Deconstructing Job Search Behavior*

Stefano Banfi
Ministry of Energy, Chile

Sekyu Choi
University of Bristol

Benjamín Villena-Roldán
CEA, DII, University of Chile,
SMAUG, MIPP

March 28, 2017

Abstract

In this paper we empirically investigate job search, specifically how a number of theoretically relevant variables impact behavior in an online setting. We take advantage of an unusually rich proprietary dataset from a Chilean job board to document and interpret a number of facts. We focus on how application behavior is influenced by (1) several demographics such as gender, age, and marital status, (2) alignment between applicants wage expectations and job ad wage offers, (3) applicant fit into job ad requirements in terms of education and experience, (4) timing variables, including unemployment duration, job tenure (for on-the-job searchers), and vacancy duration. We relate our results to a variety of theoretical models and discuss how our findings can be used to discipline current (and future) job search models.

Keywords: Online job search, Applications, Search frictions, Unemployment, On-the-job search, Networks.

JEL Codes: E24, J40, J64

Very preliminary and incomplete. Please do not circulate.

*Email: stefano.banfi@gmail.com, sekyu.choi@bristol.ac.uk and bvillena@dii.uchile.cl. We thank Jan Eeckhout, Shouyong Shi, Yongsung Chang, Francois Langot and seminar participants at Cardiff University, The University of Manchester, Diego Portales University, the 2016 Midwest Macro Meetings, the 2016 LACEA-LAMES Workshop in Macroeconomic and the Search & Matching workshop at the University of Chile for insightful discussion and comments. All errors are ours.

1 Introduction

Despite the rise in the importance of the internet in the labor market, actual details on how individual job search behavior looks like remain elusive. Although seeking jobs on line is different in several regards from other job search methods, its importance has being increasing over time, as well as its efficiency, as documented for instance by Kuhn and Mansour (2014).

In this paper, we use information from `www.trabajando.com`, a job posting website with presence in most of Latin America, in addition to Spain and Portugal. We use a comprehensive dataset on daily applications of job seekers to job postings in the Chilean labor market during the first half of 2013. The main salient feature of the data is the detailed information the website maintains on both sides of the market: we observe education, occupations and experience (among other characteristics) for individuals and for job postings (as requirements stipulated by firms). Moreover, we observe both desired and current wages for individuals (wages of last full time jobs if unemployed) and the wages firms expect to pay at jobs they are posting (this information can be made public or not by each side of the market).

The richness of our data allow us to provide evidence on many theoretically important factors that determine job search, but are generally unavailable to researchers. A first research question is how individuals, facing a set of online job ads, choose to apply to some jobs and forgo others. To do so, we estimate application decision equations, in which we try to disentangle the contribution of a large array of factors influencing the application choices we observe.

To overcome the fact that we only observe effective applications and not the entire set of relevant job positions for each candidate,[1] we use the observed *networks* formed by individual applications, where job seekers are the nodes of the network, and we define a link between nodes as having applied to the same job position. We then construct the choice set of an applicant w as the list of all job ads applied by seekers linked with w. This approach relies in reveled preferences of applicants to define similarities between jobs, instead of using adhoc criteria regarding relevant dimensions of a job.

Our empirical approach uncovers several interesting patterns with respect to job seekers' application decisions and search efforts, measured by the probability of applying to a position in the relevant set, conditional on observables. We find that some demographic characteristics are relevant (marital status and gender), and we also document the impact of age in search behavior, a factor that has been shown to be empirically relevant by Choi et al. (2015) and Menzio et al. (2016) (among others) and important for unemployment insurance design, as studied in Michelacci and Ruffo (2015).

[1] Not observing page views in the website is a shortcoming in the literature using online job board data. Unfortunately, `www.trabajando.com` nor other job search boards keep records of page views by applicant for two reasons: (i) it is very expensive to keep these records while the information is of little use (for the job board operators), and (ii) applicants need to be logged in when viewing job ads, a requirement that would reduce the likelihood of getting applicants into the board. See the references below.

We find that individuals are more likely to apply for a job offering a wage close to their expectations. This result holds even if one or both sides of the market choose to not disclose wage offers and/or wage expectations, respectively. The fact that workers react to hidden information confirms previous findings in Banfi and Villena-Roldan (2016) using the same database, on evidence of directed search. Moreover, search behavior is highly sensitive to the requirements of educational level and experience. We find that qualifications of applicants are aligned to requirements of job ads to which they apply. Nevertheless, we also find some asymmetries regarding this alignment: The probability of an application peaks when the applicant is slightly underqualified in terms of education but the pattern is reversed in the case of experience requirements. We also study how labor market status (employed vs unemployed) affects job seeking behavior. In particular, our results show that unemployed workers tend to have a better fit to qualifications than their employed counterparts and that they apply with higher probabilities to job postings in which they are over-qualified and vice versa (compared to employed seekers).

We also investigate the impact of unemployment duration (for unemployed seekers) and job tenure (for those performing on-the-job search) on the choices made by applicants. This evidence is particularly useful to understand the dynamic evolution of unemployed workers over an unemployment spell, an important input for the design of unemployment insurance policies, an aspect also considered by Faberman and Kudlyak (2013). The effect of tenure on job search is also relevant to understand factors behind job-to-job transitions, arguably an important mechanism to explain wage dispersion, as stated in Hornstein et al. (2011)).

Our results also show that individuals respond to indications of the likelihood of receiving an offer given an application: they prefer job postings where the number of advertised vacancies is higher, and dislike those which have been open for a longer period of time. This is evidence of individuals reacting to "phantoms" vacancies, a main motivation in Albrecht et al. (2015) and Chéron and Decreuse (2016). Early evidence by van Ours and Ridder (1992) suggests that applications arrive shortly after the vacancy is open.[2]

Our paper is related to several others which use data from online job-posting/search websites in order to study different aspects of frictional markets. Kudlyak et al. (2013) study how job seekers direct their applications over the span of a job search. They find some evidence on positive sorting of job seekers to job postings based on education and how this sorting worsens the longer the job seeker spends looking for a job (the individual starts applying for worse matches). Marinescu and Rathelot (2015) use information from `www.careerbuilder.com` and find that job seekers are less likely to apply to jobs that are farther away geographically. Marinescu and Wolthoff (2015) use the same job posting website to study the relationship between job titles and wages posted on job advertisements. They show that job titles explain nearly 90% of the variance of explicit wages. Gee (2015), using a large field experiment on the job posting website `www.Linkedin.com`, shows that

[2]This intuition is also supported by informal discussions with managers of `www.trabajando.com`

being made aware of the number of applicants for a job, increases ones own likelihood of making application.

On product markets, Lewis (2011) shows that internet seekers for used cars significantly react to posted information regarding automobiles' quality. Jolivet and Turon (2014) and Jolivet et al. (2016) use information from a major French e-commerce platform, `www.PriceMinister.com`, to study the effects of search costs and reputational issues (respectively) in product markets.

2 The data

We use data from `www.trabajando.com` (henceforth the website) a job search engine with presence in mostly Spanish speaking countries: as of March of 2017, the list comprises Argentina, Brazil, Colombia, Chile, Mexico, Peru, Portugal, Puerto Rico, Spain, Uruguay and Venezuela. Our data covers a sample of job postings and job seekers in the Chilean labor market, between January 1st 2008 and December 24th, 2016. The raw information in the dataset contains more than 14 million single applications, from around 1.5 million job seekers, to around 270 thousand job ads.

Our dataset has detailed information on both applicants and recruiters. First, we observe entire histories of applications from job seekers and dates of ad postings (and repostings) for recruiters. Second, we have detailed information for both sides of the market. For job seekers we observe date of birth, gender, nationality, place of residency (”comuna” and ”región”, akin to county and US state, respectively), marital status, years of experience, years of education, college major and name of the granting institution of the major.[3] We have codes for occupational area of the current/last job of the individual, information on its salary and both its starting and ending dates. In terms of the website's platform, job seekers can use the site for free, while firms are charged for posting ads. A major caveat of our dataset, is the absence of information on activities outside the website: individuals seeking for jobs through other means, and more importantly, outcomes of job search.

For each posting, we observe its required level of experience (in years), required education (required college major, if applicable), indicators on required skills (specific, computing knowledge and/or ”other”) how many positions must be filled, an occupational code, geographic information and some limited information on the firm offering the job: its size (number of employees) and its industry. Educational categories are *primary* (one to eight years of schooling), *high school* (completed high school diploma), *technical tertiary education* (professional training after high school), *college* (completed university degree) and *post-graduate* (any schooling higher than university degree).

A novel feature of the dataset, compared to the rest of the literature, is that the website asks job seekers to record their desired salary, which they can then choose to show or hide from prospective employers. Recruiters are also asked to record the expected pay for the job posting, and given the same choice whether to make this information visible or not to the applicants.

[3] This information is for any individual with some post high school education.

For the remainder of the paper, we restrict our sample to consider only individuals working under full-time contracts and those unemployed. We further restrict our sample to individuals aged 25 to 55. We discard individuals reporting desired net wages above 5 million pesos.[4] This amounts to approximately 10,300 USD per month[5], which represents more than double the 90th percentile of the wage distribution, according to the 2013 CASEN survey.[6] We also discard individuals who desire net wages below 210 thousand pesos (around 435 USD) a month (the legal minimum wage during the time). Consequently, we also restrict job postings to those offering monthly salaries within those bounds.

The unit of analysis are individual *applications*. We do not have information on individuals who create accounts but never apply to any jobs nor on job postings that don't receive any applications. We restrict our sample to individuals who were actively looking for a job (i.e., made an application) and job postings which were active (ad was available and received at least 1 application) during the time window. While we observe long histories of job search for a significant fraction of workers (some workers have used the website for several years), we consider only applications pertaining to their last job search "spell", which we define as all the applications made by job seekers *after* their last modification/update of their online curriculum vitae. We further drop individuals who apply to only one job position in our considered sample, or who apply to more positions than the 99-th percentile of job applicants.

Table 1 shows descriptive statistics for the job searchers in our sample. From the table we observe that the average age is 34.3 and that job seekers are comprised of mostly single males, with 57.1% being unemployed ($23,316$ unemployed seekers from a total of $40,813$ individuals.). Average experience hovers around 8 years. Job seekers in our sample are more educated than the average in Chile, with 46.55% of them having a college degree, compared to 25% for the rest of the country in the same age group (30 to 44), (The figure is from the 2013 CASEN survey.) although there is a big discrepancy by labor force status: unemployed seekers are significantly less educated in the website.

From the table we can also observe that most job seekers have studies related to management (around 25%) and technology (around 27%) and that average expected wages are approximately (in thousands) CLP\$ $1,170$ and CLP\$ 658 for employed and unemployed seekers, respectively. For comparison, the minimum monthly salary in Chile was 210 thousand CLP during year 2013.

In terms of search activity, an individual searches for around 62.49 days, which we measure as the time between an individual last updated her online cv, and the last observed application date

[4] A customary characteristic of the Chilean labor market is that wages are generally expressed in a monthly rate net of taxes, and mandatory contributions to health (7% of monthly wage), to fully-funded private pension system (10%), to disability insurance (1.2%), and mandatory contribution to unemployment accounts(0.6%)

[5] Using average nominal exchanges rate between January and July of 2013, http://www.x-rates.com/average/?from=CLP&to=USD&amount=1&year=2013.

[6] CASEN stands for "Caracterización Socio Económica" (Social and Economic Characterization), and aims to capture a representative picture of Chilean households.

Table 1: Characteristics of Job Seekers

	Employed	Unemployed	Total
Demographics			
Age	34.38	34.29	34.33
Fracion males	0.64	0.56	0.60
Fraction married	0.36	0.30	0.33
Experience (years)	8.66	8.40	8.51
Wages (thousand CLP)	1170	658	878
Tenure in last/current job (months)	35.81	32.55	33.99
Unemployment duration (days)	–	388	388
Education level (%)			
Primary (1-8 years)	0.12	0.21	0.17
High School	15.03	32.38	24.94
Technical Tertiary	24.77	29.27	27.34
College	58.7	37.44	46.55
Post-graduate	1.38	0.70	0.99
Occupation (%)			
Management	27.78	23.04	25.07
Technology	31.61	23.94	27.23
Not declared	15.85	34.76	26.65
Rest	24.76	18.26	21.05
Search Activity			
Days searching (website)	74.78	53.26	62.49
Number of Applications	2.83	2.84	2.83
Observations	17,497	23,316	40,813

in the website. The amount of time searching for a job is higher for those employed: 74 versus 53 days for unemployed. In terms of applications, both groups show very similar choices, with around 2.83 submitted applications.

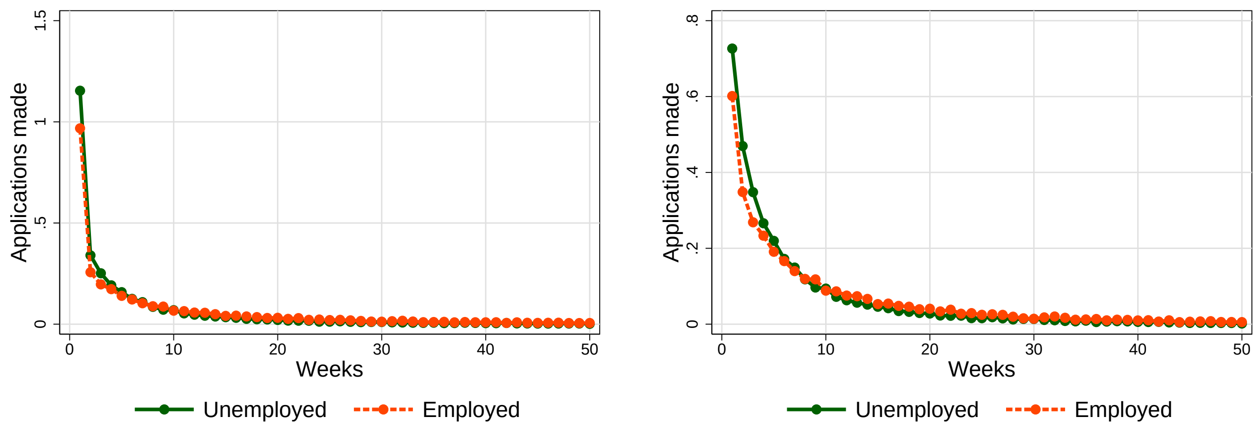
3 Empirical analysis

3.1 Life-cycle of a job search

In this section, we show some descriptive statistics computed from the dataset, where we exploit the fact that we observe both sides of the market and also the precise timing (date and hour) of submitted applications.

In figure 1 we show the number of applications submitted by individuals, by week of their job search, or more specifically, by week of website usage. We define this timing given the earliest and latest observed applications by individuals. The earlier does not always match the date of account creation in the website. Since we observe a significant fraction of individuals applying to jobs only during their first day of website usage (and a significant fraction, concentrate applications to the first hour), in panel (b) of the same figure we restrict attention to job seekers who make applications for at least two weeks. For the remainder of this section, our sample is restricted to those searchers, to avoid noise introduced by seekers who apply just one day.

As seen in the figure, the pattern of application decisions is declining in job search duration and



(a) All seekers

(b) Search for more than 1 week

Figure 1: Submitted applications by week of job search. All sample and sample restricted to individuals searching for more than 7 days in the website.

is very similar across employment status of the job seeker. A similar phenomena is found also by Faberman and Kudlyak (2013). As mentioned above, the difference between both panels in figure 1 is due to a fraction of individuals concentrating their job applications in just one day: this could be evidence that some individuals browse all their desired postings, before submitting applications all at once.

In terms of where do job seekers direct their applications, figure 2 shows an interesting phenomena. Since most job searches are bunched at the beginning of the time span of website use, these particular applications might be different from applications on following weeks. In the figure we show for each week of applications, the average number of days the ads individuals have applied to have been online (their average *online life*) along an adjusted polynomial of the number of weeks (as a non-linear trend). At the start of the job search, individuals seem to apply to all their most preferred ads, which can be randomly distributed in the distribution of "days online". After applying to all these job positions (with an average online life of 10 to 11 days), individuals start applying to only the newest job ads they can find (drop in all figures from week 1 onwards). If individuals continue searching, the ads they start considering widens, which explains an increasing trend in *online life* of ads for successive applications by week.

In figures 3 to 5 we present a similar exercise using the information on both job seeker characteristics and job posting requirements for education (levels), experience (years) and occupation (1 digit categories). For education and experience, we compute the simple difference between what is required by the job position minus the characteristic of the job seeker. Hence, positive differences mean that the requirements of the job are higher than what the candidate possesses while negative differences mean that the job seeker is over-qualified for the position in that particular dimension. With occupation, we create a dummy variable that equals one if the occupation in the job posting

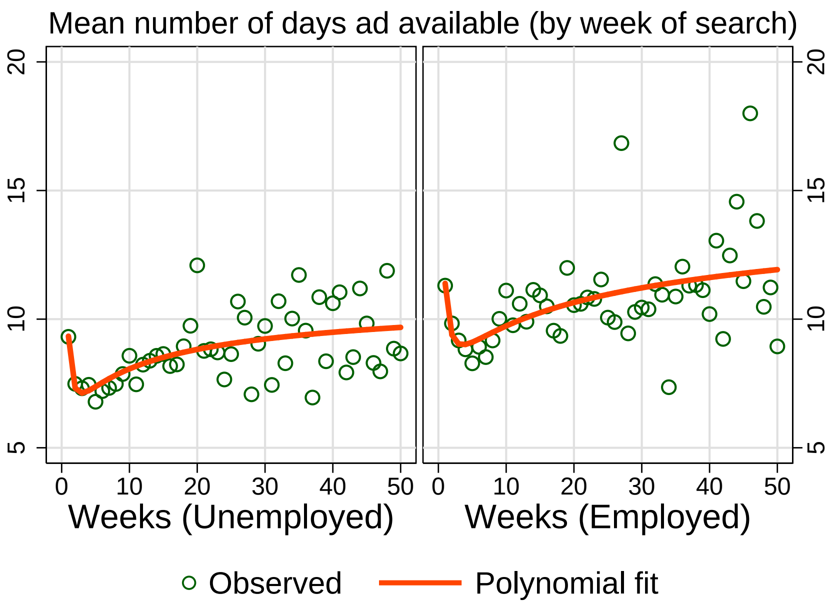


Figure 2: Average number of days an ad has been available on the website, by week of application of individuals, separated by labor force status (unemployed/employed).

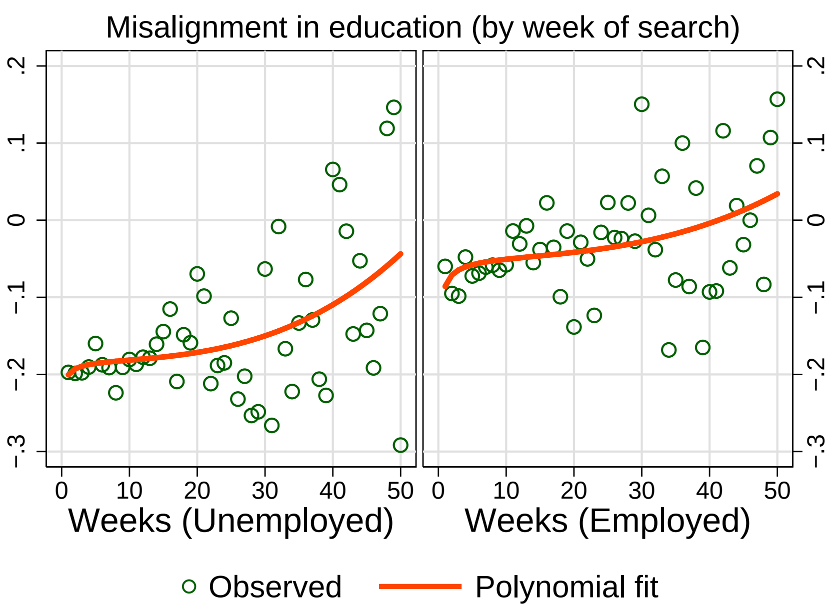


Figure 3: Difference in years of schooling between job posting requirement and job seeker characteristics, by week of application of individuals, separated by labor force status (unemployed/employed).

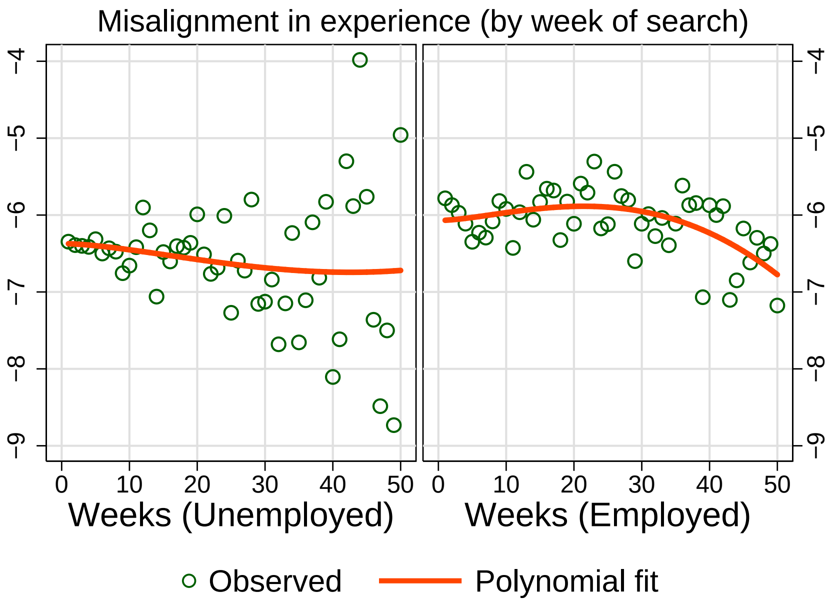


Figure 4: Difference in years of experience between job posting requirement and job seeker characteristics, by week of application of individuals, separated by labor force status (unemployed/employed).

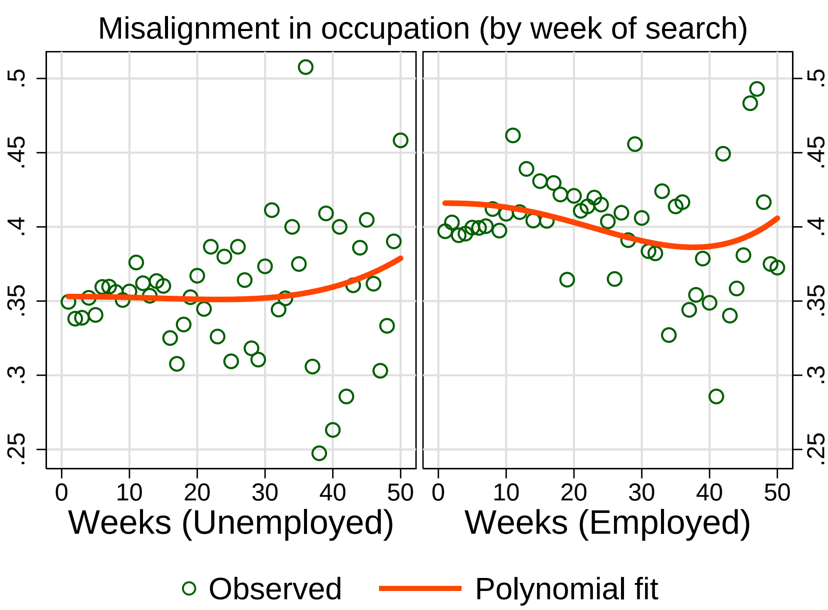


Figure 5: Difference in occupation (dichotomic variable) of job posting vs job seeker characteristic, by week of application of individuals, separated by labor force status (unemployed/employed).

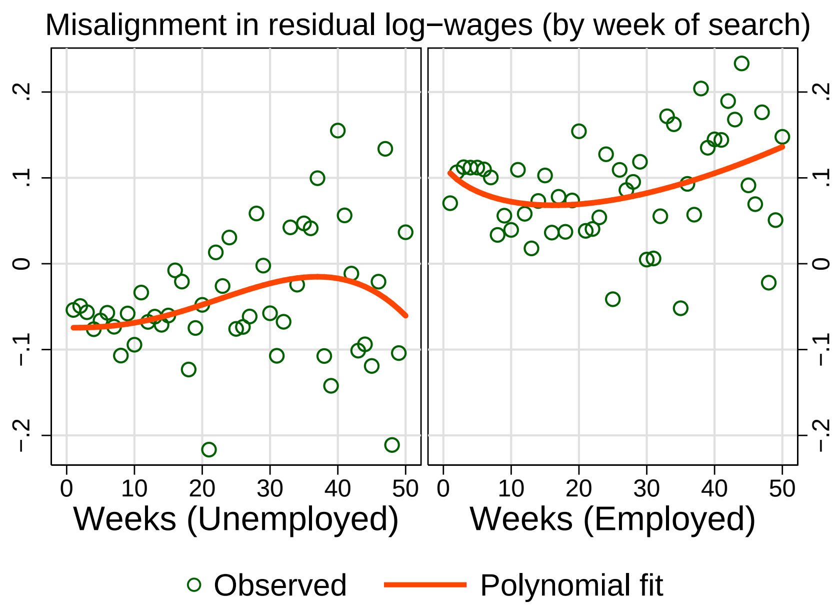


Figure 6: Difference in residual log-wages of job posting (expected to pay) vs job seeker (desired wage), by week of application of individuals, separated by labor force status (unemployed/employed).

is different from the occupation of the job seeker.

For each individual and each week of the job search life-cycle, we compute the averages of the *misalignment* measures from the above paragraph, for unemployed and employed seekers who submitted applications in that particular week. The figures confirm and expand the results in Kudlyak et al. (2013), with regards to quality of matches in terms of applications, as a function of job search duration: as time passes, individuals who have not found a job start sending applications to job postings which are farther away from their optimal desired position. This can be seen clearly from the fact that as weeks go by, the observed level of *misalignment* in applications in all relevant dimensions increases (moves away from zero), measured by the polynomial trend in each figure. There are two notable exceptions: education levels for the unemployed and occupation *misalignment* for the employed. However, at higher durations, both of these "exceptions" fit the idea that the quality of matches in applications worsens. Nevertheless, all figures show that job seekers move away from their week one target dimensions (education, experience and occupations) relaxing their search parameters. An additional feature from these figures, is the increase in dispersion observed in application decisions as a function of weeks searching. In all the figures, the spread of observed *misalignment* averages exhibits increasing dispersion.

To make a more systematic analysis, in figure 6 we perform the same exercise, but comparing levels of worker and job *type*, where we approximate the latter using simple wage regressions: for both workers and job postings, we estimate linear regressions between wages (desired and expected, for individuals and job positions, respectively) on a number of observable characteristics. From each regression, we then compute the residual, which we identify as the type of the worker/job. The figure shows two very interesting patterns: First, there is a significant level different in week one worker-job *misalignment* across employment status. Unemployed seekers seem to target positions for which their type is higher than the type of the job, which is reflected in initial negative numbers in the left panel of figure 6. For on-the-job seekers, the *misalignment* is positive: working seekers apply to positions with types higher than their own. A possible explanation for this is that unemployed individuals want to make sure of landing a job, by applying to things that they should be more likely to be matched to. On the other hand, individuals who have already have a job, could be searching to climb the wage/job/occupation ladder. The second pattern, is a nonlinear trend in type-*misalignment* with respect to weeks searching. For both types of workers, *misalignment* decreases with time (gets closer to zero), but starts increasing after some time: this happens at around weeks 35 and week 15 for unemployed and employed seekers respectively.

3.2 Market segmentation and match preferences

In this section we analyze empirically match preferences of heterogeneous job seekers (w) when confronted with heterogeneous ads (a). However, our dataset only contains information on actual applications and no direct information is collected by the website in terms of searches nor actual *clicks* on job postings by individuals. Although these pieces of information could be used to identify the relevant segment of the labor market for each job seeker, what a job seeker browses before making applications could be a very noisy signal of what her target job is.

Market segmentation through network analysis. In what follows, we use the network formed by all job seekers (w) to determine which job postings (a) are relevant to them. The idea behind this exercise, is to uncover individual preferences for job characteristics using a revealed preference argument. The key step is identifying job prospects which fit the job seeker but which were not considered (applied to).

Assume that each individual represents a node in the network, and that a link between nodes is defined as *having applied to the same job posting*. For each w, we can define the set of relevant job postings $\mathcal{A}_w^1$ as the union of all job postings applied by the set of all job seekers linked to w. This is what we define as a network of degree 1, since for each individual, we only consider their immediate links (1 degree of separation).

Following this logic, from the network of degree 0 we obtain the original dataset for individual w ($\mathcal{A}_w^0$), since the network contains only information of job seekers and their applications (no information on links is used). On the other hand, a network of degree 2 is defined as the network which considers both job seekers linked directly to w, in addition to those who are linked with the links of w (all job seekers have 2 degrees of separation), giving rise to dataset $\mathcal{A}_w^2$. We can continue with this logic iteratively, until forming the set $\mathcal{A}_w^\infty$, which is the cross between each job seeker w

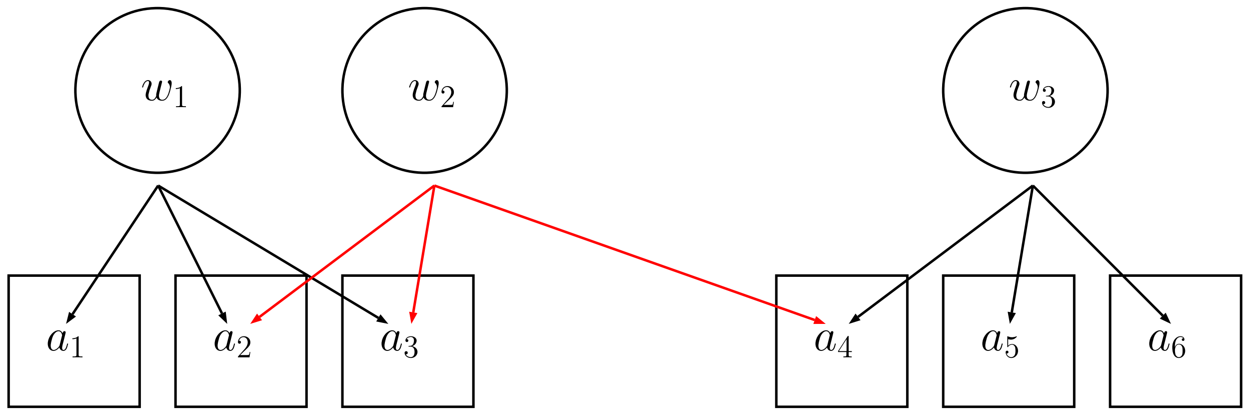


Figure 7: Example of a network formed by workers $\{w_1, w_2, w_3\}$. Worker w_1 is linked to worker w_2 by common applications to ads a_2 and a_3 but is not linked with w_3 in the network of degree 1. All workers are linked in the network of degree 2.

and all job postings a.

Figure 7 shows an example of the network algorithm and the resulting datasets. In the figure there are three workers, $\{w_1, w_2, w_3\}$ and six job postings, $\{a_1, a_2, a_3, a_4, a_5, a_6\}$. Consider worker w_1. She has applied to three jobs, thus $\mathcal{A}^0_{w_1} = \{a_1, a_2, a_3\}$ and is linked to w_2 through applications to $\{a_2, a_3\}$. Since w_2 also applied to job position a_4, one can infer that some characteristic of a_4 is not desirable to w_1. If we consider networks of degree 1, a_4 would be included in the set of relevant ads for the first worker. Notice also that in this example, w_1 is not directly linked with w_3, or in our language, the degree of separation between these two workers is higher than 1.

Again, considering the first worker, we have $\mathcal{A}^0_{w_1} = \{w_1, w_2, w_3\}$, and as discussed above, $\mathcal{A}^1_{w_1} = \{a_1, a_2, a_3, a_4\}$. Given that w_1 and w_2 are linked and that w_2 is linked with w_3, the relevant job ads for w_1, given a network of degree 2, is $\mathcal{A}^2_{w_1} = \{a_1, a_2, a_3, a_4, a_5, a_6\}$. In our simple example, the network of degree 2 is already the "exploded" network (all ads to all workers).

For each type of network, we restrict job postings to those that are actually available during the time that each job seeker is active in our dataset. Given this network procedure, we are able to construct a dataset where we can compare the characteristics of individuals and ads, when the individual made an application decision or not, and thus, estimate the relative importance individuals put in different characteristics of the job.

In table 2, we present information on the resulting number of relevant job postings per worker, given networks of degrees one and two (distance of separation between linked workers). As mentioned above, the network of degree 0 is basically our original dataset, which contains information only on job applications. The median number of relevant job postings (a) is 2 postings per job seeker (the same number applies for both employed and unemployed seekers), ranging between individuals who applied to only 2 ads, to workers who applied to 8 ads. On the other extreme, we have the network of degree 2, which spans a dataset where the median number of relevant postings

Table 2: Number of relevant ads (a) per worker (w)

Network degree	Mean	Median	St. Dev.	Min	Max
	Employed seekers				
0	2.83	2	1.29	2	8
1	18.17	12	17.89	2	198
2	86.99	31	133.4	2	1284
	Unemployed seekers				
0	2.84	2	1.3	2	8
1	13.44	9	14.26	2	197
2	69.77	12	172.55	2	2286

Notes: The table shows the number of relevant job postings per job seeker given a network of different degree (see main text). *Degree 0* refers to the original dataset (no network).

are 12 and 31, for employed and unemployed seekers respectively. The number of *relevant* job ads ranges from 2 to around $1,284$ for those employed, and from 2 to $2,286$ for the unemployed.

Preferences over heterogeneous characteristics. On each of the datasets created from the network approach, we estimate match preferences for job seekers, based on their observed characteristics along the ones posted by ads which are relevant to them. More specifically, we estimate a linear regression of the form

$$y_{aw} = X_{aw}\beta_{aw} + \sum_{k_c}\sum_{p=1}^{\bar{P}}\{\beta_{k_op}(z_{k_o})^p\} + \sum_{k_d}\{\beta_{k_d}z_{k_d}\} + \sum_{k}\sum_{\ell}\mathbf{1}_{\{k\neq\ell\}}\beta_{k\ell}z_kz_\ell + \epsilon_{aw} \tag{1}$$

where y_{aw} is a dummy variable that takes the value of one if job seeker w applies to posting a, and zero otherwise. The regression has two sets of explanatory variables: First, X_{aw} contains observed job and worker characteristics, which do not overlap, e.g., demographic characteristics of the seeker, number of vacancies the posting is offering, etc. In this set, we include polynomials for the age of the job seeker and for the amount of time (measured in days) in either the current job (for employed individuals) or in unemployment (for unemployed seekers).

Second, we include a set of controls for the *misalignment* z between characteristics required by posters vs. the characteristics of the job seeker. For continuous variables, which we denote by k_c, and encompass the level of education, years of experience and log wages, we define z_{k_c} as the simple difference between the value of the characteristic required by the position and value of the characteristic possessed by the job seeker. For discrete variables k_d (occupation), the distance z_{k_d} is defined as a dummy that takes the value of one when the category in the job posting is different from the characteristic of the worker.

In equation (1), for each of the continous dimensions, we include in the regression a polynomial of order $\bar{P} = 5$ to assess whether non-linearities exists in the effect of z_{k_c} on application decisions.

The basic idea is to try to understand if agents apply differently if they are *over*-qualified ($z < 0$) compared to when they are *under*-qualified ($z > 0$).

We estimate the above equation separating our sample between the employed and unemployed, in order to assess whether on-the-job search differs from unemployed search behavior. We further perform the estimation of a similar equation, but restricting the sample by labor force status and age/job tenure/unemployment duration groups. In this way, we aim to disentangle whether age and path dependence (time spent on the current job or unemployment duration) affects job application decisions in non trivial ways. We proceed by separating the sample between those employed and those unemployed, and by quartiles of age/job tenure/unemployment duration (separately). In each of these sub-samples, we estimate equation (1) after removing the controls for age/job tenure/unemployment duration in X_{aw}.

Table 3 shows results from estimating equation (1) using ordinary least squares, under different degrees of separation in the underlying network (degrees 1 and 2). The table shows coefficients for X_{aw} variables, and for comparability, the values are displayed as fractions of the unconditional application probability in each column. Increasing the degrees of separation expands the number of relevant postings per seeker (see table 2) while the number of actual applications remains constant, thus, y_{aw} decreases. In turn, the linear regression coefficients are affected by the scale (number of total observations) in the estimation. Results related to polynomials on *distance* terms and their interactions with path dependence are presented below. The considered dimensions are level of education, years of experience, log wages and occupational codes.

In the regression we also control for extra observable characteristics of both job seekers and posting firms. For seekers, we include a gender dummy and quintic polynomial for age and dummies for different marital states. For postings, we include controls for firm size, dummies for firm industry, a dummy variable if the posting entity is a recruiting/head hunter firm and controls for specific requirements included in the ad (e.g., specific computer knowledge). Also, since we are considering applications between 2008 and 2016, we include quarterly dummies to control for time effects.

From the table we obtain several interesting results which are unearthed given the richness of our data. On the side of individuals, we find that married agents tend to apply to fewer positions if they are searching on the job, while the effect of marital status is not significant for the unemployed seekers. We find a strong gender application gap, since both employed and unemployed males apply with higher propensity than females to job positions, everything else constant. In terms of path dependence, the table refines the evidence in figure 1 above, in that individuals who have spent more days in the website, measured by the time between the last online CV update and the date of application, apply to less ads at the margin.

On the other hand, our results show that individuals who choose to be explicit about their wage expectations at the time of an application, do not act differently from those who do not. For

Table 3: Regression Results

	Employed		Unemployed	
	Degree 1	Degree 2	Degree 1	Degree 2
married	-0.2743	-0.7751	0.1135	0.1906
	(0.1707)	(0.0014)	(0.2570)	(0.0836)
male	0.0256	0.0824	0.0452	0.1054
	(0.0539)	(0.0000)	(0.0000)	(0.0000)
explicit wage (w)	0.0022	-0.0058	-0.0119	-0.0120
	(0.8458)	(0.6348)	(0.2241)	(0.2583)
explicit wage (a)	-0.0728	-0.1026	0.0215	0.0577
	(0.0002)	(0.0000)	(0.1407)	(0.0005)
No. of vacancies	0.0004	0.0016	0.0016	0.0055
	(0.7051)	(0.2129)	(0.0159)	(0.0000)
Days searching (w)	-0.0016	-0.0034	-0.0013	-0.0025
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Days since post (a)	-0.0008	-0.0018	-0.0008	-0.0017
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Diff Occupation	-0.2031	-0.5702	-0.1597	-0.4238
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Average y_{aw}	15.3452	3.1342	19.4709	4.8588
Observations	170,790	836,191	169,720	680,132

Notes: Regression coefficients from a linear regression on application decisions. Dependent variable is y_{aw}, a dummy for the existence of a job application. Estimated coefficients are shown as fractions of unconditional application probabilities (average y_{aw}). P-values in parenthesis. Degree refers to the type of network originating the estimation dataset. Each regression controls also for polynomials and interactions in *mismatch* (see main text) as well as age of the worker, firm size, contract type, dummmies for different types of requirements of the job and industry of the firm.

job postings, table 3 shows that an explicit wage in the job ad affects negatively the decision to apply for employed seekers, but it attracts more unemployed applications. Also, the information contained in the job posting seems to provide significant information to job seekers: they react positively to posts advertising higher number of vacancies and negatively to posts which have been longer online.

Age and path dependence. Given the results above, we can show whether there are life-cycle profiles in the decision to apply for job postings in the website. In what follows we show results only for the estimation associated to the network of degree 1, but results for the network of degree 2 are very similar.[7] For comparability, we present all figures as a fraction of the unconditional mean

[7]These graphs and results are available upon request.

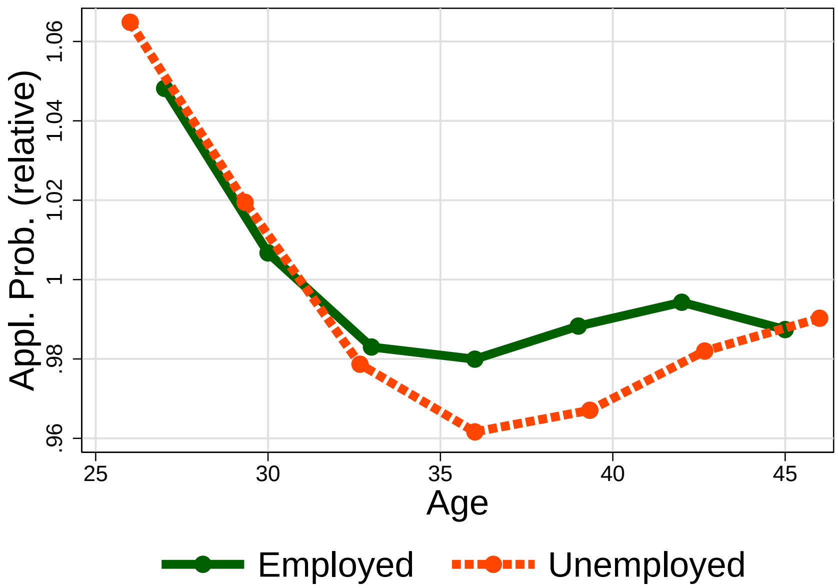


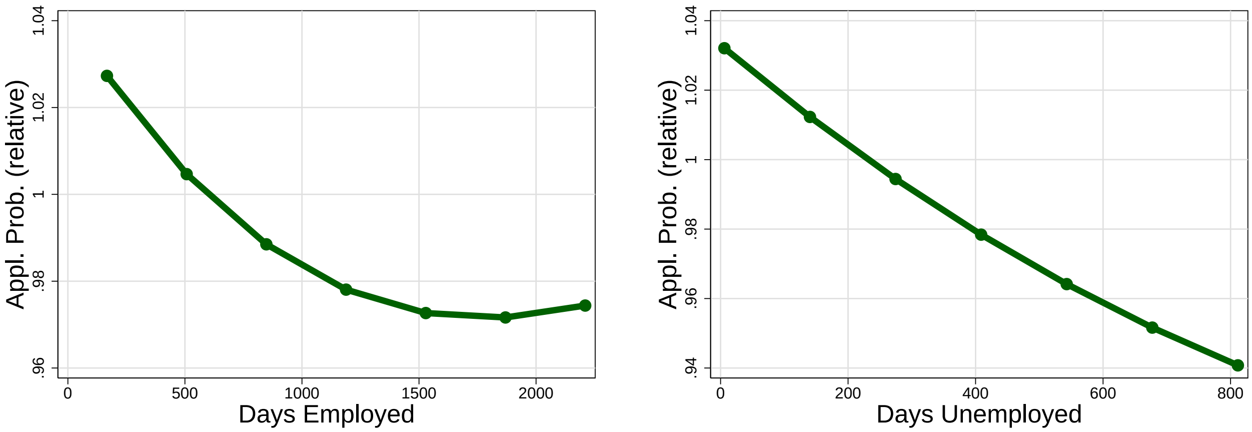
Figure 8: Predicted application probabilities for different ages, given results from eq. (1). The figure is the computed using the coefficients associated to a polynomial of order 5 on age of the applicant. Results are presented relative to unconditional application probability means and are based on a network of degree 1.

of an application for the respective sample.

Figure 8, we observe the life-cycle profile of job application decisions implied by the estimation results in equation (1): given regression coefficients on the polynomial of order 5 for the age of the job seeker, we can predict application probabilities for different ages, when the rest of variables remain at sample means. Note that since we included quarterly dummies in the regression, these results are subject only to time effects, but not cohort effect correction,

As seen from the figure, life-cycle profiles of applications decisions are very similar by labor force status, with both being downward sloping. This evidence is consistent with findings in Choi et al. (2015) and Menzio et al. (2016), among others, with respect to job finding rates and employment to employment transitions over the life-cycle. As seen from the figure, life-cycle effects alter application probabilities, making them between 6% more likely than average (for younger seekers), to 4% less likely for unemployed individuals aged 35-40.

In figure 9, we present the implied profiles of application decisions given time in the current job (for employed seekers, in the left panel) and time unemployed (for those unemployed, in the right panel). Our results show that job applications decisions for employed individuals are declining in job tenure, leveling of at around 4 years of tenure. The effects range between plus and minus 4% of mean application probabilities. In our sample, the median job tenure for on-the-job seekers is 692 days (around 1.9 years), which would imply that the average employed seeker applies to ads with *less* intensity than the average.



(a) Job tenure

(b) Unemployment duration

Figure 9: Predicted application probabilities, given number of days in the current job (left panel) or number of days unemployed (right panel). Results from a polynomial of order 5 for the respective variable (tenure/unemployment duration) in eq. (1). Results are relative to unconditional application probability means and are based on a network of degree 1.

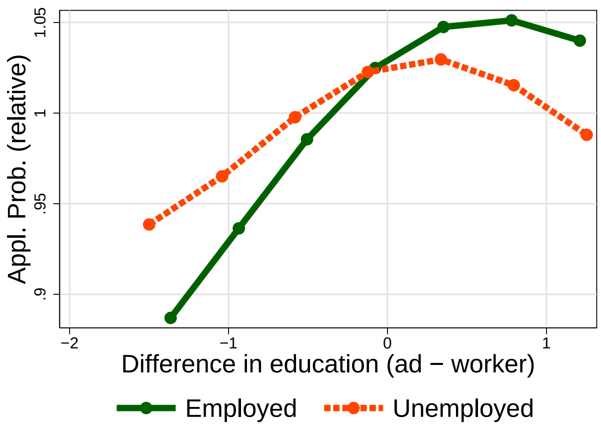
For unemployed seekers, the results are similar, but the profile of job applications is declining throughout the entire span of unemployment duration. The median unemployment duration in our sample is 166 days, which implies that the median unemployed seeker applies with *more* intensity to job postings than the average.

Misalignment and applications. Below, we present graphically the results of the effect of *distance* (z_{k_c}) in continuous characteristics (education, experience and log wages) on application decisions.

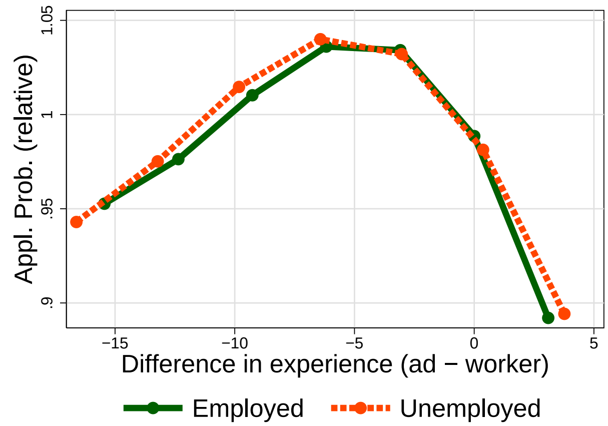
Figure 10 shows predicted application probabilities ($\hat{y}_{aw}$ from equation 1), when a particular continuous dimension varies (z_{k_c}), keeping all other observables at their sample mean (including the *misalignment* in other dimensions). The considered range is bounded by the sample mean of z_{k_c}, plus and minus its standard deviation. Again, for comparability reasons, the predicted probability is presented as a fraction of the unconditional mean of an application for each case (employed vs. unemployed samples).

As seen in the figure, job seekers in both labor states align themselves with requirements and/or characteristics of job postings. This is represented by a bell shaped function and a maximum predicted application probability (all else constant). However, this alignment is not exact: for education, job seekers tend to maximize application probabilities at levels slightly above their own, while both employed and unemployed seekers align themselves fairly well to job ad requirements in terms of experience. For the case of log-wages, the alignment seems to be very close, which is surprising since the majority of job postings do *not* reveal this information (see table **??**).

In terms of magnitude, the figure reveals that differences in log wages are the ones that affect the



(a) Education levels



(b) Years of experience

(c) log wages

Figure 10: Predicted application probabilities, given results from eq. (1) and different levels of misalignment in the selected variable. Results are relative to unconditional application probability means and are based on a network of degree 1.

most application probabilities: for unemployed seekers, the relative probabilities fluctuate between 90% and 110% of the unconditional mean, while for employed seekers, the range is wider, from around 70% to 120%. With the exception of experience levels, the ranges are wider for employed seekers than for those unemployed.

The current exercise uncovers two extra novel facts when studying application strategies of job seekers: non-linearities with respect to misalignment in some characteristics and asymmetries between the strategies of the employed versus the unemployed.

With respect to non-linearities, panels related to education and experience in figure 10 show that job seekers react differently to ads for which they are under or overqualified and that differential behavior is different in the case of education and experience. This follows from the bell shape in each of the panels, which are not symmetric around their peak. Job seekers tend to apply less to jobs to which they are over qualified in terms of education, compared to jobs for which they are under qualified; On the other hand, they tend to apply less to jobs for which they are under qualified in terms of experience, when we compare to those jobs for which they are over-qualified.

Turning to asymmetries between employed and unemployed seekers, all figures show that unemployed individuals apply slightly more (compared to their employed counterparts) when they are over qualified, with the exception of experience levels. In terms of the figures, this is reflected by the fact that the bell curves for the unemployed cross the ones for employed from the left and above. Conversely, our results show that employed individuals are more likely to apply for jobs with higher educational and/or experience requirements (and promised expected wages) than similar seekers but who are unemployed.

Misalignment, applications and time.

Next we turn our attention to the interaction between misalignment and both life-cycle forces and path dependence (measured by time in the current labor market state) and the effect of such interactions on job application decisions. The main goal of this exercise is to understand whether individuals react differently to misalignments in continuous characteristics if they are in different age groups, or different stages of the search process (on-the-job or search from unemployment).

As explained above, we estimate several times equation (1), but for different quartiles of years of age, job tenure and unemployment duration (the latter two measured in days). The actual percentiles for each variable (25, 50 and 75) for each estimation sample are presented in table 4. These numbers define the quartiles we use to restrict the estimation samples, from which we produce application probabilities as in figure 10.

Figure 11 shows the exercise for differences in log wages, when we separate the sample by different quartiles in the age, job tenure and unemployment duration distributions. As seen from the figure, all different quartiles predict a very similar pattern (bell shaped curve) of application probabilities given misalignment between offered wages by job positions and expected wages by

Table 4: Percentiles in estimation sample

	Employed sample		
	P_{25}	P_{50}	P_{75}
Age	30	34	40
Job tenure	356	692	1,261
	Unemployed sample		
Age	28	33	40
Unemployment duration	69	166	373

Notes: These percentiles define the quartiles which separate the estimation sample for restricted regressions.

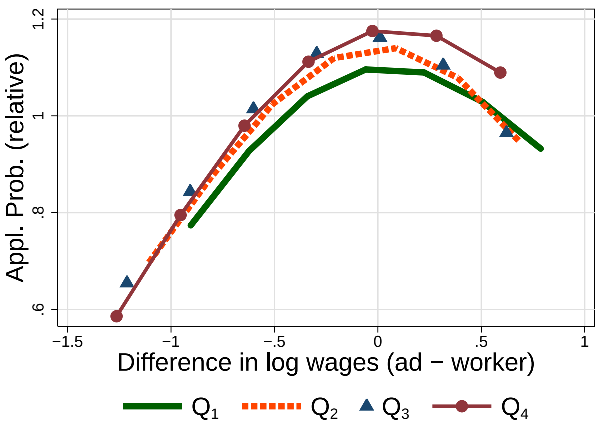
individuals. This is true for both different ages and time spent in the current labor force status for workers.

The figure also shows the existence of life-cycle effects on application probabilities in the face of *misalignment* in log-wages: this is most clear when we observe panel (b), for unemployed seekers and different age quartiles: when unemployed job seekers get older, two things happen: first, application probabilities fluctuate more, since the range of relative probabilities increases when we move from the first (Q1) to the last (Q4) age quartiles. Second, the average level of *misalignment* moves to the left, in that older unemployed workers seem to be more willing to apply to jobs with salaries below their stated desired wages. A similar, albeit but more muted effect can be seen in panel (a), for the case of on-the-job searchers and age quartiles.

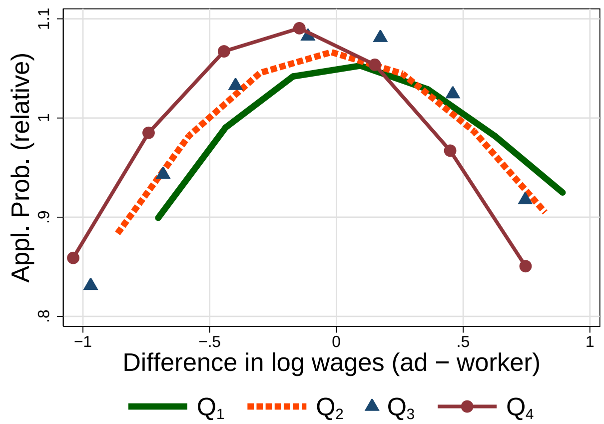
In terms of education, figure 12 shows some interesting patterns. With respect to path dependence, panels (c) and (d) of the figure show the existence of non-linear effects of both job tenure and unemployment duration on the willingness of job seekers to apply to jobs for which they are *misaligned* in terms of education.

For employed seekers, the longer their job tenure in their current position (comparing lines Q1 and Q4 in panel c), has the effect of flattening the curve of predicted applications at different levels of education *misalignment*. However, this effect is non monotonic, as observed by how lines Q2 and Q3 behave with respect to lines Q1 and Q4. On panel (d), we observe the opposite effect, when analyzing the effect of unemployment duration on application decisions: as we move up in the distribution of unemployment duration, application decisions have a wider range of probabilities (curve Q4 in panel d has the highest range) and at the same time, the maximal application probability shifts to the right: seekers in this situation apply more to jobs for which they are over-qualified.

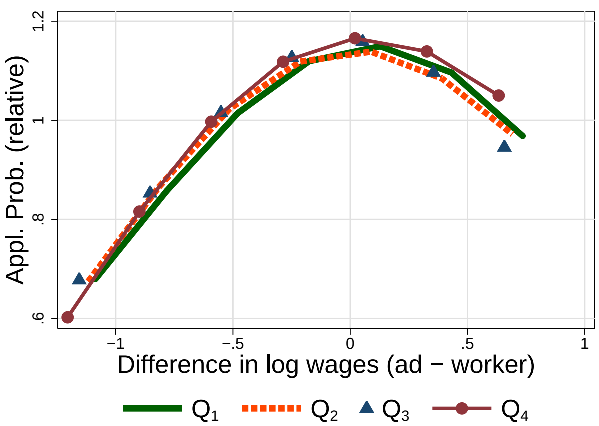
As for life-cycle effects, panels (a) and (b) do not show a clear pattern. In panel (a), the effect of age on the job decisions of employed seekers imply a lower response of application decisions in the presence of *misalignment* in education (flatter Q4 curve in that panel), but as in panel (c), the



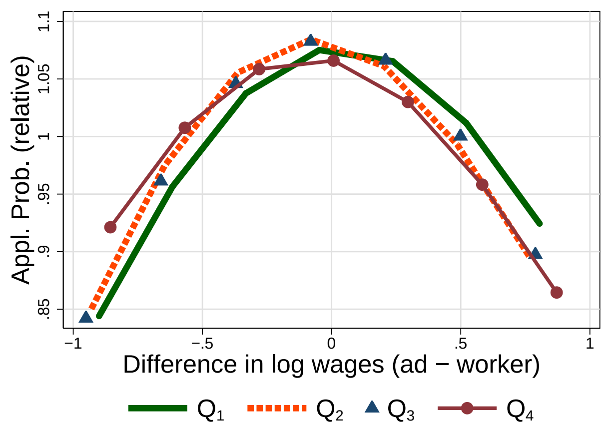
(a) Age quartiles - Employed



(b) Age quartiles - Unemployed

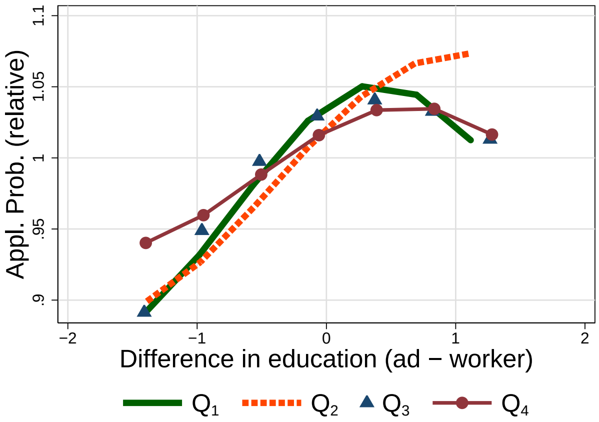


(c) Job tenure quartiles

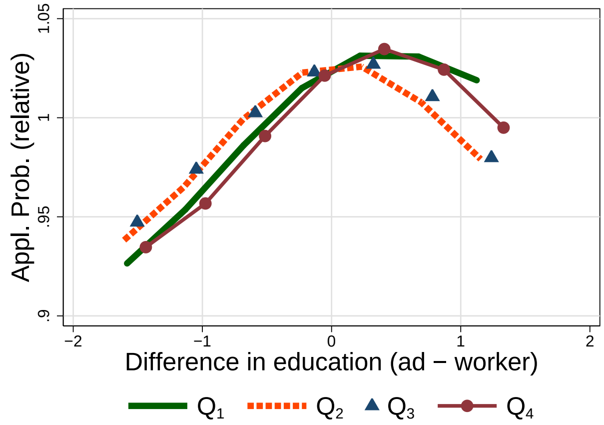


(d) Unemployment duration quartiles

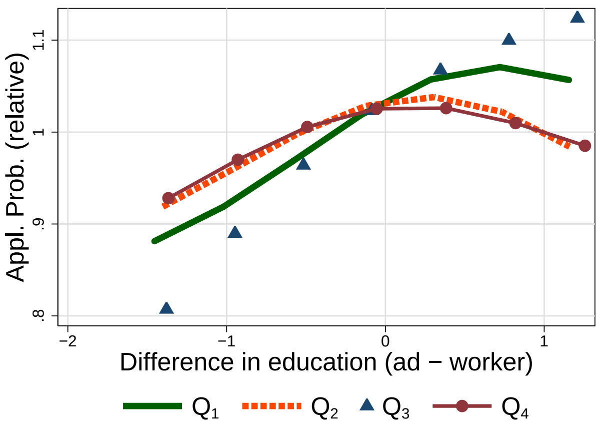
Figure 11: Predicted application probabilities, given results from eq. (1) on different samples, defined by age, job tenure and unemployment duration quartiles. All lines are relative to the unconditional application probability of each sub-sample.



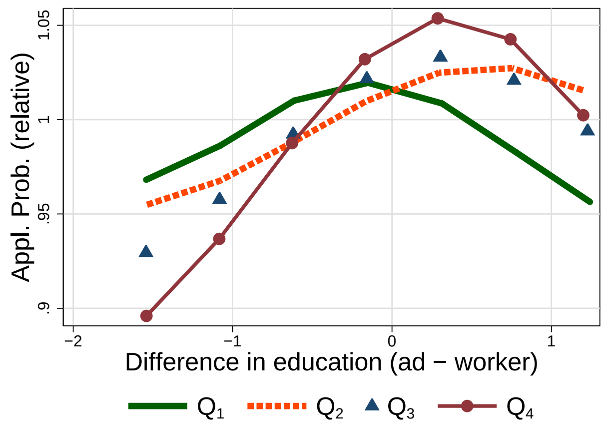
(a) Age quartiles - Employed



(b) Age quartiles - Unemployed



(c) Job tenure quartiles

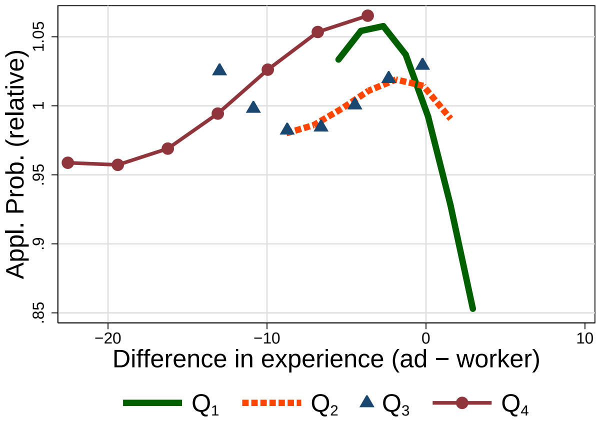


(d) Unemployment duration quartiles

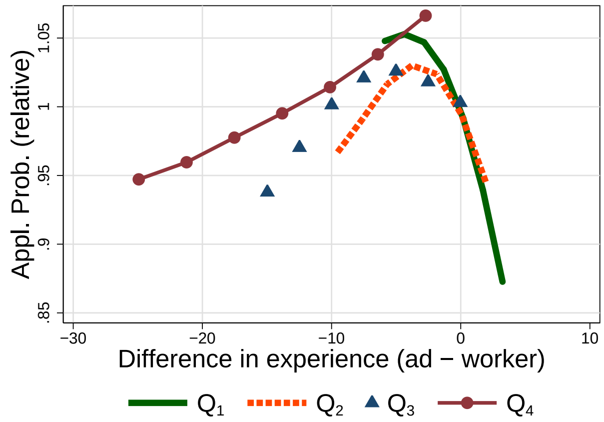
Figure 12: Predicted application probabilities, given results from eq. (1) on different samples, defined by age, job tenure and unemployment duration quartiles. All lines are relative to the unconditional application probability of each sub-sample.

effects seems to be non-linear. Life-cycle effects for unemployed individuals are not very different when comparing different age groups (Q1 and Q4 in panel b).

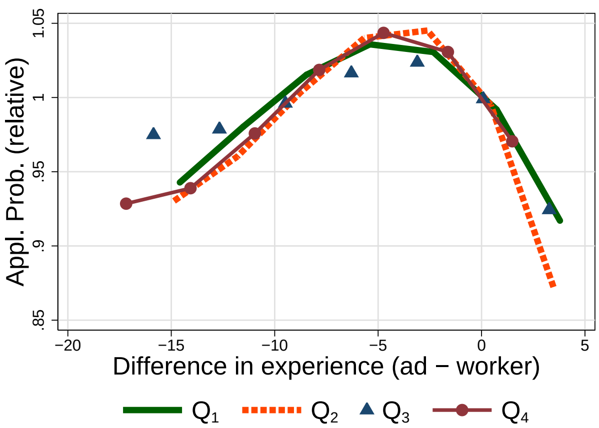
As for differences in required versus individual experience levels, figure 13 shows the exercise for this case, where we find the most dramatic differences in job search and application behavior given different life-cycle and path dependence effects. In panels (a) and (b) of the figure, we observe that the higher the age group, the application probability curve becomes flatter. The curves for different quartiles reflect also a mechanical effect of older workers having more experience (on average), and thus, finding themselves in situations where the difference between required experience and their own experience is larger than for the average ages. This effect is clearly seen in both panels, in which there is a natural progression of the application curves for each subsequent age quartile, becoming flatter each time and centered more and more towards the left (more negative values of



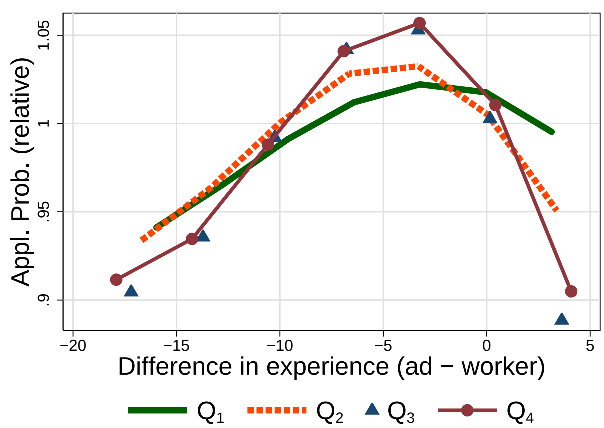
(a) Age quartiles - Employed



(b) Age quartiles - Unemployed



(c) Job tenure quartiles



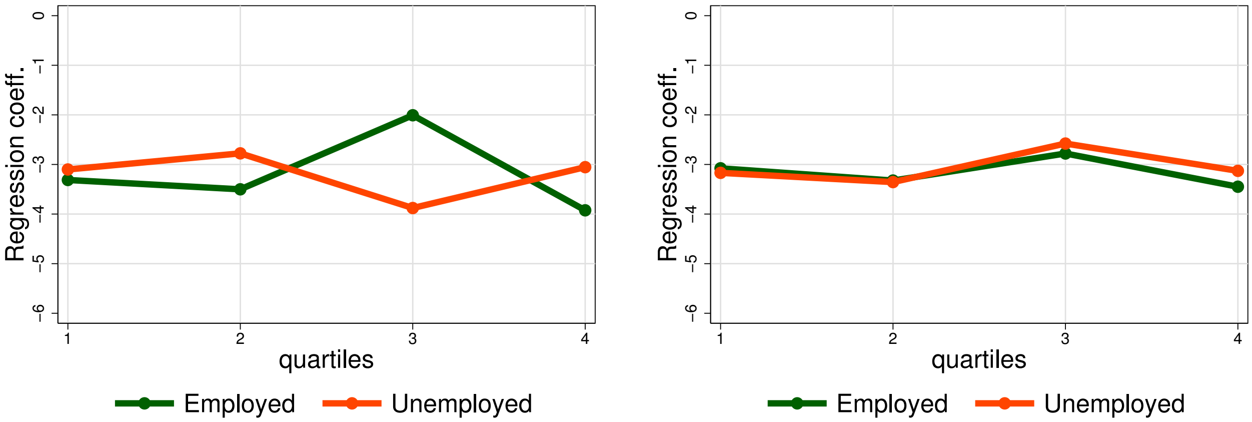
(d) Unemployment duration quartiles

Figure 13: Predicted application probabilities, given results from eq. (1) on different samples, defined by age, job tenure and unemployment duration quartiles. All lines are relative to the unconditional application probability of each sub-sample.

the *misalignment* experience level).

Panel (c) of figure 13 shows the exercise where we consider individuals with increasing job tenure. The panel shows a very muted effect of job-tenure on application decisions in the face of experience *misalignment* between job posting and individual characteristics.

Finally, in panel (d) we show application probabilities of unemployed individuals in different stages of their unemployment duration spell. As observed from the figure, as job seekers experience longer spells of unemployment, their application decisions become more focused at some average level. This can be attained when comparing curves Q1 versus Q4 of that panel, where Q4, the bell curve describing application decisions for those who have been unemployed the longest, becomes narrower than Q1. This means that individuals with longer unemployment durations are less likely to apply to jobs for which they are either too over or under qualified, and concentrate efforts in



(a) Age quartiles

(b) Job tenure/Unemployment duration quartiles

Figure 14: Effect of different occupation (ad versus job seeker) on application probabilities, given results from eq. (1) on different samples, defined by age, job tenure and unemployment duration quartiles.

some target level.

When discussing the effect of misalignment in categorical variables, as is the case with differences in occupational codes over the probability of an application, we cannot distinguish relative effects but only a extensive margin effect. In what follows, we show regression coefficients for the dummies of *misalignment* in occupations, which take a value of one if the application is to the same 1 digit code occupation as her self-reported occupation.

Figure 14 shows our results for *misalignment* in occupations. In both panels, we can observe that applications to a different occupations as the self-proclaimed one by individuals, are associated with very negative (and significant) effects, as measured by the regression coefficient for that dummy variable. In terms of life-cycle effects, the regression coefficient for different age quartiles show a non-monotonic effect, but without a clear pattern. As for job tenure or unemployment duration effects, the estimated coefficients are very stable.

4 Conclusions

Using data from a Chilean job posting website, in this paper we uncover several facts regarding the timing and nature of job applications. We find first hand evidence of stock-flow search behavior by individuals, as well as demographic heterogeneity in how individuals search for jobs. We also show how job seekers in different labor force statuses react to *misalignment* in key dimensions between own characteristics and characteristics required by job postings (level of education, years of experience, required occupation and log-wages) and find that there is significant alignment between requirements and characteristics, even though they could be hidden from each side of the market (in the case of wages mostly). Finally, we also study the interplay between *misalignment* and both life-cycle and job tenure/unemployment duration effects.

References

Albrecht, J., B. Decreuse, and S. Vroman (2015). Directed search with phantom vacancies.

Banfi, S. and B. Villena-Roldan (2016). Do high-wage jobs attract more applicants? directed search evidence from the online labor market. Cea wp 327, University of Chile.

Chéron, A. and B. Decreuse (2016). Matching with phantoms. *Review of Economic Studies* (0), 1–30.

Choi, S., A. Janiak, and B. Villena-Roldán (2015). Unemployment, Participation and Worker Flows Over the Life-Cycle. *The Economic Journal 125*(589), 1705–1733.

Faberman, R. J. and M. Kudlyak (2013). The Intensity of Job Search and Search Duration. mimeo, Federal Reserve Bank of Richmond.

Gee, L. K. (2015). Information Effects on Job Application Rates by Gender in a Large Field Experiment. Mimeo, Tufts University.

Hornstein, A., P. Krusell, and G. L. Violante (2011). Frictional Wage Dispersion in Search Models: A Quantitative Assessment. *The American Economic Review 101*(7).

Jolivet, G., B. Jullien, and F. Postel-Vinay (2016). Reputation and Prices on the e-Market: Evidence from a Major French Platform. *International Journal of Industrial Organization, forthcoming*.

Jolivet, G. and H. Turon (2014). Consumer search costs and preferences on the internet. Mimeo, University of Bristol.

Kudlyak, M., D. Lkhagvasuren, and R. Sysuyev (2013). Systematic job search: New evidence from individual job application data. mimeo, Federal Reserve Bank of Richmond.

Kuhn, P. and H. Mansour (2014). Is Internet Job Search still Ineffective? *The Economic Journal 124*(581), 1213–1233.

Lewis, G. (2011). Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors. *The American Economic Review 101*(4), 1535–1546.

Marinescu, I. E. and R. Rathelot (2015). Mismatch Unemployment and the Geography of Job Search. Mimeo.

Marinescu, I. E. and R. P. Wolthoff (2015, May). Opening the Black Box of the Matching Function: The Power of Words. Discussion Paper 9071, IZA.

Menzio, G., I. A. Telyukova, and L. Visschers (2016). Directed search over the life cycle. *Review of Economic Dynamics 19*, 38 – 62. Special Issue in Honor of Dale Mortensen.

Michelacci, C. and H. Ruffo (2015, February). Optimal Life Cycle Unemployment Insurance. *American Economic Review 105*(2), 816–59.

van Ours, J. and G. Ridder (1992). Vacancies and the Recruitment of New Employees. *Journal of Labor Economics 10*(2), 138–155.