Matching heterogeneous skills demand with heterogeneous skills supply

Jaime Montana Doncel

Université Paris I Panthéon-Sorbonne
École d'Économie de Paris - PSE
Research Master in Empirical and Theoretical Economics - MR2ETE
UFR 02 sciences économiques Master Thesis

Matching heterogeneous skills demand with heterogeneous skills supply

Nom du directeur de la soutenance:
David MARGOLIS

Présenté et soutenu par:
Jaime MONTANA DONCEL

Paris - France
June 2015

L'université de paris 1 Panthéon Sorbonne n'entend donner aucune approbation ni désapprobation aux opinions émises dans ce mémoire; elle doivent être considérés comme propre à leur auteur.

Contents

Introduction

The present document using three different data sources, proposes a theoretical and empirical model in which workers choose an occupation based on a given endowment of skills and employer requirements, and employers make a hiring decision taking in consideration several skills of heterogeneous agents. This proposal departs from the common framework because it considers the importance of the skill provision in the labor market, against the common view of an aggregated factor.

The importance of the skills had been evidenced and is part of the growth economics and labor economics literature since long time ago. In economic modeling there has been an association between the skills, knowledge and productivity, treating the concepts indistinctively. This conceptual identification has led to two main outcomes: the production function used and tested only consider one homogeneous factor of productivity (Tinbergen [1974]), or that economic models divide the productivity of workers by type, high skill workers and low skill workers[1], considers skills as an aggregation of abilities and knowledge as input to efficiently produce a good or service.

We can trace back the skill concept, as is commonly understood, in the labor economic discussion to [Becker, 1962], for which the acquired skill take the form of education, on the job training, and physical aspects of the individual (i.e. health status). These factors have a direct impact in income and earnings and the relationship has been tested empirically, being the Mincer work[2] the reference framework to assess the skill level and skill price in labor market.

[1]This division can be traced back to Adam Smith Work, which in the Wealth of nations make the difference between skilled labor and common labor based in three distinctions: agreeableness and disagreeableness, easiness or difficulty of learning how to perform the occupation and constancy and inconstancy of employment (Smith [2001]).

[2]Mincer [1974]

The literature evidence that the acquired knowledge, no matter it source, have direct impact in the productivity of workers. One of the main aims of governments and institutions is to provide the desired level of human capital for the economy to growth and allocate labor in the most efficient way. In this sense the provision of the right skill for the right job, becomes an important matter for education planning and other social services that rely on wage contribution and firm productivity.

Even if in theory this approach fits the data and present no problem, it lacks of of realism, since in practice is way more complex. Employers base their hiring decisions in a non-homogeneous set of skills for each occupation in their firm, and valuate different skills levels depending on the difficulty of the task to accomplish. And even if the education level, the health status and other variables serve as a signal, the measure they use to evaluate if the worker matches the job, rely on the possession of the endowment to be able to perform correctly the task required in the firm.

The present work using three different data sources of the Colombian job market, tries to reveal theoretically and empirically the skill importance in the hiring decision, and provide a baseline to analyze different labor policies that have effect on the skill levels of the working population. The present document analyze such effect in the Colombian labor market outcome.

The present document is divided in four parts. the first part presents the motivation and the relevancy that this document might have for economics. It does so compiling the relevant literature of occupational choice and hiring decision and explaining how the approach presented may be an interesting contribution to the economic debate. The second section presents the data used in the simulation. The third part is devoted to the description of the theoretical model and analyze the economic implications of this approach. The last part present the results, conclusions and further extensions to the presented work. An appendix complement the document with a considerations on the use of the data and a guide for the implementation of the simulation.

Section 1

Motivation

1.1 Search and matching theory

This section provides a succinct compilation of the relevant literature in search and matching theory in order to frame the scope of the proposed empirical model. I will present the ideas that are behind the structure of the model presented and which parts of the model are in common with the models previously established. The present section is divided in three parts: In the first part I present a literature review of the search and matching theory, in the following section I present the relevance of the proposal for the literature and the last section presents the relevancy from a policy point of view.

1.1.1 State of the art

Search and matching theory aim is to describe the process for which in the market under certain conditions agents form stable coalitions that are beneficial for both agents. Search and matching theory has several applications being labor market one of the main areas it has been applied. Its relevancy relies on the fact that it helps to explain many facts that friction-less supply and demand models could not. The main phenomenons that these models explain is the existence of unemployment in equilibrium, worker flows from one state to other, the number of vacancies and employment levels.

The idea behind these models is the existence of trading frictions, which affects the ability of the agents to coordinate and behave as in the classical models, playing an important role so the agents in order to decide if form a coalition

or not take all the possible information and compare the benefit of being in different states. There are two survey of literature for search and matching theory in labor economics: Yashiv [2007] and Rogerson et al. [2004] present a detailed collection of the work on the field until now. I will present in the present chapter the ideas that are relevant and support the construction of the proposed model.

The search and matching theory explore three main ideas: the search process of unemployed and job vacancies, the matching between them and the wage setting. The Mortensen and Pissarides [1994] framework is the standard model from which the matching is modelated in economic literature. The basis of this model are traced back to Diamond [1982] in which the concept of different states for workers is introduced. The existence of those different states is a consequence of the time consuming of production and a hiring behavior change because of stocks. This idea then was complemented by the introduction of a matching function in which a pool of unemployed was matched to a pool of vacancies (under the assumption of one vacancy one firm). The matching function plays the role of a matching technology. In the Mortensen Pissarides framework it takes the form of a Cobb-Douglas technology $f(U,V) = U^{\beta}V^{1-\beta}$. The economic interpretation of such form is based on bargaining power: The firm and worker will bargain the distribution of profits. the parameter β represent the wage power in the bargaining and will define the shares of firm profit. In equilibrium this technology assign to the worker a share of profit equal to the wage, being this larger that an outer option in any other state. The valuation over time of this mechanism will determine the flow of workers between states, and will allow for changes in wages, employment and unemployment. In this framework the equilibrium values depend on a key variable: the market tightness, defined as the ratio of vacancies over unemployment (one of the main results is a decreasing relationship between the unemployment level and the market tightness, consistent with the empirical relationship denoted as the Beveridge curve).

One of the main assumptions in this framework is that it exist a complete labor market, composed by an homogeneous pool of unemployed and an homogeneous pool of vacancies. This assumption is composed by two parts that have been developed in the literature: The homogeneity aspect, for which there has been several extensions but is well documented in Mortensen and Pissarides [2001] in which they include the effect of taxes in the reservation productivity level considering heterogeneity. Subsequently Albrecht et al. [2009] extend the paper

including heterogeneity in productivity for workers and including an informality state for workers to transit in. Margolis et al. [2012] extend the model to include another state, self-employment, and provide an empirical estimation of the parameters of the model for the Malaysian Economy. Heterogeneity is introduced here only in the workers types, but it still considers a complete labor market with a unique pool of vacancies and a homogeneous pool of vacancies.

This characteristic does not allow for the introduction of different kind of jobs, or the introduction of partial markets. The intuition behind this assumption is that a complete labor market assumption does not allow for different job types, or different job requirements, resulting in a unique observable wage in the economy. In order to surpass this barrier many studies focused their attention in different sector or group of jobs, assuming each market as a separated market. One example of this is the work of Stops and Mazzoni [2010], in which the authors analyze the MP model in different occupational groups, conceiving the each occupational group as a family of occupation that share the same kind of qualification requirements.

This feature of the Mortensen Pissarides model for which is possible to calculate the equilibrium only in a complete labor market or in a partial, without considering the correlations between separated markets is a constrain since in the data we observe vertical and horizontal mobility of workers. Moreover the conception of separate markets brings a new challenge since is impossible to model as many unemployment pools and status for workers as groups in the economy, and allowing the flow of people between such markets. Stops [2014] provide a theoretical and empirical demonstration for which the assumption of separated partial labor markets is not a realistic assumption. The approach presented by Stops is also relevant for this work since he bases his analysis in what he calls an 'occupational topology' for which the proximity between the occupational requirements are take into consideration to form the groups, similar to the approach I used in this document for which several occupations are considered. A natural consequence of different markets is different prices, even for the same skill level in different markets with different bargaining settings. A similar idea is proposed by Moen [1997], in which the bargaining in submarkets of the economy generate different wage conformation, independent by the heterogeneity of workers. The problem of this approach is that the number of submarkets is determined by the power of the firm to create submarkets, giving the firm the unrealistic function of

market makers.

The interesting aspect of these setups, in comparison to the Mortensen Pissarides classic framework, is that workers face a decision before the match is evaluated. Each worker in the economy choose a partial market, or in the Moen case a submarket in which the utility is maximized. This topic refers to the occupational choice, for which I present a rule of behavior for agents based on the expected wage.

An interesting aspect of the Moen competitive search equilibrium is that it allows for ex-ante posting wages, result of a trembling hand equilibrium. This factor is really interesting since the vacancy data I will use in the simulation have wage information for each occupation. The mechanism in the Moen paper is based on the ability of Unemployed workers to search only in a fraction of the jobs offered, so they will choose in base of the expected wage. In this sense the model can be interpreted as a sort of occupational choice model for which agents choose the market in which they want to participate following an utility maximization approach. The model here presented is based under a similar construction, taking into account the similarity between the skill endowments, the requirement of the jobs, the number of vacancies and the wage.

Another stream of literature in search models focused on this topic (optimal occupational choice) is due to Miller [1984]. In this model job seekers combine prior information about the characteristics of all jobs in the economy with a sample information of the jobs in which he has gain experience in the past. With this information job seeker forecast and valuates the best option. The theory is based on a construction of an index, representing a valuation function of the present value of the maximal return of a job. The index is set up in time, so is a dynamical index that is computed and compared across all possible occupations. We take this idea in the formulation of the decision model, with the difference that is not valuated trough time as in Miller's paper, but instead trough the iterations of a sequential game. Another difference of the set up presented is that the information set of experience is replaced by a subjective probability based on the number of vacancies and the share of contribution to the firm. In that sense beliefs are included in the presented model in a naive way, without considering strategic behavior. Even if the belief are naive in the construction of the model is assumed that the optimal value is the result of an utility maximization problem.

One of the main challenges to build a matching model is the way to representation of heterogeneous demand. One of the author that incorporate such feature is Lazear [2009], in which the specific human capital is introduced in form of firm specific weights for each skill. That representation is taken to buil the theoretical model presented.

The matching in the Mortensen Pissarides framework depends on the parameters of the matching function. This parameter is exogenous to the model, and in many cases is the function is considered a black box. In the presented setup the variables of the matching are endogenous, and for the simulations all of them are based on observed data.

1.1.2 Relevance of the topic for research

One of the most interesting aspects of the presented methodology is the concern of the dimensionality of the matching. As mentioned previously the matching in the labor literature is based on a single parameter that accounts for productivity or a level of human capital. Firms in the real world consider a set of characteristics and valuate them according to their needs, and is in base to that information that they match workers to jobs. Still the advancements of matching in several dimensions is not an advanced topic in economic literature.

The presented model overpass that constrain comparing the set of requirements of the firm and combine them with the observable characteristics of the individual. The result is interesting since the valuation of the occupational choice and the hiring decision can be based on that measure. Even so, the constrain exist and is interesting to design matching mechanisms that account for multidimensionality and heterogeneity, being this the problem faced by employers in the labor market. This document stress on that need.

Moreover the matching technology form is interpreted as a *black box* by many authors in economic literature (Stops [2014]). The occupational choice model and hiring model set up behavior explain the allocation of workers in firms and explain how this process occurs. This is important since many of the information systems oriented to market must have similar algorithms to show results to job seekers and employers, and is a topic that is not studied in detail, but have

an big effect on allocation.

1.1.3 Relevance for policy design and implementation

One of the main concerns of the policy planning for governments now a day is that the human capital level of the labor force don't match the requirements of the firms. Moreover the skilled workers don't possess the technical, cognitive, and socio-emotional skills to fill current vacancies or create new jobs. The World Bank report on skills Almeida et al. [2012] account that 45% of the current employers can not fill entry jobs and the same amount of youth that work, do so in jobs that don't use their acquired skills.

It does not exist a model that can serve as baseline for evaluation of the different policies that are proposed to solve the skill provision problem. The proposed solutions are based on three aspects, according to the World bank document:

- Investment in the right skill type needed.
- Increase the efficiency of the coordination.
- Training programs for the special targeted populations. The training policy recommendation is based on VET training, on the job training and training related active labor policies.

The presented model evaluates the effect of the training in the firm and present a conterfactual to evaluate the policy outcome. It also present a way to evaluate the training activities for active labor policies. Given that there is no literature on this topic the policy implication of the results are relevant for the program design.

Section 2

Data

2.1 Data Sources

This document uses three different datasets which characterize the Colombian labor market. The characterization is made in terms of occupational structure, skill requirement by occupation and skill endowment by job seeker. Using the three sources of information the document develops an empirical model which allows workers to choose optimally the occupation, and employers to hire the more suitable workers considering the skill level of the individuals.

One of the first questions that arise is why this work is based on the Colombian labor market. The reason of this choice is the author interest for the country and an opportunity window to have detailed data of labor market supply and demand for the country. Being Colombia a developing country, this kind of data is atypical, and makes this research effort more interesting, given the multiple challenges and opportunities that the results may have in terms of policy formulation and implementation.

Given that the main scope of this document, the data that is used to do the empirical simulation comes from different sources, which correspond to the necessity to have demand and supply information, and the requirement of having a characterization of the structure of the job market for Colombia. Making this differentiation, the data-base used are:

0. For the structure of the Colombian job market this document makes use of a vacancy database collected during 2014 in the Colombian Ministry of Labor. The sources of this data are the two major internet job search portals of

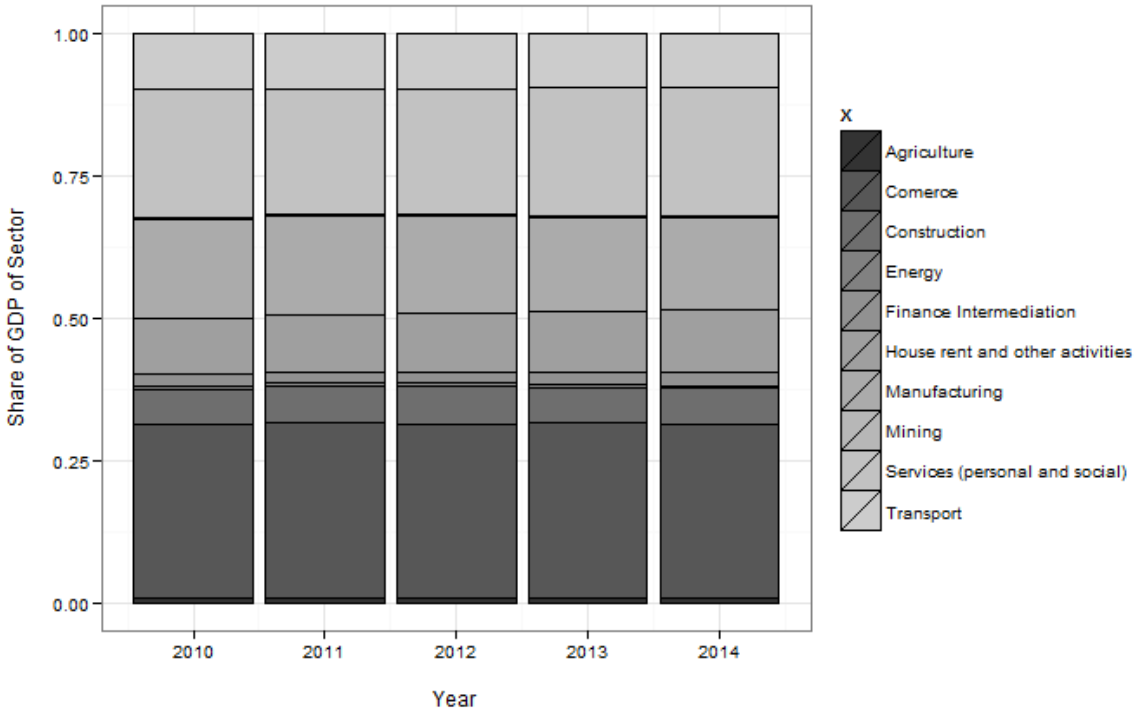
Colombia, the data from the employment agency of the national Vocational Training Institution (SENA – Servicio Nacional de Aprendizaje) and the data provided from the Colombian employment service offices (UASPE – Unidad Administrativa especial del Servicio Público de Empleo).

1. The supply of labor data corresponds to the STEP survey, driven by the World Bank in Colombia in 2012, which has information on the individual characteristics of the labor market. It provides information at the individual level of the education and level training, health status, employment, job skill requirements, personality, behaviors and preferences and language and family background.

2. O*NET is a taxonomy of occupations. One of the dimension it cover is the skill dimension.I use the skills levels and importance for each occupation of the O*NET classification. The exercise was done to the six level digits of O*NET (770 occupations), in order to have a decomposition level that was enough to give idea of a detailed structure of the economy. This document uses the last revision of O*NET database, revised in 2010, using the online information services, documentation and API. The information was attached to the vacancy dataset to characterize the demand in terms of skills.

One of the questions that arises when using collected data is the representativity of the sample. what are the observations that are no collected? Table C.2 in the appendix shows the channels used for job searching in the Colombian market. This information comes from the Household Income Survey (GEIH), and as is shown in the table the collection of the vacancy covers at least more than the 50% of the channels of job searching. Given that the occupational structure must be preserved for all the channels of search, the data is assumed to be representative. Even though arises a problem by the nature of Colombian labor market: informality. Informality in Colombia represent a big share of overall occupation and this collection can not assess the occupational structure of it, since we assume that employers don't use channels to search for job seekers if the nature of the contract violates the law. Nevertheless is assumed to preserve the same occupational structure by two reasons (that still remain as questions and are interesting to research): informality is a consequence of the skill level, and if we consider each occupation a market, the workers with low skilled in each occupation will be informal. The second option has to do with the sensitivity of employers to the hiring taxes, for which the effect must be the same for all occupations, preserving

Figure 2.1: **Occupation by sector 2010 - 2014**



Source: DANE - Household Survey (GEIH)

the structure. The both hypothesis must be tested in order to properly give answer to the question of representativeness of the database. The structure of the occupations by sector does not change much in the period 2010 - 2014 indicating the absence of a structural change in the labor market.

2.1.1 Colombian Vacancy Data

The Colombian vacancy dataset is a sample of the 2014 registries, collected by the Colombian Ministry of labor with the objective of monitoring jobs and job requirements. The idea behind taking only a year subsample is to have the monthly seasonality of job posting during the course of the year, assuming that it will match the seasonality and kind of jobs offered in 2012. The data contain information for 1.892.219 vacancies. The number of registries of the database is less than this, accounting around 1 million registries. This difference is due to the fact than a single job post can offer more than one position opened. In this sense all the analysis made here is based on the number of vacancies offered and not the number of registries.

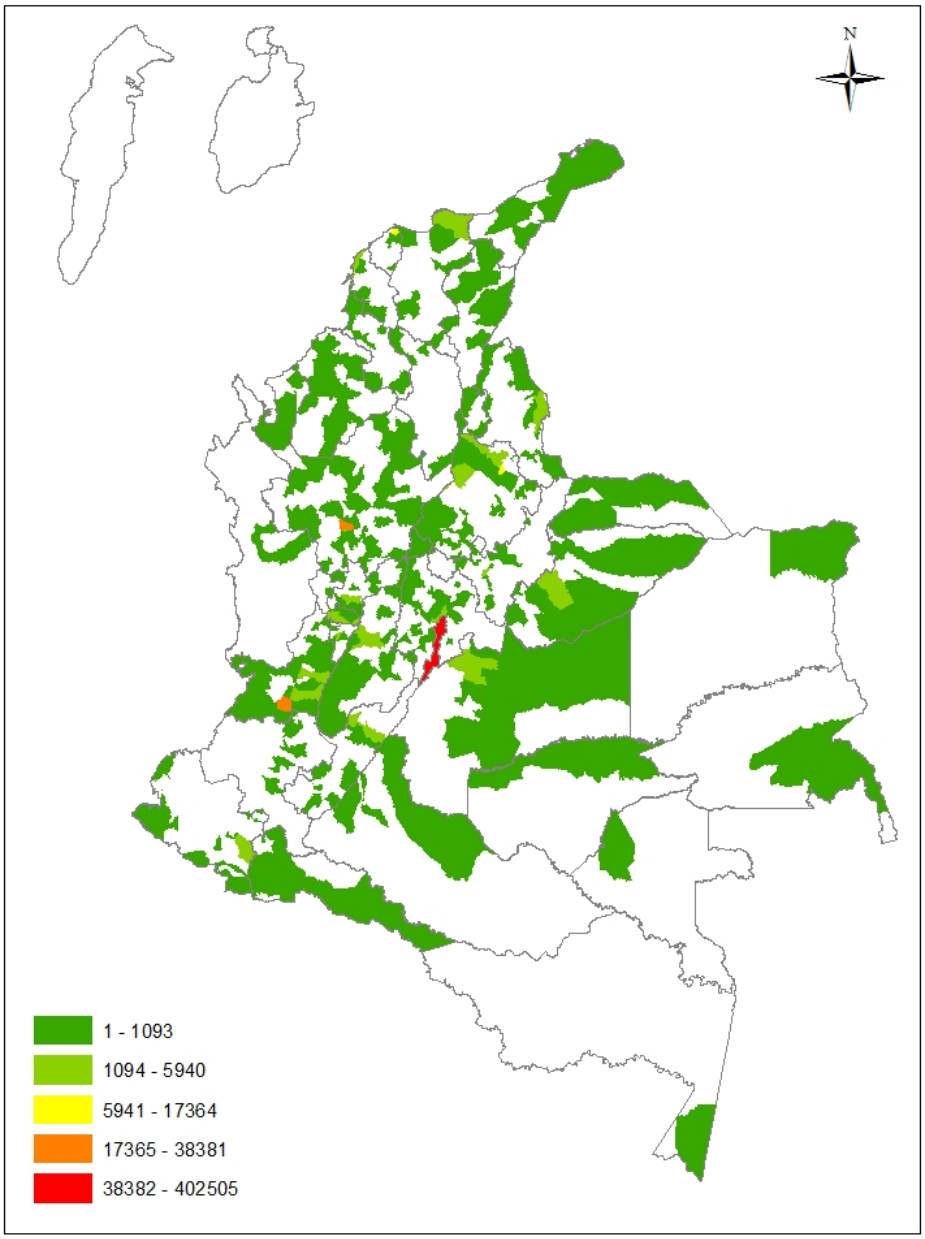
The structure allows to have information on the characteristics of the job offered by place, wage, sector, occupation, etc. Table 2.1 show the variables and the description of data.

Table 2.1: **Database Content - Variables and description**

Variable	Description
ID	Number of the job vacancy Is Requisition ID within the data warehouse. This number is unique and the role is to identify the vacancy within the warehouse.
Title	Is the "title" of the vacancy, ie, the name given to the occupation. This provides information for categorization, clustering and split the basis for the identification of skills and competencies of occupations
Company Name	Company name
Sector	Sector of the company
Position	Area where the person performs
Total years of experience	Total experience required
Experience in the position offered	Total required experience in the position
City	Location availability of vacant
Professional title	It is the title of the person requesting the vacancy ie economist
Wage	payment for work
Level of education	Degree (ie Technical, University, Bachelor)
Type of contract	Type of contract
Language	Language that requires the person to hold the position
Number of vacancies per offer	Number of vacancies that the job post offers.
Publication date	Date of publication of the vacancy
Expiring Date	Expiration date of the vacancy
Description	Description of the occupation
Occupation CIUO	classified occupation with ISCO 08
Occupation O*NET	classified occupation with O*NET

One of the main aspects that incentive me for the use of this data, as mention before, is that is a source to identify the human capital needs of the productive sector in Colombia. An important finding is that the data fits many of the measures that are observed in the surveys and previous studies. For example the wage distribution has the mode between 500.000 pesos and 1.000.000 pesos, also observed in the household survey data (GEIH). The wage distribution by schooling level required, revealing that the increasing in wages occur gradually for non-professional occupations compared to professional. This is especially true for specialized workers (one year post secondary education), master and doctorate levels.

Figure 2.2: **Vacancy distribution in Colombia**



Is also relevant that the years of experience required for the job are concentrated in the rank between one year or less; this added to the fact that the required level of education for the job is completed high school, suggest that an important share of the job offers targets low skill occupations. Regarding the education level required to enter particular job the data shows that 42.3% of jobs require at least completion of high school, 24.8% and 19.5% require technical and technological level respectively. Is worth noting that levels below high school are not in high demand, overall demand for these levels represents 2.59% of jobs, while levels of university and postgraduate added 10.1%.[1]

Regarding the gender preferences the data shows that in 81.18% of the vacancies there is not gender specification. This is also in line with the Colombian legislation, for which there must be no gender discrimination at work and in wage levels, contained in the law 1496 of 2011. However, 3.46% of the vacancies target women only while demand 8.83% of the job vacancies are targeted exclusively for men. Only the 6.53% indicate that the firm and occupation are indifferent considering gender.[2]

In order to conduct the analysis proposed in this paper I construct the occupational structure of the Colombian Economy; the constructed table is constituted by the O*NET Code in the 6 digits code format, the number of vacancies per occupation, the mean wage offered by occupation and a vector that assign a weight (observations) to each occupation to match the number of active working observations of the STEP survey. The weight was calculating taking the average persons that have one year seniority or less in the STEP survey and making the average of that. It was then refined with decimal points to match exactly the observed data. For an overview of the data structure obtain I present the top 10 most requested occupations in the following table.

2.1.2 STEP Survey 2012 - Colombia

For the supply side I use the World Bank STEP survey.The Skills Towards Employability and Productivity program - STEP - survey is a tool aimed to provided measures for quantitative data in order to build comparative databases between countries that allow an assessment of the current state and as a tool for

[1]This information is showed in table A.3 in the appendix.

[2]See table A.1 in the appendix.

Table 2.2: 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
51-6052	Tailors, Dressmakers, and Custom Sewers	720115.6	17817	4.733252

[htb]

policy analysis. The main objective of the STEP survey is to provide measures of the human capital stock of the country, in order to provide a baseline for policy implementation and international comparison.

As discussed, one of the difficulties of measuring human capital is the definition of it. How to define it? Which are the measures that matter for employability and that link human capital to work? The STEP survey is a step forward decomposing in several dimensions the human capital, dividing it in to skills. The main objective of the survey is to provide a skill supply and demand set of information, in order to understand which are the possible gaps for employment and to productivity.The skills selected were picked in based of the relevance for employers and employability, and are based on the relevancy founded in several academic articles (see for example Felstead et al. [2007], John and Srivastava [1999], Heckman et al. [2006]).

The skills that the exercise measure are the cognitive skills (reading, writing and numeracy), socioemotional skills (personality, behavior and preferences) and specific skills related to work (subset of transversal skills). The survey was realize to population in working age, between the age of 15 and 64 years, active and inactive.The implementation of the survey began on March 2012 and the results were processed and cleaned a the final database was published officially in February 2013.

Originally the survey has different modules. The first module correspond to Household level information, in which is collected the information of the household members, their relationship, characteristics (academical and level of literacy)

and labor market status (employed, unemployed or inactive). The first module also contains information about the characteristic of the household size, materials, facilities, appliances, books number and of the income sources of the household. The second module correspond to individual respondent information and covers the information on education and training (quantity and type of education), health status of the individual, employment status, job skill requirements, personality and behavior measures, family background, and the answers to the implemented test (reading literacy).

The methodology for collecting the data of the survey is based on a random representative sample of households in rural and urban areas of the country. The information of the first module is collected to the main household respondent, and collects the income, size and characteristic situation of the household. The second step of the data collection is to randomly select one of the household members and obtain the described individual measures.

For this exercise I'm going to use only of the second module, making special emphasis in the aggregate skill indicators. I will only make use of the active population, since for the sake of the empirical strategy I'm not consider the stock of inactivity. The main descriptive statistics are shown in the table 2.3 for the Colombian case. The results show that in general the variables are distributed near the average for all cognitive skills, are above the mean for the socio emotional skills and are under the average for the transversal skills.

Is should be important to remark how the final measure was taken into account in the moment to analyze the result. Each variable was homogenized in a level between 0 and 100 in order to make them comparable between them and with the other data sets. Given that each observation of the survey is weighted to make it representative, we take the rounded weight in order to specify each observation as one job seeker in the economy, being the rounded weight the number of person of that observation.

Table 2.3: 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
Think	1.289892	1.176287	0	3
Learn	1.820946	1.207189	0	3
Cognitive Challenge	1.557281	0.940956	0	3
Autonomy	2.015327	0.86067	0	3
Physical	1.901545	1.013278	0	3
Extraversion	3.047863	0.640609	1	4
Conscientiousness	3.326227	0.498628	1.666667	4
Openess	3.238473	0.513277	1	4
Emotional Stability	2.543818	0.726405	1	4
Agreeableness	3.176563	0.554637	1.333333	4
Grit	2.990806	0.613184	1	4
Desition making	3.118844	0.599811	1.25	4
Hostile bias	1.710988	0.603815	1	4
Risk	1.640442	1.080305	1	4
Discount				
Genter	0.543931	0.498192	0	1
Age	34.96111	13.16419	15	64

Source: WB STEP Survey Colombia 2012

2.1.3 O*NET - Occupations: skills importance and level requirement

In order to be able to quantify the demand of a given skill I count on information of the O*NET taxonomy. O*NET is a database that contains detail information for 965 occupation on the united States, develop to replace the Dictionary of Occupational Titles (DOT). The project started at 1991, and the idea was to collect detailed information on the different aspects of occupations, in order to be able to describe and analyze them with a quantitative approach. The methodology for collecting the information is based on continuous surveys to employers, research studies by sector and occupation, continuous revision of the estimates and actualization of the information and occupational analysis. The database has information on different occupational dimensions including: tasks, generalized work activities, knowledge, education and training, work styles, work context, skills and abilities.

O*NET is a web database, so all the information on the dimensions that were before mentioned can be accessed trough the web, through web services that are exposed to the public. To gather the information I used the web services provided and construct a request that connect to the database and obtain the information for the skills in table 2.5 for each occupation. The skills in O*NET data base are grouped in basic skills and cross-functional skills. The basic skills are the ones that facilitate the acquisition of knowledge, while the cross-functional skills are the ones that facilitate the performance in activities, so in the performance of specific task inherent to each occupation.

The O*NET skill content is divided in 35 skills, grouped in the mentioned two categories. This categories are subdivided: the basic skills are subdivided in content skills (Reading Comprehension, Active listening, writing, speaking, mathematics and science) and process skills (Critical thinking, active learning, learning strategies, monitoring). The cross-sectional skills are subdivided in social skills (Social perceptiveness, coordination, persuasion, instructing, service orientation), Complex problem solving, technical skills (operation analysis, technology design, equipment selection, installation, programming, operations monitoring, operations and control, equipment maintenance, troubleshooting, repairing, quality control), system skills (judgment and decision making, system analysis, system evaluation) and resource management skills (time management, management of financial re-

sources, management of material resources, management of personnel resources).

The skill taxonomy of O*NET presupposes that skills are the characteristics that an individual has to have in order to perform well different task, and by so the presence of a skill level in an individual, can make him able to perform different activities and be able to perform different occupations. This is one of the key assumption of the analysis presented, since the employers valuate the hiring decision, not in base on the task a worker can perform, but on the general skill level to perform those tasks.

The occupations were grouped in the 6 level digit code, in order to match the vacancy data. The values of each different level in the 8 digit level were average to make possible to have 770 titles. The information contains two different metrics for each skill value: the level, that represent the required skill level required to the employer for perform that specific occupation, and the importance, that is a specific valuation of the skill for the employer.

For the purpose of this document only 29 skills were picked, since one of the main objectives was to have a correspondence between the skills present in the skill survey and the O*NET taxonomy content. The skill that were not used in this analysis were related to resource management, since the measures of the STEP survey don't have similar information.

2.1.4 Consideration and relevant assumption in the use of the data

This section collects the main assumptions for which I justify the data usage, and by so the validity of the result of the simulation on occupational choice and hiring decision, model presented in the next section. The main assumptions for the use of the data are:

0. 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.

1. Even if the time frame in which the data of the STEP survey and the vacancies were collected doesn't coincide, should not represent a problem since the

Table 2.4: Skills requirements descriptive statistics

Variable		Mean	Std. Dev.	Min	Max
Active Learning	Importance	50.83777	12.53559	19	78
	Level	44.0941	11.09776	16	80
Active Listening	Importance	64.34634	11.05332	35	97
	Level	49.09316	9.422747	29	84
Critical Thinking	Importance	61.95727	10.81199	31	94
	Level	49.8022	9.003765	29	80
Learning Strategies	Importance	42.46033	14.45521	3	85
	Level	39.17115	12.08091	0	77
Mathematics	Importance	36.97865	14.25698	0	100
	Level	34.6804	13.40057	0	87
Monitoring	Importance	57.17222	8.987114	31	85
	Level	47.40881	8.207129	27	70
Reading Comprehension	Importance	59.54293	13.79717	25	97
	Level	50.38899	12.08755	20	86
Science	Importance	23.13736	21.57817	0	91
	Level	19.71977	19.99701	0	84
Speaking	Importance	62.90062	12.26975	31	94
	Level	47.85948	10.49566	25	77
Writing	Importance	52.37814	15.2853	10	97
	Level	45.54867	12.31633	7	75
Coordintation	Importance	53.01419	9.221817	25	81
	Level	44.71346	7.161652	27	68
Instructing Others	Importance	44.86891	15.03995	0	91
	Level	40.71048	11.30251	0	70
Negociation	Importance	40.20755	11.79493	13	91
	Level	35.91504	9.609598	12	71
Persuasion	Importance	43.46775	11.7508	16	81
	Level	39.0232	9.774566	14	68
Service Orientation	Importance	47.74338	13.21695	0	91
	Level	40.089	9.048359	2	73

Source: O*NET

Table 3.6: Skills requirements descriptive statistics (part II)

Variable		Mean	Std. Dev.	Min	Max
Social Perception	Importance	54.30171	10.96683	0	94
	Level	43.31731	9.596944	5	84
Complex problem Solving	Importance	53.91795	11.38924	22	81
	Level	43.72311	8.949525	21	73
Equipment Maintainance	Importance	18.32409	21.64172	0	81
	Level	15.28671	18.32055	0	68
Equipment Selection	Importance	18.45868	17.77134	0	75
	Level	14.67052	15.38731	0	57
Installation	Importance	6.146651	12.17319	0	78
	Level	4.656793	10.86593	0	60
Operations and control	Importance	30.94551	22.46198	0	97
	Level	25.40269	18.22881	0	80
Operations and monitoring	Importance	39.58697	19.1671	0	94
	Level	32.42439	14.51982	0	70
Operation Analysis	Importance	27.12225	15.9832	0	75
	Level	24.67528	16.02521	0	73
Programming	Importance	12.34139	11.81543	0	88
	Level	9.484008	11.4184	0	68
Quality control	Importance	35.28261	17.34819	0	78
	Level	30.56915	14.99334	0	57
Repairing	Importance	17.49324	21.80392	0	85
	Level	14.75675	18.47062	0	61
Tech Design	Importance	15.93318	9.947358	0	60
	Level	12.72916	10.517	0	60
Troubleshooting	Importance	26.1647	19.6714	0	81
	Level	22.34657	16.78092	0	75
Judgement Desicion Making	Importance	55.50743	10.28359	25	85
	Level	44.5131	9.250787	23	71

Source: O*NET

economy didn't suffer any particular shock that invalidate the comparison between the two sources.

2. 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.

3. 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.

Section 3

Method

This section aim is to provide a set up for occupational choice and hiring decision based on skills. Occupational choice and hiring decision are in labor market two of the components that allow the efficient allocation of resources between firms. One of the characteristics of this set up is that the model is a first best of rational agents, so there is no strategic interaction between agents. As first best all the information is available in all moments for all the individuals in the economy and there is no uncertainty of the outcome that a choice could take. Also, since the wage information is the average posted information of employers, the price is taken by all the agents and there is no bargain in the match like in other labor economic models(Moen [1997]). The difference of this set up is that the matching is based on the maximization of utility, given the skill endowment of each agent, its similarity with the occupation requirements and the profit maximization of the firm.

The basic matching model conceive labor as one homogeneous factor of production (knowledge, skill, productivity). Being this the foundation of the theory that describe the allocation of workers to firms, matching theory is based in an homogeneous uni-dimensional factor. The most common matching model in labor economic literature is the Mortensten-Pissarides matching technology, for which employers and workers create a cooperative coalition to distribute the rent of the production. The worker receive the rent in form of the wage and the employer receive the rent in form of the perceived profit after the match occurs. The match is a stable coalition when the perceived rent exceed the outside option of job search for the worker and for the firm. One of the main assumption is that each firm is one vacancy, so each industry has one single worker with one single wage.

Several improvement has been made to this theory. Albrecht et al. [2009] construct an extension of the model allowing heterogeneity between workers, and an inclusion of different states other than employment and unemployment(also in Margolis et al. [2012]). All these extensions of the Mortensten-Pissarides framework allow to explain the matching in labor market but are based in the assumption that we can observe productivity, and that this measure that describe the endowment of the individual is a singleton.

In reality firms can not hire in base to this single measure, since each firm value different human capital because it depends on the technology it has. This kind of approach has been introduced previously by Lazear [2009] in which each firm weight different each skill in the production function. Firm specific human capital, is opposed to the conventional view of general human capital, for which human capital augments productivity in the same amount in all firms. Instead the firms valuate different characteristic of the individual, valuating his skill endowment differently (i.e the requirement of job postings in which a set of skills are required for filling the position). Firms valuate characteristics such as education, experience among other characteristics. The other characteristics are the skill endowments that the person posses in order to perform the task of the occupation, demographic characteristics and physical factors (Becker [1962]). In this document we are going to focus in the observable characteristics that the firm can observe in the worker: skill and demographic (age, education, experience).

The following sections provide a detailed presentation of the model, taking into account the relevant economic literature to support the behavior modeling and further simulation.

In the first part we give the outlook of the model and the expected results of it. The second is a detailed description in how is performed the simulation in order to quantify the skill of each agent with respect to each occupation in the economy. The last section describe the algorithm used for the simulation.

3.1 Model

The model behind the matching mechanism I present is a two stage sequential game in which heterogeneous job seekers maximize its utility maximizing the

expected wage among all the possible occupations. Each occupation behaves as a firm and maximize its profit selecting the candidate which has the larger amount of skills among the candidates that apply for that vacancy. When the two conditions hold a stable coalition is formed and the vacancy of the firm will be filled. The process iterates until there are no more available vacancies in the market or when the number of job seekers is equal to zero. In the case the number of job seekers is larger than the number of vacancies, the remaining job seekers will be in unemployment.

This model consider matching as the process for which the skill endowment of the job seekers (supply) form a stable coalition with a firm demanding a set of skill required for the technology of the occupation (demand). In order to model this I present a model that explains the behavior of job seekers and firms. To present the model in a detailed way I present first the set up of the sequential model, followed by the optimal setting of a unique wage with multiple heterogeneous inputs. The next section presents the behavior of the job seekers, the behavior of the firms/occupations and finally the optimality condition behind the match.

3.1.1 Matching mechanism - Sequential game

The structure of the game is a two stages sequential game. The time frame for the complete allocation is equal to 0, so even if the game is repeated more than once the matching is consider momentaneous. The players of the game are:

- The job seekers: the number of job seekers in the economy is $I = \{1, 2, ..., i\}$. Each job seeker owns a set of non transferable endowments (they can not be exchanged among job seekers, nor decrease or increase during the duration of the game). The set of endowments is characterized by a skill vector $\mathbf{s} = (s_1, s_2, ..., s_k)$, where k is the number of skills. The skill are heterogeneous distributed among the job seekers, so the vector $\mathbf{s}$ is a multivariate random variable $\mathbf{s} \sim F(s_1, ..., s_k)$.

- Occupation: The main assumption is that each occupation behaves like a firm. The number of firms in the economy is $J = \{1, 2, ..., j\}$. Each firm is 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 on the production technology of the firm and $\mathbf{a}_j = (a_{(1,j)}, ..., a_{(k,j)})$ is a vector of the importance of the skill in the technology of production.The vectors $\mathbf{r}_j$ and $\mathbf{a}_j$ are multivariate random variables. This specification is an extension of the firm specific human capital Lazear [2009] for which

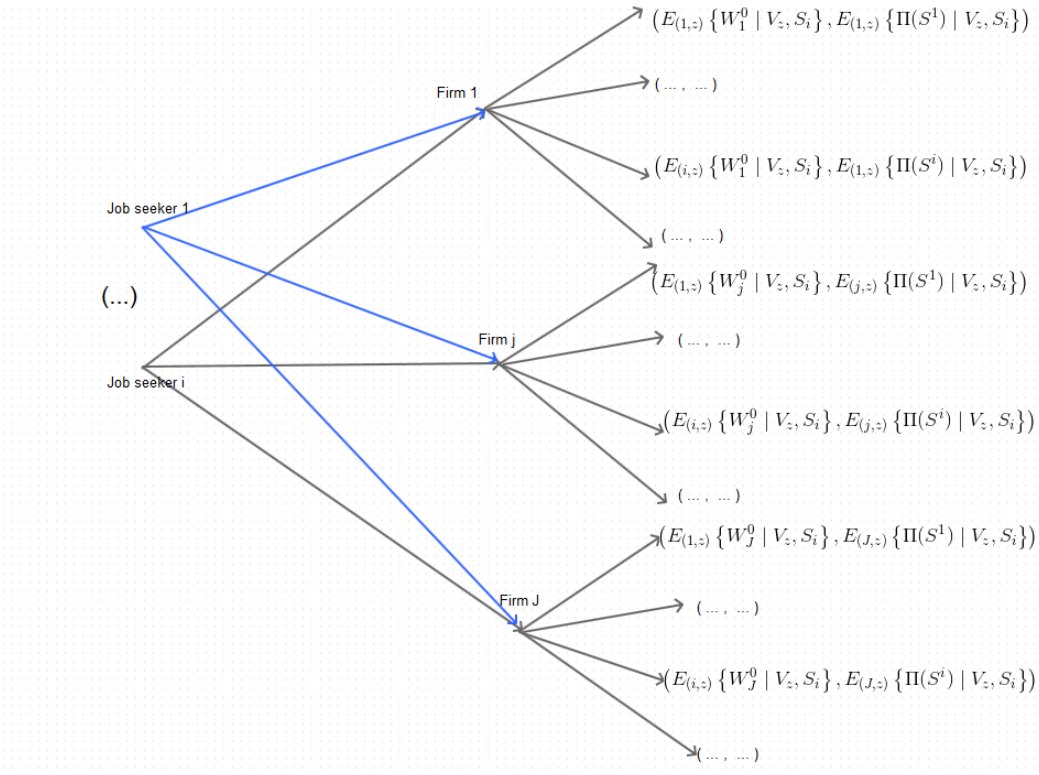
$$q_j = \sum_K a_j^k \left(\frac{s_k}{r_k^j} \right)$$

with $\sum_k a_j^k = 1$. Each occupation is endowed with a fix number of vacancies at the beginning of the game. The vector $\mathbf{V}_j^z = (V_1^z, V_2^z, ..., V_j^z)$ characterize the distribution of vacancies in the economy. This information is common knowledge.

- Order if events:

  0. Each firm posts the wage that is going to pay in exchange of the supply of one unit of labor (wich represent the complete utilization of the skill endowments). This happens only once in the beginning of the game. The wage offered by the firm is common knowledge. The individual will provide to the firm all the endowment of his skills in exchange for a Wage. I denote the offered wage by firm j at the beginning of the game as the vector $\mathbf{W}_J^0 = (W_1^0, W_2^0, ..., W_2 j^0)$. The method how a firm choose a single wage *ex-ante* for a bundle of skills is described in the next section.
  1. In the first stage of the game job seekers want to maximize their utility given the available information in the market. I consider that each job seeker choose only one vacancy in each iteration, so there is a cost associated in the search process that does not allow for multiple candidatures.Each job seeker choose a firm to make his candidature taking into account the information available, choosing the occupation that gives the largest expected wage. This value maximize the utility of each job seeker.
  2. In the second stage of the game each firm chooses among the candidates choosing the job seeker with the larger set of skills $\mathbf{S}$, considering its requirements and importance, $\mathbf{r}_j$ and $\mathbf{a}_j$. Maximizing the set of skills will maximize the profits of the firm given the announced wage W_j^0.

Figure 3.1: Structure of the game in iteration z



One of the assumptions in the model is that the firm which has more vacancies chooses first, followed by the firm that has the second larger number of vacancies and so on. This might be reasonable because of the visibility of the firm with respect to other firms.

3. If the the individual $\hat{i}$ maximize it's utility in firm $\hat{j}$, and firm $\hat{j}$ maximize its profit choosing job seeker $\hat{i}$ we have a match, so a new stable coalition is generated.The new hiring will decrease the number of vacancies available to fill in the firm.

4. The game is repeated until the number of vacancies is equal to 0 for all the firms or the number of job seekers is equal to 0. We denote each repetition as an iteration z.

The next section will explain detailed what is the behavior of each agent during the game and the concept of matching equilibrium proposed. In order to give structure to the exposition I will present first the optimal ex-ante behavior of the firm in presence of heterogeneity. The second part will be devoted to the behavior of the job seekers and the behavior of the firms. A last part will be devoted to the presentation of the equilibrium, and definition of matching.

3.1.2 Optimal unique wage setting in presence of heterogeneity

The problem of the firm to post a unique wage for the job is non trivial since the wage face heterogeneity and has specific technology (in this case represented by the importance and the requirement level). The problem relies in the fact that when a firm maximize its profits the solution would be an optimal allocation for each input that generate a price for each of the inputs in the market. Given that we assume each firm has different levels of requirements and importance valuates differently the human capital in the form of firm specific human capital Lazear [2009]. In order to choose a unique wage we propose the following cost minimization problem for firm j:

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

s.t.

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

Where ω_k^j is the associated price of the k-th skill for the given technology. The associated Lagrangean of the problem is:

$$L(\cdot) = \sum_{1}^{K} \omega_k s_k - \lambda(\bar{q} - f\left(\mathbf{s}, \mathbf{r}_j, \mathbf{a}_j\right))$$

The first $k+1$ order conditions of the problem are:
F.O.C.

$$\frac{\partial L(\cdot)}{\partial s_1} = \omega_1^j = \lambda f_{s_1}^{'}\left(\mathbf{s}, \mathbf{r}_j, \mathbf{a}_j\right))$$

$$(...)$$

$$\frac{\partial L(\cdot)}{\partial s_k} = \omega_k^j = \lambda f_{s_k}^{'}\left(\mathbf{s}, \mathbf{r}_j, \mathbf{a}_j\right))$$

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

Assuming that the optimal candidate for the firm given its technology is the individual that possess a skill level such that the skills supplied are exactly equal to the skills needed, so the distance function $d : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$:

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

For the fixed level $\bar{q}$, and the optimal skill bundle $\bar{\mathbf{s}}$, the optimal wage that ensures the profit maximization is given by[1]:

$$\widetilde{W_j^0} = \sum_1^k \omega_k^j = \sum_1^k \bar{\lambda} f'_{s_k}(\bar{\mathbf{s}}, \mathbf{r}_j, \mathbf{a}_j)$$

Given that the firm faces heterogeneity in the bundle of skill endowments of the job seekers, the final wage posted has included a risk prime to mitigate bad selection, so the firm minimize the costs of selecting a job seeker that the performance is under the desired level. The final wage posted ex-ante by firm j is defined as:

$$W_j^0 = \widetilde{W_j^0} - \Delta$$

where Δ is defined by:

$$\Delta = \sum_1^k \bar{\lambda} f'_{s_k}(\bar{\mathbf{s}}, \mathbf{r}_j, \mathbf{a}_j) - E\left\{ \sum_1^k \bar{\lambda} f'_{s_k}(\mathbf{s}, \mathbf{r}_j, \mathbf{a}_j) \mid I_s \right\}$$

Where the first part of the equation is the optimal skill value previously obtained, and the second part of the above equation is the mean endowment of the job seekers. In such way before the game starts all the occupations J set ex-ante and post a wage for its occupation.

There are two important facts to remark to this solution:

- Each firm valuates different each skill. This can be seen in the fact that the marginal productivity for a given skill k in two firms is going to be different given the firm specific human capital; each firm weight different each skill and prize it differently. This fact is interesting since two persons with the same endowments can have different wages. This fact also present a technical problem that needs to be studied since there is no equilibrium in the skills market.

- Using this setup, even if the working population is skill homogeneous, there will be difference in income. The difference come from the occupational structure, so the heterogeneity of the technology of each firm. This last fact has an implication for policy making and planning because even if the

[1]Note that $\bar{\lambda}$ is fixed since the restriction will not change given the fixed level of production.

workforce is homogeneous and provided with the maximum skills endowment (through education and training) the income difference will prevail because the occupational structure characterized by the firm specific human capital.

3.1.3 Optimal occupational choice

In the first stage of each iteration z the job seekers will maximize their utility. Given that the wage was announced and the job seekers know this information and the number of vacancies in each firm is also common knowledge, the job seeker will maximize its utility choosing the occupation with the highest expected wage. Define the utility of agent i and assuming that all wage is allocated in consumption as:

$$c_i = E_{(i,z)}\left\{W_j^0\right\}$$

The maximization of the utility for job seeker i leads to:

$$\max U_i(c_i) = \max U_i\left(E_{(i,z)}\left\{W_j^0 \mid I_v, I_s\right\}\right)$$

From which we can observe that maximizing the utility of job seeker i is equivalent to maximize the expected wage taking in consideration the available information.

$$\max_j E_{(i,z)}\left\{W_j^0 \mid I_v, I_s\right\}$$

Note that the vector $\left(E_{(i,z)}\left\{W_1^0 \mid I_v, I_s\right\}, ..., E_{(i,z)}\left\{W_j^0 \mid I_v, I_s\right\}\right)$ is the vector of all the strategies for job seeker i. The occupation $\hat{j}$ for iteration z maximizes the utility since:

$$E_{(i,z)}\left\{W_{\hat{j}}^0 \mid I_v, I_s\right\} > E_{(i,z)}\left\{W_j^0 \mid I_v, I_s\right\}$$

In order to calculate the expected value in iteration z, the probability of getting the job in occupation j is defined as:

$$E_{(\bar{i},z)}\left\{W_j^0 \mid I_v, I_s\right\} = p_(z, v, s)W_j^0 =$$

$$= W_j^0 \frac{\sum_1^k \bar{\lambda} f'_{s_k}\left(\mathbf{s}_{\bar{i}}, \mathbf{r}_j, \mathbf{a}_j\right)}{\sum_{\bar{i}\neq i}^I \sum_1^k \bar{\lambda} f'_{s_k}\left(\mathbf{s}_i, \mathbf{r}_j, \mathbf{a}_j\right)} \min\left\{\frac{v_j^z}{\sum^J \gamma_i^z}, 1\right\}$$

The expected wage is a valuation function of the chances of getting a job in round z given the number of firm specific human capital units that the job seeker

can contribute relative to the total provision for the economy in time z, times the wage for the occupation j. The expected wage is composed by two parts[2]:

- The Wage posted by the firm W_j^0. It contains information on the required skill level of the firm.
- $\frac{\sum_1^k \bar{\lambda} f'_{s_k}(\mathbf{s}_{\bar{i}},\mathbf{r}_j,\mathbf{a}_j)}{\sum_{\bar{i}\neq i}^I \sum_1^k \bar{\lambda} f'_{s_k}(\mathbf{s}_i,\mathbf{r}_j,\mathbf{a}_j)}$ which is a measure of the productivity share supplied by the job seeker with respect to the whole economy. In this part is important to remark that this probability soes not account for strategic interaction. Moreover it can be seen as inconsistent since the job seeker consider that every one applies to all firms, even if he knows that is not the case.
- The third part is the share of vacancies in occupation j with respect to the whole economy.

If the job seeker maximizes the set of strategies defined (maximize its expected wage) it maximize its utility. Moreover the optimal choice is individually optimal. Is worth to remark that the behavior of the job seekers is inconsistent [3] and does not consider strategic interaction among the job seekers by endowment level. The result of the model is the first best, the optimal allocation among the agents of the economy.

3.1.4 Hiring decision

The objective of the firm is to maximize its profit level. Given the collection of applicants $A \subset I$ for which it's utility is maximized, the firm will select the candidate with the largest marginal productivity for posted wage set ex-ante. In this way it makes sure that the profit is maximized.

The problem of profit maximization, is equivalent to solve the following problem, in which the firm selects the candidate with the maximum marginal productivity.

[2]In the proposed simulation is considered a modification of the third part ad the probability is defined as:

$$= W_j^0 \frac{\sum_1^k \bar{\lambda}\gamma_i^z f'_{s_k}(\mathbf{s}_{\bar{i}},\mathbf{r}_j,\mathbf{a}_j)}{\sum_{\bar{i}\neq i}^I \sum_1^k \bar{\lambda}\gamma_i^z f'_{s_k}(\mathbf{s}_i,\mathbf{r}_j,\mathbf{a}_j)} \min\left\{\frac{v_j^z}{\sum^J \gamma_i^z}, 1\right\}$$

This is due to the fact that in the data that we are taken from the STEP survey each job seeker has several observations. γ denotes the number of observations of that job seeker type.

[3]Inconsistent because each job seeker only applies for one job, but considers that everyone apply.

$$\max\left(E_{(j,z)}\left\{\sum_{1}^{k}\bar{\lambda}f'_{s_k}\left(\mathbf{s}_{A1},\mathbf{r}_j,\mathbf{a}_j\right)\mid S_A\right\},...,E_{(j,z)}\left\{\sum_{1}^{k}\bar{\lambda}f'_{s_k}\left(\mathbf{s}_{An},\mathbf{r}_j,\mathbf{a}_j\right)\mid S_A\right\}\right)$$

Given the technology proposed, the problem is equivalent to maximize the skill endowment level of the applicants:

$$max_{Ai}\left\{\mathbf{s}_{A1},...,\mathbf{s}_{An}\right\}$$

3.1.5 Definition of a match

A match is the stable coalition for which job seekers choosing firm $\tilde{j}$ maximize it's utility, and the firm maximize it's profit with choosing the job seeker $\tilde{i}$ among the candidates in iteration z.

A match occurs when in iteration z, the payoff of firm $\tilde{j}$ and the payoff of jobseeker $\tilde{i}$ is maximum, given the number of vacancies available in each iteration. Look that the match is a stable coalition since every other occupation different that $\tilde{j}$ decrease the utility of the job seeker for that number of vacancies. Choosing any other applicant other than $\tilde{i}$ will decrease the profits.

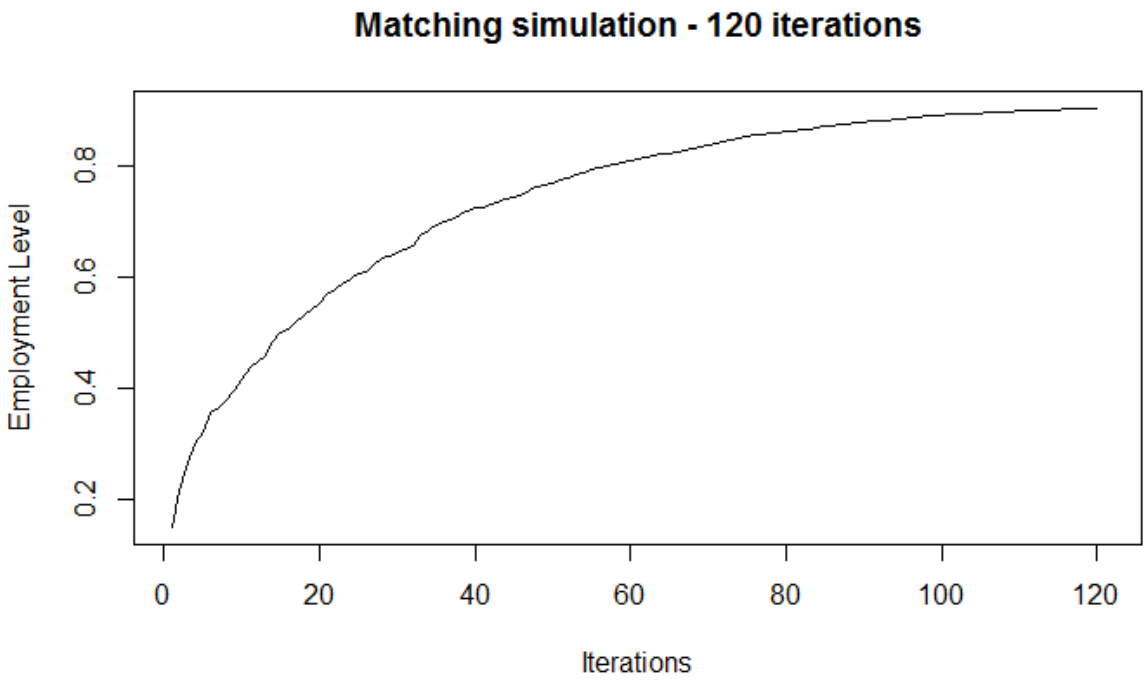
Also, the number of filled positions will be in period z in occupation j: the number of job seekers observed for $\tilde{i}$ if their number is less or equal than the number of vacancies; or the number of filled positions will be equal to the vacancies if the observed job seekers exceed the number of vacancies for that occupation.

3.2 Simulation

Since the equilibrium of the model is hard to compute using an analytical approach, is proposed a simulation that incorporate the behavior of the job seeker and the firm and uses the concept of equilibrium exposed above. The detailed design of the algorithms and how the data was used to calculate the equilibrium is reported in the appendix of the present document. In this section I will provide a brief description of the algorithm used:

- Construction a similarity matrix. Given the demand skill information of the O*NET database and the supply skill information on the STEP survey, we

Figure 3.2: Convergence of the simulation to the Colombian 2014 unemployment rate - 9,6%



define a measure of similarity for each observation and each occupation. The Index is a synthetic index of the different skill dimensions.

- I calculate the mean index for each job seeker, and regress the index on a probit function that determine the weights of a final index that include the demographic information of each job seeker. The results are provided in the appendix of the document. This index contains the information on skill endowments level and the demographic information of the observed worker.
- Match the individual maximizing the utility according to the behavior described in the above model.
- Find the unemployment level.

One of the main purposes on my interest in developing this model is be able to test different policies, and be able to assess the impact in the labor market. This framework make able to test different labor policies, in special the active labor policies that imply an effect in the level of skills.

Two main policies are of interest:

- Training in the firm, that in Colombia is part of the approve National Development plan of the second government of the former president of Colombia and has been implemented with the name UVAES, fr which the firms provide spaces for learning the task of the company, and the national vocational education training institute - SENA - will certificate the competencies of the set of skills learned in the firm for future recognition. **The learning in the firm case is our base model.**

- One of the shocks modeled is the change in requirements (In this specific case is the decrease in the overall requirements level for all occupations). This shock can be interpreted as a technological change, which makes vary the requirement level.

- The second policy of interest is the training for unemployed people. Since from the simulation we can identify the associated labor status, we can identify the unemployed and augment their skill endowment to simulate the effect of education or training. The result of such simulations are reported in the next section.

Section 4

Conclusion

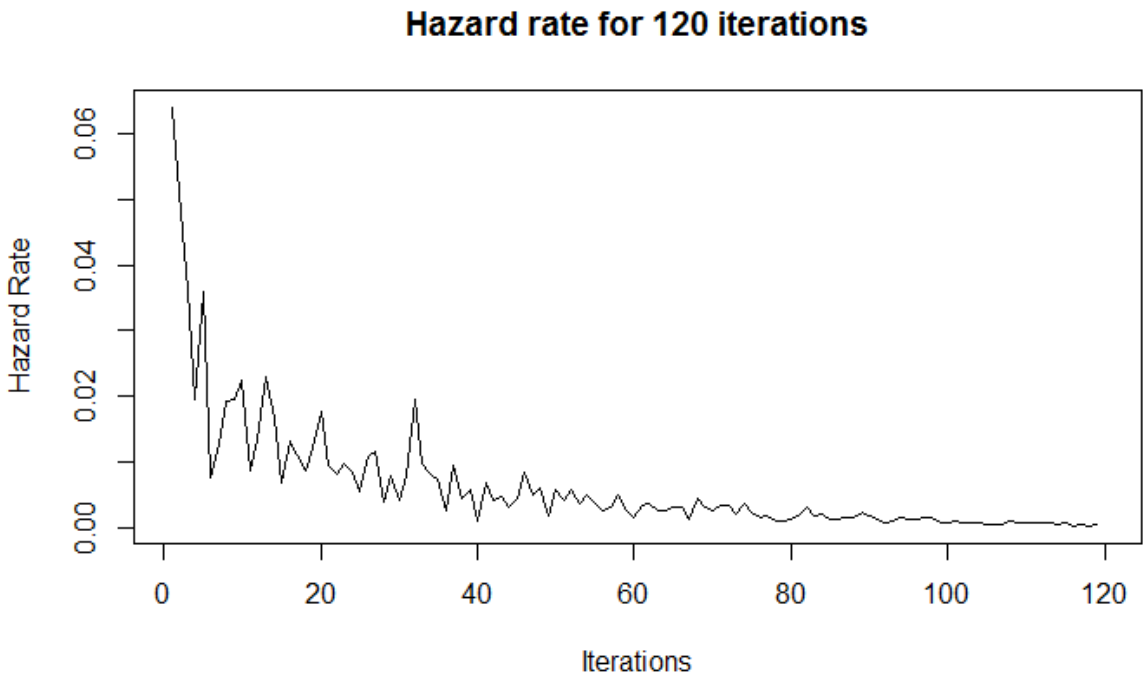
4.1 Results

Table 4.1 summarize the results of the simulation. As we can observe the allocation based on the proposed rules and indicators is individual optimal but not efficient as reported in the table. The reason for this result is that in the simulation the subjective probability valuates the number of vacancies for each occupation, and if the amount of vacancies is really large in comparison to other occupations, the choice will be towards that occupation. In this case near the 10% of the vacancies are in retail or related occupations and that high share make that option very appealing, even for the high skilled job seekers. As can be seen in the graph

From the two policy evaluated the one who has more effect is the technological change. This is mainly because the effect is larger on the whole agents of the distribution, and it reduces the skill gap making the agents compete more in the first simulations making a better assignment in the next rounds. The second policy the effect is small since the new allocation is just a replacement of the less skill employed for more skilled workers.

Making an exploration on the assignation in the first match the most skilled agents are allocated to the most demanded vacancies, and those vacancies are low and medium skill occupations. The other assignations occur in the middle and last part of the simulation. These results make us make two conclusions: even if the rules of the simulation are based on economic principles if we don't include any strategic behavior the result would be individual optimal but inefficient.

Figure 4.1: Variation of the hazard employment rate



The other conclusion has to do with the reason for inefficiency: as result of the technological change simulation we can observe that all the agents increase their skill index, being more significant for low skill agents given the asymmetry of the defined similarity function. This makes that the distribution is contracted, creating more competition and diminishing the allocation of the most skill workers in the medium and low skill occupations. This is a nice result since competition increase efficiency in the model, and is specially a nice result that a less unequal distribution of endowments increase overall efficiency.

Even if the model in its construction is aimed to explain and incorporate many complexities that we observe in the real world, and is successful in doing so, there is a lack of efficiency in the results. The main reason and one of the aims of the future extensions of this document is by the incorporation of strategic behavior.

4.2 Conclusion and extensions

One of the deficiencies of the empirical strategy adopted is that there is not strategic behavior, so even if the algorithm corresponds to an optimal allocation,

Table 4.1: Results for the base scenario and policy impact evaluation

		Baseline simulation		
		Training in the firm	Technological change	Training for unemployed
High Skill Job	High skill worker	33.1	37.5	33
	Medium Skill worker	54.4	13.7	45.7
	Low Skill Worker	12.5	48.8	21.3
Medium Skill Job	High skill worker	58.3	46.6	58.3
	Medium Skill worker	35.5	43	29.4
	Low Skill Worker	6.2	10.4	12.3
Low skill Job	High skill worker	44.3	41.3	44.3
	Medium Skill worker	30.1	29.7	29.1
	Low Skill Worker	25.6	29	26.6

another allocation would be desirable in which the most overall skilled agents are matched to the more skilled jobs. Nevertheless, the incorporation of this behavior is a feasible thing to achieve modifying the constructed set up, since I have a categorization of agents based in their overall skill and demographic index, and the incorporating asymmetry of information and the introduction of different belief to job seekers will modify the optimal occupational choice. This would change the results, giving them a more realistic output.

Several extensions could be made to the framework proposed in this document:

- Construct an agent based simulation, with different beliefs for job seekers depending on their position in the endowment skills distribution, making them consider the competition in a neighborhood of their distribution. Would be also interesting for the employers to change the hiring decision based on the secretary problem, fact that would give uncertainty to the belief of the job seekers.

- Make a modification of the constructed kernel for skills, for which is a pondered of the actual jobs of the economy. This would take a reclassification of the Colombian national household survey to the O*NET, in order to find the number of persons by occupation, measure which allow us to construct a new measure weighting the mean skill index by working population. With this we can apply the empirical strategy for finding the cut off values using the Morstenten-Pissarides framework, and specially the empirical implementation used by Margolis et al. [2012], which finds the effect of subsidies and

taxes for such results.

- Assess the quality of the proposed index using the strategy of Dupuy and Galichon [2014], and see if the Saliency Analysis can support the aggregation of the dimensions for the presented case.

- Apply the Dupuy and Galichon [2014] strategy for assessing the relevance of each skill in the Colombian market. This can give lights to which are the most effective skills to promote.

- Try to make a multi-period simulation, in order to incorporate shocks in vacancies and to see more precisely the effect of the policies evaluated.

The proposed model serve as base framework to introduce these and more interesting extensions.

Bibliography

James Albrecht, Lucas Navarro, and Susan Vroman. The effects of labour market policies in an economy with an informal sector*. *The Economic Journal*, 119 (539):1105–1129, 2009.

R. Almeida, J. Behrman, and D. Robalino. *The Right Skills for the Job?: Rethinking Training Policies for Workers*. Human Development Perspectives. World Bank Publications, 2012. ISBN 9780821387153. URL `https://books.google.fr/books?id=ZTrM4EcAYOMC`.

Gary S Becker. Investment in human capital: A theoretical analysis. *The journal of political economy*, pages 9–49, 1962.

Peter A Diamond. Aggregate demand management in search equilibrium. *The Journal of Political Economy*, pages 881–894, 1982.

Arnaud Dupuy and Alfred Galichon. Personality traits and the marriage market. *Journal of Political Economy*, 122(6):1271–1319, 2014.

Benjamin Edelman. Using internet data for economic research. *The Journal of Economic Perspectives*, pages 189–206, 2012.

Alan Felstead, Duncan Gallie, Francis Green, and Ying Zhou. Skills at work, 1986 to 2006. 2007.

James J Heckman, Jora Stixrud, and Sergio Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Technical report, National Bureau of Economic Research, 2006.

Oliver P John and Sanjay Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138, 1999.

Edward P Lazear. Firm-specific human capital: A skill-weights approach. *Journal of Political Economy*, 117(5):914–940, 2009.

GJ Li and XQ Cheng. Research status and scientific thinking of big data. *Bulletin of Chinese Academy of Sciences*, 27(6):647–657, 2012.

David Margolis, Lucas Navarro, and David A Robalino. *Unemployment Insurance, Job Search, and Informal Employment*. June, 2012.

Robert A Miller. Job matching and occupational choice. *The Journal of Political Economy*, pages 1086–1120, 1984.

Jacob Mincer. Schooling, experience, and earnings. human behavior & social institutions no. 2. 1974.

Espen R Moen. Competitive search equilibrium. *Journal of Political Economy*, 105(2):385–411, 1997.

Dale T Mortensen and Christopher A Pissarides. Job creation and job destruction in the theory of unemployment. *The review of economic studies*, 61(3):397–415, 1994.

Dale T Mortensen and Christopher A Pissarides. Taxes, subsidies and equilibrium labour market outcomes. 2001.

Richard Rogerson, Robert Shimer, and Randall Wright. Search-theoretic models of the labor market-a survey. Technical report, National Bureau of Economic Research, 2004.

A. Smith. *Wealth of Nations*. Hayes Barton Press, 2001. ISBN 9781593775414.

Michael Stops. Job matching across occupational labour markets. *Oxford Economic Papers*, page gpu018, 2014.

Michael Stops and Thomas Mazzoni. Matchingprozesse auf beruflichen teilarbeitsmärkten/job matching on occupational labour markets. *Jahrbücher für Nationalökonomie und Statistik*, pages 287–312, 2010.

Jan Tinbergen. Substitution of graduate by other labour*. *Kyklos*, 27(2):217–226, 1974.

Alejandro Vivas, Stefano Farné, and Dagoberto Urbano. Estimaciones de funciones de demanda de trabajo dinámicas para la economía colombiana. *Archivos de Economía*, (92):39, 1998.

Eran Yashiv. Labor search and matching in macroeconomics. *European Economic Review*, 51(8):1859–1895, 2007.

Appendix A

Tables and figures - Vacancy Database

Table A.1: **Gender demanded by Colombian firms**

Gender	**%**
Female	3.46%
Both Genders	6.53%
Male	8.83%
Does not specify	81.18%

Table A.2: **Experienced required by vacancies**

Experience required	**%**
Less than a year	48.34%
At least 1 year	37.23%
At least 2 year	7.94%
At least 3 year	2.49
At least 4 year	60.0%
At least 5 year	135.0%
At least 6 year	166.0%
At least 7 year	10.0%
At least 8 year	9.0%
At least 9 year	1.0%
At least 10 year	16.0%
More than ten years	3.0%

Table A.3: **Wages offered and educational attainment required**

Wages Offered	None	Elementary	Mid school	High-school	Vocational Education	Advanced Vocational education	Bachelor	Specialization	Advanced Master	PhD
less than $550.000	0.0%	0.9%	3.3%	2.5%	2.4%	1.9%	1.9%	0.4%	0.2%	0.0%
550.001– **1.000.000**	89.9%	80.2%	84.0%	83.1%	69.6%	61.0%	32.0%	2.6%	9.4%	11.3%
1.000.001– **1.500.000**	9.2%	16.0%	6.6%	12.0%	20.2%	26.8%	29.3%	4.3%	8.0%	4.2%
1.500.001– **2.000.000**	0.0%	2.9%	2.7%	1.7%	4.5%	6.2%	15.0%	14.0%	9.9%	1.4%
2.000.001– **2.500.000**	0.0%	0.0%	0.7%	0.2%	1.5%	2.0%	7.9%	21.8%	13.3%	2.8%
2.500.001– **3.000.000**	0.0%	0.0%	0.0%	0.2%	1.0%	0.9%	4.4%	13.4%	20.3%	5.6%
3.000.001– **3.500.000**	0.0%	0.0%	0.0%	0.2%	0.3%	0.4%	4.2%	8.7%	10.2%	2.8%
3.500.001– **4.000.000**	0.0%	0.0%	0.0%	0.0%	0.3%	0.3%	2.9%	11.4%	8.5%	4.2%
4.500.001– **5.500.000**	0.0%	0.0%	2.2%	0.0%	0.1%	0.2%	0.9%	9.4%	4.4%	2.8%
5.500.001– **6.000.000**	0.0%	0.0%	0.5%	0.0%	0.0%	0.1%	0.4%	4.7%	4.6%	38.0%
6.000.001– **8.000.000**	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.7%	5.2%	4.4%	4.2%
$ 8.000.001 and more	0.9%	0.0%	0.0%	0.0%	0.1%	0.1%	0.4%	4.2%	6.8%	22.5%

Table A.4: **Wages offered and years of experience required**

[htb]

Wages	Less than 1 year of experience	At least 1 year of experience	At least 2 years of experience	At least 3 years of experience	At least 4 years of experience	5 years of experience or more
Less than $550.000	1.5%	0.5%	0.0%	0.0%	0.0%	0.0%
550.001– **1.000.000**	38.6%	25.7%	2.4%	0.4%	0.1%	1.9%
1.000.001– **1.500.000**	7.7%	7.4%	1.9%	0.5%	0.1%	0.3%
1.500.001– **2.000.000**	1.3%	2.2%	1.1%	0.4%	0.1%	0.1%
2.000.001– **2.500.000**	0.4%	0.7%	0.7%	0.4%	0.1%	0.1%
2.500.001– **3.000.000**	0.2%	0.4%	0.4%	0.2%	0.0%	0.1%
3.000.001– **3.500.000**	0.1%	0.5%	0.2%	0.1%	0.1%	0.1%
3.500.001– **4.000.000**	0.0%	0.2%	0.1%	0.1%	0.1%	0.1%
4.500.001– **5.500.000**	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%
5.500.001– **6.000.000**	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6.000.001– **8.000.000**	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
$ 8.000.001 and more	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%

Appendix B

Simulation Strategy

B.0.1 Synthetic similarity index

Overall Skill Similarity Index

The idea of this section is to construct a synthetic index for which each agent in the workforce is compared to each occupation. The index is construct to provide information on the similarity of the worker endowments and the relevance of other demographic factors. Exploiting the O*NET structure for which each occupation is consider a unique identity, construct a measure that aggregate the distance between the candidate endowments to the occupation requirements considering the demographic characteristic of the agent.

I divide the construction of the index in two parts: the first part find the distance between the requirement and endowments. The second part use the previous information to find the weights of a final index and calculate it.

Consider an economy composed of i types of individuals $i \in I$, being I the set of the types of individuals that belong to the active population of the STEP survey, so I is defined by $I = \{i \in \mathbb{Z} \mid 0 < i < 1988\}$. Consider J the number of occupations in the economy. The set J is given on the six digit classification structure of the O*NET, that inherit its structure to the classified vacancy data used here. J is given by $J = \{j \in \mathbb{Z} \mid 0 < j < 771\}$, being 770 the number of occupations in the economy. For the sake of this analysis I will consider that each occupational group act like a firm, that want to allocate the number of vacancies available in based on it skills requirements. Also consider k different set of skills that are required by each occupation/firm. The number of possible skills is determined by the O*NET skills used, that in this case is determined by the 29 skills

previously described. K is given by $K = \{k \in \mathbb{Z} \mid 0 < k < 30\}$.

In order to define a similarity index we construct a measure distance defined as $d(x, y, z) : I \times J \times J \to \mathbb{R}^{++}$

$$d(x,y,z) = \begin{cases} \left(\frac{x}{y}\right) z & \text{if } x \leq z \\ z & \text{if } x > z \\ 0 & \text{if } y = 0 \end{cases}$$

This function allow to use the demand and supply information collected in the databases since the STEP survey contains the scores attained for each skill measured in the survey. the O*NET data provides the requirement for each type of skill and provide two metrics: level and importance. Applying the distance function to the measure of skill k for each active worker i and each occupation j in the economy we have, using the observations in the data 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 above distance measure the similarity for required skill k for worker type i in occupation j. The function assign a positive value if the skill level required to perform the occupation j is positive, in other case the value is 0. The measure function is not symmetrical with respect to the score and the level, since scores over the required level will obtain the same measure. The reason to built the function in this way is that workers will only contribute the required level of skill, even if they possess more endowments. The similarity between the skill requirement and skill endowment will increase as $s_{(i,j,k)}$ becomes more positive.[1]

The overall skills requirement 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:

[1]One important remark about this specific function is that does not penalize and allows for matching under the expected level of the firm; the interpretation of this fact is that the firm allows a match and assume the cost of training in the firm. A more restrictive set up would involved a larger unemployment rate and possibly unfilled vacancies.

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

Define $\mathfrak{M}_{(J,I)}$ the matrix containing all the similarity indexes for all the worker types and all the occupations in the economy. Given the matrix the rows contain the distribution of workers based on the skills endowments and the requirement of the occupation, and given that the number of people for each type of worker is known, I can calculate the kernel for each occupation of the economy.[2]

$$\begin{bmatrix} \tilde{S}_{(1,1)} & \dots & \tilde{S}_{(1,1987)} \\ \dots & \dots & \dots \\ \tilde{S}_{(770,1)} & & \tilde{S}_{(770,1987)} \end{bmatrix}$$

The above matrix would allow me to perform an empirical analysis as the one presented by Margolis et al. [2012], considering each occupation one market, in which the cutoff values for the share of employment an unemployment are a function in this case of the similarity index, element that determine the Nash bargaining equilibrium and wage. Never the less this approach would present some theoretic challenges, since in the Morstenten-Pissarides model have for each market one unemployment pool by market, and this approach would lead to to have several, a difficulty that complicate the flow of agents between markets. Given this technical impasse the best option is to execute a simulation based on the belief behind occupational choice and hiring decision. To do so we proceed with the construction of the index.

Overall Skills and Demographics Similarity Index

In order to construct an final index that incorporate the skill level and the demographic attributes of each worker type, we construct based on the marginal of a probit regression an equation that has as weights the estimated parameters of the model. The interpretation of each weight is the importance that employers

[2]Would be more precise to do the reproduction of the exercise contained in the Margolis et al. paper to use the occupied occupational structure of the household income survey classified to to the O*NET. This requires the re-classification of the raw data of the survey.

Figure B.1: **Example - Distribution of overall similarity index of active population for Chief Executives Occupation**

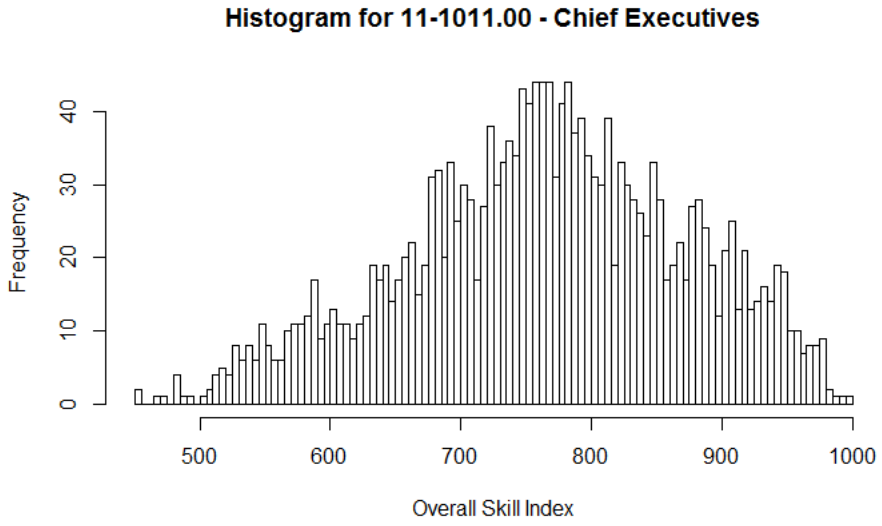
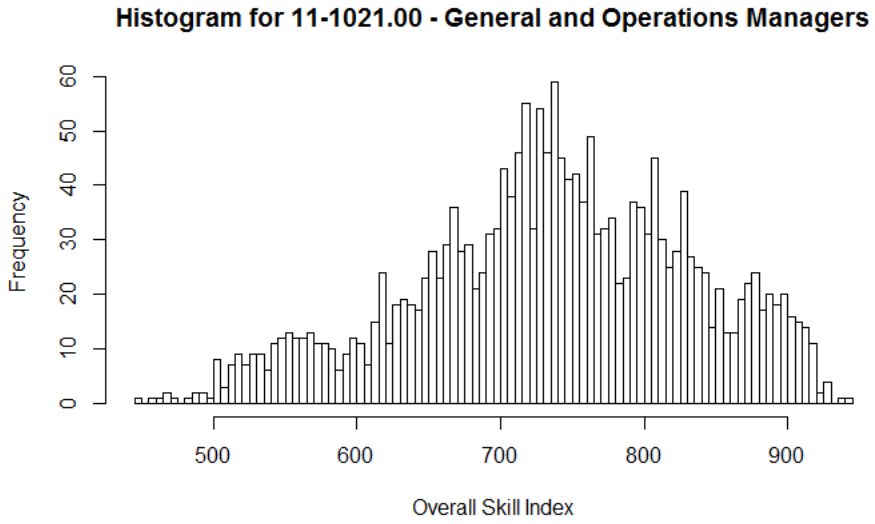


Figure B.2: **Example 2 - Distribution of overall similarity index of active population for General and operations managers**



give to each dimension in the moment of hiring a job candidate.

In order to do so first we have to construct a measure of the overall skill level for each individual. Define the vector $\mathfrak{A}_I$, as the vector containing as elements the average overall index similarity for person i [3].

$$\mathfrak{A}_I = \left(\frac{\sum^J \tilde{S}_{(1,j)}}{770}, ..., \frac{\sum^J \tilde{S}_{(i,j)}}{770}, ..., \frac{\sum^J \tilde{S}_{(1987,j)}}{770} \right)$$

This measure can be interpreted as the overall similarity of the worker with the labor market demand. The $\mathfrak{A}_I$ vector applications would be an important feature to consider. This vector, along with the STEP survey weights allow to calculate the kernel distribution of skills for all the active population of the economy. Since We don't have anymore the problem of different unemployment pools, but this kernel is for the whole economy, the Mortensten-Pissarides frame work can be used. Using the same approach of the VMR() paper we can have the results of the different states of the labor market (unemployed, employed, informal, self employed) with the cut off based on the observable skill endowments.

Coming back to the analysis the idea is to find the fix weights that based on the levels of the variables determine the overall skill and demographic index. After merging the $\mathfrak{A}_I$ in the STEP survey, I use a probit regression of the form:

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

where:

$$\beta * X^i = \beta_0 + \hat{\alpha} X^i_m + \hat{\gamma}\mathfrak{A}_I + \hat{\sigma} X^i_d$$

The independent variable of the model X_i is a vector that contain information of the status level of the worker in the active population (employed or unemployed), X^i_m is a matrix that contain information about the market preference for employers, for example a specific technology trend not captured in the skills (I put this in the regression since I consider that is important for the skill construction, but I don't have any variable to include) and X^i_d is a matrix that contain the demographic characteristics of the individual (for this specific estimation I use age, age squared, years of education, years of educations squared and the gender). In table B.1 are presented the results of this estimation, with the

[3] Is the average of each column of the $\mathfrak{M}_{(J,I)}$ matrix

probit effects.

Replacing the values calculated in the Overall Skills and Demographics Similarity Index $O_{(i,j)}$, the index is defined as:

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

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

Define $\mathfrak{O}_{(J,I)}$ the matrix containing all the Overall Skills and Demographics Similarity Indexes for all the worker types and all the occupations in the economy.

$$\mathfrak{O}_{(J,I)} = \begin{bmatrix} O_{(1,1)} & ... & O_{(1,1987)} \\ ... & ... & ... \\ O_{(770,1)} & & O_{(770,1987)} \end{bmatrix}$$

The matrix $\mathfrak{O}_{(J,I)}$ represents a characterization of all the individuals of the economy. If a basic rule based only in skill level is made one could think that agents choose an occupation for which the skill and demographic index of the worker is maximum. The firm (that in this case is each occupation) will choose the best worker type, maximizing upon the candidates index (the workers type that choose that occupation). A stable coalition is the one for which the worker choose by the firm chooses that occupation. In such situation we will have a match. Under this intuition I construct a set of rules for which the mach happen, taking into account the behavior of agent and firms (considering one occupation as one firm).

B.0.2 Simulation

In order to find the efficient allocation of workers to jobs in necessary to create a job matching set of rules to describe the behavior of the agents in the economy. A match that is based only on the job specific capital that a worker provides is considered as stable since the separation is very destructive, because any other choice would give him a lower return. The index constructed in the previous section provides a measure of job specific capital and aggregate to it the

Table B.1: Probit regression of skills and demographic characteristics

	Dependent variable:	*Marginal Effects:*
	employed	
Indexmean	0.009^{***}	0.0015
	(0.001)	
age	0.093^{***}	0.0162
	(0.020)	
age^2	-0.001^{***}	-0.0001
	(0.0003)	
yearseduc	0.056	0.0097
	(0.054)	
$(yearseduc)^2$	-0.005^{*}	-0.0009
	(0.003)	
gender	-0.180^{**}	-0.0314
	(0.084)	
Constant	-5.838^{***}	
	(0.581)	
Observations	1,880	
Log Likelihood	−576.873	
Akaike Inf. Crit.	1,167.746	

Note: *p<0.1; **p<0.05; ***p<0.01

Figure B.3: **Example 1b - Distribution of Overall Skill and Demographic Similarity Index of active population for Chief Executives Occupation**

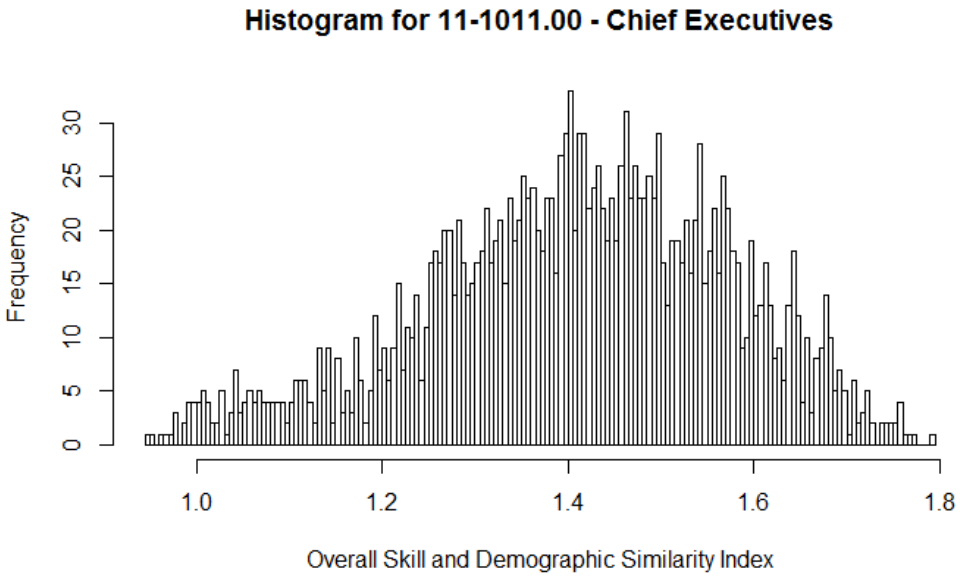
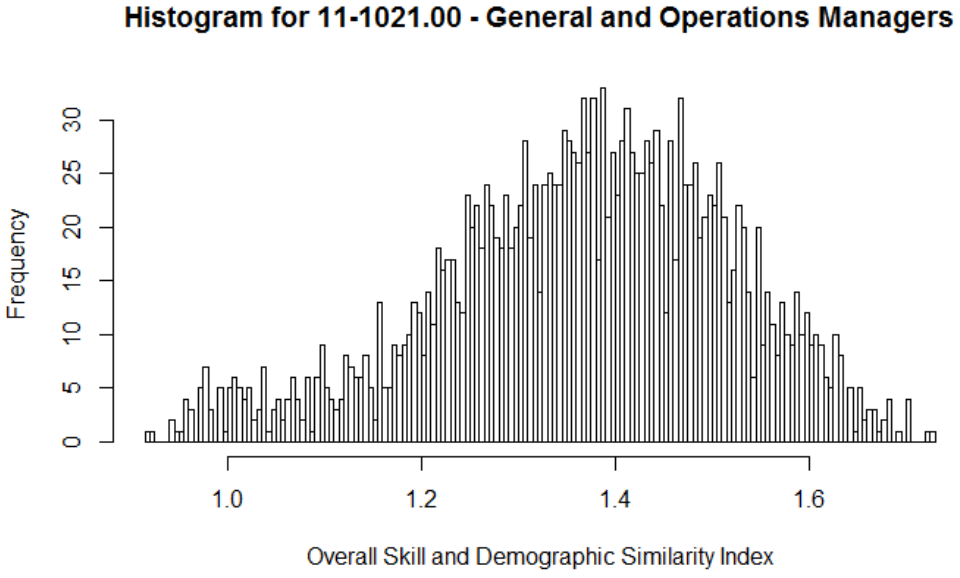


Figure B.4: **Example 2b - Distribution of Overall Skill and Demographic Similarity Index of active population for General and operations managers**



demographic information of the individual. Nevertheless considering only the index would be an error since the worker valuates other information of the market to make the decision; the wage level, the number of available vacancies, and it's skill level compared to the overall skill level play an important role in the occupational choice.

In order to understand the process I divide the algorithm in two parts: optimal occupational choice, the process for which a worker maximize its utility choosing an occupation, and hiring decision, the process for which firms (in this case the occupation) fill the vacancies maximizing it's utility with a given candidate pool.[4]

Optimal occupational choice

The way I proceed to describe the optimal occupational choice is based on the ideas presented by Miller [1984], for which the optimal decision rule for the worker its associated with the beliefs about its future returns. The optimal decision rule is characterize by the maximization of the index. A dynamic allocation index is computed for each iteration. We take this idea to construct an index that represent the behavior of the workers of the economy.

Consider $Z = \{k \in \mathbb{Z} \mid 0 < z < \bar{z}\}$ as the number of iterations required to achieve the equilibrium of the economy, in which the number of vacancies equals 0. The number of iterations are the number of negotiations that occur in the economy for 0 length unit of time.

In order to give a structure to the method presented in this paper, define the iteration invariant matrix $\mathfrak{V}_{(J,I+1)} = \left[\mathfrak{O}_{(J,I)} \mid \mathbf{W}\right]$, where $\mathfrak{O}_{(J,I)}$ is the matrix containing all the Overall Skills and Demographics Similarity Indexes and $W_{(J,1)}$ is the wage vector that contains the observed mean wage for each occupation in the vacancy data. Note that the matrix $\mathfrak{V}_{(J,I+1)}$ is defined as iteration invariant since the wages and skills and demographic are constant during the whole simulation, since it occurs in a 0 length unit of time, there is no place for learning, skill actualization or response in demand or supply of skills.

Define as $\mathfrak{J}_{(J,z)}$ as the iteration variant vector of available vacancies in the

[4]The present chapter and algorithm is proposed by my supervisor professor D. Margolis and I

economy for each iteration of the simulation. The initial iteration $\mathfrak{J}_{(J,0)}$ correspond to the total number of vacancies in the economy by occupation (Vacancy data). This vector is iteration variant since when there is a match, the number of vacancies matched in each occupation in iteration $z+1$ will be subtracted to the number of vacancies in z.

$$\mathfrak{J}_{(j,z+1)} = \mathfrak{J}_{(j,z)} - match_{(j,z)}$$

Define the matrix $\mathfrak{S}_{(I,z)}$, the variant vector of job seekers for each occupation in the economy for each iteration.The initial iteration $\mathfrak{S}_{(I,0)}$ correspond to the number of active agents by type in the economy. This information comes from the rounded weights of the STEP survey.

Define $D_{(i,j,z)}$ the dynamic allocation index that describes the behavior of the active population in the economy. It is defined by:

$$D_{(i,j,z)} = W_{(j,1)} \times \frac{O_{(j,i)} \times \mathfrak{S}_{(i,z)}}{\sum_i^I O_{(j,i)} \times \mathfrak{S}_{(i,z)}} \times \min\left\{ \frac{\mathfrak{S}_{(i,z)}}{\sum_i^I \mathfrak{S}_{(i,z)}}, 1 \right\}$$

$$D_{(i,j,z)} = W_{(j,1)} \times A_{(i,j,z)} \times B_{(i,z)}$$

$D_{(i,j,z)}$ represent the weighted personal probability of each type of worker i in the economy for each occupation j in the iteration z. The personal probability is a valuation function of the chances of getting a job in round z given the number of human capital units that the agent type can contribute relative to the total provision for the economy in time z, times the wage for the occupation j. The personal probability is composed by three parts:

$W_{(j,1)}$ is the wage expected to receive by working in the occupation j. Note that a valuable extension for this model is to allow for wage adjustment based in parameters. For doing so we would only replace the wage by the expected wage from z to $\bar{z}$ and make it change the weighted personal probability function based on this beliefs.

$$A_{(i,j,z)} = \frac{O_{(j,i)} \times \mathfrak{S}_{(i,z)}}{\sum_i^I O_{(j,i)} \times \mathfrak{S}_{(i,z)}}$$

The second part is the number of human capital units job seeker type i could potentially contributes to the occupation, based on the belief that all job seekers might apply for the same vacancy.

$$B_{(i,z)} = \min \left\{ \frac{\mathfrak{S}_{(i,z)}}{\sum_i^I \mathfrak{S}_{(i,z)}}, 1 \right\}$$

The third part is the subjective probability of getting the job without considering the endowment of skill, but only the potential number of vacancies in occupation j over the overall vacancies in the economy in iteration z.

Optimal occupation choice is defined as the maximization problem that faces the job seeker i, in which he maximizes in period z the value of the the weighted personal probability. The Optimal occupation $\tilde{j}$ is the solution to the problem:

$$\sup_j \left\{ D_{(i,j,z)} \right\}$$

,

For worker type i in iteration z. Note that if there are no available workers in type i (all type i job seekers are hired) the value of the index is 0, showing that the occupational choice problem after being matched does not exist anymore. This assumption is also a simplification, since the labor economic literature, explain many frictions by considering on the job search. The decision for not including on the job search is that the time frame of this simulation is 0, and there is no place for other reallocation.

Hiring decision - Optimal hiring

The hiring decision is built under two concepts: urgency, efficiency. By urgency I intend the relative importance that a firm that has to fill their vacancies in confront to the whole market. This might be seen as the ability of a firm to

evidence their vacancies among their competitors, and the intuition behind this is that the a firm with more vacancies will be more attractive so will choose first. To do so the order in which the firms are able to hire depend of the number of vacancies. To do so I create a ranking base on the share of vacancies by occupation and make the hiring decision in that specific order. The second concept is productivity, and is intended as the willing of the firm to choose the best candidate among the available (so in case there are more than one candidate it will choose continuously and will stop when the number of vacancies is 0.

In order to model the hiring decision consider the function $f(x) : \mathbb{R}_0^+ \to \mathbb{R}_0^+$:

$$f(D_{(i,j,z)}) = \begin{cases} O_{(j,i)} & \text{if } D_{(i,j,z)} = D_{(i,\tilde{j},z)} \\ 0 & \text{if } D_{(i,j,z)} \neq D_{(i,\tilde{j},z)} \\ 0 & \text{if } D_{(i,j,z)} = 0 \end{cases}$$

The optimal hiring for a firm is defined as$\tilde{i}$, being it the solution to the following maximization problem:

$$\sup_i \left\{ f(D_{(i,j,z)}) \right\}$$

.

We define a match the stable coalition for which job seekers $\tilde{i}$ is matched with occupation $\tilde{j}$ in iteration z. The number of matched workers of type $\tilde{i}$ will depend in the number of vacancies available in iteration z. Define the function $m(\gamma_{(\tilde{i},\tilde{j},z)}, \mathfrak{J}_{(\tilde{j},z)})$, as the function that assign the number of new employed workers in the occupation $\tilde{j}$ in based of the number of vacancies and the number of matched individuals of type $\tilde{i}$. Define $m(\gamma_{(\tilde{i},\tilde{j},z)}, \mathfrak{J}_{(\tilde{j},z)}) : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \to \mathbb{R}_0^+$ as:

$$m(\mathfrak{S}_{(\tilde{i},z)}, \mathfrak{J}_{(\tilde{j},z)}) = \begin{cases} \mathfrak{S}_{(\tilde{i},z)} & \text{if } \mathfrak{S}_{(\tilde{i},z)} \leq \mathfrak{J}_{(\tilde{j},z)} \\ \mathfrak{J}_{(\tilde{j},z)} & \text{if } \mathfrak{J}_{(\tilde{j},z)} \leq \mathfrak{S}_{(\tilde{i},z)} \\ 0 & \text{if } \mathfrak{J}_{(\tilde{j},z)} = 0 \end{cases}$$

So the number of filled positions will be equal to the number of vacancies in period z if the number of job seekers of type $\tilde{i}$ is less or equal than the number of vacancies, or the number of filled positions will be equal to the vacancies if the

job seekers exceed the number of vacancies.

B.0.3 Technological change

Redefine the function $s_{(i,j,k)}$ as:

$$s^1{}_{(i,j,k)} = d(score_i, level_j, importance_j) =$$

$$= \begin{cases} \left(\frac{score_i}{level_j}\right) * importance_j & \text{if } score_i \leq level_j \& level_j \leq 10 \\ \left(\frac{score_i}{level_j - 10}\right) * importance_j & \text{if } score_i \leq level_j \& level_j > 10 \\ importance_j & \text{if } score_i > (level_j - 10) \\ 0 & \text{if } level_j \leq 0 \end{cases}$$

A technological change is considered as the lowering in the requirement levels, so the similarity measure increases given that is a decreasing function in the $level_j$. The insight for this shock is that the occupation requirement lowers in order to make more people able to apply, and trough training in the firm the agent increases his skill endowment (there would be necessary in t+1 a wage adjustment since the offered wage depends on requirements too, this case is not considered here because here the time lapse is equal to 0).

After applying the new similarity function we define the new aggregate skill similarity index as:

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

,

and the new similarity matrix $\mathfrak{M^1}_{(J,I)}$

$$\begin{bmatrix} \tilde{S^1}_{(1,1)} & \dots & \tilde{S^1}_{(1,1987)} \\ \dots & \dots & \dots \\ \tilde{S^1}_{(770,1)} & & \tilde{S^1}_{(770,1987)} \end{bmatrix}$$

With those values we apply the weights to for the Overall Skills and Demographics Similarity Indexes post treatment for all the worker types and all the occupations in the economy.

$$\mathfrak{O}^1{}_{(J,I)} = \begin{bmatrix} O^1{}_{(1,1)} & ... & O^1{}_{(1,1987)} \\ ... & ... & ... \\ O^1{}_{(770,1)} & & O^1{}_{(770,1987)} \end{bmatrix}$$

We simulate taking in consideration this values and see how the matching results change.

B.0.4 Training for unemployed

One of the basis for active labor policies is to make sure that the unemployed keep the job search while being in a learning program. The idea of this is to increase the skill level to augment its employability. In order to construct a counterfactual we define a new similarity skill function that is conditional on being unemployed or not. If the person is employed the function is the same as in the method, if the person is unemployed the function is defined as:

$$s^2{}_{(i,j,k)} = d(score_i, level_j, importance_j) =$$
$$= \begin{cases} \left(\frac{score_i+10}{level_j}\right) * importance_j & \text{if } score_i \leq level_j \\ importance_j & \text{if } score_i + 10 > level_j \\ 0 & \text{if } level_j = 0 \end{cases}$$

As in the previous case we define the new aggregate similarity function $\tilde{S^1}_{(i,j)}$, and the new similarity matrix $\mathfrak{O}^2{}_{(J,I)}$:

$$\tilde{S^2}_{(i,j)} = \sum_{k=1}^{29} s^2{}_{(i,j,k)}$$

,

$$\begin{bmatrix} \tilde{S^2}_{(1,1)} & ... & \tilde{S^2}_{(1,1987)} \\ ... & ... & ... \\ \tilde{S^2}_{(770,1)} & & \tilde{S^2}_{(770,1987)} \end{bmatrix}$$

As in the previous case we apply the weights to for the Overall Skills and Demographics Similarity Indexes post treatment for all the worker types and all

the occupations in the economy.

$$\mathfrak{O}^{2}{}_{(J,I)} = \begin{bmatrix} O^{2}{}_{(1,1)} & \dots & O^{2}{}_{(1,1987)} \\ \dots & \dots & \dots \\ O^{2}{}_{(770,1)} & & O^{2}{}_{(770,1987)} \end{bmatrix}$$

Appendix C

Considerations on the data used on the simulation

There are two assumptions that are important to discuss, in order to justify the use of the databases used in the empirical part of the document. The first assumption validates the use of information from two different time periods in the same exercise. The second assumption regards the use of O*NET as source of information for the required skills of the occupations, and it's validity of the results for the Colombian labor market.

The first assumption that enables the use of the three data sources is that the information collected in the vacancy database, collected in 2014, can be used in for the time frame of 2012, year where the STEP survey collected the results. This assumption is justified because the Colombian economy didn't suffer any shocks that make differ the economy from 2012 to 2014, in which occupational structure and job market refers. Even if the level of unemployment felt continuously, the employment level remains constant. The main reason for this is that inactivity rate grew upon the period because of large government expenses and the introduction of conditional transfers programs focused on young population. One example of such programs is *Jóvenes en Acción*, which provided vocational education training to around 100.000 young people of poor and vulnerable families. The increasing in inactivity doesn't affect the presented model since the data used only considers the active population (this choice is discussed when the STEP database is presented).

The other reason that justifies the first assumption is the structural composition of the economy has been held relatively stable, so there are not major shift

Figure C.1: **Evolution of the Employment Rate**

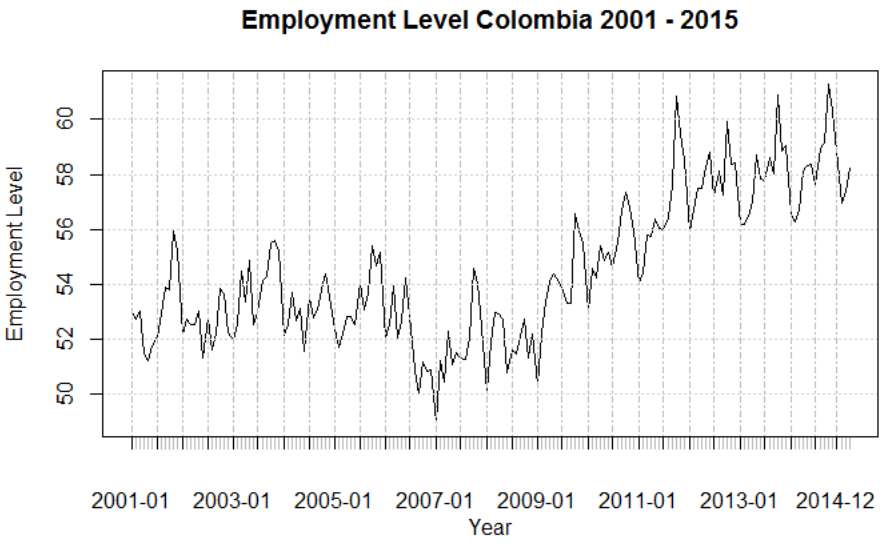
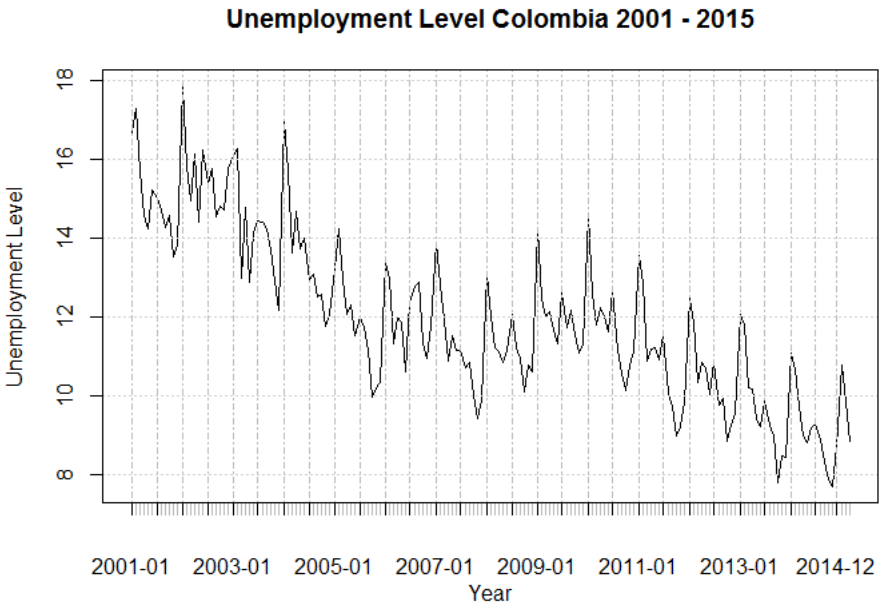


Figure C.2: **Evolution of the unemployment Rate**



in technology or production that could suggest an occupational shift during this years. The sectoral composition doesn't have a major shift. The overall growth of the Colombian economy for 2014 was of 4,3% at constant value prices, being the construction, commerce and transport sectors the ones that contribute more to growth. Anyway even if the the sectoral growth of these sector, the sectoral shares don't change too much between 2012 and 2014, and the jobs and firms characteristics are comparable.

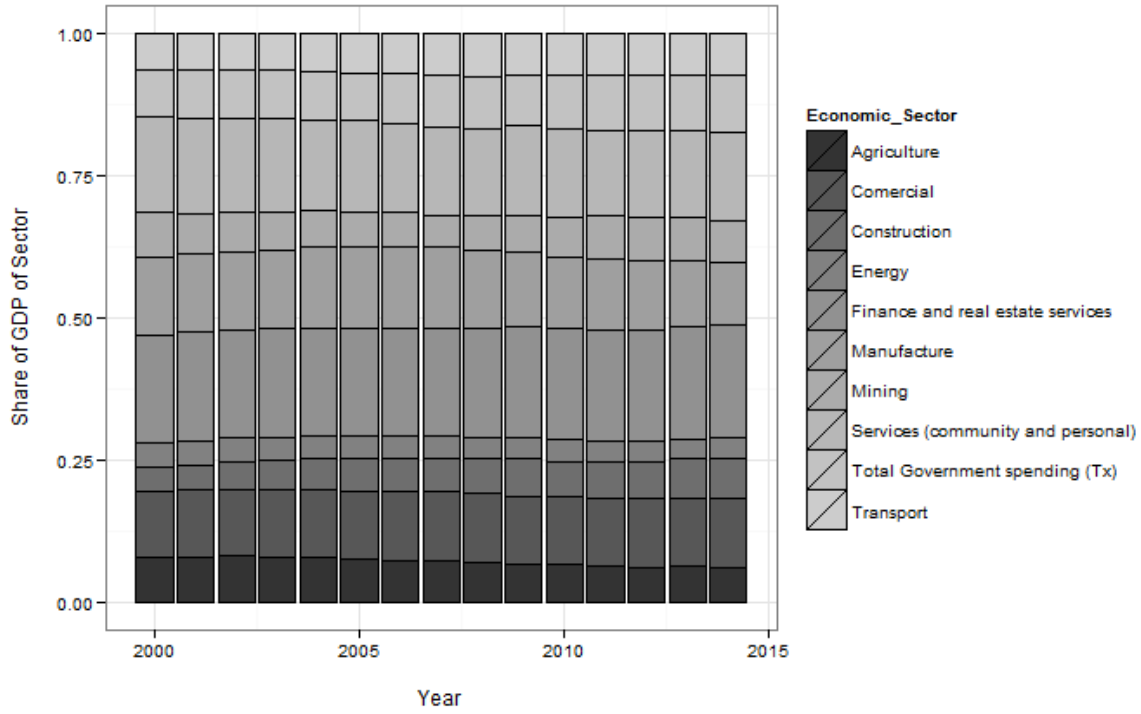
Table C.1: Share of working population by activity

	2012	**2013**	**2014**
Agriculture	0.92%	0.88%	0.80%
Mining	0.32%	0.32%	0.33%
Manufacturing	16.98%	16.28%	16.10%
Energy	0.51%	0.53%	0.55%
Construction	6.59%	6.19%	6.29%
Comerce	30.55%	30.69%	30.58%
Transport	9.69%	9.48%	9.29%
Finance Intermediation	2.02%	2.21%	2.18%
House rent and other activities	10.29%	10.79%	11.17%
Services (personal and social)	22.15%	22.61%	22.71%

Souurce: DANE household Income Survey (GEIH)

The main shock in the Colombian economy during this period was the exchange rate devaluation against the dollar and the fall in process of oil that affected the growth of the oil and mining sector. Never less this phenomenon occurred only in the last two months of 2014, so the structure didn't change trough out the year.

The last two facts lead us to two conclusions: given that the structure of the economy didn't suffer major changes and the employment levels are similar for the two periods, we should not expect any change in occupational structure. This allows us to think that the same vacancies are created for the same kind of occupation during both periods. Second, the sectoral growth of the economy is similar in both years suggesting, if jobs were created during both periods, the jobs created were similar and followed the same structure.



The only concern for the use of the two periods, is the effect that the decrease of activity of the mining and oil and gas sectors had in job creation for those sectors. Even if the decrease in price of oil and the Colombian peso devaluation at the end of 2014 was very accentuated, one may think that the effect on the industry and jobs was not instantaneous. The first reason to support this is that being oil and gas firms large companies, the operations are risk covered by futures in most of the cases, so even if the oil price drop occurred in the last quarter of 2014 and the peso depreciated in the last months of the year, one might think that the effect is felt at the beginning of 2015. If the last statement is true, one might consider no effect during 2014, being able to compare the STEP database and the vacancy data.

The second assumption, for which we analyze the occupations of the economy using the O*NET database descriptors for the Colombian economy, is justified by two practical reasons. The O*net content model is a detailed taxonomy of the different dimensions that an occupation has: abilities, skills, occupational requirements, experience requirements, task, tools and technology, between others. This kind of information is not available for the Colombian labor market. The development of this kind of information is very expensive and takes time, reason to think that it will not be available neither in the short run, because being Colombia a

developing country, it has other immediate challenges to cover.

The second argument in favor of the use of O*NET is that its use serve as a prospective tool to adequate the occupational skills, abilities and knowledge to the future requirements of the working force. It is very difficult to determine if the reality of the occupations differ so much or not between the U.S. and Colombia in terms of technology, skills, task and job values. By one side one may say that the institutions, the different environment, the values and work practices make that the occupations differ in their practice in both countries. By the other side the task, skills and abilities to make bread or to perform economic or legal analysis might be the same in the two countries. This discussion is not covered in this document but instead, O*NET is considered as a practical, and only available tool, to quantify the skills and abilities requirements in the labor market in reference to the technological leader. This view makes that the analysis and result derived from this document are considered as the medium and long term occupational requirements for the Colombian job market, giving as main characteristic that the model presented could be considered as prospective. What we can certainly say is that the use of the taxonomy, and the measures taken with the use of that tool can prepare the workforce to be as competitive as the main reference economy, being that the U.S. economy.

The following sections provide a detailed description of the databases used, providing detailed information on how the data was obtained and is used in the present work.

C.0.1 Colombian Vacancy Data

the dataset contains monthly registries of vacancy job advertisements from the two most important private job boards, the monthly registries of the national public employment offices and its job board (UASPE - Unidad Administrativa de Servicio Público de Empleo), the monthly vacancy registries of the job board national vocational education and training (SENA – Servicio Nacional de Aprendizaje), the job board of universities and the headhunter information. Since its creation in the last quarter of 2013, the database has been updated monthly with the information from the mentioned sources.

The collection method differs from each source, and for the sake of the va-

lidity of it use, and the robustness of the analysis I will explain briefly the history, methodology of collection, homologation of the information in one repository and data structure, cleaning and categorization, and analysis.

A database for job monitoring

Since the last decade there has been in Colombia an increasing concern of the competiveness of the country against its neighbors and how to include the country into the global market. By 2010 there were official policy recommendations that were consolidated in the a Colombian policy document, the CONPES 3674 of 2010, in which the country need to align the human resource strategy to the productive sector requirements and the competiveness policies implemented from the government.

Regarding the productive sector requirements is important to remark that Colombia does not have any source of information that coincides with this requirement. Colombia doesn't have a demand survey, or any source of information about the activities of the firms in Colombia, being this a problem for a correct policy implantation. The only information available of this kind is the annual manufacture survey (EAM – Encuesta Anual Manufacturera), but is representativeness is not wide, because only covers a specific sector, and does not provide any information on vacancies, hiring and skill requirements. Other source of information is the documents of the sectoral councils, activity that has been in charge of the national vocational education and training institution (SENA); even if the information is available and is extensive, the methodology is a qualitative methodology that cannot be integrated in a model, and lacks of statistical representativeness, and given the implemented methodology the information collected is not a true sample of the skills requirements of the whole industry.

In the other hand the competiveness policies are very general and don't specify the skills needed to perform a job, and this type of quantitative information is not available neither in Colombia. Even if the capital investments are made by the productive sector, the lack of skilled human resources put in jeopardy the implementation of the policies. There was needed a quantitative tool that can trace the skill requirements, and track the changes in the productive structure in order to adjust the vocational education and training offer, in order to fulfill the productive sector requirements.

As response as the lack of data there was proposed, design and implemented the project of job monitoring, which based on the collection of vacancy information and occupational monitoring aimed to identify the occupational structure and quantify the skills and ability requirements of the Colombian job market[1]. The main aim of the collection of this information was to create a methodology to assess the gap of skills and qualifications in the Colombian labor market, for which this document represent an effort to achieve this goal.

From the academic point of view this information is also relevant. The labor economics academic debate has been absent in Colombia the past decades because this lack of information, and labor demand studies are scarce (Vivas et al. [1998]). The demand labor information has been evaluated from three sources: The effective demand that does not account of the demand sources, from the household surveys (GEIH). The general equilibrium models which are based on the macro level aggregates and don't provide enough information to have account the gap of skills of the productive sector and the occupational profiles studies, that are not representative to the national level. As an alternative the vacancy information provides trough occupational analysis new data sources that could be structure in a quantitative way to detail the skill, abilities and knowledge required by employers and provide the occupational structure of the sectors.

Data collection and merging

The data sources of the database were collected through two channels. For the information that was hosted in government institutions as the SENA, universities and UASPE, there were signed agreements in order to count on the information in a monthly basis. They provided access to their information systems and the request were design in order to obtain the most quantity of information by registry. In the same way there was signed an agreement with the job hunters and human resource national association (ACRIP) to access their information in a monthly basis.

For the private job boards the methodology consisted in web scrapping, so there were built computer algorithms to download the internet published vacancies

[1]A detailed analysis of this provided in the technical documents made for the consultancy I developed in the Colombian Ministry of Labor in that moment.

and give structure to them. Working with internet published information brings new challenges, since is an atypical source of information. Never less the storage of all internet transaction data open new opportunities to research in different disciplines (Li and Cheng [2012]). Every day is more common to find new studies that used internet data as basis of its research. Edelman collect the research initiatives in economics showing that the range of impact of internet information in economic research has been applied in microeconomics, political economy and health economics Edelman [2012]. Another famous publication using this kind of data in microeconomics is the work of Levin, for which internet based purchases and consumer behavior is studied deeply, bringing new insides on preferences and taxation ,'citeeinav2014economics.

The information scrapped, is the information that is provided to the public trough the web browser. It coincides with what a job searcher reads in a web job board, so the information that was introduced by the firms that are willing to hire. In that sense the source of information is the labor demand *per se*. The procedure to obtain the information is based on the interpretation, analysis and pattern recognition of the HTML code on the web. For people with non-technical knowledge on how a web browser works, a web browser is just a visual representation of code that link images and text in a structured and defined form (tables, fields, text, images). The idea behind the web scrapping is to profit from the structured form of that language, and by this I mean recognize the pattern of the names of elements in the tagged language (HTML), to download the information. The algorithm makes use of a functionality of computer language called REGEX, that allows to identify chains of text and by similarity matching them. Doing so, identifies for each page the name and value content, and later extracting the information into a database. Since each job vacancy in each system has its own URL address, the algorithm is able to keep record of that and download only once the information. Additional filters are made after to ensure that the registry is not duplicated (there are not two identical job offers in the data), using the SOUNDEX comparison algorithm.

When all the data form different sources is obtained, the data is merged in a common data warehouse. Since every source has different fields, because every system is constructed in different ways, the data warehouse attempts to homogenize and make comparable the different sources. The structure contains information of the ID, the title of the job offer, the name of the company, the sec-

tor, the experience, the location of the job offer, the offered wage, the description of the vacancy, the number of vacancies and all the relevant information that the employer wants to make public in order to fill the position.

One important remark is that the job boards, as they don't intent to perform statistical analysis on occupations, they don't provide any occupational classifiers for the vacancies. Since the amount of data is so large there was not the possibility to do a manual classification as the one that is usually made in the statistical institutions (household survey). Instead I developed an algorithm composed by two task: using pattern string recognition classify the most common titles. This common title list was constructed in base of the data collected and assign a classifier for O*NET classification. The second step is an algorithm of automatic classification based on the Spanish version of the API of O*NET my next step. This API finds by proximity the 10 most similar occupation titles in based of an input information. Having the occupational classification for O*NET there we used the pathway for the correspondent ISCO 08 classifier. In that way is possible characterize the occupational structure of the economy, which is the main input of the simulation model developed in this document.

One of the main critics to this information is the statistical representativeness of the data. In order to demonstrate the statistical representativeness of the data I will base in two arguments: geographical coverage and means of employment search in Colombia. The first argument has to do with the geographical coverage of the collected data. The data covers 349 villages and cities. The largest survey implemented in Colombia, the household survey (GEIH) covers only the 13 biggest cities in Colombia. The amplitude of coverage of the data brings more representative of the productive sector across the country and from the human resource requirements of the productive sector. Neverless the data shows that the concentration of job offers still remains in the larger cities: Bogotá (56.4%), Medellín (6.2%), Cali (5%) and Barranquilla (2.8%). The second argument has to do with the internet sources and how the information is not representative, since Colombia is a developing country that has a huge gap in internet use, and moreover for employment search. However the data taken from the national household survey (GEIH) regarding the channels used by active labor workers to search for employment is potentially captured by the collection method and sources of the vacancy database (see table).

Table C.2: **Channels for job searching**

Means for searching personal	**Industry**	**Trade**	**Services**
Informal networks	23.80%	26.90%	18.00%
Databases / own records	17.40%	18.30%	18.70%
Web job boards	*16.70%*	*13.70%*	*20.20%*
National Apprenticeship Service (SENA) - Public Employment Service	*12.30%*	*13.70%*	*10.40%*
Advertising on media	*12.20%*	*10.80%*	*10.40%*
Job Boards of Universities and other organizations	*8.40%*	*6.90%*	*10.80%*
Headhunters / Job boards	*6.70%*	*6.50%*	*6.70%*
Contact with other educational institutions	2.10%	2.70%	4.00%
Job Fairs	0.50%	0.50%	0.80%

Source: Households Income Survey. DANE (2014)