Job Search in Thick Markets: Evidence from Italy

Sabrina Di Addario*

4th December 2004 - Preliminary version.

Abstract

I analyze empirically the effects of both urban and industrial agglomeration on men's and women's search behavior and on the efficiency of matching. The analysis is based on the Italian Labor Force Survey micro-data, which covers 520 randomly drawn Local Labor Market Areas (66 per cent of the total) over the four quarters of 2002. I compute transition probabilities from non-employment to employment by jointly estimating the probability of searching and the probability of finding a job conditional on having searched, and I test whether these are affected by population size, industry localization, labor pooling and family network size. Preliminary results indicate that urbanization always increases job seekers' chances of finding employment, while the positive externalities generated by localization often appear only beyond a certain threshold (e.g., in the industrial districts with a higher share of small manufacturing firms). Labor pooling always increases hazard rates, for both men and women, while the size of family networks increases only men's chances of employment.

*Bank of Italy and University of Oxford. E-mail address: sabrina.diaddario@bancaditalia.it. This work owes much to extensive discussions with Eliana Viviano. I am particularly grateful for her invaluable help with the attrition analysis and with some aspects of programming. I also thank Erich Battistin, Massimo Caruso, William Greene, Vernon Henderson, Barbara Petrongolo, Margaret Stevens, William Strange for their helpful comments and suggestions. Finally, I thank both the participants to the CEPR conference "Integration and Technological Change: Challenges for European Regions" (11-13 June 2004, Paris) and those of the Labor Economics seminar held at the University of Oxford on 2nd December, 2004. The views expressed in this paper are those of the author and do not involve the responsibility of the Bank.

1 Introduction

Matching models are widely used to analyze the process of job formation in the presence of labor market frictions. These models are typically taken to operate, and empirically estimated, at the national level.[1] In a context of slow mobility of labor, however, the matching of workers and jobs may occur instead at a much more localized level (e.g., at the local labor market level), and in particular, it may be affected by the degree of urban or industrial agglomeration. Local markets, for instance, may differ in the presence of skill heterogeneities: according to Marshall's "labor pooling hypothesis", agglomeration lowers the degree of mismatch between the skills required by firms and those offered by workers, improving the matching process. Also, denser markets may be characterized by a lower degree of information imperfection. On the other hand, congestion depends on population and firm density, which may vary to a great extent across local markets.

In this paper I consider two specific aspects of labor market dynamics that can be affected by agglomeration externalities: the individual search intensity and the hazard rate into employment. The former is a measure of the individual's effort devoted to job seeking and depends on individual resources, search costs and expected returns. The latter, which corresponds to the probability that a non-employed job seeker finds employment, depends on overall labor demand and supply and on the technology of matching.

On the one hand, agglomeration may increase individual search intensity by reducing certain costs of search, such as transportation costs (due to shorter distances to job interviews or more frequent "face-to-face contacts") and information-gathering costs (due to the presence of thicker informal networks that reduce information asymmetries). Moreover, search may be more intense in agglomerated areas also because of a higher cost of living (e.g. house prices), which increases the opportunity cost of staying unemployed. On the other hand, agglomeration may reduce search intensity by increasing the costs of search due to congestion (e.g., traffic jams).

On the returns side, agglomeration may increase job seekers' search intensity by raising local wages and improving hazard rates. The latter, in turn, depend on the intensity of job advertising, the thickness of the labor market, and the technology of matching. While there is some empirical evidence of higher wages in agglomerated areas, the net effect of agglomeration on labor market tightness and on the technology of matching is less clear-cut. Indeed, agglomeration may raise both the demand and the supply of labor, so that it is not obvious whether it would make markets more or less tight. With regards to the technology of matching, a higher density may actually lower

[1] See Petrongolo and Pissarides (2001) for a survey.

the contact rate if congestion effects dominate over "thick" markets externalities.[2] Even when the contact rate (per unit of search) is improved, moreover, job seekers may react by becoming choosier and accepting job offers less frequently, thereby depressing the hazard rate. Which type of external (dis)economy will prevail is, ultimately, a matter of empirical investigation.

In this paper I use the Italian Labor Force Survey micro-data to estimate the effects of agglomeration on employment probabilities and job search intensity. To measure urban agglomeration effects I first use a dummy for large city, equal to one if a local labor market system (LLM) has a population above $404,526$ inhabitants. In contrast to other studies using arbitrary cut-off points, I adopt the same threshold value devised by Di Addario and Patacchini (2004) on the basis of spatial econometric analysis applied on Italian LLMs. However, since the spatial unit of analysis is crucial to determine the existence and extent of agglomeration externalities (Arzaghi and Henderson, 2004), I also use LLMs population size.[3] Secondly, I compare the labor market dynamics of non-employed people living in urban and industrially agglomerated areas to those living in the rest of the country. In order to do so I divide the Italian territory into three sets of LLMs, namely large cities, non-industrial small towns, and small towns with a high presence of small and medium sized manufacturing firms ("super-districts"). This method enables me in particular to compare the impact of industry localization and urbanization on search behavior and employment probabilities; to my knowledge, this has not been analyzed before. I also test the effect of labor pooling, proxied by the traditional sector specialization index. Finally, I use the number of employed individuals in the household as a proxy for network size,[4] under the assumption that family networks are important to find employment and that employed individuals have access to larger networks than unemployed ones (as they presumably have more information on job offers).

My results suggest that the externalities generated by agglomeration on search behavior and employment probabilities vary according to both the type and the degree of agglomeration considered. I thus find that urbanization always increases job seekers' chances of finding employment, while the positive externalities generated by localization appear only beyond a certain threshold

[2] See Petrongolo and Pissarides (2001). Other sources of congestion may derive from local "amenities" such as more traffic jams, crowded subways, pollution etc. For a survey on agglomeration externalities see Rosenthal and Strange (2004) and Duranton and Puga (2004).

[3] According to Rosenthal and Strange (2004) the size of the area may matter, as externalities decay quickly over space (within 10 miles). However, in my regressions the log of LLM area was never significant. Most of the above described mechanisms generating agglomeration externalities on search behavior derive from density, but this is not always necessarily the case. Furthermore, these effects may occur in the continuum or only beyond certain threshold values.

[4] In a similar fashion as in Wahba and Zenou (2003), who use the number of family members who are in the labor force.

(e.g., in the industrial clusters with a higher share of small manufacturing firms). Living in a more skill-homogenous labor market, however, increases the probability of finding a job for both men and women, while having a larger family network increases only men's chances of employment. I also find that urbanization lowers the intensity of search, possibly because job seekers are discouraged by the higher congestion costs deriving from a larger population mass.

These results are not in conflict with other findings of the literature (e.g., Petrongolo, 2001) that do not find overall agglomeration effects on the unconditional probability of finding a job. In fact, to my knowledge this is one of the few studies that analyzes agglomeration effects on hazard rates conditional on having searched.

The paper is structured as follows. The next Section presents the theoretical framework; Section 3 reports the empirical model, Section 4 the data set and the variables; Section 5 discusses the estimation results; and Section 6 concludes.

2 The theoretical framework

In the standard search and matching literature (for instance, Pissarides, 2000), the number of matches M is expressed as an increasing and concave function of the amount of workers searching for employment and the number of vacant positions. To study the effects of agglomeration on search, I assume that the national labor market is geographically segmented. Thus, every geographical unit or local labor market j has a matching function specific to the area, both in terms of arguments (as in Patacchini and Zenou, 2003) and in terms of technology:

$$M_j = M_j(s_j J_j, a_j V_j) \tag{1}$$

where J_j is the number of searchers in local labor market j, s_j the area's average search intensity, V_j the amount of vacancies, and a_j the area's intensity of job advertising.

The rate of job-finding for an individual i searching with intensity s_{ij} (with $s'_{ij} > 0$ and $s''_{ij} \leq 0$) is:

$$m(s_{ij}, a_j\theta_j) = s_{ij}\frac{M_j(s_j J_j, a_j V_j)}{s_j J_j} = s_{ij} h_j(a_j\theta_j) \tag{2}$$

where h_j is the rate of matching per unit of search,[5] and $\theta_j = V_j/s_j J_j$ is a measure of the area's labor market tightness.

[5] That is, the rate at which a worker searching with unit intensity will find a job, if s_{ij} is normalized to be between 0 and 1. Under this normalization, in the empirical part of the paper (Section 3) I take s_{ij} to be the probability

Let a job seeker's budget constraint be:

$$b = C_j(s_{ij}) + c_j z_{ij} \tag{3}$$

with:

$$C_j(s_{ij}) = d_j s_{ij}^{\gamma}, \gamma > 1 \tag{4}$$

where b denotes the income of a non-employed person, $C_j(s_{ij})$ the cost of search, z_{ij} a real consumption good bundle, and c_j the area cost of living (e.g. housing costs). I assume that agents' utility from consumption $u(z_{ij})$ is an increasing and concave function of z_{ij}. The expected intertemporal utility (in steady state) achieved by an unemployed agent is therefore:

$$rW_{ij}^{U} = u\left(\frac{b - C_j(s_{ij})}{c_j}\right) + s_{ij} h_j(a_j \theta_j)(W_{ij}^{E} - W_{ij}^{U}) \tag{5}$$

where W_{ij}^{E} is her expected lifetime utility when currently employed and r the discount rate.

The optimal level of search intensity s_{ij}^{*} a job seeker will exercise is that which maximizes (5): $\partial W_{ij}^{U}/\partial s_{ij} = 0$, or (at an interior solution):

$$u'\left(z_{ij}\right) \frac{C_j'(s_{ij})}{c_j} = h_j(a_j \theta_j)(W_{ij}^{E} - W_{ij}^{U}) \tag{6}$$

Job seekers are thus faced with a trade-off between the marginal cost of increased search effort in terms of current consumption and the marginal increase in their chances of finding a job that it induces. Thus, whether search is more or less intense in agglomerated areas depends on whether density lowers the costs of search and/or increases its returns. I take this simple model as the starting point to discuss the mechanisms through which agglomeration may affect individuals' search behavior.

On the one hand, search costs may be lower in agglomerated areas because physical proximity reduces transportation costs to job interviews and decreases the costs of acquiring information on job opportunities.[6] Moreover, denser areas may be characterized by the presence of networks that facilitate the diffusion of information on vacancies (Wahba and Zenou, 2003). On the other hand,

of searching and h_j to be the hazard rate (i.e., the probability of finding a job conditional on having searched). Note that the individual's job-finding-rate can be expressed as a function of labor market tightness only under the assumption of constant returns to scale of the matching function.

[6] From the firm's perspective, in Wheeler (2001) per-worker firm recruitment costs decrease with density, as the frequency of interactions enhances the arrival rate of potential workers for a job opening, which has a fixed cost.

search costs may be higher in agglomerated areas because of the congestion costs generated by population density (e.g., more intense traffic jams, etc.).

Secondly, the higher the cost of living c_j of the area, the higher the individual's search intensity $s^*_{ij} = s(e^*_{ij})$. As more congested areas are likely to have a higher cost of living (e.g. housing costs), increasing the cost of staying unemployed with respect to lower-density areas, agglomeration should induce job seekers to search more intensively. This effect occurs whenever the unemployment benefit b is either fixed or less responsive to the local cost of living c_j than local nominal wages; in fact, there is evidence that real wages are actually higher in denser areas,[7] and b will include some nationally determined benefits that are not indexed for local cost-of-living.

With regards to the returns to search (the hazard rate), there are four main channels through which agglomeration may affect search: wages, vacancy advertisement, labor market tightness, and the technology of matching.

First, individual search intensity obviously increases with real wages, which raise the utility from employment. According to the literature on agglomeration, in denser areas wages may be higher than average,[8] which should increase search intensity.

Second, individual search intensity increases with the intensity of job advertising by firms (Pissarides, 2000). This channel operates through an improvement of the hazard rate. The impact of agglomeration on the intensity of job advertising is twofold. On the one hand, it has a positive effect, as the existence of networks in denser areas[9] may reduce the cost incurred by firms in advertising their vacant positions, and the higher total number of job seekers may allow them to more easily cover any fixed costs of advertisement. Finally, a greater labor productivity may also increase job advertising intensity.[10] In this case, job seekers exercise more effort in denser areas simply because they have better chances to find a job and are hence more encouraged to search than elsewhere.[11] On the other hand, if denser areas were characterized by tighter labor markets agglomeration would reduce job advertising intensity, since in this case a lower chance of filling their vacancies would discourage firms from advertising their positions ("discouraged-job" effect).

Third, individual search intensity increases with labor market tightness. Whether markets are

[7] Glaeser and Mare' (2001); Wheaton and Lewis (2002).

[8] See, for instance, Glaeser and Mare' (2001) for urban wages in the US and Di Addario and Patacchini (2004) on Italy, though de Blasio and Di Addario (2002) do not find evidence of higher average wages in Italian Industrial Districts.

[9] These can either be informal (Marshall's "industrial atmosphere") or real network agencies (Arzaghi and Henderson, 2004).

[10] See Pissarides (2000) for a partial equilibrium analysis of job advertising and Ciccone and Hall (1996) – among others – for the evidence on higher labor productivity in denser areas.

[11] As Pissarides (2000) notices, this is the reverse of the discouraged-worker effect.

more or less tight in agglomerated areas is itself a question of empirical investigation. Indeed, in agglomerated areas there are reasons to expect the number of both applications and vacancies to be higher than in non-agglomerated zones. Since there are no reliable data on vacancies in Italy, I cannot empirically test the existence of differentials in local labor market tightness due to agglomeration.[12] These can only be inferred from the impact of agglomeration on individual hazard rates, which are increasing in market tightness and can be measured directly.

Finally, search intensity depends on the technology of matching. Agglomeration may have an impact both on the chances and on the quality of matching. First, the greater concentration and / or specialization of matching agents in agglomerated areas may increase the effective job contact rate, and hence the hazard rate.[13] On the other hand, knowing that in agglomerated areas the probability of a meeting is higher, job seekers may be choosier, which lowers the probability of acceptance of a job offer. Which of the two effects will dominate on the hazard rate is a matter of empirical investigation. Second, the better expected quality of matches[14] may, on the one hand, increase the probability of acceptance as firms make more attractive offers, but on the other hand agents' choosiness again increases, which goes in the opposite direction.

3 The empirical model

As I showed in the previous section (equation (6)), the transition probabilities from non-employment into employment depend on two elements, one determined by agents' search behavior and the other by the matching process. In order to empirically examine the impact of agglomeration on the transition probabilities between labor market states one thus needs to find measures of both the individual's propensity to search and of the effectiveness of matching.

I shall define s_{it} as the probability that a non-employed person looks for a job at time t,[15] and h_{it} as the probability that she finds employment at time $t+1$, conditional on having searched. Each person who was not employed at time t can be in one of the possible three states at time $t+1$:

1. they sought employment between t and $t+1$ and found a job (E_{t+1});

[12] Although very few micro studies on hazards control for vacancies (Petrongolo and Pissarides, 2001).

[13] Note that agglomeration may also affect the elasticities of the matching function with respect to job seekers and vacancies, in such a way that increasing returns to scale are generated (see Petrongolo and Pissarides (2001) for a review of the empirical studies).

[14] This is Marshall's "labor pooling hypothesis". See Duranton and Puga (2004) for a survey on the agglomeration effects on the chances and quality of matching.

[15] Note that in the theoretical model presented in Section 2, s_{it} was a continuous variable ≥ 0 denoting the number of search units supplied by the individual i. Here, without loss of generality, I am normalizing search intensity to be between zero and one.

2. they sought employment between t and $t+1$ but did not find a job (U_{t+1});

3. they did not seek employment between t and $t+1$ (O_{t+1}).

Let $\tilde{s}_{it}$ be the latent variable determining whether a non-employed person looks for a job at time t (i.e., the difference in her expected utility from searching and not searching) and $\tilde{h}_{it}$ the variable determining whether a job seeker finds employment at time $t+1$ (incorporating both the likelihood of her meeting a prospective employer and the sign of the surplus generated by that match). Even though $\tilde{h}_{it}$ and $\tilde{s}_{it}$ are not observable, I can express them as a function of two non-coincident sets of individual and location-specific variables, X_{it} and Z_{it} (detailed in Section 4), using the Labor Force Survey micro-data on labor market transitions:[16]

$$\tilde{h}_{it} = \beta' X_{it} + \epsilon_{1t} \tag{7}$$

and

$$\tilde{s}_{it} = \gamma' Z_{it} + \epsilon_{2t} \tag{8}$$

The probability of observing a person who has searched at time t is thus $Pr(\gamma' Z_{it} + \epsilon_{2t} > 0|\ Z_{it})$, which I assume to be a probit $\Phi(\gamma' Z_{it})$. Similarly, the probability of observing a job seeker to find a job at $t+1$ is $Pr(\beta' X_{it} + \epsilon_{1t} > 0|\ X_{it}) = \Phi(\beta' X_{it})$.

My econometric methodology will consist in the joint estimation of s_{it} and h_{it} by maximum likelihood. To ensure robustness, two alternative econometric specifications will be estimated.

I first consider a simple search model where (after controlling for observable characteristics) individuals can be treated as identical, in the sense of being randomly matched to vacancies. In this framework, the transition probability from non-employment into employment is the product of the probability of searching s_{it} and the probability h_{it} that a job seeker finds a job. Thus, I will estimate s_{it} and h_{it} by maximizing the following likelihood function (as in Peracchi and Viviano, 2004):[17]

$$L = \prod_{i \in \{E_{t+1}\}} [\Phi(\beta' X_i)][\Phi(\gamma' Z_i)] \prod_{i \in \{Ut+1\}} [1 - \Phi(\beta' X_i)][\Phi(\gamma' Z_i)] \prod_{i \in \{Ot+1\}} [1 - \Phi(\gamma' Z_i)] \tag{9}$$

[16] Even though in the estimations I allow for location-specific effects, in this exposition I take the geographic area indices j as implicit in the individual characteristics of agent i.

[17] A large part of the empirical literature on hazard functions (see Devine and Kiefer (1991) for a review) assumes that the error terms are distributed according to a logistic function. I adopt here a normal distribution to be consistent with the second econometric model (see below). In any case, I also tested all the specifications reported in Section 5 assuming a logistic distribution and obtained very similar results (available upon request).

If there is unobservable heterogeneity among workers, however, the probabilities of searching and finding a job (conditional on the X_i and Z_i's) will not be independent. I therefore correct the above maximum-likelihood estimation to take into account the fact that the hazard-rate equation can be estimated only on the censored sample of the agents who search ($Z_{it}\gamma + \varepsilon_{i2} > 0$). To do so I adopt the method proposed by van de Ven and van Praag (1981) for bivariate probit models with sample selection. In this case, the likelihood function is:

$$L = \prod_{i\in\{Et+1\}} \Phi_2(\beta' X_i, \gamma' Z_i, \rho) \prod_{i\in\{Ut+1\}} \Phi_2(-\beta' X_i, \gamma' Z_i, -\rho) \prod_{i\in\{Ot+1\}} [1 - \Phi(\gamma' Z_i)] \tag{10}$$

where Φ_2 is the bivariate standard normal cumulative distribution of the joint probability of s_{it} and h_{it}, and ρ is the correlation between the error terms. This method corrects the bias that arises from using (9) when the error terms in equations (7) and (8) contains some common omitted variable.

The results of the two estimation methods are reported in Section 5.

4 The data

For the empirical estimation I use the Labor Force Survey (LFS), conducted in the year 2002 by the Italian National Statistical Office (Istat). This survey is the main source of information on individuals' working condition, unemployment and job search behavior, in addition to their personal characteristics.

The survey is conducted quarterly in two stages: about $1,300$ municipalities are sampled at the first stage, and about $70,000$ households at the second one. The LFS follows a rotating scheme according to which each family is interviewed for two successive waves, and then again for two other consecutive rounds after two quarters of interruption, for a total of four times. So, theoretically 50 per cent of the sample is kept constant between two consecutive rounds. The LFS has a natural longitudinal dimension with people followed up to fifteen months, but the linkage of individual records across surveys can be problematic, because of the lack of a personal identifier and because of reporting errors in the household code.

Istat currently provides yearly longitudinal files linked with a stochastic matching algorithm, but these files do not contain information about individuals' place of residence and cannot be used to study the effects of agglomeration on labor market dynamics. In this paper, I reconstructed the longitudinal quarterly transitions with a deterministic method linking individuals' records on the basis of the family identifier and some time-invariant information (i.e. the date of birth and sex; see the Appendix for further details), which enables me to recover 75 percent of the potential longitudinal sample. The loss of the remaining observations could be a potential source of bias for my estimates in case it was not randomly distributed. Even though it cannot be known whether theses losses are due to random reporting errors in the key variables or to the non-random exit of some individuals from the LFS ("attrition"), I can test the hypothesis of random loss of information. In the Appendix I describe the methodology adopted and report the test results, which confirm the validity of my deterministic matching procedure for constructing an appropriate panel dataset for the analysis of labor market dynamics.

In this paper most agglomeration units are defined on the basis of "local labor markets" (LLMs). LLMs are clusters of municipalities aggregated according to the residents' daily commuting flows to their place of work.[18] LLMs are relatively self-contained, in that, by definition, they offer employment to at least 75 per cent of their residing workers, both with respect to the total number of workers in the area and with respect to the total number of residents. Exhaustive partitions of the territory based on worker commuting have been devised in many OECD countries,[19] since they reflect local labor market conditions better than administrative areas do. The literature on matching is increasingly basing the empirical analysis on LLMs, in order to avoid a geographical aggregation bias in contexts of imperfect labor mobility. The geographical reach of agglomeration externalities is itself at the center of the literature debate, and may depend on the specific phenomenon analyzed.[20] The LLMs' characteristic of self-containment makes them particularly suited to be my spatial unit of analysis, since by definition the likelihood that a non-employed person looks for a job within its boundaries is high, especially in the Italian context of low labor mobility (see, among others, Cannari, Nucci and Sestito, 2000). As Table 1 shows, in Italy even unemployed job seekers are generally unwilling to move from their town of residence to find a job: 80 percent of the unemployed are ready to accept a job only in their LLM of residence, and 41 percent only in their own municipality. Furthermore, only 7 percent of those employed work in a province different from the one of their residence.

I consider various measures of agglomeration, both urban and industrial.

Urbanization is measured by two alternative variables: the logarithm of LLMs' population,[21]

[18] The flows are obtained from the 1991 Population Census data. I assigned each LFS observation to a LLM with an Istat's algorithm that matches LLMs to municipalities.

[19] The UK, for instance, has been divided into 308 "travel-to-work areas" (OECD, 2002).

[20] See Arzaghi and Henderson (2004) for a discussion on this issue and Petrongolo and Pissarides (2001) for a review of matching studies based on LLMs.

[21] Even though some of the literature (e.g., Coles and Smith, 1996) states that it is density that matters rather than population size in generating externalities, in my regressions the logarithm of LLM area is never significant.

and a "large city " dummy. In Italy, LLMs' population density, size and area vary greatly. Density ranges from a minimum of 10 inhabitants per square Kms (Crodo) to a maximum of 3,250 (Naples), population size from 2,901 inhabitants (Limone sul Garda) to 3,311,431 (Rome), and LLM area from 10.4 square Kms (Capri) to 3,539 (Rome). However, both population density and size increase very gradually: the largest variations occur only at the upper end of the distribution. This suggests the use of a "large city" dummy to test whether agglomeration economies manifest themselves only beyond a certain threshold value. However, the choice of threshold values is often arbitrary and should plausibly be country-specific. The Italian population, for instance, is much more evenly distributed than the one in the US, suggesting the use of different threshold values to define large cities in the two countries. Thus, while Glaeser and Mare' (2001) use the (arbitrary) cut-off point of 500,000 inhabitants, I adopt the threshold value of 404,526, suggested by Di Addario and Patacchini (2004) on the basis of spatial autocorrelation analysis[22] applied on Italian LLMs.

Industry localization is measured by the Cannari's and Signorini's (2000) "super-district" dummy. Super-districts are the "industrial district" subset with a highest incidence of LLM small-firm manufacturing employment. Industrial districts, in turn, are identified by an Istat algorithm that associates to each LLM a dummy variable equal to one if the area shows both a dominant sectoral specialization and a higher-than-average share of small and medium enterprises (SME) and manufacturing employment.[23]

The only measure of agglomeration that is not based on LLMs is the number of employed household members, that I take as a proxy of network quality.[24] The idea is that family networks are important to find employment and that employed individuals have access to better quality networks than unemployed ones (as they presumably have more information on job offers). The validity of this variable relies on the absence of unobserved characteristics (such as ability) shared among family members.

In 2002 LFS surveyed 777,248 individuals. In order to analyze transition probabilities I restricted the sample to the people who were surveyed for at least two consecutive waves. Since my analysis concerns the labor market dynamics of non-employed persons, I also excluded those already employed at time t, and those either below the age of 15 or above that of 64. After excluding those

[22] More specifically, the authors use the Moran Scatterplot in conjunction to the Local Spatial Autocorrelation Statistics.

[23] See de Blasio and Di Addario (2002) for a more detailed explanation of how industrial districts are identified by Istat and Cannari and Signorini (2000) for a description of the methodology used to single out super-districts. Note that I also tested the effect of the industrial district dummy, though I do not report the results here as this variable was never significant.

[24] As in Wahba and Zenou (2003), who use the number of family members who are in the labor force.

for whom there were missing observations, the data set comprises $71,286$ non-employed individuals.

In Italy there are 784 LLMs: of these, 20 have a population above the $404,526$ inhabitant threshold and 99 are classified as super-districts (199 as industrial districts). My sample includes 520 LLMs, of which 19 are large cities (for a total of $20,335$ observations) and 70 super-districts (for a total of $5,285$ individuals). Thus, even though the LFS was not designed to represent the super-district population, my sample distribution reflects that found at the national level (the share of LLMs classified as super-districts is 13.5 percent in my sample and 12.6 percent in Italy). Finally, as the survey is stratified to represent Italian regions and municipalities, the 19 largest LLMs are always sampled.

5 Empirical analysis

I now turn to the empirical estimation of the determinants of individual search intensities and hazard rates, examining in particular whether these probabilities differ between agglomerated and non-agglomerated areas. The estimations were conducted separately for men and women and, unsurprisingly, labor market dynamics turned out to be substantially different for the two groups.

5.1 Descriptive statistics

Table 2 reports the quarterly transition probabilities and flows both at the aggregate level and for men and women separately. The transition matrix shows that in Italy there is a high unemployment persistence, as 63 percent of the people unemployed in the quarter preceding the interview are still unemployed in the successive quarter. While these numbers are very similar for men and women, significant gender differences can be found in other respects. First, in the average probability of finding a job, conditional on being non-employed at time t: the transition probability from unemployment into employment is almost 18 percent for men and only 10 percent for women, and the respective probabilities of finding a job for those recorded as inactive at time t are 5 and 3 percent respectively.[25] Second, the transition probability from unemployment into inaction, greater than that into employment for both sexes, is much larger for women than for men (in line with other empirical results, e.g., Broersma and Van Ours, 1999).

Finally, Table 2 shows that the flows from inactivity to employment as a percentage of the working age population are generally more substantial than those from unemployment into employment

[25] However, when expressed in percentage of the working age population, the flows from inactivity to employment are larger for women than for men.

(1.4 versus 0.8 percent; in line with previous results, e.g., Petrongolo and Pissarides, 2001). In light of this fact, and consistently with the most recent literature (Broersma and Van Ours (1999); Brandolini *et al.*, 2004), I shall estimate hazards from non-employment to employment rather than from unemployment.

The Italian labor market is known to be segmented with respect to territory (see, for instance, Peracchi and Viviano, 2004). While, traditionally, labor market conditions are analyzed at the macro-area level (North, Center, and South),[26] I examine whether they also differ along the degree of urban and / or industrial agglomeration. Table 3 reports descriptive statistics for the year 2002 on the employment, unemployment and activity rates for all the agglomeration units considered in this paper (large cities, super-districts, and industry-thin small-sized towns). It also shows the share of job seekers in total non-employed population and the hazard rate into employment. The former, computed as the ratio between the sum of the employed and unemployed persons at the time of the interview and the non-employed people who actively searched in the preceding quarter, can be interpreted as a measure of average search intensity.[27] The hazard to employment is the probability that a job seeker finds a job between successive quarters, and is computed as the ratio between those moving into employment between time t and $t+1$ and total job seekers.

In 2002 the unemployment rate ranged from a minimum of 3 percent in super-districts to a maximum of 10 percent in large cities. Conversely, employment rates were lowest in large cities and highest in super-districts (55 percent against 65 percent). These patterns are largely confirmed at the macro-area level, so that they cannot be explained by the fact that most industrial districts are located in the regions of the Center-North-East of the country. With regards to labor market dynamics, the industrially denser areas show the lowest share of job seekers and the highest hazards to employment from non-employment (respectively, 11 and 57 percent). In contrast, large cities show the lowest hazards to employment, probably in large part due to the greater stock of job seekers concurring for available jobs, and relatively high unemployment rates. These offsetting effects are mostly confirmed in all the Italian macro-areas.

The descriptive statistics of Table 3 would thus suggest that agglomeration is associated with

[26] In 2002, for instance, unemployment rates ranged from 3 percent in the North-East to 16 percent in the South, while employment rates ranged, respectively, from 65 percent to 48 percent (see Table 3).

[27] Note that in this paper the pool of job seekers is larger than the set of the people recorded as unemployed according to the ILO definition. In line with a large part of the empirical literature on matching (see Petrongolo and Pissarides (2001) for a survey), I assume that each search period (the time interval between t and $t+1$) lasts three months. Thus, to ensure temporal consistency between stock and flow data (transitions to employment) the job seekers' pool must comprise all non-employed people, willing to start working immediately, whose last search action took place in the previous quarter – rather than in the previous month, as it is in the ILO definition (see Brandolini *et al.* (2004), and Peracchi and Viviano (2004) for a discussion).

specific labor market dynamics. In particular, I would expect search intensity to be highest in large cities and hazard rates to be highest in super-districts. The impact of agglomeration, however, can be better analyzed in a more comprehensive model where the features of the local labor markets and the characteristics of individuals are taken into account.

5.2 Empirical specification

The empirical models proposed in Section 3 can be used for this purpose. In the remainder of this section, I will first examine a baseline model estimating the parameters of the log-likelihood functions (9) and (10) on the basis of individual and local labor demand characteristics, then test for the existence of agglomeration effects on both hazard rates to employment and search intensity.

The hazard rate to employment depends first of all on variables affecting local labor demand conditions and the individual's productivity. The former are proxied with three set of indicators. First, two indices meant to capture contemporaneous labor demand shocks: the share of employees working overtime in total workers and the average number of extra-hours worked.[28] The coefficients on these variables should be either significantly positive or zero, depending on whether demand expansion is or is not fully compensated by overtime work increases. In the latter case, a rise of overtime work would be accompanied by an increase in the number of vacancies, which, other things being equal, would improve the hazard rate. In contrast, if all the demand increase was entirely compensated by overtime work, my indicators should not affect the hazard rate. The second local labor market variable I consider is the geographical density of job seekers (similarly to Petrongolo, 2001).[29] Since, as shown in Section 2, hazard rates are increasing in local labor market tightness, I expect job seeker density to have a negative sign. The third set of variables includes the LLM Pavitt specialization indices.[30] In particular, I expect hazard rates to be higher in the LLMs with a higher concentration of labor-intensive sectors, under the hypothesis that the areas characterized by a large presence of traditional sectors can be taken as proxies for labor pooling or, more generally, industry agglomeration. The personal characteristics that I use to control for the individual's productivity are age, age squared, and educational attainment (first degree, high school, compulsory education). I also control for search duration (0–1 month, 1–5 months, 6–11

[28] I am aware that these indices are imperfect proxy for demand, as they could also reflect supply-side conditions. Ideally, I should control for vacancies (even though the majority of hazard studies does not; Petrongolo and Pissarides, 2001), but there are no data for Italy.

[29] I also used the logarithm of the total labor force and that of the population above the age of 15, with no different results.

[30] The Pavitt specialization indices at the LLM level have been computed from the 1996 Industry Census. The four sectors are defined as follows: (1) high technology; (2) specialization; (3) scale intensive; and (4) traditional.

months), expecting it to be inversely related to the chances of finding a job. Finally, I control for a dummy denoting whether the individual had previous work experience, as well as for seasonal and geographical dummies.

As seen in the theoretical model (equation (6)), an agent's optimal search intensity $s(e_{it})$ depends on the hazard rate h_{it} into employment that he anticipates facing if he searches. In estimating the equation for search intensity, I therefore include all the individual and labor-market explanatory variables used in the hazard-rate equation. In order to proxy for the value (monetary and other) of non-search activities, which I expect to lower the probability of participation in any given application round (i.e., search intensity), I also include the individual's position within the household (single living alone, household head, and spouse), the self-perceived work status (housewife, student, or retired),[31] and the number of non-working people in the household.[32]

5.3 The results

5.3.1 Baseline model

Tables 4 and 5 present the results of the baseline model for men and for women, respectively. To show the robustness of my results, in each table I report the outcomes of both the econometric models discussed in Section 3 ((9) and (10)). In spite of the fact that the Wald-test always rejects the null hypothesis of zero correlation between the error terms, confirming the presence of a selection bias, the two estimation methods provide the same signs and statistical significance levels for almost all the regressors considered in the hazard rate equation (which is the one subject to the selection problem).

a) Hazard rates.

In the baseline model for men (Table 4), hazard rates are higher in the North-East, for those with previous work experience, the less educated, and the older population.[33] As expected, the probability of moving from non-employment into employment decreases with search duration (see, among others, Lancaster, 1979). In particular, individuals who have been searching for less than one month have a chance of finding a job twice as large as those who have been searching for more

[31] Since the household decisions are linked by a budget constraint, the position in the household may matter. Note that the sum of the three self-perceived work status dummies equals to being inactive at time t.

[32] Using data at the provincial level from the *Consulente Immobiliare*, I also controlled for house prices and rents, but these were never significant. I used data for 2002, the oldest year available (1965 for house prices, and 1993 for rents) and the average of the entire period.

[33] Even though these last two results are in contrast with some empirical studies on the U.K. (e.g., Lancaster, 1979), they are in line with previous findings on Italy (see, for instance, Peracchi and Viviano, 2004).

than one year.[34] As expected, higher LLMs' job seeker density reduces the probability of finding a job, probably because of the congestion that unemployed workers create on each other (see Burgess, 1993 or Petrongolo and Pissarides, 2001), whereas a LLM's specialization in labor-intensive sectors increases it. Surprisingly, a higher LLM share of overtime workers in total workers lowers hazard rates,[35] while average extra-hours worked do not have any significant impact. In contrast to the male population, women have a higher chance to find a job when they are younger and when they have a University degree, and a lower chance if they live in the South (Table 4).[36]

b) Search propensities.

For both men and women, search intensity increases with education, age, past work experience, and with residing in the North-East. In contrast, students, retired workers and housewives search less intensively, probably because these categories of job seekers assign a higher value to non-search activities than those who perceive themselves as unemployed. Interestingly, the position in the household matters differently for the two sexes, as being a household head or a spouse increases the probability of searching for men but decreases it for women (with respect to being an offspring or having other positions within the household). This different behavior probably reflects the tendency for wives and mothers to stay at home,[37] and a greater need for non-employed husbands and fathers, who are most often the primary earners in the household, to increase their search effort. Finally, the LLM job seeker density is positive and significant only for men, implying that women do not exercise more effort when competition for vacant jobs raises, while men do.

5.3.2 Effects of agglomeration

To examine the effects of agglomeration on s_i and h_i, I add to the baseline models the variables discussed in Section 4. Thus, I first consider the joint effect of the large city dummy, the industry

[34] In general, marginal effects have been computed at the mean for the continuous variables and for a discrete change from 0 to 1 for the dummy variables.

[35] This may be a sign that overtime work is mostly supply-driven: the extent to which people are willing to work extra hours, firms reduce the hiring rate. A possible explanation of why individuals should differ in their willingness to work extra hours is provided by Rosenthal and Strange (2002), according to whom in large markets people work more in order to signal their ability in a rivalrous context (the "urban rat race").

[36] These results are less surprising than those for men, which could possibly derive from the composition of the non-working population (e.g., a higher incidence of men difficult to employ, such as long-term unemployed, people with health problems, in the male sub-sample).

[37] Note that this may be due to child care, as Italy lacks of policies aimed at supporting mothers' employment. In order to test this hypothesis, I also ran he same regressions (not yet reported here) on the parent sub-sample, controlling for the number of children below the age of six. I find that a marginal increase in this variable lowers women's probability of searching by 1 percent (at 1 percent statistical significance), but does not affect men's behavior. This result supports the view that men and women have different behavior because the traditional household division implies that they face different (opportunity) costs of search.

localization variables and the proxy for family networks (first and third specifications of Tables 6–7).[38] I then substitute the large city dummy with the logarithm of population size (second and fourth columns).

a) Hazard rates

Table 6 summarizes the results on hazard rates and search intensity for the econometric model correcting for sample selection ((10)).[39] Specifications (6.1)–(6.2) and (6.5)–(6.6) report, respectively, the findings on men's and women's job-finding rates.

After controlling for LLM job seekers' density, which captures the negative congestion externality exercised by unemployed workers on each other (see Petrongolo, 2001), I find that urban agglomeration has an overall positive effect on the probability of finding a job, both when we measure it with the large city dummy and when we estimate it with the log of LLM population size. Indeed, both variables are positive and very significant (at the 1 percent level in the women's sub-sample and at the 1–5 percent level in the men's case).

While urbanization generates positive agglomeration economies for any level of LLM size, for localization to create significantly positive net externalities a minimum degree of firm thickness is necessary. Indeed, searching in more industrially agglomerated areas raises men's probability of finding employment only above a certain threshold of manufacturing SME concentration. Thus, other things being equal, living in a super-district increases a man's chance of finding employment (column (6.1)), while residing in an industrial district does not have any effect.[40] Since the significance level of the super-district dummy is reduced when I control for the continuous measure of population size (column (6.2)), I also split the small-sized LLMs into those that are industrially agglomerated and those that are not (which I can do since super-districts are all located in LLMs with less than $500,000$ inhabitants), in order to compare the effects of industry agglomeration on LLMs of similar size. The results, not reported in the table for space constraints, show that the super-district coefficient is greater than that in the less industrialized LLMs of the same size, but it is not statistically different from the benchmark, suggesting that in Italy urban and industrial

[38] I also considered the effect of each of these variables separately, with no different results. Note that whether the signs and the statistical significance of the urbanization and localization dummies can correctly identify agglomeration differentials in employment probabilities and search behavior clearly relies on LLMs to be separated markets (see, for instance, Coles and Smith (1996) or Duranton and Monastiriotis, 2002), as discussed in the previous section.

[39] From now on I will not report the results for (9) – available upon request – because the Wald-test always rejects the null hypothesis of no selection bias. In any case, the two models provide very similar outcomes on the sign and statistical significance of the agglomeration variables.

[40] Since the ID dummy is non-significant in all the specifications and samples I tested, I do not report the results for this variable (though they can be requested).

agglomeration economies are of a comparable size.[41]

Further evidence on the existence of positive industry localization effects is given by the traditional-sector-specialization Pavitt index. This index, taken as proxy of labor pooling / super-district agglomeration (since the latter are characterized by a high specialization in traditional sectors), should enable me to assess the effect of localization through an improvement of the *quality* of matches rather than through an increase of their *quantity*. Indeed, according to Marshall's "labor pooling hypothesis" (see Duranton and Puga (2004) and Rosenthal and Strange, 2004), labor pooling reduces the degree of mismatch between the skills required by the firms and those offered by the workers. Contact rates being equal, a better expected quality of matches may imply higher acceptance and therefore hazard rates, as job seekers become choosier but firms make more attractive offers, which increases the acceptance probability. As Table 6 shows, the labor pooling proxy is always positive and significant, suggesting that localization improves the quality of matches.

In contrast with the men's case, the super-district variable is never significant for women (columns (6.5)–(6.6)), while the index of specialization in traditional sectors keeps having a positive effect (significant at the 5 percent level).

Finally, in line with the priors, the proxy for household networks has a positive effect on the chances of employment[42] in both the specifications tested for men. However, the size of family networks does not affect women's likelihood of finding a job.

b) Search intensity

I now turn to the effects of agglomeration on men's and women's search behavior. The bottom part of Table 6 shows the results.

Urbanization has a negative effect on both men's and women's search intensity, as the log of population size is significantly negative (columns (6.10) and (6.14)).

The negative correlation between the urban agglomeration effects on hazard rates and those on search intensity may seem somewhat surprising, as job-seekers should increase their propensity to search when their chances of finding a job rise. However, in terms of the model presented in Section 2 this could be explained by the fact that in the more densely populated areas cost search increases offset the higher chances of employment. Indeed, the large commuting costs due to density (travelling on congested public transportation, spending time in traffic, etc.) may discourage people from searching even though they have a higher probability of finding a job.[43]

[41] The same can be inferred by comparing the large-city and the super-district coefficients in column (6.1).

[42] Though, as I already noticed, this variable may in fact capture unobservable ability shared by the members of the same family.

[43] Although in the model presented in Section 2 the causality runs only from search intensity to hazard rates (and

In contrast with the outcomes on hazard rates, I find that neither men nor women search any differently in large cities or in the more industrially agglomerated areas (columns (6.9)-(6.10) and (6.13)–(6.14)). Thus, industry localization does not affect men's search behavior, in spite of the fact that it raises their employment probabilities. The district variables do not have any effect on search even with respect to the other small-sized non-industrial LLMs. Also the traditional sector index is always non-significant.[44]

Finally, the higher number of employed people in the family the higher the search effort men exercise to look for a job, consistently with the previous finding on hazard rates.

Also in accord with the results on hazard rates, women do not search any differently in super-districts and are never affected by the thickness of family networks.

c) **Robustness checks**

In this Section I am concerned with whether my data are affected by problems of selection. Indeed, the positive effect of agglomeration on hazard rates and search intensity may derive from some individuals' unobserved characteristics being differently distributed between the large and the small cities.

For instance, the presence of consumption amenities specific to large cities (e.g., cultural events) may attract particularly the most educated people. Also, the extent to which the most able people are paid higher wages (see Di Addario and Patacchini (2004) for empirical evidence on Italy), they may also be better equipped to afford the crowded cities' higher cost of living. If this were the case, the results on hazard rates would be biased upwards, as more able people may also have a higher chance to find a job per unit of search. Alternatively, the more generous government support or the presence of a stronger informal labor market may in fact attract particularly the less able or lazier people to the large cities. In this case, the coefficients on hazard rates and search intensity would be biased downwards.

To deal with these issues, in this section I do two types of exercise.

First, I replicate the same estimations as in the first two columns of Table 6 after having excluded the three largest LLMS (i.e., those containing Rome, Milan and Naples).[45]

not viceversa), an alternative explanation of this finding could be that people need to exert a lower level of search effort to find a job precisely because they have higher chances of employment.

[44] Note that the reason why, contrary to the urbanization case, the industrial agglomeration variables do not show a significantly negative coefficient may reside in the fact that super-districts are characterized by a lower-than-average population density, so that they are less likely to suffer from congestion. In my sample, population density is 408 inhabitants per square kilometer in super-districts, while the corresponding figure in the largest cities is 1,360, by far much higher than the average (584).

[45] These are the three largest municipalities, respectively, in the Center, North, and South of the country. Their LLM population is above 2,400,000 inhabitants. The next most populated LLM, is that of Turin, whose population

Second, I instrument the urbanization variables using a linear probability model and LLM population size in the past as instruments.

The last two columns of Table 6 show the results of the first exercise. The number of observations drops from $25,116$ to $22,332$ in the men's sub-sample, and from $46,131$ to $40,885$ in the in the women's case.[46] The previous results are confirmed to a large extent. Urbanization increases the chances of finding a job and decreases the intensity of search for both sexes. The positive externalities deriving from (sufficiently thick) industry localization are significant only for men (columns (6.3)-(6.4)). The only difference consists in the fact that labor pooling now does not seem to increase the quality of women's matches, implying that this is possibly a phenomenon occurring only in the largest markets.

The outcome of the second exercise is reported in Table 7. The first and third columns show the result of the linear probability model without the endogeneity correction of the urbanization variables. The sign and significance level of the variables of interest are the same as in the previous model (the only exception being family networks, which now do not affect women's hazard rates).

The second and fourth columns report the corresponding outcomes of the instrumental variable estimations. I instrument the log of LLM population density in 1997 with the corresponding one in 1961. The reason of this choice is that these two variables are highly correlated, while the distribution of the individual characteristics not unobserved in 1997 along the large-small city divide should not be the same as in 1961.[47] Similarly, I instrument the large city variable with a dummy equal to one if the 1961 population is greater than $275,000$ inhabitants. This threshold was found to identify large cities in 1961 by Di Addario and Patacchini (2004), with the same methodology they use to define the large city dummy in 1997 (i.e., spatial econometrics applied on Italian LLMs).

Previous results are confirmed to a large extent also in this exercise. However, the urbanization variables in the hazard rate equation are now less significant than before (at the 12–14 percent level, even though they maintain the same signs, while in the search intensity equation the large city dummy becomes significant and negative.

The lower hazard rate coefficients in the IV estimation are consistent with the interpretation that large cities attract the most able people. This interpretation is also in line with the other finding on search intensity, the extent to which more able people exercise less effort than the others

amounts to less than $1,500,000$ inhabitants.

[46]In total, I exclude $2,848$ non-employed people in Rome, $1,835$ in Milan, and $3,530$ in Naples.

[47]It would be if people did not move, but in this case the hypothesis of separated LLMs would hold true, and there would not be any endogeneity issue.

in order to have the same probability of finding a job.

6 Conclusions

In this paper I analyze agglomeration effects on individual search intensity and hazard rates for both Italian men and women. More specifically, I empirically examine whether population size, small-sized manufacturing firm concentration, traditional sector specialization, and size of family networks generate overall net positive or negative externalities. In particular, while agglomeration effects are usually studied either at the urban or at the industry level I am able, by using an Istat algorithm that identifies the more densely industrialized LLMs, to compare urbanization and localization effects.

In general, I find search and matching to be sensitive to both the type and the degree of agglomeration of the local labor market. In particular, I find that urbanization increases job seekers' chances of finding a job independently of their gender, while industry localization raises only those of men.

While these findings hold on average, it is interesting to analyze whether they occur at any level of agglomeration, or only above certain threshold values. Thus, with respect to urbanization I find that positive externalities prevail at any level of population mass. In contrast, industry localization creates positive net economies only in super-districts (as opposed to industrial districts), that is, in the subset of industrial clusters with the highest concentration of small and medium firms in the manufacturing sector; for "regular" districts, there is no significant effect.

Residing in a LLM highly specialized in traditional sectors, on the other hand, always increases hazard rates, for both men and women. This suggests that traditional-sector intensity proxies for labor pooling, which, according to Marshall, improves the efficiency of the matching between jobs and workers.

Finally, the size of family networks, proxied by the number of employed members in the household, increases the probability of finding a job only for men.

As to search intensity, on average it lowers with urbanization and is not affected by industrial agglomeration. A possible explanation of the negative correlation between the urban agglomeration effects on hazard rates and on those on search intensity is that job seekers are discouraged from bearing the higher commuting costs produced by the presence of a large population masse (i.e., travelling on congested public transportation, spending time in traffic, etc.).

Last, consistently having higher chances of finding employment, the men who have larger family networks search more intensively.

Table 1: **Mobility attitudes**

	Employed			
Own municipality	Other municipality in same province	No fixed place	Other province or abroad	
55.2	30.7	6.9	7.1	
	Unemployed			
Own municipality	Daily commuting distance	Anywhere in Italy	Anywhere	
41.3	38.8	14.9	5.0	

Source: authors' elaboration on LFS data.

Table 2: **Average Transition Probabilities**

	Quarterly transition probabilities			
	Employed$_{t+1}$	Unemployed$_{t+1}$	Inactive$_{t+1}$	Total
			Men and Women	
Employed$_t$	96.9	0.9	2.2	100.0
Unemployed$_t$	13.9	62.6	23.6	100.0
Inactive$_t$	3.5	3.9	92.6	100.0
Population composition$_{t+1}$	54.6	5.7	39.7	100.0
			Men	
Employed$_t$	97.5	0.9	1.6	100.0
Unemployed$_t$	17.8	63.7	18.5	100.0
Inactive$_t$	4.9	4.7	90.4	100.0
Population composition$_{t+1}$	68.2	5.3	26.5	100.0
			Women	
Employed$_t$	95.9	1.0	3.2	100.0
Unemployed$_t$	10.4	61.7	27.9	100.0
Inactive$_t$	2.8	3.5	93.7	100.0
Population composition$_{t+1}$	41.5	6.1	52.9	100.0
	Quarterly transition flows			
	Employed$_{t+1}$	Unemployed$_{t+1}$	Inactive$_{t+1}$	Population composition$_t$
			Men and Women	
Employed$_t$	52.4	0.5	1.2	54.1
Unemployed$_t$	0.8	3.7	1.4	5.8
Inactive$_t$	1.4	1.6	37.1	40.0
Population composition$_{t+1}$	54.7	5.7	39.6	100.0
			Men	
Employed$_t$	66.0	0.6	1.1	67.7
Unemployed$_t$	1.1	3.4	1.0	5.4
Inactive$_t$	1.3	1.3	24.3	26.9
Population composition$_{t+1}$	68.3	5.3	26.4	100.0
			Women	
Employed$_t$	38.9	0.4	1.3	40.6
Unemployed$_t$	0.7	3.9	1.8	6.3
Inactive$_t$	1.5	1.8	49.8	53.1
Population composition$_{t+1}$	41.1	6.1	52.8	100.0

Source: elaboration on LFS (January-April 2002). Note: flows are expressed in percentage of the working age population.

Table 3: **Descriptive statistics.**

	Employment rate	Unemployment rate	Job seekers	Activity rate	Hazard into employment
			Italy		
Large city	54.7	10.2	17.0	60.9	24.7
Large city and super-district	63.3	3.7	8.0	65.7	29.5
Small town and super-district	64.6	3.0	11.0	66.6	56.9
Small town - other	54.6	9.8	17.1	60.6	32.5
Industrial district	63.3	3.5	10.5	65.7	51.2
			North-West		
Large city	61.9	5.3	11.5	65.4	36.6
Large city and super-district	63.3	3.7	8.0	65.7	29.5
Small town and super-district	63.9	2.1	6.4	65.2	60.8
Small town - other	62.7	4.3	10.7	65.5	47.5
Industrial district	63.1	3.5	8.9	65.4	48.2
			North-East		
Large city	62.2	3.3	8.7	64.3	55.5
Large city and super-district	–	–	–	–	–
Small town and super-district	65.4	2.4	11.3	70.0	62.7
Small town - other	65.3	4.0	14.2	68.1	55.3
Industrial district	65.1	2.8	10.6	66.0	59.5
			Center		
Large city	59.1	7.3	15.8	63.8	19.3
Large city and super-district	–	–	–	–	–
Small town and super-district	64.2	4.4	13.8	67.2	49.4
Small town - other	55.9	7.3	14.0	60.3	35.6
Industrial district	62.9	4.7	13.8	66.3	47.6
			South		
Large city	42.1	21.4	23.4	53.5	19.8
Large city and super-district	–	–	–	–	–
Small town and super-district	62.5	2.5	13.4	64.1	70.3
Small town - other	45.1	17.5	21.1	54.7	24.8
Industrial district	53.4	5.6	10.5	56.6	38.8

Source: elaboration on the LFS, year 2002. Note that the only LLM that is both a large city and a super-district is that of Desio.

Table 4: **Baseline models for men**

	Hazard to employment				Search intensity			
	Probit		Heckprobit		Probit		Heckprobit	
	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
LLM's job seekers (log)	-0.10	0.000	-0.09	0.000	0.02	0.374	0.01	0.383
LLM's area (log)	0.05	0.154	0.03	0.378	-0.05	0.093	-0.04	0.096
LLM's average extra hours worked	0.05	0.947	-0.14	0.841	-0.54	0.350	-0.54	0.339
LLM's share of overtime workers in total workers	-0.01	0.195	-0.01	0.131	-0.01	0.199	-0.01	0.189
LLM's sector of specialization: high technology	0.15	0.078	0.14	0.078	-0.02	0.722	-0.02	0.749
LLM's sector of specialization: specialization	0.21	0.200	0.19	0.246	-0.12	0.279	-0.12	0.267
LLM's sector of specialization: scale intensive	0.33	0.044	0.29	0.081	-0.12	0.319	-0.12	0.321
LLM's sector of specialization: traditional	0.73	0.023	0.68	0.031	-0.23	0.319	-0.22	0.344
Quarter I (seasonal dummy)	0.06	0.205	0.06	0.183	0.01	0.749	0.01	0.737
Quarter II (seasonal dummy)	0.07	0.165	0.07	0.166	0.07	0.013	0.07	0.014
North-East	0.17	0.088	0.16	0.106	0.12	0.052	0.12	0.053
Center	0.00	0.997	-0.01	0.921	-0.04	0.503	-0.04	0.449
South	-0.18	0.059	-0.14	0.117	0.03	0.612	0.03	0.610
Age	0.01	0.304	0.05	0.000	0.09	0.000	0.09	0.000
Age squared	0.00	0.684	0.00	0.000	0.00	0.000	0.00	0.000
University degree or higher	-0.16	0.102	-0.11	0.196	0.23	0.001	0.24	0.001
High school	-0.15	0.016	-0.16	0.006	0.04	0.296	0.04	0.292
Compulsory education	-0.12	0.047	-0.14	0.016	-0.01	0.697	-0.01	0.770
Past work experiences	0.20	0.001	0.26	0.000	0.10	0.058	0.10	0.039
Search duration: less than 1 month	1.38	0.000	0.80	0.000	-1.02	0.000	-1.01	0.000
Search duration: 1-5 months	0.54	0.000	0.55	0.000	0.13	0.036	0.14	0.033
Search duration: 6-11 months	0.31	0.000	0.29	0.000	-0.05	0.500	-0.05	0.513
Single living alone					0.07	0.246	0.05	0.450
Household head					0.12	0.045	0.08	0.150
Spouse					0.39	0.001	0.35	0.003
Student					-0.15	0.140	-0.25	0.009
Housewife					-1.13	0.000	-1.12	0.000
Other inactive condition					-1.34	0.000	-1.36	0.000
Number of non-working household members					0.01	0.411	0.01	0.271
Constant	-1.63	0.001	-2.20	0.000	0.03	0.946	-0.02	0.956
Number of observations:	25,116		25,116					
of which uncensored:			5,545					

Source: author's elaboration on LFS data. Note: White-robust standard errors adjusted for clustering.

Table 5: **Baseline models for women**

	Hazard to employment				Search intensity			
	Probit		Heckprobit		Probit		Heckprobit	
	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
LLM's job seekers (log)	-0.06	0.010	-0.05	0.014	0.00	0.918	0.00	0.912
LLM's area (log)	0.07	0.095	0.06	0.102	0.01	0.714	0.01	0.730
LLM's average extra hours worked	-0.45	0.496	-0.45	0.481	0.05	0.920	0.05	0.913
LLM's share of overtime workers in total workers	0.01	0.346	0.00	0.435	-0.01	0.138	-0.01	0.133
LLM's sector of specialization: high technology	0.14	0.113	0.12	0.147	-0.06	0.337	-0.06	0.334
LLM's sector of specialization: specialization	0.22	0.157	0.20	0.169	-0.05	0.573	-0.06	0.566
LLM's sector of specialization: scale intensive	0.24	0.246	0.22	0.277	-0.09	0.418	-0.09	0.414
LLM's sector of specialization: traditional	0.99	0.015	0.96	0.015	0.02	0.934	0.02	0.934
Quarter I (seasonal dummy)	0.03	0.536	0.03	0.486	0.04	0.194	0.04	0.180
Quarter II (seasonal dummy)	0.06	0.235	0.05	0.302	0.00	0.978	0.00	0.974
North-East	0.18	0.043	0.18	0.045	0.08	0.128	0.08	0.127
Center	-0.08	0.287	-0.08	0.282	-0.06	0.206	-0.06	0.202
South	-0.34	0.000	-0.33	0.000	-0.04	0.405	-0.04	0.405
Age	-0.04	0.000	-0.03	0.015	0.06	0.000	0.06	0.000
Age squared	0.00	0.000	0.00	0.010	0.00	0.000	0.00	0.000
University degree or higher	0.13	0.166	0.20	0.036	0.19	0.000	0.20	0.000
High school	-0.03	0.719	-0.01	0.903	0.08	0.011	0.08	0.010
Compulsory education	-0.09	0.239	-0.09	0.221	0.01	0.693	0.01	0.714
Past work experiences	0.23	0.000	0.29	0.000	0.13	0.000	0.13	0.000
Search duration: less than 1 month	1.36	0.000	0.95	0.000	-1.05	0.000	-1.06	0.000
Search duration: 1-5 months	0.60	0.000	0.60	0.000	0.15	0.013	0.15	0.014
Search duration: 6-11 months	0.51	0.000	0.51	0.000	0.07	0.208	0.07	0.213
Single living alone					-0.09	0.211	-0.09	0.204
Household head					-0.15	0.003	-0.14	0.005
Spouse					-0.30	0.000	-0.30	0.000
Student					-1.05	0.000	-1.04	0.000
Housewife					-1.27	0.000	-1.27	0.000
Other inactive condition					-0.97	0.000	-0.99	0.000
Number of non-working household members					0.02	0.071	0.02	0.064
Constant	-1.88	0.000	-2.09	0.000	-0.06	0.848	-0.04	0.895
Number of observations:	46,131		46,131					
of which uncensored:			5,731					

Source: author's elaboration on LFS data. Note: White-robust standard errors adjusted for clustering.

Table 6: **Hazard to employment and search intensity (bivariate probit with sample selection)**

	Hazard to employment: men							
	(6.1)		(6.2)		(6.3)(*)		(6.4)(*)	
	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
LLM's job seekers (log)	-0.122	0.000	-0.317	0.000	-0.123	0.000	-0.325	0.000
LLM's area (log)	0.026	0.435	0.002	0.959	0.020	0.551	-0.002	0.957
LLM's population (log)			0.252	0.007			0.262	0.006
Large city dummy	0.165	0.063			0.192	0.041		
Super-district dummy	0.192	0.061	0.151	0.155	0.192	0.062	0.151	0.158
Employed family members	0.051	0.026	0.048	0.038	0.045	0.069	0.041	0.094
Labor pooling	0.714	0.033	0.543	0.083	1.467	0.003	1.192	0.004
	Hazard to employment: women							
	(6.5)		(6.6)		(6.7)(*)		(6.8)(*)	
	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
LLM's job seekers (log)	-0.100	0.000	-0.284	0.001	-0.111	0.000	-0.291	0.001
LLM's area (log)	0.054	0.165	0.033	0.412	0.063	0.119	0.044	0.277
LLM's population (log)			0.259	0.007			0.247	0.011
Large city dummy	0.248	0.004			0.232	0.007		
Super-district dummy	0.099	0.281	0.056	0.549	0.095	0.299	0.055	0.562
Employed family members	0.030	0.303	0.027	0.340	0.033	0.280	0.030	0.315
Labor pooling	0.951	0.021	0.795	0.046	1.092	0.192	0.935	0.221
	Search intensity: men							
	(6.9)		(6.10)		(6.11)(*)		(6.12)(*)	
	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
LLM's job seekers (log)	0.021	0.328	0.127	0.050	0.008	0.709	0.124	0.056
LLM's area (log)	-0.043	0.118	-0.029	0.298	-0.033	0.211	-0.020	0.468
LLM's population (log)			-0.125	0.071			-0.139	0.043
Large city dummy	-0.026	0.689			-0.050	0.425		
Super-district dummy	0.030	0.560	0.048	0.363	0.028	0.583	0.047	0.360
Employed family members	0.039	0.017	0.041	0.013	0.034	0.060	0.035	0.048
Labor pooling	-0.248	0.286	-0.195	0.389	-0.418	0.308	-0.347	0.382
	Search intensity: women							
	(6.13)		(6.14)		(6.15)(*)		(6.16)(*)	
	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
LLM's job seekers (log)	0.000	0.986	0.130	0.012	-0.012	0.466	0.127	0.014
LLM's area (log)	0.008	0.731	0.026	0.270	0.017	0.472	0.035	0.128
LLM's population (log)			-0.148	0.008			-0.162	0.003
Large city dummy	-0.007	0.896						
Super-district dummy	-0.009	0.880	0.013	0.818	-0.024	0.668	0.014	0.804
Employed family members	-0.006	0.678	-0.005	0.760	-0.010	0.857	-0.002	0.876
Labor pooling	0.024	0.921	0.081	0.730	-0.004	0.786	0.031	0.942

Source: author's elaboration on LFS data. Note: White-robust standard errors adjusted for clustering.

(*) Computed on the sub-sample excluding the three largest LLMs.

Table 7: **Robustness checks (OLS, instrumental variables)**

	Hazard to employment: men							
	(OLS)		(IV)		(OLS)		(IV)	
	(7.1)		(7.2)		(7.3)		(7.4)	
	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
LLM's job seekers (log)	-0.037	0.000	-0.035	0.000	-0.112	0.000	-0.079	0.019
LLM's area (log)	0.010	0.291	0.010	0.279	0.001	0.930	0.005	0.606
LLM's population (log)					0.095	0.001	0.058	0.115
Large city dummy	0.052	0.036	0.044	0.136				
Super-district dummy	0.074	0.038	0.074	0.036	0.059	0.105	0.066	0.075
Employed family members	0.012	0.059	0.012	0.058	0.011	0.084	0.012	0.071
Labor pooling	0.215	0.015	0.211	0.016	0.156	0.056	0.167	0.044
	Hazard to employment: women							
	(OLS)		(IV)		(OLS)		(IV)	
	(7.5)		(7.6)		(7.7)		(7.8)	
	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
LLM's job seekers (log)	-0.025	0.000	-0.023	0.003	-0.090	0.000	-0.052	0.160
LLM's area (log)	0.012	0.215	0.013	0.196	0.005	0.649	0.010	0.351
LLM's population (log)					0.088	0.001	0.045	0.281
Large city dummy	0.066	0.003	0.054	0.053				
Super-district dummy	0.043	0.144	0.043	0.143	0.029	0.344	0.036	0.233
Employed family members	0.010	0.207	0.010	0.200	0.009	0.247	0.010	0.207
Labor pooling	0.247	0.021	0.243	0.021	0.195	0.057	0.210	0.038
	Search intensity: men							
	(OLS)		(IV)		(OLS)		(IV)	
	(7.9)		(7.10)		(7.11)		(7.12)	
	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
LLM's job seekers (log)	0.004	0.178	0.008	0.034	0.019	0.041	0.036	0.004
LLM's area (log)	-0.008	0.070	-0.007	0.128	-0.006	0.165	-0.004	0.400
LLM's population (log)					-0.017	0.085	-0.036	0.009
Large city dummy	-0.004	0.647	-0.025	0.052				
Super-district dummy	0.005	0.466	0.005	0.457	0.007	0.293	0.009	0.157
Employed family members	0.003	0.207	0.003	0.200	0.003	0.177	0.003	0.146
Labor pooling	-0.038	0.258	-0.041	0.251	-0.030	0.357	-0.022	0.490
	Search intensity: women							
	(OLS)		(IV)		(OLS)		(IV)	
	(7.13)		(7.14)		(7.15)		(7.16)	
	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.	Coeff.	P-val.
LLM's job seekers (log)	0.001	0.598	0.003	0.174	0.013	0.040	0.016	0.041
LLM's area (log)	0.001	0.814	0.001	0.642	0.002	0.427	0.003	0.366
LLM's population (log)					-0.014	0.049	-0.017	0.048
Large city dummy	-0.001	0.839	-0.012	0.116				
Super-district dummy	0.000	0.985	0.000	0.992	0.002	0.769	0.002	0.715
Employed family members	-0.002	0.257	-0.002	0.271	-0.002	0.294	-0.002	0.307
Labor pooling	-0.006	0.813	-0.008	0.773	0.000	0.986	0.001	0.971

Source: author's elaboration on LFS data. Note: White-robust standard errors adjusted for clustering.

Appendix 1

I reconstructed the LFS longitudinal data with the deterministic method. The loss of observations implied by this method can be due to reporting errors in the household identifier or in the other individual variables (typically, the date of birth), but it can be also due to genuine "attrition": this is the loss of information deriving from the non-availability of some of the people to be re-interviewed at time $t+1$. In what follows I use the term "attrition" for both types of losses.

If the information loss was correlated to working condition changes, attrition would be a potential source of bias for the estimation of labor market dynamics. This typically occurs when people change residence because they find employment in a different location, in which case the exit from the LFS sample is determined by a movement towards employment.

In order to test for the effects of attrition in the estimation of labor market dynamics, I follow the approach proposed by Jiménez-Martín and Peracchi (2003), looking at individuals' survey participation at time t, $t+1$ and $t+4$ (i.e., respectively, one quarter and one year after the first LFS interview). As Jiménez-Martín and Peracchi (2003), I identify two sets of individuals: (1) those participating at all the three surveys (full-time respondents); and (2) those participating at time t and $t+1$ but not at time $t+4$ (non full-time respondents). More formally, let D be an indicator equal to 1 if the person is a full-time respondent and to 0 elsewhere. Non-working individuals at time t can be either unemployed (U) or out of the labor force (O). At time $t+1$ they can be either employed (E), or unemployed (U) or out of the labor force (O). Let π_{ij}^{D} be the probability of moving from state $i = U, O$ at time t to state $j = E, U, O$ at time $t+1$, for an individual whose sample participation is denoted by $D = 0, 1$. Attrition may bias transition probabilities if

$$\pi_{ij}^{0} \neq \pi_{ij}^{1} \tag{11}$$

for $i = U, O$, $j = E, U, O$.

Consider the statistic $l_{ij} = \pi_{ij}^{0} - \pi_{ij}^{1}$. If attrition was not a source of bias for transition probabilities, under the null hypothesis l_{ij} would be equal to zero. In other words, if full time respondents and people who are subject to attrition have the same probability to move towards all the other labor market states then I can assume that attrition does not affect transition probabilities.

Critical values for l_{ij} can be easily derived. Because of the central limit theorem, l_{ij} divided by its standard error has a t-Student's distribution. Rejection at 95 percent significance level, for instance, occurs for values of l_{ij} greater than 2 in absolute value. Table A1 reports the test statistics by gender, age group (15-34 and 35+) and area of residence (North–West, North–East, Center, South). As the table shows, the test results confirm the adequacy of the adopted matching procedure in my study of labor market movements, for all the socio-demographic groups considered.

Table A1. Testing for the effect of attrition

	Men		Women	
	Age 15–34	Age 35–64	Age 15–34	Age 35–64
	North West			
l_{UE}	0.33	-0.10	0.35	0.13
l_{UU}	-0.44	-0.25	0.31	-0.27
l_{UO}	0.07	-0.05	-0.20	-0.28
l_{OE}	0.09	0.02	-0.04	0.00
l_{OU}	-0.06	0.00	-0.03	0.01
l_{OO}	0.03	0.02	-0.06	0.35
	North East			
l_{UE}	-0.39	0.21	-0.91	-0.04
l_{UU}	-0.46	-1.05	0.02	-0.43
l_{UO}	0.57	0.23	0.03	-0.21
l_{OE}	0.15	0.05	-0.07	-0.02
l_{OU}	-0.02	-0.01	-0.10	-0.04
l_{OO}	-1.31	0.12	-0.95	-0.19
	Centre			
l_{UE}	0.07	0.01	0.13	0.00
l_{UU}	-0.10	-0.54	-0.11	0.03
l_{UO}	-0.10	0.65	0.12	-0.44
l_{OE}	-0.01	-0.05	0.02	0.04
l_{OU}	0.03	0.01	0.01	0.02
l_{OO}	-0.73	0.34	-1.00	-0.66
	South			
l_{UE}	-0.06	0.22	-0.09	-0.06
l_{UU}	-0.91	-2.03	-1.14	-0.34
l_{UO}	-0.18	0.00	-0.11	-0.24
l_{OE}	-0.10	-0.01	-0.01	0.02
l_{OU}	-0.29	0.01	-0.20	0.00
l_{OO}	-0.80	0.08	-1.55	-1.38

Source: authors' elaboration on LFS data.

References

Arzaghi, M., and J. Vernon Henderson. 2004. "Networking and Location of Advertising Agencies," mimeo.

Brandolini, Andrea, Piero Cipollone, and Eliana Viviano. 2004. "Does the ILO Definінition Capture all Unemployment?," mimeo.

Broersma, L., and J.C. van Ours. 1999. "Job Searchers, Job Matches and the Elasticity of Matching," *Labor Economics*, 6, 77–93.

Burgess, S. 1993. "A Model of Competition between Unemployed and Employed Job-Searchers: An Application to the Unemployment Outflow Rate in Britain" *The Economic Journal*, 103, 1190–1204.

Cannari, Luigi, Francesco Nucci, and Paolo Sestito. 2000. "Geographic Labor Mobility and the Cost of Housing: Evidence from Italy" *Applied Economics*, 32, 1899–1906.

Cannari, Luigi, and Luigi F. Signorini. 2000. "Nuovi strumenti per la classificazione del sistemi industriali," in L.F. Signorini (ed.), *Lo sviluppo locale*, Roma, Donzelli, pp.123–151.

Ciccone, Antonio, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity," *American Economic Review*, 86, 54–70.

Coles, M.G., and E. Smith. 1996. "Cross Section Estimation of a Matching Function: Evidence from England and Wales," *Economica*, 63, 589–98.

Devine, T.J., and N.M. Kiefer. 1991. *Empirical Economics*, Oxford University Press, New York and Oxford.

de Blasio, Guido, and Sabrina Di Addario. 2002. "Labor Market Pooling: Evidence from Italian Industrial Districts," Temi di Discussione No. 453, Bank of Italy.

Di Addario, Sabrina, and Eleonora Patacchini. 2004. "Is there an Urban Wage Premium in Italy?," mimeo.

Duranton, Gilles, and Diego Puga. 2004. Forthcoming. "Micro-Foundations of Urban Agglomeration Economies," in J.V. Henderson and J.F. Thisse (eds.), *Handbook of Urban and Regional Economics*, Amsterdam and New York: North Holland.

Glaeser, Edward L., and David C. Mare'. 2001. "Cities and Skills," *Journal of Labor Economics*, 19, 316–342.

Helsley, Robert W., and William C. Strange. 1990. "Matching and Agglomeration Economies in a System of Cities," *Regional Science and Urban Economics*, 20, 189–212.

Jimenez-Martin, S., and Franco Peracchi (2002), "Sample Attrition and Labor Force Dynamics: Evidence from the Spanish Labor Force Survey," *Spanish Economic Review*, 4, 79–102.

Lancaster, T. 1979. "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47, 939–956.

OECD. 2002. *Redefining Territories. The Functional Regions*, OECD: Paris.

Paggiaro, A., and N. Torelli. 1999. "Una procedura per l'abbinamento di record nella rilevazione trimestrale delle forze di lavoro," Working Paper No. 15, Istat.

Patacchini, Eleonora, and Yves Zenou. 2003. "Search Intensity, Cost of Living and Labor Markets in England," Discussion Paper No. 3722, CEPR.

Peracchi, Franco, and Eliana Viviano. 2004."An Empirical Micro Matching Model with an Application to Italy and Spain," mimeo.

Petrongolo, Barbara. 2001. "Reemployment Probabilities and Returns to Matching," *Journal of Labor Economics*, 19, 716–741.

Petrongolo, Barbara, and Christopher A. Pissarides. 2001. "Looking into the Black Box: A Survey of the Matching Function," *Journal Economic Literature*, 39, 390–431.

Pissarides, Christopher A. 2000. *Equilibrium Unemployment Theory*, Cambridge: The MIT Press.

Rosenthal, Stuart S., and William C. Strange. 2002. "The Urban Rat Race," presented at the CEPR meeting "The Economics of Cities," London, 6-8 June 2003, mimeo.

Rosenthal, Stuart S., and William C. Strange. 2004. Forthcoming. "Evidence on the Nature and Sources of Agglomeration Economies," in J.V. Henderson and J.F. Thisse (eds.), *Handbook of Urban and Regional Economics*, Amsterdam and New York: North Holland.

van de Ven, Wynand, and Bernard van Praag. 1981. "The Demand for Deductibles in Private Health Insurance. A Probit Model with Sample Selection," *Journal of Econometrics*, 17, 229–252.

Viviano, Eliana. 2001. "Un'analisi critica delle definizioni di disoccupazione e partecipazione in Italia," Temi di Discussione No. 450, Bank of Italy.

Wheaton, William C., and M. J. Lewis. 2002. "Urban Wages and Labor Market Agglomeration," *Journal of Urban Economics*, 51, 542–562.

Wahba, J., and Yves Zenou. 2003. "Density, Social Networks and Job Search Methods: Theory and Application to Egypt," Working Paper No. 3967, CEPR.

Wheeler, Christopher H. 2001. "Search, Sorting, and Urban Agglomeration," *Journal of Labor Economics*, 19, 879–99.