Matching heterogeneous skills demand and supply under limited rationality

David N. MARGOLIS[1] and Jaime MONTANA DONCEL[2]

October 30, 2015
Preliminary, please do not cite

[1]Paris School of Economics, CNRS and IZA. E-mail: David.N.Margolis@gmail.com
[2]Turin university - Collegio Carlo Alberto. E-mail: jaimem.montana@gmail.com.

The idea

What do we know about skills on the labor market?

- Large literature on overqualification
- Unemployment patterns by education suggest possible skills mismatch
- Increasingly large literature on cognitive versus non-cognitive skills suggests that human caiptal is heterogenous
- Literature on development suggests that the types of skills demanded evolve as a country grows
- Literature on ALMPs suggests that people can be taught to search more effectively
- In order to implement correct policies it is necessary to understand how the allocation of workers to jobs occurs *in term of skills*.
- Provide theoretical support and validate implemented policies (i.e. *Bilan de compétences*)

The takeaway

- Skills are heterogeneous
- Different jobs require different skill sets
- Different individuals can provide combinations of skills
- Individuals don't know where they have the best chances to find jobs

We need a model that explains how *heterogeneous skill supply* matches to *heterogenous skill demand*, without requiring individuals to know everything about the labor market around them, i.e. in which they are at least partly *naïve*.

So what does this paper do?

1. Develop a theoretical model where heterogeneous skill supply matches to heterogeneous skill demand
   - Individuals know their own skills.
   - Job postings list number of available positions, salary and skill requirements.
   - Individuals maximize expected utility.
   - Individuals *do not* know the full distribution of skills of those with whom they compete for a job.
   - Firms choose whom to hire among applicants, providing (costly) training if necessary.
   - Unmatched workers and jobs repeat the process until all of (at least) one side of the market is matched.

So what does this paper do?

2. Estimate the parameters of the model using 3 data sets:
   - STEP household data from Colombia
   - Online job postings from Colombia
   - O*Net taxonomy of occupations

3. Simulate policy reforms designed to reduce mismatch
   - Training of unemployed
   - Training subsidy for firms
   - Technological change

Brief literature review

- Probably the most popular framework for modeling labor market matching is Mortensen-Pissarides (1994).
  - 2 states: Employment and unemployment
  - Homogeneous workers and firms within a labor market
  - Matches determined by a "black box" matching function
  - Wages are bargained

- Extensions have added some flexibility
  - Heterogenous (1-dimensional) worker productivity (MP 2001)
  - Informal vs Formal employment (Albrecht-Navarro-Vroman 2009)
  - Self employment (Margolis-Navarro-Robalino 2012)
  - And many more...

Introducing heterogeneity and wage posting

- Miller (1984) was one of the earliest to formalize the occupational choice (heterogeneous jobs) problem
- Moen (1997) uses segmented labor markets to adapt it to the MP framework
- Lazear (2009) adds multidimensional heterogeneity, but in a human capital approach
  - Workers have different skill sets
  - Firms have different returns to skills
  - This generates something that looks like firm-specific human capital
  - There is no consideration of how the workers and firms get matched, only what happens once they are matched

This paper draws on the MP literature, in that it considers the matching problem, and Lazear's approach for valuing heterogeneous skills.

Model Outline

Decisions in the model are taken sequentially:

1. *Firm HR planning:* Firms post job offers (wage and number of vacancies). This occurs only once at the beginning of the game.
2. *Job applications:* Workers decide on the job to which they will apply based on offered wages, the number of open positions, individual endowments and the size of the unemployed pool. They can only apply to one job at a time.
3. *Hiring:* Firms select among applicants. If more than one vacancy is posted, several hires can be made.
4. *Extended durations:* Unmatched vacancies and workers go back into the pool for another round of applications and hiring, starting at stage 2.

Players of the game - Job Seekers

- Job seekers are denoted $i \in \{1, 2, .., I\}$
- The set of endowments for individual i is characterized by a skill vector $\mathbf{s}_i = \left(s_{(1,i)}, .., s_{(K,i)}\right)$, where K is the number of skills
- Skills are heterogeneously distributed among the job seekers, so the vector $\mathbf{s}$ is a multivariate random variable $\mathbf{s} \sim F_s(s_1, .., s_K)$
- I_s represents the set of skills available in the economy

An important feature of the game is that each job seeker owns a set of non transferable endowments:

- Skills can not be exchanged among job seekers, nor do they decrease or increase during the duration of the game

Players of the game - Occupations

The main assumption is that each occupation behaves like a firm.

- The number of firms in the economy is $J \in \{1, 2, .., J\}$
- Each occupation is endowed with a fixed number of vacancies $\mathbf{V} = (V_1, V_2, .., V_J)$ and I_v represents the set of vacancies available in the economy
- Characterized by a production function of the form:

$$y_j = f(\mathbf{s}, \mathbf{r}_j, \mathbf{a}_j)$$

where $\mathbf{r}_j = (r_{(1,j)}, .., r_{(K,j)})$ is a vector of the specific requirements of the skill and $\mathbf{a}_j = (a_{(1,j)}, .., a_{(K,j)})$ is a vector of the importance of the skill (multivariate random variables)

- This specification is a generalization of the firm specific human capital model of Lazear [2009], where we define the value of person i in occupation j as:

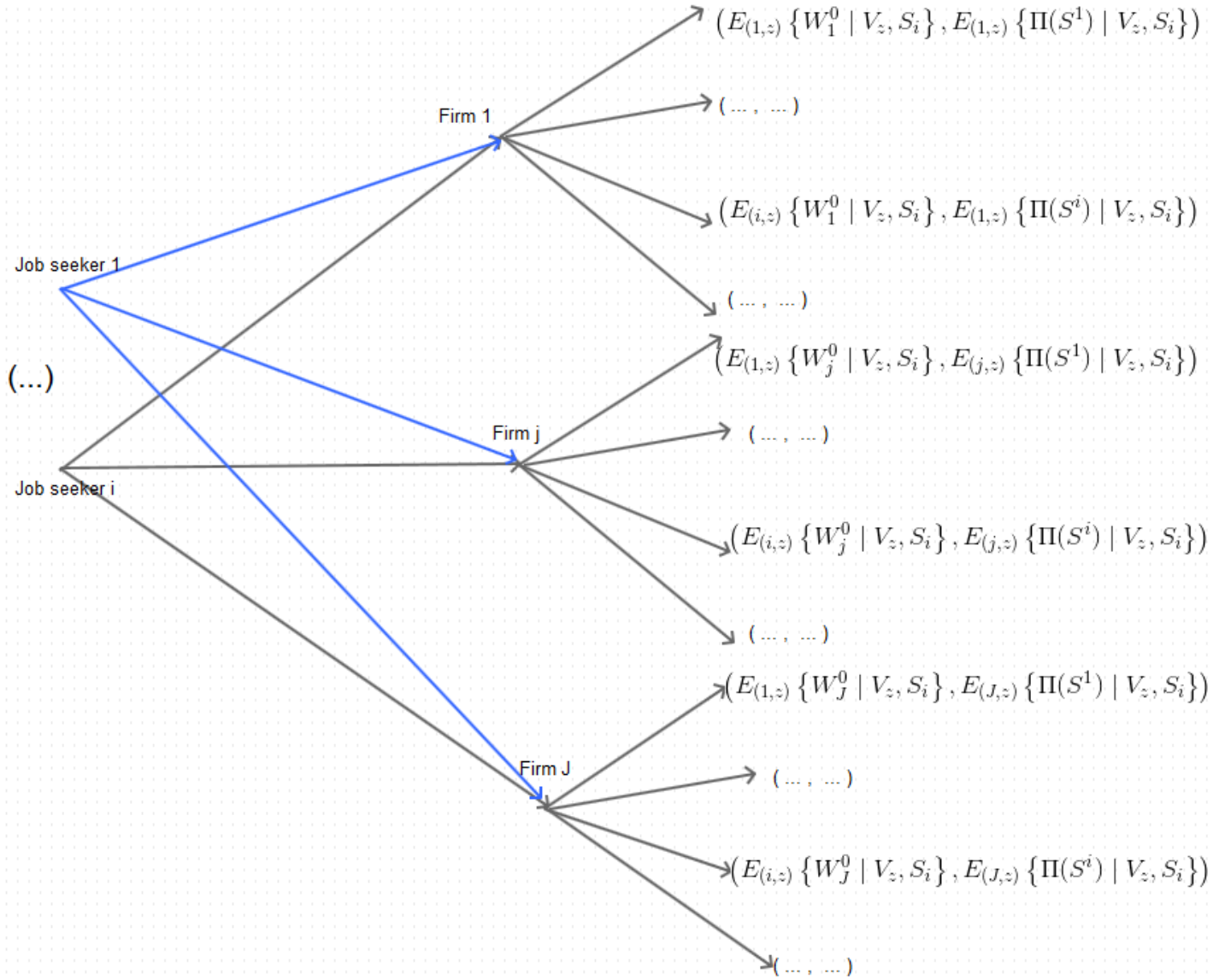
$$q_{(i,j)} = \sum_{k=1}^{K} a_{(k,j)} \left(\frac{s_{(k,i)}}{r_{(k,j)}} \right)$$

Order of Events

0. Firms post the wage for one unit of labor (a job uses all the skill endowment of the agent). $\mathbf{W} = (w_1, w_2, .., w_J)$ is the vector of posted wages (common knowledge). Firms do not compete across occupations for workers through wage posting, but rather maximize their own profits (subject to an output constraint).
1. First stage: Job seekers apply for the job that maximizes their expected wage given the available information in the market. Applying for a job takes time, so cannot simultaneously apply for multiple jobs.
2. Second stage: Firm chooses the candidate that matches best the skills requirements and importance vectors of the firm.
   - If job seeker $\hat{i}$ applied to firm $\hat{j}$ (highest expected utility), and firm $\hat{j}$ chooses job seeker $\hat{i}$ (the highest profit), a match occurs.

The game is repeated from stage 1 until all vacancies are filled or all job seekers have jobs. Denote z as the iteration of the process.

Figure: Structure of the game in iteration z



Optimal unique wage setting

The wage posting problem for the firm is non trivial

- The firm does not know ex ante which workers will apply
- It has specific technology $\Rightarrow$ The value of a hire changes with the characteristics of the person hired

The firm proceeds as if it could compensate each specific skill individually, solving the following cost minimization problem:

$$\min \sum_{k=1}^{K} \omega_k^j s_k$$

s.t.

$$\bar{q} = f\left(\mathbf{s}, \mathbf{r}_j, \mathbf{a}_j\right)$$

Optimal unique wage setting

The $k+1$ first order conditions of the problem are:

$$\frac{\partial L(\cdot)}{\partial s_k} = 0 \Rightarrow \omega_k^j = \lambda f'_{s_k}(\mathbf{s}, \mathbf{r}_j, \mathbf{a}_j))$$

$$\frac{\partial L(\cdot)}{\partial \lambda} = 0 \Rightarrow q = f(\mathbf{s}, \mathbf{r}_j, \mathbf{a}_j)$$

Assuming there is no benefit in production to hiring overqualified candidates (a restriction on $f(\mathbf{s}, \mathbf{r}_j, \mathbf{a}_j)$), the optimal candidate possesses a skill level such that the skills supplied are exactly equal to the skills needed, so for a standard distance function $d : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$, we define:

$$\bar{\mathbf{s}}_j = \min_s \in \{s : d(\mathbf{s}, \mathbf{r}_j) = 0\}$$

For the optimal skill bundle $\bar{\mathbf{s}}$, the optimal wage is:

$$\widetilde{W}_j = \sum_1^k \omega_k^j \bar{s}_k^j = \sum_1^k \bar{\lambda} f'_{s_k}\left(\bar{\mathbf{s}}, \mathbf{r}_j, \mathbf{a}_j\right) \bar{s}_k^j$$

If the firm receives no applicants with enough skills, it will incur training costs, the expectation of which it shifts to the worker through lower wages. The final wage posted is thus:

$$W_j = \widetilde{W}_j - \Delta$$

where Δ is defined by:

$$\Delta = \bar{\lambda} \sum_1^k \left[\int \left(f'_{s_k}\left(\bar{\mathbf{s}}, \mathbf{r}_j, \mathbf{a}_j\right) - f'_{s_k}\left(\mathbf{s}, \mathbf{r}_j, \mathbf{a}_j\right) \right) \bar{s}_k \, dF_s(\mathbf{s}) \right]$$

Note that the posted wage has 2 components:

- The sum of the (weighted) marginal products evaluated at the optimal skill level; and
- The sum of the (weighted) expected marginal products given the skill distribution.

The wage posting process highlights several features:

- Firms value the same skill sets differently. This implies that one would see firm effects in a wage regression, even if job seekers are homogeneous.
- Increasing skills in the population reduces the uncertainty penalty. This implies a positive return to skills (cognitive and noncognitive), which could show up as individual effects.
- Since skills have different values for different employers, some workers are better matches than others.

Consequences for policy:

- Differences in income will persist even if the whole population has an homogeneous set of endowments.
- Training the unemployed is less likely to improve matching than giving subsidizing training at the firm level, since different firms will demand different skills.

Job seeker Utility Maximization

The job seeker knows at each point in time the wages and the number of vacancies available. Assuming risk neutrality and no savings, the job seeker maximizes her utility by maximizing her expected labor income:

$$\max U_i(c_i) = \max U_i \left(E_{(i,z)} \{ W_j \mid I_v, I_s \} \right)$$

$$\max_j E_{(i,z)} \{ W_j \mid I_v, I_s \}$$

Define the occupation $\hat{j}$ for iteration z as the one that maximizes utility:

$$E_{(i,z)} \left\{ W_{\hat{j}} \mid I_v, I_s \right\} > E_{(i,z)} \{ W_j \mid I_v, I_s \} \ \forall \ j \neq \hat{j}$$

Here the expectation is defined with respect to the individual's naive subjective beliefs about the likelihood of getting a particular job if applying for it. There is *no strategic interaction* among agents.

Job seeker - Expected Subjective Utility

As the worker is naive, her evaluation of the expected value of applying to occupation j in iteration z, given the posted wage W_j, is defined as:

$$E_{(\bar{i},z)}\{W_j \mid I_v, I_s\} = W_j p_{(z,v,s)} = W_j \frac{f\left(\mathbf{s}_{\bar{i}}, \mathbf{r}_j, \mathbf{a}_j\right)}{\frac{\sum_i^I f(\mathbf{s}_i, \mathbf{r}_j, \mathbf{a}_j)}{I}} \frac{v_j^z}{\sum_{j'}^J v_{j'}^z}$$

This expectation is comprised of 3 parts:

- The posted wage, W_j
- The worker's assessment of her chances of getting any given job, as measured by the ratio of her productivity to the average productivity of job seekers in that job, $\frac{f\left(\mathbf{s}_{\bar{i}}, \mathbf{r}_j, \mathbf{a}_j\right)}{\frac{\sum_i^I f\left(\mathbf{s}_i, \mathbf{r}_j, \mathbf{a}_j\right)}{I}}$
- The number of jobs available, as measured by the share of vacancies in occupation j with respect to the whole economy .

Firms - Hiring decision

Define the collection of applicants $A_j \subset I$ as the set of individuals for whom occupation j maximizes their utility among available jobs I_v.

The problem of profit maximization is equivalent to solve the following problem, in which the firm selects the most productive candidate(s).

$$\tilde{i} \in \hat{A}_j = \left\{ i \mid f\left(\mathbf{s}_i, \mathbf{r}_j, \mathbf{a}_j\right) = \sup_{i' \in A_j} f\left(\mathbf{s}_{i'}, \mathbf{r}_j, \mathbf{a}_j\right) \right\}$$

Definition of a match

- A match occurs in iteration z, when the payoff to firm $\tilde{j}$ of hiring worker $\tilde{i}$ is maximized while simultaneously the payoff to job seeker $\tilde{i}$ of applying for a job in occupation $\tilde{j}$ is maximized, given the number of vacancies available in each iteration.
- The match is a *stable coalition* since every other occupation different that $\tilde{j}$ decreases the utility of the job seeker for the information set of iteration z. Choosing any other applicant other than $\tilde{i}$ will decrease the profits given the information set of vacancies on z.
- The game is repeated indefinitely until the number of vacancies is equal to 0 or the number of applicants is equal to 0. At the end the employment and unemployment rate is defined.

As the equilibrium of the model is analytically intractable, *the equilibrium is computed using numerical methods*.

Algorithm used to solve the model

- *Construction of a similarity index*. The index synthesizes the value of each worker type in each occupation, combining the different skill dimensions as specified by the $\mathbf{r}_j$ and $\mathbf{a}_j$ for the occupation and the $\mathbf{s}_i$ for the worker type.
- *Construction of a tie breaking index*. The tie breaking index is calculated by estimating a probit model of the probability of employment as a function of the average skill-based similarity index for the individual, the individual's demographic characteristics and other job-related characteristics.
- *Optimization*. Match the individual maximizing the utility according to the behavior described in the above model, using the probit index to break ties.
- *Stopping*. Stop the algorithm when the implied unemployment rate based on unmatched individuals is the same as that in the overall economy.

Supply side data

The second database is the STEP survey (WB). It provides information on the level and proficiency of skills of a representative sample of the Colombian population.
The specific skills that are measured include:

- Cognitive skills (reading, writing and numeracy)
- Socioemotional skills (personality, behavior and preferences)
- Skills related to work (a subset of transversal skills)

The survey samples the working age population (between the ages of 15 and 64), active and inactive.

Supply side data

Table: Some Descriptive Statistics for skills of STEP survey - Active Population

Variable	Mean	Std. Dev.	Min	Max
Read	1.889447	1.005517	0	3
Write	1.223411	0.838479	0	3
Numeric	1.779516	0.830061	0	3
Interpersonal	2.05339	1.174125	0	3
Presentation	0.233002	0.422858	0	1
Supervise	0.338616	0.473367	0	1
Computer	1.340207	1.354008	0	3
Computer type	0.559322	0.850601	0	2
Drive	0.106101	0.308051	0	1
Repair	0.053435	0.224959	0	1
Operate	0.100263	0.300431	0	1

Information on occupation requirements

O*NET is a publicly-accessible on-line taxonomy of occupations, so all the available dimensions of occupations can be accessed trough the web.
This is the source of information for *skill level requirements and skill importance* in each occupation.
The skills in the O*NET database are grouped into two broad categories:

- The basic skills are the ones that facilitate the acquisition of knowledge.
- Cross-functional skills are the ones that facilitate the performance in activities, and thereby the performance of specific tasks inherent to each occupation

The O*NET skill content of the broad categories is divided in 35 skills.

Demand side data

The demand information comes from a vacancy database. It is a collected database using web scraping and contains the vacancies information of the two major job advertisement sites and two public sources, the employment offices and SENA (Public VET provider).

Table: Structure of the Vacancy Final Database

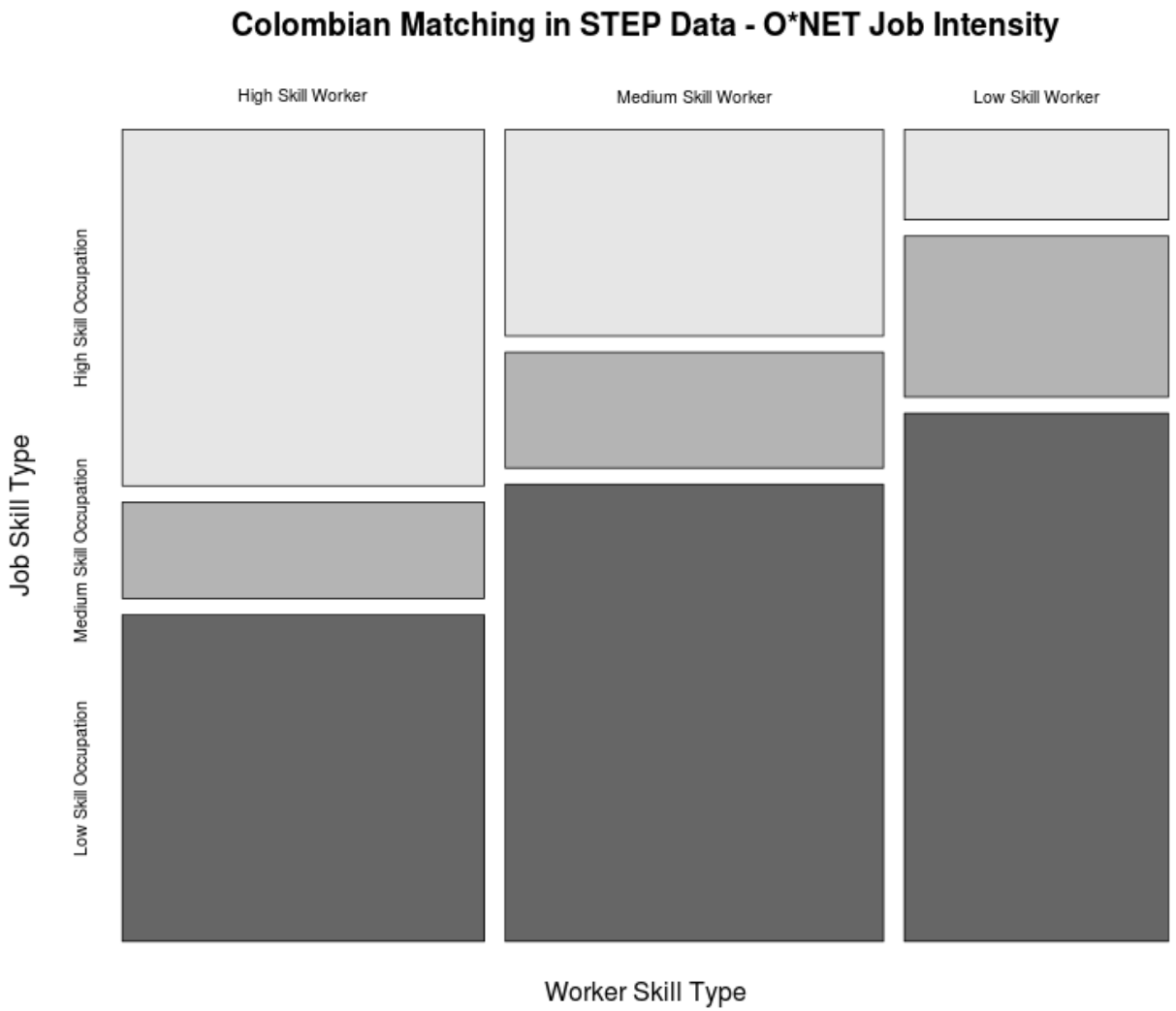
O*NET	Occupation Title	Wages	Number of vacancies	Weight
41-2031	Retail Salespersons	843525.8	324494	4.552409
43-4051	Customer Service Representatives	856134.3	130709	4.77578
41-9011	Demonstrators and Product Promoters	734387.8	92029	5.392699
43-5081	Stock Clerks	749143	63231	5.573521
51-9198	Helpers - Production Workers	745251.9	47480	4.96724
15-1152	Computer Network Support Specialists	1246871	33200	4.882156
41-2011	Cashiers	821517.2	32066	3.648313
15-1131	Computer Programmers	1121887	30627	4.594952
43-3031	Bookkeeping, Accounting, and Auditing Clerks	922857.4	24903	4.935306
43-5021	Couriers and Messengers	743303.8	18867	5.924564

Representativeness of the Vacancy Data

Table: Skill Distribution in Vacancy and STEP Data

Skill Job Type	Vacancies (%)	Filled Jobs (%)
High Skill Job	18.6	29.5
Medium Skill Job	43.9	15.5
Low Skill Job	39.2	55.0

Matching in Colombia



How much mismatch is there in Colombia?

Table: Matching in the Colombian Economy

Job Skill type	Worker Skill Type	Percentage	Underqualification(%)	Overqualification(%)
High Skill Occupation	**High skill Worker**	0.559		
	Medium Skill Worker	0.103	0.441	
	Low skill Worker	0.338		
Medium Skill Occupation	**High skill Worker**	0.288		0.288
	Medium Skill Worker	0.361		
	Low skill Worker	0.351	0.351	
Low skill Occupation	**High skill Worker**	0.275		0.677
	Medium Skill Worker	0.402		
	Low skill Worker	0.324		

Table: Source: STEP and O*NET

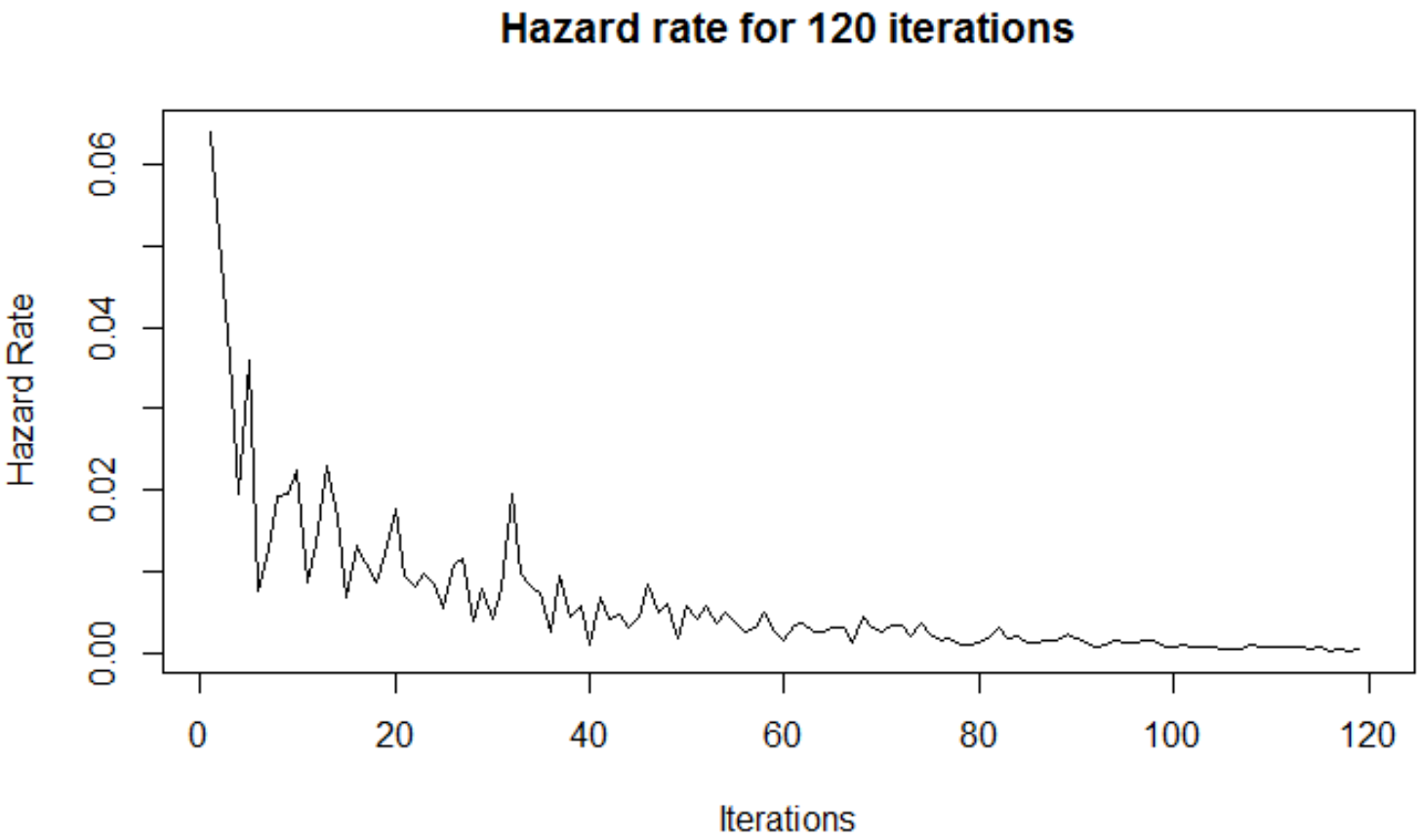
Efficiency of allocation

Table: Results for the base scenario and policy impact evaluation

		Output	Training in the firm	Technological change	Training for unemployed
High Skill Job	**High Skill Worker**	29.02	52.05	52.04	54.77
	Medium Skill Worker	34.23	31.09	31.09	30.92
	Low Skill Worker	36.74	16.86	16.87	14.32
Medium Skill Job	**High Skill Worker**	31.16	34.67	34.67	31.90
	Medium Skill Worker	37.90	39.35	39.35	41.01
	Low Skill Worker	30.93	25.98	25.98	27.09
Low Skill Job	**High Skill Worker**	37.46	20.89	20.89	22.54
	Medium Skill Worker	32.55	31.29	31.29	30.29
	Low Skill Worker	29.99	47.82	47.82	47.16

Declining hazard rate

Figure: Hazard rate



Conclusions

- This paper has presented a model in which skills are multidimensional and skills mismatch occurs as a result of optimizing behaviors of workers and firms
- Numerically solving for the equilibrium allocation shows that although each agent behaves optimally, the socially optimal allocation of workers to jobs is not reached, primarily due to the naive behavior of workers.
- This result helps explain why job search assistance is among the most effective type of active labor market program, as it allows workers to better assess their chances of finding a job and to better target vacancies for job applications.

Extensions / Future Work

- Integrate filled jobs with vacancies on the labor demand side.
- Introduce strategic behavior across firms in wage setting.
- Explicitly model training costs.
- Adjust O*Net skill composition of jobs for Colombia's level of development.
- Introduce a self-employment option.

Thank you!

Define the distance function to measure a skill k for each active worker i and each occupation j in the economy we have:

$$s_{(i,j,k)} = d(score_i, level_j, importance_j) =$$

$$= \begin{cases} \left(\frac{score_i}{level_j}\right) * importance_j & \text{if } score_i \leq level_j \\ importance_j & \text{if } score_i > level_j \\ 0 & \text{if } level_j = 0 \end{cases}$$

The overall skills Index is defined as the sum over all the skills required for the job. In order to calculate the overall similarity I define the overall similarity index $\tilde{S}_{(i,j)}$ of worker i in occupation j as:

$$\tilde{S}_{(i,j)} = \sum_{k=1}^{29} s_{(i,j,k)}$$

Create an index based on skills and demographic information. The technique will take the values of the estimated parameters of a probit as weights of the Index.

$$E\left[Employed|X_i\right] = logit^{-1}(\beta * X)$$

Where:

$$\beta * X = \beta_0 + \hat{\gamma}\bar{S}_i + \hat{\sigma}X_d$$

The vectors X_d contains information about the demographic characteristic of the individuals (age, experience, years of education) and $\bar{S_{i,j}}$ is the average index for person i.

Replacing the values, the index is defined as:

$$O_{(i,j)} = \underset{(0.0015)}{\{\tilde{Indexmean}\}} \times S_{(i,j)} + \underset{(0.0162)}{\{\tilde{age}\}} \times age_i + \underset{(-0.0002)}{\{\tilde{age^2}\}} \times (age^2)_i +$$

$$+ \underset{(0.0097)}{\{\tilde{yearseduc}\}} \times yearseduc_i + \underset{(-0.0009)}{\{\tilde{(yearseduc)^2}\}} \times (yearseduc^2)_i + \underset{(-0.0314)}{\{\tilde{gender}\}} \times gender_i$$

1. The vacancy data is representative of the demand of the Colombian firms. The structure that it provides is a representation of the needs in terms of occupations and wages of the requirements of the Colombian productive sector.
2. Even if the time frame in which the data of the STEP survey and the vacancies were collected doesn't coincide, it should not represent a problem since the economy didn't suffer any particular shock that invalidates the comparison between the two sources.
3. The occupational content of O*NET can be used in Colombia, since the occupation content should be similar for occupations in two different countries. Moreover since there is no data on job content in Colombia is a must, if I want to perform any analysis.
4. The measures for the same skill are comparable between the sources. The score in each can be homogenized and compared since they measure the same dimension.

Unemployment

Figure: Unemployment for different skill levels