) serve as useful reference for employers and job applicants. So that both employers and applicants can adjust their expectation according to the market condition. The same applies to equilibrium job quantity. It gives information about how difficult to compete for jobs and applicants by considering their potential competitors and opportunities in particular labor market. In depth analysis on labor market using this approach may answer many labor market issues such as labor shortage, over-supply, and bias/preferences in job market.

Challenges. While the labor market based on supply-demand theory is well-principled, it is challenging to construct labor markets for analysis due to several reasons. Firstly, one has to derive the supply and demand equations, the two main components of the market model, based on prior knowledge of applicants and employers in the market. Our literature study has shown that there has not been any existing work on deriving these based on job post and application data. Secondly, a labor market only covers a specific group of jobs and applicants. To analyze a wide range of jobs and applications, many labor markets and their labor supplies and demands have to be created. This is undoubtedly a daunting task which requires a fresh approach.

B. Objectives

In this paper, we address the above challenges by taking a data driven approach to model different labor markets. Instead of relying on human experts, we use real world data to determine the supply and demand of the market. This solution approach significantly lowers the barrier of a principled analysis of labor market behavior. For example, comparing salaries of different jobs and applicants, as well as insights about gender and age biases in different labor markets.

Our paper focuses on three main objectives. The first objective is to model labor markets using Supply-Demand Theory. In this approach, we make use of job post and application data collected from large job portals (online platforms aggregating the job posts from employers and job applications from job seekers). We aim to derive supply and demand of labor from these data. This modeling approach is novel and is able to scale up to many labor markets when the relevant data is available.

Our second objective is to apply our labor market model to analyze and compare different labor markets. Specifically, we want to use the model to explain salary differences, as well as biases in job seeking behavior of a large population of applicants. To the best of our knowledge, this analysis of real world labor markets using data-derived market model has not been performed in the past. It will illustrate a way to analyze labor markets in a scalable, efficient, and low cost manner.

The labor market models derived in this study will benefits different market stakeholders. From the labor research standpoint, this solution approach significantly lowers the barrier of a principled analysis of labor markets and their behavior. We can analyze, compare and explain the salaries across labor markets, labor shortage/over-supply, as well as labor responses to market situations. The model can help job seekers to determine their asking salaries for different jobs. Employers can utilize the model to set appropriate salaries to attract talent. Finally, the models can also benefit policy makers creating policies that may influence employer and applicant behaviors in the labor markets to address some issues in the markets, e.g., adjusting immigration policy [2], [3], education strategy [4], [5], or organizational culture [6], [7] to counter labor shortage.

Finally, we also want to illustrate how the labor market model can be utilized by a deployed wage dashboard system[1]. As a web-based system, users (employers and applicants) can search for equilibrium salaries and job quantities of different jobs, and explore the labor market mechanism when offer or reserved salary is changed. Beyond employers and applicants, policy makers can utilize the wage dashboard system to analyze issues in labor market such as gender and age biases.

C. Outline of Paper

The rest of the paper is organized as follows. Section II summarizes the related works. Proposed framework for modeling labor markets as well as the dataset used in this work are described briefly in Section III. We address missing offer salary and unobserved reserved salary issues, we then propose the corresponding solution methods in Section IV. Section V defines our proposed labor market model based on supply-demand Theory. We evaluate constructed labor market model and conduct analysis on them in Sections VI and VII respectively. The Wage Dashboard application system is presented in Section VIII before we conclude the paper in Section IX.

II. Related Work

A. Salary Estimation Tools

An important purpose of labor market model is to help job seekers and employers determine the market salary for different types of jobs. There are several online salary estimation tools that can be found (see Wikipedia article: Salary Calculator[2]). For example, Hudson[3] introduces an Online Salary Calculator to return different salaries of a job title based on the industry, working experience of candidate, and work discipline. To implement this calculator, salary surveys are conducted with a group of human resource experts, employers and consultants to extract their salary insights. Glassdoor[4], PayScale[5] and Emolument[6] crowdsource user contributed salaries so as to generate estimated salaries for user queried jobs. All these online tools however are based on unpublished methods. The quality of salaries provided would hinge on the quality of crowdsourced salaries and users may not always be willing to disclose their true salaries [8]. Kenthapadi et. al [9] therefore proposes Bayesian hierarchical smoothing to remove salary outliers and to derive robust insights from such user-reported data. This method is used in LinkedIn Salary to facilitate salary transparency without sacrificing security and privacy [10]. LinkedIn Salary also introduces a two-step process (company embeddings and Bayesian statistical model) to extract salary insights at the company level [11]. Finally, crowdsourced salaries do not necessarily reflect the prevailing market situation. Hence, their use is more appropriate for single job posts, instead of a labor market.

Unlike current salary estimation tools which are limited to salary estimation for a particular job without providing adequate information related to their competitors and market conditions, we propose a new framework by incorporating supply-demand theory from economics to estimate the current equilibrium salary and market condition (including potential competitors). Moreover, although our research is incorporating salary prediction models, our research is different with other salary prediction researches [12]. Because, our research focuses more on deriving labor supplies and demands (including equilibrium point), rather than predicting job post salaries only.

[1]https://research.larc.smu.edu.sg/wagedb/

[2]en.m.wikipedia.org/wiki/Salary_calculator

[3]hudson.sg/salary-hub/salary-calculator

[4]www.glassdoor.com.

[5]www.payscale.com/

[6]www.emolument.com

B. Labor Theory Research

Labor market model is largely about modeling labor supply and demand. With the two, one can determine market equilibrium by the intersection between supply and demand curves [13], [14]. The supply-demand theory and the concept of equilibrium are derived from Adam smith's famous invisible hand theory [15]. As the economic become more complex, the classical economic theory is further progressed into neoclassical economic theory [16]. According to neoclassical theory, most markets can quickly reach the equilibrium situation without excess of supply or demand. In labor market, however, there can be a long term unemployment [17], [18]. Labor market also contains wage differential even among similar workers [19]. In recent years, economists also propose a search theory to analyze frictional unemployment data [20], to explain jobs creation and jobs destruction [21].

Furthermore, neoclassical economics in general assumes a unified market for labor with open competition [1], [22], [23]. In contrast to neoclassical theory, the theory of labor market segmentation assumes the separation of the labor market into several sub-markets according to specific criteria such as occupation and location in which participants of one market group cannot easily join other market groups [24], [25], [26]. In this paper, our labor market model can be modified for specific sub market groups according to different criteria such as category and occupation types.

Economic theories may provide an explanatory analysis for specific market behavior based on several rational assumptions [27], [1], [28], [16]. However, this approach is generally conceptual and not data driven. For instance, recently people also study the influence of network structure on economic behavior [29], [30], [31] and agent based interaction [32], [33]. On the other hand, statistical/econometric approach may provide empirical estimation about current market situation, but the result is limited to purely descriptive analysis for specific market situation [34], [35], [36], [37]. Moreover, meaningful result in econometric model requires carefully controlled experiment with specific instrumental variable which may not always available in real market situation [38], [39], [40], [41]. In many cases, these studies require expensive economic experiment or a lot of effort to conduct many surveys or census on employers and employees to collect sufficient data [34], [35], [42], [43], [44]. Our attempt to construct labor market model in a data driven manner is unique, in which the supply and demand curves in our proposed labor market model are constructed based directly on data and not only conceptual. Moreover the proposed labor market model can be built very quickly and does not require any survey as traditional econometric models.

III. Datasets and Proposed Framework

A. Datasets

Job Post Dataset. The job post dataset covers all job posts advertised by employers and their attributes of Singapore's registered companies in 2015 and 2016. All these job posts together represent the labor demand. Other than (i) job title, job post j_k also mention about (ii) occupation o_{j_k} indicating the generic job title assign to the job. In total, there are 925 unique occupation labels; (iii) job category c_{j_k} indicating the professional field the job belongs to (out of a dictionary of all job categories denoted by C), there are 41 job categories, i.e., $|C| = 41$; and (iv) offer salary d_{j_k}. Every job post is assigned a salary range $[l_{j_k}, h_{j_k}]$ instead. The salary range is in the unit of Singapore Dollar. To generate the monthly offer salary, we take the average between the the minimum salary l_{j_k} and maximum salary h_{j_k}. The job post dataset also provides the following additional attributes, namely: (v) employment type, with 8 unique labels; (vi) job level, with 7 unique labels; (vii) Experience recording the years of experience suitable for the job; (viii) Industry referring to the industry the job belongs to (there are 860 unique industry labels); (ix) Education referring to the suitable education level for the job (there are 12 unique education labels in the data); (x) Company referring to the name of the company; and (xi) Number of vacancies which will be used later to determine the number of demand.

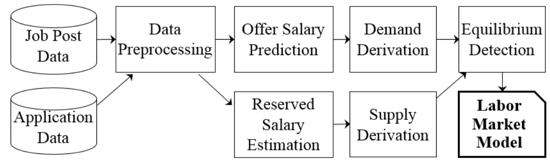

Fig. 1: Data-Driven Labor Market Modeling Framework.

Application Dataset. This dataset contains job applications for the above-mentioned job posts. We use U and A to denote the set of all observed applicants and applications respectively. An application $a_{i,k} \in A$ is an *observed application* of applicant u_i ($\in U$) to the job j_k ($\in J$). Note that observed applications are affected by market friction and do not completely represent the true set of jobs that the applicant wants to apply. In our proposed model, we would like to derive the *true* supply using some rules to match users with specific sub-market. Moreover, we also have some information about the applicants such as gender and age. The application dataset covering all applications of jobs in Singapore during 2015 only and their applicants. We unfortunately do not have applications dataset during 2016 (which limit our labor market analysis on 2015 only).

B. Framework

Unlike past works which assume analytical demand and supply curves exist beforehand, we propose to derive the labor-demand and labor-supply from the observed real world data collected from the real world labor market and construct the labor market models for market analysis. The framework takes both job post and application dataset as input to derive the demand and supply curves respectively. Figure 1 depicts our proposed framework.

1) Data Pre-processing: Our framework first performs *data pre-processing* on both the job post and application data to remove noises as well as to select the subset of job and application data for building the target labor market model. In this paper, we focus on class of labor market model, one based on job category. Another type of labor market such as occupations will be discuss briefly in Wage Dashboard in VIII

We only focus on job posts covering full-time, permanent, or contract jobs. Therefore, we remove part-time, internship, and other ad-hoc jobs. Furthermore, the same job post may be posted multiple times in the repository, hence creating duplicates. In order to distinguish whether the identical job is an extension of previous job or a new opening, we first detect job posts with identical content and were posted within a month ($\leq$ 31 days) after expiry date. We subsequently removed these posts. We assume the identical job that posted after 1 month to be a new opening of the same job.

Moreover, we also remove possible erroneous contents for cleaner job features. Specifically, we remove any jobs with unrealistically high experience (higher than 50 years) and any jobs with more than 5 categories. We find erroneous observed offer salaries by detecting job posts with invalid monthly salary ranges. These involve max salary $h_{j_k} <$ min salary l_{j_k}, and unrealistically high range between max salary and min salary, i.e., $\frac{h_{j_k}}{l_{j_k}} \geq 3$. We also remove any job posts with offer salary lower than \$500 to avoid any jobs with hourly salary. These cleaning processes are only remove about 1% jobposts

Next we want to remove salary outliers. We first remove job posts with offer salary higher than 2 standard deviation from the mean. This step will help us to remove unrealistically high salary. Next, we remove additional salary outliers for specific occupation. Outlier detection is performed by a box-and-whisker method using Tukey's fences. Specifically, for each occupation o, we compute the monthly offer salaries at the first and third quartiles of its monthly offer salary distribution as denote them as $d_{Q1}(o)$ and $d_{Q3}(o)$ respectively. We define the interquartile range, $IQR(o)$, to be $d_{Q3}(o) - d_{Q1}(o)$. The outlier job posts j_k's of occupation o are those with monthly salaries $d_{j_k} < d_{Q1}(o) - 1.5 \times IQR(o)$ and $d_{j_k}(o) > d_{Q3}(o) + 1.5 \times IQR(o)$.

Data Statistics. Table I shows the statistics about job post data. We have more than 265K job posts at the beginning. After job pre-processing, we are left about 203K job posts, of which about 20% of them do not have monthly salary range information. As we remove some job posts, we removed the associated applications. We further removed the applicants who no longer have applications. As shown in Table I, we finally have 72K applicants and their 2M applications to job posts with observed salaries.

2) Offer Salary Prediction and Demand Derivation: Next, we conduct *offer salary prediction* on job posts with missing offer salaries which often represent a significant proportion of all job posts (will be elaborated in Section IV-A). With the observed and predicted offer salaries, the *demand derivation* step determines number of job posts with offer salaries meeting at every salary point d (will be elaborated in Section V).

3) Reserved Salary Estimation and Supply Derivation: From the application dataset, we need to determine a reserved salary for every job applicant as part of the *reserved salary estimation* step (will be elaborated in Section IV-C). Once we have the reserved salary of every applicant, we carry out the *supply derivation* step which essentially constructs the supply curve by determining the number of applicants interested in the labor market at every salary point d, very much similar to the construction of demand curve (will be elaborated in Section V).

TABLE I: Dataset statistics

Job Posts (J)	
Before Preprocessing	265,553
After preprocess	203,353
Jobs with observed salaries	160,844 (79%)
Applicants (U)	
Before Preprocessing	81,656
After Preprocessing	81,656
Applicants to jobs with observed salaries	72,395 (89%)
Applications ($A = \{a_{j,k}\}$)	
Before Preprocessing	2,884,758
After Preprocessing	2,878,036
Applications to jobs with observed salaries	2,019,987 (70%)

4) Equilibrium Point Estimation: The *equilibrium point estimation* step determines the *equilibrium salary* and *equilibrium quantity* corresponding to the intersection point between demand and supply curves of a labor market model (will be elaborated in Section V). At this salary point (i.e., equilibrium salary), the number of job applicants equals that of job posts (i.e., equilibrium quantity).

IV. Salary Estimation

A. Offer Salary Prediction Models

We formulate offer salary prediction task as a regression problem using some job posts with observed offer salaries as training data. We use all available job post attribute features: *job title*, *job category*, *job level*, *employment type*, *experience*, *industry*, *education*, *occupation label*, *company*, and *skills*. Given a job title, we parse it into job function and domain words. E.g., in the job title "research manager", "research" is a domain word, and "manager" is a function word. Hence, a job title is represented by two sets of binary word features, corresponding to domain and function words respectively. Categorical attributes *job category*, *job level*, *employment type*, *industry*, *education*, *occupation label*, and *company* are represented as binary features, one for each category label. Bag attribute, *skills*, is represented as a vector of binary skill features (with 1 and 0 indicating the presence and absence of specific skill in the bag respectively). We maintain a dictionary of skills and match the skills mentioned in job post. *Experience* are represented as numerical features. Among above features, we only retain those that exist in at least 10 job posts. Finally, we obtain 18,958 features for a given job post.

To avoid the learned models predicting unrealistically high salaries and negative salaries, we implement bounded mechanism to all models, which force predicted salaries to stay within expected salary range for training job posts. Whenever a regression model predicts offer salary of a job j_k higher than expected range, we cap the prediction to be maximum observed offer salary. Similarly, we force the predicted offer salary to be minimum observed reserved salary whenever regression model predicts salary below expected range.

B. Prediction Accuracy Evaluation

Experiment Setup. To determine the accuracy of the various offer salary prediction models, we conduct an experiment to evaluate the models using 5-fold cross validation. We divide the job posts with observed offer salaries into 5 folds. When one of the folds is used as the test data, the remaining folds are used for training the regression models. *Mean absolute error* (MAE) between the predicted offer salary and ground truth of all job posts in the test fold is used as our accuracy measure. We also include *mean relative error* (MRE) defined by the mean of absolute errors between predicted and ground truth offer salaries divided by the ground truth salaries.

For comparison, we introduce prediction models: (i) **Average (Avg)** model simply estimates the offer salary of a job j_k by the average of offer salaries of all jobs in J. (ii) **Group-Average (GAvg)** model estimates the offer salary of j_k by the average of offer salaries of all jobs of the same category label (**GAvg**(C)) or occupation label (**GAvg**(O)). (iii) **Linear Regression** models, including its main variations: **Linear Regression (L-Reg), Ridge regression (R-Reg), Lasso regression (Lasso)**. These models are implemented using sklearn[45]. (iv) **Neural Network (NN)** model with several possible combinations of number of layers and nodes. For instance, **NN 2lx100** indicates fully connected network with 2 layers and 100 nodes for each layer. This model is implemented using keras[46]. (v) **Multitask Neural Network (MTNN)** model. This model is similar with **NN**, but now predicting both minimum and maximum observed offer salary, then return the midpoint of those predictions. This model is implemented using keras[46].

As shown in Table II, linear regression, ridge regression and Lasso regression outperform the average and group average models significantly. Among the regression models, Lasso performed slightly better than others. Neural network does not perform better than ridge or lasso when the number of layers and nodes are few. However, neural network outperform the regression model by adding more nodes and layers. Finally, multitask neural network that predicting both min and max salary achieved the best accuracy 13.6% mean relative error. Rigorous investigation behind better accuracy of multitask model with deeper network can be considered as future studies.

C. Reserved Salary Estimation

To estimate the supply in a labor market, we require the reserved salary of every applicant. As applicants do not explicitly reveal their reserved salaries, we propose a behavioral approach to estimate them. The user may have multiple reserved salaries depending on the market they are interested in. We therefore propose the **market specific average method (MAvg)** of estimating reserved salary of a user u_i in the labor market m with a subset of job posts denoted by $J(m)$, i.e., $\hat{d}_{u_i}^{MAvg}(m) = Avg_{a_{i,k}}\{d_{j_k \in J(m)}\}$. There are several variations of this method for instance **market specific minimum method (MMin)** or **market specific first quartile method (M1qr)**.

TABLE II: Results of Offer Salary Prediction Model

Baseline	**Avg**	**GAvg**(c)	**GAvg**(o)	**L-Reg**	**R-Reg**	**Lasso**
MAE	\$1940	\$1635	\$1255	\$ 902	\$864	\$845
MRE	0.528	0.422	0.293	0.207	0.198	0.189
NN	**2lx100**	**2lx300**	**2lx1000**	**5lx100**	**5lx300**	**5lx1000**
MAE	\$869	\$717	\$683	\$718	\$675	\$655
MRE	0.187	0.152	0.144	0.148	0.143	0.139
MTNN	**2lx100**	**2lx300**	**2lx1000**	**5lx100**	**5lx300**	**5lx1000**
MAE	\$891	\$741	\$674	\$738	\$671	**\$647**
MRE	0.189	0.157	0.143	0.157	0.142	**0.136**

V. Labor Market Modelling

Each labor market defines a scope of market covering jobs, applicants and their job applications. To derive the labor-demand curve, we predict offer salaries of jobs by using salary prediction model to learn jobpost dataset in Section IV-A. To derive labor-supply curve, we infer the reserved salaries of applicants based on their job application behavior as outlined in Section IV-C. Finally, we describe how equilibrium salary and quantity are obtained for a labor market. To apply supply-demand theory to a labor market, we define job posts from employers as the demand of labor, and job applicants as supply of labor. The labor-supply thus models the increase (or reduction) of labor supply as salary increases (or decreases). The labor-demand on the other hand models the decrease (or increase) of required labor as salary increases (or decreases).

We now define our applicant-supply and job-demand models for a given labor market $M(U, J)$ (or a sub-market $M_m(U(m), J(m))$). In this market, U and J represent the supply and demand respectively. The supply and demand vary as the salary varies. Assuming both applicants and employers are rational agents, the supply (applicants) will increase as the salary increase because the applicant want to maximize their potential salary. On the other hand, the demand (job offer) will decrease as the salary increase because company want to minimize the potential cost of production. Some company may offer higher salary to attract more applicants, however realistically they don't mind to pay lower salary if the applicants are willing to take the job with such salary.

Minimum-Market Friction (MMF) assumption. In practice, it is impossible for all applicants to read and apply all jobs that they are interested due to many reasons including confident level, work load, location, etc.. This mismatching issue is known as *market friction*. Similar to the basic supply-demand theory, we also assume that there is no market friction between applicants and jobs. Under the MMF assumption, we assume that an applicant u_i is interested in jobs from $J(m)$ as long as u_i was observed to apply for any jobs from $J(m)$.

Labor-supply and labor-demand curves. To construct the supply and demand curves of our proposed labor market

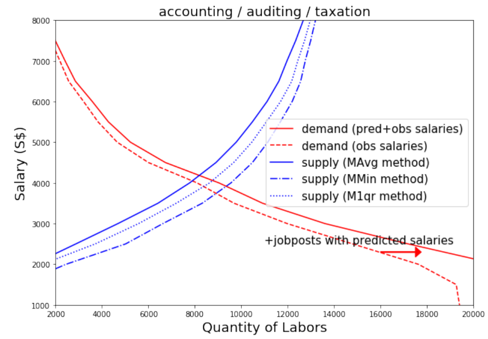

Fig. 2: Supply and Demand Curves with Different Settings.

model, we vary salary from the lowest salary in the market to the highest salary. At each salary point d, the set of jobs j_k with offer salary $d_{j_k} \geq d$, and the set of applicants u_i with reserved salary $d_{u_i} \leq d$ represent the demand and supply quantities respectively. As required by the above assumptions, we need to know the reserved salaries and offer salaries of applicants and jobs respectively (estimated in Section IV).

In Figure 2, we show typical supply and demand curve for "accounting/auditing/taxation" job category. By using all job posts with observed and predicted salaries (red line), the demand is increased slightly causing the equilibrium salary and quantity to increase compared with using only job posts with observed salaries (red dashed line). Three different supply curves represents three different reserved salary estimations in section IV-C: MAvg method (blue solid line), MMin method (blue dashed blue line), M1qr method (blue dotted line). We can also observe similar effect for other job category specific labor markets. The **equilibrium salary** d_{eq} is reached when supply and demand curves meet each other. At this d_{eq} point, the number of jobs matches that of applicants. This number is also known as the **equilibrium quantity** q_{eq}.

VI. Empirical Evaluation of Proposed Model

It is extremely difficult to validate any labor market model as the model principles may not work perfectly in the real world settings which usually involve many constraints, behaviors and external factors are not considered by the model. In this section, we propose an empirical approach to evaluate our proposed labor market model by correlating the expected salary adjustment behavior of employers with the observed salary adjustment. We assume that employers have more knowledge about the labor market's supply and demand, and hence they would adjust the offer salary upward or downward to be closer to the equilibrium salary [47], [28], [48]. Several empirical evidences suggest that labor market will experience salary adjustment and convergence across the year [49], [50].

We believe that one year response time is reasonable as it still takes quite some time for employers to learn the previous year's market, and to work out offer salaries of the current year. Moreover, the magnitude of adjustment can be different across companies, and hence we will focus on the direction only. In this evaluation, we specifically compare the gaps between offer and equilibrium salaries in 2015 (the year of which we have application data), with offer salary changes between 2015 and 2016. During those years, there were no major government policy changes that affected salaries.

We focus on gaps between offer and equilibrium salaries that are significant. We also want to consider only offer salary changes that are significant. We thus introduce z, a threshold for a significant relative salary difference. Hence, the gap between offer of a company f and equilibrium salaries of a sub-market m in year y is significant if $\frac{|d_{eq}(m,y)-d_o(m,f,y)|}{d_{eq}(m,y)} \geq z$. Similarly, $\frac{|d_o(m,f,y+1)-d_o(m,f,y)|}{d_o(m,f,y+1)} \geq z$, indicates the offer salary change of a company f for a sub-market m between years y and $y+1$ is significant[7]. We empirically are interested in at least 10% salary difference (i.e., $z = 0.1$). We then select companies which have been observed to have significant gaps between their offer salaries and equilibrium salaries as well as significant change to offer salary between 2015 and 2016.

We next examine the direction of offer salary change consistency with respect to salary gap between offer salary and equilibrium salary, by defining the **Salary Adjustment Consistency (SAC)** score for a company f in sub-market group m as follows: $SAC(m, f) = H((d_{eq}(m, y) - d_o(m, f, y)) \cdot (d_o(m, f, y + 1) - d_o(m, f, y)))$, where $H()$ is the Heavy-step function: $H(x) = 0$ if $x \leq 0$, and 1 otherwise. $SAC(m, f) = 1$ if the offer salary adjustment from year y to $y + 1$ is consistent with salary gap between offer salary and equilibrium salary in year y; and 0 otherwise. The average SAC for sub-market group m is the defined as $SAC(m) = \frac{1}{|F|}\sum_{f\in F} SAC(m, f)$. In this paper, the sub-market is limited to job category only.

In Figure 3, we list the SAC scores (for equilibrium salary using MAvg reserved salary method) of all job categories in decreasing order. The average SAC score over all job categories (shown as the green line) is about 0.67, indicating that employers in most job category-specific labor sub-markets are consistent with equilibrium salaries in their offer salary adjustments between 2015 and 2016, except a few such as "legal" and "social services". Some inconsistent salary adjustments may due to market specific context (for instance "social services" jobs are often associated with volunteering[8]). Furthermore, the average SAC scores for labor market model that assume MMin and M1qr reserved salary methods are 0.66 and 0.67, which indicates that different reserved salary method will not affect the result significantly.

VII. Analysis Using Labor Market Models

We now utilize our labor market models to analyze labor markets of different job categories. In particular, we focus on analyzing the equilibrium salaries and quantities of these labor markets, demonstrating the effectiveness of the data-driven approach. In the following analysis, we use both observed offer

[7] $d_o(m, f, y+1)$ of year $y+1$ is used as the denominator to be consistent with the use of $d_{eq}(m, y)$ as the denominator of the former which serves as the new salary benchmark.

[8] www.straitstimes.com/singapore/number-of-social-workers-doubles-in-last-4-years

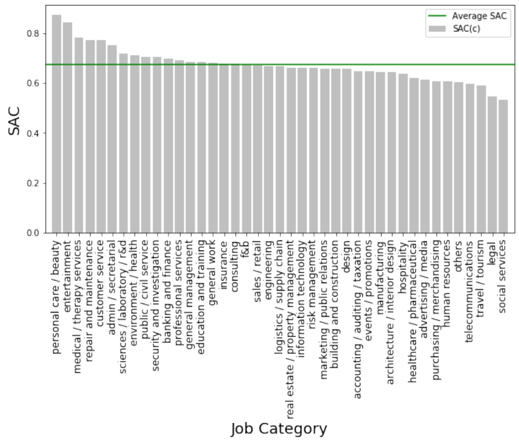

Fig. 3: SAC distribution of category-specific labor sub-markets.

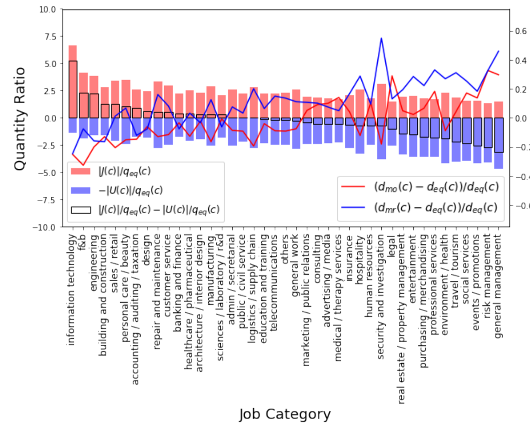

Fig. 4: Quantity Ratio and Relative Salary Ratio.

salaries and predicted offer salaries using **MTNN(5lx1000)**. Furthermore, the applicants are assumed to have reserved salaries of market specific average **MAvg**.

A **job category specific labor sub-market** $M_c(U(c), J(c))$ can be defined for each job category c. The equilibrium point indicates the intersection between supply and demand curves. The equilibrium salary and quantity of the labor market of job category c are denoted by $d_{eq}(c)$ and $q_{eq}(c)$ respectively.

A. *Equilibrium quantity analysis*

The total demand and supply of the labor market of category c refer to all job posts of job category c and all applicants interested in job category c. We denote them as $|J(c)|$ and $|U(c)|$ respectively. Over-supply occurs in the market when job posts out-number applicants, i.e., $|J(c)| < |U(c)|$.

We define the **demand quantity ratio** and **supply quantity ratio** as $\frac{|J(c)|}{q_{eq}(c)}$ and $\frac{|U(c)|}{q_{eq}(c)}$ respectively. When the quantity ratios are close to 1, the job category c sees the number of jobs and applicants close to the ideal market size. We also compute $|J(c)|/q_{eq}(c) - |U(c)|/q_{eq}(c)$ to measure the **relative difference between demand and supply**.

Figure 4 shows the above measures across job categories. Each blue or red bar represents the demand quantity ratio or supply quantity ratio respectively. The empty bar represents the relative difference between demand and supply. The job categories are ordered from the most positive relative difference demand and supply to the most negative. We observe that "information technology (IT)" has the highest demand quantity ratio around 6, implying that the number of job posts is about 6 times that of the equilibrium quantity $q_{eq}(c)$. On the other hand, its supply quantity ratio is about 1 or the total supply is about the same as $q_{eq}(c)$. It means the employers in the "IT" labor market have difficulty filling the available job positions as it is the applicant's market. This findings reflects that the "IT" labor market in Singapore experiences shortage of skill talent[9]. The figure also shows that "f&b" and "sales/retail" markets experience similar shortage of manpower. This tallies well with a business report during the same time period[10]. On the other hand, the "general management" and "risk management" markets are the opposite, as their applicants have difficulty securing job positions because of the large negative relative difference between demand and supply.

B. *Equilibrium salary analysis*

Next, we analyze the equilibrium salaries of category specific labor markets. When employers offer salaries lower than $d_{eq}(c)$, our model would suggest that they will not be able to attract enough applicants. Similarly, if applicants require salaries higher than $d_{eq}(c)$, they will not likely get a job offer. Conversely, if applicants ask for salaries lower than $d_{eq}(c)$, they are likely to secure a job.

We compare for each job category c, $d_{eq}(c)$ with *median offer salary* $d_{mo}(c)$ and *median reserved salary* $d_{mr}(c)$ derived from observed data. We also compute **relative offer salary ratio** as $(d_{mo}(c) - d_{eq}(c))/d_{eq}(c)$. When $(d_{mo}(c) - d_{eq}(c))/d_{eq}(c) > 0$, most employers offer higher than equilibrium salary. Conversely, most employers offer lower than equilibrium salaries. We call these two cases **over-offer** and **under-offer** respectively. By comparing $d_{mr}(c)$ with $d_{eq}(c)$, we derive over-reserved $(d_{mr}(c) - d_{eq}(c)/d_{eq}(c) > 0)$ and under-reserved $(d_{mr}(c) - d_{eq}(c))/d_{eq}(c) < 0)$. We call $(d_{mo}(c) - d_{eq}(c))/d_{eq}$ and $(d_{mr}(c) - d_{eq}(c))/d_{eq}$ **relative offer salary ratio** and **relative reserved salary ratio** respectively.

Figure 4 shows the relative offer salary ratio (shown as red line) and relative reserved salary ratio (shown as blue line) of each job category. Generally, the two ratios are positive correlated. Interestingly, "IT", "F&B", "engineering", and "sales" are some job categories belonging to the under-offer/under-reserved combination. This means that employers

[9]www.straitstimes.com/singapore/manpower/it-talent-in-short-supply-amid-smart-nation-push

[10]sbr.com.sg/hr-education/news/labour-intensive-industries-marred-manpower-shortage-epidemic

are not paying enough salaries and applicants are also going after jobs with less than equilibrium salaries. On the other hand, the "legal" job category enjoys both higher than expected offer salary and high than expected reserved salary (i.e., over-offer/over-reserved). Both the employers and applicants appear to offer/secure salaries higher than the equilibrium salary.

Moreover, we can also analyze the gap between relative offer salary ratio and relative reserved salary ratio. Relative reserved salary ratio is higher than relative offer salary ratio in "manufacturing", "security", "logistics", and "hospitality". In these sectors, job applicants expect higher salaries than equilibrium salary while employers offer salaries lower than equilibrium salary. On the other hand, relative offer salary ratio is higher than relative reserved salary ratio in "legal" category.

C. Gender Preference

Many studies have shown the gender has been an important factor affecting salaries. This gender-based salary gap has reduced since then but remains to be significant enough for researchers to focus on [51]. A study also found that the female-male salary difference is larger at the top wage distribution [52]. They further attempted to establish possible explanations of the difference.

In this section, we want to analyse how the applicant preferences may be determined by gender. The result here is not an indication of gender discrimination as we do not make use of hiring data from companies. Our labor market model can however show that male and female applicants have different preferences in applying jobs. In this analysis, we calculate the **median reserved relative salary ratio** for male and female as defined by $\frac{(d_{mr}^{gender}(c) - d_{eq}(c))}{d_{eq}(c)}$. Furthermore, we also calculate the proportion of male and female who applied for specific job category, $\frac{|U^{gender}(c)|}{|U(c)|}$.

The result is shown in Figure 5, where the x-axis represents job category ordered by decreasing proportion of male applicants. Job categories in the extreme left, e.g., 'repairs and maintenance', 'engineering', and 'security and investigation' dominated by male applicants. On the other hand, job categories on the extreme right, e.g., 'admin/ secretarial', 'human resources', and 'accounting/ auditing/ taxation' are mostly applied by female applicants.

Figure 5 also shows that reserved salary ratios of male (the blue line) are higher than that of female (the red line) across job categories. Female applicants appear to be more willing to apply for jobs below equilibrium salaries. Males applying for 'professional services', 'general management', and 'sales/ retail' have higher salary expectation than their female competitors. While this finding is based on personal salary expectation and does not explicitly show any pay gap, it is possible that people apply to jobs that have a comparable salary to their current job (Singapore is known for gender pay gap[11]). Moreover, the equilibrium salary can be used as indicator about how far their reserved salary disparity from

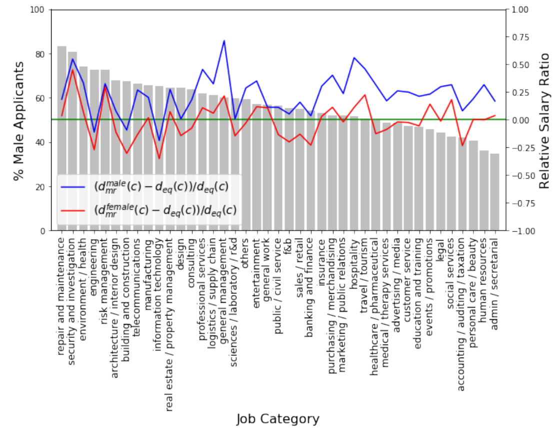

Fig. 5: Median Reserved Salary and Applications Preferences of Specific Gender.

the overall market situation. Although female median reserved salary is consistently below their male competitors, it is closer to equilibrium salary in some specific job categories.

D. Age Preference

In this section, we analyse the age-based preference of applicants. Again, this result is not related to age discrimination, but only shows applicant preferences due to age. We show the age groups below and above 30. As shown in Figure 6, the job categories are ordered based on proportion of applicants younger than 30 years old. The job categories on the left (e.g., 'events and promotions') are mostly applied by the young applicants while job categories on the right (e.g., 'information technology',general management, telecommunications) are mostly applied by older applicants.

We observe that young applicants are more willing to apply for lower salary than their older counterpart. Moreover, in many categories, applicants below 30 are willing to apply for jobs below equilibrium salary (represented by the horizontal line), while applicants above 30 generally apply for jobs above equilibrium salary.

VIII. Wage Dashboard

In this section, we introduce Wage Dashboard, a running application system that is developed based on our proposed labor market model. This system is intended to serve both employers and job applicants simplifying the salary insights using a search and visualization interface. The current version of the system covers 935 occupations from 33 industries in Singapore. For each job in an industry, the corresponding labor market model is pre-computed and stored in the system for supporting efficient interaction with end users.

Wage Dashboard has a landing page which allows user to search for a job using occupation in some industry. Note that jobs from different industries may share the same occupation, we need both occupation and industry to select a job as well as its labor market model. For the purpose of this explanation, we

[11]www.straitstimes.com/business/economy/no-improvement-in-singapores-gender-pay-gap-since-2006-report

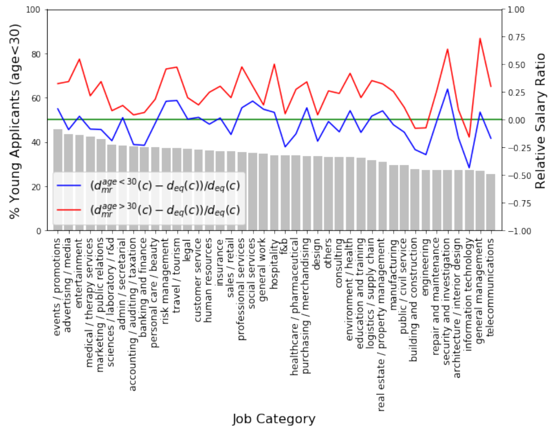

Fig. 6: Median Reserved Salary and Applications Preferences of Specific Age range.

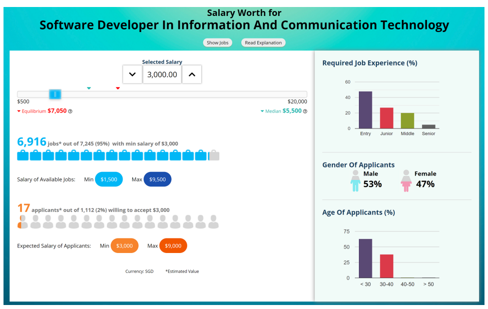

Fig. 7: Low Salary Manipulation.

select software developer in information and communication technology.

Insights visualization. Once a job is identified and its labor market model is loaded by Wage Dashboard, a insights visualization interface is presented as shown in Figure 7 and Figure 8. The insights include:

- *Salary insights*: These are the minimum offer salary, maximum offer salary, median offer salary and equilibrium salary. As equilibrium salary is higher than median offer salary, this suggests that most of the offered jobs fail to competitive market for software developer applicants.
- *Job quantity insights*: The visualization also includes total number of jobs and the equilibrium job quantity. The required job experience distribution (under entry, junior, middle, and senior) of the offered jobs is also given.
- *Applicant insights*: The dashboard displays the total number of applicants, their gender and age distribution.

The insights are useful for employers and applicants to determine the market salary, competitiveness of labor market, and general profiles of jobs and applicants.

Salary manipulation. The user can slide the salary slider and examine the non-equilibrium state of the labor market at different selected salary points. This salary point can be selected as the hypothetical offer salary for employer or reserved salary for applicant. At any selected salary point, the dashboard displays the number of offered jobs and the number

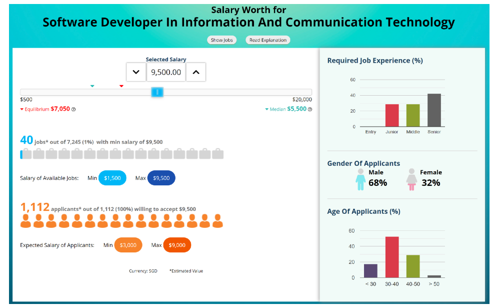

Fig. 8: High Salary Manipulation.

of applicants at the selected salary point, as well as their profile distributions. For example, Figure 7 shows the salary point of \$3,000 has been selected. There are 6,916 available offer jobs and potential 17 applicants. At this salary point, most jobs require entry level experience. Applicant's age distributions is dominated by younger applicants and applicant's gender distribution is balance.

Moreover, by sliding the slider to the right, Figure 8 shows the salary point of \$9,500 is selected. Consequently, number of available offer jobs reduces to 40, while number of potential applicants increases to 1,112. At this salary point, most jobs require senior level experience. Applicant's age distribution is now dominated by older applicants and applicant's gender distribution is now shifted towards male majority.

IX. Conclusion

We have introduced a novel data driven framework to develop principled labor market model from real world dataset. This framework helps us estimate the market equilibrium in each labor market. We also study various job category markets. With the labor market model, we can perform more comprehensive analysis and study different biases in different market segments. The framework can be extended to study other types of labor markets. Since the proposed model involves both supply and demand, the analysis can also be extended to provide other useful economic market information.

Several limitations nevertheless exist in our study. For example, reserved salaries are not observed in the data. We assume that the desired reserved salary can be estimated by some function on his/her applied jobs' offer salaries. It will be interesting to explore ways to model applicants' reserved salaries. Our proposed model also assumes ideal markets without any friction in which all related parties have complete knowledge about all available options. The assumption may be breached when there are also additional supply and demand beyond our observed data. In other words, the market is not necessarily a close system. How to account for these gaps is certainly an area for future research.

Acknowledgment

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative.

References

[1] G. Borjas, *Labor Economics*. McGraw-Hill/Irwin, 2005.

[2] P. E. Fong and L. Y. C. Lim, "Labor, Productivity and Singapore's Development Model," in *Singapore's Economic Development*, ser. World Scientific Series on Singapore's 50 Years of Nation-Building. WORLD SCIENTIFIC, Aug. 2015, pp. 135–168.

[3] M. Guzi, M. Kahanec, and L. M. Kurekova, "How Immigration Grease Is Affected by Economic, Institutional, and Policy Contexts: Evidence from EU Labor Markets," *Kyklos*, vol. 71, no. 2, pp. 213–243, 2018.

[4] P. Waring, C. Vas, and A. S. Bali, "The challenges of state intervention in singapores youth labour market," *Equality, Diversity and Inclusion: An International Journal*, vol. 37, pp. 138–150, Feb. 2018.

[5] P. H. Cappelli, "Skill Gaps, Skill Shortages, and Skill Mismatches: Evidence and Arguments for the United States," *ILR Review*, vol. 68, no. 2, pp. 251–290, Mar. 2015.

[6] A. M. Osman-Gani and Y. Paik, "Factors influencing the retention of international IT talent: an empirical investigation in Singapore," *International Journal of Human Resources Development and Management*, vol. 16, no. 1-2, pp. 1–17, Jan. 2016.

[7] W. G. Johnson, R. Butler, G. Harootunian, B. Wilson, and M. Linan, "Registered Nurses: The Curious Case of a Persistent Shortage," *Journal of Nursing Scholarship*, vol. 48, no. 4, pp. 387–396, 2016.

[8] A. N. Joinson, A. Woodley, and U.-D. Reips, "Personalization, authentication and self-disclosure in self-administered internet surveys," *Computers in Human Behavior*, vol. 23, no. 1, pp. 275–285, 2007.

[9] K. Kenthapadi, S. Ambler, L. Zhang, and D. Agarwal, "Bringing salary transparency to the world: Computing robust compensation insights via linkedin salary," in *ACM CIKM*, 2017, pp. 447–455.

[10] K. Kenthapadi, A. Chudhary, and S. Ambler, "LinkedIn Salary: A System for Secure Collection and Presentation of Structured Compensation Insights to Job Seekers," in *2017 IEEE (PAC)*, Aug. 2017, pp. 13–24.

[11] X. Chen, Y. Liu, L. Zhang, and K. Kenthapadi, "How LinkedIn Economic Graph Bonds Information and Product: Applications in LinkedIn Salary," in *ACM SIGKDD*. New York, NY, USA: ACM, 2018, pp. 120–129.

[12] S. Jackman and G. Reid, "Predicting job salaries from text descriptions," Ph.D. dissertation, University of British Columbia, 2013.

[13] G. S. Becker, *Economic Theory*, 2nd ed. New Brunswick: Transaction Publishers, Jun. 2007.

[14] M. Friedman, *Price Theory*. Routledge, Jul. 2017.

[15] A. Smith, *An Inquiry Into the Nature and Causes of the Wealth of Nations*. W. Strahan; and T. Cadell, 1778.

[16] J. M. Keynes, *The General Theory of Employment, Interest, and Money*, Dec. 2016.

[17] W. Arulampalam, A. L. Booth, and M. P. Taylor, "Unemployment Persistence," *Oxford Economic Papers*, vol. 52, no. 1, pp. 24–50, 2000.

[18] R. E. Hall, "Employment Efficiency and Sticky Wages: Evidence from Flows in the Labor Market," *The Review of Economics and Statistics*, vol. 87, no. 3, pp. 397–407, Aug. 2005.

[19] B. Kaufman and J. Hotchkiss, *The Economics of Labor Markets (Dryden Press Series in Economics) 5th edition. (2000) Hardcover*, 5th ed. Harcourt College Pub, 1705.

[20] C. A. Pissarides, "Equilibrium in the Labor Market with Search Frictions," *The American Economic Review*, vol. 101, no. 4, pp. 1092–1105, 2011.

[21] D. T. Mortensen and C. A. Pissarides, "Job Creation and Job Destruction in the Theory of Unemployment," *The Review of Economic Studies*, vol. 61, no. 3, pp. 397–415, 1994.

[22] S. Machin and A. Manning, "A Test of Competitive Labor Market Theory: The Wage Structure among Care Assistants in the South of England," *ILR Review*, vol. 57, no. 3, pp. 371–385, Apr. 2004.

[23] G. G. Cain, "The Challenge of Segmented Labor Market Theories to Orthodox Theory: A Survey," *Journal of Economic Literature*, vol. 14, no. 4, pp. 1215–1257, 1976.

[24] F. Wilkinson, *The Dynamics of Labour Market Segmentation*. Elsevier, Oct. 2013.

[25] H. Bauder, *Labor Movement: How Migration Regulates Labor Markets*. Oxford University Press, Feb. 2006.

[26] W. T. Dickens and K. Lang, "The Reemergence of Segmented Labor Market Theory," *The American Economic Review*, vol. 78, no. 2, pp. 129–134, 1988.

[27] C. A. Pissarides, *Equilibrium Unemployment Theory*, second edition edition ed. Cambridge: The MIT Press, Feb. 2017.

[28] N. Kaldor, "A Classificatory Note on the Determinateness of Equilibrium," *The Review of Economic Studies*, vol. 1, no. 2, pp. 122–136, Feb. 1934.

[29] Y. Igarashi, "Distributional effects of hiring through networks," *Review of Economic Dynamics*, vol. 20, pp. 90–110, Apr. 2016.

[30] A. Calv-Armengol and M. O. Jackson, "Networks in labor markets: Wage and employment dynamics and inequality," *Journal of Economic Theory*, vol. 132, no. 1, pp. 27–46, Jan. 2007.

[31] Y. M. Ioannides and L. D. Loury, "Job Information Networks, Neighborhood Effects, and Inequality," *Journal of Economic Literature*, vol. 42, no. 4, pp. 1056–1093, 2004.

[32] R. L. Axtell, O. A. Guerrero, and E. Lpez, "Frictional unemployment on labor flow networks," *Journal of Economic Behavior & Organization*, vol. 160, pp. 184–201, Apr. 2019.

[33] Y. Singer, "Budget Feasible Mechanism Design," *SIGecom Exch.*, vol. 12, no. 2, pp. 24–31, Nov. 2014.

[34] G. J. Borjas, "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *The Quarterly Journal of Economics*, vol. 118, no. 4, pp. 1335–1374, Nov. 2003.

[35] B. T. Heim, "The Incredible Shrinking Elasticities: Married Female Labor Supply, 1978-2002," *The Journal of Human Resources*, vol. 42, no. 4, pp. 881–918, 2007.

[36] M. K. Chen, "Dynamic Pricing in a Labor Market: Surge Pricing and Flexible Work on the Uber Platform," in *ACM EC*. New York, NY, USA: ACM, 2016, pp. 455–455, event-place: Maastricht, The Netherlands.

[37] A. Hannk, C. Wagner, D. Garcia, A. Mislove, M. Strohmaier, and C. Wilson, "Bias in Online Freelance Marketplaces: Evidence from TaskRabbit and Fiverr," in *ACM CSCW*. New York, NY, USA: ACM, 2017, pp. 1914–1933, event-place: Portland, Oregon, USA.

[38] D. S. Hamermesh and S. J. Trejo, "The Demand for Hours of Labor: Direct Evidence from California," *The Review of Economics and Statistics*, vol. 82, no. 1, pp. 38–47, 2000.

[39] D. Acemoglu, D. Autor, and D. Lyle, "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury," *Journal of Political Economy*, vol. 112, no. 3, pp. 497–551, Jun. 2004.

[40] C. D. Goldin, "The Role of World War II in the Rise of Women's Employment," *The American Economic Review*, vol. 81, no. 4, pp. 741–756, 1991.

[41] M. S. Feldstein, "Estimating the Supply Curve of Working Hours," *Oxford Economic Papers*, vol. 20, no. 1, pp. 74–80, 1968.

[42] K. B. Clark and R. B. Freeman, "How Elastic is The Demand for Labor?" National Bureau of Economic Research, Working Paper 309, 1979.

[43] J. J. Horton and L. B. Chilton, "The Labor Economics of Paid Crowdsourcing," in *ACM EC*. New York, NY, USA: ACM, 2010, pp. 209–218, event-place: Cambridge, Massachusetts, USA.

[44] A. J. Berinsky, G. A. Huber, and G. S. Lenz, "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk," *Political Analysis*, vol. 20, no. 3, pp. 351–368, 2012.

[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[46] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[47] M. Ezekiel, "The Cobweb Theorem," *The Quarterly Journal of Economics*, vol. 52, no. 2, pp. 255–280, Feb. 1938.

[48] R. B. Freeman, "A Cobweb Model of the Supply and Starting Salary of New Engineers," *Industrial and Labor Relations Review*, vol. 29, no. 2, pp. 236–248, 1976.

[49] L. F. K. a. O. Blanchard, "Regional Evolutions," Jan. 1992.

[50] R. J. B. a. X. Sala-I-Martin, "Convergence Across States and Regions," Jan. 1991.

[51] M. Chamberlain, "Demystifying the gender pay gap: Evidence from glassdoor salary data," 2016.

[52] F. D. Blau and L. M. Kahn, "The gender wage gap: Extent, trends, and explanations," *Journal of Economic Literature*, vol. 55, no. 3, pp. 789–865, 2017.