

ISSN 1390-6208  
e-ISSN 1390-7867  
Número 1 - 2015  
Enero a Junio

# Analítica

Volumen

# 9

Revista de Análisis Estadístico  
Journal of Statistical Analysis

**Analítika\*Volumen 9\*Número 1\*enero-junio 2015\*Año 5\*Quito-Ecuador.** revista semestral científica del Instituto Nacional de Estadística y Censos, Ecuador. Publicación internacional bilingüe relacionada al análisis estadístico y socio-económico con desarrollo de metodologías cuantitativas. Las ideas y opiniones expresadas en las colaboraciones, son de exclusividad de las autoras y autores.

## Editor - Director Científico

José Rosero Ph.D.  
Universidad de Amsterdam - Doctor en Economía

## Comité Editorial Internacional

Francisco Venegas Martínez Ph.D.  
Instituto Politécnico Nacional -  
México



Silviu Glavan Ph.D.  
IE University - España



Adriana Uquillas Ph.D.  
Banco Itau - Unibanco - Brazil



Carlos Machicado Ph.D.  
Instituto de Estudios Avanzados en  
Desarrollo - Bolivia



Ricardo Castro Santis Ph.D.  
Universidad del Bío-Bío - Chile



David Tomás Jacho-Chávez  
Ph.D. Emory University



Paúl Ernesto Carrillo Ph.D. George  
Washington University



## Comité Editorial Nacional - Ecuador



Rolando Sáenz Ph.D.  
Universidad Central del Ecuador



Borys Álvarez Ph.D.  
Universidad Central del Ecuador



Petronio Álvarez Ph.D.  
Universidad Central del Ecuador



Wilson Pérez Ph.D.  
Facultad Latinoamericana de Ciencias  
Sociales



Juan Carlos García M.Sc.  
Universidad de las Américas



María Luisa Granda Kuffó Ph.D.  
Escuela Superior Politécnica del  
Litoral - ESPOL



Pablo Samaniego M.Sc.  
Secretaría Nacional de Educación  
Superior, Ciencia, Tecnología  
e Innovación





## PRESENTACIÓN

**Estimados lectores,**

El Instituto Nacional de Estadística y Censos (INEC) en su afán de contribuir con la sociedad ecuatoriana tiene como objetivo fundamental convertirse en un referente nacional e internacional en la generación de estudios especializados que contribuyan a la toma, informada, de decisiones públicas y privadas y a la planificación nacional.

Desde el 2011, la Revista Analítika forma parte de esta visión institucional y se ha convertido en un espacio de referencia para la difusión de la investigación científica.

Para este año la revista se encuentra en una etapa de actualización y reorganización de su gestión con el fin de seguir siendo un producto científico que de información útil para investigadores, académicos y estudiantes, además de posicionar nuestro rol analítico dando el valor agregado a los datos y contexto para su estudio. Hemos actualizado el Comité Editorial de nuestra revista científica y nombramos un nuevo editor por ello estimamos que es importante agradecer a todos los que en este tiempo han formado parte de esta publicación; miembros del Comité, editor, articulistas y responsables institucionales, pues gracias a su trabajo esta idea del INEC se cristalizó y se fortalece.

Con este antecedente es grato para el INEC dar la bienvenida los nuevos miembros del Comité Editorial Internacional: David Tomás Jacho-Chávez PhD en Economía de la London School of Economics & Political Science, Paúl E. Carrillo PhD en Economía de la Universidad de Virginia; así como a quienes se suman al Comité Editorial Nacional: María Luisa Granda Kuffó PhD en Economía de la Universidad de Texas A&M, y Pablo Samaniego Master en Economía con especialización en Desarrollo y Política Económica de FLACSO, Ecuador. En virtud de su conocida trayectoria profesional y académica estamos seguros de que sus aportes serán valiosos para el fortalecimiento de nuestra revista científica.

En el noveno volumen de Analítika se incluyen análisis que son aporte relevante para la política pública nacional:

- Impact of Fertility on Female Labor Supply. A partir del Censo Nacional de Población y Vivienda ejecutado en el año 2010, utilizando técnicas de evaluación de impacto, se identifica el efecto de la fertilidad en la participación laboral femenina en el Ecuador.
- Efectos Macroeconómicos de la Política Fiscal en Ecuador 1993-2009. Este estudio emplea un modelo de vectores autorregresivos estructurales para identificar los efectos de la política fiscal en la economía ecuatoriana.
- Localización de Centros de empleo y su influencia sobre la distribución de la población en el Distrito Metropolitano de Quito. Este artículo se centra en identificar centros de empleo y de población en el Distrito Metropolitano de Quito, a través de distintas técnicas estadísticas y econométricas.
- Hacia un Algoritmo Óptimo de Emparejamiento de Nombres. Estudio que plantea un algoritmo óptimo de emparejamiento de nombres para la integración de registros administrativos útiles para la producción estadística en el Ecuador.

Les invitamos a seguir siendo parte de éste y de nuevos retos, a seguir construyendo juntos la historia de esta institución y del país. Así, el Ecuador podrá ver fortalecido a un Instituto que maneja sus procesos bajo los más estrictos estándares de calidad que avalan un trabajo serio, riguroso y responsable, con el libre acceso a la información antes limitada, con la didáctica como norte en la liberación de bases de datos, metodologías, sintaxis y formularios, con la entrega de cifras oportunas, útiles y de calidad, con nuestros usuarios en general como principales aliados y críticos.

Jorge García Guerrero  
Representante Institucional de Analítika



## CONTENIDO

· Impact of Fertility on Female Labor Supply .....	5
· Efectos Macroeconómicos de la Política Fiscal en Ecuador 1993-2009 .....	21
· Localización de Centros de empleo y su influencia sobre la distribución de la población en el Distrito Metropolitano de Quito .....	53
· Hacia un Algoritmo Óptimo de Emparejamiento de Nombres .....	95



# Analítica

Impact of Fertility on Female Labor Supply

Andrea Molina Vera



[www.ecuadorencifras.gob.ec](http://www.ecuadorencifras.gob.ec)





# Impact of Fertility on Female Labor Supply

Andrea Molina Vera<sup>1</sup>

## Abstract

This paper studies the effect of fertility on Ecuadorian female labor participation. It uses, as source of exogenous variability in family size, the instrument introduced by Angrist and Evans in 1998 for the United States: parental preferences for a mixed sibling-sex composition on their children. The empirical application shows that women with two boys or two girls are 4 percentage points more likely to have a third child than women with one boy and one girl (for both samples: all and married women). 2SLS estimations show that a third child causes a negative impact of 9 percentage points on female labor supply in all work definitions for married women, meanwhile for all women, there is a negative impact of 8 percentage points in “work outside home” and “paid work outside home” as measures of labor supply. On the other hand, OLS estimates are significant and negative for all definitions of work and for both samples.

JEL Classification: J13, J22

Keywords: Causality, Fertility, Female Labor Supply, Developing Countries

## 1 Introducción

For the U.S., Angrist and Evans found that fertility has a negative impact of about 12 percentage points on female labor supply (Angrist and Evans (1998)). Likewise, in Ecuador there seems to be a negative relationship between fertility and female labor supply, arising the question: can the U.S. results be extrapolated to Latin American developing countries as Ecuador?

The differences of labor markets and family contexts between U.S. and Ecuador motivates this question. Also, it is important to take into account some differences that developing countries have in comparison to the developed ones: high levels of underemployment and informal economy, high heterogeneity in work arrangements (Delpiano (2012)), higher fertility, lower levels of female education and fewer facilities for formal childcare (Cruces and Galiani (2007)).

Also, in empirical works, the problem of endogeneity between fertility and labor market outcomes is widely known. Angrist and Evans (1998) mention that this problem arises

---

<sup>1</sup>FLACSO-Ecuador. I gratefully acknowledge valuable comments from Hessel Oosterbeek.

since there are strong theoretical reasons to believe that fertility and labor supply are jointly determined. Agüero and Marks (2008) acknowledge the possibility of endogeneity for omitted variable bias which influences fertility and labor force participation simultaneously (e.g. ambition or talent). Thus, several studies have exploited exogenous changes in family size to identify the causal relationship between the number of children and female labor supply. Some examples include twins at first birth (Rosenzweig and Wolpin (1980); Bronars and Grogger (1994)) and sex of the two first children (Angrist and Evans (1998)). Most of these approaches show a reduced but still significant effect of children on female labor supply.

In Latin America there are few studies which address the endogeneity of fertility decisions and provide evidence for developing countries. Among those studies are: Cruces and Galiani (2007), Agüero and Marks (2008) and Delpiano (2012). Cruces and Galiani study the effect of fertility on maternal labor supply in Argentina and Mexico, exploiting the source of exogenous variability in family size introduced by Angrist and Evans, finding that the U.S. results can be generalized both qualitatively and quantitatively to Argentina and Mexico. Agüero and Marks using a subsample of Latin American countries and introducing female infertility as a source of variation in family size, do not find a significant relationship between fertility and mothers' employment. Finally, Cáceres Delpiano studies the impact of fertility on mothers' employment for a sample of developing countries using fertility shock (multiple births) and finds that children have a negative impact on female employment but with different impacts depending on the order that child of different sex are born.

All these results seem contradictory, while Cruces and Galiani (2007), and Delpiano (2012) show that children have a negative impact on female employment, Agüero and Marks (2008) do not find a significant impact. One reason for this is that the local result depends on the instrument used and the compliers are different for the three researches.

The present study highlights some fundamental aspects. The first one is the definition of women's employment. Delpiano (2012) underlines this issue since in developing countries labor markets have higher levels of informality and heterogeneous payment alternatives. I contribute to the discussion by presenting the impact of fertility on the following definitions of employment: a) overall labor participation (paid or unpaid), b) paid work (inside or outside the home), c) work outside the home (paid or unpaid) and d) work outside the home and paid. These definitions pretend to capture location and compensation issues. The second aspect highlighted is the exploration of the external validity of results (Angrist (2004)).

Thus, I find that using Angrist and Evans' instrument, the OLS and IV estimations are significant and negative for married women in all employment definitions used (between 8-9 percentage points). However, for all women the significance of the effect on labor supply depends on the definition of employment: there is a negative impact for "work outside the home" and "paid work outside the home" definitions (8 percentage points). These results are confirmed using boys and girls as instrument and using the number of children as a fertility measure.

The paper has the following structure: Section 2 presents the identification strategy and

the theoretical model. Section 3 presents the data, summary statistics and the construction of the instrumental variable. Section 4 presents the main impacts of fertility on labor supply. Section 5 presents exclusion restrictions related to the instruments used in this study. Finally, section 6 presents the conclusions.

## 2 Identification Strategy and Theoretical Model

### 2.1 Identification Strategy

As Rosenzweig and Wolpin (2000) indicate, impact of fertility on the participation of married women in the labor force was investigated first in the United States. Taking into account that fertility and participation on labor force are endogenous, some identification strategies were the use of twins on the first birth (Rosenzweig and Wolpin (1980); Bronars and Grogger (1994)) and the Angrist and Evans' natural instrument (1998) of the sex of the first two births, specifically sex-sameness. Some recent strategies involve infertility shocks (Aguero and Marks (2008)), exploit the multiple births in higher parities (Delpiano (2012)) or research about a particular sample (Cristia (2008)).

This paper employs the Angrist and Evans' instrument which "exploits the parental preferences for a mixed sibling-sex composition that can be thought as randomly assigned. Then, a dummy variable that indicates whether the sex of the second child matches the sex of the first one provides a plausible instrument for further childbearing among women with at least two children" (Angrist and Evans (1998):451).

This instrument must meet two conditions to be valid:

- It must be relevant. It must be found that parents whom their first two children have the same sex are more likely to have an additional child compared to those who their first two children have different sex (with the sex-sameness instrument).
- I must accomplish the exclusion restriction. It must be confirmed that to have a children of the same sex in the first two births does not directly affect the subsequent labor supply of either parent except through its effect on having an additional birth. This seems plausible since parents cannot control or manipulate their children's sex.

With this in mind, the following regression models is proposed:

$$Y_i = \alpha + w'\delta + \beta x_i + \epsilon_i \quad (1)$$

Where:

$Y_i$  is a measure of labor supply,  $w'$  is a covariates vector: age of women, age at first birth, indicators for the sex of first and second child, dummies by indigenous and urban status.

$x_i$  is the endogenous fertility measure of interest. In this case it is the third child variable or number of children variable. It is replaced in (1) by the predicted value of the following regression to obtain 2SLS estimates.

$$x_i = \rho + w'\lambda + \gamma(\text{Samesex}) + \eta_i \quad (2)$$

Where *Samesex* is a dummy for whether the sex of the second child matches the sex of the first child.

## 2.2 Theoretical Model

In a simple static model, women or families choose the levels of consumption ( $C$ ), time of leisure  $t_l$  and number of children ( $N$ ) that solves the maximization of a utility function<sup>2</sup>  $U = u(C, t_l, N)$

$$U = u(C, t_l, N) \quad (3)$$

Equation 3 is subject to following time and money constraints:

$$T = t_m + t_h + t_l \quad (4)$$

$$I + wt_m = p_c C + p_n n \quad (5)$$

Where: the time restriction has a total time( $T$ ) distributable for market ( $t_m$ ), work at home (including housework and childcare( $t_h$ )) and leisure ( $t_l$ ); the money restriction has a non-labor income( $I$ ), hour-paid wage ( $w$ ), price of goods ( $p_c$ ) and cost of child rearing ( $p_n$ ). Here the female labor supply ( $L$ ) is a function of number of children ( $N$ ) and other variables ( $Y$ ) vector, namely  $L = f(Y, N)$ .

The effect of interest is the labor supply response to changes in fertility. But as fertility can be correlated with omitted variables related with labor supply (as professional ambition, etc.), to identify the direct effect of fertility is necessary to use an instrument ( $Z$ ) that takes into account the exogenous variation of fertility but is not related with labor supply. Thus, the effect of interest can be identified as follows:

$$\partial L / \partial Z = f_y \partial Y / \partial Z + f_N \partial N / \partial Z \quad (6)$$

Since  $Z$  is exogenous with respect to  $Y$ , then  $\partial / \partial Z = 0$ . Therefore, the response of labor supply to changes in fertility is identified as:

$$f_N = (\partial L / \partial Z) / (\partial N / \partial Z) \quad (7)$$

---

<sup>2</sup>This Utility is increasing in all this arguments

### 3 Data, Summary Statistics and Variables

This research uses data on labor supply, fertility and characteristics of household members from the Ecuadorian Population Census of 2010 conducted by the National Institute of Statistics and Census (Instituto Nacional de Estadísticas y Censos INEC).

The main motivation of this paper can be seen in Figure 1, where fertility (measured as percentage of women with more than two children) and female labor supply move in opposite directions. Within the group of women aged from 21 to 35 years old and with two or more children, labor supply has increased by 13 percentage points (i.e. variation of 50% in paid work) while the percentage of women in the group with more than two children decreased by 14 percentage points (variation of 21%), all from 1990 to 2010.

Table 1 presents some labor force participation rates and probabilities of additional children for women of different ages and different marital status for 1990, 2001 and 2010 Censuses.

For the sex-mix instrument, the sample is limited to women between 21 and 35 years old with at least two children and whose oldest child was, at most, 18 years old at the moment of the census<sup>3</sup>. Following Angrist and Evans (1998), in order to match women with their own children, the sample used is conformed by females who are “heads” or “spouses” in each household, then the reported number of children alive is checked to coincide with the number of children in the household matched to the women, restraining the sample to women for whom both numbers were the same.

This is performed for two samples: all women and those married at the time of the census 455.125 observations for the first one and 404.795 observations for the last. Table 2 shows some descriptive statistics and variable definitions for covariates, instruments and dependent variables.

---

<sup>3</sup>As Angrist and Evans argue census does not track children across households; therefore, the sample is limited to mothers aged 21-35 whose oldest child was less than 18 years of age at the moment of the census.

**Table 1:** Fertility and Labor Supply Measures in Ecuador.

Fertility and Labor-Supply Measures	1990 Census	2001 Census	2010 Census
Women aged 21-35			
Mean children ever born	2.8	2.45	2.17
Percent with 2 or more children	74.1	68.61	65.12
Percent work	32.9	40.77	44.91
Percent paid work	29.6	35.1	44.17
Percent work outside the home			37.93
Percent paid employment outside the home			37.55
Observations	744321	898771	1072870
Women aged 36-50			
Mean children ever born	5.19	4.08	3.38
Percent with 2 or more children	90.28	87.85	85.79
Percent work	36.07	45.71	52.15
Percent paid work	32.38	39.77	51.43
Percent work outside the home			42.21
Percent paid employment outside the home			41.89
Observations	580660	832587	1052847
Women aged 21-35 with 2 or more children			
Mean children ever born	3.55	3.13	2.94
Percent with 2 or more children	66.28	57.87	52.17
Percent work	29.61	36.77	40.28
Percent paid work	26.28	31.19	39.53
Percent work outside the home			32.64
Percent paid employment outside the home			32.28
Observations	551520	616683	698674
Married women aged 21-35 with 2 or more children			
Mean children ever born	3.55	3.13	2.94
Percent with 2 or more children	66.42	57.92	52.14
Percent work	27.84	34.47	36.85
Percent paid work	24.5	28.97	36.15
Percent work outside the home			29.22
Percent paid employment outside the home			28.9
Observations	520635	568594	617362

Note: 1990 and 2001 census did not asked about work inside or outside from home. The married samples include women married or “common law marriage” at the time of the Census.

The variable used as a measure of fertility is the indicator of more than two children. The first instrumental variable for fertility is the indicator Same sex. Other possible instruments are Two boys and Two girls.

The labor supply variables are:

- Work: equals 1 when a mother worked at least one hour last week. This includes those mothers who: a) worked at least one hour prior to the census week, b) did not work but had a work for which she was absent, c) created a product or provided a service for one hour at least, d) helped in some business or work in the family for one hour at least or e) performed agriculture for one hour at least.
- Paid work: equals to 1 when the mother worked for pay and is not a family worker without remuneration. It includes: employees (private or public), self-employed, owners and managers (employers, partner), laborers or peons and domestic servants.
- Worked away from home: equals 1 when the mother’s job is out of home.

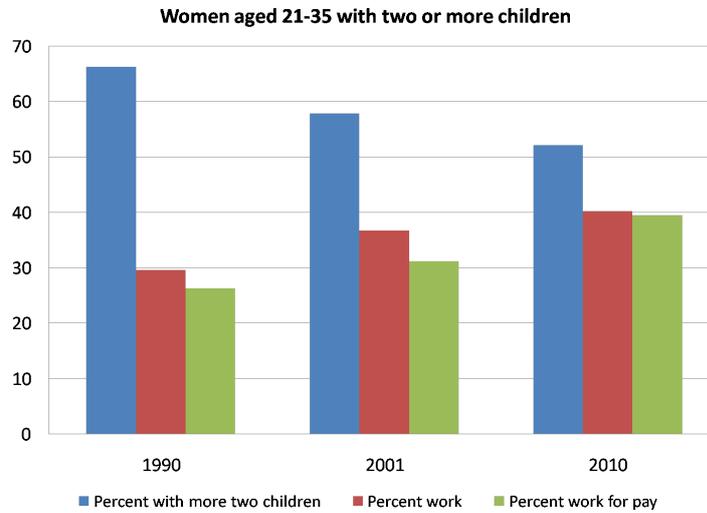


Figure 1: Ecuadorian Fertility and Female Labor Supply. Census 1990-2001-2010

Table 2: Descriptive Statistics. Women aged 21-35 with 2 or more children. 2010 Census data

Variables	All Women	Married
Children ever born	2.731 (0.967)	2.739 (0.974)
More than 2 children	0.481 (0.5)	0.484 (0.5)
Boy first (=1 if first child was a boy)	0.512 (0.5)	0.513 (0.5)
Boy second (=1 if second child was a boy)	0.509 (0.5)	0.509 (0.5)
Two boys (=1 if first two children were boys)	0.263 (0.44)	0.263 (0.44)
Two girls (=1 if first two children were girls)	0.242 (0.429)	0.241 (0.428)
Same sex (=1 if first two children were the same sex)	0.505 (0.5)	0.505 (0.5)
Age	29.477 (3.786)	29.42 (3.8)
Age at first birth	19.327 (3.08)	19.378 (3.101)
Schooling	8.855 (4.36)	8.85 (4.37)
Work (=1 if worked at least an hour in last week to census)	0.411 (0.492)	0.378 (0.485)
Paid work (=1 if worked for pay in last week to census)	0.403 (0.491)	0.371 (0.483)
Worked outside the home (=1 if worked outside the home)	0.333 (0.471)	0.300 (0.458)
Paid employment outside the home (=1 if worked for pay and outside the home)	0.329 (0.47)	0.297 (0.457)
Number of observations	455125	404795

Note: The samples include women aged 21-35 with two or more children except for women whose second child is less than a year old. Standard deviations in parentheses

- Worked for pay outside the home: equals 1 when the mother's job is paid and outside the home.

The dimensions that can be explored with these definitions are: location of job in relation to home, remuneration and labor participation.

## 4 Main Results

Following Angrist and Evans, table 3 shows estimates of the impact of child sex and sexmix on fertility, where 48.9% of all women have one girl and 51.1% have one boy at the first birth. The fraction of women with at least one child and who had a second child, conditional on the sex of the first child is 67% in both cases. This presents evidence that there is no impact of the sex of the firstborn on fertility. It is important to mention that, in married women, there are differences conditionals on the sex of the first child but this difference disappears when including controls in the regression at the first stage.

For the second analysis, table 3 presents the fraction of women who have a third child conditional on the sex composition of the first two children, where 46.4% of women with one boy and one girl have a third child, compared to 49.9% for women with two girls or two boys. That is a significant difference of 3.6 percentage points. These results are confirmed in table 4 which shows the first stage of the instrument for all and for married women, including controls, and results for the other possible instrument (two boys and two girls). The difference of 3.6 and 3.8 percentage points for all and for married women found here means that Ecuadorian women with two children of the same sex are 3.6 and 3.8 percentage points more likely to have a third child than mothers of one boy and one girl. For the United States in 1980 this difference was 6 percentage points for all women (Angrist and Evans). Cruces and Galiani found a difference of 3.5 and 3.2 percentage points (subsample of all women) for Argentina and Mexico, respectively. On the other hand, the instrument of two boys and two girls is also significant for explaining fertility.

To check a random assignment of Same-sex instrument, table 5 compares demographic characteristics of the mother among those who had a composition of same-sex (treated group) and mixed-sex (control) sibling compositions. This table includes the following variables: age of woman, age at first birth, indigenous ethnicity, years of education and residence area. None of these variables presents significant difference

For 2SLS estimates the control variables are: age of women, age at first birth, a dummy variable to indicate the sex of first and second children, a dummy variable for ethnic identification and a dummy for urban area. For the cases of two boys and two girls as instrument, the covariates exclude the sex of the second children. Table 6 shows that OLS estimates present a negative and significant impact for all measurements of work and all groups of women. The 2SLS estimates indicate there is an impact on labor supply when moving from 2 to 3 children for all women grouped in two work measurements (working outside from home and

paid work outside home) with a negative impact of 8 percentage points. Finally, for married women there is a negative impact of 8 to 9 percentage points in all work measurement<sup>4</sup>.

**Table 3:** Fraction of families that had another child

Sex of first child in women with one or more children				
	All women		Married Women	
	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had another child
(a) one girl	0.489	0.674	0.488	0.685
		[0.0008]		[0.00087]
(b) one boy	0.511	0.676	0.512	0.688
		[0.00079]		[0.00084]
difference (b)-(a)		0.0018		0.0025
		[0.001141]		[0.0012]
Sex of first two children in women with two or more children				
	All women		Married Women	
	Fraction of sample	Fraction that had another child	Fraction of sample	Fraction that had another child
two girls	0.242	0.502	0.242	0.506
		[0.0015]		[0.0016]
two boys	0.263	0.497	0.263	0.4999
		[0.00144]		[0.0015]
(a) one boy, one girl	0.495	0.464	0.496	0.466
		[0.00105]		[0.0011]
(b) both same sex	0.505	0.499	0.504	0.503
		[0.00104]		[0.0011]
difference (b)-(a)		0.036		0.037
		[0.0015 ]		[0.00157]

Note: 2010 Census data (Census-INEC). The samples are the same as in Table 2. Standard errors in brackets.

**Table 4:** First Stage: Fertility measure is third child variable

	All Women		Married Women	
	(1)	(2)	(1)	(2)
same_sex	0.0365		0.0382	
	[0.0014]***		[0.0014]***	
two_boys		0.0331		0.0347
		[0.0019]***		[0.0020]***
two_girls		0.0401		0.0417
		[0.0020]***		[0.0021]***
R <sup>2</sup>	0.14	0.14	0.15	0.15
N	455125	455125	404795	404795

Note: 2010 Census data (Census-INEC). \* p<0.1; \*\* p<0.05; \*\*\* p<0.01 Robust standard errors in brackets. Covariates: age of women, age at first birth, indicators for Boy 1st, Boy 2nd and dummies by indigenous and urban status. The variable Boy 2nd is excluded from columns (2).

<sup>4</sup>The results considering the number of children as a measure of fertility, also show a reduction in labor supply. These results are available if required

**Table 5:** Random assignment of the same sex

Variable	By Same Sex				
	Controls	sd	Treated	Sd	p-value
Age	29.482	(3.785)	29.471	(3.788)	[0.320]
Age at first birth	19.321	(3.079)	19.321	(3.081)	[0.927]
Indigenous	0.081	(0.274)	0.08	(0.272)	[0.191]
Urban area	0.608	(0.488)	0.609	(0.488)	[0.687]
Years of education	8.853	(4.365)	8.857	(4.358)	[0.832]
N	225180		229945		

Note: Data source: 2010 Census. The sample is all women. Standard errors are reported in parentheses.

It is important to mention that work measurement are relevant in the sample of all women since there is an impact only when the definition of work involves characteristics incompatible with child rearing. On the other hand, for married women there is an impact using all measurement of work. Perhaps this happens since married women are supported with their spouses' revenues.

**Table 6:** OLS and 2SLS. Estimates of Labor-Supply Models Using 2010 Census Data

	All Women			All Women		
	OLS	2SLS (1)	2SLS (2)	OLS	2SLS (1)	2SLS (2)
Instrument for More than 2 children		Same sex	Twoboys, Twogirls		Same sex	Twoboys, Twogirls
Dependent Variable						
Work	-0.1120 [0.0015]***	-0.0569 [0.0393]	-0.0527 [0.0392]	-0.1078 [0.0016]***	-0.0800 [0.0391]**	-0.0788 [0.0391]**
Work for pay	-0.1118 [0.0015]***	-0.0599 [0.0392]	-0.0549 [0.0391]	-0.1076 [0.0016]***	-0.0834 [0.0390]**	-0.0809 [0.0389]**
Work outside home	-0.1084 [0.0015]***	-0.0847 [0.0378]**	-0.0815 [0.0377]**	-0.1036 [0.0015]***	-0.0934 [0.0372]**	-0.0921 [0.0371]**
Paid Work outside the home	-0.1080 [0.0015]***	-0.0893 [0.0376]**	-0.0860 [0.0375]**	-0.1031 [0.0015]***	-0.0988 [0.0370]***	-0.0974 [0.0369]***

Note: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01. Robust standard errors in brackets. Covariates: age of women, age at first birth, indicators for Boy 1st, Boy 2nd and dummies by indigenous and urban status. In (2) the variable Boy 2nd is excluded.

Also, the OLS estimates and 2SLS are similar for married women but for all women depend on the definition of work. The results are confirmed by the two boys and two girls' instrument.

These results are local for the compliers, that is, for women whose fertility decision is changed for the instrument, which is used for identifying the impact of labor supply when children move from 2 to 3, but don't identify other increases in fertility as 0 to 1 child<sup>5</sup>.

<sup>5</sup>With the 2001 Census, the first stage is similar to 2010 for both samples. The 2SLS estimates present

**Table 7:** Exclusion restriction: Women aged 21-35 with two children

Expenditure	Controls		All women		p-value
		sd	Treated	sd	
Per-Child Clothing and shoes expenditure	35.57	(35.663)	36.63	(35.798)	[0.408]
Per-Child clothing expenditure	25.68	(27.476)	26.29	(26.986)	[0.534]
Per-Child shoes expenditure	10.41	(10.287)	10.84	(10.943)	[0.268]
Per-Child spending on clothing children 3 to 12 years	8.26	(9.887)	8.29	(9.432)	[0.930]
Per-Child spending on shoes children 3 to 12 years	4.06	(4.006)	4.08	(3.976)	[0.908]
N	1648		1506		

Note: National Household Survey of Income and Expenditure 2011-2012 INEC. Treated group is same sex equals 1 and control group is same sex equals 0.

## 5 Exclusion restriction

One concern in relation to the instrument applied in this paper is raised by Rosenzweig and Wolpin (2000), who say that the same sex instrument can affect labor supply through economies of scale, and thereby reducing the cost of childcare. To evaluate this, I used data from the Ecuadorian survey of income and expenses and I found that expenses that may involve some form of economies of scale do not have difference between households with two children of the same sex compared to households with two children of different sexes (Table 7).

## 6 Conclusions

The OLS estimates indicate that women with more than two children are 8 and 9 percentage points less likely to work than women with two children for the sample of all women and married women, respectively.

To estimate the causal effect of fertility on female labor supply I use sex composition of first two children as instrumental variable. The first stage shows that families with two boys or two girls are 3.6 and 3.8 percentage points more likely to have a third child than families with one boy and one girl, for the sample of all and married women respectively.

The 2SLS results show a causal impact 8 and 9 percentage points of decrease of female labor supply by having a third child in all and married women respectively.

Two aspects are important here: first, the results apply for the “compliers”. This means that the result is local (Local Average Treatment Effect-LATE), that is, for women who changed their fertility decision due to the instrument. Second, these results refer for moving

a negative and significant impact (at 10% of confidence level) of about 7 percentage points on the outcome “paid work” and don’t present impact on the “work” for both samples (Annex 1).

from 2 to 3 children but do not refer to other increases in fertility as in going from 0 to 1 child.

## References

- Aguero, J. and Marks, M. (2008). Motherhood and Female Labor Force Participation: Evidence from Infertility Shocks. *American Economic Review: Papers and Proceedings*, 98 (2):500–504.
- Ahn, N. and Mira, P. (2002). A Note on the Changing Relationship between Fertility and Female Employment Rates in Developed Countries. *Journal of Population Economics*, 15 (4):667–682.
- Angrist, J. (2004). Treatment Effect Heterogeneity in Theory and Practice. *The Economic Journal*, 114:C52–C83.
- Angrist, J. and Evans, W. (1998). Children and Their Parents' Labor Supply:evidence from Exogenous Variation in Family Size. *The American Economic Review*, 88 (3):450–477.
- Bargain, O., González, L., Keane, C., and Ozcan, B. (2012). Female labor supply and divorce: New evidence from Ireland. *European Economic Review*, 56(8):1675–1691.
- Bronars, S. and Grogger, J. (1994). The Economic Consequences of Unwed Motherhood:Using Twin Births as a Natural Experiment. *American Economic Review*, 84 (5):1141–1156.
- Browning, M. (1992). Children and Household Economic Behavior. *Journal of Economic Literature*, XXX:1434–1475.
- Cristia, J. (2008). The Effect of a First Child on Female Labor Supply: Evidence from Women Seeking Fertility Services. *Journal of Human Resources*, 43 (3):487–510.
- Cruces, G. and Galiani, S. (2007). Fertility and Female Labor Supply in Latin America: New Causal Evidence. *Labour Economics*, 14 (3):565–573.
- Delpiano, C. (2012). Can we still learn something from the relationship between Fertility and Mother's Employment? Evidence from Developing Countries. *Demography*, 9 (1).
- Gronau, R. (1973). The Effect of Children on the Housewife's Value of Time. *Journal of Political Economy. Part II*, 81 (2):S168–99.
- Hotz, J., McElroy, S., and Sanders, S. (2005). Teenage Childbearing and its Life Cycle Consequences Exploiting a Natural Experiment. *Journal of Human Resources*, 40 (3):683–715.

- Imbens, G. and Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62 (2):467–75.
- Maurin, E. and Moschion, J. (2009). The Social Multiplier and Labor Market Participation of Mothers. *American Economic Journal: Applied Economics*, 1 (1):251–72.
- Rosenzweig, M. and Wolpin, K. (1980). Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment. *Econometrica*, 48 (1):227–240.
- Rosenzweig, M. and Wolpin, K. (2000). Natural Natural Experiments' in Economics. *Journal of Economic Literature*, 38(4):827–874.

### Annex: 2001 Census Data

*First Stage: Fertility measure is third child variable*

	All Women		Married Women	
	(1)	(2)	(1)	(2)
same_sex	0.0375 [0.0015]***		0.0380 [0.0016]***	
two_boys		0.0355 [0.0022]***		0.0362 [0.0022]***
two_girls		0.0395 [0.0022]***		0.0399 [0.0023]***
R <sup>2</sup>	0.16	0.16	0.17	0.17
N	354272	354272	329712	329712

Note: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01. Robust standard errors in brackets. Covariates: age of women, age at first birth, indicators for Boy 1st, Boy 2nd and dummies by indigenous and urban status. In (2) the variable Boy 2nd is excluded.

*OLS and 2SLS. Estimates of Labor-Supply Models Using 2001 Census Data*

	All Women			Married Women		
	OLS	OLS (1)	2SLS (2)	OLS	2SLS (1)	2SLS (2)
Instrument for More than 2 children		Same sex	Twoboys, Twogirls		Same sex	Twoboys, Twogirls
Dependent Variable						
Work	-0.0901 [0.0018]***	-0.0504 [0.0430]	-0.0491 [0.0430]	-0.0830 [0.0018]***	-0.0448 [0.0433]	-0.0438 [0.0433]
Paid work	-0.0882 [0.0017]***	-0.0710 [0.0408]*	-0.0702 [0.0408]*	-0.0819 [0.0017]***	-0.0685 [0.0410]*	-0.0676 [0.0410]*

Note: \* p<0.1; \*\* p<0.05; \*\*\* p<0.01. Robust standard errors in brackets. Covariates: age of women, age at first birth, indicators for Boy 1st, Boy 2nd and dummies by indigenous and urban status. In (2) the variable Boy 2nd is excluded.

# Analítica

Efectos Macroeconómicos de la Política Fiscal en  
Ecuador 1993-2009

Paúl A. Carrillo Maldonado



[www.ecuadorencifras.gob.ec](http://www.ecuadorencifras.gob.ec)





# Efectos Macroeconómicos de la Política Fiscal en Ecuador 1993-2009

Paúl A. Carrillo Maldonado <sup>†</sup> \*

<sup>†</sup> *Departamento de Estudios Tributarios, Centro de Estudios Fiscales, Servicio de Rentas Internas, Quito, Ecuador*

<sup>†</sup>pacarrillom@gmail.com; <https://sites.google.com/site/paulacarrillomaldonado/>

## Resumen

La política fiscal tiene especial relevancia en economías dolarizadas. Para el caso ecuatoriano, se analizan los efectos de los impuestos y el gasto público en el desempeño general de la economía para el período 1993-2009. Para ello, se utiliza un modelo de vectores autoregresivos estructurales (SVAR) incorporando los hechos estilizados. Los principales resultados son: i) los impuestos tienen efectos solo temporales sobre la economía ecuatoriana, ii) el aumento de los impuestos indirectos influye de manera negativa en las importaciones, las exportaciones y el PIB, iii) el aumento de los tributos directos tiene un efecto positivo sobre las exportaciones, vi) un *shock* positivo del consumo de gobierno tiene un efecto sobre los impuestos y la inversión, v) los instrumentos de política fiscal tiene una alta interrelación entre ellos.

**Palabras clave:** Política Fiscal, PIB, SVAR, Hechos Estilizados, Ecuador.  
**Código JEL:** C32, E20, E62, H20.

## Abstract

Fiscal policy is particularly relevant in dollarized economies. For the case of Ecuador, we analyze the effects of taxes and public spending on the overall performance of the economy for the period 1993-2009. To do this, we use

---

\*El autor agradece los comentarios de José Ramírez, Ivan Gachet, Diego Maldonado, Miguel Acosta y Juan Carlos Serrano. Al momento de realizar esta investigación, el autor era parte del Centro de Estudios Fiscales del Servicio de Rentas Internas. Las opiniones, errores y omisiones son de responsabilidad exclusiva del autor y no necesariamente reflejan la posición oficial del Centro de Estudios Fiscales, ni de sus autoridades.

a structural autoregressive vector model (SVAR) incorporating the stylized facts of Ecuador estimated on [19]. The main results are: i) the taxes have only temporary effects on the ecuadorian economy, ii) the increase in indirect taxes affect negatively on imports, exports and GDP, iii) the increase in direct taxes has a positive effect on exports, vi) a positive shock of government consumption has an effect on taxes and investment, v) fiscal policy instruments have a high interrelation between them.

**Keywords:** Fiscal policy, GDP, SVAR, Stylized Facts, Ecuador

## 1. Introducción

Los académicos y hacedores de política fiscal mantienen una amplia discusión sobre la influencia de la política fiscal en la actividad económica. Desde la teoría, el efecto de la política fiscal sobre la demanda agregada depende de varios supuestos como la rigidez de precios, la capacidad de oferta, la apertura comercial, entre otros. Mientras que, en el enfoque empírico no existe un consenso sobre tales consecuencias ya que los resultados varían dependiendo de la metodología [37]. El único punto en común es que el sector real y fiscal tienen una alta interrelación.

Algunos estudios, como [8], [16], [3], encuentran que la política fiscal tiene importantes efectos sobre la dinámica de la economía. Para el caso ecuatoriano, [30] muestra que la política fiscal no tiene efecto sobre el ciclo económico. Estos resultados contradicen varios análisis realizados para Suramérica [13, 37, 26, 29]. Una debilidad del estudio de [30] es que utiliza el filtro de Holdrick-Prescott, un instrumento criticado por la elección arbitraria de sus parámetros y el sesgo que presenta ante variables integradas<sup>1</sup>. En este caso, la posible presencia de relaciones espurias dificulta la identificación de los efectos de la política fiscal sobre la economía del país.

El presente estudio evidencia los efectos expansivos de la política fiscal en los componentes del PIB a través de un modelo de vectores autorregresivos estructurales (SVAR, por sus siglas en inglés) de largo plazo. En particular, se estudia, mediante la función impulso-repuesta, el impacto de cambios en los impuestos y el gasto público sobre el PIB y su composición. Además, se mide la contribución de los instrumentos fiscales a la variabilidad del PIB, en el mediano y largo plazo, utilizando la descomposición de la varianza del error.

Los modelos SVAR tienen la ventaja de analizar en conjunto el comportamiento histórico de las variables con las restricciones económicas. [8], [13], [4] utilizan este tipo de restricciones teóricas para identificar la influencia de las variables fiscales sobre la economía. Para [22], “una macroeconomía realista debe basarse en las generalidades empíricas (o hechos estilizados) de la economía en estudio”. En esta línea, el presente estudio implementa las restricciones empíricas en base al trabajo de hechos estilizados de [19], que encuentra las relaciones empíricas de corto y largo plazo para algunos agregados macroeconómicos de Ecuador.

Identificar los efectos dinámicos de los impuestos y el gasto público sobre los componentes del PIB es importante para el Estado ecuatoriano, principalmente porque al ser una economía dolarizada, los instrumentos de política

<sup>1</sup>[17] amplía las críticas al filtro de Holdrick-Prescott

fiscal son los mayor relevancia para influir en la actividad económica. En el análisis se utiliza series trimestrales para el período 1993-2009. La principal limitación para extender este periodo son los datos de los impuestos directos e indirectos que se encuentran disponibles de manera oficial solo desde 1993.

Los resultados indican que el sector fiscal dinamiza al PIB y sus componentes, según sea el caso. Aumentos de los impuestos indirectos tienen un efecto negativo en las importaciones, las exportaciones y el PIB. Además, shocks positivos de los tributos directos tienen un efecto positivo sobre las exportaciones. Los tributos directos afectan también a las tasas de crecimiento de las exportaciones y del PIB. Los instrumentos de política fiscal tiene una alta interrelación entre ellos. El aumento del consumo público tiene mayor interrelación, que los impuestos, con el ingreso disponible. Los impuestos tiene una relación más estrecha con las exportaciones que con las importaciones; mientras que, el consumo gubernamental tiene más interrelación con las importaciones.

El trabajo se organiza de la siguiente manera. La sección 2 recoge la literatura acerca de las investigaciones que utilizan modelos SVAR para el análisis de la política fiscal. La tercera parte explica el modelo aplicado para Ecuador. La cuarta sección expone los principales resultados con la función de Impulso-Respuesta. La sección 5 muestra las funciones de descomposición de la varianza. La última sección concluye.

## 2. Política fiscal y modelos SVAR: revisión de literatura

[7], [10] y [36] desarrollaron modelos SVAR para identificar los efectos de la política gubernamental [23]. Sin embargo, estos modelos se han utilizado principalmente para monitorear los cambios de la política monetaria.

[8] emplean un modelo SVAR de corto plazo para analizar los efectos dinámicos del gasto de gobierno y los impuestos en la actividad productiva de los Estados Unidos después de la Segunda Guerra Mundial. El trabajo concluye que el aumento del gasto público tienen un efecto positivo sobre el producto; los shocks positivos de los impuestos tienen un efecto negativo sobre el PIB; y, finalmente, un aumento simultáneo de los impuestos y el gasto público tiene un efecto negativo sobre la inversión. Con similares resultados, [32] extiende este análisis a 5 países de la Organización para la Cooperación y el Desarrollo Económico (OCDE), e incluye 2 variables más, el nivel de precios y las tasas de interés. Para [32], la política fiscal ha sido débil en los últimos 20 años.

[3] y [4] utilizan un modelo SVAR para identificar los efectos de los impuestos y el gasto público sobre la economía. Estos estudios se basan en el trabajo de [8] para evaluar los shocks de los impuestos. [3] realizan un análisis de 4 tipos de impuestos (al ingreso, a las empresas, indirectos y a la seguridad social) para Canadá. El modelo incluye el índice de precios y la tasa de los bonos del tesoro. Dicho estudio demuestra también que el impuesto a la renta de las empresas tiene un impacto positivo sobre la producción. [4] construye un modelo estructural de vectores autorregresivos para estimar los efectos discrecionales de la política fiscal sobre la producción en Austria. [4] indica que, además del producto, el consumo y la inversión decrecen cuando se aumenta los impuestos. Los dos estudios muestran que los cambios positivos de los impuestos tienen efectos negativos temporales sobre la producción.

En Europa, [14] estiman un modelo SVAR para conocer los efectos de cambios exógenos de la política fiscal en España. El aumento de los impuestos tiene un resultado negativo sobre la actividad económica en el mediano plazo, mientras que éstos tienen un efecto positivo temporal sobre el déficit público. Además, los autores indican que la política fiscal es anticíclica en este país. Para Alemania, [21] analiza la influencia de los impuestos sobre la producción con un modelo SVAR cointegrado. Este estudio corrobora los resultados de los antes mencionados, pues demuestran que el aumento de los impuestos tiene efectos negativos sobre el producto. Además, [31] estiman los efectos macroeconómicos de largo plazo que tienen los impuestos en cuatro países de la Unión Europea (UE) utilizando un SVAR. Los principales resultados indican que un mismo cambio en los instrumentos fiscales genera diferentes reacciones en la producción y los precios de las economías europeas; los efectos keynesianos asociados a las políticas fiscales no son dominantes en el largo plazo, y los efectos nominales tienen mayor persistencia que los efectos reales.

[16] evalúan los efectos de la política fiscal y monetaria, en conjunto, para la economía de Nueva Zelanda. La investigación expone la estimación tradicional de las restricciones de los modelos SVAR, el método *sign restrictions* y las restricciones de largo plazo con cointegración. Las conclusiones a lo que se llega son que la identificación de los efectos de la política monetaria ha tenido mayor atención en la literatura dejando de lado la política fiscal. Además, indican que los shocks de política fiscal tienen mayor influencia que los de la política monetaria. En particular, se muestra que los tributos y los déficits fiscales son más importantes que el consumo de gobierno y la política monetaria en conjunto.

[13], [26], [29], [28], [30], [33], [37] analizan los efectos de los impuestos y el gasto público en la producción, con modelos SVAR, para algunos países suramericanos. [13] realiza un modelo estructural para la economía chilena en base a [8] y [32]. Los autores muestran que el aumento de los impuestos

tiene efectos negativos temporales en el producto. Además, indican que este shock no es tan significativo como en otros países como Estados Unidos y Canadá. En Perú, [29] confirma que el incremento de los impuestos tiene un efecto negativo sobre la producción. Este estudio realiza dos modelos para los períodos 1980-1990 y 1990-2006. El primer modelo muestra que el efecto es negativo aunque insignificativo; mientras que, el segundo, indica que el impacto es negativo y significativo. El trabajo de [26] identifica la influencia de los impuestos en la actividad económica de Colombia. Los autores estiman varios modelos que demuestran que algunos impuestos tienen efectos negativos temporales y de mediano plazo (6 trimestres). Además, la investigación muestra que el aumento del total de los impuestos tiene efectos positivos sobre el PIB, resultado que coincide con los encontrados para Italia, Alemania y República Checa. [37] calibra un modelo SVAR para Uruguay tanto teórico como empírico. Este estudio evidencia que el aumento de los impuestos tiene efectos negativos temporales sobre la producción uruguaya.

Por otro lado, [28] y [30] exponen resultados diferentes a los estudios suramericanos antes mencionados. [28] modela un SVAR, para Brasil, con restricciones identificadas con técnicas bayesianas. Los autores se basan en el estudio de Monford y Uhlig (2005) para calibrar los shocks en las funciones de Impulso-Respuesta. Los resultados indican que los ingresos netos no tienen efectos sobre la economía brasileña. Además, indican que el aumento del gasto público tiene un efecto negativo. [30] analiza los efectos de las variables fiscales en el ciclo del PIB de Ecuador. El estudio utiliza el filtro de Holdrick-Prescott para estimar el ciclo de las variables. Los resultados muestran que los impuestos no influyen en el ciclo del producto, lo cual implica que esta herramienta no ayuda a estabilizar la producción ecuatoriana.

Finalmente, [33] es un caso diferente a los anteriores ya que modela las economías de Chile y Colombia mediante SVAR y SVEC (Vectores Estructurales de Corrección de Error). Por un lado, los autores utilizan un SVAR mostrando que los impuestos tienen efectos negativos sobre el producto de Chile; mientras que, para Colombia los efectos son nulos. Estos resultados son consistentes con los mostrados por [13]. Sin embargo, dichos hallazgos no concuerdan con el trabajo de [26] ya que este muestra que los impuestos tienen efectos negativos o positivos. Por otro lado, realizan un modelo SVEC para identificar los efectos de largo plazo. Este modelo muestra que los efectos son los mismos que con el SVAR. Este estudio concluye que las finanzas públicas son más efectivas cuando están bajo un control más estricto como es el caso de Chile. No obstante, [37] indican que los resultados de [33] pueden explicarse por la metodología que utilizan para la evaluación de la política

fiscal<sup>2</sup>.

La literatura expone que el gasto de gobierno tiene un efecto positivo temporal sobre el producto; mientras que, los impuestos tienen efectos negativos. La presente investigación pretende corroborar o descartar estos resultados mencionados. A excepción de [37], la literatura mencionada calibra los modelos SVAR en función del trabajo de [8]. La excepción calibra el modelo con los postulados de la teoría económica y el análisis empírico de la economía uruguayana. El modelo SVAR planteado para Ecuador toma como parámetros sus hechos estilizados. Estos se encuentran con la metodología de [19], los cuales obtienen las generalidades empíricas para los agregados macroeconómicos en el corto y largo plazo; dicho estudio menciona que es importante identificar los hechos estilizados para desarrollar modelos dinámicos (como los SVAR) con “mayor consistencia económica”.

### 3. El modelo para Ecuador

#### 3.1. Evolución de los datos

Las variables que se toman para el análisis son: Producto Interno Bruto (PIB), consumo de hogares, consumo de gobierno, inversión, exportaciones, importaciones, impuestos directos (impuesto a la renta) e indirectos (impuesto al valor agregado e impuesto a los consumos especiales)<sup>3</sup>, con una periodicidad trimestral a precios del 2000<sup>4</sup>. Además, el análisis se realiza en función del crecimiento relativo. Las series se desestacionalizan utilizando la metodología TRAMO-SEATS<sup>5</sup>, ya que el componente estacional puede distorsionar el análisis [15].

En su evolución, se observa que los impuestos directos tienen una mayor

<sup>2</sup>[33] utiliza un método diferente al de [8]. Los autores utilizan una matriz de factorización que ayuda a identificar las restricciones del modelo

<sup>3</sup>Para el periodo de análisis, los impuestos indirectos incluyen: IVA e ICE; mientras que, los directos contienen: impuesto a la renta, impuesto a los vehículos motorizados, impuesto a los activos en el exterior, RISE, tierras rurales, interés por mora tributaria, multas tributarias y otros ingresos. Se escoge al impuesto a la renta, impuesto al valor agregado e impuesto a los consumos especiales ya que tienen la mayor participación en la recaudación tributaria. Además, que la inclusión del resto no influye en los resultados por su baja participación en la recaudación.

<sup>4</sup>Las fuentes de información son: Banco Central del Ecuador y Servicio de Rentas Internas

<sup>5</sup>Este método se desarrolló en el Banco de España por creada Victor Gómez y Agustín Maravall en 1996. Las siglas de TRAMO quieren decir: “Time Series Regression with ARIMA Noise, Missing Observations, and Outliers” las de SEATS “Signal Extraction in ARIMA Time Series”.

volatilidad que los indirectos y el consumo de gobierno. El crecimiento del impuesto la renta llega a tasas de hasta el 150 %, pero también tiene porcentajes negativos mayores a 50 %. Esto se debe que en dicho periodo (1999 y 2000) existieron cambios normativos que afectaron a su recaudación. En 1999, se sustituyó este impuesto por uno que grababa la circulación de los capitales. Sin embargo, este se eliminó en el año 2000, retornando el impuesto a la renta. En adelante, no existen cambios normativos que provocaron estas fluctuación (Figura 1).

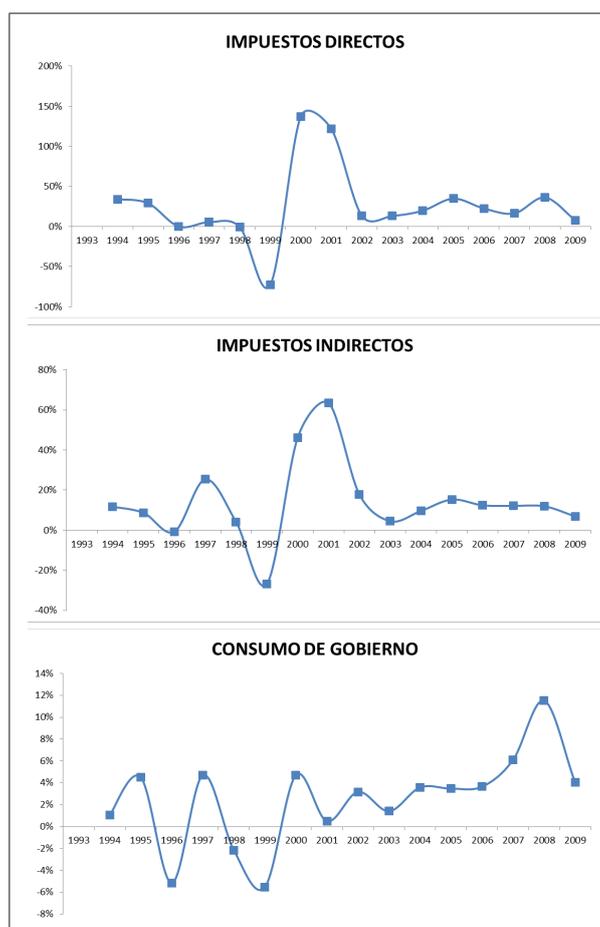


Figura 1: Evolución de las variables fiscales. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

Los impuestos directos tiene tasas de variación positivas hasta de 70 % y negativas de 30 %. La primera tasa alta de IVA e ICE fue en 1997, debido a la creación del Servicio de Rentas Internas. La tasa negativa de 1999 se debe a la crisis financiera que pasó el Ecuador en ese año. Mientras que, las altas

tasas de 2000 y 2001 se debe al aumento de la alícuota de IVA de 10 % a 12 % y de este último a 14 % temporalmente, respectivamente (Figura 1).

Antes de la dolarización se observa que el consumo final del gobierno tiene tasas positivas acompañadas, en seguida, por negativas. Desde el 2001, esta evolución cambia en su crecimiento. A partir del 2007, se observa una tasa de 6 % y 13 %, debido al cambio de política del nuevo gobierno, enfocándose en inversión y desarrollo social. Al final de la serie se muestra un bajo crecimiento debido a la crisis internacional (Figura 1).

### 3.2. El modelo SVAR

Dentro de las herramientas econométricas que permiten el análisis conjunto de variables endógenas está el modelo de vectores autorregresivos (VAR), integrando, de manera implícita, el criterio de simultaneidad entre variables. El modelo VAR es un sistema de ecuaciones dinámicas que examinan la interrelación entre variables económicas con una buena representación estadística de las relaciones pasadas y presentes de las variables. [35] introduce inicialmente estos modelos mostrando que proveían un enfoque más coherente y creíble para la descripción de los datos, pronóstico, análisis estructural de la economía y medición del impacto de decisiones de política económica [5]. Los modelos VAR asumen que la economía está descrita por la evolución pasada y el presente de las variables observadas [25].

Estos modelos presentan algunas características que hacen que sean los más adecuados para la estimación de efectos de políticas públicas a mediano y largo plazo [18]. Por una parte, los VAR permiten observar como el cambio de una variable afecta al comportamiento de las otras. Por otra parte, esta metodología trata con igualdad a las variables sin que exista una distinción a priori entre variables endógenas y exógenas; es decir, se supone que ninguna de las variables se determina de manera exógena y, por tanto, todas están interrelacionadas [31].

En su forma reducida, un modelo VAR se puede expresar de la siguiente manera:

$$Y_t = A_0 + \sum_{i=1}^p A_i Y_{t-i} + \varepsilon_t \quad (1)$$

Donde  $Y_t$  es el vector de  $K$  variables endógenas,  $A_i$  es la matriz  $K \times K$  de coeficientes del rezado  $i$  de las variables endógenas y  $\varepsilon_t$  es el vector de residuos o innovaciones que siguen un proceso de ruido blanco con media cero y varianza  $\Sigma$  ( $\varepsilon_t \sim N(0, \Sigma)$ ).

Según [23], los trabajos de [7], [10] y [36] desarrollan una extensión de los VAR. Esta segunda generación de modelos tienen la característica de introducir restricciones provenientes de la teoría económica, por ellos son conocidos como VAR Estructurales (SVAR). Esta metodología tiene la ventaja de validar de manera empírica las diferentes teorías económicas. A este modelo se lo expresa de la siguiente manera:

$$AY_t = \Gamma_0 + \Gamma_1 Y_{t-1} + \varepsilon_t \quad (2)$$

Donde  $A$  es la matrix de coeficientes  $K \times K$  contemporáneos o instantáneos de las variables endógenas. Esta forma de presentación se conoce como forma estructural [27]. Esta expresión puede mostrarse en la forma reducida:

$$Y_t = Z_0 + Z_1 Y_{t-1} + u_t \quad (3)$$

Donde  $Z_0 = A^{-1}\Gamma_0$ ,  $Z_1 = A^{-1}\Gamma_1$  y  $u_t = A^{-1}\varepsilon_t$  ( $u_t = B\varepsilon_t$ ),  $u_t$  son las innovaciones estructurales y  $\varepsilon_t$  son los residuos reducidos. Cuando las errores reducidos se conocen es posible conocer los estructurales ( $\varepsilon_t = Au_t$ ), se llama Modelo A. Si se asume que  $u_t$  es una combinación lineal de  $\varepsilon_t \sim (0, I_k)$ , o sea  $u_t = B\varepsilon_t$ , es factible la ortogonalización la matrix de covarianza de los residuos reducidos ( $\Sigma_u = B\Sigma_\varepsilon B'$ ), conocido como Modelo B. Sin embargo, no existe la mejor ortogonalización, ni el conocimiento pleno de las innovaciones reducidas por lo cual, también se utiliza la combinación de ambos modelos, denominándole como Modelo AB [27]. Además, [9] propone otra metodología para imponer estructura sobre la matrix  $A$  y  $B$ , de manera indirecta. Estos autores indican que se puede tener la estructura  $u_t = C\varepsilon_t$ , donde  $C = A^{-1}B$ , dando las restricciones en  $C$ . Los primeros modelos  $A$ ,  $B$  y  $AB$  son modelos de corto plazo porque las restricciones se imponen en las innovaciones instantáneas, mientras que la descomposición de [9] se lo llama de largo plazo [27]<sup>6</sup>. Esta investigación utiliza la metodología de [9] para construir un modelo de largo plazo, teniendo la siguiente expresión con las variables endógenas:

$$Y_t = C\varepsilon_t \quad (4)$$

Donde  $C$  es la matrix de coeficientes de largo plazo de los errores. Cada elemento de la matrix  $C$  es el coeficiente de largo plazo del error de la variable endógena, de acuerdo a la calibración del modelo. [9] calibra el modelo con restricciones teóricas de largo plazo. Por el contrario, el modelo para Ecuador aplica restricciones empíricas, ya que “una macroeconomía realista debe basarse en ‘hechos estilizados’, o generalizaciones empíricas establecidas de

<sup>6</sup>[27] realiza una discusión detallada de los modelos SVAR.

forma jerárquica a nivel macro, sectorial y micro acerca de la economía en estudio” [22].

### 3.3. Estructura del modelo

Estas generalizaciones empíricas se basan en la investigación de [19]. Los autores plantean dos pasos para identificar los hechos estilizados de Ecuador. *i)* Filtrar el ciclo de las series económicas a través del filtro de Baxter-King y Hodrick-Prescott en dos etapas. En esta investigación se utiliza el filtro de Baxter-King<sup>7</sup> para encontrar las relaciones empíricas ya que tiene más consistencia que el de Hodrick-Prescott [17]. *ii)* Identificar las relaciones de las series mediante las correlaciones cruzadas. Estas relaciones tienen patrones de tiempo y variación.

Las correlaciones se obtienen con el filtro de Baxter-King para el período comprendido entre el primer trimestre de 1993 y el último del 2009 (figura 2). Según [1], una correlación trimestral entre dos variables es significativa (diferente de cero) cuando el coeficiente de correlación es mayor o igual a  $|0,32|$  (en la figura 2, la línea roja entrecortada indica esta pauta). Además, se toma la correlación más alta en valor absoluto cuando dos o más correlaciones son significativas. Un ejemplo de hechos estilizados es que el comportamiento del consumo de gobierno y de las exportaciones es el más adelantado al PIB; estas generalizadas están en concordancia con los resultados obtenidas por [19].

---

<sup>7</sup>Este filtro es una construcción de medias móviles que separa los componentes periódicos de una serie de tiempo que está en una banda específica de frecuencias sin tomar valores de manera discrecional [17].

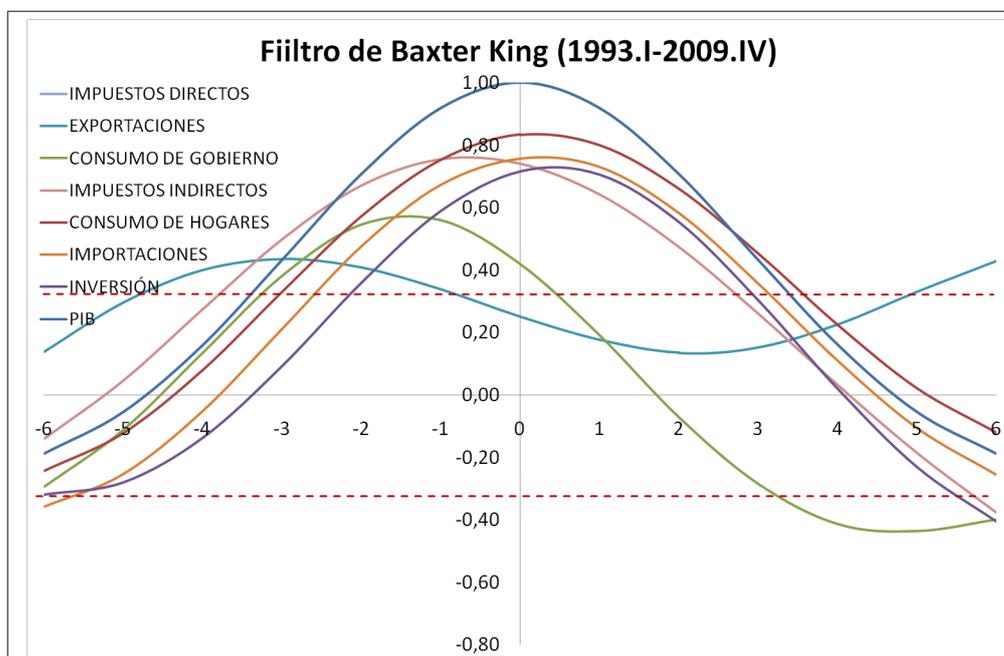


Figura 2: Correlación de las variables  $X(t + i)$  con el  $PIB(t)$ . Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

Por lo tanto, para imponer las restricciones se utiliza los patrones de tiempo de las correlaciones. Además, se supone que el shock estructural ( $\varepsilon_t$ ) está asociado a cada variable (ecuación 14). Esto indica que el shock  $\varepsilon_{cgob}$  es el impacto exógeno generado por el consumo de gobierno y así sucesivamente.

$$\begin{aligned}
 \varepsilon_{1t} &\leftrightarrow \text{Consumo de Gobierno } (\varepsilon_{cgob}) & (5) \\
 \varepsilon_{2t} &\leftrightarrow \text{Impuestos Indirectos } (\varepsilon_{ind}) \\
 \varepsilon_{3t} &\leftrightarrow \text{Inversión } (\varepsilon_{inv}) \\
 \varepsilon_{4t} &\leftrightarrow \text{Importaciones } (\varepsilon_m) \\
 \varepsilon_{5t} &\leftrightarrow \text{Consumo de Hogares } (\varepsilon_{chog}) \\
 \varepsilon_{6t} &\leftrightarrow \text{Producto Interno Bruto } (\varepsilon_{pib}) \\
 \varepsilon_{7t} &\leftrightarrow \text{Impuestos Directos } (\varepsilon_{dir}) \\
 \varepsilon_{8t} &\leftrightarrow \text{Exportaciones } (\varepsilon_x)
 \end{aligned}$$

De esta manera, se plantean las restricciones de largo plazo para el modelo (ecuación 6), de las cuales vale la pena resaltar los siguiente: *i*) todas las variables están sujetas a sus propios shocks. *ii*) el consumo de gobierno (cgob) solo está afectado por su propio shock ( $\varepsilon_{cgob}$ ). *iii*) los impuestos indirectos

(ind) dependen de los shocks del consumo de gobierno. *inv*) la inversión (*inv*) está afectada por los shocks  $\varepsilon_{ind}$  y  $\varepsilon_{cgob}$ . *v*) las importaciones (*m*) tienen el impacto exógeno del consumo de gobierno y los impuestos indirectos. *vi*) el consumo de hogares (*chog*) depende de los shocks  $\varepsilon_{cgob}$ ,  $\varepsilon_{ind}$ ,  $\varepsilon_{inv}$  y  $\varepsilon_m$ . *vii*) el PIB tiene el impacto del consumo de gobierno ( $\varepsilon_{cgob}$ ), los impuestos indirectos ( $\varepsilon_{ind}$ ), la inversión ( $\varepsilon_{inv}$ ), las importaciones ( $\varepsilon_m$ ) y el consumo de hogares ( $\varepsilon_{chog}$ ). *viii*) los impuestos directos están afectados por los shocks del PIB, el consumo de hogares, la inversión, los impuestos indirectos, el consumo de gobierno y las importaciones. *ix*) las exportaciones tiene el impacto de los shocks  $\varepsilon_{cgob}$ ,  $\varepsilon_{ind}$ ,  $\varepsilon_{inv}$ ,  $\varepsilon_m$ ,  $\varepsilon_{chog}$  y  $\varepsilon_{pib}$   $\varepsilon_{dir}$ .

$$\begin{aligned}
 cgob &= C_{11}\varepsilon_{cgob} & (6) \\
 ind &= C_{21}\varepsilon_{cgob} + C_{22}\varepsilon_{ind} \\
 inv &= C_{31}\varepsilon_{cgob} + C_{32}\varepsilon_{ind} + C_{33}\varepsilon_{inv} \\
 m &= C_{41}\varepsilon_{cgob} + C_{42}\varepsilon_{ind} + C_{43}\varepsilon_{inv} + C_{44}\varepsilon_m \\
 chog &= C_{51}\varepsilon_{cgob} + C_{52}\varepsilon_{ind} + C_{53}\varepsilon_{inv} + C_{54}\varepsilon_m + C_{55}\varepsilon_{chog} \\
 pib &= C_{61}\varepsilon_{cgob} + C_{62}\varepsilon_{ind} + C_{63}\varepsilon_{inv} + C_{64}\varepsilon_m + C_{65}\varepsilon_{chog} + C_{66}\varepsilon_{pib} \\
 dir &= C_{71}\varepsilon_{cgob} + C_{72}\varepsilon_{ind} + C_{73}\varepsilon_{inv} + C_{74}\varepsilon_m + C_{75}\varepsilon_{chog} + C_{76}\varepsilon_{pib} + C_{77}\varepsilon_{dir} \\
 x &= C_{81}\varepsilon_{cgob} + C_{82}\varepsilon_{ind} + C_{83}\varepsilon_{inv} + C_{84}\varepsilon_m + C_{85}\varepsilon_{chog} + C_{86}\varepsilon_{pib} + C_{87}\varepsilon_{dir} + C_{88}\varepsilon_x
 \end{aligned}$$

### 3.4. Proceso de estimación

La identificación de los rezagos y parámetros es un paso esencial para un modelo parsimonioso [11]. Una forma de selección del número de retardos son los criterios de información de Akaike (AIC, por sus siglas en inglés), de Hannan-Quin (HQIC) y bayesiano de Schwarz (BIC o SBIC), todos basados en Error de Predicción Final (FPE). Estos utilizan el determinante de la matriz de covarianza de los errores para un modelo de orden  $p$ , el número de parámetros y el tamaño de la muestra. El valor mínimo de los criterios entre los retardos candidatos es la elección óptima. Según [27], los criterios FPE y AIC son estadísticos convenientes para muestras pequeñas (menores a 100 observaciones) y para predicciones. Estos criterios son:

$$FPE(p) = \left[ \frac{T + Kp + 1}{T - Kp - 1} \right]^K \det \tilde{\Sigma}_u(p) \quad (7)$$

$$AIC(p) = \ln \left| \tilde{\Sigma}_u(p) \right| + \frac{2pK^2}{T} \quad (8)$$

$$HQIC(p) = \ln \left| \tilde{\Sigma}_u(p) \right| + \frac{2 \ln \ln T}{T} pK^2 \quad (9)$$

$$SBIC(p) = \ln \left| \tilde{\Sigma}_u(p) \right| + \frac{\ln T}{T} pK^2 \quad (10)$$

En función de lo mencionado, la tabla 1 muestra que se debe utilizar tres rezagos para estimar los coeficientes del modelo SVAR. El modelo con 5 rezagos no es conveniente porque tiene autocorrelación en los rezagos y no normalidad en los errores. En el cálculo de los criterios existe una interrupción en FPE indicando que no puede continuar con el proceso. Este resultado puede deberse a que la muestra de datos es pequeña, una característica de los países latinoamericanos que se caracterizan por no tener un historial adecuado de estadísticas económicas.

Rezagos	FPE	AIC	HQIC	SBIC
1	4.50E-26	-58.4477	-57.476	-55.5163
2	3.00E-26	-59.5428	-57.5994	-53.6798
3	<b>4.9e-28*</b>	-66.4413	-63.5262	-57.6468
4	.	-540.303	-536.417	-528.578
5	.	<b>-577.427*</b>	<b>-572.569*</b>	<b>-562.77*</b>
6	.	-575.408	-569.578	-557.819

Tabla 1: Orden de Rezagos para el Modelo. Fuente: Banco Central Del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

Después de la identificación del rezago, se debe obtener una estimación eficiente de los parámetros. Para [27], los parámetros del proceso VAR no pueden ser estimados por Mínimos Cuadrados Ordinarios (MCO), por lo que generalmente se utiliza la técnica de Máxima Verosimilitud (MV). En especial, cuando los parámetros están sujetos a restricciones (SVAR) es conveniente utilizar MV [25]. Esta técnica asume que los errores tienen una función de distribución normal. Con la forma reducida de la ecuación 3, la función de verosimilitud es la siguiente:

$$L(Z_1, \dots, Z_p) = \prod_{t=1}^T \frac{1}{(\sqrt{2\pi})^K \sqrt{\det \Sigma_u}} \exp \left[ -\frac{1}{2} \sum_{t=1}^T \left( Y_t - \sum_{i=1}^p Z_i Y_{t-i} \right)' \Sigma_u^{-1} \left( Y_t - \sum_{i=1}^p Z_i Y_{t-i} \right) \right] \quad (11)$$

El logaritmo de la ecuación anterior se expresa de la siguiente manera:

$$LL(Z_1, \dots, Z_p) = -\frac{KT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma_u| - \frac{1}{2} \sum_{t=1}^T u_t' \Sigma_u^{-1} u_t \quad (12)$$

Después, esta expresión se maximiza para obtener estimaciones de  $Z_0, \dots, Z_p$  y de  $\Sigma_u$ . Una característica de los modelos VAR es que los coeficientes ( $Z_i$ ) son difíciles de interpretar debido a la naturaleza multivariante. Por lo tanto, se desarrollaron las funciones de Impulso-Respuesta (IR) y de Descomposición de la Varianza del Error (FEVD, por sus siglas en inglés) para superar estas limitaciones [20]. La función de IR indica la reacción que tiene una variable ante un shock de otra variable; es decir, responde a ¿qué sucede en una variable si se aumenta en 1% el crecimiento de otra variable?, mientras que, la FEVD muestra el porcentaje de contribución de los shocks de las variables sobre una variable, responde a la pregunta ¿con qué porcentaje el shock de cada variable del modelo ayuda a explicar a otra variable en el largo plazo?. Después de la estimación, se utiliza la descomposición del Wold para estimar estas funciones. Con la ecuación 3, se puede suponer un modelo con un rezago:

$$Y_t = Z_0 + Z_1 Y_{t-1} + u_t \quad (13)$$

Si este proceso comienza en el tiempo  $T = 1$  y se sigue hasta  $T = t$ , se obtiene:

$$\begin{aligned} Y_1 &= Z_0 + Z_1 Y_0 + u_1, \\ Y_2 &= Z_0 + Z_1 Y_1 + u_2 = Z_0 + Z_1(Z_0 + Z_1 Y_0 + u_1) + u_2, \\ Y_2 &= (I_K + Z_1)Z_0 + Z_1^2 Y_0 + Z_1 u_1 + u_2 \\ &\vdots \\ Y_m &= (I_K + Z_1 + \dots + Z_1^{m-1})Z_0 + Z_1^m Y_0 + \sum_{i=0}^{m-1} Z_1^i u_{m-i} \\ &\vdots \\ Y_t &= \mu + \sum_{i=0}^{\infty} Z_1^i u_{t-i}, \quad t = 0, 1, 2, \dots, \end{aligned} \quad (14)$$

Donde  $\mu := (I_K - Z_1)^{-1}Z_0$ . Las distribuciones marginales y conjuntas de los  $Y_t$ 's son determinados únicamente por la distribución de los procesos  $\varepsilon_t$ . Aunque a veces, se asume que un proceso inicia en un periodo específico, también conviene aceptar que inicia en el pasado infinito ( $i = -\infty$ ). La representación de medias móviles es una herramienta especialmente utilizada para examinar la interrelación entre las variables. Los coeficientes de  $Z_i$  pueden utilizarse para generar los shocks  $\varepsilon_{1t}$  y  $\varepsilon_{2t}$  en toda la trayectoria del tiempo de  $x_{1t}$  y  $x_{2t}$  [12]. La combinación de los elementos  $z_{jk}$  de  $Z_i$  y  $u_t$  es conocida como la función de impulso-respuesta, los cuales son los multiplicadores de impacto. Además, la sumatoria ponderada de dichos shocks estiman la función de FEVD. Los intervalos de confianza se estiman mediante simulaciones de bootstrap con 200 repeticiones. [12] propone el siguiente algoritmo para dicho proceso:

1. Obtener  $Z_i$  y  $u_t = Y_t - \sum_{i=1}^p Z_i Y_{t-i}$ .
2. Realizar un muestreo (con o sin reemplazo) con  $R$  repeticiones y construir  $Y_t^r = \sum_{i=1}^p Z_i Y_{t-i}^r + u_t^r$ .
3. Estimar  $Z_i^r$  con los datos contruidos en 2) y calcular los IR para cada  $r$ .
4. Reportar los percentiles de la distribución los IR, para este caso se reporta el 5 % y 95 %.

## 4. Resultados

Para exponer el efecto de la política fiscal en los componentes del PIB, se utiliza las funciones de Impulso-Respuesta y FEVD. Estas se construyen después de la estimación de los parámetros del modelo VAR. Como antes se indicó, se utiliza 3 rezagos para dicha estimación. El horizonte de análisis es 20 trimestres (5 años) con intervalos de confianza del 95 % y 5 %, dados por 200 réplicas según el algoritmo de [12].

El impulso del consumo de gobierno (cgob) tiene efectos temporales y de mediano plazo (Gráfico 3). Este shock provoca una aumento de los impuestos indirectos en el mismo trimestre. Esto puede suceder porque el gobierno aumenta su gasto en bienes y servicios, lo cual podría conllevar el aumento de la recaudación del IVA e ICE. Además, este cambio tiene un efecto negativo en la inversión que tiene una duración de 1 trimestre. En los impuestos directos, el shock induce un aumento en la recaudación, la cual exhibe una duración

de 3 trimestres. Este efecto se puede explicar también por el aumento del consumo público y los mayores ingresos de las empresas en consecuencia. Este shock provoca, además, un aumento en el consumo de hogares, las exportaciones, las importaciones y el producto, aunque no son significativos.

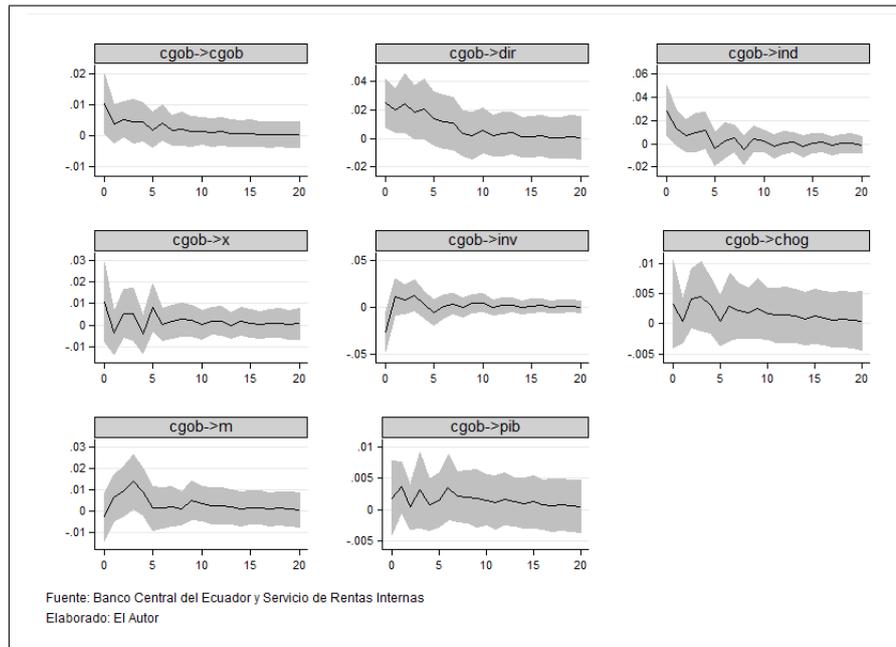


Figura 3: Función Impulso-Respuesta del Consumo de Gobierno. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

El crecimiento de los impuestos indirectos tiene un efecto negativo de 1 trimestre en las exportaciones (Gráfico 4), ya que los exportadores son el último nivel del encadenamiento del IVA, pues no pueden trasladar este impuesto. Este shock también tiene un efecto negativo sobre las importaciones, ya que encarece el precio del producto externo. Además, induce un efecto negativo en el PIB que dura 1 trimestre. En los impuestos directos y el consumo de hogares, el efecto es negativo en el primer y segundo trimestre respectivamente, aunque ninguno es significativo.

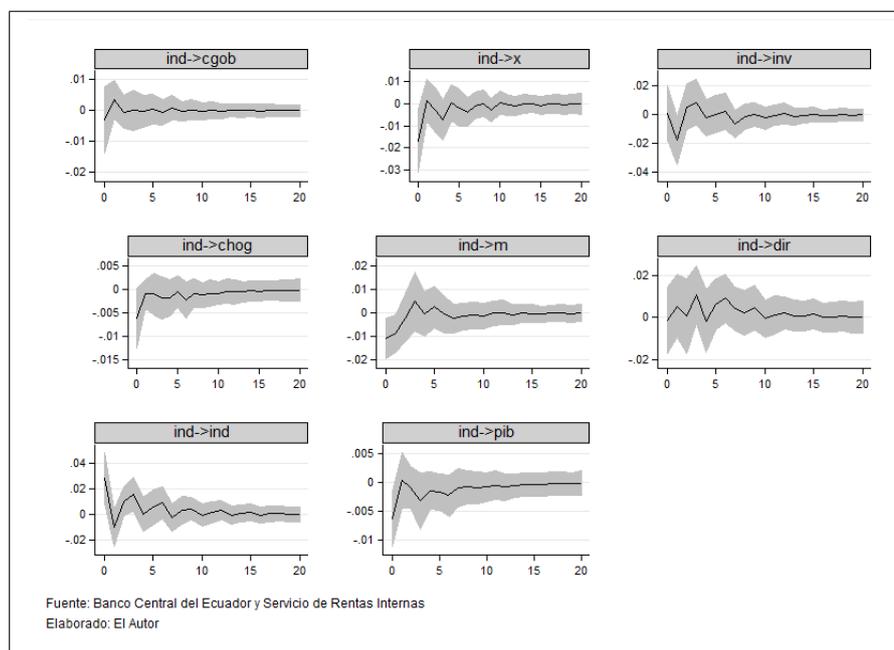


Figura 4: Función Impulso-Respuesta de los Impuestos Indirectos. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

El shock de los impuestos directos no tiene efectos significativos en la mayoría de variables (excepto en las exportaciones). Según [34], este impacto se puede dar porque no existe una adecuada recaudación de este tipo de impuestos. En efecto, [2] muestran que el Impuesto a la Renta<sup>8</sup> es el tributo con mayor evasión en el país: “la brecha de recaudación del impuesto a la renta de las empresas en términos globales es de 61 %”. El único efecto que este aumento de impuestos directos provoca es el aumento de las exportaciones porque, probablemente, los exportadores reciben mayores beneficios en el mercado externo que en el interno.

<sup>8</sup>El Impuesto a la Renta representa el 99 % de la recaudación directa

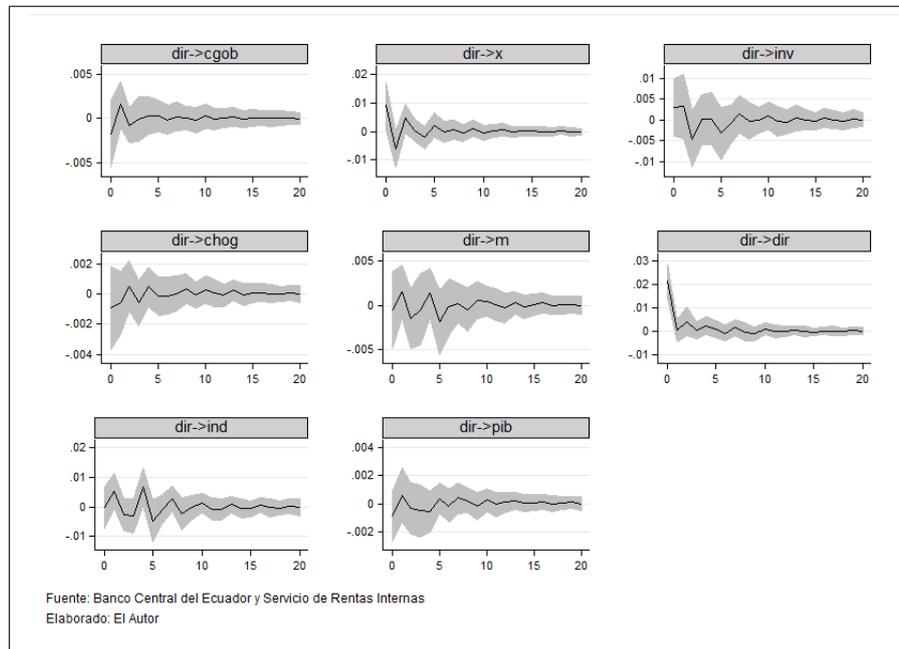


Figura 5: Función Impulso-Respuesta de los Impuestos Directos. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

En comparación con los países de la región, los resultados son consistentes con [13], [37], [26], [29]. Estos autores también indican que el aumento de los impuestos tiene efectos negativos sobre la producción, y son temporales los de los impuestos indirectos. Mientras que, los impuestos directos tiene efectos diferentes a los países analizados debido a la alta evasión de estos tributos en Ecuador.

Además, se muestra cómo impacta el shock de todas las variables en los impuestos y el consumo de gobierno. Los agregados que afectan a los impuestos indirectos son el consumo de gobierno y las exportaciones. El primero provoca un aumento temporal y el segundo tiene un shock positivo temporal en el cuarto trimestre. Las demás variables no presentan efecto alguno ya que se mantienen alrededor de cero y no tienen significancia.

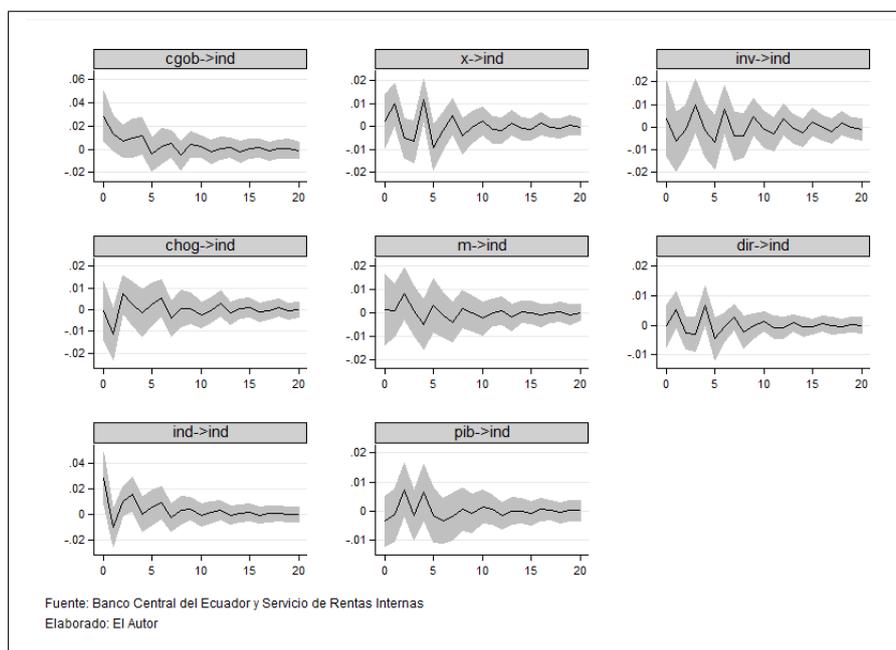


Figura 6: Funciones Impulso-Respuesta en los Impuestos Indirectos. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

En los impuestos directos, los agregados macroeconómicos que tienen efecto son el consumo de gobierno, las exportaciones, las importaciones y el PIB. Las importaciones y el producto provocan un efecto temporal en el segundo y quinto trimestre, respectivamente, aunque el shock de importaciones tiene su intervalo de confianza cerca de cero. El consumo de gobierno tiene un efecto positivo de mediano plazo que disminuye período a período. Las exportaciones tienen un efecto negativo al instante y positivo en el siguiente trimestre.

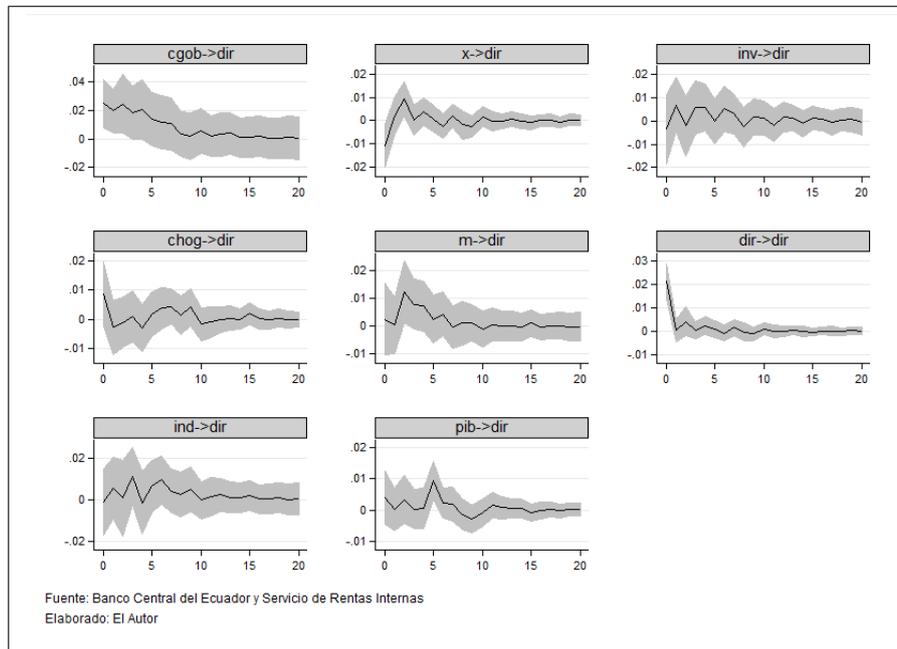


Figura 7: Funciones Impulso-Respuesta en los Impuestos Directos. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

Los agregados que tienen un efecto negativo sobre el crecimiento del consumo de gobierno son la inversión, el consumo de hogares y las importaciones. Un shock positivo de la inversión tiene provoca una disminución del consumo ya que el gobierno proporciona el 60 % de la inversión total [6]. El crecimiento del consumo de las familias y las importaciones tienen un efecto temporal ya que los hogares consumen más bienes y servicios importados desplazando al consumo de gobierno. Además, el aumento de los impuestos no tiene un efecto significativo sobre el gasto público, indicando que aunque exista mayor recaudación el gobierno mantendrá su gasto corriente de acuerdo a lo presupuestado. Los shocks del PIB y las exportaciones tampoco tienen un efecto significativo en esta variable.

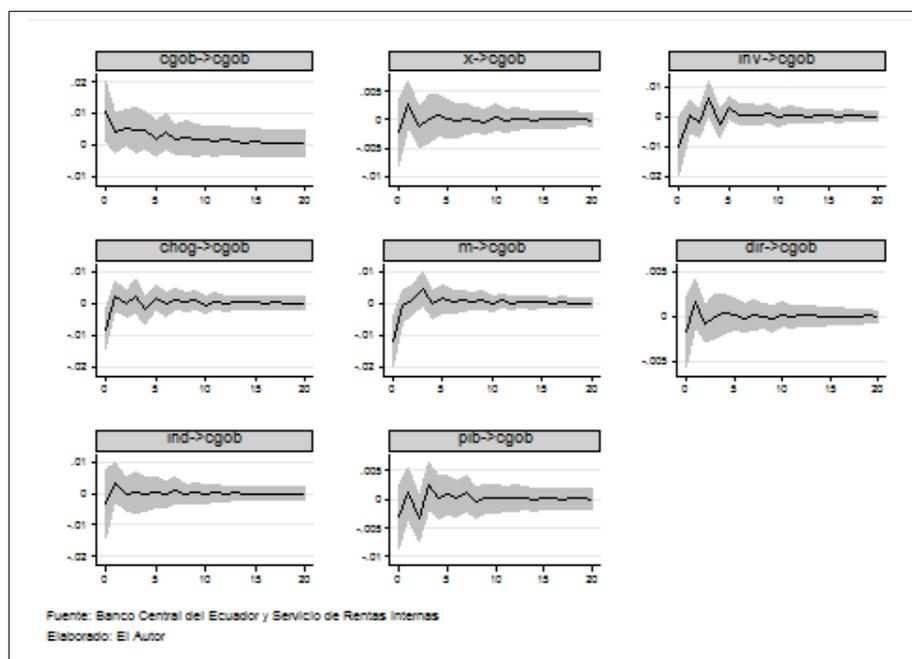


Figura 8: Funciones Impulso-Respuesta en el Consumo de Gobierno. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

Para examinar más sobre las interrelaciones que tiene las variables se utiliza la función de descomposición de la varianza. Esta herramienta además permite identificar la exogeneidad de los shocks en el modelo. Los resultados se muestran en grupo de variables: cuentas fiscales, ingreso disponible, comercio exterior y PIB.

Las variables que están relacionadas con el consumo de gobierno son los impuestos indirectos, el mismo consumo público, las exportaciones y la inversión (gráfico 9). La relación con la imposición indirecta se da porque estos tributos son los mayores aportantes a los ingresos gubernamentales [6]. La relación que tiene el consumo público con las exportaciones podría responder a que el Estado exporta petróleo, otra de las fuentes importantes de los ingresos en el presupuesto. Finalmente, el gasto gubernamental tiene una interrelación con la inversión total ya que el Estado proporciona el 60% de ésta [6].

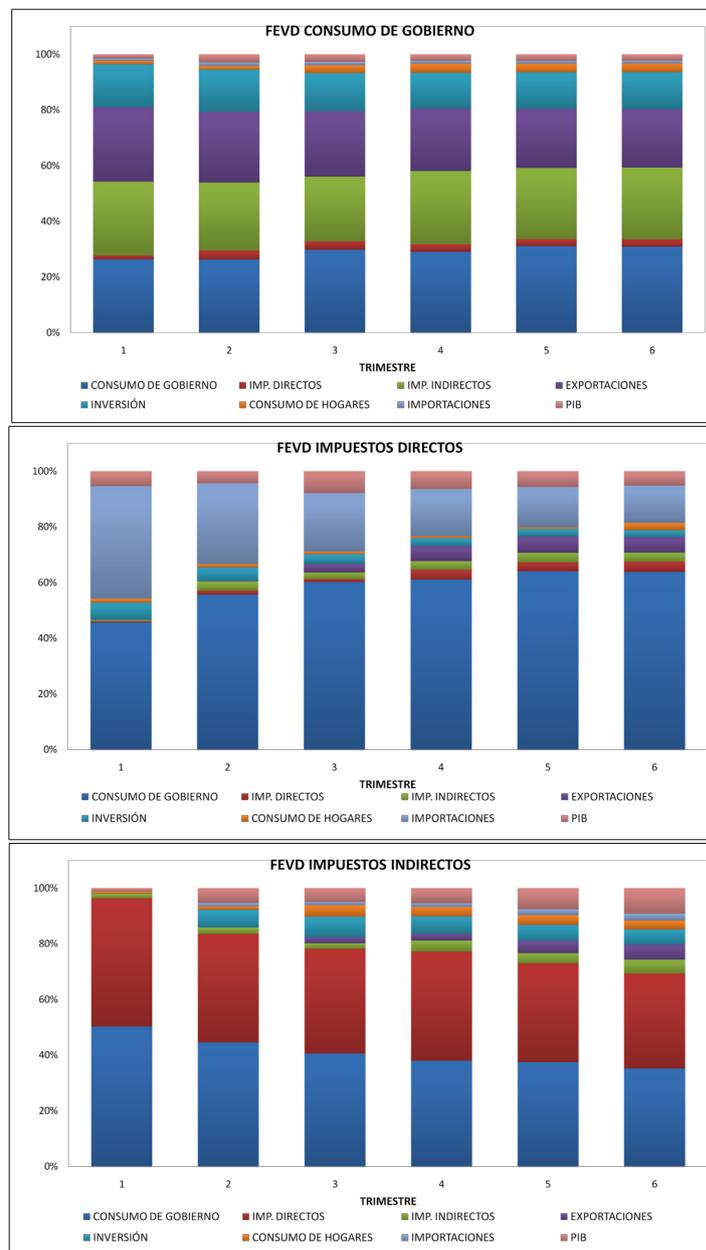


Figura 9: FEVD de las Variables del Gobierno. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

Los impuestos directos tienen una alta interrelación de largo plazo con el consumo de gobierno; éste tiene una participación de 64% en la varianza. Además, las importaciones participan con un 14%. Por otro lado, los impuestos indirectos tienen una alta interrelación con el gasto público y los impuestos directos, estas dos variables aportan a la predicción con 36% y

35 %, respectivamente (gráfico 9).

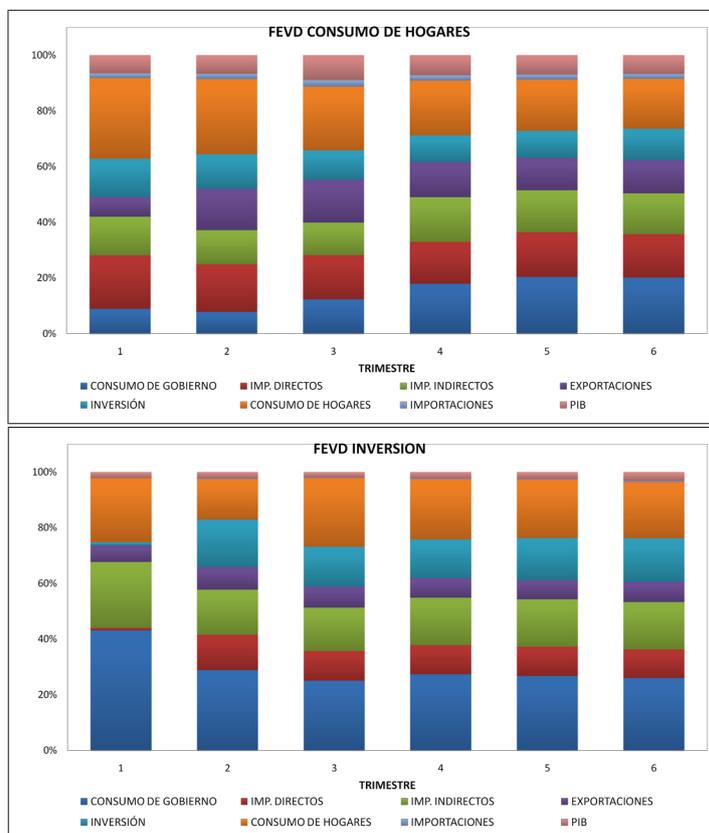


Figura 10: FEVD del Ingreso Disponible. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

Según [24], el ingreso disponible está dividido en consumo y ahorro siendo este último igual a la inversión en una economía. Por ello, se analiza en conjunto la descomposición de la varianza del consumo de hogares y la inversión (gráfico 10). Ambas variables tienen interrelaciones con el gasto público, los impuestos y ellas mismas. De manera individual, el consumo privado tiene interrelaciones con el consumo de gobierno (20 %), su propia evolución (18 %) y los impuestos directos e indirectos (16 % y 15 %). Por otro lado, la inversión total está relacionada con el gasto gubernamental (26 %), el consumo de los hogares (20 %), los impuestos indirectos (17 %) y su crecimiento (15 %).

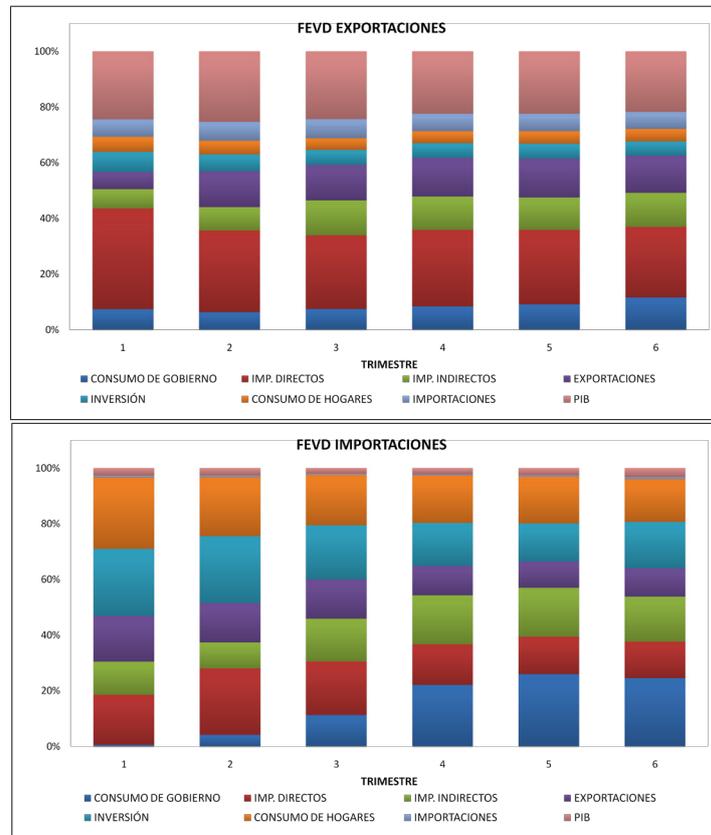


Figura 11: FEVD del Comercio Exterior. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

Las variables de comercio exterior exhiben comportamientos diferentes. Sin embargo, ambas tienen interrelaciones con los impuestos y el consumo de gobierno (gráfico 11). Por un lado, las exportaciones tienen una relación con los impuestos directos (25%), la producción total (22%), la propia evolución (13%), los impuestos indirectos (12%) y el consumo público (12%). Por otro lado, las importaciones están interrelacionadas con el gasto gubernamental (25%), el consumo de hogares (17%), la inversión total (16%) y los tributos indirectos (16%).

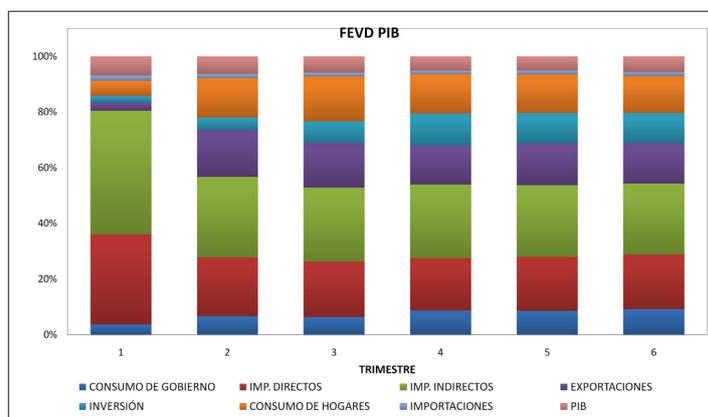


Figura 12: FEVD del PIB. Fuente: Banco Central del Ecuador y Servicio de Rentas Internas. Elaborado: El Autor.

Por último, más del 50 % del crecimiento del PIB está interrelacionado con los impuestos directos e indirectos. Las exportaciones y el consumo de hogares son los componentes que tienen la mayor relación el producto; cada variable explican el 13 % de la varianza. Las demás variables aportan el 22 % (gráfico 12).

## 5. Conclusiones

Este documento muestra el efecto de los impuestos y el consumo público en el PIB y sus componentes a través de un modelo de vectores autorregresivos estructurales (SVAR) de largo plazo. La investigación de [19], sobre los hechos estilizados de la economía ecuatoriana, es la base para imponer las restricciones del SVAR. Esta herramienta permite tener un acercamiento más realista a la estructura económica del país para que los hacedores de política económica puedan tomar decisiones más acertadas.

En este sentido, los principales resultados de la investigación son: *i*) los impuestos tienen una relación dinámica con los componentes del PIB. *ii*) los shocks de los impuestos indirectos afectan negativamente en las exportaciones y el PIB. *iii*) el crecimiento de los impuestos directos provoca un aumento en las exportaciones. *iv*) el shock del consumo de gobierno presenta efecto sobre los impuestos y la inversión. *v*) los agregados de comercio exterior tienen diferentes efectos sobre los impuestos; las importaciones muestran efectos positivos en la recaudación total y las exportaciones provocan efectos negativos sobre los tributos directos. *vi*) los instrumentos de política fiscal tienen una alta interrelación entre ellos. *vii*) el aumento del consumo público está estrechamente interrelacionado con el ingreso disponible que los impuestos. *viii*)

los impuestos tienen mayor relación con las exportaciones que con las importaciones; mientras que, el consumo gubernamental muestra una estrecha interrelación con las importaciones.

La generación de políticas económicas a partir de estos resultados no implica que tenga una fácil implementación y reacción, ya que el estudio es una evaluación económica de los instrumentos mas no social y política. Se debe tomar en cuenta todas las consideraciones planteadas a lo largo del estudio que limitan la modelación de la realidad, es decir, que representen de manera precisa el comportamiento de las variables observadas.

## Referencias

- [1] P. AGÉNOR, *The economics of adjustment and growth*, Harvard University Press, New York, second ed., sept 2004.
- [2] M. ANDINO AND J. C. PARRA, *Estimación de la brecha de recaudación del iva e impuesto a la renta de las sociedades por industria*, Revista Fiscalidad, (2007), pp. 101–139.
- [3] K. P. ARIN AND F. KORAY, *Are some taxes different than others? an empirical investigation of the effects of tax policy in canada*, Empirical Economics, 31 (2006), pp. 183–193.
- [4] H. BADINGER, *Fiscal shocks, output dynamics and macroeconomic stability: an empirical assessment for austria (1983 - 2002)*, Empirica, 33 (2006), pp. 267–284.
- [5] BANK OF ENGLAND, *Economic Models at the Bank of England*, Bank of England, 1999.
- [6] BCE, *Información estadística mensual*, Reporte Estadístico 1896, Feb 2010.
- [7] B. S. BERNANKE, *Alternative explanations of the money-income correlation*, Working Paper 1842, National Bureau of Economic Research, February 1986.
- [8] O. BLANCHARD AND R. PEROTTI, *An empirical characterization of the dynamic effects of changes in government spending and taxes on output*, Working Paper 7269, National Bureau of Economic Research, July 1999.
- [9] O. BLANCHARD AND D. QUAH, *The dynamic effects of aggregate demand and supply disturbance*, The American Economic Review, 79 (1989), pp. 655–673.
- [10] O. J. BLANCHARD AND M. W. WATSON, *Are business cycles all alike?*, Working Paper 1392, National Bureau of Economic Research, December 1987.
- [11] G. BOX, G. JENKINS, AND G. REINSEL, *Time Series Analysis: Forecasting and Control*, John Wiley, cuarta ed., 2008.
- [12] F. CANOVA, *Methods for Applied Macroeconomic Research*, Princeton University Press, 2007.

- [13] R. CERDA, L. F. LAGOS, AND H. GONZÁLEZ, *Efectos dinámicos de la política fiscal*, Cuadernos de Economía (Latin American Journal of Economics), 42 (2005), pp. 63–77.
- [14] F. DE CASTRO AND P. H. DE COS, *The economic effects of exogenous fiscal shocks in Spain: a svar approach*, Banco de España Working Papers 0604, Banco de España, Feb. 2006.
- [15] DEPARTAMENTO DE ESTUDIOS TRIBUTARIOS, *Índice de actividad empresarial no petrolera (iae-np)*, Nota tributaria 2010-05, Centro de Estudios Fiscales (CEF) - Servicio de Rentas Internas (SRI), Sept 2010.
- [16] M. DUNGEY AND R. FRY, *The identification of fiscal and monetary policy in a structural var*, CAMA Working Papers 2007-29, Australian National University, Centre for Applied Macroeconomic Analysis, Dec. 2007.
- [17] M. FLORES, *El filtro de baxter king, metodología y aplicaciones*, Documento de trabajo DIE-NT-01-2000, Banco Central de Costa Rica, Departamento de Investigaciones Económicas, Costa Rica, Marzo 2000.
- [18] J. FREEMAN, J. WILLIAMS, AND T. LIN, *Vector autoregression and the study of politics*, American Journal of Political Science, (1989), pp. 842–877.
- [19] I. GACHET, D. MALDONADO, N. OLIVA, AND J. RAMIREZ, *Hechos estilizados de la economía ecuatoriana: El ciclo económico 1965-2008*, Revista Fiscalidad, 6 (2013), pp. 59–122.
- [20] I. GACHET, D. MALDONADO, AND W. PÉREZ, *Determinantes de la inflación en una economía dolarizada: El caso ecuatoriano*, Cuestiones Económicas, 24 (2008), pp. 5–28.
- [21] F. HÖPPNER, *A var analysis of the effects of fiscal policy in Germany*, Institute for International Economics, University of Bonn, (2001), p. 29.
- [22] N. KALDOR, *Capital Accumulation and Economic Growth*, New York: St. Martin's, 1963.
- [23] J. KEATING, *Structural information in recursive var orderings*, Journal of Economic Dynamics and Control, 20 (1996), pp. 1557 – 1580.
- [24] J. M. KEYNES, *The general theory of employment, interest and money*, Macmillan for the Royal Economic Society, London :, 1935.

- [25] S. LARDIC AND V. MIGNON, *Econométrie des séries temporelles macroéconomiques et financières*, vol. 418, Economica, 1 ed., April 2002.
- [26] I. LOZANO AND K. RODRÍGUEZ, *Assessing the macroeconomic effects of fiscal policy in colombia*, (2009).
- [27] H. LÜTKEPOHL, *New introduction to multiple time series analysis*, Springer, Berlin [u.a.], 2005.
- [28] M. MENDONCA, J. CARDOSO, L. A. MEDRANO, AND A. SACHSIDA, *Evaluating the fiscal policy effects in brazil using agnostic identification*, (2008).
- [29] W. MENDOZA AND K. MELGAREJO, *La efectividad de la política fiscal en el Perú: 1980-2006*, consorcio de investigaciones económicas y sociales, Centro Internacional de Investigaciones para el Desarrollo (IDRC), 2007.
- [30] D. PACHECO, *Ecuador: Ciclo económico y política fiscal*, Cuestiones Económicas, 22 (2006), pp. 7–56.
- [31] A. PEREIRA AND O. ROCA-SAGALÉS, *Efectos macroeconómicos de las políticas fiscales en la ue*, Papeles de Trabajo 1/07, Instituto de Estudios Fiscales, Ministerio de Economía y Hacienda., España, 2007.
- [32] R. PEROTTI, *Estimating the effects of fiscal policy in oecd countries*, Tech. Report 168, Germany, 2002.
- [33] J. E. RESTREPO AND H. RINCÓN, *Identifying fiscal policy shocks in chile and colombia*, Working Papers Central Bank of Chile 370, Central Bank of Chile, Aug. 2006.
- [34] E. REZK, C. AVRAMOVICH, AND M. BASSO, *Dynamic effects of fiscal shocks upon diverse macroeconomic variables: A structural var analysis for argentina*, Annals of the Public Finance Workshop on Fiscal Indicators, (2006).
- [35] C. SIMS, *Macroeconomics and reality*, Econometrica, 48 (1980), pp. 1–48.
- [36] C. SIMS, *Are forecasting models usable for policy analysis?*, Quarterly Review, (1986), pp. 2–16.
- [37] I. TISCORDIO AND E. BUCACOS, *Efectos de la política fiscal en uruguay: una aproximación a través de shocks fiscales*, Documento de Trabajo 2-2008, Banco Central de Uruguay, Uruguay, Julio 2008.

# Analítica

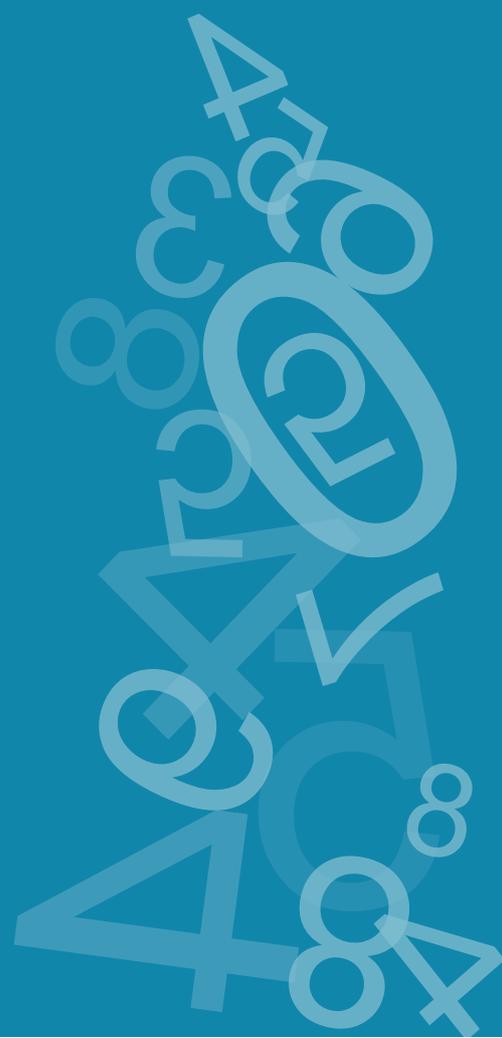
Localización de centros de empleo y su influencia  
sobre la distribución de la población en el Distrito  
Metropolitano de Quito

Diego Rojas

Coordinación de Investigación y Análisis. Instituto  
Nacional de Estadística y Censos



[www.ecuadorencifras.gob.ec](http://www.ecuadorencifras.gob.ec)





# Localización de centros de empleo y su influencia sobre la distribución de la población en el Distrito Metropolitano de Quito

Diego Rojas<sup>1</sup>

*Coordinación de Investigación y Análisis. Instituto Nacional de Estadística y Censos*

---

## Resumen

Este trabajo estudia la influencia de la actividad económica, representada por centros de empleo, sobre las preferencias de localización de la población. Para esto analiza los patrones de distribución de la población en 2001 y 2010. Además analiza los patrones de distribución del empleo e implementa una técnica de identificación de centros de empleo. Finalmente estudia la relación de los centros de empleo y la población a través de análisis descriptivos y econométricos. El trabajo encuentra 9 centros de empleo que tienen influencia sobre la distribución de la población, el comportamiento que predomina es el cambio de uso del suelo de habitacional a económico en la localización de los centros de empleo.

*Palabras clave:* población, centros de empleo, localización, identificación .

## Abstract

This document explores the influence of economic activity, represented by employment centers, over the population localization preferences. To achieve this I inquire about the population distribution patterns in 2001 and 2010. I also analyze the distribution patterns of employment and implement a technique to identify employment subcentres. Finally I study the relation between the employment subcentres and the population through descriptive statistics and econometric models. I found nine employment subcentres that wield some influence over the distribution of population, the dominant behavior is the change in the land use from housing to economic ends at the location of employment subcentres.

*Keywords:* population, location, employment subcentres, identification.

---

<sup>1</sup>diego\_rojas@inec.gob.ec

## 1 Introducción

La existencia de múltiples centros de empleo en una ciudad es el resultado de la interacción de muchos elementos en las decisiones de sus habitantes. El crecimiento de las ciudades, la evolución de la tecnología, el transporte, la actividad industrial y la actividad inmobiliaria han determinado que el espacio urbano cambie. Y el resultado de la evolución es que en las ciudades es un proceso de dispersión de la actividad económica y la población que contradice al fenómeno de concentración que sufren las ciudades en sus primeras etapas de crecimiento. La forma como estos centros de empleo definen la localización de las personas en las ciudades, no puede ser ajena del interés de una ciudad como el Distrito Metropolitano de Quito que registra un crecimiento importante tanto en términos poblacionales, como urbanos en las últimas décadas.

El intento de caracterizar la forma de la ciudad dio origen a lo que se conoce como el modelo de ciudad monocéntrica. Esta ciudad es una forma de organización del espacio que responde a un modelo de uso de la tierra propuesto por Von Thünen (1996). En su obra *El Estado Aislado*, propone un modelo que define una forma de toma de decisiones de localización de los productores agrícolas. Estos se distribuyen en el territorio en anillos concéntricos alrededor de un mercado principal ubicado en el centro de la ciudad.

El trabajo de Alonso (1964), que es una adaptación del modelo del uso del suelo agrícola de Von Thunen, pretende demostrar que las ciudades se ajustan a un prototipo monocéntrico; alrededor de lo que él llama el *Central Business District*, -CBD-; concepto que ocupa el lugar del mercado de la ciudad en una concepción más moderna de aquella de la ciudad tunecina del siglo XIX. Las principales características de este CBD serán: precios altos en el suelo y la mayor densificación del espacio tanto en población como en de la distribución de la actividad económica. Además determinará la existencia de una disposición espacial que se subordina a la localización presupuesta de este. Esta distribución del suelo es pareto eficiente (Fujita and Thisse (2002)) debido a la simplicidad del modelo que no incorpora externalidades sino que están basadas puramente en las decisiones de los agentes privados; esta característica desaparece cuando se incorpora en el análisis cuestiones como medio ambiente, congestión, amenidades<sup>2</sup> y otros factores (Anas et al. (1998)). El modelo de ciudad monocéntrica entonces es una buena representación de una ciudad pequeña (Redfean (2007)) que no sufre de las externalidades antes mencionadas.

La estructura policéntrica de las metrópolis es una hipótesis que se ha contrastado ampliamente desde la economía urbana (Cervero and Wu (1997); Anas et al. (1998); Fujita and Thisse (2002); McMillen (2003)), obteniendo resultados concluyentes: las grandes ciudades están sufriendo un proceso de policentralización. La discusión ahora gira alrededor de la interacción que existe entre los varios centros de empleo que se pueden encontrar en una ciudad y la distribución del empleo, la población, las amenidades, recursos financieros y otros. La

---

<sup>2</sup>Se llama amenidades a equipamientos de tipo cultural, social, entre otros; que presentan ventajas para las personas que se localizan más cerca de ellos. Ejemplo: Parques, cines, teatros, etc.

existencia de estos centros de empleo sería causado por una tensión en la relación de fuerzas de aglomeración y de dispersión, que resultan del proceso de producción y de los factores internos y externos para motivar su aglomeración (Anas et al. (1998)). Tanto la perspectiva de la Nueva Economía Urbana, que sigue la línea del modelo monocéntrico, como la teoría del lugar central<sup>3</sup> concuerdan en la existencia de este fenómeno, aunque proveen una explicación distinta. El modelo de ciudad policéntrica supone la existencia de una relación estable entre distintas características alrededor de los centros de empleo (Muñiz et al. (2008)). Esta población tomará sus decisiones de localización en función de su restricción presupuestaria y esto determinará entonces la aparición de una configuración social alrededor de este hecho. De la misma manera como lo haría con respecto al CBD en un modelo de ciudad monocéntrica; la diferencia radica en que la existencia de las múltiples centralidades suponen la existencia de efectos de aglomeración y dispersión que están vinculados a los procesos de decisión, óptima y pareto eficiente, de elección de una localización de los individuos.

Mientras el estudio de la dispersión y policentralización de las ciudades es un cuerpo amplio de estudios; por otra parte la forma de identificar la localización y la extensión de los centros de empleo es todavía un espacio en desarrollo temprano (Redfearn (2007)). (Redfearn 2007). La necesidad de entender como los distintos actores y fuerzas que convergen en la ciudad se comportan con respecto a la existencia de estos centros de empleo hace que el desarrollo de métodos de identificación de centros de empleo sea un objeto de estudio de interés para académicos y actores públicos. Los métodos de identificación de centros de empleo se pueden calificar en tres grandes ramas: los basados en análisis de clúster; los basados en técnicas paramétricas; y los basados en técnicas no paramétricas (Redfearn (2007)).

La evidencia que se muestra en las metrópolis del mundo desarrollado es contundente y apunta a procesos de dispersión de la actividad económica en el espacio urbano; las ciudades de otras regiones del mundo tienen un nivel de investigación mucho menor en el tema. En el caso latinoamericano la noción de policentralismo está presente entre el discurso de los estudiosos. Sin embargo los esfuerzos que existen para implementar métodos robustos para la identificación de centros de empleo son pocos. Dentro de lo más significativo en la línea de estos esfuerzos se encuentra los trabajos de Rojas et al. (2009), usa el método de splines cúbicos para el área metropolitana de concepción siguiendo a Muñiz et al. (2003). Avedaño (2012) utiliza también la metodología desarrollada en la Universidad autónoma de Barcelona (Muñiz et al. (2003)) para el caso de Bogotá. Para el caso de Quito existe un trabajo de identificación de centros de empleo (Mancheno and Rojas (2013)) que sigue un método no paramétrico basado en ajuste de superficies de la densidad de empleo (McMillen (2001)). Existen trabajos en términos de la descripción de la policentralidad de las ciudades

<sup>3</sup>La explicación de esta teoría es la integración de centros poblados debido a la ampliación del mercado, que supone la especialización de los centros poblados en un sistema de ciudades. Este trabajo no desarrolla la perspectiva de la teoría del lugar central de la policentralidad debido a que tiene un enfoque de corte de la NEU, para trabajos al respecto de esta ver: Clark and Kuijpers-Linde (1994); Champion (2001)

latinoamericanas, pero que usan técnicas y métodos de orden cualitativo, histórico u otros para determinar esta cualidad (Vallejo (2008); Link (2008); Segre (2008); Schelotto (2008)).

Esta mínima presencia de estudios cuantitativos que den cuenta de la validez de la hipótesis de la policentralidad en las metrópolis latinoamericanas supone al menos tres falencias en el estudio de la forma de la ciudad. El primero es la ausencia de evidencia que muestre que este fenómeno está presente en toda la región y en qué medida. El segundo aspecto es que si bien puede existir esta realidad es necesario entender cuál es el mejor método para identificar la presencia de estas centralidades en el contexto latinoamericano, pues la mayor parte de las técnicas desarrolladas son probadas en escenarios con condiciones muy diferentes a las presentes en el contexto de la región. Por último, es necesario entender cómo la población de las ciudades de Latinoamérica se ve afectada por la existencia de múltiples centros de empleo en las urbes. Este trabajo pretende aportar a estas tres falencias al implementar varias técnicas de identificación de centros de empleo en el sector urbano del Distrito Metropolitano de Quito y evaluar la influencia de estos sobre la distribución espacial de la población de esta ciudad.

La sección dos analiza, a través mapas y técnicas estadísticas, la distribución espacial de la población en los años 2001 y 2010 para encontrar cambios importantes en los patrones de distribución. A través de análisis de correlación espacial y ajustes de superficies a través de métodos no paramétricos se encuentra una población que se dispersa desde el Central Business District de la ciudad hacia las zonas rurales conurbadas y los extremos de la ciudad. La sección tres realiza un análisis similar para la distribución espacial del empleo. El ajuste no paramétrico de superficies permite confirmar la existencia de un CBD que se extiende desde el Centro Histórico, hasta el distrito comercial y financiero cerca al parque la Carolina. Para encontrar picos menores de empleo se utiliza la transformación logarítmica de la densidad de empleo; sitios de importancia son identificados fuera del CBD. La sección cuatro implementa el método de identificación de centros de empleo usando varios escenarios de ajuste no paramétrico de superficies para encontrar el la especificación de parámetros de ajuste que mejor concuerden con la realidad de la ciudad. El método no paramétrico (McMillen (2001)) identifica 9 centros de empleo que se ajustan a la realidad; pero no provee una extensión de cada uno. Los resultados encontrados son consistentes con la realidad.

La sección cinco hace un análisis descriptivo de la localización de los centros de empleo y la distribución de la población. A través de indicadores de correlación espacial bivariada encuentra que los centros de empleo se sitúan en espacios donde la densidad de población tiende a ser alta. Sin embargo no provee de resultados concluyentes de la influencia de los centros de empleo sobre la población.

La sección seis implementa dos modelo de influencia de los centros de empleo sobre la población en la línea de Muñoz et al. (2008). Estos modelos se desprenden de la función inversa exponencial pero incorpora variables de distancia de las unidades territoriales a los centros de empleo identificados a través del método de McMillen (2001); encontrando el modelo 2 que la mayor parte de centros de empleo tienen una influencia importante sobre la

distribución de la población en el DMQ.

## 2 Patrones de distribución de la población en Quito

Esta sección intenta ilustrar a través de mapas y técnicas espaciales estadísticas las lógicas de comportamiento de la población en el DMQ, específicamente en un espacio urbano. En primer lugar se hará una descripción de los datos usados para el análisis y se hará el análisis de los mapas de densidad poblacional y su evolución. Por último se hará un análisis de clústers a través mapas de correlación estadística con el método de la I local de Moran.

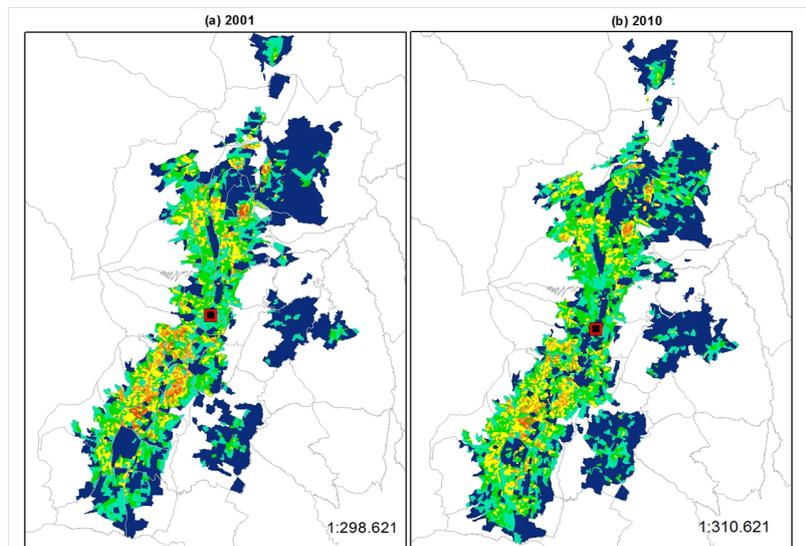
### 2.1 La población en Quito: un análisis exploratorio espacial

Las fuentes de la que provienen los datos son los censos de Población y Vivienda 2001 y 2010. Se utiliza los mapas de la recolección de cada uno de estos. El área que es objeto de estudio está conformada por sectores seleccionados que corresponden a los sectores censales amanzanados correspondientes a 32 parroquias urbanas y 9 parroquias rurales<sup>4</sup> escogidas por su nivel de conurbación y proximidad<sup>5</sup>. El tamaño mínimo por sector censal es de 0,32 hectáreas, el máximo es de 191,24 hectáreas, con un promedio de 6,24 hectáreas y un total de superficie de 30601,99 hectáreas; la densidad promedio en estos sectores es de 132,92 habitantes por hectárea. Para 2001 existe un total de 3386 sectores censales, con un tamaño mínimo de 0,54 ha, máximo de 174,7 ha, un área total de 27654,1 ha y un tamaño promedio de 8,17 ha; existe una densidad promedio por sector censal de 141,9 habitantes por ha.

La densidad poblacional se escoge para estudiar la población pues permite ilustrar una dotación de espacio por habitante y hace por tanto comparables a las distintas unidades geográficas. Es además, la habitual variable para el estudio de la población en trabajos previamente mencionados. Este análisis sin embargo resulta insuficiente para entender los cambios en la densidad espacial de la población en Quito en el tiempo. Las escalas en las que están agrupados los sectores censales por niveles de densidad en cada año serían disímiles; además el crecimiento y evolución propios de la ciudad en el tiempo plantea la dificultad de que Quito muestre dos densidades promedio y generales distintas en los dos años de análisis. Para resolver esto Millward (2008) propone utilizar lo que él denomina cocientes de localización, estos no son más que la representación de la densidad de cada sector censal como porcentaje de la densidad total para el área de estudio en la misma fecha. Este ejercicio se replicó para Quito y se obtuvieron los resultados expresados en el Figura 1. El panel (a) muestra los resultados 2001 y el panel (b) los de 2010.

<sup>4</sup>Calderón, Conocoto, Cumbayá, Llano Chico, Nayón, Pomasqui, San Antonio, Tumbaco y Zámbriza

<sup>5</sup>Se escoge esta área de estudio pues es los análisis propuestos por la metodología suponen una superficie casi continua; además el concepto de centro de empleo es por sí mismo un concepto urbano.



**Figura 1:** Densidad por sector censal como porcentaje de la densidad general para 2001 y 2010  
**Fuente:** INEC (2001,2010). Censo de Población y Vivienda.

La densidad general en 2001 para el área de estudio alcanzó 58.1 habitantes hectárea y en 2010 alcanzó los 64.1 habitantes hectárea. Además ilustra un fenómeno de densificación general en un panorama de aumentos y disminuciones localizadas de la densidad.

Los resultados muestran varias peculiaridades, entre las más importantes se encuentra la característica de un Central Business District cuya densidad poblacional es muy inferior a la de las áreas que le rodean; podemos observar además una disminución de la densidad en los sectores censales correspondientes a este entre 2001 y 2010. Lo que da cuenta de un proceso de relocalización de la población de esta parte de la ciudad, posiblemente atribuible a los usos económicos que va teniendo el suelo en dicho sector de la ciudad.

Otra importante característica que se debe resaltar es la prominentemente superior densidad poblacional en la parte sur de la ciudad en el 2001, particularmente en la parroquia de Solanda. Esta parte de la ciudad muestra niveles de densidad en el rango más alto de la escala. Su evolución hacia 2010, sin embargo, muestra una leve disminución y se observa a la par una densificación importante hacia el extremo sur de la ciudad. Lo que da cuenta de un proceso de relocalización de la población en el espacio desde el Centro-Sur hacia el extremo.

En cuanto a las zonas en que se muestra aumentos en los cocientes de localización, el extremo norte de la ciudad muestra un crecimiento notable, en específico: Calderón, Carcelén, el Condado y Pusuquí. De igual forma el extremo sur muestra un aumento importante en su densidad relativa en la zona de Quitumbe y sus alrededores. Existe un crecimiento importante en la densidad relativa de áreas de los valles, ubicados al este de la ciudad, de Cumbayá, Tumbaco, Conocoto y Nayón muestra un alza en su densidad que da testimonio de un proceso acelerado de urbanización por parte de clases acomodadas hacia estas zonas.

## 2.2 Autocorrelación espacial y análisis LISA

El análisis de la densidad relativa a través de los cocientes de localización es una técnica que ilustran los patrones de distribución de la población. Justamente por ser descriptivos estos métodos utilizados carecen de un criterio que dé orientación sobre el nivel de aleatoriedad o, en su defecto, de un orden de los patrones de localización de la población; además sobre la significancia de estos patrones Millward (2008).

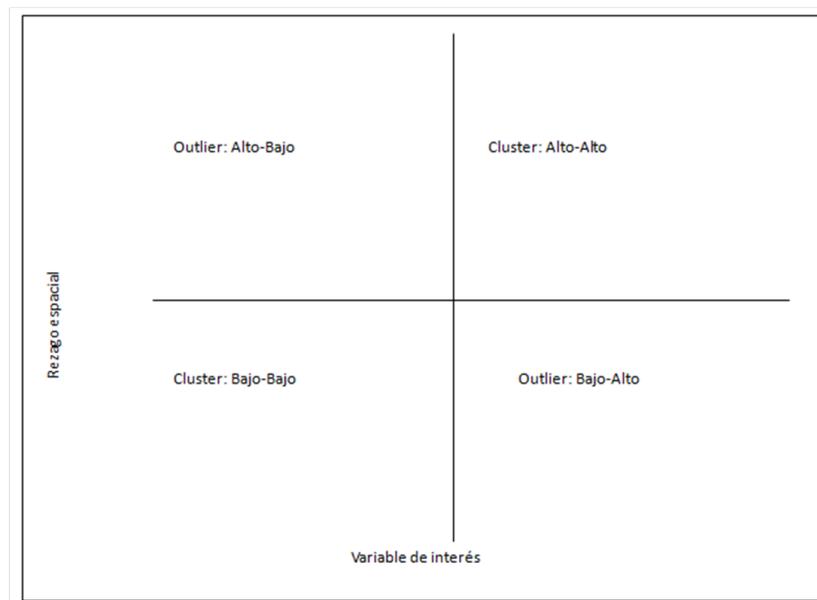
El análisis de la distribución espacial de una variable debe tener en cuenta la formación de patrones que a través de relaciones de proximidad entre lugares pueden influenciar de manera no aleatoria otras características del mismo lugar. Para entender esto se puede decir que las relaciones espaciales entre dos lugares pueden determinar características similares entre estos. Un ejemplo que se puede mencionar es la influencia que tiene la proximidad de dos espacios en la diseminación de enfermedades o patrones socio-culturales. Es decir si una ciudad X tiene una relación de proximidad mas fuerte con la ciudad Y que con la ciudad Z las características de X y Y deberían ser similares en un mayor grado que entre X y Z.

Siguiendo a Millward (2008), este trabajo selecciona el análisis de la I de Moran, tanto general como local. El método general permite obtener un índice cuyos valores toman un rango entre -1 y 1. Cuando el valor del índice es igual a 0 no existe autocorrelación espacial y la distribución espacial de la variable en cuestión es un fenómeno aleatorio; -1 y 1 representan los valores de perfecta correlación negativa y positiva respectivamente. Además este índice permite analizar el nivel de significancia del resultado.

Esta prueba, que se utilizará de manera transversal en este trabajo, se calcula como el ratio del producto de la variable de interés y su rezago espacial con el producto cruzado de la variable de interés y ajustado a través de pesos espaciales (Bivand et al. (2008)). Esta es una extensión espacial del coeficiente de correlación espacial de Pearson y se calcula como:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

Donde  $y_i$  es la observación  $i$ ,  $\bar{y}$  es la media de la variable de interés y  $w_{ij}$  es el peso espacial que liga al  $i$  con  $j$ . Lo que este coeficiente de correlación asume es que el modelo ideal tiene una media constante y que cualquier patrón que sobre después del centrado debe ser un efecto de la relación espacial expresada en  $w_{ij}$ . Este coeficiente puede calcularse para dos variables sustituyendo el valor de otra variable rezagada espacialmente, esto da origen a la I de Moran bivariada.



**Figura 2:** Gráfico de dispersión de Moran (Moran Scatterplot)

La utilización de índices globales sin embargo determina una pérdida en la identificación de patrones locales que son ignorados por este tipo de indicadores. Para solucionar este problema Anselin (1995) introduce los indicadores locales de asociación espacial o LISA por sus siglas en inglés. Los define como un indicador que da indicios sobre el grado de aglomeración espacial entre la observación y sus vecinos; además la suma de todos los LISA debe ser proporcional al indicador global. La fórmula de este indicador puede expresarse como sigue:

$$I_i = (y_i - \bar{y}) \sum_{j=1}^J w_{ij}(y_j - \bar{y}) \quad (2)$$

El análisis que se desprende de este indicador es la conformación de clusters con valores Altos y con valores Bajos. Además permite la identificación de outliers que muestran la asociación de valores disímiles en las distintas combinaciones que se pueden realizar. Esta clasificación se realiza alrededor del gráfico de dispersión de Moran o *Moran Scatterplot*, ilustrado en la Figura 2. Los valores se clasifican según en los cuadrantes en los que se ubique según la clasificación del análisis LISA. El valor del indicador global puede observarse en la pendiente de una regresión que se ajusta a los datos del gráfico de dispersión.

Los clusters son conjuntos, en este caso, de sectores censales que tienen valores estadísticamente similares, y que por su situación de proximidad presentan autocorrelación espacial positiva. Cuando en estos clusters, un sector censal que está por encima de la media está rodeado por sectores censales que están por encima de la media la categoría LISA para este

cluster es Alto-Alto; si, en el caso contrario el cluster tienen sectores censales que se encuentran por debajo de la media este será de la categoría LISA Bajo-Bajo.

En el caso de los outliers, estos son conjuntos de sectores censales que presentan valores estadísticamente disímiles y por tanto muestran autocorrelación negativa; si un sector, bien sobre la media, se encuentra rodeado por sectores que se encuentran bajo la media o viceversa las categorías LISA que les corresponden son Alto-Bajo y Bajo-Alto respectivamente. Este análisis también identifica aquellos sectores censales cuyas relaciones con los otros no son estadísticamente significativas. Para realizar este análisis se utilizó en este trabajo el software especializado en econometría espacial de software libre creado por Luc Anselin<sup>6</sup> llamado Geoda<sup>7</sup>.

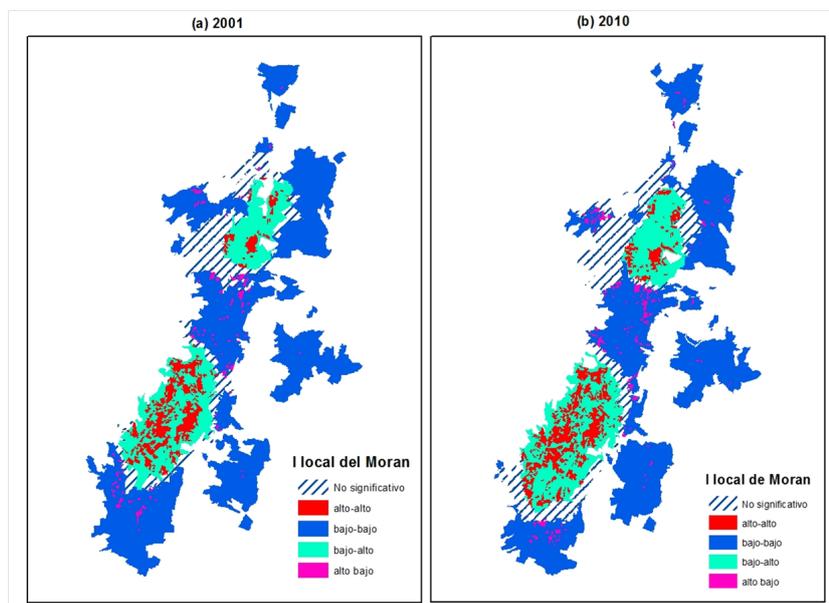
Para el cálculo de este índice se necesita la existencia de una matriz de pesos espaciales<sup>8</sup> que permita establecer cómo se relacionan las unidades territoriales entre sí. Para esto se experimentó con varios tipos de matrices espaciales: de continuidad tipo, torre o reina; o con los inversos de la distancia de cada vecino dentro de un radio determinado. Los resultados que muestran una mayor consistencia son aquellos que utilizan el peso de distancia inversa. Millward (2008) utiliza dicha forma de construcción de matriz de pesos espaciales en un análisis similar al de este trabajo para 5 ciudades canadienses mostrando resultados consistentes. Se escoge el método de la distancia inversa con un radio de 2km de interrelación entre sectores, pues en experimentos con radios mayores se pierden particularidades entre los espacios.

El análisis de la I de Moran muestra la existencia de correlación espacial positiva, esto confirmando lo que los análisis descriptivos muestran sobre la conformación de clusters poblacionales. El valor del indicador es de 0.28 y 0.24 en 2001 y 2010 respectivamente; ambos tienen una pseudo-probabilidad, calculada con 999 permutaciones, menor al 1% de que la distribución espacial sea un fenómeno aleatorio. Se puede entonces intuir que existe de hecho formaciones de clusters estadísticamente significativos en la ciudad en ambos periodos. Una vez que se conoce esta evidencia se procede a realizar el cálculo del análisis local cuyos resultados se plasman en un mapa para cada periodo, esto se encuentran en la Figura 3.

<sup>6</sup>Para ver una muestra de las capacidades de este programa ver Anselin et al. (2006)

<sup>7</sup>Disponible gratuitamente en <https://geodacenter.asu.edu>

<sup>8</sup>Ver Anselin (1998)



**Figura 3:** Análisis LISA de la densidad poblacional a nivel de sector censal en Quito, 2001-2010  
**Fuente:** INEC (2001,2010). Censo de Población y Vivienda.

Como se ve en la Figura 3 la mayor parte del territorio del DMQ muestra la formación de clusters, tanto en 2001 como en 2010. Los clúster más abundantes son aquellos de categoría Bajo-Bajo en ambos periodos. Predominan en el centro de la ciudad y hacia los extremos norte, sur y este. Una particularidad relevante es observar que en el CBD de la ciudad existen efectos de asociación en ambos periodos, sin embargo, muestran comportamientos de la densidad por debajo de la media; este clúster, como es evidente, se agranda entre 2001 y 2010. Este hecho da testimonio de una disminución de la densidad en esta área, estadísticamente significativa. Podemos observar que el área del tipo Bajo-Bajo hacia el extremo sur de la ciudad se reduce en el tiempo y es reemplazado por sectores tipo Alto-Alto y Bajo-Alto; por otra parte al noroeste de la ciudad existe una reducción del tipo Bajo-Bajo que es reemplazado por sectores no significativos.

Los clúster tipo Alto-Alto se encuentran principalmente en la parte sur de la ciudad y se observa que la extensión sobre la que están distribuidos aumenta en el tiempo; también una dispersión de estos. En la parte norte de la ciudad existen lugares bien definidos para la conformación de estos clusters, específicamente en las zonas del Comité del Pueblo, Carapungo, Cotocollao y Carcelén. La evolución de estas cuatro localidades en el tiempo parece ser de un estable y leve crecimiento.

El resumen de las áreas por categoría LISA se presenta en la Tabla 1. Se observa que en efecto los sectores de tipo Bajo-Bajo representan 59,3% del área de estudio en 2001 y presentan una reducción hasta el 51,5% en 2010. Mientras, el área del tipo Alto-Alto muestra un crecimiento de representar el 6,9% en 2001 al 8,7% en 2010. Por otra parte, se observa

que los outliers que predominan son los del tipo Bajo-Alto con un 18,5 % y 20,5 % en 2001 y 2010 respectivamente; los del tipo Alto-Bajo representan solamente un 1,6 % y un 2,1 % en 2001 y 2010. La superficie de tipo no significativa aumenta de 14,1 % al 17,1 %; la superficie no significativa es la que corresponde a conjuntos de sectores censales que tienen densidades muy diversas entre si y por ende presentan un patrón aleatorio de distribución.

**Tabla 1:** Superficie de Quito por categoría LISA, 2001-2010

Categoría LISA	2001		2010	
	Área	%	Área	%
No significativo	3906	14,10 %	5234	17,10 %
Alto-Alto	1919	6,90 %	2677	8,70 %
Bajo-Bajo	16397	59,30 %	15764	51,50 %
Bajo-Alto	4985	18,00 %	6281	20,50 %
Alto-Bajo	431	1,60 %	646	2,10 %
Total	27638	100,00 %	30602	100,00 %

**Fuente:** INEC (2001,2010). Censo de Población y Vivienda.

Estos resultados muestran principalmente tres grandes rasgos: i) el incremento del área de tipo no significativa, específicamente sobre áreas antes de tipo Bajo-Bajo, dan señales de una posible densificación progresiva y desordenada de estas, dando como resultando patrones de distribución de la densidad aleatorios; ii) la dispersión de los clusters en la parte sur muestra un posible proceso de redistribución de la población en esa área de la ciudad o de una migración intensa ; iii) el centro financiero y de negocios de la ciudad, así como el centro-norte de la ciudad son zonas que parecen estar correlacionadas sin embargo muestran comportamientos de baja densidad.

### 2.3 La densidad poblacional en Quito

Para entender la distribución de la densidad de la población se utilizan varias técnicas. Una de ellas es a través de métodos paramétricos usando la función inversa exponencial introducida por Clark (1951) y catalogada como un buen método para mapear esta relación (McDonald (1989); Joseph and Wang (2010)). Ésta, sin embargo, supone la imposición de una forma funcional a la densidad de población con respecto a la distancia que le separa del CBD de la ciudad y no da cuenta de las particularidades de la densidad dentro de la ciudad sino que estima un gradiente de decaimiento de la densidad de población a medida que se aleja la unidad territorial del CBD. Otros autores amplían el análisis al uso de técnicas no paramétricas, para mapear la relación entre la distancia y la densidad de población. Este tipo de análisis ha sido utilizado por varios autores para describir por ejemplo: la forma de la densidad poblacional Anderson (1982); la delimitación de ciudades satélites (Muñiz et al. (2003)).

Un método de ajuste como los antes descritos está limitado a entender la distancia como una variable en un espacio de dos dimensiones, por lo que los resultados se limitan a explicar la densidad en función de la cercanía sin importar en qué dirección se encuentra la unidad espacial que se está analizando; en síntesis anulando los efectos de sesgo direccional o sectorial<sup>9</sup>. Por este motivo se flexibilizará aun más la relación densidad-distancia, con una estimación en tres dimensiones, que permita analizar los cambios de la densidad con respecto a la distancia, pero en dos ejes. Esto permitirá analizar cómo evoluciona la densidad cuando la distancia ya no tiene magnitud sino también dirección.

El método que se escoge específicamente para mapear esta relación es el llamado *Locally Weighted Regression*<sup>10</sup> o más comúnmente conocido como LOESS. Este método estima una función de densidad utilizando regresiones localmente ponderadas para cada punto del dominio de los regresores<sup>11</sup> <sup>12</sup>. Para este trabajo se han utilizado varios span en función de observar cuales son aquellos que dan una mejor idea sobre el tema en cuestión: la densidad poblacional. El modelo que entonces se estimará será:

$$DEN_i = m(Dist_i) + \varepsilon \quad (3)$$

Para poder hacer la estimación en tres dimensiones se ha realizado una descomposición de la distancia en sus componentes rectangulares. En vez de realizar el cálculo de la distancia entre dos puntos como una línea recta, ahora se entiende a la distancia entre dos puntos como la distancia norte-sur y la distancia este-oeste:

$$DEN_i = m(N_i, E_i) + \varepsilon \quad (4)$$

Donde  $N_i$  y  $E_i$  son las distancias entre cada centroide de la cartografía censal con respecto al CBD. Lo que esto permitirá es realizar el estimado ya no de una curva sino de una superficie en el espacio. Para realizar este ajuste se utilizó el mismo método de regresión localmente ponderada que en la sección anterior; los resultados se expone en la Figura 4. La estimación se corrió con un span de 0.1 para guardar comparabilidad entre los datos de las curvas ajustadas y los de las superficies.

El resultado que se obtiene es un gráfico en tres dimensiones. La Figura 4 muestra la superficie, en los ejes de la caja se encuentran los valores de la distancias norte, este y la

<sup>9</sup>El sesgo direccional hace referencia a que una medición de distancia entre dos puntos debe tomar en cuenta la dirección en que esa distancia se mide. Esto pues en una ciudad o territorio real los fenómenos no suceden de manera homogénea en el espacio en todas direcciones.

<sup>10</sup>Para más detalles sobre el método de ajuste LOESS, véase Cleveland and Devlin (1988).

<sup>11</sup>El kernel utilizado en este trabajo es el tricúbico y se puede expresar como:  $K_i = \left(1 - \left(\frac{d_i}{d_{max}}\right)^3\right)^3 \text{iff } I(d_i < d_{max})$  donde  $d_i$  es la distancia del punto a ser estimado con respecto al punto  $i$ ,  $d_{max}$  es la distancia máxima determinada por la ventana escogida y  $I()$  es una función binaria que es igual a 1 cuando al condición se cumple o igual a cero si no.

<sup>12</sup>Los ensayos que se realizaron consideraron tres valores para el span: 0,1; 0,25; y 0,5. Tras el ejercicio se encontró que el span más apropiado para usarse es el de 10%.

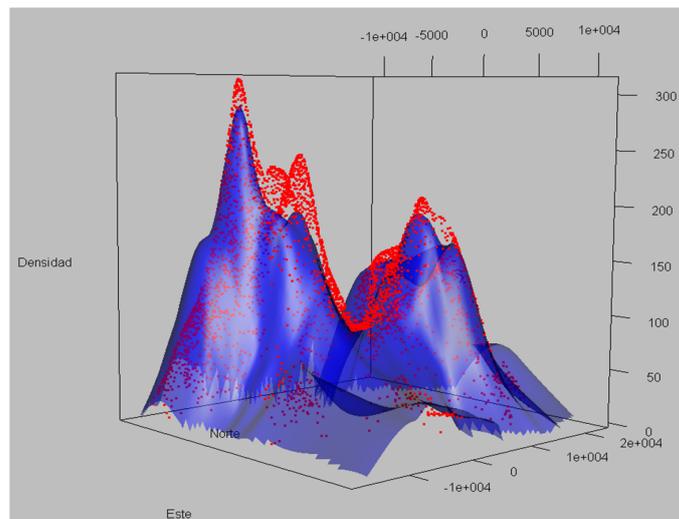
densidad estimada a través del método no paramétrico de ajuste. Cada punto rojo representa el valor de la densidad para un sector censal de 2001; mientras la superficie celeste representa la superficie estimada para 2010. Esta forma de representar resultó ser la más comprensible de todas las ensayadas. Como se observa los puntos rojos replican la forma del mapa de Quito.

La Figura 4 es la vista lateral de la superficie desde el este. Lo que muestra es el fenómeno ilustrado por la sección anterior pero en el espacio. Lo primero que es notorio es el cráter de densidad que se forma en el centro de la ciudad; a partir de este comienza una tendencia de crecimiento hacia las direcciones norte y sur donde son visibles varios picos de densidad. Esta estructura parece estable durante el periodo pues los cambios en la forma de la superficie entre 2001 y 2010<sup>13</sup> no presenta picos de densidad que aparezca de manera inesperada; si existe un cambio importante en la forma de la curva que concierne a un pico que se encuentra hacia el sur de la superficie y que parece desaparecer casi por completo en la superficie de 2010.

El fenómeno de densificación de las zonas más periféricas de la zona de estudio es evidente que en los extremos de la ciudad los puntos rojos de la superficie de 2001 se encuentran por debajo de la superficie de 2010. De igual forma las zonas de Tumbaco, Cumbayá y Conocoto muestran niveles estimados superiores en la actualidad. Los picos de densidad identificados hacia el sur son los sectores de la Villaflora y Solanda, que presentan bajas en sus densidades mientras el extremo sur correspondiente a las zonas de crecimiento urbano más reciente, relativamente, muestran crecimientos importantes. De igual manera hacia el norte de la ciudad el pico de densidad más importante parece bifurcarse formando dos picos distintos, lo que hace presumir que la reducción de la densidad es localizada a diferencia del sector sur.

---

<sup>13</sup>Es necesario aclarar que ambas predicciones son modelos que se corren de manera independiente. Que las formas de las superficies sean similares en el tiempo no es más que evidencia de que el método de ajuste es coherente y que produce resultados estables y comparables entre los periodos.



**Figura 4:** Vista de la superficie ajustada de la densidad poblacional, vista desde el este  
**Fuente:** INEC (2001,2010). Censo de población y vivienda

El uso de métodos no paramétricos de ajuste permitió entender básicamente tres cosas:

- i** la existencia de un cráter de densidad en el centro de la ciudad, es decir una densidad relativamente muy baja en el CBD, que se profundiza y expande su frontera en el tiempo debido a un probable desplazamiento del uso del suelo habitacional por uno económico;
- ii** la existencia de picos de densidad, máximos locales de densidad en la superficie observada, apostados alrededor del cráter, antes mencionado, que tienen la cualidad de mostrar una tendencia a la disminución de su densidad poblacional en el tiempo;
- iii** un aumento de la densidad para las localidades que están más allá de los picos antes mencionados, es decir un crecimiento de la densidad de las zonas rurales conurbadas y los extremos de la ciudad que ilustran un fenómeno de dispersión de la población desde los picos de densidad antes mencionados hasta éstas zonas

Estas cualidades conjuntamente con las mencionadas en los párrafos anteriores permiten describir al patrón de distribución de la población de Quito como uno que se asemeja en gran medida al que estudios empíricos han logrado identificar para ciudades estadounidenses<sup>14</sup>. Es decir un centro de la ciudad que pierde a sus habitantes para dar paso a un uso económico; que además mira un fenómeno de dispersión de su población al observarse procesos de crecimiento de zonas no tradicionalmente pobladas.

<sup>14</sup>El patrón típico de ciudad estadounidense muestra “un gradiente decreciente en general, la aparición de un cráter central de densidad, un desplazamiento hacia fuera del borde del cráter, una disminución en la densidad de picos, y un menor grado de ajuste (es decir, una mayor variabilidad) alrededor del gradiente de tendencia”. (Millward (2008))

### 3 Empleo

Esta sección divide el trabajo en tres secciones. La primera comprende un ejercicio de mapeo de la variable de densidad de empleo. La segunda es el uso de técnicas de análisis de correlación espacial para dar cuenta de la formación de sitios “calientes” de empleo. La tercera estima una función de densidad de empleo usando un análisis no paramétrico, similar al de la sección anterior.

#### 3.1 Un análisis exploratorio de la densidad de empleo

El análisis que se realiza en esta sección se encuentra en base a los datos del Censo Nacional Económico 2010. El censo económico anterior a este fue realizado en 1984. La información correspondiente a este censo, tanto la cartografía como la base de datos, no se encuentran digitalizados por lo que no existe la posibilidad de realizar un análisis intertemporal. Esto no representa sin embargo un obstáculo para los fines de esta investigación.

La desagregación de los datos es a sector censal, la cartografía es la misma utilizada para el análisis del censo de población y vivienda; se utilizará la misma área de estudio definida en la sección 1. El número de sectores censales que presentan registros de empleo en el área de estudio son 4520. Estos aglomeran 530495 empleos; la densidad de empleo promedio por hectárea es de 23,21 y su valor máximo es de 2138. Esta evidencia permite concluir que la distribución de la densidad de empleo tiene un severo sesgo hacia la derecha, es decir, la distribución muestra un número importante de sectores en los niveles más bajos de densidad. Esta peculiaridad obliga a preguntar si este fenómeno podría llevar a las técnicas de análisis propuestas a ignorar los pequeños detalles de la distribución espacial de la densidad por la naturaleza de esta distribución. La existencia de una macrocentralidad cuyo tamaño es superlativamente superior hace que por motivos de escala los comportamientos de otros lugares del territorio no sean fácilmente visibles. Estas pequeñas particularidades son las que dan evidencia sobre los procesos de localización del empleo en los territorios fuera del CBD. Por esto es necesario utilizar un artificio para poder visualizar de mejor manera las pequeñas particularidades; se transformará a la variable a logaritmo para que su distribución se asemeje más a la normal y esto permita visualizar de mejor manera la información en el espacio. El análisis se realizará sobre la densidad de empleo y sobre el logaritmo de la densidad de empleo.

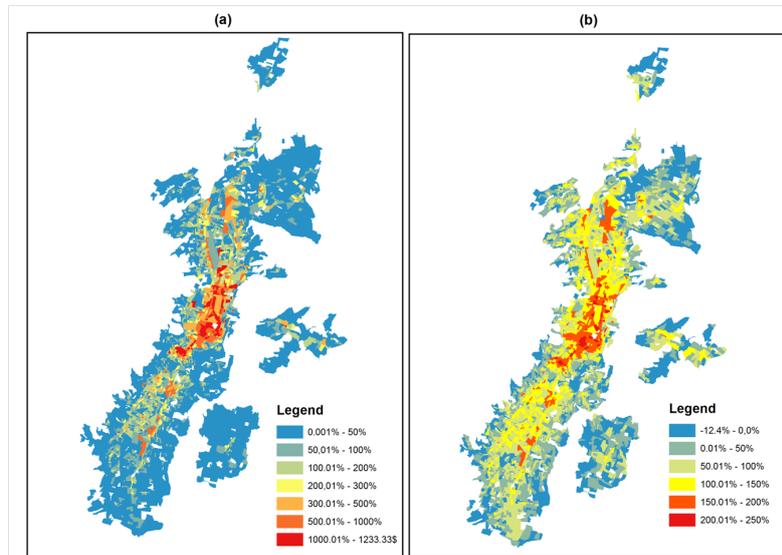
El mapa de la izquierda, en el panel (a), se tendrá siempre el mapa que contenga la información de la densidad de empleo; mientras el mapa de la derecha, en el panel (b), contendrá la información correspondiente al logaritmo de la densidad de empleo. La Figura 5 muestra el mapa de la densidad de empleo en el área de estudio, el Quito urbano; se utiliza los cocientes de localización para entender la distribución de la información en una escala que guarda relación con su propia distribución. El cociente de localización da cuenta de la relación entre la densidad registrada en la unidad territorial y la densidad de empleo de

toda el área de estudio; que es de 17,33 empleados por hectárea; expresada como porcentaje. Como se puede observar cuando no se aplica la transformación a la variable el mapa de densidad muestra una estructura ciertamente hipercéntrica que concentra en el CBD y en sus alrededores todos los sectores con valores importantes de densidad. A pesar de esto existen pequeñas alzas en la densidad de empleo hacia los extremos de la ciudad y las zonas rurales conurbadas que podrían dar cuenta de un proceso de policentralización de la ciudad desde una perspectiva económica.

Esta hiper centralidad se presenta en una zona de la ciudad donde existen tres lugares de importancia económica, que en otros estudios ha sido catalogada como la *macrocentralidad* de la ciudad Vallejo (2008). Estos lugares son: i) la zona circundante al parque La Carolina, consolidada como el centro financiero de la ciudad; ii) la Mariscal, que históricamente consolidado como un centro de empleo importante en la ciudad desde los 60 Vallejo (2008), cualidad que al parecer no ha perdido; iii) el Centro Histórico, que por su cualidad de concentrar las actividades administrativas dentro de sí ha visto proliferar a su alrededor una economía de comercio y servicios.

Se debe mencionar también la relativa menor importancia que se observa en las áreas rurales conurbadas, tales como Calderón, Tumbaco y Cumbayá, que muestran patrones de densidad de empleo que permiten advertir la posible presencia, dentro de ellos, de dinámicas económicas propias con una influencia local y tal vez incluso global sobre la ciudad. En otras palabras, en cada una de estas localidades se pueden advertir comportamientos de atracción de empleo que pueden tener una influencia sobre el área que les rodea o incluso influir sobre toda la distribución del empleo de la ciudad. Es decir, que existen puntos en el espacio, fuera de la macrocentralidad, que parecen tener cierta influencia sobre la concentración geográfica del empleo, al menos, sobre el área circundante. Esto da cuenta de la posible conformación de centros de empleo.

Una peculiaridad interesante es la que se observa en la formación de los patrones de densidad de empleo fuera del CBD. Estos se encuentran localizados a lo largo de la extensión de ejes viales. Hacia el norte se pueden apreciar las formaciones sobre la Avenida La Prensa al oeste; sobre la avenida Galo Plaza Lasso al centro; y sobre la avenida Eloy Alfaro hacia el extremo este; hacia el sur sobre dos ejes viales, la Avenida Pedro Maldonado y con menor intensidad sobre la avenida Mariscal Sucre.



**Figura 5:** Mapa de cocientes de localización de la densidad de empleo y su logaritmo  
**Fuente:** INEC (2010). Censo Nacional Económico.

El panel (b) permite rescatar las particularidades de las densidades en las zonas rurales conurbadas. En el panel (a) la importancia de estas zonas es relativamente baja por lo que se invisibiliza los cambios en las densidades dentro de ellas. Cumbayá y Tumbaco parecen revelar ciertos patrones de distribución que invitan a pensar en la posibilidad de encontrar centros de empleo en estas localidades; esto sucede también en Calderón y Llano Chico.

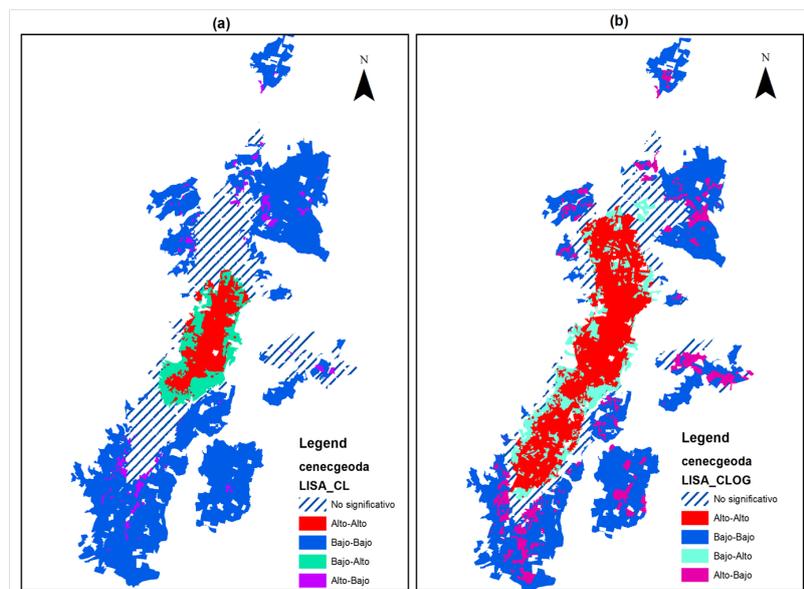
### 3.2 Análisis LISA de autocorrelación espacial de la densidad de empleo.

El análisis de clusters y outliers se usa para validar, en un sentido estadístico, las particularidades identificadas a través del análisis previos. Como ya se mencionó, este análisis intenta graficar los lugares donde los patrones de distribución muestren valores similares de tal manera que puedan clasificarse dentro de una de sus 4 categorías. El nivel de confianza de los resultados es del 95 % y las pseudo probabilidades fueron calculadas con un proceso de 999 permutaciones.

En este caso en particular los valores de la I global de Moran son de 0,17 para el análisis de la densidad de empleo y 0,30 para el caso del logaritmo de la densidad de empleo. Ambos resultados muestran un nivel de significancia mayor al 99 %. Esto da cuenta de la existencia de patrones de distribución no aleatorios, lo que hace presumir la existencia de clusters. Como se observa el estadístico es mayor para el caso del logaritmo de la densidad de empleo, es decir existe un proceso de autocorrelación espacial más acentuado, por lo que se asume que esta identificará más sectores dentro de la categoría de clusters.

La Figura 6 muestra en el panel (a) el análisis de clusters para la densidad de empleo y el panel (b) para el logaritmo de la densidad. Se puede ver que el resultado para la densidad de empleo genera un segmento de clusters tipo alto-alto en el CBD. Hacia los extremos de la ciudad se hallan clusters de tipo bajo-bajo, de igual manera sucede en las áreas, antes rurales o conurbadas. Este resultado era de esperarse por todo lo antes visto. Esta identificación de clúster sin embargo se encuentra influenciada por la distribución de la densidad de empleo.

La alternativa de utilizar el logaritmo de la densidad resulta de mucha utilidad en este caso. Los resultados de este escenario muestran que de hecho existen hacia el norte y el sur del CBD un número importante de sectores censales clasificados en la tipología alto-alto. La reducción de superficie categorizada como no significativa es importante. Además la aparición de clusters tipo alto-bajo en las zonas rurales conurbadas dan testimonio de procesos de aglomeración leve del empleo a nivel local. Volviendo a la situación de la parte interna de la ciudad, lo que se observa es una superficie cuasi continua cubierta por este tipo de sectores censales; se distingue una leve separación entre el CBD y la parte sur de la ciudad. Esta evidencia soporta, ya con validez estadística, las primeras suposiciones sobre la existencia de picos de densidad a lo largo de la ciudad; que pueden ser considerados como centros de empleo, salvando la enorme diferencia de tamaño entre estos y la macrocentralidad.

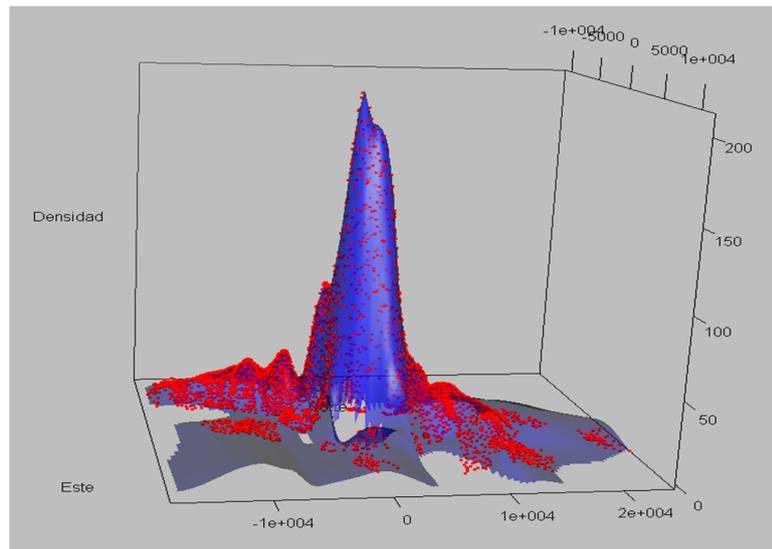


**Figura 6:** Análisis LISA de la densidad de empleo y la transformación logarítmica de la densidad del empleo  
**Fuente:** INEC (2010). Censo Nacional Económico.

### 3.3 Una estimación de densidad de empleo

Al igual que en la sección anterior se realiza la estimación de la densidad en tres dimensiones, teniendo en cuenta la posición exacta de cada sector censal, al igual que se realizó para la

densidad poblacional. Esto permite visualizar una superficie estimada de la densidad de empleo en Quito. De igual manera el análisis se realiza tanto para la densidad de empleo como para su logaritmo con el fin de mantener la tónica de la importancia de la escala en la comprensión de esta variable en particular; los regresores en ambos casos fueron la distancia en sentido norte-sur y la distancia en sentido este-oeste.

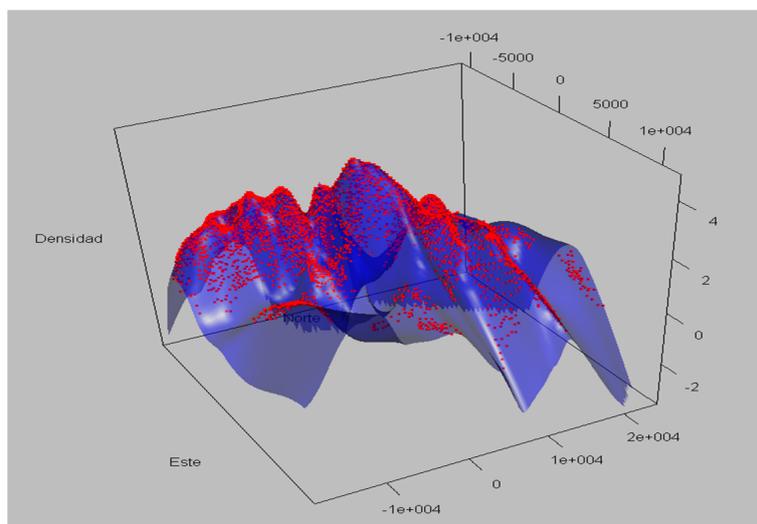


**Figura 7:** Curvas ajustadas por loess del logaritmo de la densidad de empleo, varios span  
**Fuente:** INEC (2010). Censo Nacional Económico.

Los resultados del análisis de la densidad de empleo se encuentran en la Figura 7, esta muestra la vista desde el nor-este de la densidad estimada para la ciudad con un span de 0,05. Se puede observar la existencia de una concentración de la densidad de empleo desproporcionada con respecto al resto de la ciudad en el área que se conoce como hiper centro. Este pico de densidad era el que dificultaba observar ciertas peculiaridades que ahora son notables y que se encuentran distribuidas a lo largo de la ciudad. Se logran distinguir al menos 6 picos en toda la superficie de la ciudad que presentan serios indicios sobre la existencia de centros de empleo en la ciudad, que puedan tener influencia local o general. Se divisa un pequeño pico dentro de la figura del pico más grande, este correspondería al Centro Histórico que a pesar de manifestarse, apenas, por fuera de la superficie no logra salir del área de influencia de este. Las conjeturas que se desprenden de la Figura 7 se reafirman cuando se hace el análisis con el logaritmo de la densidad de empleo; los resultados se encuentran en la Figura 8. La evidencia es más contundente en este caso los picos de empleo se distinguen con mayor facilidad y replican las posiciones de estos con respecto a la Figura 7.

Los picos más notables son seis y corresponden a las localidades de Tumbaco, Cum-bayá y Calderón en cuanto tiene que ver con las zonas rurales conurbadas, Conocoto no es distinguible; La Villaflora y Solanda en el sur de la ciudad; por ultimo al norte se encuentra

Cotocollao, existen otras alzas locales, pero que visualmente no se pueden calificar como máximos. Lo que estos sitios muestran potencial para ser considerados como centros de empleo. Lo que determina que se les califique de esta manera es la forma como estos determinan una lógica de distribución alrededor de ellos que se asemeja a los patrones de distribución visualizados en la primera sección a través de los análisis exploratorios.



**Figura 8:** Curvas ajustadas por loess del logaritmo de la densidad de empleo, varios span  
**Fuente:** INEC (2010). Censo Nacional Económico.

## 4 Identificación de Centros de empleo

Esta sección hará un análisis de identificación de los centros de empleo utilizando el método no paramétrico de McMillen (2001) que se ha explicado en la sección de fundamentos metodológicos y que ya ha sido aplicada para el DMQ (Mancheno and Rojas (2013)). (Mancheno y Rojas 2013). Es importante entender en qué medida se manifiesta la policentralidad en el DMQ. Un centro de empleo puede ser bien un viejo CBD que ha sido reemplazado; podría ser un pueblo aledaño a una ciudad que por características de conurbación y desarrollo económico ha adquirido esta cualidad, otros pueden ser territorios dentro de la ciudad que previamente no estaban conectados a la dinámica económica por motivos de transporte y que el desarrollo tecnológico ha logrado dinamizar. La definición de un centro de empleo puede variar según el enfoque y los fines del autor<sup>15</sup>; casi todas concuerdan en que un centro de empleo es un área cuyo empleo o densidad de empleo es relativamente distinta, mayor, con respecto a la de la ciudad o su vecindad. En lo que difieren es la forma como identificar

<sup>15</sup>Ver McDonald (1989), Cervero (1989), McMillen (2001), McMillen (2003), Redfearn (2007)Giuliano and Small (1991), entre otros.

esta área y como saber si esta alza en la densidad es significativa para la densidad de toda la ciudad.

La identificación de centros de empleo en Quito ya se ha realizado (Mancheno and Rojas (2013)) a través de una metodología no paramétrica (McMillen (2001)). Esta identificación determina la existencia de 8 lugares como centros de empleo en Quito; sin embargo, no agota todas las versiones del modelo de identificación, que la variación de ciertos parámetros permitiría. Se aplica la metodología no paramétrica (McMillen (2001)) utilizando varios escenarios para dar cuenta de la sensibilidad de los resultados obtenidos. Se realizará un ejercicio de prueba error con varios parámetros para probar la consistencia de los datos con la realidad. Tras este ejercicio se escogerán los sitios para subcentros que muestren la mayor solidez tanto de manera metodológica como de consistencia con el conocimiento de la ciudad.

#### 4.1 Identificación a través del método no paramétrico de McMillen

Este presenta la ventaja de tener una forma flexible de identificar los centros de empleo a través de métodos no paramétricos sin tener que contar con un stock enorme de conocimiento específico sobre el territorio a ser analizado y con un nivel menor de arbitrariedad que los otros métodos. Se utiliza un procedimiento de dos fases para poder identificar los subcentros de empleo. En la primera parte se estima una superficie del logaritmo de la densidad bruta<sup>16</sup> de empleo como una suerte de vitrina para mostrar los lugares candidatos a centros de empleo. La estimación se hace a través de un método no paramétrico de ajuste de superficies, que se utilizará en otros puntos de este trabajo, denominado regresión localmente ponderada<sup>17</sup>. Esta regresión estima una densidad de empleo para cada uno de los centroides de las unidades territoriales. Para realizar esta estimación es necesario especificar la forma como se ponderan los casos en el momento del ejecutar el procedimiento. En este trabajo se utiliza un *kernel tricubico*<sup>18</sup> similar al utilizado por McMillen (2001) que pondera las observaciones haciendo que las que estén más próximas del punto a ser estimado se les confiera una ponderación más alta.

Un factor que es definitorio en este tipo de métodos de ajuste de superficies es el tamaño de la ventana o span que se escoge para la estimación. Este valor representa el porcentaje de

<sup>16</sup>Se utiliza la densidad bruta de empleo pues como se expone en McDonald (1989):  $E/L = (E/L_e)/(L_e/L)$  donde  $E$  es el empleo total en un sector,  $L$  es el total de la extensión del sector,  $L_e$  es la extensión dedicada exclusivamente a esta actividad dentro del sector. Si el valor de  $L_e$  es bajo puede llevar a alzas en el  $E/L_e$  (Densidad Neta) que, sin embargo, podrían no tener efecto sobre las zonas circundantes. Desde esta perspectiva, la densidad neta de empleo permite evitar esa posibilidad.

<sup>17</sup>Para más detalles sobre la Locally Weighted Regression (LOESS), Cleveland and Devlin (1988)

<sup>18</sup>Un kernel tricubico se representa así:  $K_i = \left(1 - \left(\frac{d_i}{d_{max}}\right)^3\right)^3$  *iff*  $I(d_i < d_{max})$  donde  $d_i$  es la distancia del punto a ser estimado con respecto al punto  $i$ ,  $d_{max}$  es la distancia máxima determinada por la ventana escogida y  $I()$  es una función binaria que es igual a 1 cuando la condición se cumple o igual a cero si no.

casos que se utilizan el momento de la ponderación con respecto al total de observaciones. Este parámetro es definitorio pues de este depende el nivel de suavizado y el tamaño del error de ajuste que la superficie. Los casos que se encuentren fuera de esta ventana recibirán una ponderación igual a cero. A continuación se seleccionan aquellas unidades territoriales que tienen residuos significativamente positivos al 5% de confianza, es decir que  $(y_i - \hat{y})/\sigma > 1,96$  y se identifican los máximos locales de la superficie predicha en un radio o distancia determinada por el investigador para encontrar a los candidatos a centros de empleo, cuya ubicación será posteriormente examinada para determinar si en efecto son centros de empleo. Para realizar la estimación no paramétrica se utilizan como variable dependiente el logaritmo de la densidad del empleo o población; como regresores las distancias de los puntos en el eje este-oeste y en el eje norte-sur con respecto a lo que se denomina como el CBD o hipercentro, de tal manera que se obtiene un modelo del tipo<sup>19</sup>.

Una vez que se ha realizado el proceso de selección de candidatos, en la segunda fase se procede a realizar una regresión semiparamétrica para determinar si la localización de los candidatos identificados tiene alguna influencia sobre el total de la densidad de empleo. Se emplea un modelo de la forma:

$$\ln(Demp_i) = g(dcbd_i) + \sum_{j=1}^s (\delta_{1j} decent_{ij}^{-1} + \delta_{2j} decent_{ij}) + \varepsilon_i \quad (5)$$

Donde  $dcbd$  es la distancia de la observación  $i$  hasta el CBD y  $dcent_{ij}$  es la distancia entre cada observación con respecto al candidato a subcentro  $j$ ; y  $\varepsilon_i$  es el término de errores-tocástico. Como se ve la relación entre la distancia entre la distancia al CBD y la densidad de empleo es una función a ser determinada por un método no paramétrico. El término no paramétrico es estimado a través del uso de una expansión de Fourier. Para poder hacer esto primero se debe transformar la variable  $dcent$  de tal manera que se encuentre en un rango entre 0 y  $2\pi$ . La expansión de Fourier entonces sería:

$$g(dcbd_i) \approx \lambda_0 + \lambda_1 z_i + \lambda_2 z_i^2 + \sum_{q=1}^Q (\gamma_q \cos(qz_i) + \delta_q \sin(qz_i)) \quad (6)$$

La variable  $z$  denota la variable  $dcbd$  transformada para estar dentro del rango ya especificado. El  $q$  que se escoge es aquel que minimiza un criterio estándar de información. McMillen (2001) usa originalmente el criterio de información de Schwartz (SIC), en este trabajo al igual que en Mancheno and Rojas (2013) se utiliza el *criterio de información de Akaike* (AIC). Esto pues el criterio de Akaike presenta características de mejor desempeño práctico en la selección de modelos<sup>20</sup>. Esto pues el criterio de Akaike presenta características

<sup>19</sup>Esta técnica también se utiliza dentro de la estimación de una función de densidad poblacional

<sup>20</sup>Donde  $AIC = \log(\sigma^2) + 2m/n$ , teniendo en cuenta que además  $m = 3 + 2Q$ .

de mejor desempeño práctico en la selección de modelos<sup>21</sup>.

En el modelo se incluye  $D_{ij}^{-1}$  y  $-D_{ij}$ , para que los estimadores lineales en la regresión semiparamétrica tengan una relación positiva con el logaritmo de la densidad si ejercen alguna influencia sobre la densidad de empleo. A continuación se corre el modelo utilizando un proceso de regresión de selección inversa en que se incluyen todas las variables de distancia antes mencionadas para todos los candidatos en el modelo y se elimina aquella cuyo valor  $t$  sea el más bajo de todos y se vuelve a correr el modelo. Así hasta que todos los estimadores sean significativos al 10 %<sup>22</sup> <sup>23</sup>. En todos los pasos se forza a los términos de la expansión y al intercepto a permanecer en el modelo. Es así que son subcentros aquellos candidatos que tengan variables de distancia con efectos significativamente positivos en el logaritmo de la densidad del empleo al final del proceso de regresión de selección inversa.

El span utilizado para el ajuste es de 0,05; esto a pesar de que el trabajo original de McMillen (2001) establece un span de 0,5. La decisión de usar un span tan pequeño se toma, en primer lugar, debido a la evidente diferencia que se reflejó, tanto en el análisis exploratorio como en la función estimada, de la macrocentralidad y del resto de la ciudad; en segundo lugar debido a que un span tan grande como 0,5 puede hacer que no se capturen las particularidades de las localidades más pequeñas (Redfearn (2007)). Tras la detección de los lugares con residuos significativos se realiza una identificación de candidatos, estos son máximos locales dentro de un radio de 2.000 metros. Tras esto se procede a la segunda fase del método que es una evaluación de los efectos de cada uno de estos centros de empleo sobre la densidad de empleo para confirmar que estos candidatos son en efecto centros de empleo.

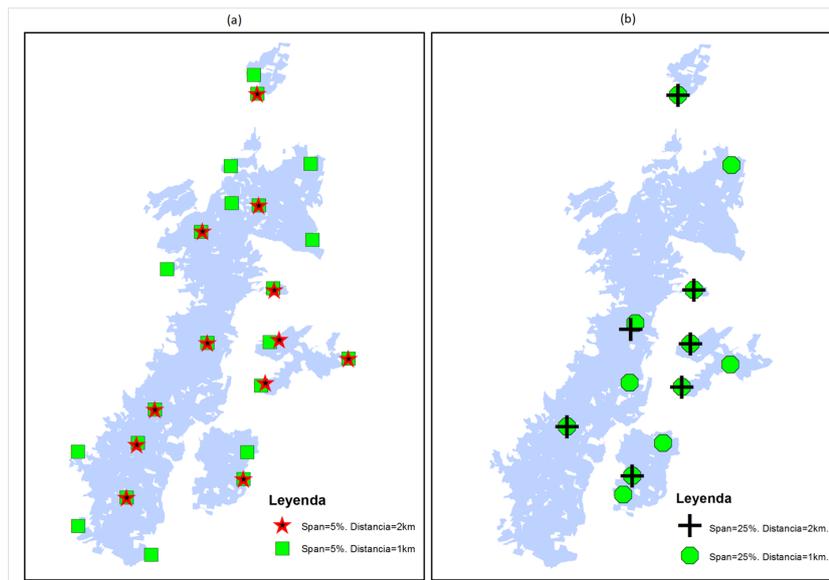
Los resultados obtenidos en trabajos previos con esta metodología muestran resultados consistentes al usar este span; sin embargo el ejercicio de identificación de candidatos se realizó usando distintos span. Específicamente a las distancias de dos kilómetros; además los span utilizados son de 0.05, 0.15, 0,25 y 0.5. Los resultados muestran que niveles de span muy grandes la técnica no logra detectar todas las localidades que el conocimiento general de la ciudad indicaría como centros de empleo. Pero más allá de eso además es evidente por el análisis exploratorio que el número de subcentros y sus localizaciones son más consistentes con el span más pequeño de 0.05. El cambio de distancia solo cambia el número de candidatos mas no la localización de los que ya han sido detectados; esto se puede apreciar en el Figura 9. En el panel (a) se pueden ver los candidatos obtenidos con un span de 5 % de las observaciones con distancias de 1.000 y 2.000 metros; en el panel (b) los obtenidos con un span de 25 % de las observaciones con distancias de 1.000 y 2.000 metros. Es evidente que el span es lo que determina la localización de los candidatos y en cierta medida el número. La distancia

<sup>21</sup>Para ver las ventajas del AIC sobre el SIC (o BIC), véase Burnham and Anderson (2002); Burnham and Anderson (2004).

<sup>22</sup>La metodología original McMillen (2001) considera un nivel del 20 % de confianza sin embargo por la reducción del span antes explicada, se utiliza un filtro más riguroso en la significancia de la relación entre las variables dependientes y los regresores.

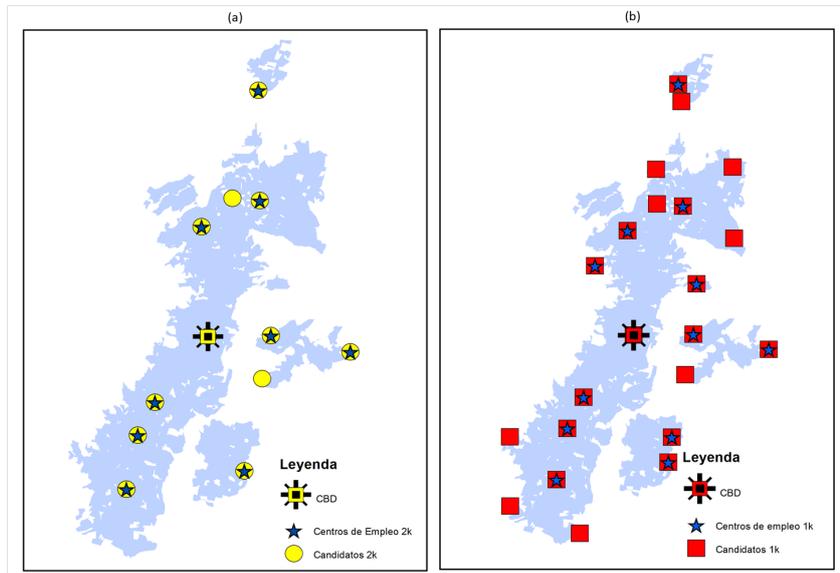
<sup>23</sup>La variable que se borra puede ser significativa si su coeficiente es negativo.

por otra parte solo hace variar el número de ellos, más no su localización. Como se observa el procedimiento con un span más grande representa pérdida de precisión en la localización además de ignorar, o simplificar particularidades locales. Por esto se escoge un span del 5 % de las observaciones que es conveniente para los fines de esta investigación y consistente con el análisis exploratorio.



**Figura 9:** Candidatos a centros de empleo varias distancias y span  
**Fuente:** INEC (2010). Censo Nacional Económico.

El procedimiento en la segunda parte del método se aplicó solo para los candidatos obtenidos con el span de 0.05; tanto para las distancias de 1.000 y 2.000 metros. Lo que se hace es utilizar un proceso de regresión *stepwise*, es decir que descarta variables del modelo original de manera que se corre el modelo cada vez que una variable es descartada en función de un criterio de significancia. El modelo de regresión utilizado, como ya se mencionó en los fundamentos metodológicos, es de tipo semiparamétrico utilizando una expansión de Fourier para poder representar la relación entre la distancia al CBD y la densidad del empleo; usa la parte paramétrica para estimar la relación entre la densidad y la distancia a cada uno de los candidatos a centro de empleo identificados en la primera parte. Este proceso permite identificar aquellas localidades que además de tener una densidad de empleo alta, en términos estadísticamente significativos, ejercen también influencia en la densidad de empleo bien de toda el área de estudio o bien del área próxima a este.



**Figura 10:** Candidatos y centros de empleo varias distancias y span  
**Fuente:** INEC (2010). Censo Nacional Económico.

Una de las particularidades de los resultados que cabe mencionar son los lugares comunes entre los dos escenarios. En otras palabras, los mismos centros que son identificados con el escenario de distancia de 2km son identificados a su vez en el escenario de distancia de 1km, lo que varía entre los dos es que el segundo escenario identifica otros sitios a parte de los del primer escenario. Esto se ilustra en la Figura 10, que tiene en el panel (a) los candidatos y los centros del escenario con span igual a 0.05 y distancia igual a 2km; en el panel (b) se registran los candidatos y centros identificados en el escenario con span igual al 0.05 y distancia igual a 1km.

El resultado del primer escenario identifica nueve centros de empleo distribuidos a lo largo del área de estudio. En el sur de la ciudad identifica tres centros de empleo en las zonas: Villaflores, Solanda y Quitumbe. En el norte se identifican tres localidades de cuatro posibles en las zonas de: Cotocollao, San Antonio y Carapungo; mientras el candidato de la zona industrial norte es descartado. En las zonas de los valles podemos identificar, de igual manera, tres centros de empleo de cuatro posibles localizaciones: Cumbayá, Tumbaco y Conocoto; por otra parte el candidato de la zona sur de Cumbayá se descarta.

En el segundo escenario el resultado no varía demasiado, los centros de empleo identificados por el primer escenario también son identificados por el segundo. El problema que surge con este escenario es el hallazgo de centros que resultan inconsistentes. Es el caso de un subcentro que se encuentra sobre el borde noroeste de la superficie, donde la ciudad comienza a escalar por las estribaciones del Pichincha. De igual manera sucede con la identificación de un centro en el área de Nayón. Otro caso de este tipo es la aparición de un segundo subcentro de empleo en Conocoto. Estos tres sitios no siguen los patrones que el análisis exploratorio

mostró y que daban los primeros indicios de la existencia de centros de empleo subordinados a la macrocentralidad. Esto ha llevado a pensar que un umbral muy pequeño de distancia podría llevar a incluir localidades en el análisis que pueden distorsionar la segunda fase del método. A pesar de esto el método logra excluir de manera certera a 10 candidatos y coincide en 9 localidades con el primer escenario.

Este análisis es importante, y se diferencia del visto previamente, porque ha permitido recoger las pequeñas alzas de empleo que se han presentado tanto dentro de la mancha urbana como en las zonas rurales conurbadas. Además de esto ha identificado localidades que tienen efectos estadísticamente significativos sobre la densidad de empleo; que al fin y al cabo son los que más se acomodan a los fines de esta investigación. Hasta este punto este método resulta el más consistente con el análisis exploratorio y con el conocimiento convencional que se tiene sobre la ciudad. Es necesario sin embargo explorar otra herramienta que permita confirmar la superioridad de este método o que en su defecto permita encontrar una mejor opción.

## 5 Localización de centros de empleo y la distribución de la población

Existe un acuerdo en la literatura de la nueva economía urbana sobre la existencia del fenómeno de la policentralidad que se encuentra acompañada por una baja progresiva en los costos de transporte y por los avances tecnológicos comunicacionales. La sección anterior mostró que Quito no es ajeno a este fenómeno. Siguiendo la línea de la Nueva Economía Urbana se esperaría que los centros de empleo tengan una influencia local o global sobre la distribución del empleo pero además sobre la población (Anas et al. (1998)). Esto debido a que son el resultado, de un juego conjunto de fuerzas de expulsión y atracción que estos determinan y que configuran aglomeraciones en el espacio. Esta sección pretende contrastar esta regularidad empírica y teórica para la ciudad de Quito; intentando dar un primer acercamiento al efecto de esta distribución sobre la configuración social de la ciudad desde una perspectiva espacial.

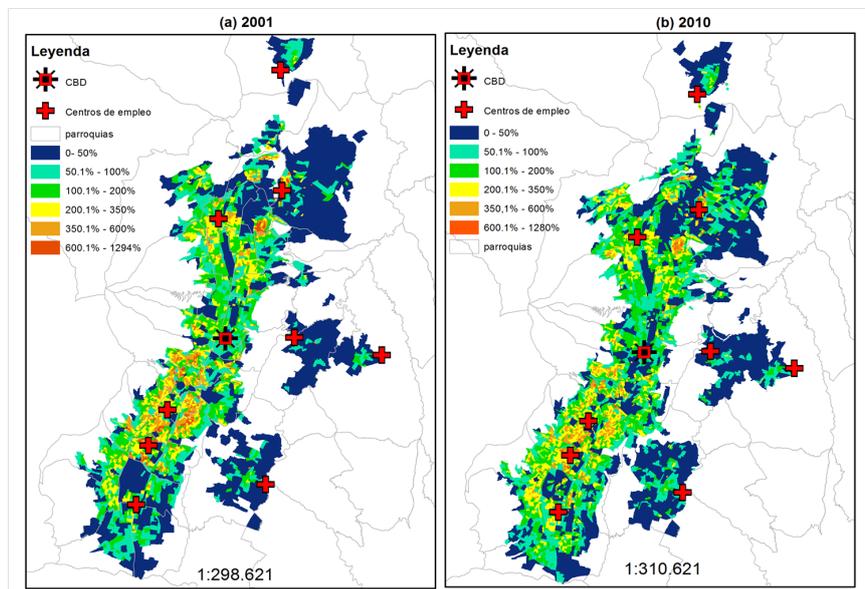
**Tabla 2:** Centros de empleo ubicados por el método de McMillen (2001)

Centros de Empleo
Quitumbe
Solanda
La Villaflores
Cotacollao
Conocoto
San Antonio
Carapungo
Cumbayá
Tumbaco

**Fuente:** INEC (2010). Censo Nacional Económico.

## 5.1 Análisis descriptivo

El simple análisis visual de la distribución de la densidad de población, expresada en los cocientes de localización, en relación a la localización de los centros de empleo identificados a través del método de McMillen (2001) da un buen primer panorama de la posible existencia de una relación de influencia de la distribución de la actividad económica sobre la distribución de la población. Esto entendiendo que los centros de empleo son aquellos puntos donde las actividades económicas tienen un mayor dinamismo y por ende ejercen fuerzas atractoras y expulsoras a partir de ellos. El mecanismo a través del cual se ejercen estas fuerzas, al menos teóricamente, son los precios del suelo. Los centros de empleo encontrados se exponen en la Tabla 2.



**Figura 11:** Mapa de densidad poblacional como cociente de localización con centros de empleo superpuestos, 2001-2010.

**Fuente:** INEC (2001,2010). Censo de Población y Vivienda; Censo Nacional Económico.

La Figura 11 contiene los mapas de densidad poblacional para 2001 y 2010 en el área de estudio, además de la localización de los centros de empleo. Como se puede observar, todos los centros de empleo se encuentran ubicados cerca a alzas locales de densidad de población. Se usa el mapa de densidad para observar si la distribución de 2001 guarda alguna relación con los centros de empleo de 2010; sin embargo, este ejercicio dejará de realizarse pues solo sirve para ilustrar el fenómeno de movimiento de la población alrededor de un centro de empleo. Si bien se puede presumir que la ubicación de los centros de empleo puede ser estable en el mediano plazo no se puede asegurar que su emergencia fue consecuencia, más bien, de un aumento previo en la densidad de población en esa misma localidad. Es decir este

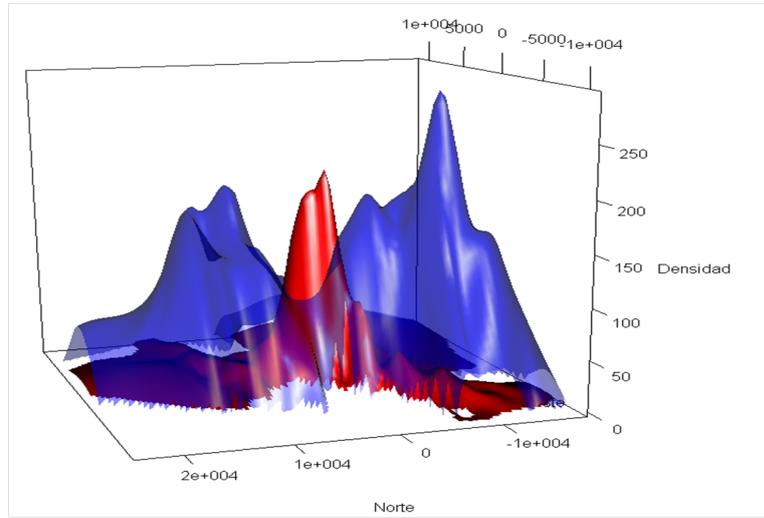
aumento de densidad poblacional atrajo actividades económicas en busca de mercados de consumo y fuerza laboral; y, no al revés, como lo sugiere la teoría. En esta medida la lógica general que se puede observar dos casos<sup>24</sup>: i) El primero alrededor de los centros de empleo que sufren un proceso de dispersión de una población que antes se encontraba concentrada alrededor de estos; este es el caso específico de Solanda, la Villaflores y Cotacollao. Estos tres parecerían estar en una etapa de reemplazo del uso habitacional del suelo por un proceso de dinamización de la economía en estas localidades, al igual que sucede en el CBD. El segundo caso es del resto de centros que parecen estar en una etapa todavía de atracción de población y que muestran alzas de la densidad de la población en el tiempo; estos se encuentran en partes de la ciudad cuyo desarrollo urbanístico es más joven que las zonas del primer caso. Un buen ejemplo de esto es Quitumbe o las zonas de los valles en Conocoto, Cumbayá o Tumbaco.

Otra particularidad que se observa claramente es que ninguno de los centros identificados se encuentra dentro del área del *cráter de densidad poblacional* que se encuentra en el centro de la ciudad, en el CBD. Lo que muestra claramente que “este cráter” se produce a partir del desplazamiento del uso del suelo habitacional por un uso del suelo por parte de actividades económicas; y, esto a una escala superior que en los centros de empleo antes mencionados. La importancia de esta macrocentralidad ha generado incluso un proceso de desplazamiento que supera a los de los centros de empleo subordinados a él. Este fenómeno se puede ver con claridad en la Figura 12, que tiene la representación de las curvas ajustadas de densidad poblacional y de empleo a través del método LOESS de ajuste de superficies.

El decrecimiento de los picos más cercanos de densidad poblacional al CBD, podrían explicarse entonces por dos sucesos: el crecimiento del área del cráter de la macrocentralidad y un posible crecimiento de los picos de empleo más próximos a este. Es decir una intensificación y extensión del uso del suelo en actividades económicas en el centro de la ciudad y además un simultáneo proceso de consolidación y crecimiento de centros de empleo ha determinado un desplazamiento del uso del suelo habitacional de estas áreas. Este desplazamiento ha representado un movimiento de la población hacia las parroquias rurales conurbadas y sectores menos poblados en el sur de la ciudad; esto a su vez ha determinado la aglomeración de la actividad económica alrededor de estos puntos poblados que han ganado tamaño entre 2001 y 2010. Esta constatación es una primera descripción de la relación entre la actividad económica y la población; esta sin embargo posee mucho de análisis intuitivo y conjeturas, por lo que es necesario pasar a un análisis con una base formal que dé una validación estadística sobre los rasgos identificados en esta sección.

---

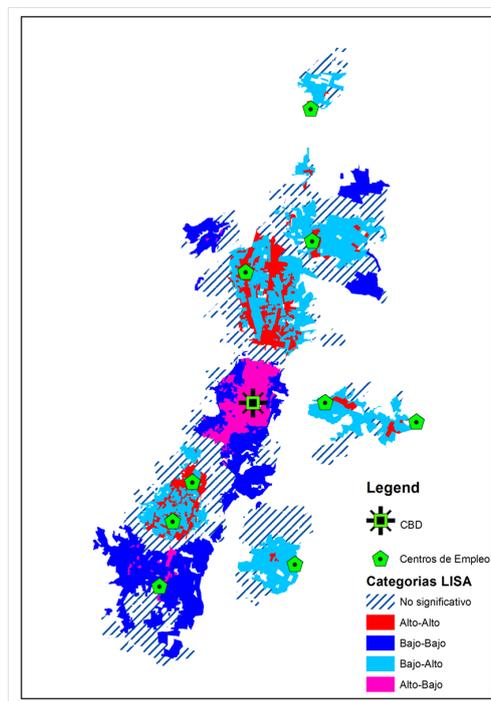
<sup>24</sup>Para entender esta lectura de mapas es indispensable que el lector se haya familiarizado con la cartografía del Anexo A que tiene como fin orientar con respecto a las localidades en el DMQ.



**Figura 12:** Superficie de densidad poblacional y de empleo ajustadas por LOESS, 2010.  
**Fuente:** INEC (2010). Censo de Población y Vivienda; Censo Nacional Económico.

## 5.2 Análisis LISA bivariado

En esta sección se usa la extensión del análisis LISA esta técnica para identificar clusters y outliers entre dos variables distintas. La técnica analiza al valor de una variable en un sector censal y le compara con el valor de otra variable en sus sectores vecinos. Por ejemplo, para este caso en particular, este método identifica un cluster tipo Alto-Alto cuando un sector censal tiene correlación espacial positiva con los sectores vecinos y cuyo valor de densidad de empleo sea mayor al promedio; al mismo tiempo el valor de la densidad de población de sus sectores vecinos debe ser mayor al promedio. Las categorías funcionan de la misma manera que en las secciones anteriores pero ahora la relación es entre las dos variables: densidad de empleo y densidad de población. Los resultados se consiguen tras 999 permutaciones para darles significación estadística a los resultados; la selección de sectores censales son los 4520 que registraban al mismo tiempo empleo y población. El análisis de la I de Moran bivariada global mostro un estadístico de  $-0.037$  lo que muestra una correlación espacial negativa entre las variables a un nivel global. Esto significa que existe una tendencia a nivel global de encontrar en el espacio valores altos de empleo y bajos de población. El análisis de las superficies de la Figura 12 hace pensar que este resultado podría variar al explorar esta correlación espacial en el territorio debido a la existencia de picos de densidad de empleo subordinados al CBD, los centros de empleo.



**Figura 13:** Superficie de densidad poblacional y de empleo ajustadas por LOESS, 2010  
**Fuente:** INEC (2001,2010). Censo de Población y Vivienda; Censo Nacional Económico.

El output de esta categorización se muestra en la Figura 13. Se especifican las categorías por colores mientras las zonas no significativas, que son las que no muestran patrones de distribución definidos, se encuentran en la zona a rayas. Una primera buena indicación sobre el método de identificación es que todos los centros se encuentran dentro de zonas significativas. Por ende, esto implica que la distribución conjunta de la población con respecto a la actividad económica no es un fenómeno aleatorio; sino que responde a una lógica espacial de aglomeración. Además casi todos los centros de empleo se encuentran en zonas con clusters de tipo Alto-Alto, esto da indicios del buen desempeño del modelo de identificación y de que existe, en efecto, una lógica de aumento de densidad de población a medida que aumenta el empleo; sobre todo en estas localidades clave. Los dos centros que se encuentran en clusters tipo Alto-Bajo son el CBD y el centro de empleo de Quitumbe. En este sentido se tiene que el primero debe su comportamiento al ahora confirmado fenómeno de desplazamiento de la población del cráter en el CBD, es decir la alta actividad económica que se concentra en este espacio ha repelido a la población que vivía en este espacio, tal vez a través de una lógica de precios; esto reafirma lo que se mostró con la superposición de las superficies estimadas en la Figura 12 . El segundo, el centro de Quitumbe, podría tener influencia solamente sobre un área muy pequeña circundante, o su efecto no se ejerce hacia la población sino solo hacia la actividad económica, en otras palabras por sus características de ubicación o tamaño su influencia todavía no es notable sobre la distribución de la población en lo que se esperaría

fuese su área de influencia. La Figura 13 confirma de alguna manera lo observado con el examen visual, corrobora a un nivel estadístico de confianza del 95 % que las distribuciones de la población y la actividad económica se encuentran relacionadas en ciertas localidades; teniendo cada localidad su particularidad. Sin embargo en relación con los movimientos de la población no se puede decir mucho por la inexistencia de un censo económico para el periodo 2001 y debido a que este análisis supone tener datos para una misma cartografía; es imposible sin un método de extrapolación de datos extender el análisis para ese periodo.

## 6 Un modelo de ciudad policéntrica para la densidad de población

El análisis exploratorio ha aportado sobremanera para entender la relación de la ubicación de los centros de empleo con respecto a la distribución espacial de la población. Con el objeto de elevar el rigor metodológico y estadístico en los resultados se correrán dos modelos econométricos que pueden aportar con información relevante. Los modelos son básicamente la extensión de la función inversa exponencial, que fue implementado por el trabajo seminal de Clark (1951), pero utilizando varias variables de distancia a los distintos centros de empleo. El motivo de escoger esta función es su inmensa popularidad y la descripción de varios autores como uno de los mejores métodos para la estimación de la función de densidad poblacional en el ámbito paramétrico (McDonald (1989); Joseph and Wang (2010)). El modelo en su versión original tiene la forma que expresa la siguiente ecuación:

$$DEN = D_0 e^{-\gamma x + \varepsilon} \quad (7)$$

Donde:  $DEN$  representa la densidad;  $x$  es la distancia de la unidad, en este caso sector censal, al centro de negocios de la ciudad;  $D_0$  y  $\gamma$  son las constantes a ser estimadas por el modelo; y  $\varepsilon$  es el termino de error estocástico<sup>25</sup>. Para poder ser implementado este modelo suele ser sujeto de una transformación logarítmica (McDonald (1989)) de tal manera que la forma de estimación real es:

$$DEN^* = D_0^* - \gamma x + \varepsilon \quad (8)$$

Donde el asterisco denota a aquellos términos que han sido transformados por su logaritmo natural<sup>26</sup>.

<sup>25</sup>McDonald (1989) hace referencia a la posibilidad de incluir el termino de error multiplicativamente y no aditivamente como método para mejorar la calidad del ajuste del modelo. Debido a la naturaleza ilustrativa de este análisis dentro de este trabajo solo se tomará en cuenta de manera aditiva.

<sup>26</sup>El principal problema de este método de estimación es que por la desigualdad de Jensen la transformación de la esperanza de una variable puede ser menor o igual a la esperanza de la transformación. Ver Verbeek (2008).

Para poder dar cuenta de la influencia de la localización de los centros de empleo sobre la población se modifica la especificación para que tome en cuenta la localización de estos. Este procedimiento se extrae parcialmente de Muñiz et al. (2008) quien realiza un ejercicio similar para el Área Metropolitana de Barcelona. El primer modelo es exactamente la extensión del modelo lineal solo que se incorpora las variables de distancia de cada sector censal hasta cada uno de los centros de empleo identificados. A diferencia de Muñiz et al. (2008) no se incluye en el modelo lo que él denomina una variable de control, que es la distancia a la infraestructura vial más cercana; por falta de datos. A pesar de lo antes mencionado el modelo 1 debería ser suficiente para nuestros fines usando la forma:

$$\ln(Den_i) = D_0 - \rho_i dist_{CBD} - \sum_{j=1}^s \rho_j dist_{SUB_{ij}} + \varepsilon_i \quad (9)$$

donde  $S$  es el número de centros identificados. Por motivos de conveniencia las variables de distancia entran al modelo multiplicadas por  $-1$ . Este modelo permite a través del signo de cada uno de los  $\rho$  identificar el efecto que tiene el centro de empleo sobre la población. El efecto esperado de un centro de empleo que ejerce influencia sobre la población arrojaría un  $\rho$  con signo positivo. Esto pues a medida que la distancia con respecto a este disminuye debería existir un mayor nivel de densidad. El modelo 2 es por otra parte una réplica del modelo 1 sin embargo se utiliza el inverso de la distancia a cada centro de empleo para poder detectar de manera más localizada los efectos de la distancia sobre la población. El modelo 2 es de la forma:

$$\ln(Den_i) = D_0 - \rho_i dist_{CBD} + \sum_{j=1}^s \rho_j (dist_{SUB_{ij}}^{-1}) + \varepsilon_i \quad (10)$$

De igual manera en el modelo 2 se espera un signo positivo de un centro de empleo que ejerza influencia sobre la población. Los resultados para los modelos se exponen en la Tabla 3, en este se puede apreciar el  $\rho$  estimado para cada variable, entre paréntesis el error estándar y el nivel de significancia con los símbolos seleccionados. El método utilizado para la estimación es el de mínimos cuadrados ordinarios. Para dar cuenta del error de varianza heterocedástica y autocorrelación de los errores se utiliza el método de White (1980) para el cálculo del error estándar y la matriz de covarianza.

Los resultados del modelo 1 arroja una distancia al CBD cuyo efecto no tiene el signo esperado y que además presenta un nivel muy bajo de significancia. Esto puede deberse principalmente a la formación del cráter de densidad en la localización del CBD, que ahora es más evidente debido a la inclusión de las otras variables de distancia. En cambio, de los centros de empleo solamente dos de ellos tienen un signo distinto al esperado, los de Cumbayá y Conocoto; y, existe solo un centro de empleo, Solanda, que no presenta efectos significativos. Este escenario muestra una densidad de población que responde de manera positiva a la mayoría de centros de empleo. Lo que significa que a medida que se acerca el

sector censal hacia uno de estos centros de empleo su densidad de población tiende a subir. La relación que se intuía dentro de los análisis exploratorios se confirma en la mayoría de los casos, incluso la influencia contraproducente de la presencia del cráter en el CBD.

El caso de Conocoto y de Cumbayá es interesante dada su característica de no pertenecer directamente a la mancha urbana. Los motivos para que el signo de un centro no sea el esperado son cinco y se explican de manera detallada en la fundamentación teórica. Aparentemente estos centros no ejercen el efecto esperado sobre la población; es decir la distribución de la población y de la actividad económica en estas localidades son independientes. Por ejemplo, este sería el caso de Cumbayá o Solanda. Por otra parte el caso de Conocoto se puede explicar por su proximidad a, Sangolquí, un centro poblado no considerado en este análisis.

El modelo 2 por otra parte muestra el modelo que trata de ubicar la relación de los centros y la densidad de población con un enfoque más local. Matemáticamente esto sucede debido a que en la relación  $\frac{1}{D}$  a medida que la distancia aumente el valor de la relación tenderá a cero. Los resultados tienen mayor consistencia, la distancia al CBD tiene en este modelo significancia estadística y además el signo esperado, incluso a pesar de la existencia del cráter de densidad. Se obtienen signos distintos a los esperados en los centros de empleo ubicados en las zonas de Tumbaco, Conocoto, Cumbayá y San Antonio. Lo que a resumidas cuentas dice que estas localidades no tienen un efecto significativo sobre la población que les rodea, esto puede deberse a una actividad económica relativamente joven, que no ha terminado de consolidarse y su influencia local sobre la distribución de la población es incipiente.

**Tabla 3:** Estimados para los modelos MCO 1 y 2

	Modelo 1	Modelo 2
Intercepto (Do)	9.7215930*** (0.511591)	5.0990155*** (0.0821991)
Distancia CBD	-0.0195280 (0.019392)	0.05063*** (0.0049082)
Dist. Quitumbe	0.1501370*** (0.018833)	
-		
Dist. Solanda	0.0176340 (0.022021)	
Dist. La Villaflora	0.0897380*** (0.018711)	
Dist. Cotacollao	0.0372820*** (0.010225)	
Dist. Conocoto	-0.0631080*** (0.011091)	
Dist. San Antonio	0.0933570*** (0.017697)	
Dist. Carapungo	0.0535380** (0.018466)	
Dist. Cumbayá	-0.1270140*** (0.036957)	
Dist. Tumbaco	0.0801240** (0.028811)	
1/Dist. Quitumbe		0.4154786*** (0.0591619)
1/Dist. Solanda		0.3882762*** (0.0556879)
1/Dist. La Villaflora		0.3479281*** (0.0672492)
1/Dist. Cotacollao		0.1980508*** (0.038453)
1/Dist. Conocoto		-1.0461961*** (0.1902883)
1/Dist. San Antonio		-0.1579994 (0.1222771)
1/Dist. Carapungo		0.2883554*** (0.0493084)
1/Dist. Cumbayá		-1.5344794*** (0.32847)
1/Dist. Tumbaco		-0.9322526** (0.3238656)

Los niveles de significancia se leen: \*\*\*=100; \*\*=0.99; y \*=0.95

**Fuente:** INEC (2001,2010). Censo de Población y Vivienda; Censo Nacional Económico.

Lo que ambos modelos tienen en común es la confirmación de que en Quito existe una influencia de la localización de la actividad económica, representada por los centros de empleo, y la distribución de la población. Este análisis permitió además validar a un nivel estadístico las conjeturas realizadas en el análisis LISA y en la visualización de los patrones de distribución de la densidad de población. El resultado de estos desarrollos muestran una ciudad que dentro de la macrocentralidad muestra un claro desplazamiento del uso habitacional,

al mismo tiempo muestra señales claras de influencia de los centros de empleo dispersos en toda la ciudad sobre la distribución de la población. A medida que los centros de empleo se consolidan se observa un fenómeno de desplazamiento del uso habitacional; este es el caso, por ejemplo, de Cotocollao o la Villaflores.

Se logra apreciar un gradiente de decaimiento, es decir una proporción a la que cae el nivel de densidad con respecto a la distancia, con el signo esperado en los centros consolidados, que además determinan patrones de distribución de la densidad de población alrededor de ellos que se apreciaron en la Figura 11 y en la Figura 12. Es decir cada uno de estos centros de empleo se comporta como un pequeño CBD, alrededor del cual la población se ubica para poder acceder a los servicios, empleos y otros equipamientos que pueden estar presentes en estos centros; la lógica se puede analizar también al revés: los centros que concentren dentro de sí empleos y equipamientos de interés de la población serán los que tengan mayor influencia sobre la densificación de la población alrededor de ellos. Mientras más cerca se encuentre el sector censal del centro de empleo entonces la población estará más densificada. Estos patrones de distribución deberían presentarse también en otras dimensiones como el salario de los empleados, la relación capital-trabajo, el precio del suelo y otras variables cuya distribución supone una relación estrecha con la ubicación de estos centros<sup>27</sup>. Estas particularidades deberían entonces determinar un patrón social de distribución de la población en función de sus ingresos y su renta ofertada; esto debido a que el precio del suelo alrededor del centro de empleo debería presentar un patrón también decreciente con un respectivo gradiente de decaimiento haciendo que el nivel de ingresos de la familia siga el mismo patrón de distribución que el de los precios.

## 7 Conclusiones

La población de Quito ha sido desplazada progresivamente de las localizaciones donde se consolidan centros de empleo. En el lugar donde se asienta la macrocentralidad se aprecia un cráter de densidad poblacional que da testimonio de esto. El crecimiento de la ciudad de manera extensiva y longitudinal ha desembocado en un fenómeno de baja de la densidad poblacional en los anteriormente consolidados centros de población. Si bien se identificó un gradiente de decaimiento de la población con respecto a la distancia al CBD; el modelo de ciudad monocéntrica no parece ser la mejor opción para explicar la distribución de la población.

La distribución de la densidad de empleo es eminentemente hipercéntrica, con alzas pequeñas locales que se subordinan a la macrocentralidad. Se confirma la existencia de múltiples centros de empleo distribuidos a lo largo de la ciudad, esto además por varios métodos de identificación. El método no paramétrico de McMillen (2001) McMillen (2001) arrojó los

<sup>27</sup>Comprobar esto no le compete a esta investigación de manera directa por lo que se deja abierto para posteriores trabajos

mejores resultados, de estos se desprende la existencia de 9 centros de empleo para Quito, cinco en zonas rurales conurbadas y cuatro dentro de la zona urbana de la ciudad. El norte de la ciudad depende en mayor medida de la macrocentralidad por lo que se identifican menos centros de empleo en esta área.

La relación entre la población y el empleo es clara a un nivel visual, y además utilizando técnicas estadísticas de análisis. La consolidación de centros de empleo en localidades como La Villaflora y Cotocollao determina, así como el crecimiento de la macrocentralidad, el desplazamiento del uso del suelo habitacional. Se confirma la estructura policéntrica al observar patrones de decaimiento de la población alrededor de cada centro de empleo. Se confirma la influencia de casi todos los centros de empleo sobre la densidad de empleo ya sea a un nivel local o global a través de un modelo de influencia de la ubicación de los centros de empleo sobre la densidad de la población .

## Referencias

- Alonso, W. (1964). *Location and land use: toward a general theory of land rent*. Editado por Harvard University Press. Publications of the Joint Center for Urban Studies of the Massachusetts Institute of Technology and Harvard University, Harvard University Press. <http://www.cabdirect.org/abstracts/19641802976.html>.
- Anas, A., Arnott, R., and Small, K. (1998). Urban Spatial Structure. *Economic Literature* 36, 3:1426–64.
- Anderson, J. E. (1982). Cubic-Spline Urban-Density. *Journal of Urban Economics*, 12:155–67.
- Anselin, L. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis* 27 (2), Wiley Online Library:93–115. doi:10.1111/j.1538-4632.1995.tb00338.x.
- Anselin, L. (1998). *Spatial econometrics: models and methods*. Dordrecht: Kluwer Academic.
- Anselin, L., Syabri, I., and Kho, Y. (2006). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* 38, (1):5–22.
- Avedaño, A. J. (2012). Three essays on urban spatial structure in Bogota D.C. *Universitat Autònoma de Barcelona*. <http://hdl.handle.net/10803/117436>.
- Bivand, R., Pebesma, E., and Rubio, V. (2008). *Applied Spatial Data Analysis with R*. Editado por Robert Gentleman, Kurt Hornik, y Parmigiani. Giovanni. New York, NY: Springer New York. doi:10.1007/978-0-387-78171-6.

- Burnham, K. and Anderson, D. (2002). *Model Selection and Multimodel Inference: a Practical Information-theoretic Approach*. New York Springer, volume 60. New York: Springer-Verlag. 2nd ed. [http://www.amazon.co.uk/Model-Selection-Multi-Model-Inference-Information-Theoretic/dp/1441929738/ref=sr\\_1\\_1?ie=UTF8&qid=1325808253&sr=8-1](http://www.amazon.co.uk/Model-Selection-Multi-Model-Inference-Information-Theoretic/dp/1441929738/ref=sr_1_1?ie=UTF8&qid=1325808253&sr=8-1).
- Burnham, K. and Anderson, D. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* 33, 2:261–304. doi:10.1177/0049124104268644.
- Cervero, R. (1989). *Americas suburban centers: The land use-transportation link*. Boston: Unwin Hyman.
- Cervero, R. and Wu, K.-L. (1997). Polycentrism, commuting, and residential location in the San Francisco Bay area. *Environment planning A* 29, 5:865–86. <http://ideas.repec.org/api/oenviron/29y1997i5p865-886.html>.
- Champion, A. G. (2001). A Changing Demographic Regime and Evolving Polycentric Urban Regions: Consequences for the Size, Composition and Distribution of City Populations. *Urban Studies* 38, 4:657–77. doi:10.1080/00420980120035277.
- Clark, C. (1951). Urban Population Densities. *Journal of the Royal Statistical Society. Series A (General)* 114 (4). Wiley for the Royal Statistical Society, pages 490–96. doi:10.2307/2981088.
- Clark, W. and Kuijpers-Linde, M. (1994). Commuting in Restructuring Urban Regions. *Urban Studies* 31, 3:465–83. doi:10.1080/00420989420080431.
- Cleveland, W. and Devlin, S. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association, Applications & Case Studies*, 83 (403)., American Statistical Association:596–610. doi:10.2307/2289282.
- Fujita, M. and Thisse, J.-F. (2002). *Economics of agglomeration: cities, industrial location, and regional growth*. New York: Cambridge University Press.
- Giuliano, G. and Small, K. (1991). Subcenters in the Los Angeles region. *Regional Science and Urban Economics* 21, 2:163–82. doi:10.1016/0166-0462(91)90032-I.
- Joseph, M. and Wang, F. (2010). Population density patterns in Port-au-Prince, Haiti: A model of Latin American city? *Cities* 27 (3)., Elsevier Ltd:127–36. doi:10.1016/j.cities.2009.12.002.

- Link, F. (2008). De la policentralidad a la fragmentación en Santiago de Chile. *Centro-h*, 2:13–24. <http://www.redalyc.org/articulo.oa?id=115112535002> .
- Mancheno, D. and Rojas, D. (2013). Subcentros de empleo en el DMQ y la creación de centralidades en el Plan Metropolitano de Ordenamiento Territorial 2012-2022. *Questiones Urbano Regionales 1*, 2:95–119.
- McDonald, J. F. (1989). Econometric studies of urban population density: a survey. *Journal of Urban Economics*, 26 (enero):361–85.
- McDonald, J. F. (1989). Econometric studies of urban population density: a survey. *Journal of Urban Economics* 26, (enero):361–85. <http://www.ncbi.nlm.nih.gov/pubmed/12283233>.
- McMillen, D. (2001). Nonparametric Employment Subcenter Identification. *Urban Economics* 50, 3:448–73. doi:10.1006/juec.2001.2228.
- McMillen, D. (2003). Identifying Sub-centres Using Contiguity Matrices. *Urban Studies*. doi:10.1080/00420980220080161.
- Millward, H. (2008). Evolution of population densities: Five canadian cities, 1971-2001. *Urban Geography* 29, 7:616–38. doi:10.2747/0272-3638.29.7.616.
- Muñiz, I., Garcia, M., and Galindo, A. (2003). Cubic spline population density functions and satellite city delimitation: the case of Barcelona. *Urban Studies* 40, 7:1303–21. doi:10.1080/0042098032000084613.
- Muñiz, I., Garcia, M., and Galindo, A. (2008). The Effect of Employment Sub-centres on Population Density in Barcelona. *Urban Studies* 45, 3:627–49. doi:10.1177/0042098007087338.
- Redfearn, C. (2007). The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area. *Urban Economics* 61, 3:519–41. doi:10.1016/j.jue.2006.08.009. doi:10.1016/j.jue.2006.08.009.
- Rojas, C., Muñiz, I., and García, M. (2009). Estructura urbana y policentrismo en el área metropolitana de concepción. *Eure* 35, 105:47–7. doi:10.4067/S0250-71612009000200003.
- Schelotto, S. (2008). La ciudad de Montevideo: ¿una metrópoli policéntrica? *Centro-h*, 2:37–46.
- Segre, R. (2008). Río de Janeiro: una vocación histórica policéntrica. *Centro-h*, 2:25–36. <http://www.redalyc.org/articulo.oa?id=115112535003>.
- Vallejo, R. (2008). Quito: capitalidad y centralidades. *Centro-h*, 2:47–54.

Verbeek, M. (2008). *A Guide to Modern Econometrics*. Editado por John Wiley Sons Ltd. Text. John Wiley & Sons. <http://www.amazon.de/p/0470517697>.

Von Thünen, J. (1996). *Von Thünen's Isolated State: An English Edition of Der Isolierte Staat*. Editado por P Hall. Elsevier Science & Technology. <http://books.google.com.ec/books?id=VJfpPAAACAAJ>.



# Analítica

Hacia un Algoritmo Optimo de Emparejamiento  
de Nombres

Towards an Optimum Name Matching  
Algorithm

Juan Carlos Delgado Loyola



[www.ecuadorencifras.gob.ec](http://www.ecuadorencifras.gob.ec)





# Hacia un Algoritmo Óptimo de Emparejamiento de Nombres<sup>1</sup>

## Towards an Optimum Name Matching Algorithm

Juan Carlos Delgado Loyola<sup>2</sup>

*Dirección de Registros Administrativos. Instituto Nacional de Estadística y Censos*

Quito - Ecuador

---

### Resumen<sup>3</sup>

La gran variedad de algoritmos de emparejamiento de nombres no son suficientes por sí mismos para proveer de una juntura óptima entre bases de datos administrativas pertenecientes a los mismos ciudadanos. De acuerdo con otras comparaciones experimentales de técnicas de emparejamiento, las variantes en la composición de nombres de personas tienen un rol importante en el proceso de emparejamiento, así que se ha hecho primero un análisis previo de lo mismo. El presente estudio realiza una evaluación de un algoritmo de emparejamiento con una muestra representativa de nombres de personas tomada de la población del Ecuador en el año 2010. Esta muestra se empareja con una similar que contiene nombres de ciudadanos de la base de datos del Registro Civil; y se asume que estuvieron presentes durante el día del censo. Se incluye un modelo de procesos para el nuevo algoritmo que combina técnicas fonéticas y de distancia de edición. Finalmente, se obtiene evidencia estadística a través de la diferencia significativa en el factor de exactitud para valores antes y después de la ejecución del algoritmo.

Palabras Clave: emparejamiento de nombres, técnicas fonéticas, distancia de edición.

JCL: 1: Fonética/Fonología; 2: Morfología<sup>4</sup>

---

<sup>1</sup>Primero quiero agradecer a todas las personas que asistieron a Diálogo Estadístico en INEC y a todos aquellos que tuvieron la oportunidad de conocer en detalle este trabajo, y alentaron a su publicación, especialmente a Boris Espinoza, Cesar Vicuña, Stalyn Flores, Yandré Jaime, Lorena Moreno y Karla Pasquel. En segundo lugar me es grato reconocer a Juan Fernando Galarraga, como colaborador docente en mis estudios relacionados con estas investigaciones y a su aporte profesional en la revisión de este trabajo.

<sup>2</sup>juancarlos.delgado@inec.gob.ec

<sup>3</sup>El presente estudio utiliza datos individuales con fines investigativos y estadísticos, por lo tanto no vulnera el principio de confidencialidad y reserva de la Información establecido en el artículo 21 de la Ley de Estadística, el artículo 6 de la Ley Orgánica de Transparencia y Acceso a la Información Pública y demás normas conexas.

<sup>4</sup>Journal Computational Linguistic (JCL): Niveles 1: Fonética/Fonología; 2: Morfología

## 1. Introducción

Las técnicas de emparejamiento de nombres tienen su fundamento teórico en el Procesamiento del Lenguaje Natural (PLN) y el reconocimiento de entidades (RE) a partir de textos escritos en cualquier idioma y generalmente las utilizan los algoritmos de búsqueda. Aquellas técnicas que tienen el propósito de emparejar nombres provenientes de diferentes fuentes se las ha denominado aquí: Técnicas de Reconocimiento de Nombres de Entidad (RNE). Los trabajos de Reynar (1998), y Huang et al. (2007) son útiles en este análisis.

La aplicación sistemática de las técnicas de emparejamiento de nombres en grandes bases de datos ha hecho posible algunas aplicaciones como: el cobro efectivo de impuestos, la ubicación de historias clínicas, la verificación de datos para chequeo de visas, el seguimiento a refugiados y personas sospechosas de terrorismo, la identificación de clientes potenciales, el censo basado en registros, etc. Dichos estudios han sido tratados por Hermansen (2006). Son útiles también las aplicaciones y técnicas RNE tratadas en Schay (2011). Finalmente, hay estudios de estadísticas basadas en registros administrativos que sugieren estos tipos de emparejamiento como los nombrados por Wallgren (2012).

Con la finalidad de obtener un registro único de ciudadanos partiendo de la información del último censo de población y vivienda de Ecuador y los registros de cedulados del Registro Civil, se ha propuesto un algoritmo optimizado de emparejamiento de nombres de personas, pues no se cuenta con cédulas de identidad en el censo. Este artículo trata sobre la evaluación de dicho algoritmo y se ha organizado de la siguiente forma: Una referencia a las técnicas RNE se presenta en el capítulo II. En el capítulo III se presenta los resultados de la evaluación del algoritmo RNE a través de la medición de su factor de exactitud. La confirmación de validez estadística de resultados propone el rechazo de la hipótesis nula, relacionada con la uniformidad de eventos antes y después del experimento, con una certeza de significación del 95 %.

## 2. Marco conceptual

### 2.1. Técnicas de emparejamiento

Establecer una comparación aproximada entre dos textos diferentes que tienen diferentes orígenes de datos es una tarea compleja que requiere más de una técnica. Para el proceso de emparejamiento de nombres se consideran tres grupos de técnicas: 1) fonéticas; 2) de deletreo y distancia; y 3) combinadas. Las técnicas fonéticas establecen comparaciones de palabras por similitud en la percepción de sonidos, cuando hay más de una forma escrita para representar el mismo nombre. Estas técnicas asignan códigos a cada secuencia de caracteres basados en el sonido que estos producen. El emparejamiento se hace entre las formas canónicas de los nombres. Las técnicas de deletreo y distancia, por su parte, generan un valor máximo de similitud equivalente al valor mínimo de distancia que resulta de operaciones de inserción,

borrado o sustitución de caracteres hechas para equiparar dos palabras; esta técnica se conoce como Distancia de Leveinshtein-Damerau. Otras técnicas de deletreo obtienen el valor de similitud mediante otros métodos como el relacionado con el reconocimiento de patrones de texto y la división de palabras en sub-unidades de  $N$  caracteres (N-grams). Estos métodos no necesitan de ninguna transformación fonética. Finalmente, todo lo que se fusione a través de métodos fonéticos con métodos de distancia de edición se conoce como técnicas combinadas. La más conocida es Editex. Esta última introduce verificación de sonidos iguales en las operaciones de distancia de la Técnica de Levenshtein y Damerau para descartar errores en la medición debido a la presencia de similitud fonética en los nombres.

Lo relevante de las técnicas de emparejamiento fonético son las adaptaciones que de ellas se han hecho a lo largo de los años para obtener un mayor número de similitudes aproximadas de nombres. Lo que comenzó con un simple algoritmo de indexación de apellidos en lenguas anglosajonas, mediante reducción de su representación escrita a 6 dígitos (Soundex), ha dado origen a una serie de algoritmos similares con reglas adicionales de representación adaptados a nombres en lenguas europeas (Metaphone, Phonex, NYSIIS), hindúes y ahora también, asiáticas. De interés experimental para el presente estudio son las dos nuevas adaptaciones al código Soundex para la lengua castellana. La primera es propuesta por Fernandez L. (2010) conocida como Soundex-SP y contempla reglas de indexación para las letras Y, LL, y CH. La segunda es de Mazariegos O. (2012) y va más allá al introducir reglas adaptadas a la pronunciación centroamericana como la asignación del dígito 7 a las letras Q y J y la reducción del sonido de la ‘CH’ a ‘V’ y el de la ‘LL’ a ‘J’.

La similitud de dos cadenas de caracteres es determinada por el valor de retorno de la función que calcula la distancia mínima de edición entre las cadenas de caracteres  $s$  y  $t$ , denominada distancia de Leveinshtein ( $distld$ ). La ecuación 1 muestra el cálculo de esta distancia. El valor mínimo se obtiene de la sumatoria de transformaciones entre todas las combinaciones posibles entre las posiciones de la cadena  $s$  y la cadena  $t$ . En la sumatoria;  $x$ , y son los valores absolutos para las operaciones de inserción, borrado y sustitución de caracteres.  $W_i$  es un valor de peso aplicado a cada operación  $i$ .

$$distld(s, t) = \min \sum_{i=1}^N W_i(|x|, |y|) \quad (1)$$

Par comprender mejor este cálculo, a cada transformación de  $(|x|, |y|)$  se la registra en una matriz  $d(i = 1..s, j = 1..t)$ , donde la posición  $i$  es de la primera cadena de caracteres ( $s$ ) y la posición  $j$  es de la segunda cadena ( $t$ ). Las operaciones de una transformación se expresan en la ecuación 2. El vector  $c(i, j)$  tiene los valores asignados durante el proceso. El vector  $W_i$  es el peso asignado a cada operación. Damerau introdujo en la ecuación original una nueva operación: la de transposición. Esta identifica cuando un carácter ha ocupado el lugar que le correspondía al siguiente o anterior en el nombre. Por ejemplo, la transposición en los nombres ‘Gabriel’ y ‘Grabiél’ son muy comunes.

$$\begin{aligned}
 d(i, j) = \text{mín}\{ & d(i - 1, j) + 1, \textit{insercion} \\
 & d(i, j - 1) + 1, \textit{borrado} \\
 & d(i - 1, j - 1) + c(i, j), \textit{sustitucion} \\
 & d(i - 2, j - 2) + c(i, j - 1) + c(i - 1, j) + 1\} \textit{transposicion}
 \end{aligned}
 \tag{2}$$

Cabe mencionar otras técnicas de deletreo como la de Guth, en la cual se obtiene un valor de similitud por acumulación de resultados en las variables dicotómicas de cumplimiento ( $1 = Si; 0 = No$ ) de hasta doce reglas de similitud entre posiciones anteriores o posteriores a cada caracter en dos palabras supuestamente similares. Sin embargo sus resultados no son muy convincentes para nombres cortos.

De las técnicas de deletreo que analizan caracteres comunes en los nombres para obtener un valor máximo de similitud, las más exhaustivas en la búsqueda y, por lo general; más precisas son las de Jaro y Jaro-Winkler, analizadas por Christen (2006). El algoritmo de Jaro calcula un valor de similitud entre dos cadenas, aceptando los caracteres que están dentro de la mitad de la longitud de la cadena más larga. La similitud de Winkleres una medida mejorada a la de Jaro. La ecuación 3 muestra este valor de similitud.

$$\textit{sim}_{\textit{jaro}}(s_1, s_2) = \frac{1}{3} \left( \frac{c}{|s_1|} + \frac{c}{|s_2|} + \frac{c - t}{c} \right)
 \tag{3}$$

Dónde  $s_1$  y  $s_2$  son las dos cadenas de caracteres a comparar,  $c$  es el número de caracteres comunes y  $t$  es el número de transposiciones:

## 2.2. Trabajos relacionados

Al igual que Kumar et al. (2010), este estudio propone utilizar el potencial de las técnicas fonéticas en la indexación de nombres, con el fin de reducir el gran volumen de datos iniciales a un conjunto de registros similares en pronunciación en poco tiempo de proceso. Ante la pregunta ¿Puede desarrollarse una codificación fonética adaptada al origen etimológico de los nombres?, se encontró que existen ciertos estudios de lenguajes latinos, los cuales proveen una descripción de procesos fonológicos en la creación de nombres personales y cómo los cambios lingüísticos y culturales afectan a dichos nombres. En trabajos futuros pueden ser de utilidad los estudios de Fall and Giraud-Carrier (2005), pues construyen chequeadores de deletreo fonético para mejorar en la precisión del emparejamiento fonético. También se puede encontrar algo similar en los experimentos de Mendoza and Zamudio (2005), Christen (2006) y Tibón (2005).

El emparejamiento de nombres a través de los algoritmos de distancia de edición ha tenido una amplia aceptación en la comunidad científica que se dedica a la fusión probabilística de registros. Tomando en cuenta el trabajo iniciado por Cohen et al. (2001), dónde las

en Registro Civil y que se presume estuvieron presentes en el día del censo, es decir el 28 de noviembre de 2010. Las muestras se prepararon para emparejar los nombres de la base de empadronados con los nombres de la base de cedulados utilizando el algoritmo RNE, sin que intervenga otro campo adicional ni tampoco las cédulas, únicamente los nombres de ciudadanos.

### 3.2. Construcción del algoritmo RNE

El algoritmo RNE se construyó para aplicarlo en un caso de estudio destinado a evaluar similitudes de a) nombres propios y b) apellidos a la vez entre los registros N1 (1...n) de la base de empadronados (CPV2010) y los registros de la muestra N2 (1...m) de cedulados (RCIVIL). En este caso  $n = m$ . El objetivo de la implementación fue fusionar registros por nombres de las personas, por lo cual el algoritmo realizaría un promedio general de los valores probabilísticos de similitud obtenidos en (a) y en (b). Antes de la utilización de los resultados en el análisis se descartaron casos de homónimos, es decir, cuando para un mismo registro de N1 (1...n) le corresponde más de un registro similar en N2 (1...m).

Para la construcción del algoritmo fue necesario realizar un experimento preliminar para comparar e identificar las técnicas de emparejamiento a aplicar acordes con el caso de estudio citado. Una muestra de 2,375 nombres propios, y otra de 2,253 apellidos principales, cada uno con al menos diez variantes en su escritura y pronunciación en el país, se seleccionaron entre los más frecuentes para comparar dichas técnicas.

En cuanto a técnicas de emparejamiento fonético, los resultados de este primer experimento indicaron que un alto porcentaje (82.08 %) de ciudadanos con apellidos de origen hispano en el país, y un porcentaje similar (70.03 %) de estas personas con nombres propios también de origen hispano, influían significativamente en la cantidad de coincidencias detectadas mediante una u otra técnica fonética. Por consiguiente, se tomó como pivote a la técnica Soundex-SP para comparar el número de coincidencias aproximadas que eran capaces de reconocer las otras técnicas. Se obtuvo que la técnica Soundex (71.97 % de casos) era la que mayor se acercaba al número de casos de nombres propios detectados por la técnica Soundex-SP, seguida de las otras técnicas: Phonex (78.38 %), Metaphone (41.77 %) y NYSIIS (21.83 %). En forma similar se compararon los números de casos de similitud aproximada de apellidos entre la técnica Soundex-SP y las otras técnicas, encontrando la misma distribución con similares porcentajes: Soundex (77.36 %), seguida de Phonex (71.42 %), Metaphone (28.82 %) y NSIIS (18,09 %).

En cuanto a técnicas de emparejamiento de deletreo y distancia, para el mismo experimento se estableció un límite porcentual de similitud aproximada de  $\geq 95\%$  tanto para nombres propios como apellidos. Dicho porcentaje corresponde a valores mínimos de distancia de edición entre 1 y 2 puntos entre las palabras comparadas y para técnicas de deletreo corresponde a los valores máximos de su coeficiente de proximidad superior al dicho límite porcentual. De todas estas técnicas, con la de Editexse obtuvo un rango mayor de casos de apellidos que su-

peraron el límite (45.10 %) seguida de Levenshtein (43.40), Levenshtein-Damerau (43.16 %), Guth (13.71 %), de N-gram (3.70 %), de Jaro (4.65 %) y la de Jaro-Winkler (5.31 %). Para casos de nombres propios, se obtuvo una distribución de resultados similar: Editex (69.87 %), Levenshtein (66.67 %), Levenshtein-Damerau (67.25 %), Guth (24.82 %), de N-gram (2.66 %), de Jaro (6.90 %) y la de Jaro-Winkler (8.71 %).

De la comparación de tiempos de procesamiento entre técnicas se encontró que las técnicas fonéticas son relativamente mucho más rápidas que las técnicas de deletreo y distancia. Tomando en cuenta que la indexación de nombres mediante códigos fonéticos se la hizo una sola vez en cuestión de pocos minutos, la operación de juntura entre nombres similares entre la muestra M1 y la muestra M2 se la hizo a su vez en cuestión de segundos, con un máximo de 80 segundos en apellidos y de 159.5 segundos en nombres propios para la técnica Metaphone, seguido de las otras técnicas Soundex, Soundex-SP, Phonex y NYSIIS, respectivamente. Sin embargo el tiempo empleado en la ejecución de técnicas de deletreo y distancia, aplicadas para emparejar las mismas muestras, fue del orden de horas de procesamiento, siendo la de Levenshtein-Damerau (6.15 horas) la que mayor tiempo obtuvo. En cuanto a la técnica combinada Editex, esta superó a todas en tiempo de ejecución: (10.38 horas).

Con los resultados de la comparación de técnicas se analizó la factibilidad de construir el algoritmo RNE en 5 fases. El flujo de dicho algoritmo se muestra en la Figura 1.

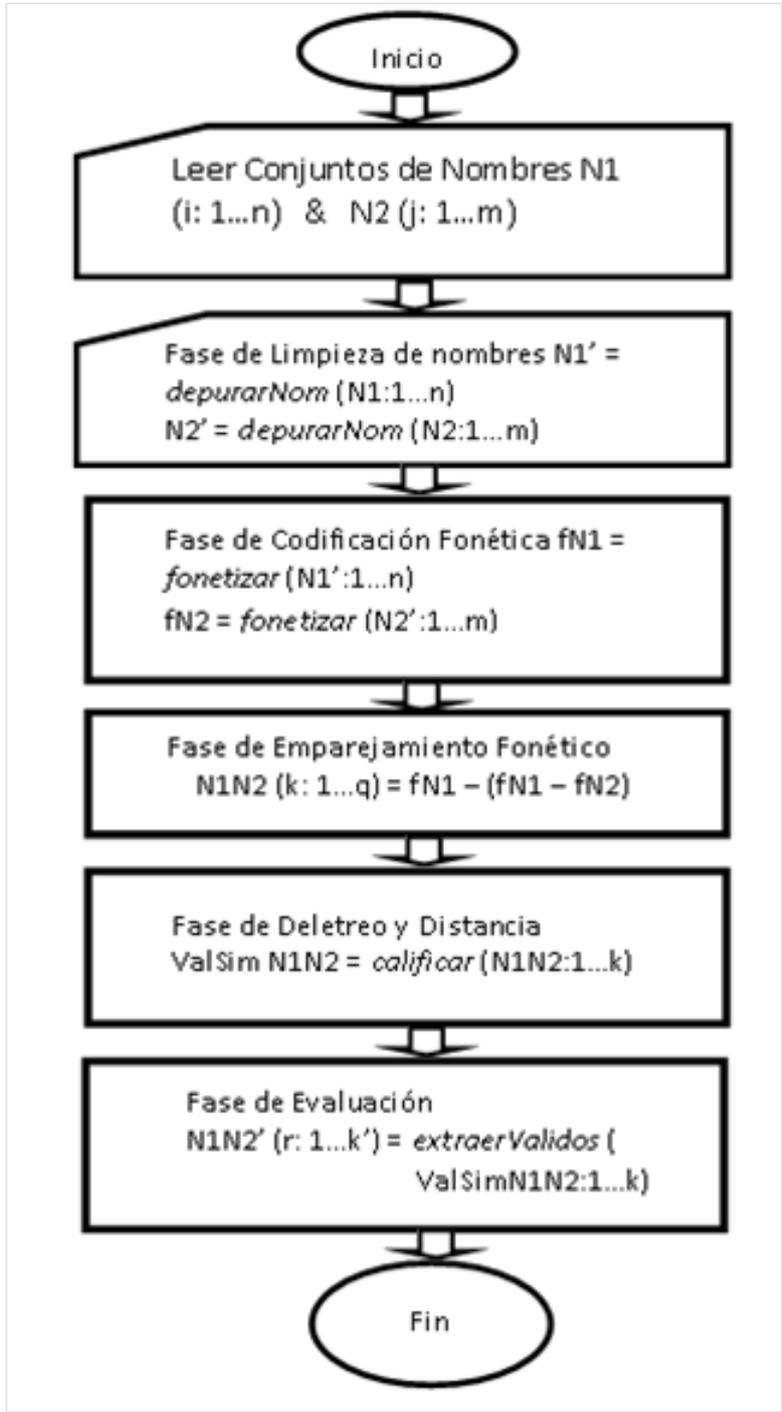


Figura 1: Algoritmo RNE

- **Fase de limpieza de nombres.**

En esta fase, se trató de que los nombres se simplificaran hacia un alfabeto estándar. En efecto se hizo una traducción de ciertos fonemas a su equivalente fonético único utilizado en la mayoría de lenguajes. Por ejemplo, para las letras que representan el sonido 'S' cuando tienen la misma pronunciación que la C o la Z se las convirtió al carácter 'S'. Además, tuvieron que ser retirados los espacios dentro del texto y se trató de evitar al máximo todos los caracteres especiales tales como "-", ".", ",", "(;)", etc. El resultado de esta fase fue la transformación a datos estandarizados y limpios.

- **Fase de codificación fonética.**

Tanto los nombres propios como los apellidos en cada conjunto de datos debieron ser traducidos a su equivalencia de código fonético. Por consiguiente, se utilizó el algoritmo Soundex-SP por el hecho se producían más emparejamientos que con otros algoritmos. Lo que se consiguió en esta fase es un resultado preliminar de registros por similitud fonética.

- **Fase de emparejamiento fonético.**

Se aplicó una simple operación de juntura para emparejar nombres propios y apellidos codificados con la técnica fonética. Para tal efecto se dividió la muestra grande de 15,746 registros en 17 pequeñas muestras. A través de esto consiguió mayor rapidez en la ejecución del algoritmo. Para reducir el número de coincidencias entre los conjuntos de la muestra N1 ( $i = 1 \dots n$ ) y la muestra N2 ( $j = 1 \dots m$ ), donde  $n = m$  se compararon los códigos fonéticos en lugar de los nombres originales y no se introdujo ninguna condición adicional con otros campos de la base de datos. Una operación de diferencia entre conjuntos ( $N1 - (N1 - N2)$ ) permitió simular la juntura de dos bases de datos. Lo que se consiguió en esta fase fue una gran cantidad emparejamientos aproximados por similitud fonética con un total de  $k$  registros, donde  $k < (m * n)$ .

- **Fase de deletreo y distancia.**

Una vez que el conjunto de datos de emparejamiento fonético fue generado por la operación de juntura en la fase previa, la similitud de cada par de nombres emparejados fonéticamente ( $N1 - (N1 - N2)$ ) fue calificada por la operación de distancia de Damerau-Levenshtein que bien pudo haber sido también un algoritmo de deletreo más complejo tal como Editex o Ngrams. Un coeficiente de similitud expresado en porcentaje se mostró con los resultados de distancia calculada para cada componente de nombres de persona: apellidos y nombres propios. Luego, un coeficiente promedio se calculó para todo el nombre. Si se hubiera requerido una mayor precisión con menos emparejamientos, Jaro y Jaro Winkler pudo haberse usado. Lo que se consiguió en esta fase fue una alta completitud, exactitud y precisión. Además, se

obtuvo un conjunto de registros menor del que se obtendría con el producto cartesiano entre  $N1$  y  $N2$  equivalente a la operación  $N1 + N2$  de  $(m * n)$  si es que solamente se aplicara una técnica de deletreo con todos los registros  $N1$  ( $i = 1..n$ ) y  $N2$  ( $j = 1..m$ ). En lugar de esto se aplicó la distancia solo a los registros fonéticamente coincidentes.

#### ▪ Fase de evaluación.

Los registros obtenidos en la fase previa tuvieron que ser clasificados por coeficiente de similitud en orden descendente. Aquellos emparejamientos con el más alto coeficiente se separaron en un conjunto de datos final. El resto de valores bajos de emparejamientos fueron rechazados. Los emparejamientos duplicados con valores altos o bajos tuvieron que ser analizados para detectar la presencia de homónimos. No se obtuvieron casos de homónimos pero en caso de que se hubieren presentado, se pudo haber utilizado una condición de restricción (ej. lugar de nacimiento). Lo que se consiguió en esta fase fue un conjunto emparejamientos altamente aproximado. Al final se encontró, por cada persona, el promedio de los valores de similitud alcanzados para las cadenas de caracteres de nombres propios y de apellidos. Luego se ordenaron los registros finales en forma descendente por el valor promedio de similitud para cada persona y se descartaron los registros que tenían valores de similitud inferiores al 95 %. El criterio para elegir este valor fue experimental y se basó en una revisión histórica de casos correctos en muestras de prueba mientras se probaba el algoritmo. Los casos de sinonimia entre un registro de persona en  $N1$  con varios idénticos de los similares en  $N2$  pudieron haberse comparado con relación a otra variable común como el lugar de nacimiento, sin embargo para efectos de este caso de estudio, solo fue indispensable evaluar las similitudes a través de las técnicas sin introducir otro tipo de comparaciones que involucraran otros campos, es decir, se eliminó el ruido en el algoritmo.

## 4. Resultados de ejecución del algoritmo RNE

Para aplicar el algoritmo RNE a la búsqueda de coincidencias de personas, se dividió la muestra aleatoria  $N1$  de 15,746 empadronados en 32 sub-muestras homogéneas de alrededor de 524 casos. A cada una de las sub-muestras se la emparejó con la muestra  $N2$  de 15,746 cedulados. Como resultado se obtuvo un total de 9,092 registros emparejados de 4,097 coincidencias exactas y 4,995 coincidencias aproximadas. Así, la completitud de casos emparejados fue del 57.7% del total de  $N1$ , con un 45.0% coincidencias exactas antes de aplicar el algoritmo y un 54% de coincidencias aproximadas después de aplicar el algoritmo.

## 5. Evaluación del algoritmo RNE

Se estableció como unidad de análisis el algoritmo RNE medido a través de su factor de exactitud (F1). Se hicieron varias corridas sucesivas con las 32 sub-muestras tomadas de la

base de datos de empadronados CPV2010 emparejados con la muestra de 15,746 cedulados de Registro Civil a nivel Nacional. La hipótesis nula  $H_0$  y la hipótesis alternativa  $H_1$ , antes y después de aplicar el algoritmo se describen en las ecuaciones 1 y 2, con un nivel de significación del 95 %.

$$H_0: p > 0,05 \rightarrow F1 \text{ antes} = F1 \text{ después}$$

$$H_1: p \leq 0,05 \rightarrow F1 \text{ antes} < F1 \text{ después}$$

$$F1 \text{ antes} < F1 \text{ después}$$

, siendo  $p$  el valor significativo de la prueba  $t$  para muestras relacionadas. El Factor F1 se obtuvo mediante la ecuación 3:

$$F1 = \frac{(2 * P * R)}{P + R}, \text{ Dónde } P = \text{ precisión y } R = \text{ Relevancia}$$

Los valores alcanzados antes y después de la ejecución del algoritmo para el factor F1 en las sub muestras se presentan en la tabla 1. Además se describen los falsos positivos y falsos negativos detectados luego de una revisión manual realizada sobre los resultados.

**Tabla 1:** Resultados de cálculo de los Factores F1

MUESTRA	EXACTOS ANTES	APROX ANTES	APROX DESPUÉS	FACTOR DE EXACTITUD PRE	FACTOR DE EXACTITUD POST	FALSO POSITIVO PRE	FALSO NEGATIVO PRE	FALSO POSITIVO POST	FALSO NEGATIVO POST
MCN01RC	174	0	255	0.770889488	1	0	255	0	0
MCN02RC	208	0	104	0.858725762	0.9967846	0	102	2	0
MCN03RC	120	0	28	0.918238994	0.9931973	0	26	2	0
MCN04RC	121	0	166	0.775675676	1	0	166	0	0
MCN05RC	173	0	156	0.808353808	1	0	156	0	0
MCN06RC	124	0	162	0.781163435	0.9929577	0	158	4	0
MCN07RC	116	0	141	0.788643533	0.9861933	0	134	7	0
MCN08RC	81	0	145	0.759930915	0.9865471	0	139	6	0
MCN09RC	120	0	170	0.775510204	0.9913043	0	165	5	0
MCN10RC	117	0	172	0.783258595	0.9509982	0	145	27	0
MCN11RC	124	0	153	0.786647315	0.9890511	0	147	6	0
MCN12RC	114	0	2	0.991452991	1	0	2	0	0
MCN13RC	137	0	0	1	1	0	0	0	0
MCN14RC	165	0	337	0.765765766	0.9169364	0	260	77	0
MCN15RC	115	0	174	0.780923994	0.9509982	0	147	27	0
MCN16RC	120	0	166	0.779661017	0.9822064	0	156	10	0
MCN17RC	97	0	148	0.768025078	1	0	148	0	0
MCN18RC	107	0	142	0.789022298	0.960334	0	123	19	0
MCN19RC	121	0	155	0.781740371	0.9963636	0	153	2	0
MCN20RC	115	0	170	0.775568182	0.9784946	0	158	12	0
MCN21RC	113	0	163	0.777614138	0.9777778	0	151	12	0
MCN22RC	181	0	219	0.816936488	0.9010989	0	147	72	0
MCN23RC	131	0	166	0.790896159	0.9669565	0	147	19	0
MCN24RC	124	0	151	0.796850394	0.9583333	0	129	22	0
MCN25RC	118	0	147	0.787692308	0.9827255	0	138	9	0
MCN26RC	117	0	201	0.762254902	0.9888712	0	194	7	0
MCN27RC	130	0	156	0.785714286	1	0	156	0	0
MCN28RC	127	0	216	0.776623377	0.9314642	0	172	44	0
MCN29RC	141	0	154	0.79776848	0.9845095	0	145	9	0
MCN30RC	120	0	313	0.786248132	0.7557471	0	143	170	0
MCN31RC	117	0	142	0.786482335	0.9941748	0	139	3	0
2MCN32RC	109	0	21	0.930909091	0.9922481	0	19	2	0
TOTALES	4097		4995				4420	575	

**Fuente:** Caso de estudio emparejamiento N1 y N2

La prueba de Kolmogorov Smirnov (KS) para una muestra dio como resultado valores de significancia menores a 0.05 para las series de valores de F1 correspondiente a los emparejamientos N1 vs. N2 de la Tabla 1. De esta manera, ambas series para F1 antes y F1 después no se aproximan a la curva Normal, por lo tanto no justifica aplicar la prueba normal *t* de muestras relacionadas. En su lugar se aplicó la prueba no paramétrica de rango de Wisconsin para muestras relacionadas.

De acuerdo con los resultados de la Tabla 2, en los 32 casos de ejecución del algoritmo en las sub-muestras, el estadístico significativo (2 colas) es inferior a 0.05.

**Tabla 2:** Prueba t del signo rango de Wilcoxon

	N	Rango Medio	Suma de Rangos
POST.Factor de Exactitud – Rangos Negativos	1 <sup>a</sup>	2.00	2.00
PRE.Factor de Exactitud Rangos Positivos	30 <sup>b</sup>	16.47	494.00
Lazos	1 <sup>c</sup>		
Total	32		
Test Estadísticas			
	POST.Factor de Exactitud	PRE.Factor de Exactitud	
Z			-4.821 <sup>a</sup>
Asymp. Sig. (2-colas)			.000

a. Basado en rangos negativos

b. Test del Signo Rango de Wilcoxon

**Fuente:** Resultados SPSS de Prueba t para muestras Relacionadas

## 6. Conclusiones

- Las técnicas fonéticas permiten reducir el número de casos de nombres coincidentes para un volumen de datos muy grande en un tiempo relativamente corto, del orden de unos cuantos segundos. Además funcionan eficientemente como métodos de indexación para búsquedas de nombres similares. Sin embargo, carecen de precisión en la valoración de la similitud entre nombres y solamente detectan semejanzas de escritura debido a una pronunciación equivocada utilizando a reglas conocidas del lenguaje.
- Las técnicas de deletreo y distancia son independientes del lenguaje en la que están escritas las palabras. Sin embargo necesitan del máximo de comparaciones posibles entre los caracteres presentes en los nombres y requieren de mucho más tiempo de procesamiento. Un caso de excepción de mejora a estas técnicas es la técnica combinada Editex, sin embargo, al introducir un peso de similitud fonética en las operaciones de la distancia de edición, multiplica considerablemente el tiempo de procesamiento y se vuelve dependiente de las reglas del lenguaje en que están escritas las palabras.
- Un algoritmo combinado de técnicas de emparejamiento como el propuesto aquí tiene un efecto óptimo para encontrar similitudes entre nombres de personas en las bases institucionales del Estado, sumando las ventajas de ambos tipos de técnicas. El análisis

del factor de exactitud de dicho algoritmo (F) demuestra que efectivamente se presenta un cambio de optimización al utilizarlo y que este no es debido al azar, ya que se tienen evidencias suficientes para rechazar la hipótesis nula  $H_0$ : Si F antes es igual a F después, con un nivel de significación del 95 %. Esto se explica por el p valor  $\leq 0.05$  ó sigma obtenido de la prueba del signo rango de Wisconsin para muestras relacionadas aplicada a los valores de F antes y después de aplicar el algoritmo. El p valor y los datos de la Tabla 2 permiten concluir que hay una diferencia significativa entre los valores antes y después del factor de exactitud F. Lo cual permite confiar en la hipótesis alternativa  $H_1$ : F antes < F después en el nivel de significación del 98.5 %.

## Referencias

- Christen, P. (2006). A Comparison of Personal Name Matching: Techniques and Practical Issues. *Department of Computer Science, The Australian National University*.
- Cohen, W., Ravikumar, P., and Fienberg, S. (2001). A Comparison of String Distance Metrics for Name-Matching Tasks. *Carnegie Mellon University, Carnegie Mellon University, Carnegie Mellon University*, pages 1– 6.
- Fall, C. and Giraud-Carrier, C. (2005). Searching trademark databases for verbal similarities. *World Patent Information*, pages 135–143.
- Hermansen, J. (2006). Advanced Global Name Recognition Technology - Entity Analytics Solutions. *IBM Corporation*, pages 1–13.
- Huang, C.-R., Simon, P., Hsieh, S.-K., and Prévot, L. (2007). Rethinking Chinese word Segmentation: Tokenization, Character Classification, or Wordbreak Identification.
- Kumar, A., Rawat, S., and Garg, S. (2010). Based Search of Indian Names in Databases. *ITT Kampur, India*, pages 1–14.
- Mendoza, A. and Zamudio, R. (2005). Nombres propios de procedencia latina. *AÑO VIII, No.17*, pages 153–182.
- Navarro, G. (2001). Guide Tour to Approximate String Matching. *Dept. of Computer Science, University of Chile, Blanco Encalada 2120*, pages 1–68.
- Nayan, A., Kiran, R., and P, S. (2002). Named Entity Recognition for Indian Languages. *Institute of Information Technology*, pages 1–103.
- Peng, T., Li, L., and Kennedy, J. (2001). A Comparison of Techniques for Name Matching. *Edinburg, UK*, pages 1–7.

- Reynar, J. (1998). Topic segmentation: Algorithms and applications.
- Schay, W. (2011). A Generic Framework for the Matching of Similar Names. *Faculty of Engineering and the Built Environment. University of Witwatersrand*, pages 11–203.
- Tibón, G. (2005). Diccionario Etimológico Comparado de Nombres Propios de Persona. *3ed.Fondo de Cultura Económica*, pages 1–248.
- Wallgren, A. (2012). Estadísticas basadas en Registros Administrativos: Aprovechamiento estadístico de datos administrativos. *México: INEGI, 2012. INEC, Biblioteca, Administración Central*.





# Analítica

Administración Central  
Quito

Juan Larrea N15-36 y José Riofrío  
Teléfonos: (593) 22544 326 - 22544 561  
Fax: (593) 22509 836 Casilla postal: 135C  
correo-e: planta\_central@inec.gob.ec

[analitika@inec.gob.ec](mailto:analitika@inec.gob.ec)

ISSN 1390-6208  
e-ISSN 1390-7867  
Número 1 - 2015  
Enero a Junio



[www.ecuadorencifras.gob.ec](http://www.ecuadorencifras.gob.ec)