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

October 10, 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, the Search & Matching workshop at the University of Chile and the 2017 SAM Meetings in Barcelona, Spain 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 (minimized search cost for example), its importance has being increasing over time, as well as its efficiency, as documented for instance by Kuhn and Mansour (2014). The recent focus of equilibrium unemployment in macroeconomic models[1] calls for new evidence to disentangle the mechanisms that are at play when individuals decide how, when and what type of job to search for.

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 period 2008 to 2016. Our dataset contains some features which, to the best of our knowledge, are novel to the profession. 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 for individuals and for job postings (as requirements stipulated by firms). Moreover, we observe detailed information of seeker and job characteristics, as well as 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.

The richness of our data allow us to provide evidence on many theoretically important factors that determine job search, but are generally unobserved by researchers. The main research question in this paper 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 disentangle the contribution of a large array of factors influencing application choices.

To overcome the fact that we only observe effective applications and not the entire set of relevant job positions for each candidate,[2] we use the *network* formed by individual applications. Taking job seekers as individual nodes of a network, we define a link between nodes as the case when two job seekers have applied to at least one common job. Then, we construct the choice set of an applicant as the list of all ads applied by seekers linked with her. Our methodology creates sets of job ads for each individual where we observe variation in application decisions, overcoming the problem of only observing effective applications in the original data.

This network approach relies on revealed preferences of applicants to define similarities between jobs, instead of defining the relevant local market to the job seeker with a particular geographic

[1]See Rogerson, Shimer, and Wright (2005) for a survey.

[2]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.

location or an occupational/industry category, as is the case with Şahin, Song, Topa, and Violante (2014) and Herz and van Rens (2015). By considering the latter rules, the choice set of individuals which we could consider would not allow for either occupational or geographical mobility of workers (or both jointly), which may be important issues to consider.[3]

Our empirical exercise reveals interesting patterns with respect to job seekers' application decision making and search effort, measured by the probability of applying to a position in the relevant set. We find that some demographic characteristics are quite relevant at the margin: men apply to more job ads than women, while married individuals apply less than their single counterparts if they are searching on the job. We also document that search effort decreases with age, which is consistent with evidence in Choi, Janiak, and Villena-Roldán (2015) and Menzio, Telyukova, and Visschers (2016), among others. This decrease in effort is of crucial importance for unemployment insurance design, as studied in Michelacci and Ruffo (2015).

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. The latter is evidence of individuals reacting to "phantom" vacancies, a main motivation in Albrecht, Decreuse, and Vroman (2015) and Chéron and Decreuse (2016).[4] Closely related to this, we find evidence of "stock-flow" matching behavior:[5] new job seekers in the website (the flow) seek to match with the observed stock of job adverts during their initial time on the platform. When time passes, the inflow of job seekers becomes part of the stock of individuals, who then try to match with the new flow of job positions.

One of our main findings relates to how job seekers align themselves with heterogeneous types of jobs. We find that search behavior is highly sensitive to the requirements of educational level and experience and that job seekers target an optimal or most preferred type of job, which is not necessarily the one that matches perfectly their current characteristics: 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 (years). An implication of this result, is that job seekers' incentives to apply to jobs that are further away from the most preferred job (in terms of these characteristics) decrease, although this decrease is not symmetrical.

In terms of wages, 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 or *implicit* information confirms previous findings in Banfi and Villena-Roldan (2016) using

[3]See for example, Carrillo-Tudela and Visschers (2014).

[4]Early evidence by van Ours and Ridder (1992) suggests that applications arrive shortly after the vacancy is open. This intuition is also supported by informal discussions with managers of `www.trabajando.com`.

[5]See Taylor (1995), Coles and Muthoo (1998), Coles and Smith (1998) and Ebrahimy and Shimer (2010), among others.

the same database. Going one step further, and given that in our empirical approach we control for an exhaustive array of observable characteristics, one could argue that wages in our setup are proxies for inherent *types* of both workers and firms. Thus, our results are suggestive of positive assortative matching patterns at the application stage, but more importantly, summarizes well the main distinction between search strategies of the unemployed versus the employed: while the unemployed target wage offers (types) that are very close to their own stated salary expectations, employed seekers target wages (types) which are on average above their expectations. Our reading from this is that the unemployed are trying to maximize the chances of obtaining a job offer, while the employed (performing on-the-job search) are more likely trying to climb the job ladder.

Our results also show that application decisions decline with either unemployment duration (for unemployed seekers) and job tenure (for those performing on-the-job search). 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, Krusell, and Violante (2011)).

Our paper is related to a growing literature which use data from online job-posting/search websites in order to study different aspects of frictional markets. Kudlyak, Lkhagvasuren, and Sysuyev (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 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, Jullien, and Postel-Vinay (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 September 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.[6] 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. Job advertisements are posted for a minimum of thirty days, but firms can pay additional fees to extend this period to sixty and ninety days.

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. Naturally, one could question the reliability of wage information which will be ultimately hidden from the other side of the market. Banfi and Villena-Roldán (2016) address the potential issue of ”nonesensical” wage information in job ads by comparing the sample of explicit vs. implicit (job ads without any salary information) postings by firms, and find that observable characteristics predict fairly well implicit wages. In the case of job seekers... [COMPARISON BETWEEN JOB SEEKER'S EXPECTATIONS AND NESI?!] On the other hand, a major caveat of our dataset is the absence of information on activities performed outside the website: individuals seeking for jobs through other means, and more importantly, outcomes of job applications.

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

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

to 55. We discard individuals reporting desired net wages above 5 million pesos.[7] This amounts to approximately 10,300 USD per month[8], which represents more than double the 90th percentile of the wage distribution, according to the 2013 CASEN survey.[9] 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 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 the time betweenthe last modification/creation of their online curriculum vitae (cv) and the time of their last submitted application. Since individuals maintain information about their last job in their online profile, as well as contact information and salary expectations, we assume that any modification of this information is done primarily when individuals who are currently working or who have already used the website are ready to search in the labor market again. We further drop individuals who apply to only one job position in our considered sample, or who are above the 99-th percentile of job applicants in term of number of submitted applications (this leads to a maximum of 8 submitted applications by individual).

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 2013.[10]

[7] 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%)

[8] 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.

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

[10] The minimum wage has increased to 270 thousand CLP, by July 2017. See http://www.dt.gob.cl/consultas/1613/w3-article-60141.html.

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 terms of search activity, the average search spell amounts to 62.49 days. The amount of time searching for a job is higher for those employed than for the unemployed: 74 versus 53 days, respectively. In terms of applications, both groups show very similar choices, with around 2.83 submitted applications.

3 Life-cycle of a job search

In this section we present some descriptive statistics showing the overall process of job applications by individuals using the website. Below we present information about job search activity during a standard job search spell, which confirms evidence found elsewhere in the literature, but also uncovers some new evidence.

In figure 1 we show the average number of applications submitted by individuals, by week of their job search, or more specifically, by week of website usage (since last modification of their online profile). As seen in the figure, the pattern of application decisions is declining in job search duration and is very similar across employment status of the job seeker. A similar phenomena is found also by Faberman and Kudlyak (2013). We can also observe from the graph that there is a significant difference between the number of applications submitted in the first week versus the rest of weeks. This is due to the fact that a significant fraction of individuals concentrate their job applications in just one day: In our sample, 32.4% of individuals apply for jobs only during their

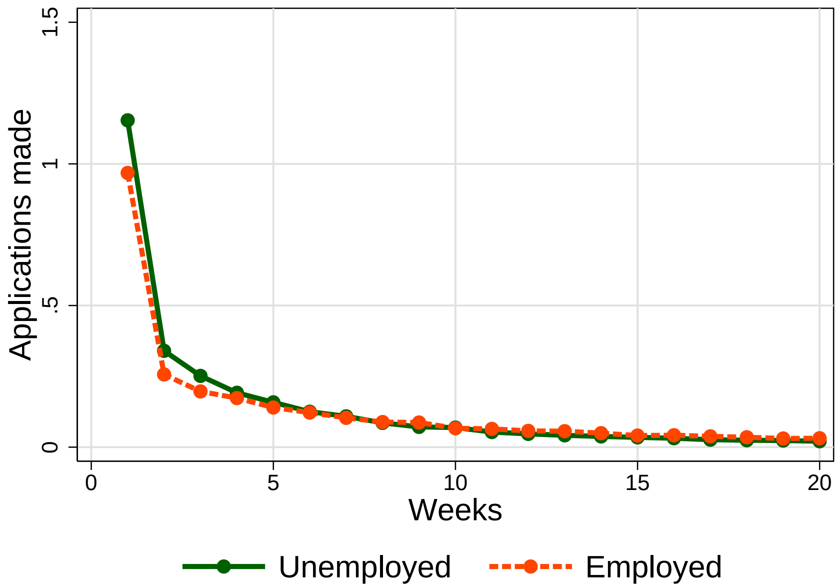


Figure 1: Average number of submitted applications by week of job search (since date of last online CV modification).

first week of search and 17.6%, use the website only one day.[11] This could be evidence that some individuals browse all their desired postings, before submitting applications in bunches.

In terms of what type of ads job seekers prefer, figure 2 shows a first overview. In the graph, we show the relationship between length of the job search spell and length of time ads have been available in the website (their average *online life*) along an adjusted polynomial of the number of weeks (as a non-linear trend). For example, if we take the unemployed sample, applications submitted by this group during the first week of search, are directed to job ads that are on average 11 days old. For employed seekers, the number is closer to 14 days. For both samples, these averages decrease significantly after the first week.

Given that job ads have an average *online life* of thirty days, it is striking that the averages in the figure are much lower than half that number. We interpret this as job seekers' distaste for "old" job ads, which can be filled in a matter of days after the ad goes online. This phenomena is referred in the literature as *phantom* vacancies, and has been analyzed by Albrecht, Decreuse, and Vroman (2015) and Chéron and Decreuse (2016).

Related to distaste for *phantom* vacancies, the drastic decline in average number of days of ads applied by individuals in figure 2 can be rationalized as follows: at the start of the job search process, individuals apply to their most preferred ads, which are randomly distributed in terms of the "days online" universe of postings. Job seekers who submit applications after that first week,

[11]Figure 8 in the appendix shows that the qualitative results are maintained when restricting the sample to individuals searching for *more* than two weeks.

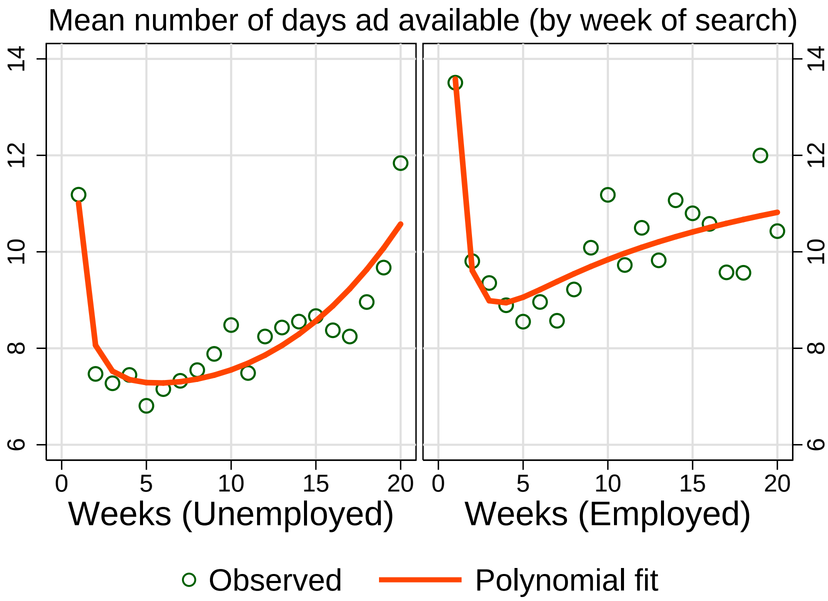


Figure 2: Average number of days an ad has been available on the website, by week of application of individuals, separated by labor force status (unemployed/employed).

choose only the newer job adverts they face (an average of eight and ten days, for unemployed and employed respectively). This could be due to job seekers not receiving callbacks from applications made during their first week of search or just making more applications from the new flow of vacancies to maximize their chances of landing a job. After the initial weeks, the ads job seekers start considering grow older (in number of days, which explains an increasing trend in the *online life* of considered ads (polynomial average).

Figure 2 shows evidence of "stock-flow" matching: the inflow of new job seekers seeks to match with the stock of job adverts (first week of search). If these seekers are unsuccessful and time passes, they become part of the stock of job searchers, who then try to match with the flow of job positions (newer posts in the website). This type of behavior has been studied by Taylor (1995), Coles and Muthoo (1998), Coles and Smith (1998) and Ebrahimy and Shimer (2010), among others.

4 Market segmentation and job seeker preferences

In this section we analyze empirically which attributes of heterogeneous jobs are more valued by heterogeneous job seekers. To do this, we need to determine which is the relevant market segment for each of the job seekers in our sample. However, our dataset only contains information on actual applications and no information is collected by the website on total number of searches nor *clicks* on job postings by individuals. Thus, we do not have sample variation in terms of job ads: we only observed those that individuals chose to apply to, but not those which are observed but then discarded by seekers.

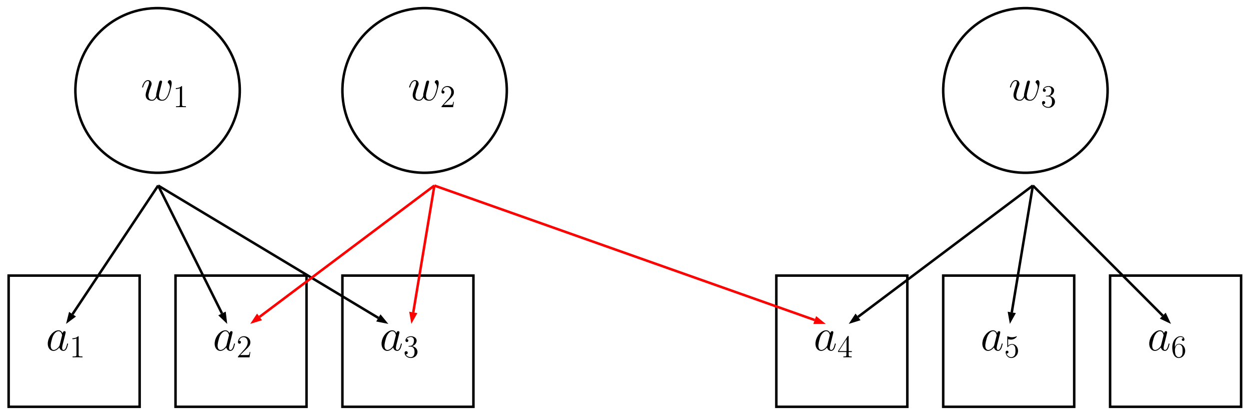


Figure 3: Example of a network formed by workers $\{w_1, w_2, w_3\}$. Worker w_1 is linked to worker w_2 by common applications to ads a_2 and a_3 but is not linked with w_3 in the network of degree 1. All workers are linked in the network of degree 2.

4.1 Market segmentation through network analysis.

One way to obtain the variation described above, would be to consider the cross between all job applicants and all job ads in our sample, what is usually referred to as the *exploded* dataset. However, there are some drawbacks from this approach: first, the *exploded* dataset makes comparisons between job seekers and job positions which may be objectively too different to consider; second, the size of the estimating sample increases exponentially with the initial number of job seekers and job ads, making the task of even simple calculations infeasible in most cases.

In what follows, we use the network formed by job seekers to determine which job postings are relevant to them. 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 job seeker 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, the network of degree 0 is 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 (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 and all job postings a, or the *exploded* dataset.

Figure 3 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 $\mathcal{A}_{w_1}^0 = \{a_1, a_2, a_3\}$ 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 (are less than 30 days old) during the time that each job seeker is active in our dataset, which we define by the time period between the last modification of his/her online profile and the last observed application submitted. Given this 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.

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).

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 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.

4.2 Preferences over heterogeneous characteristics.

On each of the datasets created from the network approach, we estimate preferences of 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

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

where y_{aw} is a dummy variable that takes the value of one if job seeker w applies to posting a, and zero otherwise. In X_{aw}, we control for observed job and worker characteristics, which do not overlap. The list of variables for the job includes firm size, firm's location (region), firm's industry, dummies for specific job requirements (computer knowledge, or some other form of specific knowledge) and controls for specific job characteristics: type of contract (full/part time), dummies for quarter in which the job ad was posted, number of vacancies needed to be filled and controls for job titles.[12] For individuals, we control for marital status (dummy variable for marriage), gender (dummy for male) region of residence and quintic polynomials for the age of the job seeker and for the amount of time (measured in days) in either the current job (for those employed) or in unemployment (for unemployed seekers). For both seekers and ads, we include a variable of whether the wage expectation (for seekers) or the wage expected to be paid (for jobs) is made explicit or not.

On the other hand, 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 continuos dimensions k_c, we include in the regression a polynomial of order $\bar{P} = 5$ to assess whether non-linearities exist 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 ($z_{k_c} < 0$) compared to when they are *under*-qualified ($z_{k_c} > 0$). 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 behaviour. Finally, we also consider interaction effects between different *misalignment* levels.

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} (in pctg.)	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 *misalignment* as well as age of the worker, firm size, contract type, dummmies for different types of requirements of the job and characteristics of the firm (see details in the main text).

5 Results

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 *misalignment* terms and their interactions are presented later. The considered dimensions are level of education, years of experience, log wages and a one-digit occupational code.

[12] We follow work by Marinescu and Wolthoff (2015) and Banfi and Villena-Roldán (2016).

In the table 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 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. The latter confirms the finding in figure 2, where we find a seemingly distaste of individuals for older job posts and shows that this effect is not driven by sample composition or other covariates.

5.1 Life-cycle and duration effects.

Given the results from the regression, here we show the existence of life-cycle profiles and duration dependence in the decisions 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.[13] For comparability, we present all figures as a fraction of the unconditional mean of an application for the respective sample.

Figure 4, we observe the life-cycle profile of job application decisions implied by the estimation results in equation (1): given regression coefficients on the polynomial for the age of the job seeker, we can predict application probabilities for different ages, when the rest of the variables in the regression remain at their 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, Janiak, and Villena-Roldán (2015) and Menzio, Telyukova, and Visschers (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.

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

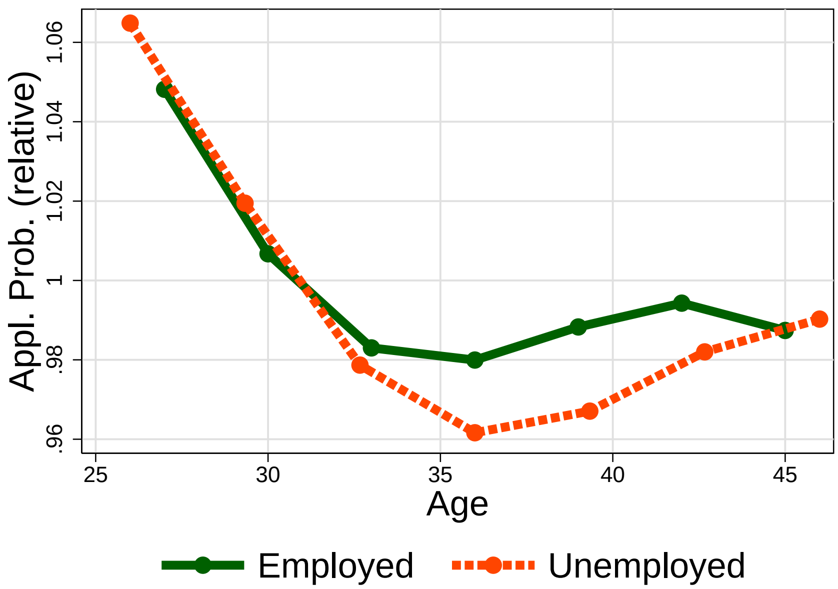
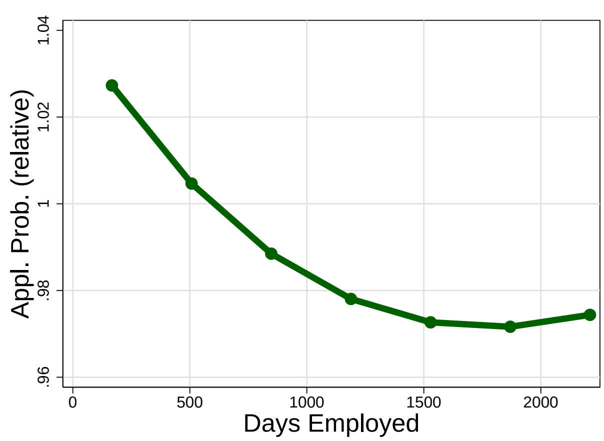
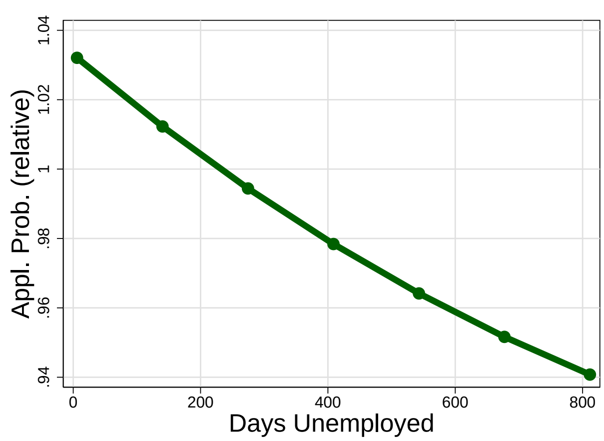


Figure 4: Predicted application probabilities for different ages, given results from eq. (1). The figure is the computed using the coefficients associated to a polynomial of order 5 on age of the applicant. Results are presented relative to unconditional application probability means and are based on a network of degree 1.



(a) Job tenure



(b) Unemployment duration

Figure 5: 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.

In figure 5, 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 decline 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.

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.

5.2 Misalignment and applications.

Below, we present graphically results of the effect of *misalignment* in continuous characteristics (education, experience and log wages) on application decisions. Figure 6 shows predicted application probabilities ($\hat{y}_{aw}$ from the estimates of equation 1), when a particular continuous dimension (z_{k_c}) varies, keeping all other observables at their sample mean (including the *misalignment* in other dimensions). To produce these figures, we compute the marginal effect of altering a particular z_{k_c} on the application outcome. Given that each *misalignment* dimension enters the equation as a fifth-order polynomial and that there are interactions between them, the computed effect is highly non-linear and depends on which value the other control variables take. For the figures, we set all control variables at their sample mean.

The considered range for z_{k_c} is bounded by its sample mean 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 figure 6, job seekers in both labor states align themselves with particular levels of requirements of job postings. This is represented by an inverted U shaped relationship between *misalignment* and application probability (all else constant). Notice that since the *misalignment* variable is defined as the value from the job ad minus the value of the worker, negative values imply that individuals are over-qualified in the specific dimension, while positive values indicate that they are under-qualified for the position. As seen from the figure, both the education and experience dimensions are centered on a negative number: the average *misalignment* in education is slightly below zero, while for experience, the number is slightly below negative five. This means that the average job seeker is slightly more educated and has five more years of experience than what is requested in the average job ad.

The alignment in both education and experience dimensions is not exact: for education, job seekers tend to maximize application probabilities at levels slightly above their own, while both

(a) Education levels

(b) Years of experience

Figure 6: 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.

employed and unemployed seekers apply with more frequency to jobs for which, on average, they have more experience than what is required by the job ad. The main reason for the average *misalignment* in this dimension is the fact our sampling strategy makes the average website user, someone attached to the labor force, with significant number of years of experience.

With respect to non-linearities, figure 6 shows that job seekers react differently to ads for which they are under or overqualified. Moreover, this behavior is different in the case of education and experience. This follows from the inverted U shape curves, 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, as opposed to those jobs for which they are over-qualified.

The figure also shows another difference between preferences of unemployed versus the employed. From the education panel, we observe that unemployed seekers are less sensitive to being under/over qualified to job positions, which is reflected in the curve for them being flatter in the panel than that for the employed. As for experience, the differences between seekers in different labor states are minimal.

In figure 7 we show the same result, but for log-wages. Given that our estimates control for all other observables across job positions and job seekers, and that the regression controls for interactions, the *misalignment* in log-wages could be interpreted as *misalignment* in job and worker types: controlling for all observables, higher paying jobs and job seekers with higher earnings expectations must be of higher skill on average, and viceversa.

The figure reveals that differences in log-wages (types) are the ones that affect application

Figure 7: Predicted application probabilities, given results from eq. (1) and different levels of misalignment in log-wages. Results are relative to unconditional application probability means and are based on a network of degree 1.

probabilities the most: 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%. As with the case of the education dimension, the figure shows that employed seekers are more sensitive to the misalignment in log-wages than the unemployed, fact which is reflected in the curvature of the application probability function.

Finally, the figure summarizes well the overall search strategy that seems to be behind those seekers in different labor market states. For the unemployed, the curve predicting applications reaches its peak very close to zero. This means that those who currently are looking from unemployment try to get matched to jobs for which they are a best fit, in the sense that the absolute difference between log-wages of ads versus that of seekers are minimal. On the other hand, job seekers who already have a job direct their search towards job ads with types slightly above theirs, which is reflected in the solid curve in figure 7 reaching a peak at a positive level of the difference in log-wages. Thus, an intuitive takeaway from the evidence is that unemployed seekers try to maximize the probability of getting hired, while individuals performing on-the-job search, are interested in climbing the job ladder.

6 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 distaste of phantom vacancy, stock-flow matching behavior, and 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.

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.

BANFI, S., AND B. VILLENA-ROLDÁN (2016): "Do High-Wage Jobs Attract more Applicants? Directed Search Evidence from the Online Labor Market," Discussion paper, Working Paper 327, Center for Applied Economics, University of Chile.

CARRILLO-TUDELA, C., AND L. VISSCHERS (2014): "Unemployment and Endogenous Reallocation over the Business Cycle," Cesifo working paper 4079.

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.

COLES, M. G., AND A. MUTHOO (1998): "Strategic Bargaining and Competitive Bidding in a Dynamic Market Equilibrium," *Review of Economic Studies*, 65(2), 235–260.

COLES, M. G., AND E. SMITH (1998): "Marketplaces and Matching," *International Economic Review*, 39(1), 239–254.
ΩŞahin, Song, Topa, and Violante

ŞAHIN, A., J. SONG, G. TOPA, AND G. L. VIOLANTE (2014): "Mismatch Unemployment," Forthcoming, American Economic Review.

EBRAHIMY, E., AND R. SHIMER (2010): "Stock-flow matching," *Journal of Economic Theory*, 145(4), 1325–1353.

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.

HERZ, B., AND T. VAN RENS (2015): "Accounting for Mismatch Unemployment," Discussion paper, University of Warwick.

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): "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): "Optimal Life Cycle Unemployment Insurance," *American Economic Review*, 105(2), 816–59.

Rogerson, R., R. Shimer, and R. Wright (2005): "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, 43(4), 959–988.

Taylor, C. R. (1995): "The Long Side of the Market and the Short End of the Stick: Bargaining Power and Price Formation in Buyers', Sellers', and Balanced Markets," *The Quarterly Journal of Economics*, 110(3), 837–855.

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

A Extra Figures

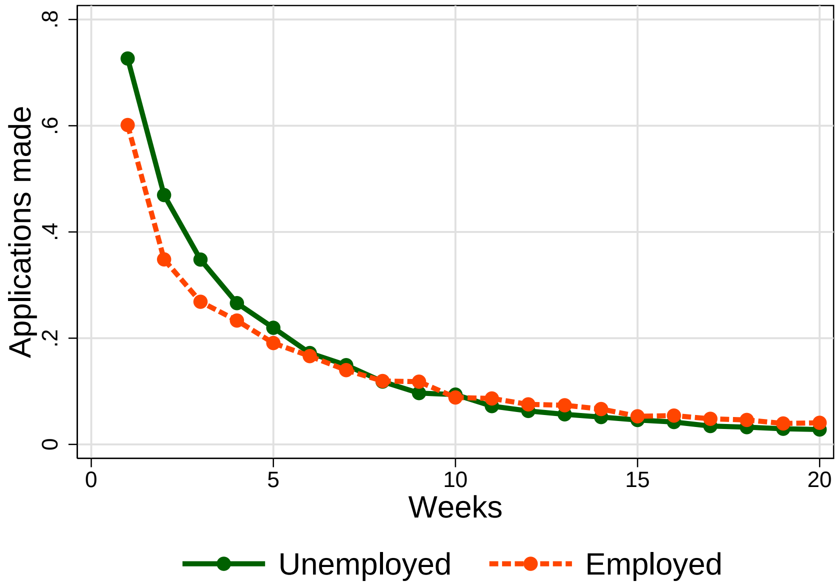


Figure 8: Average number of submitted applications by week of job search (since date of last online CV modification). Sample of individuals who submit applications two or more weeks after the start of their search.

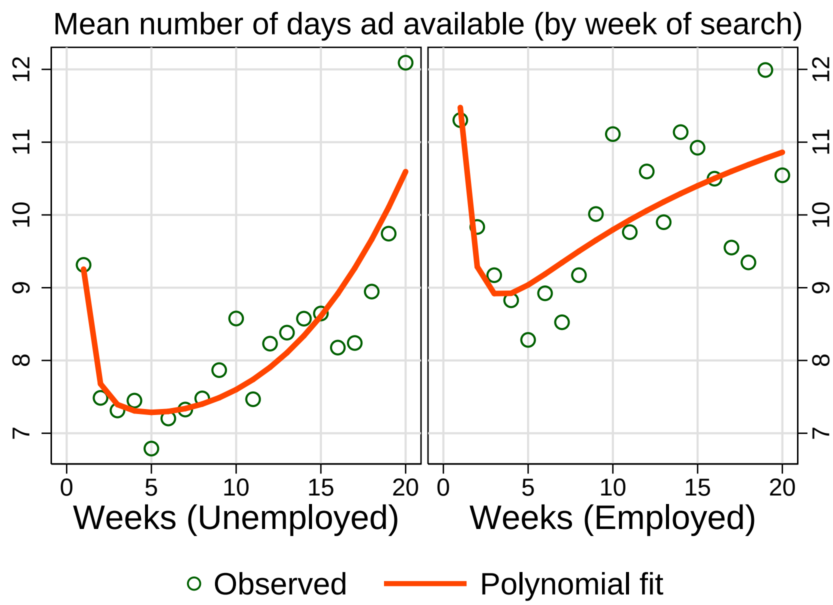


Figure 9: Average number of days an ad has been available on the website, by week of application of individuals, separated by labor force status (unemployed/employed).