Sorting On-line and On-time*

Stefano Banfi
Chilean Ministry of Energy

Sekyu Choi
University of Bristol

Benjamín Villena-Roldán
CEA, DII, University of Chile

January 22, 2019

Abstract

Using proprietary data from a Chilean online job board, we find strong, positive assortative matching at the *worker-position* level, both along observed dimensions and on unobserved characteristics (OLS Mincer residual wages). We also find that this positive assortative matching is robustly procyclical. Since we use information on job applications instead of final matches, we use the generalized deferred-acceptance algorithm to simulate tentative final allocations. Under all considered scenarios for the algorithm, positive assortative matching is preserved from the application stage to the realized matches.

Keywords: Online search, assortative matching, labor markets.

JEL Codes: E24, E32, J24, J60

*Email: sbanfi@minenergia.cl, sekyu.choi@bristol.ac.uk and bvillena@dii.uchile.cl. We thank Jan Eeckhout, Shouyong Shi, Yongsung Chang, Chao He, for comments and discussions with colleagues at Cardiff University, The University of Manchester, Diego Portales University, 2016 Midwest Macro Meetings, 2016 LACEA-LAMES, 2016 Workshop in Macroeconomic, Search & Matching at the University of Chile, 2016 Shanghai Workshop in Macroeconomics, and Alberto Hurtado University. Villena-Roldán thanks for financial support the FONDECYT project 1151479, Proyecto CONICYT PIA SOC 1402, and the Institute for Research in Market Imperfections and Public Policy, ICM IS130002, Ministerio de Economía, Fomento y Turismo de Chile. All errors are ours.

1 Introduction

The way in which heterogeneous workers match with heterogeneous jobs has crucial implications for the economy. Several papers show the intertwined relation between sorting patterns, inequality, and efficiency, positing a trade-off between the latter two. Using different approaches, Card, Heining, and Kline (2013), Bagger and Lentz (2014), Lise, Meghir, and Robin (2016) find sizable welfare or productivity improvements of reallocating workers optimally in Germany, Denmark, and the US, respectively. Allocation patterns also matter for aggregate productivity, unemployment, and transitions between different labor market states and overall occupational mobility. Şahin, Song, Topa, and Violante (2014) show that mismatch between employers and job seekers across occupation markets translates in lower job finding rates and substantially higher unemployment in the US after the Great Recession. In general, resource misallocation accounts for a sizable share of productivity differences across countries, as shown by Restuccia and Rogerson (2017).

Despite of its importance, we know little about such allocation patterns. Lack of appropriate data precludes researchers from learning about *worker-job* matches, arguably the closest simil of the marriage sorting problem of Becker (1973) applied to labor markets. Instead, the profession has had to settle for studying *worker-firm* allocations. Given the large diversity of positions in corporations, what we could learn about sorting just by looking at matched employer-employee data is limited and often uninformative about the relevant allocation patterns generated in the labor market. For instance, consider `amazon.com`, who hires many of the best computer scientists, but not necessarily the best janitors. Why? `amazon.com` critically depends on talented computer scientist, while its janitors perform the same tasks they would do in other firms.

Instead, our approach departs from the existing literature by focusing in matches between workers and job positions, *not* firms. To this end, we analyze information from `www.trabajando.com`, a job posting website with presence in most of Latin America. The website provides a comprehensive dataset on applications of job seekers to job postings in the Chilean labor market, between 2008 and 2014. We have detailed information for both sides of the market such as education, occupations and experience for individuals and for job postings (as requirements stipulated by firms). A key advantage of the dataset, is that we observe both expected 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.[1].

The richness of the data allows us to circumvent the problem of making inferences on strength and sign of assortative matching based on realized wages from matched employer-employee data as in Abowd, Kramarz, and Margolis (1999), Hagedorn, Law, and Manovskii (2017) and Card, Cardoso, Heining, and Kline (2018), among others. In our data, we observe all positions which the worker applies to, which is akin to observing her *acceptance set*, in the language of Shimer and Smith (2000). For both workers and firms, we can use the explicit information they provide

[1]The wage information is required for all users of the website, although users can choose whether to make this information public or not to the other side of the market

with respect to expected and offered wages, which we use to create a direct measure of types to construct relative rankings.

In Section 3, we find strong, positive assortative matching (PAM) between job seekers and job postings, both along observed dimensions and along unobserved types, which we proxy using residual wages from linear regressions. We improve the measurement of productivity in the literature, largely based on firm and worker fixed effects correlations for three reasons. First, we obtain a measure of productivity at the job instead of at the firm level. Second, we observe a much richer set of traits for workers (major, self-reported experience, etc) and for jobs (required education, major, experience, on top of firm size, industry, etc) which allow us to estimate residuals much more precisely as the unexplained part of expected and offered wages. Third, since our wage and type measures reflect information before an actual match occurs, they are not subject to ex post compensations, which would make identification of sorting almost impossible, as pointed out for example in Eeckhout and Kircher (2011), Hagedorn, Law, and Manovskii (2017) and Lopes de Melo (2018).

We also analyze how assortative matching fluctuates with business cycle conditions in Section 4. Keeping constant observable characteristics of both workers and firms with the DiNardo, Fortin, and Lemieux (1996) method, we find that both differences in observed characteristics and residual wage *mismatch* is counter-cyclical: assortative matching becomes more positive when aggregate conditions are good. The results are quite robust and are found when we both use aggregate unemployment series or monthly GDP indexes as indicators of business cycle conditions. Our findings should help the estimation and identification of models of sorting over the business cycle, as in Lise and Robin (2017).

One limitation of our analysis, is that we do not observe the resulting patterns of worker-job sorting nor wages paid and we cannot link pre-match levels of sorting with realized ones. Hence, we attempt to provide some bounds on the level of assortative matching for realized worker-ad matches by simulating outcomes using the generalized deferred-acceptance algorithm (henceforth GDA) in Gale and Shapley (1962) in Section 5. As in the Hitsch, Hortaçsu, and Ariely (2010) online dating matching application, we rely on the two-sided search model of Adachi (2003) in which the decentralized market achieves the GDA allocation as search costs vanish. While the online job board does not fit exactly these conditions, the GDA allocation is a useful Pareto-efficient benchmark for the side of the market who makes offers in the deferred acceptance procedure.[2]

Results from our simulation exercise show that, even though we consider a wide array of different scenarios and assumptions, we find no cases of negative assortativeness and most of the simulations show significant levels of PAM. Although the underlying preferences we assume for workers and job positions can either amplify or attenuate the ex ante sorting pattern observed in the application stage, the fact that we find mostly PAM as the result of our simulations, indicates that already at

[2] See Roth and Sotomayor (1990) and Hitsch, Hortaçsu, and Ariely (2010).

the

Our approach and findings are complementary to the literature searching for identification and quantification of assortative matching in labor markets using realized wages. Conclusions there are mixed. Using matched employer-employee panel data from France, and a two-way fixed effects approach, Abowd, Kramarz, and Margolis (1999) find weak positive sorting between workers and firms, in terms of unobserved productivity levels. Andrews, Gill, Schank, and Upward (2008) find a negative bias for the correlation between worker and firm fixed effects, especially if job movers are scarce. Card, Cardoso, Heining, and Kline (2018) synthesize this literature and interpret that firm-specific effects on wages not only reveal productivity but also idiosyncratic preferences over employers. Eeckhout and Kircher (2011) argue that "using wage data alone, it is virtually impossible to identify whether assortative matching is positive or negative". On the other hand, Hagedorn, Law, and Manovskii (2017) show that assortative matching is indeed recoverable using information on co-workers, wages and labor market transitions. They apply their framework to matched employer-employee data from Germany and find signs of strong PAM. Lopes de Melo (2018) shows that the correlation between worker and firm effects understates the true sorting since over-qualified workers need to compensate their firms and vice versa, an argument stressed in Eeckhout and Kircher (2011). Instead, analyzing co-worker sorting is a superior way to study PAM. Yet another approach comes from Bartolucci, Devicienti, and Monzón (forthcoming), who use Italian data to identify sorting patterns using firm profit information.

This paper is also related to the growing literature using online job-posting 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 (forthcoming) 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 an application. Lewis (2011) and Banfi and Villena-Roldan (forthcoming) show internet seekers significantly react to posted information for car and labor markets, respectively. 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.[3] Our data covers daily job postings and job seeker activity in the Chilean labor market, between September 1st, 2008 and June 4th, 2014. We observe entire histories of applications (dates and identification numbers of jobs applied) from job seekers and dates of ad postings (and re-postings) for firms. A novel feature of the dataset, is that the website asks job seekers to record their expected salary, which they can then choose to show or hide from prospective employers. Recruiters are also asked to record the expected wage for the advertised job, and are also given the same choice of whether to make this information visible or not to applicants. The wage information in the website is *mandatory* for all users, in that they cannot apply/post a vacancy if they have not written something in the relevant field. Admittedly, our analysis relies on self-reported wage information that applicants or employers may choose to keep private, which may rise doubts about its accuracy. However, Banfi and Villena-Roldan (forthcoming) find that the informational content of wages that are kept private (implicit wages) is high, given that they can be predicted quite accurately using observable characteristics and an estimated model with a sample of explicit wages only.

For each posting, besides offered wage (which can or cannot be visible by applicants) we observe its required level of experience (in years), required education (required college major, if applicable), indicators of 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 (brackets with number of employees) and an industry code.

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.[4] We have codes for occupational area of the current/last job of the individual as well as information on the monthly salary of that job and both its starting and ending dates.

As with many self-reported data sources, there is some amount of measurement issues in our analysis. Individuals may lie in their CV’s in order to being more appealing in the selection of firms, but this may be a dangerous strategy for job seekers. One objective problem we face in our analysis, is that the worker information given to us is a snapshot of current online CV’s by individuals on June 4th, 2014. For job seekers who use the site to search only once, this creates no issues. However, if an individual uses the website to apply in two different points in time, with several months (or even years) between them, this creates a measurement issue when we correlate information of job ads (measured without error) and information of candidates if applying to jobs

[3]The list of countries as of January of 2016 is: Argentina, Brazil, Colombia, Chile, Mexico, Peru, Portugal, Puerto Rico, Spain, Uruguay and Venezuela.

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

prior to the update of his/her online profile. However, this measurement issue probably decreases the level of assortative matching we estimate, since it most likely makes job requirements and job seeker characteristics more *dissimilar* than what they actually are.

For the remainder of the paper, we restrict our sample to consider 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.[5] This amounts to approximately 9,745 USD per month,[6] which represents more than double the 90th percentile of the wage distribution, according to the 2013 CASEN survey.[7] We also discard individuals who desire net wages below 159 thousand pesos (around 310 USD) a month, the monthly minimum wage in Chile during 2008. Consequently, we also restrict job postings to those offering monthly salaries in those bounds. Additionally, we restrict our sample to active individuals and job postings: we consider those workers who make at least one application and job postings which receive at least one application respectively, during the span of our dataset.

Table 1 shows some descriptive statistics for the job searchers in our sample. From the table we observe that the sample is a young one (average age is 32.9), comprised of mostly single males, with 47% being unemployed ($117,951$ unemployed from a total of $250,796$ job searchers). Given the age group we consider, most individuals in the sample have some working experience, with mean number of years of experience hovering around 8. Job seekers in our sample are more educated than the average in Chile, with 43.6% of them claiming a college degree, compared to 25% for the rest of the country in a similar age group (30 to 44), (The figure is from the 2013 CASEN survey.)

We can also observe that most job seekers have studies related to management (around 20.4%) and technology (around 28.3%), but a significant fraction (around 29.9%) does not declare any occupation. In terms of salaries, average expected wages are (in thousand of CLP) $1,129$ and 601 for employed and unemployed seekers, respectively. For comparison, the average minimum monthly salary in Chile was around CLP 187 thousand for the time span considered in our sample.

For our sample, the average number of submitted applications is 12.4 per worker, with employed seekers (on-the-job searchers) applying to 14.9 job ads, versus 10.1 applications for the unemployed group. Additionaly, we can observe that the vast majority of applications is to job ads that do not show an explicit wage, with the average job searcher sending around 0.97% of her applications to ads that show explicitly how much they expect to pay.

Table 2 shows sample statistics for job postings. We separate our sample between postings with

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

[6]Using the average nominal USD-Chilean peso exchange rate between the first quarter of 2008 and the second quarter of 2014.

[7]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
Age: Mean/(S.D.)	32.69	32.99	32.85
	(6.91)	(7.55)	(7.26)
Males (%)	61.58	50.50	55.71
Married (%)	34.18	29.25	31.56
Years of Experience: Mean/(S.D.)	8.33	7.64	7.97
	(6.15)	(6.69)	(6.45)
Education level (%)			
Primary (1-8 years)	0.25	0.46	0.36
High School	18.14	37.55	28.42
Tech. (tertiary) educ.	25.79	27.90	26.91
College	54.85	33.61	43.60
Graduate	0.97	0.47	0.71
Occupation (%)			
Management	23.10	17.93	20.36
Technology	35.23	22.19	28.32
Non-declared	17.82	40.81	29.99
Rest	23.86	19.07	21.32
Wages in thousands CLP			
Expected wages: Mean/(S.D.)	1129.10	600.85	849.29
	(765.69)	(469.88)	(679.85)
Number of applications per worker			
Applications: Mean/(S.D.)	14.91	10.13	12.38
	(21.86)	(16.91)	(19.54)
App. to explicit wages ads: Mean/(S.D.)	0.97	1.20	1.09
	(2.00)	(2.42)	(2.24)
Observations	117,951	132,845	250,796

implicit wages (do not post information on salaries) and with explicit wages (ads show salary to be paid). From a total of $139,921$ active job postings in our sample period, only $18,096$ (12.9%) are classified as having an explicit wage.

Implicit wage postings are characterized for requiring higher levels of experience, higher levels of education and being associated with higher salaries. They also tend to concentrate more on technology related occupations: 30.7% of ads with implicit wages are related to technology sectors, versus 15.4% of job postings with explicit wages. Job postings in our sample receive a mean of 22.2 applications, with a significant difference in the number received by implicit wage postings (23.2) versus those received by explicit ones (15.1).

3 Sorting on-line

In this section, we analyze how similar are the characteristics of job seekers (w, for workers) versus the requirements of the job postings (a, for ads) they consider. We compute correlations between characteristics of job postings against the characteristics of job seekers. We do so for observable characteristics (education, experience, and wages), and for a measure of worker and ad *types*, computed using residuals from linear wage regressions which are conceptually comparable to worker and job fixed effects in Abowd, Kramarz, and Margolis (1999).

Table 2: Characteristics of Job Postings

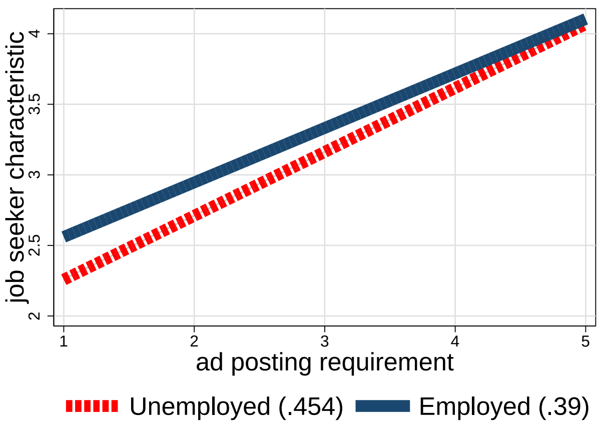
	Implicit wage	Explicit wage	Total
Required experience in years: Mean/(S.D.)	2.20	1.55	2.12
	(1.82)	(1.39)	(1.78)
Required education level (%)			
Primary (1-8 years)	0.97	1.99	1.10
High School	32.31	53.83	35.10
Tech. (tertiary) educ.	28.30	26.62	28.09
College	37.79	17.36	35.15
Graduate	0.63	0.19	0.57
Occupation (%)			
Management	24.42	25.33	24.54
Technology	30.65	15.47	28.69
Non-declared	36.74	52.46	38.77
Rest	8.19	6.73	8.00
Wages in thousands CLP			
Offered wages: Mean/(S.D.)	734.64	436.66	696.11
	(603.73)	(342.89)	(585.28)
Applications received: Mean/(S.D.)	23.28	15.14	22.23
	(37.49)	(26.90)	(36.40)
Observations	121,825	18,096	139,921

3.1 Sorting on observables

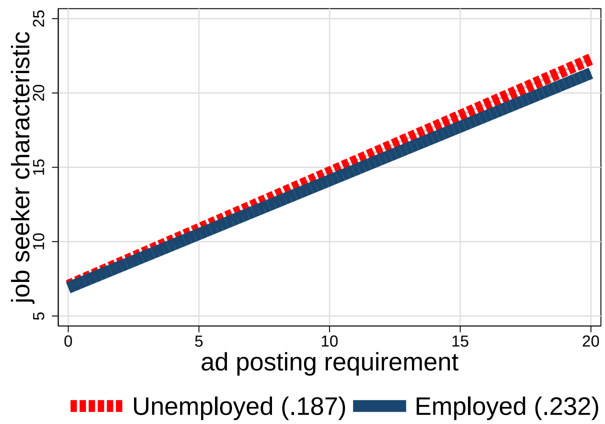
In figure 1 we present correlations between observed characteristics requested by job ads, versus the characteristics of individuals applying to those jobs. In the figure we plot the implied linear relationship between observable characteristics for all applications/matches in our sample. We do so for education (ordered categorical variable), experience (in number of years) and monthly salary (in logs), respectively. Additionally, in each of the three panels, we perform the exercise separately for unemployed and employed job seekers and show the sample correlation in parenthesis.

As seen from the figure, there is PAM along all three dimensions, and this fact is remarkably similar for both employed and unemployed individuals. The strength of PAM is highest for log wages, followed by education and then experience levels. The figure also shows that job seekers in our sample seem to be, on average, *overqualified* for the positions they apply to (this is the case for both the employed and unemployed) given that for each observable characteristic, the predicted level of job seeker characteristic is on average higher than the posting requirement.

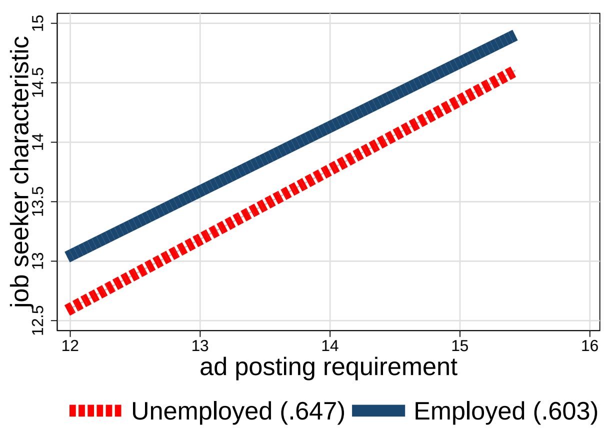
One could argue that it is not a surprise that workers sort coherently across vacancies in terms of these objective and observable characteristics. With the exception of salaries (which are often not included as information in job postings, as seen in table 2), requirements on education and number of years of experience are almost always available in the information for postings, making it easy for seekers to direct applications. However, the existence of *phantom* vacancies and the fact that workers can use the website free of charge, an thus potentially play random strategies in their applications, does not immediately lead us to determine that sorting would be an equilibrium result. Furthermore, as discussed earlier, we measure worker characteristics with some degree of



(a) Education



(b) Experience



(c) Log wages

Figure 1: Linear prediction of job seeker characteristic given ad posting requirements, for different observable characteristics and employment status of individuals. Simple correlations in parenthesis. All applications between 1-Sep-2008 and 4-Jun-2014.

error, which would put into question whether one should expect assortative matching ex ante.

3.2 Sorting on types

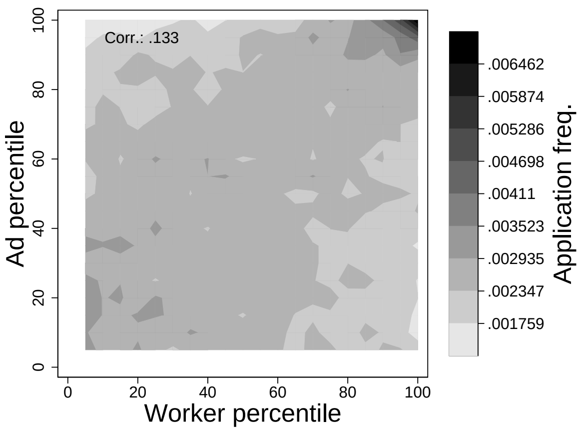
To study sorting beyond observable characteristics, we rank workers and job postings by some notion of *type*: a ranking which might be comparable across individuals and positions. For each individual, we run linear regressions between their stipulated desired wages and observable characteristics. We do the same for wages expected to be paid by job posters and their observables. By doing so, we obtain measures that are comparable to worker and firm fixed effects in Abowd, Kramarz, and Margolis (1999). For job seekers, the regression includes a polynomial of order 5 for age, a categorical variable for gender, educational attainment (levels), dummies for area of study (in case of tertiary education), dummies for marital status, professional experience (in years), dummies for region of residence and number of days in their current labor force status (duration of unemployment vs. tenure at current job); for job posters, we control for the number of vacancies related to the posting, required educational attainment, required area of study, industry (categorical variable) of the firm posting the vacancy, type of contract offered (full time/part time), requested amount of experience and region of the vacancy. From the linear regressions we obtain residuals (unexplained variation in wages), which we define as the worker/ad type, and thus are subject to be ranked.

In Figures 2 and 3, we show the frequency of applications given percentiles of worker and ad types. One peculiarity of the data set we are using, is the fact that both job seekers and job posters can decide to either show or hide their desired and posted salaries. When agents decide to show wage information, we denote the case as an *explicit* case; similarly, if an agent chooses not to divulge wage information, we denote it as an *implicit* case, since other pieces of information can implicitly but not perfectly convey this information. We make this distinction, since both the existence and strength of PAM in wages could be due to a composition effect: job seekers might be directed by explicit wages in job postings and/or the fraction of seekers in a particular group (employed versus unemployed) might be more or less subject to sorting by explicit wages, leading to a spurious correlation in the figures.

From the figures we observe that even after controlling for observables, PAM between workers and job postings remains but its strength lessens when compared to the case using observable characteristics. The figure for the case when job postings do not show wages explicitly contains the least well defined joint density figures. This could be due to sample size: from table 2, these postings amount to 12.9% of all considered postings. A different explanation could be that job postings with explicit wages represent low requirement, low salary jobs on average[8] which may be "easier to get" and thus, a safe bet/fallback option for individuals across the worker type distribution.

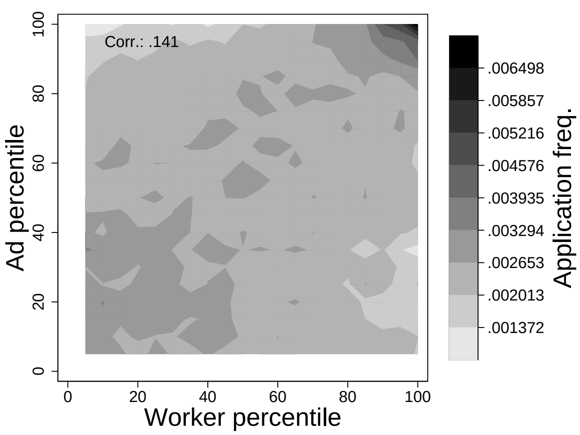
For both employed and unemployed workers, the interesting case arises when vacancies do not show explicitly the offered wage. In these cases, represented by both panels (b) and (d) of figures

[8] See Banfi and Villena-Roldan (forthcoming).

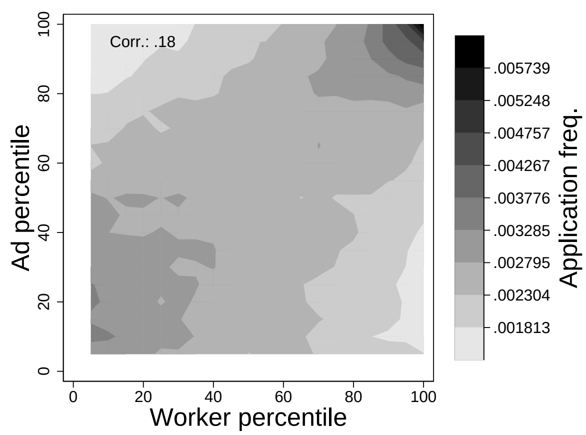


(a) Explicit applicant to Explicit ad

(b) Explicit applicant to Implicit ad

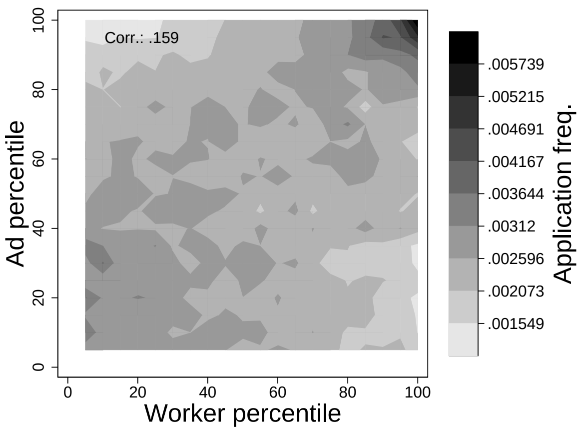


(c) Implicit applicant to Explicit ad



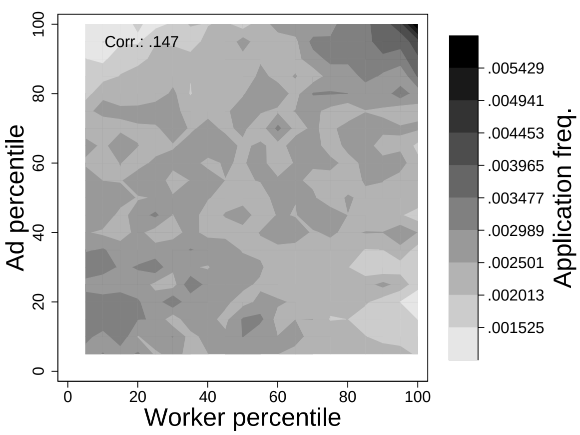
(d) Implicit applicant to Implicit ad

Figure 2: Frequencies of applications, by worker and ad percentiles of wage residuals (types) for UNEMPLOYED individuals. Residuals are obtained using a linear regression on wages controlling for observable characteristics. All applications between 1-Sep-2008 and 4-Jun-2014.

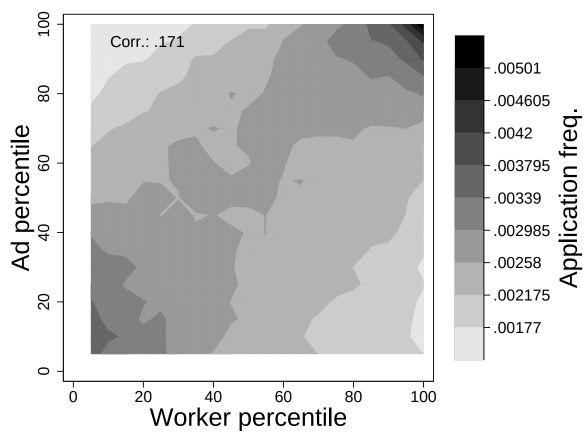


(a) Explicit applicant to Explicit ad

(b) Explicit applicant to Implicit ad



(c) Implicit applicant to Explicit ad



(d) Implicit applicant to Implicit ad

Figure 3: Frequencies of applications, by worker and ad percentiles of wage residuals (types) for EMPLOYED individuals. Residuals are obtained using a linear regression on wages controlling for observable characteristics. All applications between 1-Sep-2008 and 4-Jun-2014.

2 and 3 respectively, we can observe a clear area around the diagonal where most applications concentrate, which represents evidence of assortative matching. Also, it can be seen from all the figures that some concentration takes place in both the southwest and northeast corners of each figure, which reflects another layer of sorting.

It is also interesting to observe that the overall dispersion in applicant's desired wages is marginally higher for the unemployed, as seen by the wider application regions in figure 2 as opposed to figure 3, which may be a sign that on the job searchers have narrower fields of search.

In appendix A, we graph the same relationship, but between a job seeker and a *firm* for the sake of providing correlations that are more directly comparable to those reported in the sorting literature with matched employer-employee data, as in Abowd, Kramarz, and Margolis (1999), Card, Heining, and Kline (2013) and Hagedorn, Law, and Manovskii (2017), among many others. A firm is defined as the average of all job postings by the same firm. Thus, the procedure to obtain firm types is modified by obtaining a desired posted wage, as the average of all posted wages by all ads from the same firm. This average wage is then regressed using firm level observables only, and the resulting residual is treated in the same way as in the case with ads. As observed in the figures in the appendix, the qualitative results remain, even aggregating at the firm level, although the correlation decreases, as expected.

3.3 Different measures for firm type

In this section we explore sorting when we alter the definition of job "types". We propose four alternative measures which do not use the own expected wage to be paid at the position as a robustness check. The measures are as follows:

F_1: Average of the residual wage expectation of individuals applying to the ad.

F_2: Average of the residual last wage of individuals applying to the ad.

F_3: Average of the residual wages offered by job ads applied by individuals applying to the ad. That is, what other wage offers (cleaned from their observable characteristics) potential candidates are considering from competing job positions.

F_4: Fraction of job applicants to the ad who are currently employed

Measures F_1 to F_3 use related information on wage residuals, of co-applicants or co-applied ads in order to obtain a measure of type which does not depend on the wage posting stated by the firm posting the job. Measure F_4 could also be called a "poaching index", in the spirit of Postel-Vinay and Robin (2002): higher type firms attract more workers, especially from other firms.

Table 3 shows the correlations between our measure of worker type (own expectation wage residual) and the four alternative measures for the firm. In the table we separate by employment status and by submarket: whether the applicant/ad are explicit or implicit about their wages. For all cases, sorting patterns are positive and higher than our baseline measure (residual wages for both applicants and ads), thus our current results are the most conservative ones when thinking of positive assortative matching.

Table 3: Sorting, different firm type definitions

		Explicit applicant	Implicit applicant
Unemployed	explicit ad	0.46	0.47
		0.38	0.38
		0.27	0.27
		0.20	0.24
	implicit ad	0.50	0.50
		0.45	0.45
		0.35	0.35
		0.30	0.30
Employed	explicit ad	0.46	0.47
		0.38	0.38
		0.27	0.27
		0.20	0.24
	implicit ad	0.50	0.50
		0.45	0.45
		0.35	0.35
		0.30	0.30

Notes: The table shows correlations between worker type (wage expectation residuals) and measures F_1 to F_4 of firm's type (see main text).

4 Sorting on-time

In this section, we study how the correlation between characteristics of workers and jobs vary with aggregate business cycle conditions. To assess this, we run the following standard specification using all applications (pairs of workers w and job ads a) at time t:

$$y_{w,t} = \alpha + \rho\, y_{a,t} + \delta\, y_{a,t} z_t + \nu z_t + \sum_{\tau} \lambda_\tau \mathbb{I}[t = \tau] + \epsilon_{a,w,t} \tag{1}$$

where $y_{w,t}$ is the statistic of interest of the worker at time t, $y_{a,t}$ is the statistic of the job posting and z_t is a variable capturing aggregate economic conditions at quarterly frequency (time t). These variables are standardized, so their mean is zero and their standard deviation is one. The specification also includes quarter dummies to introduce secular trends and seasonality in a flexible way.

In this specification, the estimate for ρ is the average correlation between $y_{w,t}$ and $y_{a,t}$ when the cyclical variable is at its sample mean value $\bar{z}$, and matches the notion of sorting in the previous section. The coefficient δ, in turn, measures how assortative matching is affected when the cyclical variable z_t increases in one sample standard deviation. In what follows, we use the average unemployment rate for the Chilean economy as an aggregate indicator, but our results are robust to the use of other measures, such as aggregate economic activity indicators, results which are left in the appendix.

To properly interpret these regressions as evidence of sorting, the composition of job ads and workers should be unchanged over the business cycle. Otherwise, estimated changes in correlations during the cycle may be generated by cyclical composition changes of postings and job seekers. We address this possibility by controlling for compositional changes in our sample using the reweighing technique of DiNardo, Fortin, and Lemieux (1996) (DFL henceforth). We implement the method by first choosing the composition of jobs and workers in 2011Q3, the quarter with an unemployment rate closest to the sample average (6.83%). We run a probit model estimating the probability of being at 2011Q3 as a function of observables on the applicant side X_w (gender, age, marital status,etc) and on the job ad side X_a (firm industry, firm size, region, etc). We compute a predicted probability and define a weight for a worker w and a job a in time t as

$$\omega_{awt} = \frac{1 - \Phi(\widehat{\pi}_w X_w + \widehat{\pi}_a X_a)}{\Phi(\widehat{\pi}_w X_w + \widehat{\pi}_a X_a)},$$

where $\Phi(\cdot)$ stands for the cumulative density of a standard normal distribution.

Table 4 shows the estimates for ρ and δ, when we consider log-wages, types (Mincer residuals), years of education, and reported experience as sorting dimensions. In the table, we also report results for individuals who were searching while unemployed and employed (on-the-job search). The estimates show that for all measures and sub-samples, assortative matching is positive and significantly different from zero. As expected, sorting is stronger for log-wages compared to types, which is natural, since worker and job types are the remanent part of a productivity measure once observables are controlled for. Observables characteristics such as years of education and experience show a clear PAM pattern, too, being education the more accentuated.

In terms of the interaction between sorting and business cycle conditions, on the other hand, the reported estimates for δ are negative and significant in all cases. This shows that sorting is *procyclical*: the correlation between characteristics of workers and jobs increases by δ as the unemployment rate decreases by one standard deviation (business cycle conditions improve). Education is the variable with the most sensitivity of sorting with respect to aggregate conditions, while worker-position types show the least sensitivity.

Focusing on labor market status in the last two panels of table 4, the overall picture does not change. In the case of log wages, after splitting the sample between labor status, the correlation

Table 4: Assortative matching (ρ) and cyclical correlation (δ) between job seekers and postings, constant composition sample for log wages and types

All applicants					
	ρ	δ		ρ	δ
Log Wages	0.660***	-0.016***	Education	0.367***	-0.062***
	(0.16)	(0.14)		(0.24)	(0.20)
Types	0.152***	-0.007***	Experience	0.210***	-0.011***
	(0.23)	(0.18)		(0.21)	(0.16)
Unemployed					
	ρ	δ		ρ	δ
Log Wages	0.590***	-0.027***	Education	0.362***	-0.060***
	(0.27)	(0.25)		(0.35)	(0.33)
Types	0.143***	-0.014***	Experience	0.216***	-0.016***
	(0.40)	(0.35)		(0.38)	(0.33)
Employed					
	ρ	δ		ρ	δ
Log Wages	0.585***	-0.022***	Education	0.322***	-0.061***
	(0.21)	(0.16)		(0.36)	(0.25)
Types	0.159***	-0.011***	Experience	0.210***	-0.007***
	(0.28)	(0.20)		(0.24)	(0.18)

Notes: 100X Standard error in parenthesis. We report mean correlation ρ and cyclical sensitivity δ as defined by equation (1). Types refers to log-wage residuals, as explained in the main text. The cyclical measure is the Chilean non-seasonally adjusted unemployment rate according to the OECD database. Regressions use DFL weights (see main text), which are computed using a probit model in which the dependent variable is an indicator for 2011Q3, and independent variables are for applicants: age, age squared, gender, gender interacted with age terms, and a full array of indicators of nationality, marital status, region, educational mayor. Independent variables for job ads are indicators for region, industry, economic activity (job board classification), educational area required, and firm size category.

decreases a little and the procyclicality of PAM substantially increases. For types, the average sorting pattern is somewhat higher for the employed, while procyclicality is much larger within each group when splitting the sample. In contrast, observable characteristics (education and experience) show stable PAM and procyclicality across different labor force status.

In Table 5 we decompose the sorting in wages and types according to the choice of whether to display wage information or not. As discussed above, workers must record their wage expectations while employers are asked about the expected offered salary. Both sides of the market are then given the choice of whether to make this information public or not. If either job seekers or job ads choose to make wage information public, we label them as *explicit*; if they keep that information private, we label the case as *implicit* since job seekers act as if they are accurately guessing them Banfi and Villena-Roldan (forthcoming).

The table shows that the main conclusions remain: assortative matching, measured by the correlation coefficient ρ is positive and significantly different from zero, both for log-wages and worker/firm types (being lower for the latter). On the other hand, the estimates of δ are negative for all the cases: whether we consider employed/unemployed applicants, and whether we consider implicit or explicit markets.

To show the robustness of our findings, in appendix B we show the analogue of tables 4 to 5, for estimations where no re-weighting of observations takes place, and hence variation comes from cyclical behavior as well as compositional changes of applicants and ads. In tables A1 and A2 we can observe that the overall results remain mostly invariant, implying a very modest compositional effect.

In tables of Section C of the appendix, we perform the same analysis using an alternative measure of aggregate business cycle conditions: the monthly economic activity index (IMACEC)[9] which is a monthly proxy for GDP in the Chilean economy. The tables in this section show DFL weighted and unweighted estimates.

Since the IMACEC activity index is negatively correlated with aggregate unemployment, estimates of δ in equation (1) need to be re-interpreted. As seen in tables A3 and A4, the sign of the estimates for δ are positive. Thus, the procyclicality of sorting is maintained, even with a completely different aggregate indicator variable.

5 Applications vs. Outcomes: a simulation approach

Our analysis up to this point is focused on the application stage and the correlations we present are between job positions and *potential* matches with job seekers. While providing with a new perspective on the issue of labor market sorting, this makes the results not directly comparable to the rest of the literature. In lieu of not observing the actual job relationships that are formed outside the website, in this section we apply the generalized deferred-acceptance (GDA henceforth)

[9] *Indice mensual de actividad económica.*

Table 5: Average and cyclical sorting, constant composition sample, given different sub-markets

			Explicit applicant		Implicit applicant	
			ρ	δ	ρ	δ
All applicants	explicit ad	Log Wages	0.645***	-0.019***	0.834***	-0.040***
			(0.24)	(0.22)	(0.27)	(0.22)
		Types	0.159***	-0.016***	0.508***	-0.017***
			(0.36)	(0.30)	(0.52)	(0.38)
	implicit ad	Log Wages	0.702***	-0.004**	0.873***	-0.031***
			(0.67)	(0.79)	(0.72)	(0.82)
		Types	0.156***	0.000**	0.630***	-0.047***
			(0.94)	(0.94)	(1.30)	(1.11)
Unemployed	explicit ad	Log Wages	0.595***	-0.020***	0.872***	-0.030***
			(0.42)	(0.41)	(0.54)	(0.47)
		Types	0.164***	-0.021***	0.530***	-0.013*
			(0.64)	(0.58)	(0.87)	(0.76)
	implicit ad	Log Wages	0.634***	-0.021*	0.884***	-0.023**
			(1.05)	(1.23)	(0.95)	(1.16)
		Types	0.146***	-0.005*	0.627***	-0.052***
			(1.50)	(1.60)	(1.70)	(1.59)
Employed	explicit ad	Log Wages	0.565***	-0.027***	0.733***	-0.045***
			(0.29)	(0.25)	(0.33)	(0.25)
		Types	0.159***	-0.020***	0.491***	-0.015***
			(0.44)	(0.33)	(0.58)	(0.33)
	implicit ad	Log Wages	0.623***	-0.029***	0.769***	-0.055***
			(0.91)	(0.96)	(1.07)	(0.99)
		Types	0.178***	-0.020*	0.630***	-0.051***
			(1.30)	(1.18)	(2.01)	(1.37)

Notes: 100X standard errors in parentheses. Types refers to log-wage residuals, as explained in the main text. The cyclical measure is the Chilean non-seasonally adjusted unemployment rate according to the OECD database. Regression are weighted using DFL weights (see main text), which are computed using a probit model in which the dependent variable is an indicator for 2011Q3, and independent variables are for applicants: age, age squared, gender, gender interacted with age terms, and a full array of indicators of nationality, marital status, region, educational mayor. Independent variables for job ads are indicators for region, industry, economic activity (job board classification), educational area required, and firm size category.

algorithm in Gale and Shapley (1962) to simulate outcomes in terms of which job seekers end up at which job positions.

It is worth noting that the GDA algorithm produces an arbitrary final assignment of worker to jobs. Moreover, our exercise simulates the level of assortative matching of the *flow* of new hirings (for both unemployment and job-to-job movers). This may differ from the sorting of the *stock* of employed workers (the object considered in most of the literature) depending on the selectivity of layoffs, something we do not observe in our data. On the other hand, it is very hard to identify the forces behind the simulated ex post sorting: it is influenced by the entire observed network formed by applicants and ads, and the interaction of the assummed preferences and the specifics of the GDA algorithm.

Even with these caveats, our exercise provides a useful benchmark given the existence of multiple matches (applications) linking the two sides of the market (multiple job positions may receive multiple job applications from the same group of job seekers) and because the algorithm has several desirable features: it is a good description of online markets and provides an outcome which under some mild assumptions is efficient.[10]

The algorithm to solve the generalized stable marriage problem, also known as the hospitals/residents problem, suits our setup well: job adverts can have more than one position to fill (as hospitals can accept more than one resident) while job seekers (residents) can send multiple applications to different job ads (hospitals). Also, the number of job seekers is not the same as the number of job positions, while matching is one-to-one in the standard marriage problem.

To make the algorithm operational in our setting, we must assume preferences by both workers and firms. Given the absence of any firm decisions in our sample with respect to worker selection, we are unable to realistically infer much about their preferences. Thus, in what follows, we make assumptions on these unobserved preferences in order to make the algorithm feasible and scalable to the most number of possible scenarios.

First, we assume that workers and firms have preferences over types only (we ignore log-wage, education and experience dimensions) and that types for both sides of the market are observable. Second, we assume that preference rankings for both firms and workers can be one of the following cases: i) monotonically increasing with respect to partner's types, ii) monotonically decreasing with respect to partner's types, or iii) decreasing in absolute type distance (*mismatch*). Given these preferences, there are nine different scenarios under which we can run the GDA algorithm. Note that in this exercise we are not interested in the rationalization of the preference rankings we assume, but we think of these as the sufficient number of cases that allow us to map a significant range of different outcomes in terms of ex post assortative matching.

We apply the GDA algorithm to the observed pattern of applications of workers to firms, which

[10] See the discussion in Hitsch, Hortaçsu, and Ariely (2010) who use the algorithm in a similar way to predict who ends up with whom in an online dating website.

Table 6: Outcome of GDA-algorithm

Worker	Ad	% **w** matched	% **a** matched	ρ	δ
H	H	0.456	0.396	0.266	-0.021^{*}
H	L	0.534	0.424	0.092^{***}	-0.027^{***}
H	M	0.593	0.46	0.539^{***}	-0.015
L	H	0.435	0.379	0.143^{***}	-0.034^{**}
L	L	0.516	0.409	0.124^{***}	-0.004
L	M	0.565	0.44	0.662^{***}	0.015
M	H	0.462	0.394	0.456^{***}	0.026^{*}
M	L	0.549	0.422	0.401^{***}	0.056^{***}
M	M	0.572	0.445	0.745^{***}	0.040^{***}

Notes: Simulation results from applying the generalized deferred-acceptance algorithm (GDA) in Gale and Shapley (1962). First two columns represent worker and ad (employer) preferences: H = highest type, L = lowest type, M = minimum distance between own and partner type. The last two columns represent the estimates of ρ and δ for types in equation (1), where we use the same weighting and formulation as in table 4.

is the data we use for all the previous empirical results. The algorithm is separated into rounds. The algorithm goes as follows:

1. At the beginning of each round, each job position still available (vacancies to be filled) make one offer to their most preferred worker in their queue and who has not yet received an offer for the position.

2. Workers who receive offers, decide on which offer they like the most and keep it for the next round (workers can receive multiple, one or zero offers) while discarding the rest (workers are free to renege on job offers at hand if they receive a better one).

3. Check if jobs that are still not matched have more offers to make and whether there are any unmatched workers. If any of these is false, the process ends.

Outcomes for simulations under different preference rankings are in table 6. Preferences for both workers and ads are labelled as **H**, for highest type of partner, **L** for lowest type of partner or **M**, for the lowest mismatch (absolute difference) between own and prospective partner's type. In the table, we show the percentage of workers and ads that get matched in each scenario, and results from performing the estimation of equation (1) on the resulting counterfactual samples.

The exercise shows a sharp prediction with respect to average sorting (estimates for ρ), in that the range of possible simulated sorting patterns is always positive, ranging from 0.092 for the combination of **H** and **L** preferences for worker and ad, respectively, and going as high as 0.745, when agents have preferences for lowest mismatch. The latter correlation is close to what Hagedorn, Law, and Manovskii (2017) find, using their methodology on German matched employer-employee data. A caveat for a fully direct comparison is that the latter result refers to worker-firm matches whereas ours refers to worker-job matches.

Table 7: Outcome of GDA-algorithm by labor force status

Worker	Ad	Sorting (ρ) All	U	E	Cyclical effect (δ) All	U	E
H	H	0.266***	0.258***	0.263***	−0.021*	−0.052***	0.017*
H	L	0.092***	0.076***	0.105***	−0.027***	−0.058***	−0.005***
H	M	0.539***	0.486***	0.592***	−0.015	−0.058***	0.011*
L	H	0.143***	0.164***	0.112***	−0.034**	−0.041*	−0.014*
L	L	0.124***	0.114***	0.135***	−0.004	−0.005***	−0.011***
L	M	0.662***	0.631***	0.695***	0.015	0.016**	0.002**
M	H	0.456***	0.482***	0.419***	0.026*	0.039*	0.024*
M	L	0.401***	0.395***	0.411***	0.056***	0.062***	0.044***
M	M	0.745***	0.725***	0.769***	0.040***	0.056***	0.016***

Notes: Simulation results from applying the generalized deferred-acceptance algorithm (GDA) in Gale and Shapley (1962). First two columns represent worker and ad (firm) preferences: H = highest type, L = lowest type, M = minimum distance between own and partner type. ρ and δ are the estimates for Types in equation (1), where we use the same weighting and formulation as in table 4.

In terms of cyclical responses of sorting to aggregate conditions (unemployment rate), the results in table 6 are mixed: while the majority of cases exhibit pro-cyclical sorting (negative effect of unemployment on worker-ad type correlation), there are cases where the estimated relationship is opposite, i.e. workers and job positions are less similar to each other than on average in booms. On the other hand, while estimates of ρ are all significantly different from zero (at the 99% level), the estimates of δ are less precise.

In table 7, we present the same exercise splitting up the sample by labor force status of job seekers at the time of application decisions. In terms of average sorting (ρ), the table shows that for most cases, ex post or realized sorting for the unemployed group is lower than for the employed group, with the exception of cases **LH** and **MH**, i.e., when workers prefer either the lowest type of job position or the smallest *mismatch* with them, while ads have preferences for the highest type of worker. As before, the results for cyclical sorting (δ) do not show clear consistent difference between estimates using the unemployed or employed pools.

6 Conclusions

In this paper we use data from the job posting website `www.trabajando.com`, to study sorting between heterogeneous workers and heterogeneous firms. Our approach differs from the literature in that we focus in ex ante assignments, i.e. the revealed preference of a worker to be attached to a particular job. In contrast, virtually all the literature studies realized matches, an ex post assignment.

While studying realized matches is relevant, several issues preclude researchers from learning the underlying assortative pattern driving assignments. First, lack of information regarding job positions in matched employer-employee databases forces the profession to study worker-firm assortative patterns instead: however, worker-job matching patterns are conceptually the real object of interest in the spirit of Becker (1973). Moreover, there is a non-trivial link between size, span of control, productivity of firms that should be considered when studying sorting of larg firms, as shown by Eeckhout and Kircher (2018).

In addition, Eeckhout and Kircher (2011), Hagedorn, Law, and Manovskii (2017), Lopes de Melo (2018) among others, have noticed that theoretical models generate a non-monotonic effect of types on wages implying that the standard worker- and firm-fixed effects model is inappropriate to study sorting. The key intuition is that partners need to compensate over-qualified counterparts for being misplaced in a particular match. Our approach surmounts these problems since we base our analysis in information provided by employers and workers before a particular match takes place. Finally, the information we have from the online job board is richer than the one available in other databases used to study sorting. For these reasons, our measurement of types is likely to be more accurate and free from compensations that agents may offer to their counterparts to form a match.

The obvious limitation for us is the lack of observable ex post assignments: we do not see who get the jobs. Under the assumption that jobs are assigned only to applicants in www.trabajando.com, we simulate realized matches using the Gale and Shapley (1962) algorithm following the rationale of Hitsch, Hortaçsu, and Ariely (2010) (in turn, based on Adachi (2003)). Under different worker and employer (ad) preferences, we show that the ex post assignment would generate PAM ex post, although the magnitude substantially varies under different preference scenarios.

We also study the cyclical behaviour of sorting. Robust evidence shows a clear procyclical pattern, suggesting a mechanism for increasing efficiency in the labor market during booms. The ex ante productivity gains due to this pattern are somewhat higher for the unemployed. Nevertheless, simulations for ex post assignment shows no clear procyclical pattern across preference scenarios.

In sum, we offer a new approach to study sorting in labor markets. While there are limitations, we also offer novel information that help surmount setbacks other researchers have found in the conventional empirical assessment of sorting. Our simulations are an agnostic way to minimize the shortcomings and build a bridge between our results and those in the literature.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 2(67), 251–333.

Adachi, H. (2003): "A search model of two-sided matching under nontransferable utility," *Journal of Economic Theory*, 113(2), 182–198.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008): "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 171(3), 673–697.

Bagger, J., and R. Lentz (2014): "An Empirical Model of Wage Dispersion with Sorting," Working Paper 20031, National Bureau of Economic Research.

Banfi, S., and B. Villena-Roldan (forthcoming): "Do High-Wage Jobs Attract more Applicants? Directed Search Evidence from the Online Labor Market," *Journal of Labor Economics*.

Bartolucci, C., F. Devicienti, and I. Monzón (forthcoming): "Identifying sorting in practice," *American Economic Journal: Applied Economics*.

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4), 813–846.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36(S1), S13–S70.

Card, D., J. Heining, and P. Kline (2013): "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly journal of economics*, 128(3), 967–1015.
ΩŞahin, Song, Topa, and Violante

Şahin, A., J. Song, G. Topa, and G. L. Violante (2014): "Mismatch Unemployment," *American Economic Review*, 104(11), 3529–64.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64(5), 1001–1044.

Eeckhout, J., and P. Kircher (2011): "Identifying Sorting–In Theory," *The Review of Economic Studies*, 78(3), 872–906.

——— (2018): "Assortative Matching With Large Firms," *Econometrica*, 86(1), 85–132.

Gale, D., and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69(1), 9–15.

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

HAGEDORN, M., T. H. LAW, AND I. MANOVSKII (2017): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85(1), 29–65.

HITSCH, G. J., A. HORTAÇSU, AND D. ARIELY (2010): "Matching and Sorting in Online Dating," *American Economic Review*, 100(1), 130–63.

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.

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

LISE, J., C. MEGHIR, AND J.-M. ROBIN (2016): "Matching, sorting and wages," *Review of Economic Dynamics*, 19, 63 – 87, Special Issue in Honor of Dale Mortensen.

LISE, J., AND J.-M. ROBIN (2017): "The macrodynamics of sorting between workers and firms," *American Economic Review*, 107(4), 1104–35.

LOPES DE MELO, R. (2018): "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence," *Journal of Political Economy*, 126(1), 313–346.

MARINESCU, I. E., AND R. RATHELOT (forthcoming): "Mismatch Unemployment and the Geography of Job Search," *American Economic Journal: Macroeconomics*.

MARINESCU, I. E., AND R. P. WOLTHOFF (2015): "Opening the Black Box of the Matching Function: The Power of Words," Discussion Paper 9071, IZA.

POSTEL-VINAY, F., AND J.-M. ROBIN (2002): "The Distribution of Earnings in an Equilibrium Search Model with State-Dependent Offers and Counteroffers," *International Economic Review*, 43(4), 989–1016.

RESTUCCIA, D., AND R. ROGERSON (2017): "The causes and costs of misallocation," *Journal of Economic Perspectives*, 31(3), 151–74.

Roth, A. E., and M. A. O. Sotomayor (1990): *Two-sided matching*, vol. 18 of *Econometric Society Monographs*. Cambridge University Press.

Shimer, R., and L. Smith (2000): "Assortative Matching and Search," *Econometrica*, 68(2), 343–369.

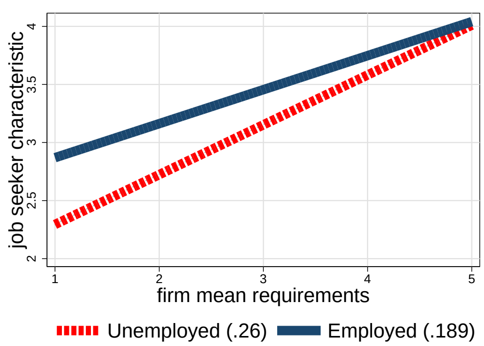
A Appendix: Firm-Worker matches

In this section, we consider matches between workers and firms. Since in our dataset we have firm identifiers for each job posting, we concentrate information of ads by individual firms. Our goal is to make our estimates comparable to those obtained in the two-way worker and firm fixed effects (Abowd, Kramarz, and Margolis, 1999).

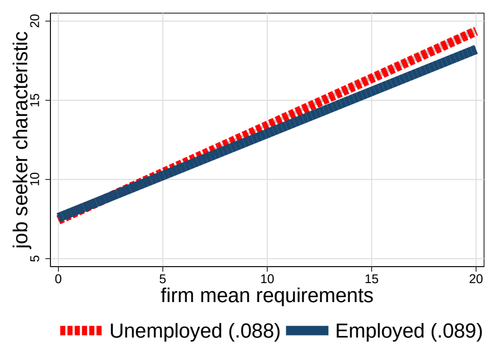
Our approach is as follows:

For types, we run wage regressions between the average wage posted by all job adverts from a firm, on its observables: firm sector, number of employers, region, etc.

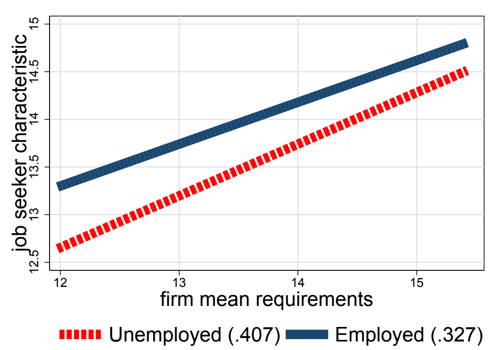
For wages, we use the same residuals as in the worker-job analysis. See the description at the main text.



(a) Education

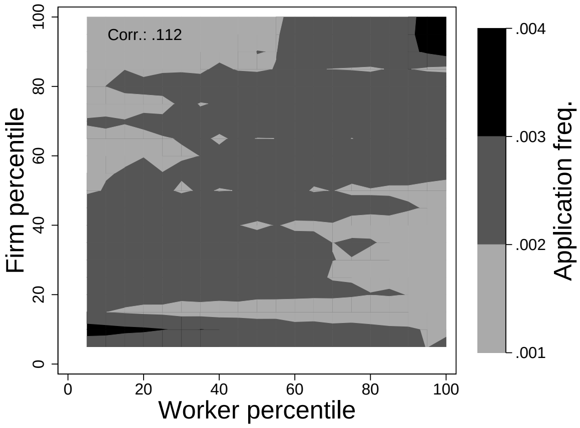


(b) Experience

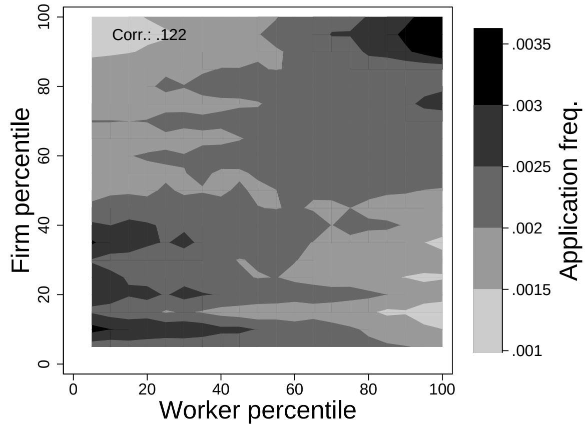


(c) Log wages

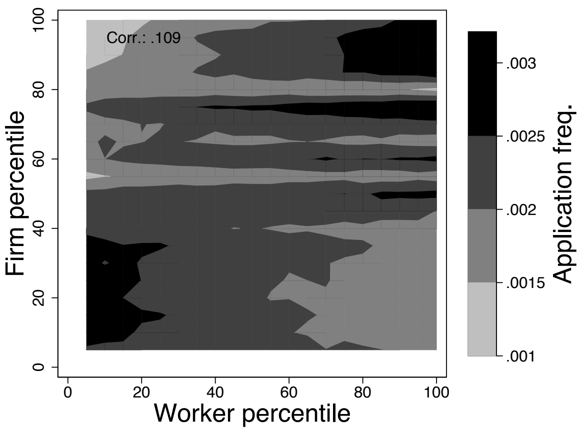
Figure 4: Correlation between actual education of job seekers and required education by job postings. The dotted line is the fitted linear relationship between both axis. All applications between 1-Jan-2013 and 1-Jul-2013.



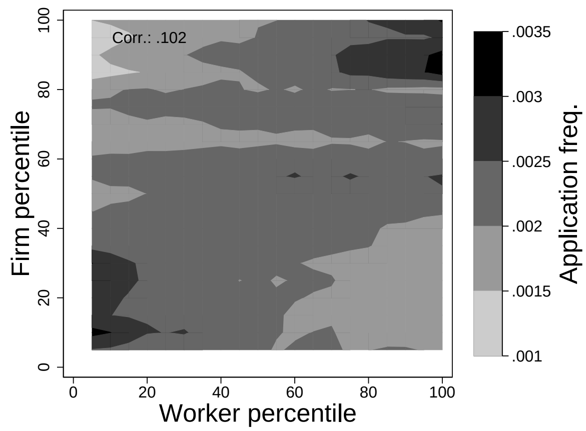
(a) Unemployed explicit applicant



(b) Unemployed implicit applicant



(c) Employed explicit applicant



(d) Employed implicit applicant

Figure 5: Empirical contour plots between desired wage *residuals* by job seekers and *residual* salaries offered by firms (both in logs). Residuals are obtained using a linear regression on wages controlling for worker and firm observable characteristics. All applications between 1-Jan-2013 and 1-Jul-2013.

B Unweighted sorting on time

For the sake of completeness, we provide estimates of cyclical behavior without using DFL weights. Overall, results seen here in table A2 are remarkably close to those in table 4, where we use a reweighing scheme of DiNardo, Fortin, and Lemieux (1996) to keep constant the composition of ads and workers. These evidence implies that varying composition of these populations over time does not affect in an economically significant way our conclusions.

Table A1: Assortative matching (ρ) and cyclical correlation (δ) between job seekers and postings, varying composition sample for log wages and types

	ρ	δ		ρ	δ
All applicants					
Log Wages	0.675***	-0.021***	Education	0.395***	-0.041***
	(0.04)	(0.05)		(0.06)	(0.06)
Types	0.163***	-0.007***	Experience	0.209***	-0.001**
	(0.06)	(0.06)		(0.06)	(0.06)
Unemployed					
Log Wages	0.612***	-0.020***	Education	0.392***	-0.037***
	(0.07)	(0.08)		(0.08)	(0.09)
Types	0.150***	-0.009***	Experience	0.216***	-0.012***
	(0.10)	(0.10)		(0.11)	(0.11)
Employed					
Log Wages	0.590***	-0.027***	Education	0.337***	-0.043***
	(0.06)	(0.06)		(0.09)	(0.09)
Types	0.173***	-0.013***	Experience	0.211***	0.005***
	(0.08)	(0.08)		(0.07)	(0.07)

Notes: 100X standard errors in parentheses. First column is the sorting dimension in equation (1). Types refers to log-wage residuals, as explained in the main text. The cyclical measure is the Chilean non-seasonally adjusted unemployment rate according to the OECD database. Regressions are unweighted.

Table A2: Average and cyclical sorting, constant composition sample, given different information availability sub-markets

			explicit applicant		implicit applicant	
			ρ	δ	ρ	δ
All applicants	explicit ad	Log Wages	0.661***	-0.025***	0.853***	-0.038***
			(0.07)	(0.07)	(0.07)	(0.07)
		Types	0.168***	-0.012***	0.500***	-0.016***
			(0.09)	(0.09)	(0.14)	(0.12)
	implicit ad	Log Wages	0.717***	0.006**	0.894***	-0.012***
			(0.24)	(0.27)	(0.23)	(0.25)
		Types	0.163***	-0.004*	0.627***	-0.057***
			(0.33)	(0.33)	(0.50)	(0.48)
Unemployed	explicit ad	Log Wages	0.619***	-0.014***	0.891***	-0.017***
			(0.11)	(0.13)	(0.11)	(0.12)
		Types	0.163***	-0.013***	0.528***	-0.008***
			(0.16)	(0.17)	(0.22)	(0.21)
	implicit wage ad	Log Wages	0.642***	0.002***	0.898***	-0.007*
			(0.39)	(0.46)	(0.36)	(0.40)
		Types	0.150***	-0.007***	0.633***	-0.054***
			(0.46)	(0.50)	(0.71)	(0.72)
Employed	explicit ad	Log Wages	0.573***	-0.032***	0.756***	-0.044***
			(0.09)	(0.09)	(0.09)	(0.09)
		Types	0.175***	-0.018***	0.483***	-0.017***
			(0.12)	(0.11)	(0.18)	(0.15)
	implicit ad	Log Wages	0.646***	-0.021***	0.794***	-0.031***
			(0.34)	(0.36)	(0.34)	(0.34)
		Types	0.198***	-0.023***	0.618***	-0.060***
			(0.50)	(0.46)	(0.73)	(0.65)

Notes: 100X standard errors in parentheses. First column is the sorting dimension in equation (1). Types refers to log-wage residuals, as explained in the main text. The cyclical measure is the Chilean non-seasonally adjusted unemployment rate according to the OECD database. Regressions are unweighted.

C Alternative cyclical measures

We check whether our cyclical patterns holds for other cyclical measures, such as the log of IMACEC (Monthly Index of Economic Activity, in Spanish *Índice Mensual de Actividad Económica* or better known as IMACEC is a monthly GDP measure covering nearly 90% of Chilean economic activity reported by the National Accounts Division of the Central Bank of Chile). Since we control for quaterly dummies, there is no need for detrending the indicator to assess the impact of its cyclical behavior on sorting.

The results in table A3 are fully consistent with those in the main text, in which we use the unemployment rate as cyclical variable. Of course, the sign of the interaction term δ to asses the cyclical pattern has a positive sign since the unemployment rate is known to be highly counter-cyclical.

Table A3: Assortative matching (ρ) and cyclical IMACEC correlation (δ) between job seekers and postings, varying composition sample for log wages and types

All applicants					
	ρ	δ		ρ	δ
Log Wages	0.665***	0.013***	Education	0.387***	0.051***
	(0.13)	(0.15)		(0.18)	(0.21)
Types	0.154***	0.006***	Experience	0.213***	0.010***
	(0.17)	(0.20)		(0.15)	(0.18)
Unemployed					
	ρ	δ		ρ	δ
Log Wages	0.597***	0.023***	Education	0.376***	0.047***
	(0.23)	(0.25)		(0.28)	(0.33)
Types	0.147***	0.010***	Experience	0.219***	0.011***
	(0.32)	(0.36)		(0.31)	(0.35)
Employed					
	ρ	δ		ρ	δ
Log Wages	0.592***	0.017***	Education	0.347***	0.052***
	(0.15)	(0.18)		(0.25)	(0.28)
Types	0.165***	0.010***	Experience	0.213***	0.008***
	(0.19)	(0.23)		(0.17)	(0.21)

Notes: 100X standard errors in parentheses. First column is the sorting dimension in equation (1). Types refers to log-wage residuals, as explained in the main text. The cyclical measure is the log of the Monthly Index of Economic Activity (see text). Regressions use DFL weights (see main text), which are computed using a probit model in which the dependent variable is an indicator for 2011Q3, and independent variables are for applicants: age, age squared, gender, gender interacted with age terms, and a full array of indicators of nationality, marital status, region, educational mayor. Independent variables for job ads are indicators for region, industry, economic activity (job board classification), educational area required, and firm size category.

Table A4: Average and cyclical sorting (IMACEC), constant composition sample, given different information sub-markets

All applicants					
	ρ	δ		ρ	δ
Log Wages	0.675***	0.022***	Education	0.395***	0.044***
	(0.04)	(0.05)		(0.06)	(0.06)
Types	0.164***	0.009***	Experience	0.209***	0.003***
	(0.06)	(0.06)		(0.06)	(0.06)
Unemployed					
	ρ	δ		ρ	δ
Log Wages	0.611***	0.023***	Education	0.390***	0.037***
	(0.07)	(0.08)		(0.08)	(0.09)
Types	0.150***	0.010***	Experience	0.215***	0.014***
	(0.10)	(0.10)		(0.11)	(0.11)
Employed					
	ρ	δ		ρ	δ
Log Wages	0.591***	0.030***	Education	0.340***	0.049***
	(0.06)	(0.06)		(0.10)	(0.09)
Types	0.174***	0.017***	Experience	0.211***	-0.005***
	(0.08)	(0.08)		(0.07)	(0.07)

Notes: 100X standard errors in parentheses. Types refers to log-wage residuals, as explained in the main text. The cyclical measure is the log of the Monthly Index of Economic Activity (see text). Regression are weighted using DFL weights (see main text), which are computed using a probit model in which the dependent variable is an indicator for 2011Q3, and independent variables are for applicants: age, age squared, gender, gender interacted with age terms, and a full array of indicators of nationality, marital status, region, educational mayor. Independent variables for job ads are indicators for region, industry, economic activity (job board classification), educational area required, and firm size category.