



Desarrollo y Sociedad

ISSN: 0120-3584

revistadesarrolloy sociedad@uniandes.ed

u.co

Universidad de Los Andes
Colombia

Cárdenas Hurtado, Camilo Alberto; Hernández Montes, María Alejandra; Torres Gorrón,
Jhon Edwar

A Statistical Analysis of Heterogeneity on Labour Markets and Unemployment Rates in
Colombia

Desarrollo y Sociedad, núm. 75, enero-junio, 2015, pp. 153-196

Universidad de Los Andes
Bogotá, Colombia

Available in: <http://www.redalyc.org/articulo.oa?id=169140726004>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

A Statistical Analysis of Heterogeneity on Labour Markets and Unemployment Rates in Colombia

Un análisis estadístico de la heterogeneidad en los mercados laborales y las tasas de desempleo en Colombia

Camilo Alberto Cárdenas Hurtado¹
María Alejandra Hernández Montes²
Jhon Edwar Torres Gorron³

DOI: 10.13043/DYS.75.4

Abstract

In this paper, we study the structural factors that determine the differences in unemployment rates and in labour market performance for Colombian cities. Using cross-sectional data for 23 metropolitan areas, we apply an extension of a principal axes method—Multiple Factor Analysis for Multiple Contingency Tables (MFACT)—in order to identify unobservable factors that are relevant when disentangling the heterogeneity observed among groups of variables considered explanatory of regional unemployment differentials. Our findings suggest that differences in qualified labour supply levels, participation incentives and age structure are important when it comes to understanding regional heterogeneity in terms of labour markets and unemployment rates in Colombia.

1 Banco de la República. Bogotá, Colombia. Corresponding author. ccardehu@banrep.gov.co.

2 Banco de la República. Bogotá, Colombia. mhername@banrep.gov.co.

3 Banco de la República. Bogotá, Colombia. jtorrego@banrep.gov.co.

Este artículo fue recibido el 28 de febrero de 2014, revisado el 24 de Junio de 2014 y finalmente aceptado el 13 de mayo de 2015.

In addition, clustering methods reveal that cities that display high unemployment rates do not necessarily share the same structural characteristics; that is, labour market frictions that give rise to unemployment are not the same across Colombian cities.

Key words: Heterogeneous unemployment rates, regional labour markets, factor analysis.

JEL classification: R23, J40.

Resumen

En este artículo estudiamos los factores estructurales que determinan las diferencias en las tasas de desempleo y en los mercados de trabajo de las ciudades colombianas. Utilizando información de corte transversal para 23 áreas metropolitanas, aplicamos una extensión de un método de ejes principales, análisis factorial múltiple para múltiples tablas de contingencia (MFACT), con el fin de identificar factores no observables que son relevantes para entender la heterogeneidad observada entre grupos de variables que se consideran explican las diferencias regionales en las tasas de desempleo. Nuestros resultados sugieren que las diferencias en los niveles de mano de obra calificada, incentivos a la participación laboral y la estructura etaria son importantes para entender la heterogeneidad de los mercados de trabajo y de las tasas de desempleo en Colombia. Además, un ejercicio de *clustering* revela que las ciudades con altas tasas de desempleo no necesariamente comparten las mismas características estructurales, esto es, las fricciones en el mercado de trabajo que dan origen al desempleo no son las mismas en las ciudades colombianas.

Palabras clave: tasas de desempleo heterogéneas, mercados laborales regionales, análisis factorial.

Clasificación JEL: R23, J40.

Introduction

The high levels and persistence of unemployment rates, together with the complex dynamics observed in labour market structures in Colombia, have puzzled local economists for decades now. Although some issues have been studied

over the past few years (see, for example, Arango and Hamann, 2013; Urrutia, 2001), there are still several unanswered questions that, if solved, might lead to a better understanding of the convoluted particularities of labour market institutions in our country.

One of the most unexplored topics in Colombian labour market literature is regional unemployment, as stated by Arango (2013). Some pioneer works explaining regional and urban unemployment in Colombia are those by Jaramillo, Romero and Nupia (2000), Galvis (2002), Gamarra (2005) and Barón (2013). However the topic has not been fully explored. Arango (2013) points out that there are noticeable differences between Colombian cities when analysing labour market performance over the past few decades. His findings show that there is an evident heterogeneity between cities in terms of labour market indicators such as the unemployment rate, participation rate, occupation rate, underemployment rates, salaries and education.

He shows that some cities, such as Pereira, Popayán and Quibdó, have persistently displayed high unemployment rates over the past few years, while others, like Bogotá, Barranquilla, Bucaramanga and Cali, have seemingly performed better over the same time span. There are several *feasible* explanations for these differences, but still not a single definite one. Very few articles (if any at all) have explored the driving factors that determine the contrasts in unemployment rates between regions or cities in Colombia. For example, to the best knowledge of the authors, only Díaz (2011) has provided valuable evidence of spatial clustering of different types of municipalities that share differences in economic and socio-demographic attributes. The author claims that unemployment and labour market performance in these cities rely heavily on the geographical distribution of those attributes. However, we argue that factors that determine urban and regional unemployment rate differentials do not necessarily depend on spatial interactions between labour markets, but are inherent to the labour market structure itself. We expect to find that the municipalities that share common factors in labour market structures also display similarities in their unemployment rate levels.

This article explores such differences by analysing the determinants of differentials in unemployment rates for a set of Colombian metropolitan areas, following the framework proposed by Elhorst (2003). We build a high dimensional dataset for these cities and by studying the relationships between variables and,

among observations; we aim to find the structural factors that help understand the regional heterogeneity in labour market indicators described by Arango (2013). In order to identify such factors, we use exploratory multivariate statistical analysis techniques. These methods, which are well known for their suitability for dimensionality reduction, allow for the synthesis of information encoded in a high dimensional dataset into a lower dimensional space of factors that admit graphical representations and an easier interpretation. The resulting factors will be interpreted as structural variables that explain regional differences in unemployment rates.

We rely on Multiple Factor Analysis methods (MFA, Escofier and Pagés 2008) and their extension to tables containing various contingency tables (MFACT), introduced by Bécue-Bertaut and Pagés (2004, 2008). In contrast to other factor methods, the main characteristic of this methodology is that it summarizes a dataset composed of both continuous and discrete variables, and various contingency (frequency) tables, into a new set of factors that can be projected in a lower dimensional space. Therefore, we can take advantage of different types of data that might be useful to understand labour market structures and unemployment in Colombian metropolitan areas. We are also interested in discovering whether cities can be grouped into different clusters that share common structural determinants of regional unemployment differentials. These clusters are built based on the resulting factors, which imply that geographical location is not necessarily determinant on their construction.

This article consists of six sections, including this Introduction. In Section I, we describe the determinants of differentials on regional labour markets proposed by Elhorst (2003), enriched by a complementary literature review. Section II describes the statistical methodology used in this paper and the data. Section III covers the main results of the MFACT exercise. Clustering results are shown in Section IV. The final section concludes and suggests that in order to reduce unemployment rates and assure better labour conditions in Colombian cities, it is important to count for the heterogeneity observed in regional labour markets. Our results also suggest that unemployment is the result of several different frictions in labour markets and should not be studied from a single perspective.

I. Explaining Regional Labour Market Differentials

Regional heterogeneity in labour markets and unemployment rates are topics that have been long addressed from both theoretical and empirical perspectives.

The academic literature on this subject has benefited from the contributions made by the so-called *new economic geography (NEG)* and the equilibrium-disequilibrium theories. The former suggests that the presence of economies of scale in a certain location might foster productivity gains, industrial clusters and urban development, which in turn allow for lower unemployment rates when comparing these regions to sparse, non-developed peripheral regions. Recent advances in NEG suggest that the factors that yield agglomeration and regional productivity differences are also the ones that induce unemployment disparities (Epifani and Gancia, 2005). On the other hand, equilibrium-disequilibrium theories argue that unemployment rate differentials will arise as a result of labour mobility restrictions and the presence of amenities that might attract labour supply to a certain city or region (Blanchard and Katz, 1992; Marston, 1985). However, we do not aim to discuss theoretical models concerning regional unemployment differentials; instead, we focus on the empirical perspective.

According to Elhorst (2003), variables that explain differentials in regional unemployment fall into one or more of the categories here presented. On one hand, there are endogenous variables that are related to the city's population and the dynamics of regional labour markets; on the other, there are exogenous variables that are not directly related to the labour force or the equilibrium reaching mechanism. We stress that no attempt is made to be exhaustive in reviewing the existing literature, since it is not the main goal of this paper. We focus on influential papers on regional unemployment topics that have enriched labour economics literature over the past few decades. Accordingly, Elhorst states that variables can be categorized into one of the following groups:

A. Demographic Structure

Variables such as birth rate, age structure and other related demographic indicators have been found to be determinant on the labour supply size in the long run (Biffl, 1998; Chawla, Betcherman and Banerji, 2007; Lerman and Schmidt, 1999). A region will display persistence in its unemployment rate if its population growth is higher than the employment creation rate. In addition, when the age structure of the population is skewed towards young and old individuals, the region is more likely to display high unemployment rates (Lottman, 2012).

B. Participation

Mixed results have been found when assessing the significance of these kinds of variables in explaining regional unemployment differentials. Authors usually think of a positive (non-linear) relationship between unemployment and participation rates. However, it has also been found that higher unemployment rates are usually accompanied by low participation rates. Several explanations arise: according to Fleisher and Rhodes (1976), low participation rates might reflect low levels of human capital investment and low levels of labour commitment. Also, lower female participation rates are often explained by the presence of children in the household. The latter implies a trade-off for female workforce between having a family and pursuing a career (Martínez, 2013). Finally, changes in participation rates greater than those in occupation rates might also yield higher unemployment levels (Blundell and MaCurty, 1999; Da Rocha and Fuster, 2006).

C. Migration

Immigrant flows influence participation rates and reinforce the effects reported for participation variables. Also, these flows have been found to be correlated with regional disparities in economic performance and labour market conditions (Blanchard and Katz, 1992; Pissarides and Wadsworth, 1989). However, the effect depends heavily upon the initial endowments (both human and physical capitals) of the incoming population: If high, demand for qualified workforce is likely to increase, as are net investment rates and aggregate productivity (Eggert, Krieger and Meier, 2010; Moretti, 2012). If low, however, new inhabitants will enter low skilled unemployment lines, as demand for this type of labour might not increase as fast as supply does (Walden, 2012). For the Colombian case, Barón (2013) reported that workforce mobility was limited and did not have a significant effect on labour market indicators, but responded to economic differences between regions.

D. Commuting

Détang-Dessendre and Gaigné (2009) found that long traveling times and long-distance commuting have significant effects on unemployment duration and labour market mismatching. Also, Brueckner, Thisse and Zenou (2002) argued that firms' market power when hiring new personnel is higher when workers incur on high commuting costs, measured in both time and money spent.

E. Wages

Theoretically, higher wages usually have a positive effect on labour supply and a negative effect on labour demand and, in frictionless models; wages are the result of the labour market equilibrium reaching mechanism (e.g. Cahuc and Zylberberg, 2004, Ch. 5-7). Also, frictions related to workforce mobility between regions or cities yield regional wage differentials (Bande, Fernández and Montuenga, 2008). Lastly, wages serve as a productivity measure: differentials in wages across regions can occur due to differences in labour productive skills (Burdett and Mortesen, 1998).

F. Regional Growth

Regions with good economic performance usually display low (structural) unemployment rates and high productivity indicators. This result can be encompassed in Okun's law framework (Okun, 1962), but at regional level, as in Oberst and Oelgemöllér (2013).

G. Market Potential

Location factors matter for labour market dynamics: firms tend to settle in regions where there is growth potential in terms of sales and stable household consumption perspectives, among other reasons (Krugman, 1995). As a consequence, unemployment rates will be lower in those regions. In addition, some approaches have argued that innovation plays a key role in unemployment reduction. Innovative sectors attract skilled labour force and have multiplier effects on employment in other sectors (Moretti, 2010, 2012).

H. Economic Structure

Regions with a diversified productive structure may be less affected by sector-specific shocks and, therefore, exhibit lower unemployment rates throughout the business cycle, as argued by Malizia and Ke (1993) and Izraeli and Murphy (2003).

I. Economic and Social Barriers

These are unobservable economic and social variables that discourage workforce mobility between regions or cities and, therefore, act as frictions in

regional labour markets (Elhorst, 2003). Frictions in real-estate markets, welfare and social security programs, and general tightness of labour markets are some of the variables in this group. Lottman (2012) and Walden (2012) provide some recent empirical evidence on this topic.

J. Education

Higher levels of educational attainment lower the risk of unemployment, increase the likelihood of higher wages, and promote labour mobility between regions (Mincer, 1991). Also, it has been empirically tested that high levels of human capital stocks have spillover effects over the non-educated population in labour market outcomes (Winters, 2013). Although the overall quality of workforce skills cannot be entirely measured by the average number of years spent in education, it is a sufficient indicator that has been found to be negatively correlated with the unemployment rate, even at regional level (Eggert, Krieger and Meier, 2010).

K. Unionisation

From a theoretical perspective, the bargaining power of unions has been treated as a distortion that deviates the labour market from its competitive equilibrium (Cahuc and Zylberberg, 2004, Ch.7). Unionisation has been found to be correlated with lower labour demand and also to be an influential variable in the wage setting mechanism, as argued by Mincer (1981), Lewis (1986), and Farber (1986). More recently, the role of unionisation in the labour market has been explored by Albagli, Garcia and Restrepo (2004), Freeman (2009), and Krusell and Rudanko (2013).

L. Regional Natural Unemployment Rate and Persistence

Some authors argue that heterogeneity in regional unemployment arises due to differences in persistence and natural unemployment rate measurements between regions. This approach has been treated as a purely statistical problem in a wide range of empirical studies, such as Brunello, Lupi and Ordine (2000), Gomes and da Silva (2009), Lanzaflame (2010) and de Figueiredo (2010).

II. Methodology and Data

A. The General Factor Analysis Method

Exploratory data analysis (EDA) is the process by which a researcher extracts vital information from a large dataset, allowing him to understand the underlying structure that rules the relationships between observations and variables as well as determine how 'related to' or 'different from each other' they are. Those relationships and contrasts are often thought to be driven by a series of non-observable variables, known as *factors*, which are obtained from the data using statistical methods.

Factor analysis methods are exploratory multivariate statistical techniques that aim to produce a simplified representation of the variance (inertia) structure of a high dimensional dataset. Factors themselves form a set of variables that belong to a lower dimensional space; *i.e.*, factor analyses reduce the dimensionality of the original problem, thus, allowing for more easily formulated conclusions and interpretations of the observed data. A high inertia structure yields greater heterogeneity among the individuals within the sample, which, in turn, is evidenced in the values that the factors attain for each observation. For example, if observation i scores high on the first factor, and observation i' scores low, then it means that i and i' are different (heterogeneous) along the set of observed variables that are summarized by this factor. Therefore, the analysis is restricted to a one-dimensional problem, rather than over the whole set of dimensions (variables) that were initially considered.

However, dimensionality reduction comes at a cost. In order to ensure that the database's original inertia structure remains as unchanged as possible, we have to place some restrictions on the exercise of dimensionality reduction. Factor analysis reduces to a restricted maximization problem: we aim to maximize the information contained in the original dataset into a lower dimensional set of variables, guaranteeing that each one of the resulting factors carry different pieces of information (*i.e.*, are independent from each other). A mathematical approach is readily presented, but for introductory, yet comprehensive, references on this topic see Escofier and Pagés (2008), Johnson and Wichern (2007), and Peña (2002).

Following, we explain how these factors are obtained. Let X be a I (individuals) $\times K$ (variables) matrix. Since $X \subseteq \mathbb{R}^K$, we can define a metric on X in order to measure the distance between any two points x_i and x_j , $i, j \in I$. The weights used when computing these distances, labelled m_k for each variable⁴, are defined by the $K \times K$ matrix M . Usually M is defined as an identity matrix of size K . In fact, when M is diagonal (i.e. $m_k \in \text{diag}(M)$), the distance between points i and j is computed as $d^2(i, j) = \sum_{k \in K} (x_{i,k} - x_{j,k})^2 \cdot m_k$. Since m_k weights the influence of each variable $k \in K$ in the computation of this distance, M is usually understood as the "columns' weights" matrix.

The shape of the individuals' cloud in \mathbb{R}^K is completely defined by X and M . However, when calculating the inertia structure (variance) of X , the weight associated to an individual i , p_i , enters into the computation. These weights are ordered in a diagonal matrix D of rank I , i.e. $p_i \in \text{diag}(D)$. The more heterogeneous the individuals, the richer the inertia structure of X .

Now let $F_{u_h} = XM u_h$ be the projection of X over a single vector u_h in \mathbb{R}^K . The variance of X projected over u_h is $\sum_{i \in I} p_i [F_{u_h}(i)]^2 = F_{u_h}' D F_{u_h} = u_h' M X' D X M u_h$. Factor analysis methods aim to find a new set of orthonormal vectors u_h , $h \in H$ ⁵, such that the inertia projected over each one of them is maximized. The set of unitary vectors u_h that satisfy

$$u_h \in \arg\max_{u_h \in \mathbb{R}^K} \{\text{Inertia}(X M u_h) := u_h' M X' D X M u_h\}, \text{ s.t. } u_h' u_h = \|u_h\| = 1; \quad (1)$$

are the eigenvectors of the diagonalizable matrix $X' D X M$ ordered according to their associated eigenvalues ranging from the highest (in absolute value), λ_1 , to the lowest, λ_K . Note that, by construction, the inertia projected over u_h will be λ_h , $\forall h \in H$, i.e., $\text{Inertia}(X M) = \sum_{k \in K} \lambda_k$.

Principal Component Analysis (PCA), Correspondence Analysis (CA) and Multiple Correspondence Analysis (MCA) are specific cases of this general factor method, and each one has its own specification for matrices X , D and M . For a detailed presentation of each method, see Escofier and Pagès (2008, Ch. 1-4) and Greenacre (2007).

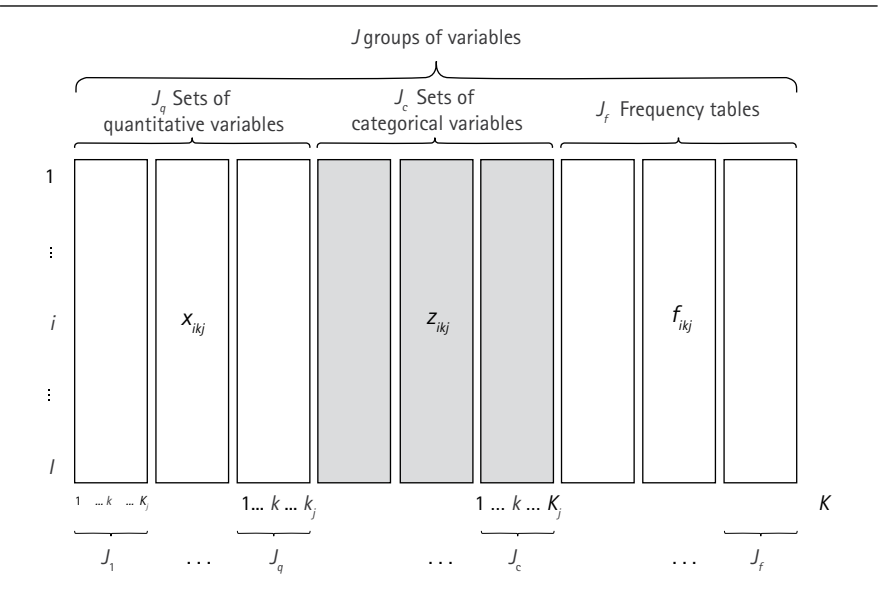
4 The subindex k denotes the k^{th} element of $\text{diag}(M)$.

5 Again, subindex h denotes the h^{th} vector belonging to a set of cardinality $H \leq K$.

B. Dealing With Mixed Datasets

Usually observations are simultaneously described by joint juxtaposed sets of *quantitative* (numerical) or *qualitative* (categorical) variables, and *contingency tables*, as shown in Figure 1. In this $I \times K$ matrix, we have J groups of variables: J_q quantitative groups, J_c categorical groups and J_f frequency tables. Each group has K_j variables, which means that $\sum_{j \in J} K_j = K$. For notational purposes, x_{ikj} corresponds to a numerical realization and z_{ikj} is a dichotomous variable that assigns 1 if X_i belongs to category k in K_j or 0 if not. f_{ikj} is the ratio of the number of occurrences of x_i for variable $k \in K_j$ to the total number of realizations on the contingency table, i.e., $f_{ikj} = x_{ikj} / \sum_i \sum_k x_{ikj}$.

Figure 1. Global Table



Source: Author's adaptation from Bécue-Bertaut and Pagès (2008).

Multiple Factor Analysis for Contingency Tables (MFACT) was developed by Bécue-Bertaut and Pagès (2008, 2004) to deal with mixed datasets (Figure 1). In MFACT, the distance between individuals is determined by the information available on the numerical and categorical variables, and the contingency

tables. This represents an advantage in comparison to the separate analysis approach using PCA, MCA, and CA, respectively. In MFACT, the weight of each variable k belonging to a group j , m_k^j , is standardized by the first eigenvalue computed in each individual analysis on j , λ_1^j , i.e. new columns weights are $m_k^j / \lambda_1^j, \forall k \in K, \forall j \in J$. Readers are referred to Escofier and Pagès (1994, 2008) and Pagès (2002, 2004) for an explanation of MFA methods.

In sum, MFACT can be understood as a general factor method applied to a global table X subject to some previous transformations (which depend on the nature of the variables), with a specific metric M and the rows weights D . Matrices are specified in Table 1. Supplementary projections and graphical representations are also supported.

Table 1. MFACT Matrices

	Quantitative Variables	Categorical Variables	Frequency Tables
X	$\frac{x_{ikj} - \bar{x}_{kj}}{s_{kj}}$	$\frac{z_{ikj} - (\sum_{i \in I} p_i \cdot z_{ikj})}{\sum_{i \in I} p_i \cdot z_{ikj}}$	$\frac{f_{ikj} - \left(\frac{f_{i \cdot j}}{f_{\cdot j}} \right) \cdot f_{\cdot kj}}{f_{i \cdot \cdot} \cdot f_{\cdot kj}}$
M	$\frac{1}{\lambda_1^j}$	$\frac{\sum_{i \in I} p_i \cdot z_{ikj}}{\sum_{k \in K_j} \sum_{i \in I} p_i \cdot z_{ikj} \lambda_1^j}$	$\frac{f_{kj}}{\lambda_1^j}$
D	p_i		

Source: Author's adaptation from Bécue-Bertaut and Pagès (2008).

Dealing with a mixture of quantitative, categorical and frequency tables in the global analysis brings forth a number of issues when deciding which weights are assigned to individuals. On PCA and MCA (quantitative and categorical tables) individual weights are set according to the user's preferences and are usually fixed to be uniform across all rows ($p_i = 1/I$). However, on a multiple contingency table, individual weights are determined by the row margins ($p_i = f_{i \cdot}$, where $f_{i \cdot} = \sum_{k \in K} \sum_{j \in J} f_{ikj}$). MFACT can operate under any specifica-

tion of matrix D . We set D as in Bécue-Bertaut and Pagès (2008) (i.e. $p_i = f_{i..}$), to favour cities with greater populations and to avoid distorted results influenced by uniform individual weights.

C. Data

We build a large dataset consisting of 182 variables measured for 23 Colombian urban areas⁶, which are further categorized into 23 groups.

The 182 variables are classified into groups that belong to two broad categories: *quantitative variables* (119) and *contingency (frequency) tables* (63). Quantitative groups are: demographic variables ($Demo_c$), participation ($Part_c$), inter-regional migration (Mig_c), commuting (Mob_c), market structure ($Mktst_c$), regional growth ($Regw_c$), market potential ($Mktp_c$), educational attainment (Edu_c), wages (Wag_c), unionisation ($Unio_c$) and economic and social barriers ($Esbr_c$). In addition, 12 contingency tables that count for 63 variables are constructed: age structure ($Demo_f1$), age structure for men ($Demo_f2$), age structure for women ($Demo_f3$), marital status for men ($Part_f1$), marital status for women ($Part_f2$), waged employment structure ($Mktst_f1$), employment structure by occupational position ($Mktst_f2$), employment structure by economic sector ($Mktst_f3$), educational attainment structure for unemployed population (Edu_f1), educational attainment structure for employed population (Edu_f2), educational attainment structure for working age population (Edu_f3) and educational attainment structure for inactive population (Edu_f4). The dataset was constructed with information obtained from the National Statistics Administrative Department (DANE), Ministry of Education (MEN), Ministry of Finance (MHCP), Department for Social Prosperity (DPS), Economic Commission for Latin America and the Caribbean (ECLAC), Observatory for the Colombian Caribbean (Ocaribe) and the Central Bank of Colombia (Banco de la República). A description of the dataset is presented in Appendix 1 and is available upon request. Due to the lack of information availability in

6 8 metropolitan areas (Bogotá, Medellín, Cali, Barranquilla, Bucaramanga, Cúcuta, Pereira and Manizales) that sum a total of 52 municipalities, and 15 capital cities (Pasto, Ibagué, Montería, Cartagena, Villavicencio, Tunja, Florencia, Popayán, Valledupar, Quibdó, Neiva, Riohacha, Santa Marta, Armenia and Sincelejo) where representative samples were obtained.

eight variables for two cities⁷, we used the method presented in Husson and Josse (2013) to handle missing data in our sample.

Given that according to Elhorst (2003), these variables are the structural determinants of regional unemployment differentials, we expected the results not to depend heavily on the year for which this exercise was computed. We therefore chose 2010 for the analysis.

III. Results

Before presenting any results, we recall the fact that no assumption is made on the multivariate distribution of the data. This means that no probabilistic results shall arise from an MFACT exercise and, therefore, we will not make any kind of statistical inference from the dataset. Computations were made using the statistical software R (R Core Team, 2013), and the FactoMineR package (Husson, Josse and Lê, 2008). We also point out that we project variables belonging to groups *Demo_f1*, *Mktst_f1* and *Edu_f3* as supplementary, so these groups do not add any extra information to the principal axes computations⁸.

The results for each separate analysis reveal a rich variance structure for each group of variables, providing strong evidence for a MFACT approach. The number of factors that summarize the total inertia of the original dataset ($\sum_{j \in J} \lambda_j = 26,6$) are the projections of matrix **XM** over the eigenvectors whose eigenvalues are greater than unity. 76.1% of this is summarized in the first five principal axes (Figure 2).

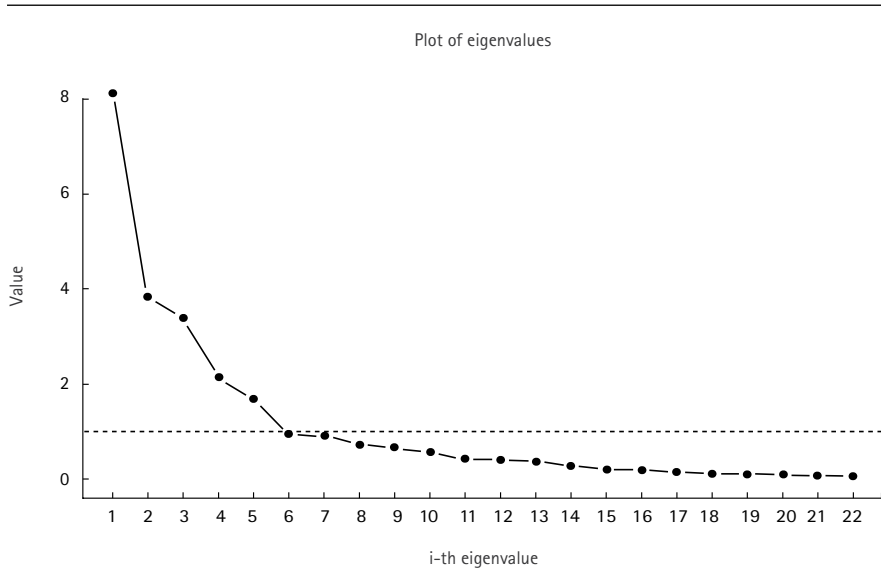
A. Interpreting Principal Axes

We name the resulting factors after the groups of variables that contribute the most to the inertia projected onto each dimension, as highlighted in Table 2. High correlations are also observed for the contributing groups.

7 Herfindahl and Hirschman's index for exports diversity, weighted distance to closest markets, Herfindahl and Hirschman's index for market diversity, firm's efficiency index, industrial density, store construction costs, registration costs and sale taxes for Quibdó and Florencia.

8 By construction $Edu_f1 + Edu_f2 + Edu_f4 = Edu_f3$, $Mktst_f2 + Mktst_f3 = Mktst_f1$ and $Demo_f2 + Demo_f3 = Demo_f1$

Figure 2. Eigenvalues: Scree Plot for Global Factor Analysis



Source: Author's own calculations.

The first principal axis ranks cities according to their population's educational attainment, workforce productivity and their occupational positions. This axis explains 32.8% of the total variance ($\lambda_1 = 8,7$) and is associated with variables such as number of waged workers, people with 13 or more years of formal education or nominal and real wages. We label this dimension an "*index for quality of labour supply*".

The second factor has high loadings on participation variables and educational attainment of the unemployed and inactive population. This dimension counts for almost 16.0% of the total variance ($\lambda_2 = 4,2$). Cities that display negative values in this factor are those with high remittances per capita, a demographic structure biased towards the older population and high unemployment rates for low skilled workers. In contrast, cities that display positive values are those that exhibit higher unemployment rates in the skilled population and show low participation in the labour market. For these reasons, the second axis has been labelled as the dimension for "*participation and skilled job demand frictions*".

Table 2. Inertia Contribution and Correlations of Each Group of Variables

Groups	Contribution (%)					Correlation				
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Demo_c	3.13	1.69	2.80	15.22	3.26	0.61	0.38	0.34	0.57	0.24
Part_c	0.15	13.88	5.16	0.26	0.10	0.12	0.77	0.42	0.28	0.06
Mig_c	2.16	0.04	13.85	0.99	11.79	0.44	0.08	0.69	0.43	0.44
Mob_c	5.95	2.31	1.32	0.08	0.62	0.72	0.31	0.21	0.04	0.10
Mktst_c	6.18	1.13	6.89	3.66	4.39	0.74	0.34	0.51	0.44	0.63
Regw_c	2.39	0.72	3.00	6.08	11.46	0.53	0.35	0.33	0.41	0.45
Mktp_c	6.68	0.60	3.00	4.75	5.14	0.76	0.16	0.32	0.32	0.29
Edu_c	5.48	0.83	17.63	18.04	2.70	0.89	0.26	0.79	0.70	0.42
Wag_c	7.54	2.71	0.12	3.91	5.62	0.81	0.37	0.17	0.29	0.31
Unio_c	1.16	0.60	4.31	15.00	0.33	0.32	0.16	0.38	0.57	0.07
Esbr_c	6.22	2.67	8.82	4.04	7.90	0.77	0.34	0.65	0.41	0.36
Demo_f2	7.36	9.23	1.11	0.75	0.66	0.82	0.68	0.33	0.23	0.12
Demo_f3	6.69	9.71	0.70	0.34	0.47	0.78	0.67	0.22	0.18	0.12
Part_f1	5.90	4.98	8.65	0.24	7.06	0.73	0.52	0.62	0.14	0.39
Part_f2	6.29	5.53	6.80	0.33	6.13	0.75	0.59	0.50	0.22	0.39
Mkts_f2	7.97	1.72	3.60	3.61	7.99	0.84	0.44	0.43	0.47	0.39
Mkts_f3	6.40	3.26	1.71	7.64	11.69	0.81	0.55	0.28	0.54	0.53
Edu_f1	1.57	17.59	3.25	6.07	2.95	0.43	0.87	0.49	0.58	0.34
Edu_f2	8.63	6.03	4.56	4.09	1.01	0.88	0.57	0.68	0.57	0.23
Edu_f4	2.14	14.74	2.71	4.88	8.75	0.49	0.80	0.68	0.56	0.40
Supplementary groups										
Demo_f1	6.96	9.60	0.87	0.47	0.54	-	-	-	-	-
Mkts_f1	7.82	0.29	1.39	1.77	5.88	-	-	-	-	-
Edu_f3	6.24	8.95	3.78	4.51	1.27	-	-	-	-	-

Source: Author's own calculations.

The third axis, which explains 12.9% of the total variance ($\lambda_3 = 3,4$), is related to education, migration and economic and social barriers groups. In this dimension, cities are projected according to their middle and higher public education coverage, net migration rates between and within (from rural to urban spaces) regions, and (negatively) royalties per capita. This dimension summarizes the differences between cities on a basis of opportunity. We label this axis as a *"public education efficiency and migration vulnerability index"*.

The fourth axis, that accounts for 8.1% of the total inertia ($\lambda_4 = 2,2$), has high loadings on basic public education coverage, demographic (race) and unionization variables. Positive values in this dimension suggest high proportions of Afro-descendant populations and a low proportion of unionised workers. We think of this axis as a *"non-wage rigidities dimension"* in the labour market. The fifth axis explains 6.3% of the total variance ($\lambda_5 = 1,7$), and it has high loadings on migration, regional growth groups and labour market structure. This axis is interpreted as the *"economic diversity axis and labour absorption capacity"*.

B. Interpreting Cities' (Individuals) Projections

The heterogeneity of a cloud of I individuals in \mathbb{R}^K is best understood by analysing the inertia of separate clouds in \mathbb{R}^2 . Cities' projections are the ordered pairs $(F_{u_h}(i), F_{u_{h+1}}(i))$, for $i = 1, \dots, 23$ and $h = 1, \dots, H$.

1. First Principal Plane (First and Second Dimensions – Figure 3A)

Cities projected along the first dimension are ordered according to their work-force skills, wages and occupational positions. On the positive side of the axis, we identify cities with high levels of qualified labour supply, better salaries and more stable, productive and promising job positions, as in Bogotá, Medellín and Bucaramanga. In contrast, cities with low human capital stocks, low wages and poor educational conditions, such as Quibdó, Florencia, and Riohacha, are projected onto the left side of the axis.

Pereira, Cali and Manizales are projected onto the negative side of the second dimension, opposed to cities like Tunja, Cartagena and Quibdó. That is, cities where low (high) unemployment rates prevail for the educated population are located on the negative (positive) side of the axis. Also, Pereira, Cali and Armenia are among the cities that receive greater inflows of remittances, in contrast to Tunja, Riohacha and Quibdó. Remittances are thought to discourage participation in urban labour markets.

Projections for the first principal plane give an initial insight into the structure of labour markets in Colombia (Figure 3). The distance from each projection to the origin measures likeliness to the average city on that dimension. We explain the results for Tunja, Quibdó and Pereira.

Tunja, which is projected onto the first quadrant, shows a highly qualified workforce, along with Bogotá, Medellín and Bucaramanga; but it also displays high unemployment rates for the skilled workforce, as in Quibdó, Valledupar and Riohacha. Several arguments lead us to believe that supply characteristics and demand needs on this labour market do not match. On one hand, the local economy is biased towards agricultural activities, as suggested by the information on departmental GDP provided by DANE; while, on the other, the economic activity might not have evolved as rapidly as the educational attainment, acting as a barrier to the creation of proper job positions for educated people. Quibdó, for instance, is a city where economic opportunities are scarce. Poor economic performance and few job positions for skilled people are, among others, factors that determine the lack of willingness of its population to commit themselves to build up better human capital stocks. Results for Pereira and some other cities located in the coffee-growing region suggest that these cities are characterized by an aging population, high remittance dependence, and low skilled workforce. The coexistence of these factors represents difficulties for the accumulation of human capital, since the population pyramid is already old and the incentives to enrol in training programs are not sufficient for the working age population. As a result, these cities have experienced poor economic growth over the past few years, especially in those sectors that are intensive in low skilled labour force (*i.e.* construction, retail, among others).

2. Second Principal Plane (Third and Fourth Dimensions – Figure 3b)

Along the third dimension, striking differences arise between cities like Quibdó, Popayán or Florencia versus others such as Cali or Bogotá. The incentives for migration are seemingly higher in the former group of cities: In addition to violence and other political issues, low wages, poor education quality, higher shares of single youngsters and poor economic conditions for low skilled workers and young populations are, among others, the main reasons that encourage migration in these cities.

Projections along the fourth dimension distinguish cities with a high percentage of unionised workers, such as Popayán, Florencia or Neiva, from those with a relatively low share, such as Barranquilla or Cartagena. This finding suggests that negative valued cities in this axis face rigidities originated in the labour supply side of market power. Also, this axis classifies cities depending on the average size of households and other participation variables: most of those

located in the Caribbean region are characterized by larger families and very low female participation rates. Finally, there are some demographic characteristics that also contribute to the computation of this axis. Cities such as in Barranquilla, Quibdó and Cartagena in which the Afro-descendant proportion of population is higher are projected onto the positive side.

In sum, these results provide solid evidence of heterogeneity in the determinants of regional unemployment differentials, as suggested by Arango (2013). However, one of the most interesting findings in this paper is that cities with high unemployment levels do not necessarily share the same underlining structure on an economic, demographic, educational or even cultural basis. It is clear that there are great disparities between regions in terms of unemployment rates, but not all are due to the same reasons.

IV. Clustering

Clustering and multivariate data analysis techniques are complementary methods (Lebart, 1994, p.162), since studying the similarities between individuals in a lower dimensional space leads to a better understanding of the structure of the data. We group cities that share the same characteristics along the five principal axes we found in the MFACT step. Following Husson, Josse and Pages (2010), we combine MFACT results and both hierarchical (Ward's criterion) and partitional (*k-means* algorithm) clustering methods.

To interpret each partition, we measure the association between the cluster (understood as a categorical variable) and each (group of) quantitative variable and each frequency (contingency) table, and check its significance as in Bécue-Bertaut and Pagès (2008, pages 3261–3262). The resulting clusters suggest that differentials in unemployment rates are associated with different factors across Colombian cities.

A. First Cluster: Quibdó, Florencia, Riohacha and Valledupar

Cities belonging to this cluster are Quibdó, Florencia, Valledupar and Riohacha. Although their individual unemployment rates were not the highest among the sample and are not significantly different from those observed for the urban areas (12.4% on average for 2010), there are notorious differences

in other variables that determine quality of life and human capital formation. MFACT results suggest that this cluster groups cities that are, on a statistical basis, different from the others because of their outstandingly low educational attainment and their poor economic and social perspectives that influence on participation and human capital accumulation decisions.

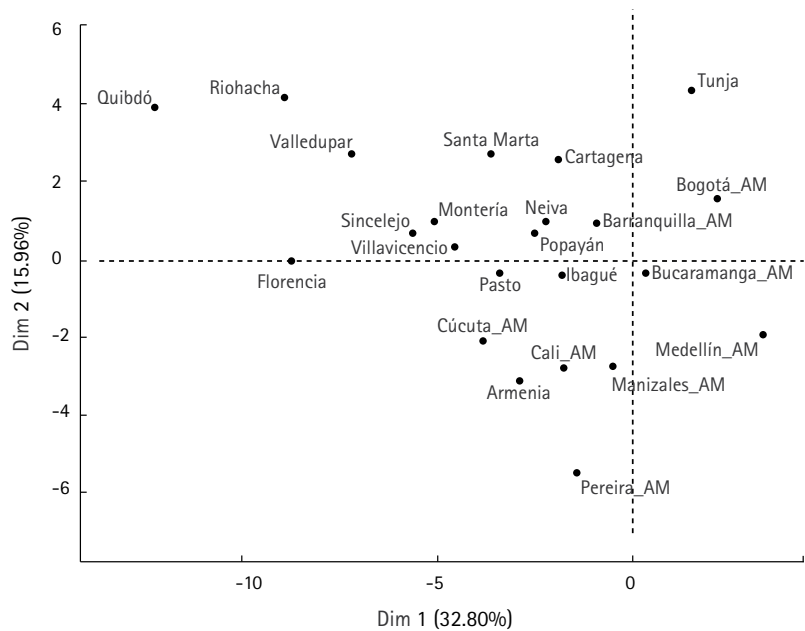
Cities in this cluster are statistically different on a demographic basis: over 50% of the population are young (0 to 25 years old), and the Afro-descendant and indigenous populations are the most representative ethnic groups. Younger people are more likely to be unemployed, mostly due to lack of expertise and education (Furnham, 1985). Also, there is empirical evidence of race discrimination in employment and wage setting (Darity and Mason, 1998), meaning higher unemployment rates for Afro-descendants and indigenous people.

Results show that these cities are net migrant recipients, as these are capital cities of departments where conditions are not favourable for the rural population, mostly due to security problems, lack of rural development and poor health and education coverage. Average Educational attainment is very low for these cities: illiteracy rates are the highest in the sample and the occupied workforce has the lowest levels of years of schooling, as projections over the first principal axis confirm. In addition, access to communication services and technology is scarce, as revealed by Internet service coverage and computer usage indicators.

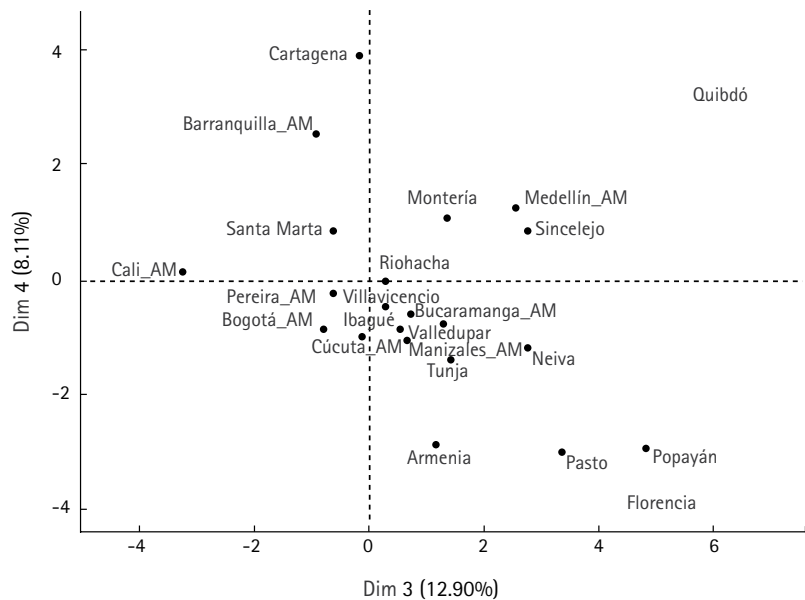
The economic activities that play an important role in local GDP are less productive and add less value as compared to other departments. Mining activities, for example, counted for about a third of their GDP (30%) on average over the past few years, almost ten times higher than the total national share over the 2000–2010 period (3.4%). Shares of industry, commerce and finance are significantly below the national average (3.2% vs. 14.0%; 8.7% vs. 12.4%; and 6.0% vs. 20.7%, respectively). Finally, average wages are about 20% lower than those paid in other cities of our sample.

In sum, this cluster is made up of cities where inhabitants have low educational and productive skills, while economic structure is biased towards activities that are not workforce intensive and that are not chained to other sectors that add more aggregated value to the economy. We recall that poverty and inequality have deeper roots in political and economic issues that are not entirely related to poor labour conditions.

Figure 3. Projections over the First and Second Principal Planes



a) Individuals Representation: First and Second Axis



b) Individuals Representation: Third and Fourth Axis

Source: Author's own calculations.

B. Second Cluster: Popayán, Pasto, Montería, Neiva, Villavicencio and Sincelejo

Cities belonging to this cluster are Popayán, Pasto, Montería, Neiva, Villavicencio and Sincelejo. Unemployment rates for these cities are somewhat heterogeneous, but they do still share underlining characteristics in their labour market structure. For example, the average ratio of non-salaried workers to total workforce in these cities is above the total national ratio (68.1% vs. 53.9%), as well the average share of self-employed working population (51.7% vs. 46.8%). This feature is a relevant characteristic of this cluster: results show that cities in this group exhibit high levels of self-employment and low workforce enrolment in formal firms.

In addition, there is a larger proportion of unionised workers (6.3% vs. 3.4% national), which can be considered as friction for the equilibrium reaching mechanism in these labour markets. It has been shown that unionised manufacturing firms tend to expand at a lower speed than the non-unionised ones (Hirsch, 1997; Long, 1993), which might contribute to overall lower economic performance and a lower labour demand expansion over time. We point out that these levels are low in comparison to other countries in Latin America and around the world (Blanchflower, 2006; Visser, 2006).

Another characteristic that they share is that the tertiary sector (*i.e.* retail, transport and services) has gained importance in these economies over the past few years: the average growth rate for the last decade is 6.6%, greater than the average for the national case (4.3%). This performance has been achieved in great part due to the dynamics of the financial sector.

Other results suggest that poor labour conditions might exist in these cities. Both average nominal and real incomes are below the national average, even if that difference is not statistically significant. In addition, although not included in the principal axes computations, it is also important to report that underemployment rates (both subjective and objective) are above average for urban areas: 33.9% and 15.5% vs. 30.1% and 12.9%, respectively. This leads us to think that cities belonging to this cluster are characterized by dysfunctional formal labour markets where prevailing working conditions encourage self-employment and informality, but are not necessarily reflected in the unemployment rate itself.

C. Third Cluster: Barranquilla, Santa Marta and Cartagena

This group is constituted by Barranquilla, Santa Marta and Cartagena and is the cluster with the lowest unemployment rates in the sample. It is also comparatively low in terms of occupation and participation rates (52.3%, 57.8% vs. 57.2%, 65.5%, respectively).

Global participation rates for under 25s and for women are outstandingly below the corresponding national averages. Among other reasons, this could be related to the average household size, which is the highest among the clusters in our sample (4.1 persons vs. 3.7 for the national average). According to the data, women continue to take over parenting responsibilities and domestic tasks at home, supporting the evidence of lower female participation rates.

Regarding education variables, this group displays the largest number of non-public institutions per 100,000 inhabitants (43 vs. 30 for the national average). According to Vilorio (2006), education coverage has increased over the past few years in the Caribbean region but results suggest that coverage efforts have not been accompanied by quality improvements. In fact, the average number of years of schooling for the inactive and unemployed populations is higher than the national averages, supporting the idea of mismatching between supply characteristics and demand needs in these labour markets.

To confirm the latter, data on educational attainment for the unemployed population is the highest among the clusters (11.1 years spent in education vs. 10.1 for the national average). In addition, the number of qualified unemployed people (with college or graduate education) is much higher than the national average, suggesting that skilled employment absorption in this cluster is insufficient and much lower than for other cities in our sample.

Summing up, unemployment and participation rates in this cluster are, on average, the lowest in our sample. Also, young people and women participate less in the labour market than in other cities. This is consistent with the high enrolment rates in educational and parenting and housekeeping activities. However, it is important to pay special attention to the quality of both higher education and job positions, since low unemployment rates may be due to the lack of dynamic and inclusive institutions in labour markets, in addition to the fact that overall participation rates are already low.

D. Fourth Cluster: Pereira, Armenia, Manizales, Ibagué, Cúcuta and Cali

The cities belonging to this cluster are Pereira, Armenia, Manizales, Ibagué, Cúcuta and Cali, and are mainly characterized by their demographic composition. Such characteristics tend towards an older population and the population's educational attainment, which is biased towards the population having few years of schooling (less than 10).

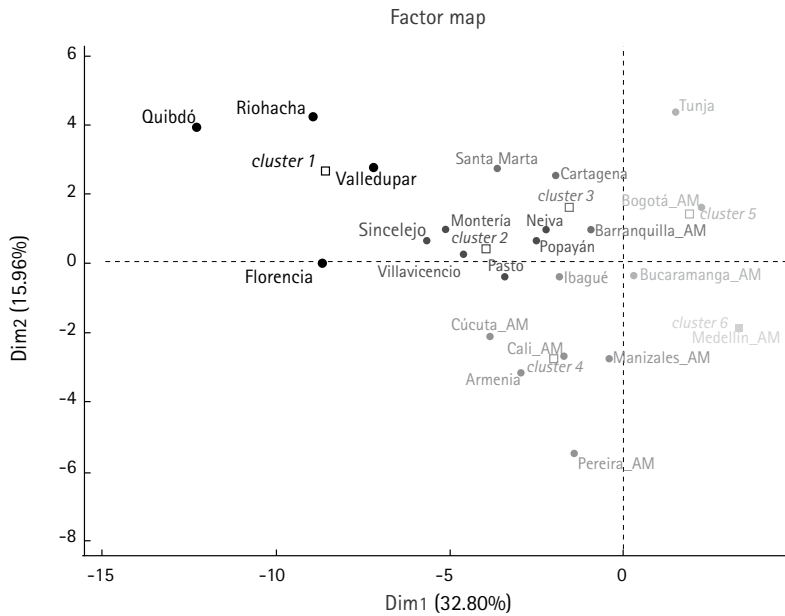
These cities exhibit lower gross birth rates than the average for urban areas (15 vs. 22), which suggests that the population pyramid tends to reverse faster in this group. This feature will lead to a greater proportion of dependent population and lower levels of education in the long run, since educational levels are already low in these cities and, given the progressive ageing of the population, incentives for human capital training are decreasing.

On the other hand, remittances per capita, which are three times the national average, and high hidden unemployment rates suggest a possible discouragement phenomenon that lowers people's incentives to participate in the labour market. Data supports this hypothesis: participation rates for adult males and for over 45s are below the national average. In fact, the unemployment rate for the latter population segment is the highest in the sample (11% compared to the national average of 8%).

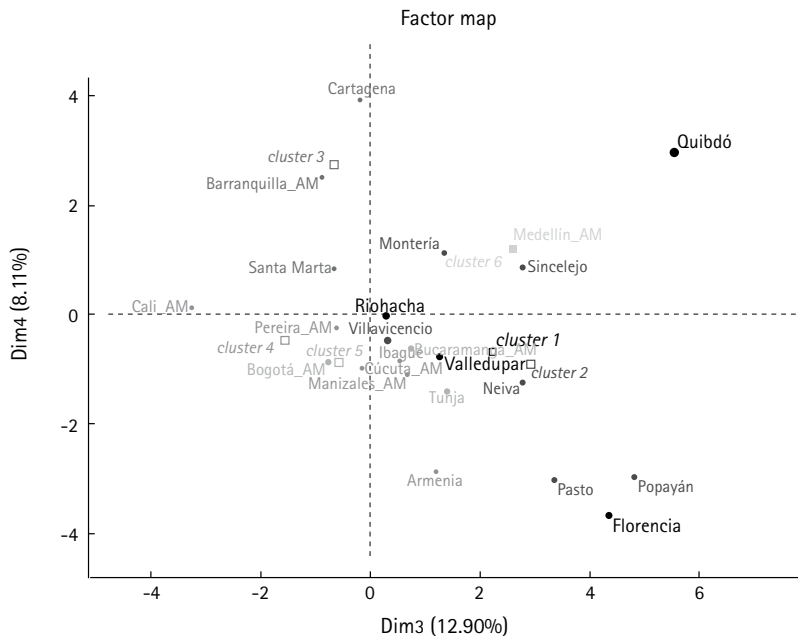
Another remarkable fact is that living costs (both in levels and annual variations) are, on average, lower than the rest of the economy for the 2008-2010 period, as deduced from both the food CPI inflation (1.1% vs. 1.9% national average) and total CPI inflation (2.3% vs. 2.6%). Cheap living costs lower incentives for people to improve their income levels and to participate in the labour market by increasing the average reservation wage (Arango, Montenegro and Obando, 2013).

Migration variables play an important role in this cluster as well. The average net migration rate is negative, revealing qualified workforce migration to other places with better economic and labour conditions. This "brain drain" leads to, for example, low economic development, low productivity and low wages, which cause second round effects on labour market performance (Eggert *et al.*, 2010). In fact, the average share of qualified working age population in these cities is the lowest among the clusters.

Figure 4. Cluster Projections Over the First and Second Principal Planes



a) City Clusters Representation: First Principal Plane



b) City Clusters Representation: Second Principal Plane

Source: Author's own calculations.

On the labour demand side, we find that this cluster has experienced the lowest economic growth in the 2000-2010 period (3.1% on average vs. 4.1% for the Colombian economy). Weak economic performance is generalized for all sectors, but it is most worrying in the secondary (industry and construction) and tertiary (commerce and services) sectors, which are labour intensive.

Interestingly, this cluster groups together cities that exhibit the highest unemployment rates in the sample (Pereira, 20.5%; Ibagué, 17.6%; Manizales, 17.6%; and Armenia 16.3%), along with Cúcuta (14.0%) and Cali (13.9%). Our hypothesis is that high unemployment rates in these cities arise due to the coexistence of high non-skilled labour supply levels, low incentives for participation, older population predominance, rigidities in terms of human capital accumulation and an economic structure that is not inclusive for non-qualified available workforce. Results show that the mismatch between supply characteristics and demand needs is definitely a major issue in labour markets in these cities, which can, in turn, determine long run structural unemployment (Yarce, 2000).

E. Fifth Cluster: Bogotá, Tunja and Bucaramanga

Bogotá, Bucaramanga and Tunja belong to this group, which is characterized by both higher educational levels and higher wages. The average school years of the working age population stands out as an important characteristic for this cluster. On average, 31% of the working age population is qualified, in contrast to the national average of 22%.

Cities in this group also display the highest average GDP per capita, and both real and nominal incomes, which are about 30% higher than the rest of the cities in the sample, only surpassed by Medellín and its metropolitan area (Cluster 6).

In this cluster, the percentage of workers employed in the financial intermediation sector is higher than the national average (2.1% vs. 1.3%), as well as those employed in real estate activities (9.1% vs. 6.3%). These shares reflect the degree of specialization of these economies in service provision activities. It also highlights the industry participation and its good performance during the 2000-2010 decade (5.1% on average vs. 3.5% national average).

The average global participation rate in this group of cities is also higher than the national average (67.4% vs. 65.5%), mainly because of increased female participation in comparison to the rest of cities (61.3% vs. 54%). We point out that female unemployment rates are the lowest among all the other clusters (13% on average), and that total unemployment rates are among the lowest in the country.

In sum, this cluster has very high levels of skilled workforce supply and a higher demand for this kind of labour than that observed for the rest of the cities. Results show that mismatching is low in these cities⁹, since labour supply responds to the demand for a skilled and productive workforce. This scenario has recently driven good labour market performance, which in turn allows the average unemployment rate for this group to be lower than that reported for urban areas (11.4% vs. 12.4%).

F. Sixth Cluster: Medellín

Medellín and its metropolitan area form a cluster by themselves, mainly characterized by market potential variables associated to population density, very high industrial density and the lowest average distance to major markets. Such factors would, in principle, yield lower levels of unemployment (because of the matching and higher labour demand, as explained in section 2). Also, the average household income (both nominal and real) for this cluster is above the national average by about 30%.

However, this city does not exhibit an unemployment rate below the national average (13.9% versus 12.4%). Despite the fact that the industrial sector absorbs a greater percentage of the working population than the national average (21.2% vs. 12.2%) and even though services oriented sectors count for almost 50% of the economic activity, mismatching exists in this city as well. It is noteworthy that although educational attainment levels are above the national average, demand for a qualified workforce seems to be just partially fulfilled: the skilled unemployed population share is just 28%, lower than 30% for our sample, and below 40% in cluster 5. Our hypothesis is that the demand is presumably requiring more skilled workforce than labour supply

9 Except for Tunja, where the unemployment rate is somewhat higher than the urban areas average (12.9% vs. 12.4%), despite of the very high levels of skilled workforce in this city.

can provide in Medellín. This may be due to the low incentives that many young people have to invest in human capital, given the violent environment in which they live, as argued by Medina, Posso and Tamayo (2011).

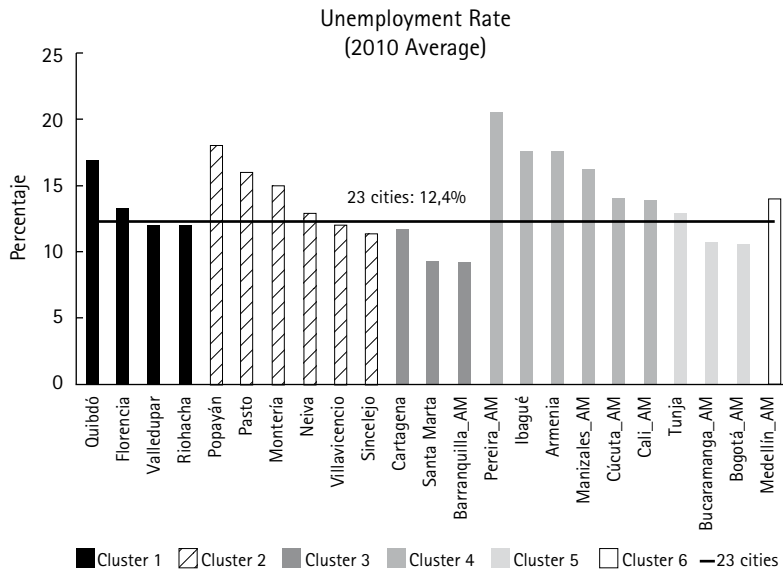
V. Conclusions

The heterogeneity found in urban labour market indicators in Colombia has not been widely studied. This paper aims to explore the relations between variables that have been theoretically and empirically assessed to determine the differentials in regional unemployment. Following Elhorst (2003), we studied a large dataset in order to establish similarities and differences between Colombian cities based on principal axes methods (MFACT, Bécue-Bertaut and Pagès 2004, 2008), clustering techniques and statistical criteria (Husson *et al.*, 2010). Our results suggest that there is evidence of disparities in structural variables that define the performance of regional labour markets. Particularly, our most relevant result is that cities that display high unemployment rates do not necessarily share the same characteristics; that is, frictions that give rise to unemployment are not the same across Colombian cities.

Clustering results give an important insight into the Colombian labour market structure. For example, we find that high unemployment rates in cluster 4 obey primarily to the mismatch between labour supply and demand resulting from the lack of educated workforce and the need for qualified workers, and also from low participation incentives due to high levels of per capita remittances; while unemployment problems in cluster 2 originate in the high levels of self-employment and the risks associated to this type of work. As suggested in the influential work of Overman, Puga and Hylke (2002), bearing in mind that not all cities or regions share the same structural problems and that they do not react to the same national-based labour institutions allows for policy makers to propose and execute better local policies focused on unemployment and inequality reduction. Therefore, this type of analysis matters and provides arguments for a better national and local government policy formulation. It is worth noting that, in many cases, clusters are made up of cities located near each other, suggesting that regional effects are also influenced by geographical positions, as mentioned in Overman *et al.* (2002) and Garcilazo and Spiezia (2007).

In sum, *Cluster 1* is made up of cities where poverty and a lack of strong institutional background prevail, while *Cluster 2* is characterized by high rates of labour informality, low average income and high underemployment rates. *Cluster 3* is statistically different with low participation rates, especially for the female working age population. This situation yields low unemployment rates, but average wages and income suggest low quality of job positions.

Figure 5. Average Unemployment Rates for 23 Colombian Cities (2010)



Source: DANE (GEIH), author's calculations.

Cluster 4 is perhaps the most interesting and complex group, since it is made up of cities that displayed very high unemployment rates in 2010. Statistical tests suggest that cities in this cluster are different from others insofar as their frictions in participation and in skilled workforce demand, but also their migration and opportunity vulnerabilities, are quite particular. For example, the coexistence of low educational attainment and a population pyramid biased towards an ageing population pose challenges to the successful implementation of conventional public policy programs.

Cluster 5 is made up of cities with high levels of educational attainment, wages, productive population and a prosperous economic structure. We believe that human capital accumulation is the main factor driving the good dynamics of labour markets and economic growth in this cluster. We claim that particular characteristics in these cities have fostered human capital accumulation over the past few decades, as in Díaz (2013). Finally, *Cluster 6* shares some labour market and economic characteristics with cluster 5. However, there are still some unresolved social and cultural issues in Medellín that influence labour market performance and yield a higher unemployment rate than the average for the metropolitan areas (Medina *et al.*, 2011).

Our results provide a useful insight into labour market structures in Colombia. However, there are still some minor differences in unemployment rates between cities belonging to the same cluster (as shown in Figure 9) that are not fully captured by differentials in variables used in this paper. We encourage future works to give a deeper insight into each one of these clusters in order to explore such inner heterogeneities. Our findings suggest that some cities share common structural characteristics that allow for variety in unemployment rates in Colombian urban areas. However, it is clear that unemployment rates will likely decline over time with the implementation of city-based actions designed to encourage participation, local incentives for low-skilled labour intensive sectors, and regional youth educational programs.

Acknowledgments

The authors are currently working as economists at Banco de la República. A previous version of this peer-reviewed paper was published in the working paper series *Borradores de Economía*, issue 802. The opinions, statements, findings and interpretations presented in this paper are responsibility of the authors and do not represent those of Banco de la República nor of its Board of Directors. Usual additional disclaimers apply. We thank Daniel Quintero Castro (dquintca@gmail.com), who participated actively in the early stages of this paper. Comments from Luis Eduardo Arango, Adolfo Cobo and two anonymous referees were very helpful, appreciated and acknowledged. Valuable assistance was received from Jackeline Piraján and Natalia Solano.

The research undertaken to write this paper did not have any kind of institutional funding.

References

1. ALBAGLI, E., GARCIA, P., AND RESTREPO, J. (2004). *Labor market rigidities and structural shocks: An open-economy approach for international comparisons* (Working Paper Series 263). Central Bank of Chile.
2. ARANGO, L. E. (2013). "Mercado de trabajo en Colombia: suma de partes heterogéneas", in L. E. Arango and F. Hamann (Eds.), *El mercado de trabajo en Colombia: hechos, tendencias e instituciones*. Bogotá: Banco de la República.
3. ARANGO, L. E., AND HAMANN, F., (Eds.) (2013). *El mercado de trabajo en Colombia: hechos, tendencias e instituciones*. Bogotá: Banco de la República.
4. ARANGO, L. E., MONTENEGRO, P., AND OBANDO, N. (2013). "El desempleo en Pereira: ¿solo cuestión de remesas?", in L. E. Arango and F. Hamann (Eds.), *El mercado de trabajo en Colombia: hechos, tendencias e instituciones*. Bogotá: Banco de la República.
5. BANDE, R., FERNÁNDEZ, M., AND MONTUENGA, V. (2008). "Regional unemployment in Spain: Disparities, business cycle and wage setting", *Labour Economics*, 15:885-914.
6. BARÓN, J. (2013). "Sensibilidad de la oferta de migrantes internos a las condiciones del mercado laboral en las principales ciudades de Colombia", in L. E. Arango and F. Hamann, (Eds.), *El mercado de trabajo en Colombia: hechos, tendencias e instituciones*. Bogotá: Banco de la República.
7. BÉCUE-BERTAUT, M., AND PAGÈS, J. (2004). "A principal axes method for comparing contingency tables: MFACT". *Computational Statistics & Data Analysis*, 45(3):481-503.
8. BÉCUE-BERTAUT, M., AND PAGÈS, J. (2008). "Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data". *Computational Statistics & Data Analysis*, 52(6):3255-3268.
9. BIFFL, G. (1998). "The Impact of Demographic Changes on Labor Supply". *Austrian Economic Quarterly*, 3(4):219-228.

10. BLANCHARD, O., AND KATZ, L. (1992). "Regional Evolutions", *Brookings Papers on Economic Activity*, 23(1):1-75.
11. BLANCHFLOWER, D. (2006). *A cross-country study of union membership* (Discussion Papers 2016). IZA.
12. BLUNDELL, R., AND MACURTY, T. (1999). "Labor Supply: A review of alternative approaches", in O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics* (vol. 3a, chap. 27, pp. 1559-1695). Amsterdam: Elsevier Science.
13. BRUECKNER, J., THISSE, J. F., AND ZENOU, Y. (2002). "Local labor markets, job matching and urban location", *International Economic Review*, 43(1):155-171.
14. BRUNELLO, G., LUPI, C., AND ORDINE, P. (2000). "Regional disparities and the Italian NAIRU", *Oxford Economic Papers*, 52:146-177.
15. BURDETT, K., AND MORTENSEN, D. (1998). "Wage differentials, employer size and unemployment". *International Economic Review*, 39(2):257-273.
16. CAHUC, R., AND ZYLBERBERG, A. (2004). *Labor economics*. Cambridge, MA: MIT Press.
17. CHAWLA, M., BETCHERMAN, G., AND BANERJI, A. (2007). *From red to gray: The "third transition" of aging populations in Eastern Europe and the former Soviet Union*. The World Bank.
18. DA ROCHA, J., AND FUSTER, L. (2006). "Why are fertility rates and female employment ratios positively correlated across OECD countries?", *International Economic Review*, 47(4):1187-1222.
19. DARITY, W., AND MASON, P. (1998). "Evidence on discrimination in employment: Codes of color, codes of gender", *The Journal of Economic Perspectives*, 12(2):63-90.
20. De FIGUEIREDO, F. A. (2010). "Dynamics of regional unemployment rates in Brazil: Fractional behavior, structural breaks and Markov switching", *Economic Modelling*, 27:900-908.

21. DÉTANG-DESSENDRE, C., AND GAIGNÉ, C. (2009). "Unemployment duration, city size and the tightness of the labor market". *Regional Science and Urban Economics*, 39:266-276.
22. DÍAZ, A. M. (2011). *Spatial unemployment differentials in Colombia* (Discussion Papers 14). Institut de Recherches Économiques et Sociales de l'Université catholique de Louvain.
23. DÍAZ, A. M. (2013). "The employment advantages of skilled urban municipalities in Colombia", *Ensayos sobre Política Económica*, 31(70):315-366.
24. EGGERT, W., KRIEGER, T., AND MEIER, V. (2010). "Education, unemployment and migration", *Journal of Public Economics*, (94):354-362.
25. ELHORST, P. (2003). "The mystery of regional unemployment differentials: Theoretical and empirical explanations", *Journal of economic surveys*, 17(5):709-748.
26. EPIFANI, P., AND GANCIA, G. A. (2005). "Trade, migration and regional unemployment", *Regional Science and Urban Economics*, 35(6):625-644.
27. ESCOFIER, B., AND PAGÈS, J. (1994). "Multiple factor analysis (afmult package)", *Computational Statistics and Data Analysis*, 18(1):121-140.
28. ESCOFIER, B., AND PAGÈS, J. (2008). *Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation*. Paris: Dunod.
29. FARBER, H. (1986). "The analysis of union behavior", in O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics* (vol. 2, chap. 18, pp. 1039-1089). Amsterdam: Elsevier Science.
30. FLEISHER, M., AND RHODES, G. (1976). "Unemployment and the labor force participation of married men and women: A simultaneous model", *The Review of Economics and Statistics*, 58(4):398-406.
31. FREEMAN, R. (2009). *Labor regulations, unions and social protection in developing countries: Market distortions or efficient institutions?* (Working Paper Series 14789). NBER.

32. FURNHAM, A. (1985). "Youth unemployment: A literature review". *Journal of Adolescence*, 8(2):109-124.
33. GALVIS, J. (2002). *Integración regional de los mercados laborales en Colombia: 1984-2000* (Documentos de Trabajo sobre Economía Regional 27). Banco de la República.
34. GAMARRA, J. (2005). *¿Se comportan igual las tasas de desempleo de las siete principales ciudades colombianas?* (Documentos de Trabajo sobre Economía Regional 55). Banco de la República.
35. GARCILAZO, J., AND SPIEZIA, V. (2007). "Regional unemployment clusters: Neighborhood and state effects in Europe and North America", *The Review of Regional Studies*, 37(3):282-302.
36. GOMES, F. A., AND DA SILVA, C. G. (2009). "Hysteresis versus NAIRU and convergence versus divergence: The behavior of regional unemployment rates in Brazil", *The Quarterly Review of Economics and Finance*, 49:308-322.
37. GREENACRE, M. (2007). *Correspondence analysis in practice*. Boca Raton, FL: Chapman & Hall.
38. HIRSCH, B. (1997). "Unionization and economic performance: Evidence on productivity, profits, investment and growth", in E. Milhar (Ed.), *Unions and Right-to-Work Laws* (pp. 35-70). Vancouver: The Frazer Institute.
39. HUSSON, F., AND JOSSE, J. (2013). "Handling missing values in multiple factor analysis", *Food Quality and Preference*, 30:77-85.
40. HUSSON, E., JOSSE, J., AND LÊ, S. (2008). "FactoMineR: An R package for multivariate analysis", *Journal of Statistical Software*, 25(1):1-18.
41. HUSSON, E., JOSSE, J., AND PAGÈS, J. (2010). *Principal component methods, hierarchical clustering, partitional clustering: Why would we need to choose for visualizing data?* (Technical Reports – Laboratoire de Mathématiques Appliquées, AgroCampus- OUEST, Rennes, France).

42. IZRAELI, O., AND MURPHY, K. (2003). "The effect of industrial diversity on state unemployment rate and per capita income", *The annals of Regional Science*, 37:1-14.
43. JARAMILLO, C., ROMERO, C., AND NUPIA, O. (2000). *Integración en el mercado laboral colombiano: 1945-1998* (Borradores de Economía 148). Banco de la República.
44. JOHNSON, R., AND WICHERN, D. (2007). *Applied multivariate statistical analysis*. Upper Saddle River, NJ: Pearson-Prentice Hall.
45. KRUGMAN, P. (1995). *Development, geography and economic theory*. Cambridge, MA: The MIT Press.
46. KRUSELL, P., AND RUDANKO, L. (2013). *Unions in a frictional labor market* (Working Paper Series 18128). NBER.
47. LANZAFAME, M. (2010). "The nature of regional unemployment in Italy", *Empirical Economics*, 39:877-895.
48. LEBART, L. (1994). "Complementary use of correspondence analysis and cluster analysis", in M. Greenacre and J. Blasius (Eds.), *Correspondence analysis in the social sciences* (pp. 162-178). San Diego, CA: Academic Press.
49. LERMAN, R., AND SCHMIDT, S. (1999). *An overview of economic, social and demographic trends affecting the U.S labor market: Final report* (Technical report). The Urban Institute-United States Department of Labor.
50. LEWIS, H. G. (1986). "Union relative wage effects", in O. Ashenfelter and D. Card, (Eds.), *Handbook of Labor Economics* (vol. 2, chap. 20, pp. 1139-1181). Amsterdam: Elsevier Science.
51. LONG, R. (1993). "The effect of unionization on employment growth of Canadian companies", *Industrial and Labor Relations Review*, 46(4):691-703.
52. LOTTMAN, F. (2012). *Explaining regional unemployment differences in Germany: A spatial panel data analysis* (Discussion Papers 26). SFB 649 Humboldt-Universität zu Berlin.

53. MALIZIA, E., AND KE, S. (1993). "The influence of economic diversity on unemployment and stability", *Journal of Regional Science*, 33(2):221-235.
54. MARSTON, S. (1985). "Two views of the geographic distribution of unemployment", *The Quarterly Journal of Economics*, 100(1):57-79.
55. MARTÍNEZ, C. (2013). *Descenso de la fecundidad, participación laboral de la mujer y reducción de la pobreza en Colombia, 1990-2010*. Estudio a profundidad basado en las Encuestas Nacionales de Demografía y Salud (ENDS) 1990-2010. Bogotá: Profamilia.
56. MEDINA, C., POSSO, C., AND TAMAYO, J. (2011). *Costos de la violencia urbana y políticas públicas: algunas lecciones de Medellín* (Borradores de Economía 614). Banco de la República.
57. MINCER, J. (1981). *Union effects: Wages, turnover and job training* (Working Paper Series 808). NBER.
58. MINCER, J. (1991). *Education and unemployment* (Working Paper Series 3838). NBER.
59. MORETTI, E. (2010). "Local multipliers", *American Economic Review, Papers and Proceedings*, 2(100):1-7.
60. MORETTI, E. (2012). *The new geography of jobs*. New York, NY: Houghton Mifflin Harcourt Publishing Company.
61. OBERST, C., AND OELGEMOLLER, J. (2013). *Economic growth and regional labor market development in German regions: Okun's law in a spatial context* (Working Paper Series 5). FCN.
62. OKUN, A. (1962). *Potential GDP: Its measure and significance* (Working Papers 190). Cowles Foundation, Yale University.
63. OVERMAN, H., PUGA, D., AND HYLKE, V. (2002). "Unemployment clusters across Europe's regions and countries", *Economic Policy*, 17(34):115-147.

64. PAGÈS, J. (2002). "Analyse factorielle multiple appliqué aux variables qualitatives et aux données mixtes", *Revue de Statistique Appliquée*, 50(4):5-37.
65. PAGÈS, J. (2004). "Multiple factor analysis: Main features and application to sensory data", *Revista Colombiana de Estadística*, 27(1):1-26.
66. PEÑA, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw Hill.
67. PISSARIDES, C., AND WADSWORTH, J. (1989). "Unemployment and the Inter-regional mobility of labour", *The Economic Journal*, 99(397):739-755.
68. R CORE TEAM (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
69. URRUTIA, M., (Ed.) (2001). *Empleo y economía*. Bogotá: Banco de la República.
70. VILORIA, J. (2006). *Educación superior en el Caribe colombiano: análisis de cobertura y calidad* (Documentos de Trabajo sobre Economía Regional 69). Banco de la República.
71. VISSER, J. (2006). "Union membership statistics in 24 countries", *Monthly Labor Review*, Issue 3. pp. 38-49. Bureau of Labor Statistics.
72. WALDEN, M. (2012). "Explaining differences in state unemployment rates during the great recession", *The Journal of Regional Analysis and Policy*, 42(3):251-257.
73. WINTERS, J. (2013). "Human capital externalities and employment differences across metropolitan areas of the USA", *Journal of Economic Geography*, 13:799-822.
74. YARCE, W. (2000). *El desempleo estructural y la tasa natural de desempleo: algunas consideraciones teóricas y su estado actual en Colombia* (Lecturas de Economía 52). Universidad de Antioquia.

Appendix 1

Table A1.1. Dataset Description

Group	Variable	Type of var.	Description	Source
Demographic Structure	TCP	q	Population Growth Rate (Average 1986–2010)	2005 census (DANE)
	POBT_0–25	f	Pop. under the age of 25 (y.o) (2010 estimate)	
	POBT_25–45	f	Pop. between 26 and 45 y.o (2010 est.)	
	POBT_45–65	f	Pop. between 46 and 65 y.o (2010 est.)	
	POBT_65+	f	Pop. older than 65 y.o (2010 est.)	
	POBM_0–25	f	Male under the age of 25 y.o (2010 est.)	
	POBM_25–45	f	Male pop. between 26 and 45 y.o (2010 est.)	
	POBM_45–65	f	Male pop. between 46 and 65 y.o (2010 est.)	
	POBM_65+	f	Male pop. older than 65 y.o (2010 est.)	
	POBF_0–25	f	Female pop. under the age of 25 y.o (2010 est.)	
	POBF_25–45	f	Female pop. between 26 and 45 y.o (2010 est.)	
	POBF_45–65	f	Female pop. between 46 and 65 y.o (2010 est.)	
	POBF_65+	f	Female pop. older than 65 y.o (2010 est.)	
	PPIN	q	% Indigenous population (2010 est.)	
	PPAF	q	% Black population (2010 est.)	
	TMI	q	Infant mortality rate (2010)	
	TBN	q	Gross birth rate (2010)	
Participation	TGP	q	Labour participation rate (2010)	GEIH (DANE)
	TGPM	q	Male labour participation rate (2010)	
	TGPF	q	Female labour participation rate (2010)	
	TDO	q	Hidden unemployment rate (2010)	
	TGPHJ	q	Young males labour participation rate (< 25 y.o) (2010)	
	TGPH	q	Male labour participation rate (< 25 y.o) (2010)	
	TGPMJ	q	Young females labour participation rate (> 25 y.o) (2010)	
	TGPM	q	Female labour participation rate (> 25 y.o) (2010)	
	EC_H_UL	f	Male – Marital Status: Free union (> 10 y.o)	
	EC_H_C	f	Male – Marital Status: Married (> 10 y.o)	
	EC_H_SE	f	Male – Marital Status: Divorced (> 10 y.o)	

Table A1.1. Dataset Description (continued)

Group	Variable	Type of var.	Description	Source
Participation	EC_H_V	f	Male – Marital Status: Widowed (> 10 y.o)	GEIH (DANE)
	EC_H_SOL	f	Male – Marital Status: Single (> 10 y.o)	
	EC_M_UL	f	Female – Marital Status: Free union (> 10 y.o)	
	EC_M_C	f	Female – Marital Status: Married (> 10 y.o)	
	EC_M_SE	f	Female – Marital Status: Divorced (> 10 y.o)	
	EC_M_V	f	Female – Marital Status: Widowed (> 10 y.o)	
	EC_M_SOL	f	Female – Marital Status: Single (> 10 y.o)	
	TH	q	Average size of household	Profamilia ENDS
	EPN	q	Women's median age on the first birth	
Migration	REM	q	Remittances per capita	BanRep
	TMN	q	Net migration rate (2010)	DANE
	PREC	q	Incoming (displaced) people (per 100K hab.)	DPS
	NPDEP	q	Displacement Register (Arrivals–people) (per 100K hab.)	
	NHDEP	q	Displacement Register (Arrivals–households) (per 100K hab.)	
Commuting	ICVSP	q	Public transport use Index (Author's Calculations) (2010)	ETUP (DANE)
	PSSP	q	# of pass. using Buses (per 100K hab.) 2010	
	PVSP	q	# Public use veh.in service (2010) (per 100K hab.)	
	IPCVIV	q	Housing CPI (2008–2010 av.)	DANE
	IPCTR	q	Transport services CPI (2008–2010 av.)	
	INFVIV	q	Housing CPI change (2008–2010 av.)	
	INFTR	q	Transport services CPI change (2008–2010 av.)	
Wages	INP	q	Av. Nominal Income	GEIH (DANE)
	INM	q	Median Nominal Income	
	IRP	q	Av. Real Income	
	IRM	q	Median Real Income	
	SNPA	q	Av. nominal wage (waged–salary workers)	
	SNMA	q	Median nominal wage (waged–salary workers)	
	SRPA	q	Av. real wage (waged–salary workers)	
	SRMA	q	Median real wage (waged–salary workers)	

Table A1.1. Dataset Description (continued)

Group	Variable	Type of var.	Description	Source
Wages	IPCALI	q	Food CPI (2008–2010 av.) – Living cost	DANE
	INFALI	q	Food CPI change (2008–2010 av.) – Living cost	
	IPCTot	q	Total CPI (2008–2010 av.) – Living cost	
	InflalPCTot	q	Total CPI change (2008–2010 av.) – Living cost	
Regional Growth	PIBT	q	Departmental total GDP Growth (2000–2011 av.)	DANE
	PIB1	q	Departmental Primary GDP Growth (2000–2011 av.)	
	PIB2	q	Departmental Secondary GDP Growth (2000–2011 av.)	
	PIB3	q	Departmental Tertiary GDP Growth (2000–2011 av.)	
	PIB_agr	q	Departmental Agr. GDP Growth (2000–2011 av.)	
	PIB_min	q	Departmental Mining GDP Growth (2000–2011 av.)	
	PIB_ind	q	Departmental Manufact. GDP Growth (2000–2011 av.)	
	PIB_ele	q	Departmental Energy GDP Growth (2000–2011 av.)	
	PIB_con	q	Departmental Construc. GDP Growth (2000–2011 av.)	
	PIB_com	q	Departmental Retail GDP Growth (2000–2011 av.)	
	PIB_tra	q	Departmental Transport GDP Growth (2000–2011 av.)	
	PIB_fin	q	Departmental Financial GDP Growth (2000–2011 av.)	
	PIB_ssv	q	Departmental Services GDP Growth (2000–2011 av.)	
	PIB_tax	q	Departmental Taxes GDP Growth (2000–2011 av.)	
Market Potential	DDM	q	Demographic density (2005)	DANE
	ESML	q	Deviation of unemployment rate from its long-term trend (tightness in labour market, Author's own calculations)	
	SEMD	q	Av. unemployment duration	
	AC	q	Trade Openness (Author's own calculations)	Ocaribe –SID
	DIST_MERC	q	Weighted distance to major markets (Bogotá, Medellín, Cali and Barranquilla)	
	IHHM	q	Herfindahl – Hirschman Market Index	

Table A1.1. Dataset Description (continued)

Group	Variable	Type of var.	Description	Source
Market Potential	EPE_2009	q	Efficiency of Business Processes	Ocaribe -SID
	IDI_2009	q	Industrial Density Index	
	DIST_MERC	q	Weighted distance to major markets (Bogotá, Medellín, Cali and Barranquilla)	
	EPE_2009	q	Efficiency of Business Processes	
	IDI_2009	q	Industrial Density Index	
	CostConst	q	Av. cost of building a warehouse	
	Aemp	q	Starting a business (costs and requirements to register a new company)	
	TIT	q	Total Taxes rate 2010	
Economic Structure	IHHP	q	Herfindahl – Hirschman Product Index	Ocaribe -SID
	PART_agr	q	Agricultural GDP Share (2000–2011 av.)	Authors' own calculations from Departmental GDP (DANE)
	PART_min	q	Mining GDP Share (2000–2011 av.)	
	PART_ind	q	Manufacturing GDP Share (2000–2011 av.)	
	PART_ele	q	Electricity GDP Share (2000–2011 av.)	
	PART_con	q	Construction GDP Share (2000–2011 av.)	
	PART_com	q	Retail GDP Share (2000–2011 av.)	
	PART_tra	q	Transport GDP Share (2000–2011 av.)	
	PART_fin	q	Financial GDP Share (2000–2011 av.)	
	PART_ssv	q	Services GDP Share (2000–2011 av.)	
	PART_tax	q	Taxes GDP Share (2000–2011 av.)	
	VAR_PART_agr	q	Change on agricultural GDP share (2000–2011 av.)	
	VAR_PART_min	q	Change on mining GDP share (2000–2011 av.)	
	VAR_PART_ind	q	Change on industry GDP share (2000–2011 av.)	
	VAR_PART_ele	q	Change on electricity GDP share (2000–2011 av.)	
	VAR_PART_con	q	Change on construction GDP share (2000–2011 av.)	
	VAR_PART_com	q	Change on trade GDP share (2000–2011 av.)	
	VAR_PART_tra	q	Change on Transport GDP share (2000–2011 av.)	

Table A1.1. Dataset Description (continued)

Group	Variable	Type of var.	Description	Source
Economic Structure	VAR_PART_fin	q	Change on financial GDP share (2000–2011 av.)	Authors' own calculations from Departmental GDP (DANE)
	VAR_PART_ssv	q	Change on services GDP share (2000–2011 av.)	
	VAR_PART_tax	q	Change on taxes GDP share (2000–2011 av.)	
	AS	f	# of wage–salary workers (2010)	GEIH (DANE)
	NAS	f	# of unpaid wage workers (2010)	
	EMPA	f	# of private employees	
	EMGOB	f	# of government employees	
	EMDOM	f	# of domestic employees	
	EMCP	f	# of self employed	
	EMPT	f	# of bosses	
	EMFSR	f	# of unpaid family workers	
	EMSR	f	# of unpaid workers	
	EMJ	f	# of workmen	
	EMO	f	# of other workers	
	NPAI	f	Ppl. on Real estate activities	
	NPAG	f	Ppl. on Agr.	
	NPCM	f	Ppl. on Retail	
	NPCN	f	Ppl. on Construction	
	NPEG	f	Ppl. on Gas and Water Supply	
	NPFN	f	Ppl. on financial intermediation	
	NPIN	f	Ppl. on manufacturing	
	NPMN	f	Ppl. on Mining activities	
	NPSS	f	Ppl. on Services	
	NPTN	f	Ppl. on Transport sector	
Economic & Social Barriers	ICV	q	Life Quality Index (2005)	DNP
	DHV	q	% of households w/o housing (2005)	2005 census (DANE)
	DHVC	q	% of households w/o housing (2005) – quantitative	
	DHVQ	q	% of households w/o housing (2005) – qualitative	
	NBI	q	Unsatisfied basic needs (2010)	

Table A1.1. Dataset Description (continued)

Group	Variable	Type of var.	Description	Source
Economic & Social Barriers	PIBPC	q	GDP per capita (2010)	Authors' computations from Dept. GDP series
	PIBPC_SMIN	q	GDP per capita w/o mining (2010)	
	C_PIBPC	q	GDP per capita Growth (2001–2010 av.)	
	C_PIBPC_SMIN	q	GDP per capita Growth w/o mining (2001–2010 av.)	
	NH_V	q	Number of households per housing	GEIH (DANE)
	COB_ELEC	q	% of housing w/o electricity	
	COB_ACU	q	% of housing w/o Aqueduct	
	COB_SAN	q	% of housing w/o sewage system	
	COB_BAS	q	% of housing w/o garbage disposal	
	COB_GAS	q	% of housing w/o natural Gas	
	EJ_FBKF	q	Gross fixed capital formation per capita Execution (2010)	Authors' Computations based on information from MHCP.
	REGA	q	Income from royalties per capita (2010)	
	TRANS	q	National transfers income per capita (2010)	
	NATH	q	# of requests for Humanitarian Ass. (2010) per 100K habs.	DPS
	NNICBF	q	# of children aided by ICBF (2010) – per 100K habs.	
	HOM	q	Murders per 100.000 inhabitants (2010)	Ocaribe –SID
Education	COBP	q	Primary education Coverage (2010)	MEN
	COBS	q	Secondary education Coverage (2010)	
	COBM	q	High school Coverage (2010)	
	IENO	q	Non–public Institutions per 100K habs. (2010)	
	IEO	q	Public Institutions per 100K habs. (2010)	
	POBA	q	% Literate population (2010)	
	POBAN	q	% Illiterate population (2010)	GEIH (DANE)
	HOG_SININT	q	% of household w/o Internet (2010)	
	HOG_SINPC	q	% of household w/o PC (2010)	
	APED_PET	q	Av. # of school years – Working age pop. (2010)	
	APED_OC	q	Av. # of school years – Employed (2010)	
	APED_DE	q	Av. # of school years – Unemployed (2010)	

Table A1.1. Dataset Description (continued)

Group	Variable	Type of var.	Description	Source
Education	APED_DE_0-5	f	% of Unemployed between 0 and 5 years of education (y.o.e) (2010)	GEIH (DANE)
	APED_DE_6-11	f	% of Unemployed between 5 and 11 y.o.e (2010)	
	APED_DE_12-13	f	% of Unemployed between 12 and 13 y.o.e (2010)	
	APED_DE_14-15	f	% of Unemployed between 14 and 15 y.o.e (2010)	
	APED_DE_15+	f	% of Unemployed more than 15 y.o.e (2010)	
	APED_OC_0-5	f	% of Employed between 0 and 5 y.o.e (2010)	
	APED_OC_6-11	f	% of Employed between 6 and 11 y.o.e (2010)	
	APED_OC_12-13	f	% of Employed between 12 and 13 y.o.e (2010)	
	APED_OC_14-15	f	% of Employed between 14 and 15 y.o.e (2010)	
	APED_OC_15+	f	% of Employed more than 15 y.o.e (2010)	
	APED_PET_0-5	f	% of Working age population between 0 and 5 y.o.e (2010)	
	APED_PET_6-11	f	% of Working age population between 6 and 11 y.o.e (2010)	
	APED_PET_12-13	f	% of Working age population between 12 and 13 y.o.e (2010)	
	APED_PET_14-15	f	% of Working age population between 14 and 15 y.o.e (2010)	
	APED_PET_15+	f	% of Working age population more than 15 y.o.e (2010)	
	APED_INA_0-5	f	% of Inactive population between 0 and 5 y.o.e (2010)	
	APED_INA_6-11	f	% of Inactive population between 6 and 11 y.o.e (2010)	
	APED_INA_12-13	f	% of Inactive population between 12 and 13 y.o.e (2010)	
	APED_INA_14-15	f	% of Inactive population between 14 and 15 y.o.e (2010)	
	APED_INA_15+	f	% of Inactive population more than 15 y.o.e (2010)	
Unionisation	PSIN	q	% of Unionised labour force	GEIH (DANE)