Sorting On-line and On-time

Banfi, Stefano and Choi, Sekyu and Villena-Roldán, Benjamin

Chilean Ministry of Energy, University of Bristol, UK, Universidad Andres Bello

January 2019

Sorting On-line and On-time*

Stefano Banfi
Fiscalía Nacional Económica de Chile

Sekyu Choi
University of Bristol

Benjamín Villena-Roldán
Departament of Economics, Universidad Diego Portales

January 13, 2022

Abstract

Using proprietary data from a Chilean online job board, we compute sorting between workers and job positions during the application stage (ex ante) and predict sorting in the flow and stock of created matches (ex post) for different type measures. We find strong evidence for positive and procyclical correlations between workers and job types. Since ex ante and ex post sorting are very similar, we conclude that sorting is largely generated at the application stage. This suggests that theoretical models of sorting with directed search are a promising path for future research.

Keywords: Online search, assortative matching, labor markets, applications.

JEL Codes: E24, E32, J24, J60

*Email: stefano.banfi@gmail.com , sekyu.choi@bristol.ac.uk and benjamin@benjaminvillena.com. We thank Katarina Borovickova, Pieter Gautier, Ronald Wolthoff, Jan Eeckhout, Shouyong Shi, Yongsung Chang, Chao He, Ilse Lindenlaub, Chinhui Juhn, and Germán Cubas, the editor Florin O. Bilbiie and two anonymous referees for comments and discussions. We acknowledge the support of the SABE (Sistema de Análisis de Bolsas de Empleo) project team (Juan Velásquez, Rocío Ruiz and Felipe Vera) for online job board data curatory and financial support of SENCE and OTIC SOFOFA. We are also in debt to www.trabajando.com for data provision, and especially to Ramón Rodríguez, Jorge Vergara, and Ignacio Brunner for informative conversations and support. We also thank colleagues at 2021 SaM Annual Meeting, University of Houston, 2018 SOLE Meeting, Cardiff University, 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 gratefully acknowledges financial support from the Institute for Research in Market Imperfections and Public Policy, MIPP (ICS13_002 ANID), FONDECYT grant projects 1151479 and 1191888, Proyecto CONICYT PIA SOC 1402, as well as Universidad de Chile. All errors are ours.

1 Introduction

Sorting in labor markets has been extensively studied in economics: whether high wage workers work (or not) for high wage firms is relevant for questions of efficiency and inequality. However, little is known about the way these allocations are generated: do workers and firms meet randomly, with sorting arising ex post by selective hiring and separations? is the ex ante self-selection of workers into different positions at the application stage relevant for sorting?

Using data from the Chilean online job board `www.trabajando.com` from 2010 to 2019, we study the importance of ex ante versus ex post forces in determining labor market sorting. Using information of job seekers and job ads linked through applications on the job board, we can estimate several notions of worker and job types as well as their correlations, something we label *ex ante sorting*. Using an external, longitudinal household survey for the Chilean economy and a simple imputation procedure to match observed allocations of new hires, we predict the amount of ex ante sorting that is likely to prevail ex post.

As a first contribution, we find a large and robust correlation between worker and job types: around 0.6 using our preferred measure. When comparing ex ante and ex post correlations, we find that the two numbers are remarkably similar, leading us to conclude that most of labor market sorting is due to application decisions of workers (the ex ante sorting). We also provide evidence that sorting is significantly procyclical, thus economic downturns have a sullying (rather than cleansing) effect on matches.[1]

We provide a second contribution in terms of measurement of types. Since the seminal work of Abowd, Kramarz, and Margolis (1999) (AKM henceforth) economists have looked for ways to identify and estimate productivity types from realized wages. The key idea is that, in expectation, high productivity workers earn high wages regardless of the specific match they are in and high productivity positions expect to pay systematically higher wages independent of the worker. Instead of relying on a particular theoretical, statistical or structural model to identify types, we use the fact that our data contains these wage expectations directly: employers declare wages they expect to pay before hiring anyone in particular, and workers declare a wage they expect to receive before applying to any job.[2] Additionally, we show how we can compute with our data a measure similar to the AKM fixed effects, which we label *ex ante AKM*. We show that the sorting implied by this measure is substantially lower than the one obtained from our preferred types, as expected in theory and in practice.[3] Unlike the standard AKM measures, our ex ante types are immune from ex post compensation issues (see Eeckhout and Kircher, 2011; Hagedorn, Law, and Manovskii, 2017) and low mobility bias (see Andrews, Bradley, Stott, and Upward, 2008; Bonhomme, Lamadon, and

[1] Our evidence is related to the findings of Crane, Hyatt, and Murray (2018).

[2] We see our results as complementary to the structural frameworks in Lise and Robin (2017) and Bagger and Lentz (2018).

[3] See Eeckhout and Kircher (2011), Hagedorn, Law, and Manovskii (2017), Lopes de Melo (2018), Bonhomme, Lamadon, and Manresa (2019), among others.

Manresa, 2019).

Our results show the need of expanding the focus of labor market sorting from worker-firm pairs to include information from worker-job allocations, which seems conceptually closer to the original idea in Becker (1973). While worker-firm assignment is interesting on its own regard because of its links to raising inequality in different contexts,[4] worker-job matches can provide a richer picture of the labor market in terms of skills mismatch, occupational mobility, and overall efficiency, specially in the context of technological change affecting the task content of jobs. This focus is closely related to a rapidly growing strand of research on job search in frictional environments where individuals match with jobs or occupations rather than with firms per se, as in Baley, Figueiredo, and Ulbricht (2020), Guvenen, Kuruscu, Tanaka, and Wiczer (2020), Taber and Vejlin (2020), Dvorkin and Monge-Naranjo (2019), Cubas and Silos (2020) among others. We provide further evidence of the relevance of worker-job assignments by estimating job types, both with and without controlling for the posting firm's fixed effects. Sorting patterns remain independent of the choice of fixed effects, which shows that jobs alone are a meaningful margin to study.

This paper is related to the growing literature using online job-posting and retail websites in order to study different aspects of frictional markets. This literature is already vast, so we refer to papers that provide evidence on how posted wages or prices affect behavior of applicants or buyers. 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 Wolthoff (2020) use the data from the `Careerbuilder.com` posting website to study the relationship between job titles and wages posted on job advertisements. Lewis (2011) and Banfi and Villena-Roldán (2019) 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.

In sum, we offer a new approach to study sorting in labor markets. While there are some limitations to our exercise (mainly, we do not observe the final allocations of workers to jobs), our setting offers a transparent and straightforward way to measure sorting. The fact that sorting at the application stage is remarkably similar to ex post sorting, suggests that theoretical settings, along the lines of Shimer (2005) and Abowd, Kramarz, Pérez-Duarte, and Schmutte (2018), in which directed search is a key ingredient, are a desirable path for future research in this area.

The next section describes thoroughly our main data set which is needed to explain how we intend to rationalize our results given the analytical framework we present in section 3. In sections 4 and 5 we present our results for the cross section and cyclical conditions, respectively. Section 6

[4]See for example Card, Heining, and Kline (2013) and Song, Price, Guvenen, Bloom, and Von Wachter (2019).

presents our description of the imputation procedure we use to predict ex post sorting and presents results from that exercise. Section 7 relates our findings to existing theories of sorting in labor markets. The last section concludes.

2 The Data

We use data from `www.trabajando.com` (henceforth TC or the website). Our data covers daily dated job postings and job seeker activity in the Chilean labor market, between January 1st, 2010 and December 31st, 2019. We observe entire histories of applications (dates and identification numbers of jobs applied) from job seekers and dates of ad postings for employers. After some cleaning and applying restrictions, our final dataset contains 39,371,368 applications linking 1,687,578 applicants and 1,135,080 job ads. We also use information from the *Encuesta Nacional de Empleo* (National Employment Survey in Spanish, ENE henceforth)[5] for some representativeness analysis and the ex post sorting exercise later in the paper.

A novel feature of the dataset, is that the website asks job seekers to record a salary figure ("expectativa de renta" in Spanish), customarily a monthly amount, net of taxes.[6] Employers are also asked to record the net monthly salary for the advertised job. Agents on both sides of the market can then choose to show or hide the salary from potential partners on the other side. While there is no unique rationalization about why individuals and firms may choose to hide this information, Banfi and Villena-Roldán (2019) provide some suggestive evidence: hidden/implicit job ads usually post higher job requirements (education and/or experience), expect to pay higher wages and receive more applications on average, all evidence of strategic signaling, along the lines of Michelacci and Suarez (2006). In the next section, we provide a lengthy discussion as to how to use this information for sorting analysis.

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-Roldán (2019) 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 (not aligned with international standardized codes), geographic information, and some limited information on the firm offering the job: its size (brackets

[5] See the Appendix for further details of this survey.

[6] 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%).

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 and educational attainment. We have codes for a referential occupational area (also not aligned with standardized taxonomies) 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 (if unemployed).

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, to avoid dealing with education and retirement decisions. We discard individuals reporting desired monthly net wages above 5,000,000 CLP, an amount well above the 99th percentile of the wage distribution in Chilean household surveys. We also discard individuals who desire net wages below 150,000 CLP a month, somewhat below the minimum wage for full-time workers in Chile in 2010 (165,000 CLP), which has been steadily increasing ever since. 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 workers who make at least one application and job postings which receive at least one application during the span of our dataset.

As with many self-reported data sources, there are measurement issues. Individuals may misrepresent information in their CVs to look more appealing for employers, but this may be a dangerous strategy for job seekers in a concrete hiring situation as their credentials are likely verified. One caveat with our analysis is that the worker information is a snapshot of their last CV as of Dec 31st, 2019. For job seekers who have never updated their CV, this is not an issue. However, if job seekers change their CV in between applications, we may correlate information of a newer CV with information of the job ads they applied to. Since we focus on job seekers between 25 and 55 years old, most of them have already finished their education and are still sufficiently far from retirement so that we should not expect dramatic CV changes over the time we observe their applications, which is ten years at most.[7] In addition, 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 at the application time (we offer a robustness check in the appendix which confirms our intuition).

Table 1 shows some descriptive statistics for individuals in our sample. From the table we observe that males are a majority of all job seekers, especially among employed job seekers. The sample is young, with an average age of 33.8 years at the time of application. Job seekers above 50 years are scarce. Given the age group we consider, most individuals in the sample have some working experience, with the mean number of years of experience hovering around 8. Job seekers in our sample are more educated than the average in Chile, with 34.8% of them claiming a college degree

[7] The only variable that is updated for an individual is age, since we compute it as the difference between the application date and the candidate's birthday.

or more, compared to around 22% in a similar age group and time frame (2010-2019) according to the ENE.

We can also observe that most job seekers have studies related to management (around 17.2%) and technology (28.1%), but a significant fraction (around 26.9%) does not declare any specialization. In terms of salaries, average expected monthly salaries are (in thousands of Chilean Pesos) $1,033.9$ and 626.2 for employed and unemployed seekers, respectively ($1,762$ and $1,067$ USD).[8] Compared to the salary distribution in Chile according to ENE,[9] the average expected job wage of 813.8 thousands of CLP is 1.24 and 1.78 times the average and median salaries, respectively, over 2010-2019.

Table 2 shows sample statistics for job postings. We separate our sample between postings with hidden or implicit wages (which do not post information on salaries) and with explicit wages. From a total of $1,135,080$ active job postings in our sample period, only $183,858$ (16.2%) have an explicit wage.

Hidden wage postings are characterized for requiring higher levels of experience and higher levels of education. They also tend to concentrate more on technology related occupations: 30.9% of ads with hidden wages are related to technology, versus 18.8% of job postings with explicit wages. Job postings in our sample receive 34.7 applications on average, with a significant difference in the number received by implicit wage postings (37.2) versus those received by explicit ones (21.6). Banfi and Villena-Roldán (2019) properly show that employers tend to hide wages when ads post a high wage, and require high education and experience. This evidence is consistent with employers signaling openness to ex post bargaining, along the lines of the model of Michelacci and Suarez (2006). The average of posted wages is near 709 thousands of CLP (1,208 USD), being the explicit-wage subsample considerably lower. This average TC job ad salary is 1.08 and 1.55 times the average and median salary, respectively, for the whole Chilean economy. In any case, jobs posted in the website offer wages that are substantially higher than those measured by household surveys in Chile. In Appendix A1 we show further statistics for both applicants and job ads.

[8]The CLP/USD average exchange rate for 2010-19 was 586.88.

[9]We use information from the *Encuesta Suplementaria de Ingresos* (ESI), a supplement to the ENE, akin to what the Merged Outgoing Rotational Group (MORG) is to the Current Population Survey (CPS) in the US. See http://webanterior.ine.cl/estadisticas/laborales/ene and http://webanterior.ine.cl/estadisticas/ingresos-y-gastos/esi.

Table 1: Characteristics of Job Seekers

	Employed	Unemployed	Total
Gender (%)			
Females	41.60	49.30	45.75
Males	58.40	50.70	54.24
Age (%)			
25 - 29	40.67	44.16	42.54
30 - 34	23.72	20.93	22.21
35 - 39	15.33	13.59	14.39
40 - 44	9.85	9.49	9.66
45 - 50	6.25	6.84	6.57
$>$ 50	4.18	4.98	4.62
Applicants Age			
Age (Mean/(S.D.))	33.79	33.68	33.73
	(7.41)	(7.80)	(7.62)
Experience (%)			
0 - 3	23.64	33.47	28.93
4 - 7	29.16	25.39	27.10
8 - 12	24.59	21.12	22.70
13 - 20	17.04	14.53	15.67
$>$ 21	5.51	5.29	5.39
Applicants Experience			
Years of Experience (Mean/(S.D.))	8.59	7.64	8.08
	(6.51)	(6.87)	(6.73)
Education level (%)			
Primary (1-8 years)	0.11	0.36	0.24
High School	28.11	39.78	34.41
Tech. Tertiary Educ.	23.11	23.98	23.57
College	44.00	26.90	34.77
Post-Graduate	4.63	8.89	6.93
Education area (%)			
Commerce and Management	17.93	16.50	17.15
Agropecuary	1.11	0.81	0.95
Art and Architecture	2.45	1.78	2.09
Natural Sciences	1.21	1.38	1.30
Social Sciences	5.73	4.20	4.90
Law	2.03	1.72	1.86
Education	3.92	4.13	4.03
Humanities	1.08	0.98	1.03
Health	5.80	6.67	6.27
Technology	32.12	24.77	28.14
Non-declared	18.54	34.09	26.94
Other	8.09	2.96	5.32
Expected wages, thousands of CLP (%)			
150 - 300	6.23	18.19	12.68
300 - 600	28.96	48.00	39.24
600 - 1,000	31.79	23.07	27.08
1,000 - 1,500	15.74	5.99	10.47
1,500 - 2,500	12.32	2.96	7.26
$>$ 2,500	4.70	1.06	2.73
Not declared	0.24	0.07	0.15
Wages			
Expected wages (Mean/(SD))	1,033.94	626.16	813.81
	(757.60)	(481.20)	(656.08)
Observations	775,960	909,889	1,687,578

Table 2: Characteristics of Job Ads

	Hidden wage	Explicit wage	Total
Required Experience (%)			
0	15.72	22.75	16.86
1	34.83	44.75	36.43
2 - 3	36.78	26.55	35.12
4 - 7	10.94	5.28	10.02
8 - 12	1.56	0.58	1.40
> 12	0.17	0.09	0.16
Required Experience			
Experience (Mean/(SD))	1.92	1.39	1.84
	(1.81)	(1.45)	(1.77)
Required educ. level (%)			
Primary (1-8 years)	1.33	5.08	1.94
High School	33.07	51.96	36.12
Tech. Tertiary Educ.	29.20	27.29	28.89
College	35.78	15.48	32.49
Graduate	0.62	0.19	0.55
Major study area (%)			
Commerce and Management	13.02	12.60	12.95
Agropecuary	0.53	0.71	0.56
Art and Architecture	0.36	0.26	0.34
Natural Sciences	0.59	0.90	0.64
Social Sciences	3.41	2.19	3.21
Law	0.67	0.45	0.63
Education	0.91	0.96	0.92
Humanities	0.20	0.29	0.22
Health	1.78	1.51	1.74
Technology	30.89	18.81	28.93
Non-declared	47.48	61.16	49.69
Other	0.16	0.16	0.16
Offered wages, thousands of CLP (%)			
150 - 300	18.60	29.29	20.33
300 - 600	36.42	50.25	38.66
600 - 1,000	24.14	12.79	22.30
1,000 - 1,500	11.05	3.78	9.87
1,500 - 2,500	6.67	1.52	5.83
> 2,5001	1.79	0.51	1.58
Not declared	0.10	0.05	0.10
Offered Wages			
Wages (Mean/(SD))	752.67	486.01	709.48
	(603.98)	(381.36)	(582.17)
Observations	951,222	183,858	1,135,080

To assess the representativeness of our data when compared to the Chilean Economy, in table 3 we compare some salary statistics computed from ENE. We use the same sample period (2010-2019) and apply the same age and salary (minimum/maximum) restrictions as in our TC data. We also focus only on full-time workers not on short-term/temporary contracts.

In table 3 we consider several cases: all workers, new workers with job tenure of less than twelve months (which represent a sample who has recently performed a job search), new workers who do

Table 3: Wage distribution comparison: household survey vs TC

	p10	p25	p50	p75	p90	mean	sd	N
all	250	319	456	753	1248	653	578	91,195
newly hired	229	292	381	553	878	505	427	22,409
newly hired, excl. agric, fishing, construction, public	230	300	390	570	910	520	457	15,239
newly hired, excl. agric, fishing, construction, public & workers with high school educ or more	240	300	400	600	999	542	476	13,210
TC workers	300	400	600	1000	1500	814	656	1,687,578
TC ads	250	340	500	900	1500	709	582	1,135,080
TC vacancies	200	270	350	500	900	487	435	4,672,821

not work in agriculture, fishing, construction, and public sectors (poorly represented in TC, as seen in table 2) and new workers, not working in those sectors, who have at least completed high school because job seekers with lower education are scarce in TC data.

As discussed above, the sample of workers and job ads in TC is described by monthly wages which are higher than the overall distribution observed in the Chilean data. As the last row of Table 3 makes clear, when ads are weighted by the number of vacancies the distribution of salaries matches the one in ENE much closely. This is explained because job ads with multiple vacancies often aim at attracting unskilled labor, with lower salaries and requirements.

Besides this level differences between TC wages and the rest of the Chilean economy, wages in the website still present a significant spread, as seen by the different percentiles in the table and relatively in line with the dispersion observed in the Chilean economy. What is more important, there is alignment between wage expectations by workers and job positions in the website.

3 Analytical framework

As discussed in the introduction, we are focused on *pre-match* allocations.[10] In this section we attempt at providing links between our results and current empirical frameworks.

3.1 Types definitions

In our data we focus on two key variables: (i) *the offered salary* that employers expect to pay to a hired worker for the advertised position *before knowing any specific applicant*, and (ii) the anticipated worker salary, i.e. the net monthly figure job seekers expect to receive when hired by an employer (job ad) *before applying to any specific job ad*. Therefore, we interpret the first variable as the average net monthly salary that the employer intends to pay to job seekers in situations when both sides *likely accept* to form a match. The second variable, described as "renta esperada" in Spanish (expected salary) is naturally interpreted as the wage the job seeker expects to earn in a potential match that both sides *likely accept*.

[10]To the best of our knowledge, there is no model in the literature that encompasses all features of our novel setup. Building a new model is beyond the scope of our paper.

Matches are "likely" acceptable because job seekers apply to multiple jobs and job ads receive multiple applications. Therefore, a match between a job ad a and a worker w is formed depending not only on their types, but also on the types of other employers to which w applies to and/or the other job seekers who apply to job a. For example, a job ad a would have hired w if another more qualified applicant w' had not applied to a and accepted the offer. In turn, the offer of employer a would have been taken if w' had not applied to a more attractive job a'. Assuming that job seekers face a positive marginal cost for applying to a job, their applications are an expression of interest in a potential match that could occur with strictly positive probability, a likely acceptable one.

To formalize these ideas, define $Q(a, w)$ as the indicator function when w has applied to a (worker w "is on the queue" for job a) and s as the expected salary for that (potential) match. A *likely acceptable* set of job ads by worker w is thus

$$L_w \coloneqq \{a \in \mathcal{A} : Q(a, w) = 1\}, \tag{1}$$

where $\mathcal{A}$ is the set of available job ads at the time the worker w seeks for a job. The set for job ad a, L_a is defined analogously, with $\mathcal{W}$ being the set of workers in queuing for job a:

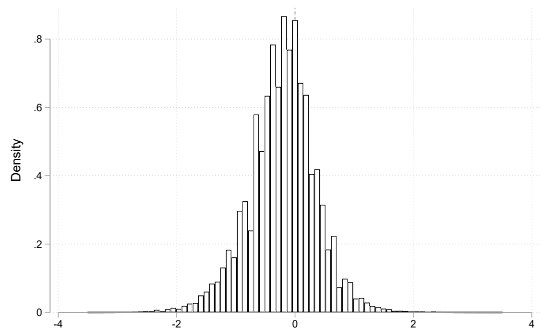
$$L_a \coloneqq \{w \in \mathcal{W} : Q(a, w) = 1\}, \tag{2}$$

Given this, we can define *expected salary* types for both workers and ads as

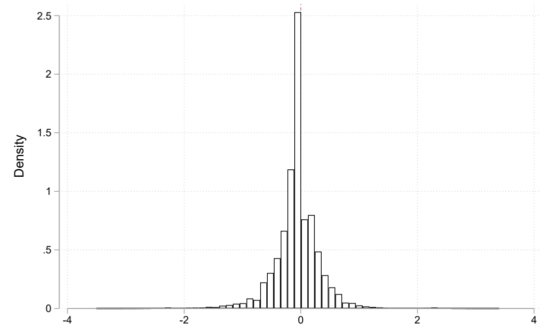
$$ES_w = \log E[s|L_w]$$
$$ES_a = \log E[s|L_a]$$

Before establishing whether these variables are meaningful, we offer some independent evidence that these declared expectations provide an approximation to the expected wage, conditional on the acceptable set of matches. In the left panel of figure 1 we show that the (log) difference between the wage of an ad and the expected wage of a job seeker (who has applied to the position, i.e. $Q(a, w) = 1$ in the data) is not that far from symmetrically distributed around zero, although the average gap is -0.211 log points, i.e. workers on average anticipate wages that are somewhat higher than the actual jobs they apply to (see more details descriptive statistics in appendix A2). A negative gap is not a surprise: workers may apply more often low-wage jobs as high-wage jobs receive more applications and are harder to obtain (Banfi and Villena-Roldán, 2019). The symmetric shape of the distribution suggests that workers *do not* take the *expected salary* question as either a maximum or minimum (reservation) expected salary.

This rationale is reinforced by the right panel of the same figure, where we plot the average difference (across individuals) between salaries in their last job (or current job for those employed job seekers) and their current expected salary. The average gap in this case is -0.053 log points,



(a) Gap ad posted-worker anticipation ($ES_a - ES_w$)



(b) Gap last wage - worker anticipation

Figure 1: Differences in log wages. Left panel: wage posted in ad minus anticipated worker wage. Right panel: wage in last/current job (for unemployed/employed individuals) minus anticipated wage. Sample consists of 39,371,368 applications between 1-Jan-2010 and 31-Dec-2019 in `www.trabajando.com`

with a lot of mass of the distribution around zero (high kurtosis) suggesting that workers take their realized previous or current wage as an important reference for their anticipated wages.

A final remark on the reliability of expected wages as types: posted and expected wages are consequential for agents in the search process, as argued by Banfi and Villena-Roldán (2019). The search engine of the website filters job ads by wage brackets even if employers choose to conceal wages from applicants. On the other side of the market, employers may observe the job seeker's wage declared expectation if visible, potentially affecting contact decisions.

Our ES type definition offers two advantages with respect to traditional measures. First, ES types are not match-specific, which avoids the ex post mismatch compensation mechanism highlighted in many theoretical models of labor market sorting (see for example Eeckhout and Kircher, 2011; Hagedorn, Law, and Manovskii, 2017). Second, as ES types are directly obtained from self-declared wage anticipations, we can circumvent the "low mobility bias" which produces attenuation of the types' correlation in the standard two-way fixed effects (AKM) of Abowd, Kramarz, and Margolis (1999). This frequent problem occurs when workers have very few employers over the typical span of matched employer-employee databases, making the estimation of fixed effects noisy (see Andrews, Bradley, Stott, and Upward, 2008; Bonhomme, Lamadon, and Manresa, 2019).

3.2 ES types and the AKM framework

Below we compare ES types to the AKM framework. Our analysis links the two and suggests a novel formulation, which we denote the *ex ante* AKM (eAKM) type. Let

$$\log s = w + a + \epsilon \tag{3}$$

where s is the salary of worker w at job a. According to this AKM specification (and abusing notation) $\log s$ depends on the type of the worker w, the type of the position a and an error term ϵ, which is assumed to satisfy $E[\epsilon|a,w] = 0$. Given the definitions in (1) and (2), we obtain

$$E[\log s|L_w] = w + E[a|L_w]$$
$$E[\log s|L_a] = a + E[w|L_a]$$

Term $E[a|L_w]$ is the expectation of the job type worker w could likely match with given her applications. Analogously, $E[w|L_a]$ is the expectation of the worker type job a may likely match with. In other words, they represent the average type of partners each side of the market expects. We approximate these definitions by

$$E[\log s|L_w] \approx w + \overline{X}_{L_w}\gamma$$
$$E[\log s|L_a] \approx a + \overline{X}_{L_a}\delta$$

where vector $\overline{X}_{L_w}$ contains average characteristics of jobs for which worker w is applying to: indicators for explicit wage posting, firm size, firm's geographic location, educational requirements, contract types (long term, fixed term, etc), time arrangements (full- or part-time), number of vacancies, required experience, and number of employed and unemployed applicants received by the job ad. On the other side of the market, $\overline{X}_{L_a}$ includes averages of: dummy variables for gender, nationality, marital status, region of residence, educational attainment, educational area, age and self reported experience of job seekers, as well as indicators for whether individual applicants disclose wage expectations. Both γ and δ are parameters.

Note that expected salaries on both sides of the market reflect expectations in levels rather than in logarithms. Therefore, we approximate $E[\log s|L_w]$ by $\log E[s|L_w] = ES_w$ and $E[\log s|L_a]$ by $\log E[s|L_a] = ES_a$ using a first-order Taylor expansion.[11] Using these arguments, we estimate by ordinary least squares the parameters γ and δ:

$$ES_w = w + \overline{X}_{L_w}\gamma + \epsilon_w$$
$$ES_a = a + \overline{X}_{L_a}\delta + \epsilon_a$$

[11] Allowing for a second-order Taylor expansion yields that $E[\log s|L_w] \approx \log E[s|L_w] - CV[s|L_w]^2$, where CV is the coefficient of variation, which is typically increasing in w, i.e. larger salaries have higher relative variance conditional on types. If this is the case, the downward bias is specially large for higher types, suggesting that positive assortative matching (PAM) is likely to be underestimated.

where ϵ_w and ϵ_a are estimation error terms. Thus, we compute *ex ante AKM types* (eAKM) as

$$eAKM_w = ES_w - \overline{X}_{L_w}\widehat{\gamma} \tag{4}$$

$$eAKM_a = ES_a - \overline{X}_{L_a}\widehat{\delta} \tag{5}$$

These eAKM types capture the time-invariant or permanent productivity component in the original AKM framework using self-declared expected salaries on both sides of the market. To construct these types, we assume that the average characteristics of potential partners contain relevant information regarding the kind of partner an agent expects to match. To illustrate how the data variation identifies types, consider a simple example. Suppose that two workers of similar declared expected salary apply to a different set of jobs. The first worker aims at firms and positions offering very valuable characteristics or amenities, while the second applies to jobs that seem less desirable. This variation identifies that the first worker has a lower permanent salary because a larger part of his expected wage is due to job characteristics or amenities. In contrast, we infer that the second worker has a higher specific time-invariant or permanent wage component because he applies to less attractive jobs. The principle is quite similar to the AKM identification source in matched employer-employee datasets, in which time-invariant fixed effects are regarded as types conveying a permanent component of productivity. We obtain this information in our procedure to the extent that anticipated wages reflect a permanent component of agent productivity. In our example, both workers end up with similar salaries, but the first worker is matched to a high-wage firm, implying that his type is lower than the type of the second worker.

Since we are focusing on the concept of a worker-position match instead of the more traditional worker-firm match, we must consider how much controlling for firm's fixed effects alters our results. If position productivity is firm-dependent, or if the firm is an attribute of the position in itself, we should not include firm fixed effects in our estimates above. Under this assumption, (4) and (5) correctly approximate eAKM types. On the contrary, if we think that firm fixed effects are additively separable from the position type a, then we should include a term f in equation (3). We call this second measure eAKM-f types. We remain agnostic and produce both estimates in the next section.

The key difference between ES/eAKM and AKM types is the distribution of available partners on the other side of the market. While the derivation of ES and eAKM types conditions on the distribution of applications, i.e. $Q(a, w) = 1$, the traditional AKM setting is conditional on a realized hiring, which we indicate as $H(a, w) = 1$. Define $M_w = \{a \in \mathcal{A} : H(a, w) = 1\}$ as the set of jobs that potentially hire worker w, and $M_a = \{w \in \mathcal{W} : H(a, w) = 1\}$, as the set of job seekers who are potentially hired by the employer posting the job ad a. How are these ex ante AKM types related to the standard (ex post) AKM types? if we consider that any hiring has been generated through an application, any hired worker w must have applied to the job a. Conversely, any job

a taken by a worker w received an application from her. We logically conclude that any ad a in M_w must also belong to L_w, i.e. $M_w \subseteq L_w$. Likewise, any job seeker w in M_a must be in L_a, i.e. $M_a \subseteq L_a$.

Thus, conceptually, we just need to take expectations over the set of potential hires given w, M_w in (3) under the stated assumptions, which yields

$$E[\log s|M_w] = w + E[a|M_w]$$
$$E[\log s|M_a] = a + E[w|M_a]$$

To perform a strict comparison between these measures (eAKM vs standard AKM), one would need a matched employer-employee dataset *along with* a dataset similar to ours. While no such combination exists (to the best of our knowledge), we provide an alternative in section 6, where we use the DiNardo, Fortin, and Lemieux (1996) semi-parametric procedure to approximate the distribution of ex ante potential matches to the realized ones using household survey data.

4 Sorting on-line

In this section, we report cross-sectional sorting between workers (w) and job ads (a) using the measures described above: expected salary types (ES) and ex ante AKM, with (eAKM-f) and without (eAKM) firm fixed effects.

In Figure 2 we show the joint frequency of observed applications given 30 quantiles of worker and ad types, that is, their empirical bivariate distribution. A darker area reflects high percentage of applications made by workers in a given quantile (projected onto the vertical axis) and received by job ads in a another quantile (projected onto the horizontal axis). The first subfigure shows significant alignment of ES types. Lighter gaps can be seen due to the fact that most wages posted are round numbers. Some bunching of applications is observed for low worker and job ad ES types. As we move further away from the 45 degree line, more lighter tones appear, reflecting that applications combining low and high types are nearly 10 times more infrequent than the most abundant cases, depicted in black. Overall, the correlation of ES types is around 0.6, suggesting strong positive assortative matching. The number is highly significant.[12]

Both eAKM and eAKM-f types show smoother joint densities exhibiting some bunching on the extremes of the 45 degree line. As we move away from the latter line, we observe progressively lighter tones that are consistent with positive sorting. For these measures we find weaker assortative patterns than the one found for ES types: correlations of eAKM and eAKM-f types are lower than

[12] We compute White (1980) robust standard errors in our regressions. Following the method of Cameron, Gelbach, and Miller (2011), we also compute standard errors using a two-way cluster (at the job ad and applicant levels). These standard errors are between 17% and 31% larger than White's, and therefore do not really affect p-values (given that our sample consists of several million observations). Thus, we focus on economic rather than statistical significance.

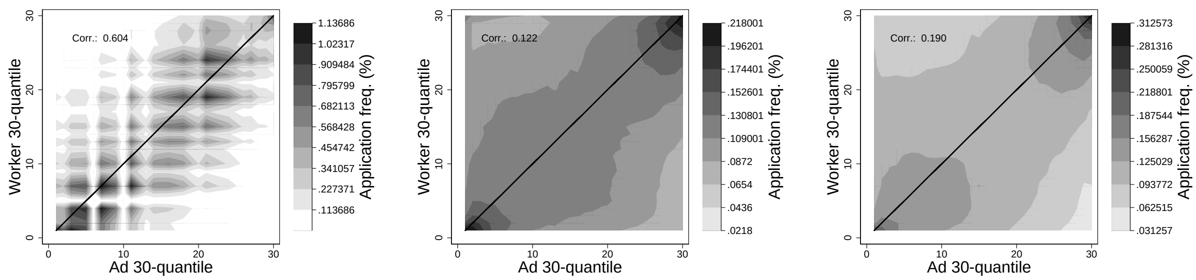
that of ES types (0.12 and 0.19, respectively). Whether we define firm identity as a job's attribute or not seems to be not empirically relevant for the degree of sorting. This is evidence that sorting at the worker-job level represents a relevant margin, given that we find sorting even after controlling for the effect of the firm.

The fact that there is a sizable difference in the strength of sorting between ES and eAKM types is a natural consequence of controlling for average observable characteristics of agents linked through an application to a counterpart on the other side of the market. Part of the anticipated wages depends on the "likely acceptable" matches they can find on the other side, as defined in the previous section. Therefore, taking away the variation of attributes of likely matches, which also guide application decisions, is likely to decrease the alignment of types.

To assess the importance in sorting of anticipated likely acceptable partners, we pose a simple correlation comparison: our ES sorting estimate is roughly 0.6 and the eAKM-f is 0.19. The ES estimate reflects the correlation of anticipated wages considering all factors: fixed and time-variant, observed or not. The eAKM-f estimate reflects the correlation of time-invariant (either observed and unobserved) characteristics. Then the difference between these estimates must reflect the contribution of time-variant aspects that are mostly unobservable in our data, which accounts for $(0.6 - 0.19)/0.6 = 68\%$ of the observed correlation.

Setting aside the discussion about whether AKM or our eAKM types allow to appropriately measure sorting, their use is a standard way to empirically quantify this feature of labor markets. Hence, our version of AKM types can be compared to other studies in the literature as a basic assessment of how our results on ES types would look like in other settings. The correlation of eAKM types obtained here are on the upper bound of estimates found in the literature. For example, Lopes de Melo (2018) (see table 1 of the paper) summarizes results for estimates using data from the US, France, Germany, Italy, Denmark and Brazil: all estimates are either close to zero or even negative, as is the case for France. Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler (2020) in their Table D1 also report estimates ranging from -0.24 to 0.23. One of the issues in this literature is that low job-to-job mobility of workers biases types' correlations to zero, as explained in Andrews, Bradley, Stott, and Upward (2008), Lamadon, Mogstad, and Setzler (2019), and many others. To the extent that individuals manifest their interest in potential jobs in our data, our eAKM results probably attenuate the low mobility problem. Therefore, obtaining an estimate in the upper part of the range in the literature should not be a surprise. However, a strict comparison between our setting and matched employer-employee data used in the literature is difficult because job seekers are already self-selected workers with relatively high likelihood of moving.

Robustness. We study sorting in different subsamples. Our results for the strength of the types correlation are robust. First, we divide our sample between applications made by unemployed and employed job seekers. By doing so, the correlation of the two subsamples goes down compared



(a) ES (b) eAKM (c) eAKM-f

Figure 2: Frequencies of applications, by worker and ad percentiles of ES, ex ante AKM and ex ante AKM (with firm fixed effect) types. Sample consists of 39,371,368 applications between 1-Jan-2010 and 31-Dec-2019 in `www.trabajando.com`

to the benchmark of 0.6 obtained for the whole sample. Results in Table A5 show that ES types are higher for unemployed than for employed, but this is reversed when we consider eAKM types (sorting is stronger for the employed seekers).

In addition, Table A5 also shows estimates for a subsample of applications that were made after the last CV update of the worker (set of most recent applications for each worker). As expected, the correlation between types is larger in this group (0.658) because of less measurement error (albeit, in sacrifice of sample size): our full sample considers old applications between job ads and workers for whom we might observe the *innacurate* information. The same result is observed with eAKM and eAKM-f types. Nevertheless, there are caveats about using only applications made after the online CV is updated as a baseline result. Newly CV updates may generate selection bias in terms of composition of job seekers, because some demographic groups may update CVs more frequently. Moreover, this sample contains relatively fewer older applications, which may be problematic to assess the cyclical properties of sorting.

Table A6 contains correlation estimates for samples with varying degrees of visibility for job seekers. As mentioned in Section 2, workers and employers may choose to hide wages to the other side of the market. Therefore, it is interesting to investigate whether releasing information spurs or dampens assortative matching in applications. The upper panels show that the ES correlation mildly decreases when splitting applications between ads with and without explicit wages. As explicit wage posting is consistently related to lower wages and requirements (Banfi and Villena-Roldán, 2019), we should expect lower correlation when conditioning on such a variable. When we split applications by the visibility of worker wage expectations, we observe that correlations remain at roughly the same level as in the baseline scenario. Therefore, we conclude that applicants have correct expectations or private information of wages, regardless the fact they are explicitly posted online or not.

5 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:

$$y_{w,t} = \alpha + \rho\, y_{a,t} + \delta\, y_{a,t} z_t + \phi z_t + X_t \lambda + \epsilon_{a,w,t} \tag{6}$$

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 (either the ES, eAKM or eAKM-f types) and z_t is a variable capturing aggregate economic conditions at monthly frequency (time t). These variables are standardized, so their mean is zero and their standard deviation is one. The specification also includes monthly seasonal dummies and a linear trend to capture long run movements in both types and the cyclical variable.

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.[13] During our sample period (2010-2019), the Chilean economy experienced an economic recovery from the 2009 global financial crisis, a period of low and stable unemployment roughly between 2014 and 2016 and an uptick in joblessness by the end of the sample, all of which produce a fair amount of variation in the considered aggregate conditions.

To properly interpret these regressions as evidence of cyclical 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/or 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 2017, the year with an unemployment rate (6.93%) closest to the sample average (6.92%). We run a probit model estimating the probability of being part of the 2017 sample as a function of observables on the applicant side X_w (gender, cubic age profiles, marital status categories, region of residence, and educational level categories) and on the job ad side X_a (firm industry, firm size, region, required educational area and educational level categories, and required minimum experience). We then compute a predicted probability and define a weight for a worker w and a job a pair in time t as

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

[13] Our source is ENE. We use the official updated version frequency weights from the 2017 Census.

Table 4: Cyclical assortative matching, constant composition

	ES	eAKM	eAKM-f
ρ	0.618***	0.120***	0.195***
	(0.014)	(0.018)	(0.018)
δ	-0.007***	-0.002***	0.000
	(0.012)	(0.016)	(0.016)
Obs	39,049,242	39,039,301	39,039,179
R^2	0.400	0.044	0.051
Δ_c	-0.0541	-0.0125	0.0000

Note: 100X Robust standard error in parenthesis. We report mean correlation ρ and cyclical sensitivity δ as defined by equation (6). The cyclical measure is the Chilean non-seasonally adjusted unemployment rate reweighed according to the 2017 Census, as published by Instituto Nacional de Estadísticas. Regressions use DFL weights (see the main text)

where $\Phi(\cdot)$ stands for the cumulative density of a standard normal distribution and $\widehat{\pi}$ represent estimates. To appropriately implement this method, we consider that being in 2017 to be a treatment, and its probability to be a propensity score for treated (2017) and non-treated (not 2017) groups. In Appendix A4, Figure A2 we depict kernel estimates of these propensity scores densities. Due to the common support assumption (i.e, each observation must have a non-zero probability of being in both groups), we trim a very few number of observations that have extreme probabilities of being observed in 2017.

Table 4 shows the estimates for ρ and δ, when we consider ES, eAKM and eAKM-f types. The estimated ES type correlation when the unemployment rate is at its average is 0.615, very close to the cross-section correlation in Section 4. The table also shows that all estimates of ρ are negative, which indicates that sorting is *procyclical*: when aggregate business cycle conditions improve (and unemployment decreases), sorting at the application level increases, and vice versa.

To gauge the variability of sorting over the cycle, we compute $\Delta_c = \delta(z^*_{max} - z^*_{min})$ where z^* is the unemployment rate which has been partialled out from time effects (using linear trend and monthly seasonal dummies). From the table, we see that $\Delta_c = -0.054$ for ES types, which implies that the average correlation at the nadir of the cycle (maximum unemployment rate) is 0.054 lower than the average correlation at the peak of the cycle (minimum unemployment rate). As shown in the previous section, the average correlations of eAKM and eAKM-f types are substantially lower. On the other hand, cyclicality of sorting is relatively more important for eAKM compared to ES types, since the value of Δ_c relative to that of ρ is higher in the earlier case: $0.10 = 0.0125/0.120$ versus $0.087 = 0.0541/0.618$, respectively. Finally, we find that cyclicality of sorting according to the eAKM-f measure is negligible.

Robustness. In Appendix A4, Table A7 we present a more flexible specification with year fixed effects. Although the latter may absorb some of the cyclical variation of z, results are robust. In the same Table we also report estimates without DFL weights, which shows that changing composition of the sample (due to aggregate conditions in labor markets) by itself is not substantially driving the results. This also shows that the procyclicality of sorting is genuinely a behavioral change of agents, and not a compositional change of ads or workers.

In Table A8 we also report results for three subsamples of interest: unemployed job seekers, employed job seekers, and applications done after the last CV update. As in the previous section, the average ES correlation is not far from the benchmark baseline of 0.6. Both employed and unemployed ES sorting is procyclical, but the pattern is greater for employed job seekers. To summarize this result, the statistic $\Delta_c = -0.072$ (for employed) shows that a sizable change in sorting among the employed job seekers is due to cyclical fluctuations, while the number for the unemployed is $\Delta_c = -0.0484$. These findings suggest that the observed procyclicality for the entire market is particularly driven by the application behavior of on-the-job seekers. Ex ante AKM types, separated by employment status, exhibit procyclical sorting as well, with the employed sample having a smaller estimate of δ.

Sorting is acyclical in practice for eAKM-f types. To the extent that in eAKM-f types we take away the firm specific component, the aforementioned finding means that the pure job title / occupational component barely changes in response to business cycle conditions. In constrast, eAKM types are clearly procyclical. The two facts together suggest that the increase in positive assortative matching in response to expansions is mainly due to workers applying to ads with job titles similar to the ones they often have applied in the past, but in firms that are more aligned to their own types. In other words, within the same occupation or job title, high-wage workers apply more to high-wage firms in expansions.

In the last panel of Table A8 we show that average sorting for the post CV update sample is higher than in the baseline, reaching 0.644 when the unemployment rate is at its average. While sorting remains procyclical, the business cycle is not as important as in the other samples to drive assortativeness of applications. The degree of procyclicality is higher for eAKM types in this sample, while eAKM-f sorting remains acyclical.

In sum, studying how sorting evolves over the business cycle leads to two main results. First, positive assortative matching can be regarded as high in comparison to the literature, around 0.6, and the cyclical behavior of the labor market may affect this figure but not overturn it. Second, sorting in ES and eAKM types is robustly procyclical, while eAKM-f types are acyclical.

Note once more that this is procyclical sorting at the *pre-match* stage and may reflect sorting of the flow of new job positions rather than the stock (what the rest of the literature focuses on). We see our results as complement to current literature. For example, in Lise and Robin (2017) there are strong forces to produce pro-cyclical sorting. However, these forces apply mostly during the *post-match* stage: in their model, the employed have an easier time finding better matches during an expansion due to on-the-job search. On the other hand, higher aggregate productivity during an expansion, increases the acceptance regions of workers and firms, posing an additional force against sorting.

6 From applications to realized matches

Our analysis so far cannot be directly compared to the existing sorting literature which provides evidence based on administrative matched employer-employee information. Considering that sorting in the stock of worker-job matches can be due to ex ante and ex post channels, it is an empirical question which is more important. Nevertheless, our previous results are suggestive since they show that a high degree of sorting exists ex ante (searching and pre-match stage).

In this section, we use a representative Chilean employment survey (*Encuesta Nacional de Empleo*, henceforth ENE) to forecast unemployment-to-employment (U2E) and job-to-job (J2J) transitions in the Chilean economy, conditional on a number of observable characteristics.[14] We use these models to predict the likelihood that a given application (worker-job pair observed in the TC website) ends up as a hire. With these transition probabilities we create sample weights to recompute our sorting estimates from the previous sections. These weights give us an approximation of realized match sorting of the flow of new job creation, making them more comparable to other studies in the literature. Of course, we are unable to trace specific hirings out of applications since we do not observe who gets hired in TC, but we can infer the average degree of sorting generated after real matches form.

To be more concrete, we select a vector of covariates $\mathbf{X}$ which are present both in the online job board and in ENE. Using a probit model we estimate the probability of being hired from unemployment (U2E) and from another job (J2J) in the ENE dataset. For U2E transitions, we consider individuals who have exerted some job search effort in the past four weeks (a standard definition). For the J2J calculations, we identify job movers as individuals who report being employed for two consecutive quarters, but that their most recent employment duration is less than three months.

Variables in $\mathbf{X}$ related to the worker consist of gender, age, educational level, and region[15] of residence. For the firm side, we have information on industry and firm size. We also use year and month dummies to control for time, business cycle and seasonal effects. After estimating the models using ENE data, we forecast the likelihood of each application becoming an actual hire for each worker-job pairs observed in TC. In the Appendix, Figure A3 shows kernel estimates for these predicted probabilities for applicants in `trabajando.com`. Equation (7) shows the formula we use for the imputation procedure

[14] The ENE is a quarterly rotating panel data, akin to the Current Population Survey (CPS) in the US. We report average transition rates for different groups in Table A9 in the appendix.

[15] Geographic administrative divisions of Chile, akin to US states

$$\text{Prob}\,(\text{Hire}|\mathbf{X}) = \mathbb{I}[\text{Employed}]J2J(\mathbf{X}) + (1 - \mathbb{I}[\text{Employed}])U2E(\mathbf{X}) \tag{7}$$

where observables $\mathbf{X}$ on the right hand side of the equation are from TC applications, $\mathbb{I}[\cdot]$ represents an indicator function for being an employed user of TC and $J2J$ and $U2E$ represent the predicted probabilities from models estimated using ENE data.

We also estimate a model to forecast employment-to-unemployment (E2U) transitions using roughly the same covariates we describe above, except for the omission of firm size.[16] With this information, we can forecast the likelihood of a job separation given covariates. Our ultimate goal here is to approximate a long-run employment probability under the assumption that transition rates remain stable over time. Using standard dynamic equations that are common in search models in steady state, we approximate the long-run probability of being employed by

$$\text{Prob}(\text{Long-run Employment}|\mathbf{X}) = \frac{U2E(\mathbf{X})}{U2E(\mathbf{X}) + E2U(\mathbf{X})\,(1 - J2J(\mathbf{X}))} \tag{8}$$

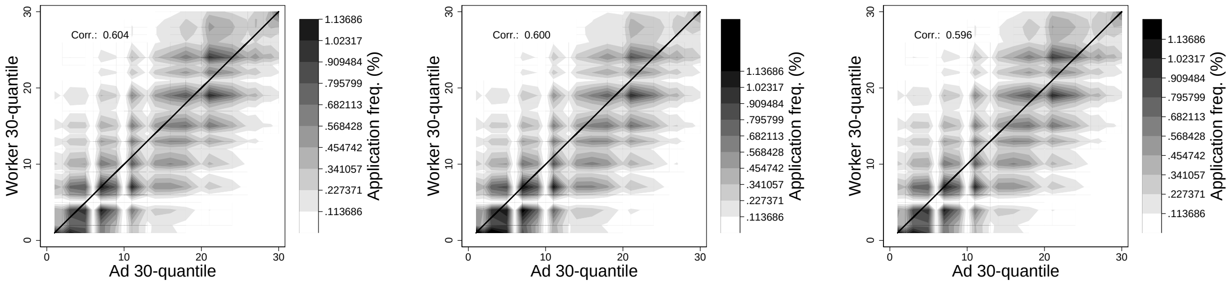
in which we have assumed that a job-to-job transition precludes a job separation (E2U transition) within a quarter. The distribution of predicted hiring and long run employment probabilities are shown in Figure A3 in the Appendix.

This approach has some limitations, as the hiring process may also depend on a number of factors that are not observed in ENE data, such as job requirements and their fit with the applicant's profile. These estimations have a predictive purpose and do not try to uncover any causal relationship between applicant and firm traits on hiring probabilities. Covariates in these predictive equations have a reasonably high explanatory power for transitions. In the appendix we present a summary of estimated equations in Table A10 while Figure A4 depicts the predicted average probability by age and educational level. In general, mobility decreases with age and education.

With these predicted probabilities, we redo the sorting analysis in preceding sections to assess the degree of assortative matching in a sample that is closer to the actual flow of matches (using hiring probabilities in eq. 7) and in the employment stock (using long-run employment probabilities in eq. 8). As in previous sections, we also study how these features evolve over the business cycle.

From Figure 3 we observe the original pre-match correlation, in subfigure (a). Subfigure (b) depicts the empirical joint probability distribution of types when weighted by the hiring probability. In subfigure (c) we observe the same distribution, but weighted by the long-run probability of employment. The shapes look remarkably similar with slight variations in areas near the median of ad types and low and high types of workers. The computed correlations for ES types vary only marginally and remain around 0.6. As the overall sorting does not change when we perform the reweighting procedure, we conclude that the application or search stage is, by far, the most

[16] For the unemployed we use information regarding the sector of their last employer. Nevertheless, the questionnaire in ENE does not include information regarding the size of the last employer.



(a) ES pre-match (b) ES hiring (c) ES long-run employment

Figure 3: Frequencies of applications, by worker and ad percentiles of ES types reweighted by the estimated probability given observables. Details of the computation of predicted probabilities are in the text. Sample consists on 39,371,368 applications between 1-Jan-2010 and 31-Dec-2019 in `www.trabajando.com`

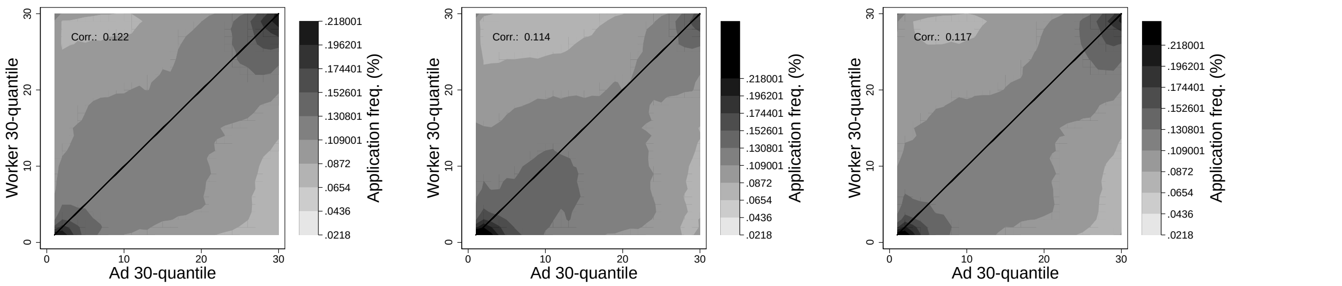
important stage in the labor market to generate the observed ex post sorting. Although the point estimates suggest that the hiring stage marginally decreases sorting, we believe that such a conclusion would be far-fetched as there is considerably sampling uncertainty in the transition probability imputation.[17]

Figure 4 shows a marginal decline in sorting of eAKM types. While the correlation of types remain quite stable nearly 0.12, the shape of the joint distribution of types slightly changes between application and hiring stages as there is a larger alignment for lower types, reflected in a darker area at the bottom-left of subfigure (b), and also an enlargement of the area for average joint frequencies. Subfigure (c) looks similar to (a) indicating a marginal increase of sorting in the long term. Despite these differences, there is no economically relevant variation between these subfigures.

Results in Figure 5 are qualitatively similar to those in Figures 3 and 4. Despite some changes in the shape of the joint distribution of types at the pre-match and hiring stages, especially for low types, the correlations remain very similar and close to 0.18. Again, combining evidence from ES, eAKM and eAKM-f types leads us to conclude that sorting is generated at the application stage to a great extent.

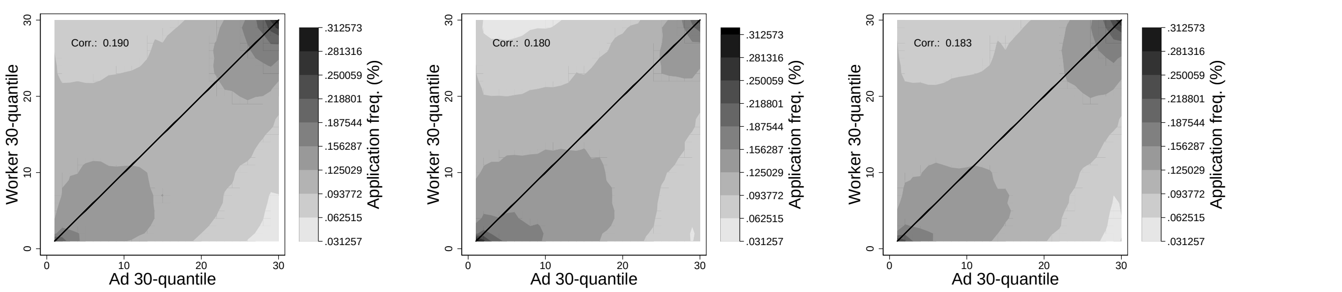
Our ex post approximation to AKM (reweighted eAKM types) can be compared to types defined by Borovickova and Shimer (2020). These authors show that the correlation of AKM types is able to recover the true strength of sorting only if the data generating process is truly AKM. In contrast, these authors find that the correlation of ex post ES types (using the jargon of our paper) captures the true sorting under a wider set of circumstances, including when AKM is the true underlying model. Therefore, the correlation gap between the two measures, ex post ES and ex post AKM (by

[17]While unfeasible in this paper due to computational constraints, we envision a jointly bootstrap procedure for both the TC and ENE samples to generate an empirical confidence interval of the ratio between the ex ante correlation and the ex post correlation. Nevertheless, the strength and robustness of the ex ante correlation makes it very hard to overturn its relevance in the sorting process.



(a) eAKM pre-match (b) eAKM hiring (c) eAKM long run employment

Figure 4: Frequencies of applications, by worker and ad percentiles of ex ante AKM types reweighted by the estimated probability given observables. Details of the computation of predicted probabilities are in the text. Sample consists on 39,371,368 applications between 1-Jan-2010 and 31-Dec-2019 in `www.trabajando.com`



(a) eAKM-f pre-match (b) eAKM-f hiring (c) eAKM-f long run employment

Figure 5: Frequencies of applications, by worker and ad percentiles of ex ante AKM (with firm fixed effect) types reweighted by the estimated probability given observables. Details of the computation of predicted probabilities are in the text. Sample consists on 39,371,368 applications between 1-Jan-2010 and 31-Dec-2019 in `www.trabajando.com`

means of reweighting), suggests that the AKM setup misrepresents the true sorting in the data.

Table 5 shows the cyclical correlation obtained when weighting observations by the hiring and the long-run employment probabilities. For ES types, the hiring probability weighting implies somewhat lower procyclicality of sorting although it remains significant. This is reflected in the comparison of Δ_c in both estimations: -0.0541 in the pre-match (table 4) versus -0.0392 at the hiring stage, which refers to the maximum variation of types correlation attributable to business cycle fluctuations in the sample. In contrast, the long-term weighted correlation shows a slightly higher procyclicality of sorting than the ex ante case discussed in Section 5.

For eAKM types, results show the reverse pattern. Sorting weighted by the hiring probability exhibits larger procyclicality, while long-term employment exhibits a lower value. Both remain significant. In the case of eAKM-f types, sorting turns out to be acyclical when weighted as it is in the ex ante stage.

Table 5: Cyclical assortative matching, constant composition, ex post reweigh

	Hired prob weight			Employment prob weight		
	ES	eAKM	eAKM-f	ES	eAKM	eAKM-f
ρ	0.611***	0.109***	0.182***	0.612***	0.116***	0.190***
	(0.015)	(0.020)	(0.020)	(0.014)	(0.018)	(0.018)
δ	-0.005***	-0.003***	0.000	-0.009***	-0.001***	0.000
	(0.013)	(0.017)	(0.017)	(0.013)	(0.016)	(0.016)
Obs	39,026,649	39,016,709	39,016,587	39,026,649	39,016,709	39,016,587
R^2	0.392	0.043	0.047	0.395	0.047	0.051
Δ_c	-0.0392	-0.0216	0.0021	-0.0601	-0.0062	0.0004

Note: 100X robust standard errors in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Robustness: We estimate ex post sorting on additional subsamples of interest: unemployed and on-the-job seekers. We place in the appendix the Table A11 and Figures A6 and A7 in which we report the weighted correlations and bivariate empirical distributions of types weighted to approximate the sorting patterns of the flow and stock of realized matches. These estimates ought to be compared to those in Table A5. For the unemployed, the ex ante sorting of ES types (0.587) slightly decreases when weighted by hiring (0.575) and long-term employment probabilities (0.581). For on-the-job seekers, the sorting of ES types for the recently hired remains almost the same with respect to the the pre-match stage (0.565 vs 0.567), but decreases to 0.557 when correlation is computed using long-term employment probabilities.

For eAKM and eAKM-f types both ex post weights marginally decrease sorting (for unemployed and employed seekers alike). These estimates point in the same direction as the ones obtained for the whole sample. From the application stage no much additional sorting is gained, if any, in realized matches. As Faberman, Mueller, Şahin, and Topa (2019) have shown, on-the-job search may be quite different from unemployed search in terms of effort and outcomes. Thus, we should expect notorious differences in realized matches for these two groups. Although we observe little effect of hiring and long-term employment probability weighting in sorting for the whole sample, it could be possible to observe different behavior because J2J and U2E flows are estimated separately. Even so, we find that splitting the sample by labor force status barely affects the sorting of realized matches. Hence, we conclude that our evidence for the whole sample is not driven by labor force status.

Another concern we address is that the long-run employment probability formula assumes that a potential new match in another job seeking spell will have the same observables of the current application. Another approach is to weight applications by the expected duration given the observables. Assuming that the probability given worker-job observables remains constant over time and considering that a job separation may occur due to a layoff or to a job-to-job move, the expected duration is

$$\text{E(Employment Duration}|\mathbf{X}) = \frac{1}{E2U(\mathbf{X}) + J2J(\mathbf{X})}.$$

Using employment duration weights, we redo our previous exercises and obtain very similar results. In the appendix Figure A5 depicts the joint distribution of types weighted by the long-run employment and the expected duration formulas. The second method generates slightly higher sorting estimates, so the overall picture remains unchanged.

7 Discussion

While the online job board data in TC help us study sorting patterns in the labor market, strictly speaking, there is no model in the literature containing all key elements in our empirical framework: we observe job seekers and workers who are heterogeneous in several dimensions, while at the same time, there are multiple applications sent and received by job seekers and job positions respectively. Add the fact that job positions may be hiring several candidates, the exact equilibrium mechanism leading to a *sorted* stock of matches is unknown. However, in this section we link our exercise to different papers in the literature and assess how our results could be useful to further understand sorting in labor markets.

7.1 Relation to random search models

ES types can be useful to study sorting in a variety of theoretical settings, as they reflect an expected salary regardless of potential matches. First, for random search structural models along the lines of Shimer and Smith (2000) and Card, Cardoso, Heining, and Kline (2018) as well as for the canonical statistical AKM framework, the article by Borovickova and Shimer (2020) proposes a type measurement closely matching the wage expectations from our online job board data. In particular, they suggest to use as the worker's type "the expected log wage she receives in an employment relationship conditional on taking the job" and for the firm type, "the expected log wage that it pays to an employee conditional on hiring the worker". They also suggest that the degree of assortative matching can be inferred by the correlation of these expectations. Our ES type and those proposed by Borovickova and Shimer (2020) are similar in spirit, although there is a contrast between the ex ante nature of the ES definition and the ex post nature of the definition in Borovickova and Shimer (2020). The former defines the expected wage conditional on a likely mutual acceptance, whereas the latter hinges on an expectation definition conditional in hiring. Moreover, an ES type is a log of expected salary while Borovickova and Shimer (2020) type is a expected log wage.

The literature has also evolved to study sorting in dynamic cyclical settings. Two leading examples are Lise and Robin (2017) and Bagger and Lentz (2018), which rely on random search and ex post mutual acceptance as the key mechanism leading to sorting. In contrast, our data allows us to study labor market sorting in a *pre-match* stage. We see our evidence as complementary to these studies. Assuming that no one would apply to a job whose probability of match is zero, our notion of ex ante sorting based on applications defines a set containing all mutually acceptable matches, which is relevant to this class of models.

7.2 Relation to directed search models

ES types also make sense for a directed search framework in that activity inside an online job board can be rationalized naturally using the competitive search framework, along the lines of Moen (1997): firms post vacancies independently of each other, which can lead to coordination frictions of potential applicants. These online job advertisements, along the wealth of information they provide about working conditions, may be taken as particular *sub-markets* designed to attract appropriate applicants. At any point in time (after positions are posted), job seekers observe all relevant job ads and decide where to apply, choosing job ads which maximize the expected value provided to them.

There are theoretical reasons to believe that posted wages are a credible measurement for realized wages in addition to the empirical regularities portrayed in Section 3. In a competitive auction setting, akin to competitive search, Kim and Kircher (2015) show that auctioneers' posted prices (wages) truthfully reveal their types although their announcements are essentially cheap talk. On the other hand, there is empirical evidence that high-wage ads attract more applicants (all else equal), which clearly indicates competitive search behavior as noted by Banfi and Villena-Roldán (2019) and Marinescu and Wolthoff (2020).

ES types could be useful to understand both random and directed search environments. However, we find that the ex ante sorting is strong and the ex post approximation does not increase types' correlations. Therefore, theoretical models relying on the role of posted wages and information driving allocations seem more appropriate to explain our evidence. For instance, the model in Shimer (2005) portrays competitive search paired with heterogeneity of workers and jobs. Under assumptions of monotonicity and supermodularity, he finds that in the competitive equilibrium firms hire the most productive type of worker, decentralizing the social optimum. Posted wages (akin to job ES types in our setup) are increasing in job productivity. However, the same cannot be said about expected wages for workers without imposing restrictions on the exogenous distribution of worker and job types. In another contribution, the model of Eeckhout and Kircher (2010) has also heterogeneity on both sides of the market and directed search, i.e. agents are allowed to seek the wages and qualities they prefer in the market. Agents narrow down their search to the more suitable partners, reducing the chances of match failure in comparison to the random search case. Therefore, the complementarity required for positive assortative matching is weaker than in markets with random search.

Recent papers add relevant elements for sorting into the picture. Cai, Gautier, and Wolthoff (2021) introduce simultaneous meetings with screening technology (interviews, tests, etc) on top of a directed search setup. The model distinguishes between sorting in contacting and in matching. It also shows that lower screening capacity eases positive sorting, conferring more importance in determining allocations to the application stage. Yet another contribution is Bartolucci and Monzón (2019) who keep the random search and ex post bargaining setup, but introduces on-the-job search, a key empirical feature in our data. They conclude that search frictions and on-the-job search create an endogenous preference for high-type partners even if the production technology is submodular.

Although these models capture essential aspects we see in the data, the implications of these features altogether in sorting are far from trivial. Moreover, they cannot fully account simultaneously for several essential facts of online job boards (and the labor market more generally): workers sending multiple simultaneous applications, employer screening, on-the-job search, not to mention the lack of dynamics to study business cycle effects. Hence, it is a formidable challenge to infer modularity properties of the production function from the sorting patterns we empirically observe, as a sequel of theoretical papers have done since Becker (1973) and Shimer and Smith (2000). More research is needed to better understand sorting taking into account these facts altogether.

8 Conclusions

In this paper we revisit the question of whether workers and job positions are sorted in any meaningful way in the labor market. The vast majority of the literature concerns about the sorting of realized matches, especially in matched employer-employee administrative datasets. However, there is little direct evidence on how these allocations are generated, and especially, an assessment of the role applications play in the sorting process. Online job board data is particularly well-suited to make several contributions along these lines.

We show a high correlation (nearly 0.6) between expected salary (ES) types, which is a meaningful indication of ex ante positive assortative matching, or following Cai, Gautier, and Wolthoff (2021), sorting in *contacts*. The conceptualization of types as expectations is similar to Borovickova and Shimer (2020), although their notion applies to ex post matching rather than the pre-match definition we use. Our results also show that this correlation increases when unemployment decreases, i.e., the positive assortative matching is procyclical.

We approximate ex post sorting with our expected salary types using transition probabilities estimated from an external employment survey (ENE) in order to make our results comparable to the standard literature following Abowd, Kramarz, and Margolis (1999). We find correlations similar to the one obtained at the application stage, from which we conclude that sorting is mainly produced by application decisions, with other selection mechanisms (such as screening, hiring, and layoffs) having a secondary role.

We also discuss how the types we propose have several advantages to what is used in the literature: expected salary types reveal sorting patterns at the worker-job level, even after controlling for hiring firm characteristics. This is in line with novel research which dissociates jobs from firms (breaking the standard "one firm equals one job" assumption), as in Guvenen, Kuruscu, Tanaka, and Wiczer (2020); Taber and Vejlin (2020) among others. Moreover, our ex ante measure of sorting relies on expectations that are declared before meeting any partner, so they are free from ex post compensation, which obscures the relationship between wages and productivity as pointed out by Eeckhout and Kircher (2011); Hagedorn, Law, and Manovskii (2017). Additionally, our estimations are free from low mobility bias which is pervasive in matched employer-employee databases (Andrews, Bradley, Stott, and Upward, 2008; Bonhomme, Lamadon, and Manresa, 2019).

Since our results can be interpreted as directed search behavior, theoretical settings along the lines of Shimer (2005) and Abowd, Kramarz, Pérez-Duarte, and Schmutte (2018), in which directed search is a key ingredient, are a desirable path for future research in this area.

References

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

Abowd, J. M., F. Kramarz, S. Pérez-Duarte, and I. M. Schmutte (2018): "Sorting Between and Within Industries: A Testable Model of Assortative Matching," *Annals of Economics and Statistics*, (129), 1–32.

Andrews, M. J., S. Bradley, D. Stott, and R. Upward (2008): "Successful Employer Search? An Empirical Analysis of Vacancy Duration Using Micro Data," *Economica*, 75(299), 455–480.

Bagger, J., and R. Lentz (2018): "An Empirical Model of Wage Dispersion with Sorting," *The Review of Economic Studies*, 86(1), 153–190.

Baley, I., A. Figueiredo, and R. Ulbricht (2020): "Mismatch Cycles," Working paper.

Banfi, S., and B. Villena-Roldán (2019): "Do High-Wage Jobs Attract more Applicants? Directed Search Evidence from the Online Labor Market," *Journal of Labor Economics*, 37(3), 715–746.

Bartolucci, C., and I. Monzón (2019): "Frictions lead to sorting: a partnership model with on-the-match search," *Available at SSRN 3393545*.

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

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2020): "How Much Should we Trust Estimates of Firm Effects and Worker Sorting?," Working Paper 27368, National Bureau of Economic Research.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87(3), 699–739.

Borovickova, K., and R. Shimer (2020): "High Wage Workers Work for High Wage Firms," Unpublished manuscript.

Cai, X., P. Gautier, and R. Wolthoff (2021): "Search, Screening and Sorting," *Unpublished Manuscript*.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference With Multiway Clustering," *Journal of Business & Economic Statistics*, 29(2), 238–249.

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.

Crane, L., H. Hyatt, and S. Murray (2018): "Cyclical Labor Market Sorting," Discussion paper.

Cubas, G., and P. Silos (2020): "Social Insurance and Occupational Mobility," *International Economic Review*, 61(1), 219–240.

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.

Dvorkin, M. A., and A. Monge-Naranjo (2019): "Occupation Mobility, Human Capital and the Aggregate Consequences of Task-Biased Innovations," *FRB St. Louis Working Paper*, (2019-13).

Eeckhout, J., and P. Kircher (2010): "Sorting and Decentralized Price Competition," *Econometrica*, 78(2), 539–574.

——— (2011): "Identifying Sorting–In Theory," *The Review of Economic Studies*, 78(3), 872–906.

Faberman, J., A. Mueller, A. Şahin, and G. Topa (2019): "Job Search Behavior among the Employed and Non-Employed," Working paper, Federal Reserve Bank of New York.

Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2020): "Multidimensional Skill Mismatch," *American Economic Journal: Macroeconomics*, 12(1), 210–44.

Hagedorn, M., T. H. Law, and I. Manovskii (2017): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85(1), 29–65.

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.

Kim, K., and P. Kircher (2015): "Efficient competition through cheap talk: the case of competing auctions," *Econometrica*, 83(5), 1849–1875.

Kudlyak, M., D. Lkhagvasuren, and R. Sysuyev (2013): "Systematic Job Search: New Evidence from Individual Job Application Data," mimeo, Federal Reserve Bank of Richmond.

Lamadon, T., M. Mogstad, and B. Setzler (2019): "Imperfect competition, compensating differentials and rent sharing in the US labor market," Discussion paper, National Bureau of Economic Research.

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., 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., and R. Wolthoff (2020): "Opening the Black Box of the Matching Function: The Power of Words," *Journal of Labor Economics*, 38(2), 535–568.

Michelacci, C., and J. Suarez (2006): "Incomplete Wage Posting," *Journal of Political Economy*, 114(6), 1098–1123.

Moen, E. R. (1997): "Competitive Search Equilibrium," *The Journal of Political Economy*, 105(2), 385–411.

Shimer, R. (2005): "The assignment of workers to jobs in an economy with coordination frictions," *Journal of Political Economy*, 113(5), 996–1025.

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

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019): "Firming up inequality," *The Quarterly journal of economics*, 134(1), 1–50.

Taber, C., and R. Vejlin (2020): "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," *Econometrica*, 88(3), 1031–1069.

White, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, pp. 817–838.

Appendix: Sorting on-line and on-time
For on-line publication only

A1 Additional Descriptive Statistics

Table A1: Characteristics of Job Seekers

	Employed	Unemployed	Total
Gender (%)			
Females	41.60	49.30	45.75
Males	58.40	50.70	54.24
Marital status (%)			
Married	30.73	25.43	27.87
Partner	2.08	2.24	2.17
Divorced	2.49	2.53	2.51
Separated	2.28	2.45	2.38
Single	62.22	67.05	64.83
Widow	0.20	0.29	0.25
Nationality (%)			
Argentina	0.50	0.38	0.44
Bolivia	0.14	0.18	0.16
Brazil	0.18	0.17	0.17
Chile	89.22	87.65	88.36
Colombia	0.84	1.21	1.04
Ecuator	0.16	0.18	0.17
U.S.A.	0.05	0.03	0.04
Spain	0.32	0.30	0.31
France	0.05	0.04	0.05
Italy	0.08	0.04	0.06
Mexico	0.07	0.06	0.07
Other	1.88	2.04	1.97
Peru	0.54	0.68	0.61
Uruguay	0.08	0.06	0.07
Venezuela	5.91	6.97	6.48
Region of Residence (%)			
I Tarapaca	1.48	1.57	1.53
II Antogafasta	4.10	4.09	4.09
III Atacama	1.47	1.34	1.40
IV Coquimbo	2.79	3.25	3.04
V Valparaiso	8.60	10.05	9.38
VI OHiggins	3.11	3.29	3.21
VII Maule	2.76	2.91	2.84
VIII Bio Bio	6.11	6.78	6.47
IX Araucania	2.34	2.85	2.62
X Los Lagos	2.46	2.71	2.59
XI Aysen	0.26	0.25	0.26
XII Magallanes	0.54	0.48	0.51
RM Metropolitana	59.82	55.94	57.72
XIV Los Rios	1.11	1.31	1.22
XV Arica y Parinacota	0.87	1.11	1.00
Foreigners	0.97	1.19	1.09
Observations	775,960	909,889	1,687,698

Table A2: Characteristics of Job Seekers

	Employed	Unemployed	Total
Days job searching (%)			
1 day	22.80	26.76	24.96
2-10 days	2.90	5.00	4.04
11 - 25 days	2.49	4.18	3.40
26 - 60 days	3.63	5.75	4.77
61 - 90 days	2.24	3.24	2.78
91 - 140 days	3.05	4.17	3.66
141 - 210 days	3.63	4.51	4.10
> 210 days	59.26	46.38	52.27
Job searching			
Days (Mean/(S.D.))	855.93	597.66	716.03
	(996.20)	(867.90)	(937.85)
Ads applied (%)			
1	16.74	17.87	17.37
2	9.44	10.54	10.04
3	6.62	7.53	7.11
4 - 6	12.89	14.45	13.73
7 - 10	10.13	11.01	10.60
11 - 20	13.65	13.81	13.73
21 - 30	7.41	6.91	7.14
Ads Applied			
Applications per worker (Mean/(S.D.))	26.37	20.74	23.32
	(46.34)	(39.05)	(42.64)
Applications to explicit wages per worker (Mean/(S.D.))	2.12	2.55	2.35
	(4.71)	(5.65)	(5.24)
Explicit Wages			
% Appl. WP/Total Appl. (Mean/(S.D.))	9.059	13.614	11.519
	(17.325)	(21.323)	(19.723)
Observations	775,960	909,889	1,687,698

Table A3: Characteristics of Job Ads

	Hidden wage	Explicit wage	Total
Applications per ad (%)			
1	9.95	14.99	10.76
2	7.06	10.52	7.62
3	5.48	8.04	5.89
4 - 5	8.19	11.49	8.72
6 - 10	13.20	16.17	13.68
11 - 20	15.09	14.59	15.01
21 - 50	20.44	14.12	19.41
> 50	20.60	10.09	18.90
Applications per ad			
Applications (Mean/(SD))	37.20	21.57	34.67
	(76.16)	(56.47)	(73.55)
Observations	951,222	183,858	1,135,080

A2 Analytical framework: statistics from Figure 1

Table A4: Wage gap statistics from Figure 1

	log difference of wages	
	$ES_a - ES_w$	last wage - ES_w
mean	-0.211	-0.053
std dev	0.576	0.409
skewness	-0.133	-0.132
kurtosis	4.352	10.883
perc 25	-0.547	-0.223
perc 50	-0.182	0.000
perc 75	0.134	0.118

A3 Sorting on line: additional analysis

Table A5: Cross-sectional assortative matching: Additional samples I

	All			Post CV update		
	ES	eAKM	eAKM-f	ES	eAKM	eAKM-f
ρ	0.604***	0.122***	0.190***	0.658***	0.141***	0.213***
	(0.013)	(0.017)	(0.017)	(0.045)	(0.061)	(0.061)
Obs	39,371,248	39,360,822	39,360,694	3,507,132	3,501,580	3,501,524
R^2	0.365	0.015	0.036	0.432	0.020	0.046

	Unemployed			Employed		
	ES	eAKM	eAKM-f	ES	eAKM	eAKM-f
ρ	0.587***	0.098***	0.160***	0.567***	0.104***	0.176***
	(0.020)	(0.025)	(0.025)	(0.018)	(0.023)	(0.023)
Obs	18,882,036	18,875,900	18,875,787	20,470,751	20,466,503	20,466,488
R^2	0.345	0.010	0.026	0.322	0.011	0.031

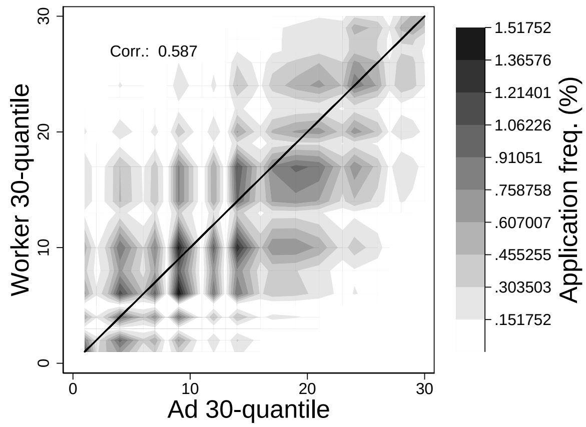
Note: 100X robust standard errors in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Table A6: Cross-sectional assortative matching: Additional samples II

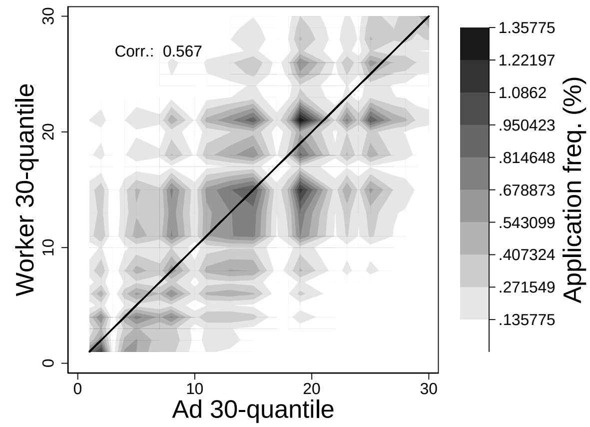
	Explicit wage ad			Hidden wage ad		
	ES	eAKM	eAKM-f	ES	eAKM	eAKM-f
ρ	0.555***	0.113***	0.163***	0.597***	0.120***	0.190***
	(0.410)	(0.213)	(0.274)	(0.116)	(0.077)	(0.103)
Obs	3,968,265	3,965,265	3,965,202	35,402,983	35,395,557	35,395,492
R^2	0.308	0.013	0.026	0.357	0.014	0.036

	Explicit wage worker			Hidden wage worker		
	ES	eAKM	eAKM-f	ES	eAKM	eAKM-f
ρ	0.617***	0.138***	0.192***	0.595***	0.109***	0.184***
	(0.110)	(0.085)	(0.101)	(0.113)	(0.077)	(0.103)
Obs	9,261,789	9,257,550	9,257,500	30,109,459	30,103,272	30,103,194
R^2	0.381	0.019	0.037	0.354	0.012	0.034

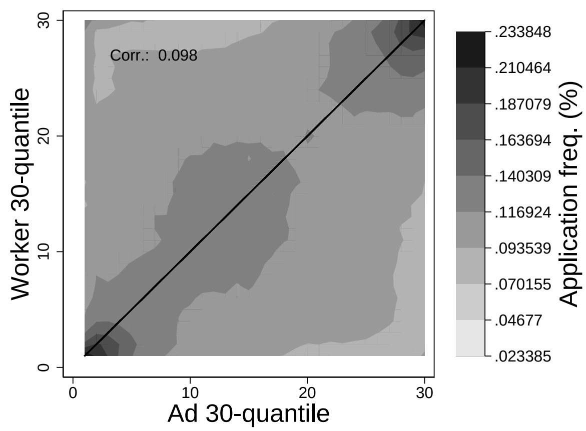
Note: 100X robust standard errors in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$



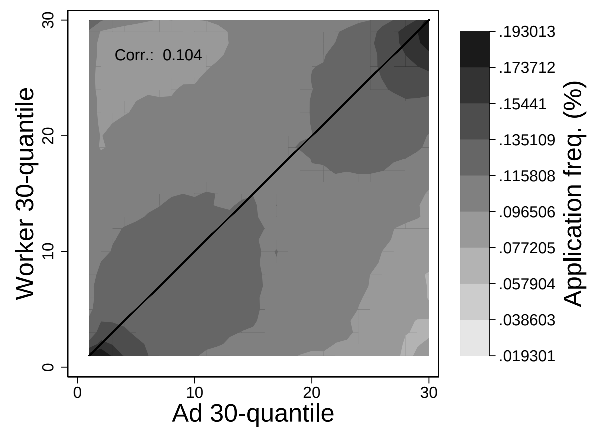
(a) Unemployed, ES types



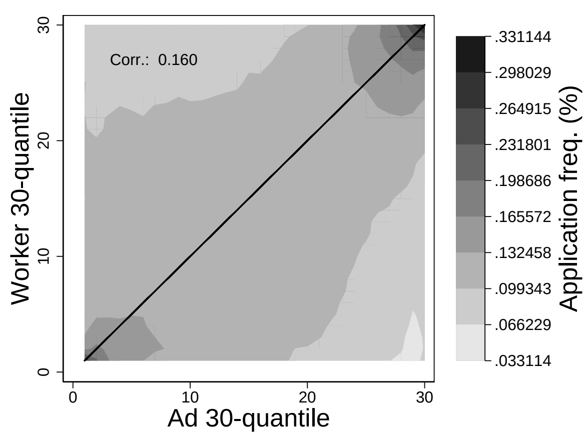
(b) Employed, ES types



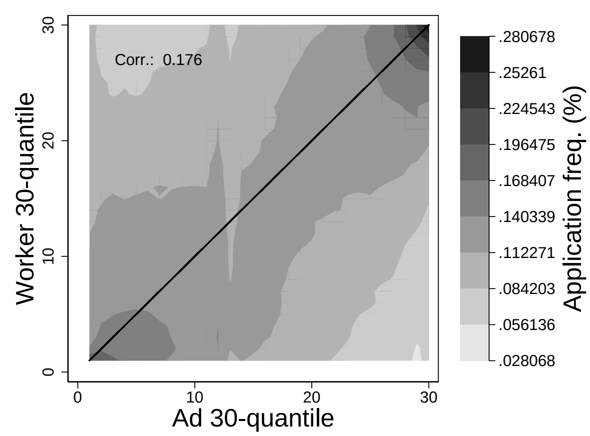
(c) Unemployed, eAKM types



(d) Employed, eAKM



(e) Unemployed, eAKM-f types



(f) Employed, eAKM-f types

Figure A1: Frequencies of applications, by worker and ad percentiles of ES, ex ante AKM and ex ante AKM (with firm fixed effect) types by employment status. All applications between 1-Jan-2010 and 31-Dec-2019 in `www.trabajando.com`.

A4 Sorting on time: additional analysis

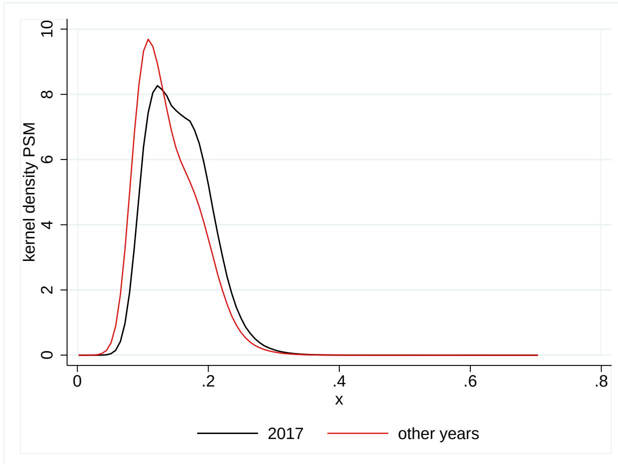


Figure A2: Estimated propensity score matching for the DiNardo, Fortin, and Lemieux (1996) reweighting compositional adjustment. Observations outside of the common support of the distribution are trimmed.

Table A7: Cyclical assortative matching: Additional samples I

	Year FE / DFL			no DFL		
	ES	eAKM	eAKM-f	ES	eAKM	eAKM-f
ρ	0.618***	0.119***	0.195***	0.623***	0.127***	0.197***
	(0.014)	(0.018)	(0.018)	-0.013	-0.017	-0.017
δ	-0.007***	-0.001***	0.000	-0.008***	-0.001***	-0.001***
	(0.012)	(0.016)	(0.016)	-0.012	-0.016	-0.015
Obs	39,049,242	39,039,301	39,039,179	39,066,742	39,056,692	39,056,570
R^2	0.400	0.046	0.051	0.399	0.046	0.051
Δ_c	-0.0585	-0.0100	0.0020	-0.0533	-0.0091	-0.0057

Notes: 100X Standard error in parenthesis. We report mean correlation ρ and cyclical sensitivity δ as defined by equation (6). The cyclical measure is the Chilean non-seasonally adjusted unemployment rate reweighted according to the 2017 Census, as published by the *Instituto Nacional de Estadísticas*. Regressions use DiNardo, Fortin, and Lemieux (1996) weights (see main text), which are computed using a probit model in which the dependent variable is an indicator for 2017, 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, and educational major. Independent variables for job ads are indicators for region, industry, economic activity (job board classification), educational area required, and firm size category.

Table A8: Cyclical assortative matching: Additional samples II

	Unemployed		
	ES	eAKM	eAKM-f
ρ	0.581***	0.079***	0.153***
	(0.022)	(0.026)	(0.027)
δ	-0.006***	-0.003***	0.001***
	(0.019)	(0.023)	(0.023)
Obs	18,756,023	18,750,091	18,749,980
R^2	0.371	0.045	0.039
Δ_c	-0.0484	-0.0239	0.0108

	Employed		
	ES	eAKM	eAKM-f
ρ	0.590***	0.111***	0.187***
	(0.019)	(0.025)	(0.025)
δ	-0.011***	0.000	-0.001***
	(0.017)	(0.022)	(0.022)
Obs	20,279,795	20,275,797	20,275,786
R^2	0.371	0.043	0.050
Δ_c	-0.0721	0.0024	-0.0049

	Post CV update		
	ES	eAKM	eAKM-f
ρ	0.644***	0.107***	0.197***
	(0.050)	(0.066)	(0.067)
δ	-0.002***	-0.004***	0.001
	(0.046)	(0.060)	(0.062)
Obs	3,486,726	3,481,384	3,481,328
R^2	0.428	0.033	0.046
Δ_c	-0.0188	-0.0416	0.0090

Notes: 100X Standard error in parenthesis. We report mean correlation ρ and cyclical sensitivity δ as defined by equation (6). The cyclical measure is the Chilean non-seasonally adjusted unemployment rate reweighted according to the 2017 Census, as published by the *Instituto Nacional de Estadísticas*. Regressions use DiNardo, Fortin, and Lemieux (1996) weights (see main text), which are computed using a probit model in which the dependent variable is an indicator for 2017, 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, and educational major. Independent variables for job ads are indicators for region, industry, economic activity (job board classification), educational area required, and firm size category.

A5 Sorting ex post: additional analysis

Description of Encuesta Nacional de Empleo (ENE): The ENE is the official employment survey in Chile, conducted by the *Instituto Nacional de Estadísticas* (INE) to produce official labor force statistics. It is a quarterly rotating panel survey in which urban households remain up to 6 quarters in the sample and rural ones, up to 12 quarters. There are unique person identifiers which allow any researcher to trace labor market transitions on a quarterly basis. Table A9 reports the average transition rates by gender, age groups, educational attainment, and region. The total number of transitions is reported at the end.

Table A9: Descriptive statistics *Encuesta Nacional de Empleo*

	J2J	U2E	E2U
Gender			
Females	3.9%	5.6%	5.3%
Males	7.8%	6.5%	4.0%
Age			
25 - 29	6.6%	9.5%	5.4%
30 - 34	6.6%	6.8%	4.7%
35 - 39	6.2%	5.8%	4.3%
40 - 44	5.8%	5.1%	4.4%
45 - 49	5.3%	4.5%	4.5%
>50	4.5%	3.9%	4.4%
Education level			
Primary (1-8 years)	8.9%	7.4%	6.3%
High School	6.6%	6.8%	5.1%
Tech. Tertiary Educ.	4.3%	5.0%	4.0%
College	3.7%	5.2%	3.4%
Post-Graduate	2.9%	2.9%	2.4%
Region			
I Tarapaca	5.0%	6.1%	4.7%
II Antofagasta	4.1%	5.0%	4.1%
III Atacama	5.7%	6.6%	4.7%
IV Coquimbo	7.2%	6.8%	4.7%
V Valparaiso	6.2%	5.9%	4.6%
VI O'Higgins	8.6%	7.4%	5.5%
VII Maule	8.7%	8.2%	5.5%
VIII Biobio	6.1%	6.3%	4.8%
IX La Araucania	7.0%	6.4%	5.6%
X Los Lagos	5.6%	5.2%	4.7%
XI Aysen	6.1%	6.3%	4.0%
XII Magallanes	4.2%	5.1%	3.4%
RM Metropolitana	4.9%	5.5%	4.3%
XIV Los Rios	6.4%	6.6%	5.1%
XV Arica y Parinacota	4.9%	5.3%	4.9%
Number of transitions	108,071	107,175	86,655
Observations		1,391,479	

Notes: Reported numbers are average quarterly transition rates by different categories. The time span is March 2010 to December 2019.

Table A10: Estimation quarterly transition probabilities *Encuesta Nacional de Empleo*

	(1) N2E	(2) J2J	(3) E2N
age	-0.378***	-0.256**	0.0274
	(0.106)	(0.119)	(0.0764)
age^2	0.00805***	0.00677**	-0.000695
	(0.00280)	(0.00309)	(0.00198)
age^3	-5.70e-05**	-5.89e-05**	5.59e-06
	(2.40e-05)	(2.61e-05)	(1.67e-05)
female	-0.400	-5.297***	1.549
	(1.656)	(1.826)	(1.268)
female × age	0.00414	0.451***	-0.107
	(0.134)	(0.146)	(0.102)
female × age^2	-2.69e-05	-0.0119***	0.00242
	(0.00352)	(0.00378)	(0.00264)
female × age^3	3.34e-07	0.000102***	-1.72e-05
	(3.01e-05)	(3.19e-05)	(2.22e-05)
educ - high school	-0.0695	-0.0722	-0.0412
	(0.0427)	(0.0491)	(0.0271)
educ - technical high schhol	-0.201***	-0.182***	-0.102***
	(0.0555)	(0.0572)	(0.0328)
educ - technical tertiary	-0.298***	-0.297***	-0.137***
	(0.0492)	(0.0565)	(0.0320)
educ - college	-0.323***	-0.366***	-0.246***
	(0.0469)	(0.0632)	(0.0344)
educ - graduate	-0.628***	-0.540***	-0.390***
	(0.0842)	(0.134)	(0.0729)
female × high school	0.0922*	-0.0370	0.0205
	(0.0511)	(0.0565)	(0.0352)
female × tech high school	0.136**	-0.0340	0.0359
	(0.0645)	(0.0647)	(0.0427)
female × tech tertiary	0.179***	-0.0736	-0.0342
	(0.0606)	(0.0668)	(0.0433)
female × college	0.176***	-0.145**	-0.0235
	(0.0563)	(0.0687)	(0.0435)
female × graduate	0.187	-0.155	-0.175*
	(0.114)	(0.151)	(0.101)
regional effects	Y	Y	Y
industry efects	Y	Y	Y
firm size category	Y	Y	N
year effects	Y	Y	Y
monthly effects	Y	Y	Y
Observations	98,097	98,090	178,315
Pseudo R2	0.0410	0.0695	0.0235

Note: The time span is March 2010 to December 2019. $***p < 0.01$, $**\,p < 0.05$, $*p < 0.1$

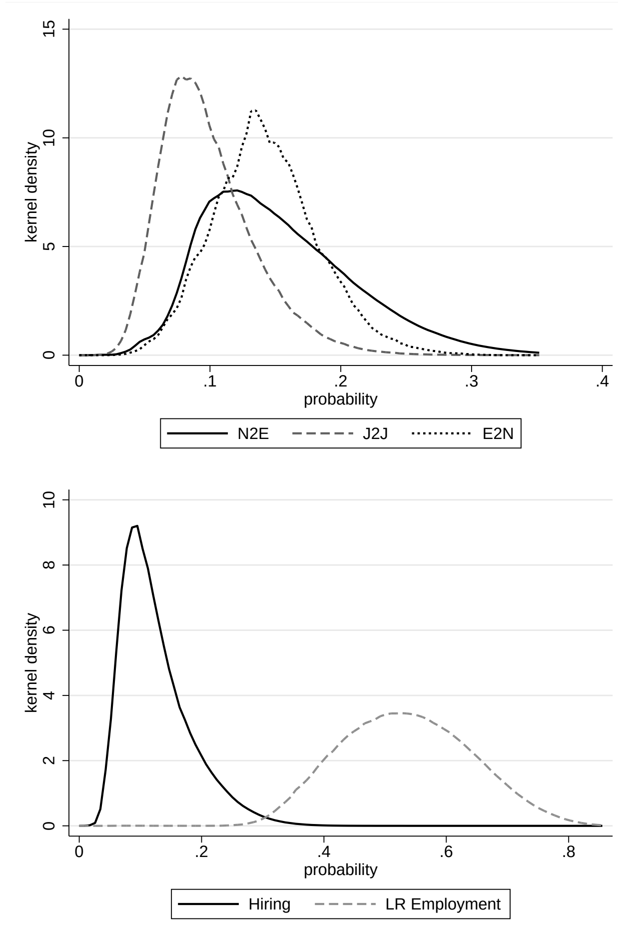


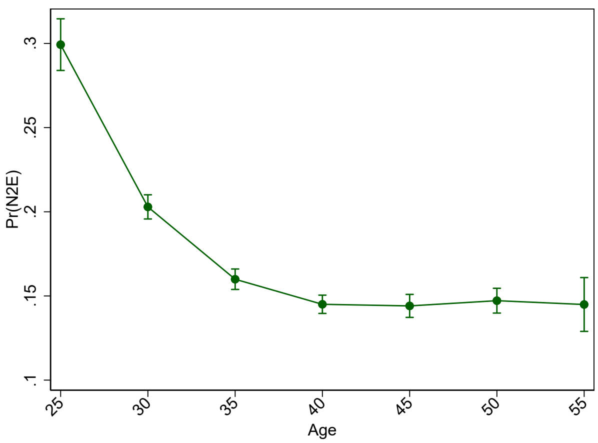
Figure A3: Predicted transition probabilities using ENE data given observables in the website (upper panel) and predicted hiring and long term employment probabilities, i.e. ex post weights (bottom panel)

Table A11: Ex post reweighted cross-section sorting estimates: additional samples

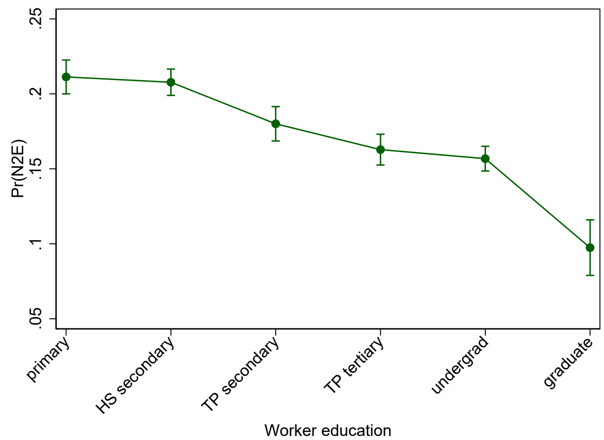
Unemployed						
	Hired prob weight			Employment prob weight		
	ES	eAKM	eAKM-f	ES	eAKM	eAKM-f
ρ	0.575***	0.086***	0.149***	0.581***	0.095***	0.155***
	(0.022)	(0.028)	(0.027)	(0.021)	(0.026)	(0.026)
Obs	18,871,284	18,865,148	18,865,035	18,871,284	18,865,148	18,865,035
R^2	0.331	0.007	0.022	0.337	0.009	0.024

Employed						
	Hired prob weight			Employment prob weight		
	ES	eAKM	eAKM-f	ES	eAKM	eAKM-f
ρ	0.565***	0.101***	0.171***	0.557***	0.098***	0.167***
	(0.020)	(0.025)	(0.024)	(0.019)	(0.024)	(0.023)
Obs	20,458,835	20,454,588	20,454,573	20,458,835	20,454,588	20,454,573
R^2	0.319	0.010	0.029	0.310	0.010	0.028

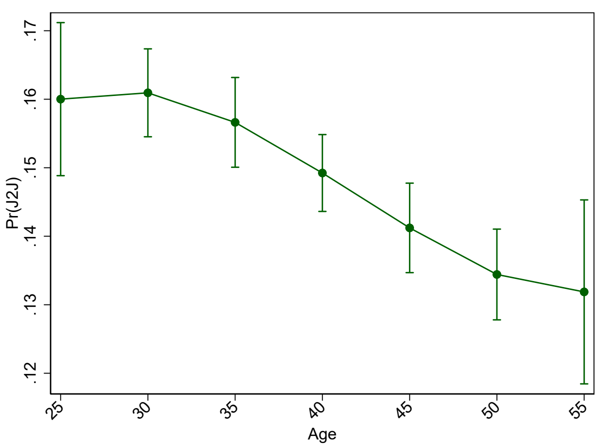
Note: 100X robust standard errors in parentheses. $***p < 0.01$, $**p < 0.05$, $*p < 0.1$



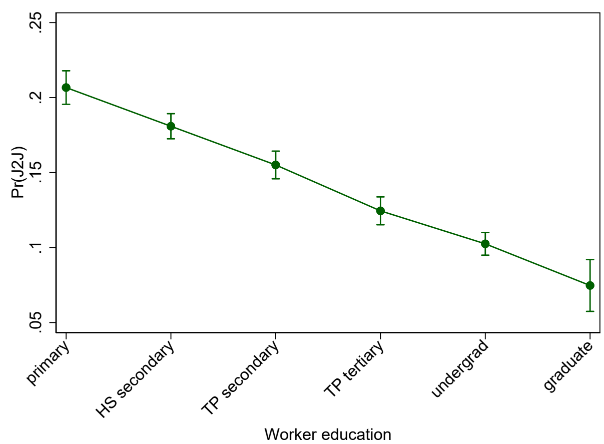
(a) Predicted U2E transition by age



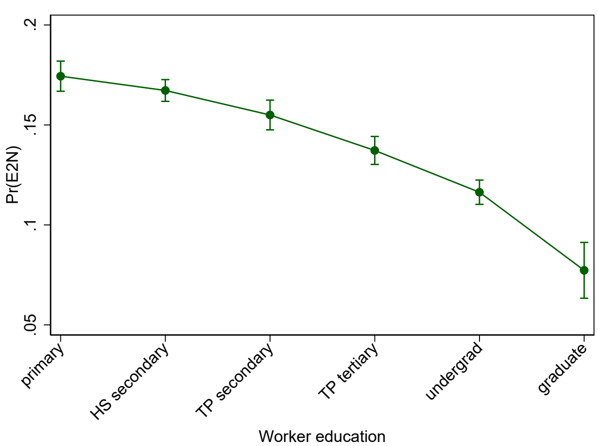
(b) Predicted U2E transition by education



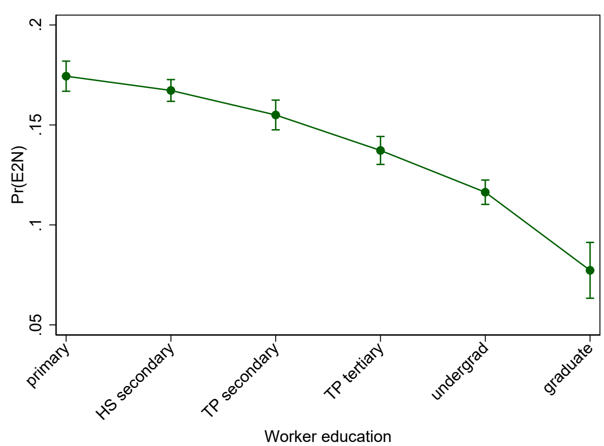
(c) Predicted J2J transition by age



(d) Predicted J2J transition by education

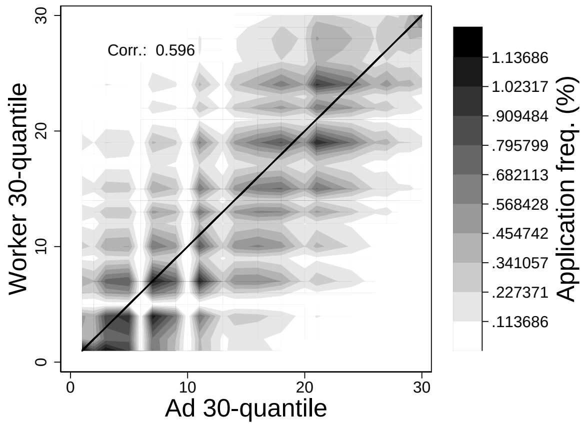


(e) Predicted E2U transition by age

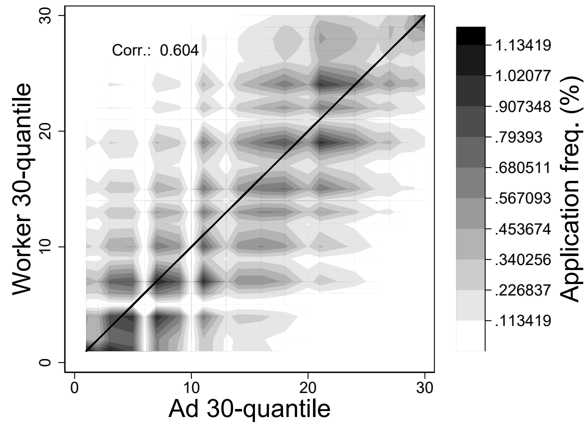


(f) Predicted E2U transition by education

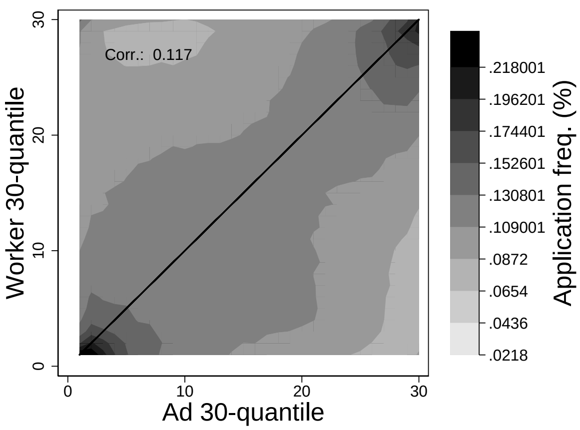
Figure A4: Note: Average predicted probabilities according to the forecast transition equations in *Encuesta Nacional de Empleo* in Table A10. Vertical bars at each point represent 95% confidence intervals.



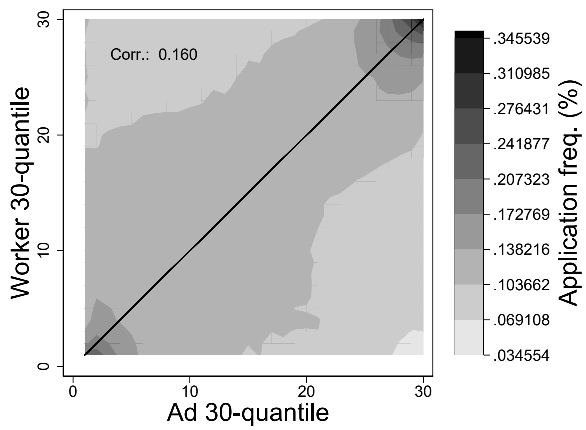
(a) ES long-run employment



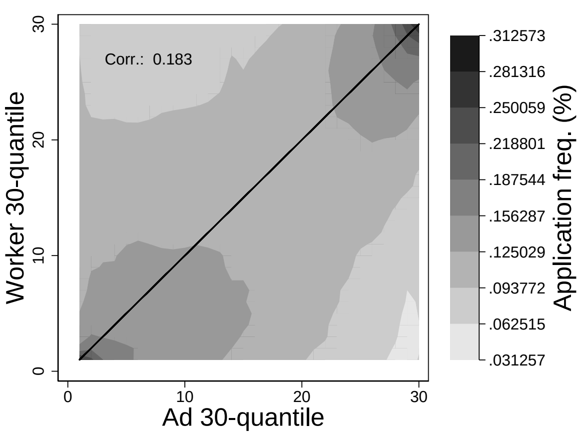
(b) ES expected employment duration



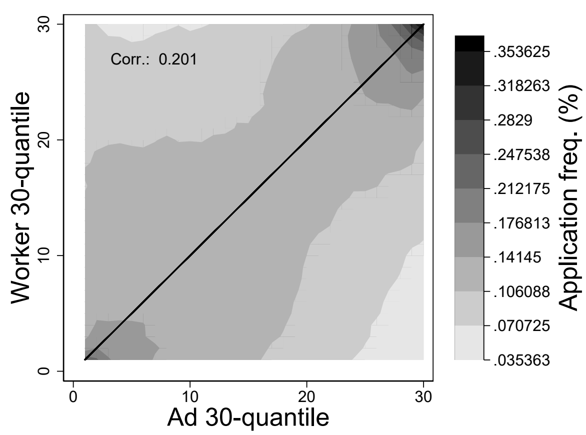
(c) eAKM long-run employment



(d) eAKM expected employment duration

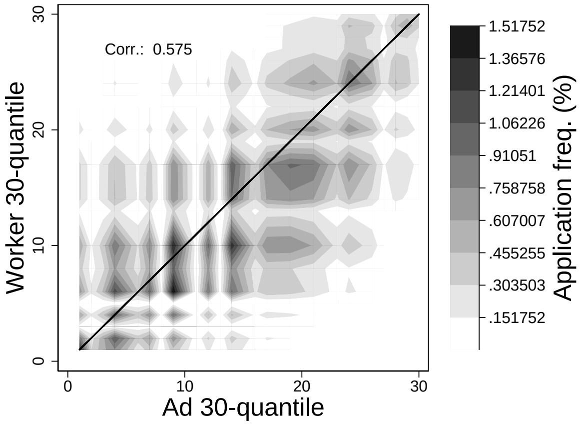


(e) eAKM-f long-run employment

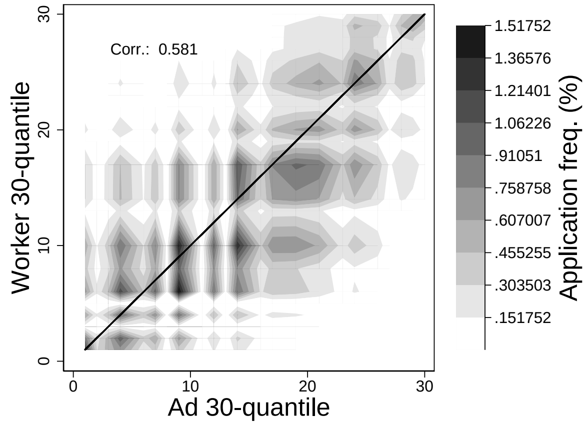


(f) eAKM-f expected employment duration

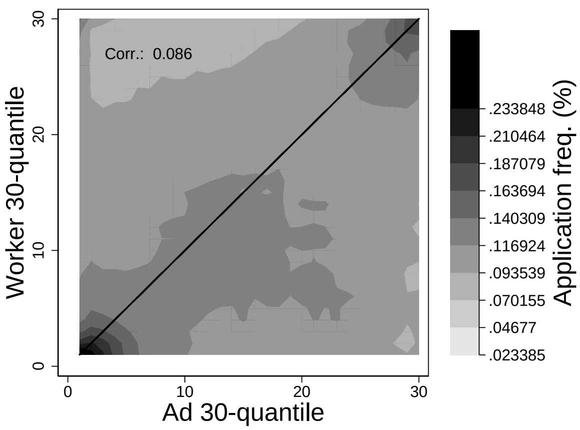
Figure A5: Frequencies of applications, by worker and ad percentiles of ES types reweighted by the estimated long-run employment probability or expected match duration given observables. Details of the computation of predicted probabilities are in the text. Sample consists on 39,371,368 applications between 1-Jan-2010 and 31-Dec-2019 in `www.trabajando.com`



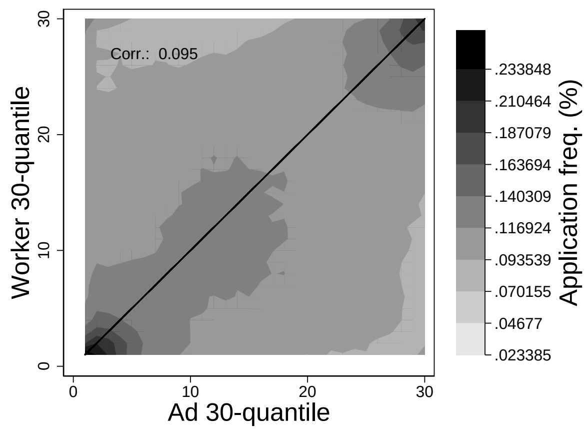
(a) Unemployed, ES types, hirings



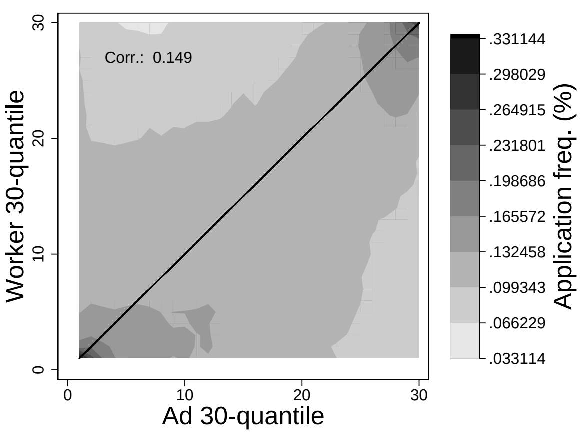
(b) Unemployed, ES types, long-term employment



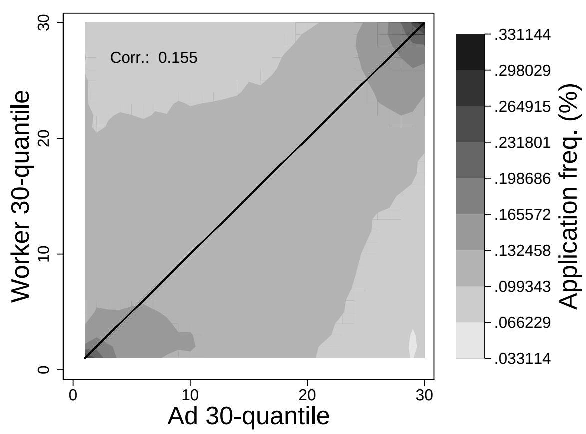
(c) Unemployed, eAKM types, hirings



(d) Unemployed, eAKM types, long-term employment

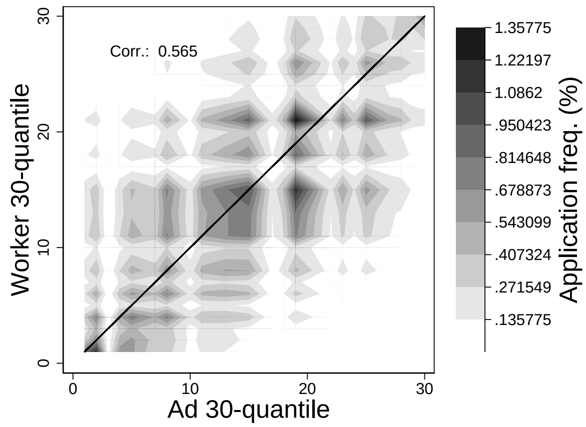


(e) Unemployed, eAKM-f types, hirings

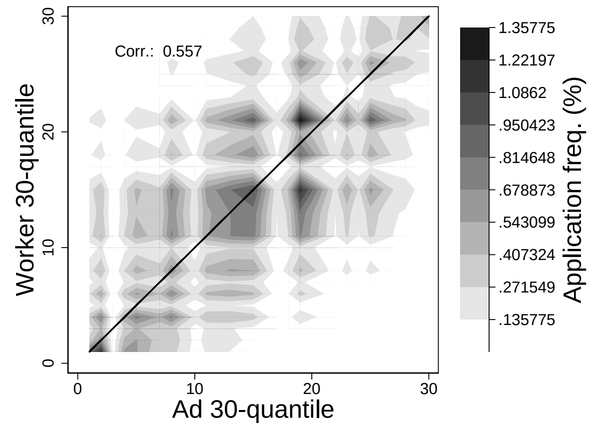


(f) Unemployed, eAKM-f types, long-term employment

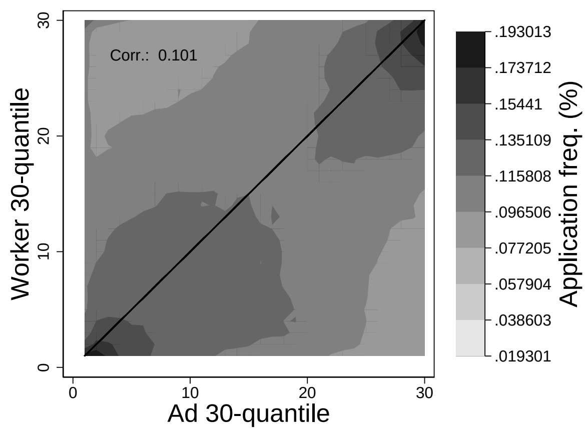
Figure A6: Note: Frequencies of applications, by worker and ad percentiles of ES, ex ante AKM, and ex ante AKM (with firm fixed effect) types reweighted by the estimated probabilities of being hired or being employed on the long term, given observables. All applications between 1-Jan-2010 and 31-Dec-2019 in www.trabajando.com.



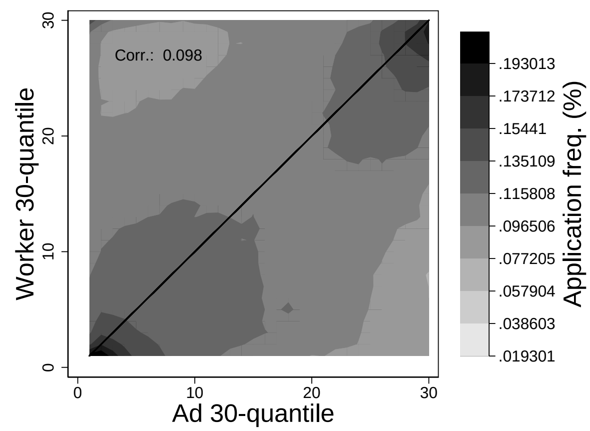
(a) Employed, ES types, hirings



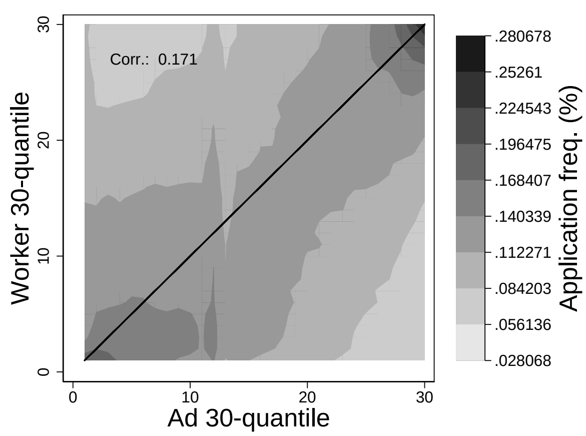
(b) Employed, ES types, long-term employment



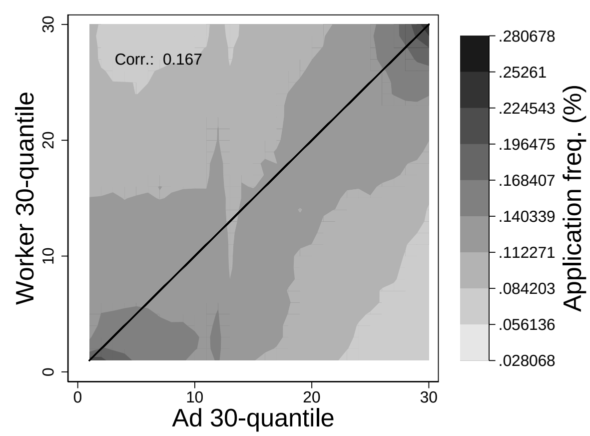
(c) Employed, eAKM types, hirings



(d) Employed, eAKM types, long-term employment



(e) Employed, eAKM-f types, hirings



(f) Employed, eAKM-f types, long-term employment

Figure A7: Note: Frequencies of applications, by worker and ad percentiles of ES, ex ante AKM, and ex ante AKM (with firm fixed effect) types reweighted by the estimated probabilities of being hired or being employed on the long term, given observables. All applications between 1-Jan-2010 and 31-Dec-2019 in `www.trabajando.com`