The estimation of urban premium wage using propensity score analysis: some considerations from the spatial perspective

Dusan Paredes

Marcelo Lufin

Patricio Aroca

Serie de Documentos de Trabajo en Economía - UCN

<https://sites.google.com/a/ucn.cl/wpeconomia>

Identificador: WP2012-07

Antofagasta – Abril 2012

www.ucn.cl

Avda. Angamos 0610, Pab. X10, Antofagasta, Chile

Tel: + 56 55 355746 - Fax: + 56 55 355883

MAGISTER EN
CIENCIA REGIONAL

Ciencia Regional y Políticas Públicas
UN NÚCLEO DE iniciativocientificamilenio

Abstract

This paper estimates the Urban Wage Premium for the Chilean case, but proposing three contributions to the literature. First, the geographical space is reconfigured using functional regions instead administrative regions. This process is carried out using techniques from the spatial econometric literature. Second, we exploit the use of micro data. We estimate the wage equation using a survey at individual level that avoids problems associated with regional aggregated specifications. Finally, we set the comparability among observations using matching comparison. Thus, we isolate the urban effect from other alternative sources. Our results suggest a Urban Wage Premium between 4 and 21%.

Este artículo estima el Premio Urbano en Salarios para el caso de Chile, pero proponiendo tres contribuciones a la literatura. Primero, el espacio geográfico es reconfigurado usando regiones funcionales en vez de regiones administrativas. Este proceso es llevado a cabo usando técnicas de econometría espacial. Segundo, nosotros explotamos la riqueza del micro dato. La especificación de la ecuación de salario es a nivel individual para eliminar problemas de agregación regional. Finalmente, la comparabilidad entre individuos es garantizada usando comparaciones entre contra factuales lo que permite aislar el efecto de la prima sobre otras fuentes. Los resultados sugieren un Premio Urbano de Salarios entre 4 y 21%.

Palabras Claves: Size and Spatial Distributions of Regional Economic Activity

Codigo JEL: R12

1. Introduction

The urban economics literature supports that thick labor markets pay higher wage levels than thin labor markets. Glaeser and Mare (2001) estimate the elasticity wage-city size larger than 1 million inhabitants around of 36 per cent higher than smaller areas, while Glaeser and Resseger (2009) identify a elasticity of 45 percent for the case of skilled workers. This positive relation also exists within industries, but with an uneven impact (Elvery, 2010). In spite of the extensive empirical evidence, the most of the applications have been focused on North American, European and Asian contexts. In this paper we extend the analysis toward the Latin American case, where the ONU-Wider has strongly recommended focusing on “*increasing inequalities partly as a consequence of the uneven impact of trade openness and globalization*” (Kanbur et al. 2005). We use the Chilean case and provide a first estimation of wage differentials between thick and thin labor markets. Although the extension toward new contexts could be considered a contribution as itself, the particular scenario of Latin American realities must be discussed.

Puga (1998a), Rosenthal and Strange (2004) and Venables (2005) among other authors suggest that developing countries, such as the Chilean case, are characterized by high primacy index. This fact implies that a unique region, in the most of the cases the capital of the country, concentrates a large portion of the total economic activity and population. For example, the Metropolitan Region of Santiago concentrates around the 40% of the total population and the 50% of production in the 2% of the total territory¹. That indicates than developing countries present less thick labor markets than developed countries and even, some countries could present only one dense labor market. In the Chilean context, the MR rises as a natural benchmark. Given that MR is a thick labor market, then we expect that this region pay a wage premium in

¹ Annual National Manufacturing Survey, Chile, 1997

comparison with any other Chilean region². In spite of the simplicity of this statement, some additional discussion about the labor markets is needed.

We expect to contribute with three empirical discussions for the estimation of spatial wage differentials in the Chilean case. The first is the issue raised when the aggregated wage at regional level is used as a dependent variable. Glaeser and Mare (2001), Berry and Glaser (2005), Glaeser and Resseger (2009) and Elvery (2010) are examples of spatial wage differentials using averages wages as dependent variable. The authors add an independent variable indicating the size of the region and its coefficient identifies the wage premium. This measure can be accepted when there are not data for capturing the heterogeneity in human capital, but its use presents several problems. Firstly, the theory of spatial sorting suggests that high skills workers are concentrated in large cities (Combes et al. 2008 and Mion and Naticchioni (2009)). This theory rejects the direct comparison of wages at aggregated level due to the worker heterogeneity could also explain the wage differentials. Additionally, the skills concentrations vary across industries and economic sectors, implying a wage variation at intra-city level. Even controlling by worker characteristics, the industrial mix affects the estimation of wage differentials. Both problems put in evidence the advantages of considering the wages at micro data instead of any aggregated level. We partially overcome this problem using a rich database for the Chilean case which is disaggregated at individual level. Our survey allows identifying characteristics of the worker such as age, education (years), educational level, economic sector, occupation, type of contract, and marital status among other personal characteristics.

² The Chilean administrative division presents fifteen regions including a Metropolitan Region (MR).

A second dimension is how the literature has defined the geographical scale of a thick labor market. For the Chilean case, and the most of the empirical literature, the researcher would use the administrative division of the geographical space and the MR, again, appears a natural candidate for benchmark area. However, this administrative approach is problematic. A thick labor market emerges due to the existence of economics of agglomerations that work to maintain a spatial concentration of demand and labor supply. The continuous interaction in the labor market generates increasing returns to scale and peer effect networks that increase the worker productivity and, consequently, generates higher wages. In this framework, the economics of agglomeration works in a functional space, which does not necessarily fit with the administrative division. In other words, the thick labor market in the MR could be larger than the administrative division because of the interaction MR with other contiguous regions. We cover this gap proposing a stepwise tool for spatial delineation of thick labor markets following a combination of methods. We use commuting flows among counties for building the geographical scope of labor markets. The proposed method is rooted on the spatial analysis literature (Anselin, 1995; Getis & Griffith, 2002) and also social networks analysis (Bonacich, 1972; Green, 2007). We generate a new division of the geographical space where we confirm the existence of only one thick labor market, namely the MR, but we show a larger geographical scope where some communes of other regions are absorbed by this thick labor market.

The third dimension emerges when the wage differential is considered as a causal effect of thick labor market instead of other control variables such as human capital. Two workers could present different wages under different labor markets, let's say thick and thin labor markets, but we cannot identify this causality if the workers also have different human capital. We face this problem bringing the counterfactual framework inside the regional science literature. The

counterfactual framework makes a comparison just among comparable workers. For example, a thick labor market could group very high skilled workers and its comparison with any other thin labor market would be biased. In order to improve the previous literature, we use a Coarsened Exact Matching (CEM) to find identical workers between a thick and thin labor markets as a previous step to estimate the wage differential. We use CEM instead other matching approach such as Propensity Score Analysis (PSA) due to the exact matching find the similar worker, but the PSA could match two workers with similar propensity score and still with different characteristics.

According to the three empirical contributions previously discussed, we set the structure of our article. Our hypothesis is that thicker labor markets generate a wage premium in comparison with thin labor markets. To carry out our testing strategy, we require a redefinition of the geographical space to identify functional thick labor markets. We used spatial and social network analysis to identify the functional area. With this result, we are able to compare two geographical divisions: an administrative space (fourteen regions and the Metropolitan Region of Santiago labeled as thick labor market) and a functional division (fourteen regions and a new thick labor market which captures the old MR plus a portion of other two regions). Next, we estimate the spatial wage differentials using both divisions and we test if the functional division increases the wage premium in comparison with the previous spatial structure.

Our results support the established hypothesis. First, the functional approach shows a new geographical space. The administrative perspective implies fifteen different markets, but our approach suggests a different configuration of the space. The Metropolitan Region of Santiago is a huge labor market, which also captures a portion of other regions. Moreover, a portion of the Metropolitan Region, particularly the rural areas, does not belong to the new thick labor

market. These results suggest evidence how the geographical extent of labor markets is not defined by administrative reasons, but rather by its functionality. Using the administrative and functional geographical division, the wage differential was estimated for Chilean regions. The results show how, after a matching procedure, a similar set of workers earn higher wages in the thick labor markets comparison for almost any region. This premium is bounded between 4 and 21%. Alternatively, a set of three regions show a higher wage than thick market, but its difference is reduced with our functional approach. These shown results suggest the existence of spatial wage differentials in the Chilean case. This output opens new discussion for policy makers, especially for concerns about how the density of labor market affects the welfare measurements such as wages.

2. Methodology

Our empirical strategy implies two steps: a functional definition of thick labor markets and the estimation of the wage premium through matching estimator at micro data level. The administrative division of Chile presents three administrative divisions: 346 counties, 53 provinces and 15 regions. The densest region is the Metropolitan Region of Santiago (MR) which comprises 6 provinces and 53 counties. The MR contains the 40 percent of the total population with a density of 390 habitants per square kilometers when the national average density is 19.9. These characteristics posit the MR as the largest region in Chile and a perfect spatial unit to use as a thick labor market. However, the choice of the MR as a dense urban area could be arbitrary. According to Marshall (1920) the agglomeration economics, which are the sources to argue higher wages in dense areas, operate in a geographical space defined by the extent of the labor market. Thus, several critiques appear by using the MR as a benchmark area.

First, the MR could represent more than one labor market and an aggregated perspective could bias the estimation. Second, other large geographical labor markets could exist in others places of the country. Third, the labor market extent must be defined with homogeneous labor characteristics more than political-administrative areas.

This paper contributes with an alternative methodology for building a new “space” dimension to measure the scope of economics of agglomeration. Firstly, we identify a thick labor market using functional regions based on geographic contiguity and people flow interaction among the 346 counties. This methodology keeps a geographic dimension, but it is built according to some relation-measure, leaving out the a priori spatial administrative division. Here, as a complement to flows between areas, the presence of a significant level of spatial local autocorrelation (Getis & Griffith, 2002) is proposed to identify thick labor markets in order to maximize areas that behave similar in terms of their importance in the networks of workers flows between locations.

To identify the relative importance of each location in the system of flows, we propose a functional design based on the social networks literature, there the Eigen-centrality has been used to identify relevant actors in a relational structure (Bonacich, 1972 and Borgatti, 2005), it is defined as the principal eigenvector of the matrix defining the network. Using that index, each node (location) is ranked base on its eigenvector score, in a manner in which those more centrals are adjacent to nodes that are themselves more centrals. This idea has been also explored and used in geography literature (Boots & Kanaroglou, 1988; Boots & Tiefelsdorf, 2000). In that fashion, locations that are similar in terms of their centrality in the system of workers’ flows, but also their constitute a spatial cluster of contiguous units associated to a hotspot in terms of commuting rates, belong together to in a new functional geographical area defined as a thick labor market.

a. Eigen-Centrality Index and functional region Identification

This centrality measure was proposed by Bonacich (1972) as a modification of the degree centrality perspective, traditionally the centrality approach establishes that nodes (counties in our case) who have more connections are more important in the relational system, because they can affect more other actors. Bonacich argued that each node's centrality is a function of the extent of its connections (degree), but also it is depending of the number of connections that its neighbor locations hold. This new centrality perspective rest in the idea that being connected to well connected others makes an actor more central.

The eigenvector centrality was developed by Bonacich (1972) to measure of how well connected is an actor in the whole system. The method works as the factor analysis (FA), assuming that each municipality in the commuting flows matrix is a variable, FA identifies "dimensions" of the distances, based on commuting flows, among them. The location of each municipality with respect to each dimension is called an "eigenvalue," and the collection of such values is called the "eigenvector." Usually, the first dimension captures the "global" aspects of distances among actors; second and further dimensions capture more specific and local sub-structures. Higher scores indicate that actors are "more central" to the main pattern of social network distances among all of the actors, lower values indicate that actors are more peripheral.

Later, those Eigen-Centrality indexes by municipality, jointly with the proportion of total population that commutes and population density were mapped in their geographical

representation, to identify thick labor markets areas a spatial cluster analysis were performed using Lisa Indicators (Anselin, 1995), to test those areas three adjacency matrices based on contiguity type queen order 1, shortest distance between polygons, and six-nearest neighbors were used, with 9,999 permutations. Those municipalities systematically associated with the same hotspot were included in the same functional thick labor region.

b. Counterfactual framework and matching comparison

Assume that the worker can be assigned to two exposures: $T = 1$ if he is at a thick labor market and $T = 0$ if he is not. Assume that the wages can be observed under $T = 1$ and $T = 0$, with W_1 and W_0 , respectively. Using the Neyman-Rubin counterfactual framework, the potential wage for worker i is:

$$W_i = T_i W_{1i} + (1 - T_i) W_{0i} \quad (1)$$

Note that the researcher cannot observe W_i under $T_i = 1$ and $T_i = 0$ which is known as the fundamental problem of causal inference (Holland, 1986). Using the counterfactual framework, the wage premium is defined by:

$$UWP = E(W_0|T = 0) - E(W_1|T = 0) \quad (2)$$

While the $W_0|T = 0$ can be easily observed (the wage of a worker in a thin labor market), the counterfactual $W_1|T = 0$ (the wage of the same worker in thick labor market) is unobservable. Note that treatment T is not a random selection, and there are several literatures discussing how the workers are moved toward a particular labor market. Moreover, the treatment $T_i = 1$ or

$T_i = 0$ is motivated by some observables and unobservable variables (X). We assume that we can observe X and the orthogonal condition (3) is satisfied.

$$(W_0, W_1) \perp T | X \quad (3)$$

Equation (3) is the Ignorable Treatment Assignment Assumption (ITAA)³, which allows identifying the spatial wage differential. The ITAA can be perfectly assumed in the case of randomized experiments, where the treatment is absolutely orthogonal to the output, but this assumption must be carefully considered in the case of quasi-experimental methods. A second assumption to be maintained is the Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1986). This assumption imposes that W for the worker i exposed to the treatment T is the same no matter what mechanism is used to assign the treatments to the rest of workers. According to Heckman (2005), this assumption rules out social interactions and general equilibrium effects. While the ITAA can be considered with a correct selection of the set X in the propensity score framework, we must necessarily maintain the SUTVA.

Both assumptions help to establish (2) as a measure of wage differential, but it cannot be empirically estimated: we cannot observe the worker under both conditions simultaneously. The matching estimator is proposed as an alternative to get an estimation of (2). The method proposes to identify a clone worker in the MA ($E(W_1 | T = 0)$) with a similar set of characteristics than the worker in a region r . The workers at r are called “treated”, while the counterfactual workers at thick labor markets are called “controls”. According to (3), the search of clones, hereafter called matching, must use the set of observable X . The matching is highly

³ This assumption has received several names in the literature: unconfoundness (Rosembaum and Rubin, 1983), selection on observables (Barnow, Cain and Goldberger, 1980), conditional independence (Lechner, 1999), and exogeneity (Imbens, 2004).

conditioned by the dimension of the set X . If X is a large matrix, the matching procedure can be problematic. The Coarsened Exact Matching is a new technique inside the family of matching estimators that leads the dimensionality problem. Iacus King and Porro (2008) proposes a two-step technique where in a first stage the variables are portioned in blocks in order to reduce the number of potential matching workers. In a second step, an exact matching is warranted among workers. Finally, with the new sample of workers the UWP is computed.

3. Data

a. Data for functional thick labor markets

A set of variables is used to identify thick labor market. Such as it was described previously; we define the new MA in order to represent the geographical area where the thick labor market is located. Firstly, we used a matrix of commuting flows between municipalities, including just workers, normalizing the row flows by the size of the total population in the origin area, this relative flows matrix was used to calculate Eigen-centrality indexes. Secondly, the population density by municipality was calculated considering total population divided by its area measured in kilometers squared. Thirdly, to represent the relative flows as a network, a particular geographical layout was setting using roads distances between municipalities, those distances were kilometers taken to reach a municipality's local capital from another, that matrix was reduced to a unique pair of vectors by metrical multidimensional scaling, and in that manner was possible to integrate the geographical space into the commuting network flows.

With the purpose to visualize the relative importance of each municipality in the commuting system of flows, two different kinds of algorithms were used. The first algorithm, the Fruchterman & Reingold (1991) layout (F-R) was used to produce a representation of the

networks of flows, in which those municipalities more central were located in the core of the system.

The F-R algorithm operates under two principles for graph drawing: (1). Vertices connected by an edge should be drawn near each other, and (2) Vertices should not be drawn too close to each other. To determine how close vertices should be placed some gravitational laws analogies are used, F-R assumed that vertices behave as celestial bodies, exerting attractive and repulsive forces on one another; those forces are modeled making only vertices that are neighbors attract each other, and at the same time, all vertices repel each other. The algorithm defines a particular static equilibrium to such system, in that way be close means to be connected but also it means to play a similar role in the relational system, in opposition be far means to be played an opposite role in that system.

The second algorithm, it is following Castell's ideas about the network society as a dichotomy between the space of flows and the space of places (Castell, 2006), its main purpose is to be able to mix geographical topology associated with the spatial distribution of municipalities with their role in the system of flows, indicated by their hierarchy in terms of Eigen-Centrality, it is assumed here, that thick labor markets are those areas more important in terms of flows system and to show their importance into the conventional geographical terms, it is necessary to weight their physical areas by their centrality. To mix both dimensions, the physical municipality area was weighted using its Eigen-Centrality importance using the Diffusion-Based method proposed by Gastner & Newman (2004). The diffusion process generates a "diffusion cartogram"; where areas of municipalities have been rescaled to be proportional to their Eigen-Centralities, in that manner thick labor markets appears a bigger areas in the map, in opposition thin labor markets are reduced to reflex their less importance in the system.

b. Data for Coarsened Exact Matching

This paper uses information from the Characterization National Survey 2009 (CASEN 2009), which is available at micro data level. The selected workers are 14 and 64 years and all of them receive a wage/hour greater than zero. We do not consider the army workers to properly represent the market equilibrium of labor demand and labor supply in the regional markets. The workers with economic sectors labeled as “Unspecified sectors” were also deleted. Finally, the outliers were treated using three different procedures. We run a regression by region where the dependent variable is the logarithm hourly wage and the independent variables are described in the table one. In a second step, we estimate the studentized residuals, standard residuals and *dfbeta*. We choose only those observations with *dfbeta* smaller than one and studentized and standard errors lower than three (absolute value).⁴ The table 1 reports a set of descriptive statistics for worker characteristics after the treatment of missing.

Table 1: Worker characteristics.

REGIÓN ADMINISTRATIVA	Sxh	Esc	Exper	Head	Casado	Contrac t	Permanent	Alphabe t	Participatio n	Observatio ns
	XV. ARICA Y PARINACOTA	2156,34	12,40	20,61	0,48	0,62	0,87	0,81	1,00	0,50	420
	I. TARAPACÁ	1997,28	11,82	20,40	0,52	0,63	0,88	0,80	0,99	0,52	877
	II. ANTOFAGASTA	2303,95	11,65	20,93	0,42	0,61	0,90	0,84	0,99	0,54	1,689

⁴ The Stata® code is available upon autor request.

III. ATACAMA	1925,03	11,30	22,15	0,53	0,66	0,87	0,77	1,00	0,49	1,129
IV. COQUIMBO	1754,28	11,29	22,53	0,50	0,58	0,80	0,73	1,00	0,53	1,597
R.M. METROPOLITANA DE SANTIAGO	2871,26	12,29	21,34	0,46	0,61	0,87	0,87	0,99	0,57	13,647
V. VALPARAÍSO	1820,20	11,71	22,12	0,50	0,61	0,86	0,79	0,99	0,51	5,775
VI. LIBERTADOR GENERAL BERNARDO O'HIGGINS	1886,84	11,18	21,85	0,48	0,62	0,88	0,67	0,98	0,55	3,444
VII. MAULE	1945,00	10,96	23,12	0,54	0,62	0,83	0,70	0,99	0,52	2,777
VIII. BÍO-BÍO	1752,75	11,68	22,00	0,52	0,64	0,86	0,75	0,99	0,48	6,415
XIV. LOS RÍOS	1854,44	11,24	22,00	0,56	0,64	0,86	0,78	0,99	0,48	1,103
IX. LA ARAUCANÍA	1999,98	11,42	21,78	0,52	0,67	0,83	0,75	0,98	0,49	2,385
X. LOS LAGOS	1731,35	10,98	22,09	0,50	0,64	0,84	0,77	0,98	0,54	2,26
XI. AYSÉN DEL GENERAL CARLOS IBÁÑEZ DEL CAMPO	2178,00	10,68	22,67	0,52	0,66	0,89	0,74	0,99	0,58	625
XII. MAGALLANES Y DE LA ANTÁRTICA CHILENA	2362,54	11,33	21,64	0,51	0,68	0,89	0,72	1,00	0,56	487

Such as it was previously discussed, Chile presents fifteen regions and they are presented in the first column of the table 1. Given its particular geographical shape, the geographical structure of the country is similar as the column is presented: “XV: Arica and Parinacota” is the region in the extreme north of the country and “XII: Magallanes y de la Antartica Chilena” is the extreme south. The second column describes an average hourly wage comparison, where the higher wage is observed in the Metropolitan Region of Santiago (MR). This exercise shows how the data suggest that the thick labor market generates a wage premium. Only a subset of three regions seem to have a wage close to the MR: II: Antofagasta, XI: Aysen and XII: Magallanes. This three are very specialized regions where the copper (II), salmon and oil (XI, XII) are the main industries. The second and third column represents the education (years) and experience. For the case of education, the MR presents the highest average education with almost 12.30 years. Head household and marital status represent two socio characteristics variables. We use this variable because of the responsibilities of the workers push the labor participation and the

incentive the workers to improve its human capital in order to increase the wages. The rest of the variables area standard set of human capital characteristics. Second sets of independent variables are related with the occupation of the worker: two workers can earn different wages according to the type of occupation. CASEN 2009 provides a rich data set of information and it is shown in Table 2

Table 2: Occupations

REGIÓN ADMINISTRATIV A	Poder Ejecuti vo	Profesionales científicos	Técnicos y Prof. nivel medio	Empleados de oficina	Vendedores, comerciantes	Agricultores y trab. calificado	Oficiales, operarios y artesanos	Operadores y montadores	Trabajadore s no calificados
XV. ARICA Y PARINACOTA	0,05	0,09	0,13	0,06	0,21	0,02	0,14	0,10	0,19
I. TARAPACÁ	0,03	0,10	0,10	0,08	0,21	0,01	0,15	0,18	0,15
II. ANTOFAGASTA	0,02	0,08	0,11	0,07	0,18	0,01	0,22	0,14	0,17
III. ATACAMA	0,02	0,06	0,10	0,08	0,15	0,02	0,22	0,16	0,19
IV. COQUIMBO	0,02	0,07	0,09	0,07	0,17	0,04	0,19	0,10	0,24
R.M. METROPOLITANA DE SANTIAGO	0,04	0,15	0,12	0,11	0,19	0,01	0,13	0,09	0,17
V. VALPARAÍSO	0,03	0,09	0,11	0,08	0,20	0,02	0,15	0,10	0,22
VI. LIBERTADOR GENERAL BERNARDO O'HIGGINS	0,01	0,09	0,09	0,08	0,14	0,03	0,13	0,11	0,31
VII. MAULE	0,02	0,11	0,09	0,06	0,18	0,03	0,13	0,09	0,30
VIII. BÍO-BÍO	0,02	0,11	0,10	0,08	0,17	0,02	0,17	0,12	0,20
IX. LA ARAUCANÍA	0,02	0,12	0,09	0,07	0,19	0,02	0,16	0,10	0,21
XIV. LOS RÍOS	0,02	0,10	0,10	0,12	0,17	0,05	0,13	0,14	0,17
X. LOS LAGOS	0,02	0,08	0,10	0,09	0,17	0,03	0,17	0,11	0,22
XI. AYSÉN DEL GENERAL CARLOS IBÁÑEZ DEL CAMPO	0,03	0,09	0,10	0,07	0,18	0,03	0,16	0,13	0,21

XII. MAGALLANES Y DE LA ANTÁRTICA CHILENA	0,06	0,07	0,09	0,09	0,13	0,06	0,21	0,13	0,16
---	------	------	------	------	------	------	------	------	------

An interesting pattern is detected in the table 2: the largest proportion of scientific and professionals are spatially sorted in the MR. A 15 % of the total sample of professional scientific are concentrated in this region against a 6% in III: Atacama. This value reveals the potential of matching techniques because a direct comparison between workers in the MR and the III region would be highly biased due the heterogeneity in human capital. Both set of workers are not comparable and, moreover, the large set of workers in the MR could be do not exist in the III region. The table 3 shows a descriptive statistics to analyze the economic sector of the regionals markets.

Table 3: regional Industrial mix composition.

REGIÓN ADMINISTRATIVA	Agricultur a, caza y silvicultur a	Explotaci ón minas y canteras	Industria Manufacture ra	Electricida d, gas y agua	Construcci ón	Comerc io / Rest. Hoteles	Transporte y comunicacio nes	Estableci mi. Financier os	Servicios comunal es y sociales
XV. ARICA Y PARINACOTA	0,05	0,07	0,09	0,00	0,13	0,24	0,06	0,08	0,28
I. TARAPACÁ	0,02	0,08	0,05	0,01	0,09	0,24	0,13	0,08	0,31
II. ANTOFAGASTA	0,02	0,17	0,09	0,02	0,11	0,18	0,11	0,07	0,24
III. ATACAMA	0,08	0,20	0,05	0,01	0,10	0,18	0,10	0,04	0,23
IV. COQUIMBO	0,14	0,09	0,05	0,01	0,13	0,20	0,08	0,05	0,24
R.M. METROPOLITANA DE SANTIAGO	0,02	0,01	0,13	0,01	0,09	0,22	0,10	0,13	0,30
V. VALPARAÍSO	0,08	0,02	0,08	0,01	0,11	0,22	0,10	0,08	0,30
VI. LIBERTADOR	0,22	0,07	0,09	0,01	0,08	0,19	0,07	0,04	0,24

GENERAL BERNARDO O'HIGGINS									
VII. MAULE	0,22	0,01	0,10	0,01	0,07	0,24	0,06	0,05	0,25
VIII. BÍO-BÍO	0,09	0,01	0,13	0,01	0,12	0,18	0,09	0,06	0,31
IX. LA ARAUCANÍA	0,08	0,00	0,09	0,01	0,12	0,22	0,08	0,06	0,32
XIV. LOS RÍOS	0,12	0,00	0,10	0,01	0,11	0,19	0,11	0,04	0,32
X. LOS LAGOS	0,10	0,00	0,10	0,01	0,12	0,24	0,09	0,04	0,30
XI. AYSÉN DEL GENERAL CARLOS IBÁÑEZ DEL CAMPO	0,07	0,01	0,07	0,00	0,18	0,18	0,08	0,06	0,36
XII. MAGALLANES Y DE LA ANTÁRTICA CHILENA	0,06	0,02	0,08	0,01	0,12	0,25	0,08	0,05	0,33

The table 3 completes the spatial sorting picture around the MR. Again; we can see a highest proportion of financial sector with a representation of the 13 percent. An interesting fact is the highly mining specialization of the II and III region where around of the 18 percent of the population is related with this sector. This event also supports the high level of specialization around a particular sector, which almost does not exist in the MR (1%). This set of tables support our hypothesis about the spatial sorting of human capital and about how this sorting must be controlled in order to warranty an appropriate wage comparison

4. Results

a. Thick and thin labor markets.

In Map No.1 is displayed the Chilean administrative political division at municipal level, considering 334 spatial units (left side). It is important to remark that Chilean territory is extremely long and narrow, near 4300 kilometers from North to South, and 177 kilometers as

average wide. The municipality's areas are also heterogeneous, typically those located in both extremes (south and north regions) are larger than those located in central area, this feature determines in an important manner average travelling time between municipalities, and it also affects the commuting flows among them. In that map, it is also remarked those municipalities (52) that are part of the Political Administrative Metropolitan Region of Santiago (where the Chile's capital city is located) because they are highly populated and they concentrate an important share of the total commuting flows. Using the method previously discussed a thick labor market was identified as a local spatial cluster of highly correlated values in terms of population density, high rates of commutation and higher values in terms of Eigen-Centrality indicators. In the map No1 (right side), the thick functional labor market was named Functional Metropolitan Area (FMA), its presence reduces the original number of administrative counties to a new total of 283 remaining municipalities, in that manner FMA implies to collapse 51 spatial original municipalities into FMA, those municipalities come from the administrative Region of Santiago (44), Valparaíso (1) and O'Higgins Region (6).

In Map No.2 (left side) is displayed a standard deviation map of population density, there should be noticed that central municipalities are highly dense populated and smaller in terms of geographical areas, in opposition extreme regions are bigger in terms of area but also less populated. In that map (right side), the results from the "diffusion cartogram" are presented, there those bigger areas represent municipalities more relevant in terms of their hierarchy of commuting flows (greater Bonacich's Eigen-Centrality). In that particular map, the identified thick labor market area (FMA), matches exactly with the more central spatial units, and for this reason they appear with greater weighted areas, indicating that it is the Core of the commuting

flows system and also the thicker labor market, in contrast the thin labor markets appear with reduced areas.

To understand why those municipalities are collapsing together into a FMA, it is necessary to describe the main characteristics of Chilean system of commuting flows. In Network 1, those flows are plotting in a F-R layout. That network corresponds to a valued and directed type, there nodes represent each municipality, and edges represent a connection among them as the relative commuting flow between them normalized by total population at origin place, to simplify that network connections a cutoff equal to 1% was used. The network shows that municipalities tend to cluster into cohesive groups, implying that cross-groups commuting are rare.

To clarify the role that geographical topology is playing there, in Network No. 2, the same flows are plotted, but using a layout based on first 2 dimensions from a Metrical Multidimensional Scalling (MMS), the MMS was applied over the matrix of travel distances among municipalities, in that manner vertexes located close means that are geographically close and/or between them it is relatively easy to commute, fact the normally occurs when they are located into the same political administrative region or between geographically contiguous areas. There, central – and densely interlinked- area is occupied by vertexes (municipalities) located in the Chilean central regions (Valparaiso, Metropolitana of Santiago, and O`Higgins), in contrast the extreme and less connected nodes are municipalities located in extreme South and North regions, in that fashion it is clear that geographical location (or the Castellian System of Places) plays an important role to configure the System of Flows, in our case commuting interactions.

Both networks contribute to explain why the Administrative Metropolitan Region of Santiago is not an appropriate benchmark region for estimating the UWP. The identified Functional Thick

Labor Market is absorbing a portion of other two regions and the rural side is left out. This new region is considered the new densest labor area of the country. We use this new geographical division (FMA) as the second scenario discussed in the introduction and also it was used for estimating a new UWP.

b. Estimation of UWP using both geographical divisions.

Our hypothesis establishes that the wage premium is better identified when the thick labor market is defined in a functional sense. We ground our hypothesis on the logic of economics of agglomeration: if the spatial spillovers explain the higher productivity of workers, then they exist in a labor market where the interaction between economic agents is warranted. However, the geographical scope of the market could not coincide with the administrative division. If the administrative area does not fit with the functional, then the wage gap could not be properly identified.

In order to test our hypothesis, we follow a two-step procedure. First, we estimate the wage differential between a thin labor market (each one of the fourteen Chilean political administrative regions) and the MR of Santiago (a region selected as benchmark because it is traditionally considered a thick labor market). In a second step, we repeat the same exercise, but using the FMA instead of the administrative one as control.

This testing strategy rests on the critical role played by a fair comparison of wages among a set of heterogeneous workers. In this sense, we estimate the differentials using a Coarsened Exact Matching. The CEM allows a resampling where only a set of comparable workers is considered. Given that, the balance is the core of CEM, we must provide a statistical measurement to evaluate how comparable are the workers after matching. The balance between both groups of

workers is evaluated with the L_1 statistics proposed by Iacus, King and Porro (2008). Instead of considering only a mean test, this statistics considers the proximity of multivariate histogram between groups. The statistics is:

$$L_1(f, g) = \frac{1}{2} \sum_{l_1 \dots l_k} |f_{l_1 \dots l_k} - g_{l_1 \dots l_k}|$$

The procedure builds a crosstab of k covariates for treated (thin labor market) and control (thick labor market) group. The frequencies for treated $f_{l_1 \dots l_k}$ and control $g_{l_1 \dots l_k}$ are used to estimate the balance measure. If L_1 is close to zero, then the balance property is satisfied. Otherwise, the matching procedure does not satisfy the core assumption. Formally, if the matched frequencies are labeled f^m and g^m , then we expect :

$$L_1(f^m, g^m) < L_1(f, g)$$

Assuming the frequencies without matching is labeled as $L_1(f, g)$ then we expect $L_1(f^m, g^m) < L_1(f^c, g^c)$. We estimate the L_1 for each region and the MR with and without matching.

Figure 1: L1 statistics for a particular region and MR with and without CEM

The Figure 1 contains the Chilean regions, on the horizontal axis they are displayed in a geographical order from North to South, and the vertical axis shows the L1 statistics. Those statistics represent the heterogeneity of workers before and after matching, using the administrative division as control. The CEM finds similar workers using the complete set of variables described on tables 1, 2 and 3. The black line (circles) represents the L1 statistic for each region in comparison to the MR as a thick labor market. If L1 is close to one, it means a higher heterogeneity between both groups. The figure 1 shows a high heterogeneity level for the whole set of regions, the most of L1 statistics are around 0,9. This result supports our hypothesis about the bias when the workers in a region are compared with workers in the MR without a proper matching procedure. Additionally, the statistics is reduced for those regions close to the MR that indicates that a higher bias is carried out for the extremes of the country. In the same figure 1, the gray line represents the L1 statistics, but when the CEM is applied. Clearly, the new

subsets of workers are more comparable and the bias is considerably reduced. These results show the advantages of the matching procedure and how when it is applied the worker's heterogeneity is minimized. After to get an homogenous groups of workers, the next step consists of estimating the wage differential between regions, using both definition of thick labor markets, the administrative one (MR) versus the functional one (FMA). As it was discussed in the introductory section, we expect that using MR is going to bias the **UWP** estimation because the presence of spatial mismatch affecting the selected control group.

After the matching procedure, each region owns a new subset of workers and the MR also presents a set of comparable workers. This group is equivalent according to the variables used to carry out the CEM. With this new set of comparable workers, we can estimate the wage differential as a simply means test between each thin and thick labor market. However, this approach would assume that there is not any kind of differential in human capital among workers. This strong assumption implies an exact and perfect matching, which is not truth by a statistical definition. For this reason, we estimate the mean test, but controlling again for the same variables utilized in the CEM procedure. This exercise is equivalent to Mincer equation with a dummy indicating if the worker is located in a thin or thick labor market, but assuring set comparable workers. We show the R-square for each Mincer regression using the MR as a thick labor market.

Figure 1: R square for the regression with and without economic sectors.

The Figure 2 shows the R squares of the regression after CEM. The fit level of the complete set of regressions is around of 0.5. This implies that almost the 50 percent of the wage variance is correlated with the variance provided by the set of independent variables plus the dummy indicator. This level of fit is increased for those regions located on the extremes, Norther and Souther part of Chile, and it shows the importance of controlling by economic sector such as mining or fishing, that are highly located in those extreme regions. Our final results are displayed in the figure 3 and 4.

The figure 3 shows the estimation of wage differential using both identified thick labor market areas (RM and FMA). The first result suggests the existence of a wage premium for almost the complete set of regions. In both cases, thick labor markets show a strong wage premium against the regions III, IV, V, VI VII, VIII, IX and X. It means that any set of comparable workers earns a higher wage in the thick market than in those thin labor markets.

This differential is between 10 and 20% for that set of regions. In contrast, for those two regions (XV and I) located in the extreme north also present a negative wage effect in comparison with the thick labor market, but it is small around 4.5%. Only three regions do not show wage premium, namely the II, XI and XII, but where all of them are characterized by particular industrial conditions. The existence of specialization in mining, fishing and fuel industry generate higher wages that thick markets.

Figure 2: Spatial wage differential using administrative and functional division

Figure 3: Bias of wage differentials using administrative versus functional division

The figure 4 provides the difference of UWP for both types of regions. The administrative system is always under-estimating the UWP. For the most of the regions, the bias is around 0 and 1 percent. However, the bias increases in the south extreme regions such as X, XI and XII. These results suggest that the spatial sorting of workers implies a high heterogeneity between these workers and those located in the Metropolitan Region of Santiago

5. Conclusions

This paper contributes in several dimensions. First, a new functional approach is built for the Chilean case. The Metropolitan Region of Santiago is a dense labor market and its geographical scope is broader than administrative boundaries. Moreover, a portion of the Metropolitan Region, particularly the rural areas, does not belong to the new thick labor market. These results

suggest evidence how the geographical extent of labor markets is not defined by administrative reasons, but rather by its functionality. U

The UWP is estimated using both spatial divisions. The results show how, after a matching procedure, a similar set of workers earn higher wages in the thick labor markets comparison for almost any region. We estimate this premium around 4 and 21%. Only three regions show a higher wage than thick market, but its difference is reduced with our functional approach. This exercise provides a first look for the Chilean case and the estimations suggest how the persistent concentration of economic activity increases the wage through the existence of a urban premium.

6. Bibliography

Network 1 - Commuting Flows Networks Representation.

Network 2- Commuting Flows Networks and Geographical Space

Map1: Administrative RM and Functional RM

CHILEAN MUNICIPALITIES

334 ADMINISTRATIVE SPATIAL UNITS

283 FUNCTIONAL SPATIAL UNITS

Map2: Administrative RM Characteristics and Functional RM relative importance

POPULATION DENSITY AND RELATIVE IMPORTANCE IN COMMUTING FLOW SPACE

Distribution of Total Commuting Rate by municipalities

Municipality Areas Weighted by Bonacich's Eigencentrality Importance in Commuting Flows Network

Legend

Municipality's W. Area

Other Regions

REGION

Adm. Metropolitan Reg

referencia:

Castells, M. (2006). "The Theory of The Network Society". MPG Books Ltd, UK.

Gastner, M.T. & M.E.J. Newman (2004). "Diffusion-Based Method for Producing Density-Equalizing Maps". PNAS, 101(20), 7499-7504.

Green, N. (2007). "Functional Polycentricity: A Formal Definition in Terms of Social Network Analysis". Urban Studies, 44, 2077-2103, doi:10.1080/00420980701518941

Bibliography

Anselin L. 1995. Local indicators of spatial association – LISA. *Geographical Analysis* **27**: 93-115.

Barnow B. S, Cain G.G, Goldberger A.S. 1980. Issues in the Analysis of Selectivity Bias. En Stromsdorfer, E. W. y Farkas, G. (eds.), *Evaluation Studies: Review Annual*, vol. 5, Sage Publications, Beverly Hills.

Berry C, Glaeser E. 2005. The Divergence of Human Capital Levels Across Cities. NBER Working Papers **11617**. *National Bureau of Economic Research, Inc.*

Bonacich P. 1972. Factoring and weighing approaches to clique identification. *Journal of Mathematical Sociology* **2**: 113–120.

Boots B, Kanaroglou P. 1988. Incorporating the effects of spatial structure in discrete choice models of migration. *Journal of Regional Science* **28**: 495-507.

Boots B, Tiefelsdorf M. 2000. Global and local spatial autocorrelation in bounded regular tessellations. *Journal of Geographical Systems* **2(4)**: 319-348.

Borgatti, S. 2005. Centrality and network flow. *Social Networks* **27(1)**: 55-71.

Castells M. 2006. *The Theory of The Network Society*. UK: MPG Books Ltd.

Combes P, Duranton G, Gobillon L. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* **63**: 732-742.

Elvery J. 2010. City size and skill intensity. *Regional Science and Urban Economics* **40(6)**: 367-379.

Fruchterman, T, Reingold, E. 1991. Graph Drawing by Force-Directed Placement. *Software: Practice and Experience* 21(11).

Gastner M, Newman M. 2004. Diffusion-based method for producing density-equalizing .

Getis A, Griffith D. 2002. Comparative spatial filtering in regression analysis. *Geographical Analysis* **34(2)**: 130-140.

Glaeser E, Mare D. 2001. Cities and Skills. *Journal of Labor Economics* **19(2)**: 316-42.

Glaeser E, Resseger M. 2009. The Complementarity between Cities and Skills. NBER Working Papers **15103**. *National Bureau of Economic Research, Inc.*

Green N. 2007. Functional Polycentricity: A Formal Definition in Terms of Social Network Analysis. *Urban Studies* **44**: 2077-2103, doi:10.1080/00420980701518941.

Heckman J. 2005. Lessons from the Technology of Skill Formation. NBER Working Papers **11142**. *National Bureau of Economic Research, Inc.*

Holland P. 1986. Statistics and causal inference. *Journal of the American Statistical Association* **81**: 945-970.

Iacus S, King G, Porro G. 2008. Matching for Causal Inference Without Balance Checking. *UNIMI - Research Papers in Economics, Business, and Statistics* unimi-**1073**, Università degli Studi di Milano.

Imbens G. 2004. Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *The Review of Economics and Statistics* **86(1)**: 4-29.

Kanbur R, Venables A. 2005. Spatial inequality and development overview of unu-wider project. URL <http://www.arts.cornell.edu/poverty/kanbur/WIDERProjectOverview.pdf>

Lechner M. 1999. Nonparametric bounds on employment and income effects of continuous vocational training in East Germany. *Econometrics Journal* **2(1)**: 1-28.

Marshall A. 1920. *Principles of Economics*. London, UK: MacMillan and Co.

Mion G, Naticchioni P. 2009. The spatial sorting and matching of skills and firms. *Canadian Journal of Economics* **42(1)**: 28-55.

Puga D. 1998a. Urbanisation patterns: European vs. less developed countries. *Journal of Regional Science* **38**: 231-252.

Puga D. 1998b. The rise and fall of regional inequalities. *European Economic Review*, forthcoming.

Rosenbaum P.R, Rubin D.B. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **70(1)**: 41-55.

Rosenthal S, Strange W. 2004. Evidence on the nature and sources of agglomeration economies. In Henderson J, Thisse J (eds.) *Handbook of Regional and Urban Economics* 2120-2171.

Rubin D.B. 1986. Which ifs have causal answers. *Journal of the American Statistical Association* **81**: 961-962.

Venables A. 2005. Spatial disparities in developing countries: cities, regions and international trade. *International Economic Review* **5**: 3-21.