



Munich Personal RePEc Archive

# **Persistent Spatial Structure Analysis. Regional Unemployment in Argentina.**

Herrera Gómez, Marcos

CONICET-IELDE, UNSa

August 2013

Online at <https://mpra.ub.uni-muenchen.de/49407/>

MPRA Paper No. 49407, posted 01 Sep 2013 09:40 UTC

# Análisis de Estructuras Espaciales Persistentes. Desempleo Departamental en Argentina.

Marcos Herrera<sup>(1)\*</sup>

<sup>(1)</sup>CONICET- IELDE, Universidad Nacional de Salta

## Resumen

Este trabajo presenta un conjunto de herramientas de econometría espacial que detectan dependencia espacial a nivel global y local. Estas herramientas son utilizadas para analizar la estructura espacial de la tasa de desempleo en Argentina en los Censos de 2001 y 2010. El estudio detallado permite la identificación y comparación de estructuras espaciales persistentes en la distribución departamental del desempleo.

**Palabras Clave:** *Autocorrelación Espacial, LISA, Test Getis-Ord, Entropía Simbólica, Desempleo.*

**Código JEL:** *C12, C21, J64.*

## Abstract

This paper presents a collection of spatial econometrics tools to detect global and local spatial dependence. These tools are used to analyze the spatial structure of the unemployment rate in Argentina in Census 2001 and 2010. Detailed study enables identification and comparison of persistent spatial structures in the departamental distribution of unemployment.

**Keywords:** *Spatial Autocorrelation, LISA, Getis-Ord Test, Symbolic Entropy, Unemployment.*

**JEL Codes:** *C12, C21, J64.*

---

\*El autor agradece los comentarios realizados en el 27º Seminario de Investigación del IELDE, Universidad Nacional de Salta. E-mail: mherreragomez@gmail.com

# 1. Introducción

En las últimas décadas una creciente literatura ha indagado sobre las características que presenta el desempleo regional. Trabajos pioneros como Blanchard y Katz (1992) han motivado un mayor análisis acerca de las causas y consecuencias de las disparidades regionales en las tasas de desempleo en Estados Unidos. A su vez, en Europa, los estudios centrados en esta problemática se incrementaron notoriamente: Decresin y Fatás (1995), Taylor y Bradley (1997), Faini *et al.* (1997) y Overman y Puga (2002), entre otros.

Entre las preocupaciones de los mencionados estudios se encuentra el fenómeno de la persistencia del desempleo regional. La comparación respecto a la evolución de las tasas de desempleo entre Estados Unidos y Europa permiten bosquejar diferentes explicaciones sobre la existencia y persistencia de las disparidades regionales. En Estados Unidos se observan diferencias entre estados pero no son persistentes a través del tiempo. Blanchard y Katz (1992) muestran evidencia sobre la existencia de un mecanismo de ajuste hacia el equilibrio competitivo, de tal manera que, las regiones con elevado desempleo en pocos años ven reducidas sus tasas aún debajo del promedio nacional (véase Bertola e Ichino, 1995). En contraposición, en Europa se observan tasas de desempleo que son persistentes en el tiempo como el caso de España (Jimeno y Bentolila, 1998) o el de Italia (Faini *et al.*, 1997). Taylor y Bradley (1997), realizando un estudio comparativo entre países europeos, encuentran evidencia sobre las disparidades regionales en el desempleo en Alemania, Italia y Reino Unido. Para una extensa revisión de las posibles explicaciones teórico-aplicadas sobre este fenómeno en países desarrollados véase Elhorst (2003).

En Argentina, las cuestiones regionales tales como las diferencias socio-económicas, demográficas y la persistencia de las mismas han sido destacadas en diferentes oportunidades. Díaz Cafferata y Figueras (1996) mencionan la existencia de una marcada desigualdad en la evolución del desempleo regional y llaman la atención sobre las particularidades de cada región como un problema detrás del diferencial del desempleo. Lamarche *et al.* (1998) mencionan que el desempleo regional es determinado por factores comunes a todas las regiones, factor nacional, pero que existen factores específicos de cada región que explican el comportamiento diferencial. En un trabajo más reciente, Galiani *et al.* (2004) analizan la persistencia del desempleo regional a nivel provincial. La evidencia encontrada refleja que la estructura del desempleo regional no es muy persistente, asemejándose a los resultados encontrados para Estados Unidos y alejado de los evidenciados en España y el Reino Unido.

Nuestro interés se centra en el analizar la distribución geográfica de la tasa de desempleo y su evolución en los censos más recientes, 2001 y 2010. En particular, estamos interesados en detectar clusters persistentes identificados a nivel departamental. Para ello utilizaremos diferentes estadísticos de dependencia espacial, globales y locales, así como un herramental basado en análisis simbólico que permite capturar la complejidad de la distribución espacial y compararla entre diferentes puntos temporales. Más específicamente, la idea es contrastar si existe evidencia de permanencia temporal una vez que se ha detectado la persistencia espacial. Al hablar de persistencia espacial nos referimos al comportamiento similar de los departamentos que se encuentran en un radio de distancia cercano en el espacio para un periodo temporal determinado.

El trabajo se encuentra estructurado de la siguiente manera. En la sección 2 se presenta la motivación de la investigación basada en la aparente similitud de la distribución espacial del desempleo en el año 2001 y 2010. En la sección 3 se desarrollan las principales herramientas espaciales. En primer lugar, se muestran los estadísticos de dependencia global como el I de Moran,  $c$  de Geary y  $G$  de Getis y Ord. Luego, la exposición continúa con el análisis de indicadores espaciales locales, I de Moran y  $G^*$  de Getis y Ord. Estos últimos estadísticos son utilizados para establecer una estrategia de comparación intertemporal. En la sección 4, se presenta un herramental novedoso proveniente de la teoría del caos que permite complementar la estrategia desarrollada en la sección previa. La sección 5 se encarga de utilizar los desarrollos estadísticos para detectar la persistencia la tasa de desempleo a nivel departamental. Finalmente, la sección 6 presenta las principales conclusiones.

## 2. Motivación

Los estudios empíricos por lo general se basan en datos sub-nacionales, considerando la diferencia del desempleo entre cada región. Esta metodología es la aplicada en la mayoría de las investigaciones en Argentina (lo mismo puede decirse para los estudios pioneros de Estados Unidos y Europa). En el caso de Argentina, los datos provienen principalmente de la EPH (Encuesta Permanente de Hogares). Este tipo de encuesta es representativa de los grandes aglomerados urbanos y permite un número de observaciones temporales abundante para cada uno de los mismos. Sin embargo, los resultados no son representativos de toda la población de cada región y solo pueden realizarse inferencias a nivel urbano, el componente rural no se encuentra representado.

En este trabajo se abordará la problemática de la persistencia del desempleo con algunas particularidades. En primer lugar, el análisis se basará sobre un nivel de agregación de la información inferior al presentado en anteriores estudios. La reciente publicación de datos a nivel departamental permite un análisis más detallado en cuanto a su representatividad geográfica. En segundo lugar, los métodos que se aplicarán provienen de la econometría espacial que consideran a las unidades interrelacionadas y toman en cuenta explícitamente la ubicación geográfica de cada observación. Esto contrasta fuertemente con el análisis habitual en donde se intenta capturar la proximidad entre áreas pero sin considerar explícitamente la interdependencia regional. A su vez, el uso de información censal permite mayor representatividad en cuanto a la población ya que incluye a la totalidad de la misma, sea urbana o rural. Por otra parte, es necesario mencionar algunas limitaciones de los datos utilizados: (1) la estimación del desempleo en los censos puede no ser tan precisa respecto a la obtenida por la EPH en donde existe un conjunto de preguntas que sirve de filtro para definir la situación laboral cada encuestado; (2) no hay suficiente información censal para establecer un conjunto de variables explicativas sobre el desempleo departamental y solo se dispone de dos puntos temporales.

Teniendo en cuenta las características de la información utilizada, a continuación se presenta un cuadro resumen de los principales estadísticos en ambos censos, Cuadro 1. La tasa de desempleo se ha definido como el porcentaje de desempleados respecto a la población económicamente activa de cada departamento siguiendo la metodología del INDEC. El tamaño muestral es de 510 observaciones georeferenciadas para ambos censos, siendo suficientemente grande como para realizar inferencias espaciales válidas. Los distritos de Capital Federal han sido agregados en una única unidad, en ambos censos, a los efectos de mantener la comparabilidad.

Cuadro 1: Desempleo Departamental. Estadísticos Descriptivos.

Estadísticos	Desempleo 2001 (%)	Desempleo 2010 (%)
Media	23.9	5.2
Mediana	23.2	5.0
Varianza	7.9	1.9
Mínimo	2.1	1.2
Máximo	51.0	19.2

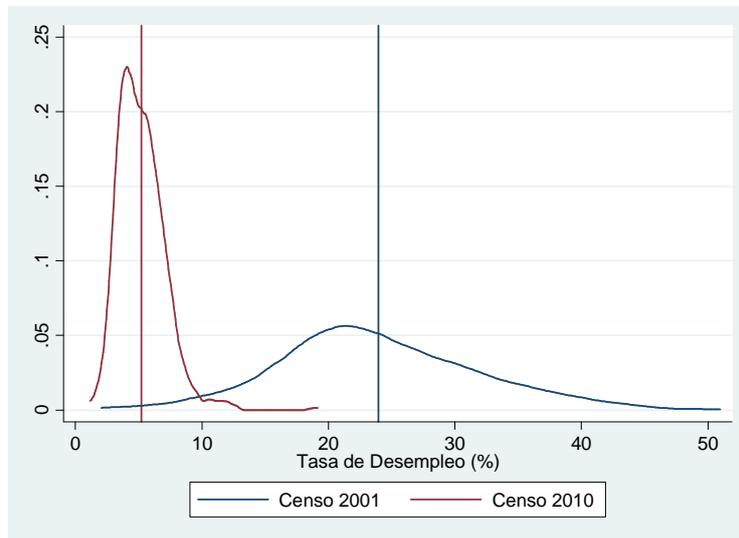
Fuente: Censo 2001, 2010. Base de Datos Redatam. INDEC.

Puede observarse la elevada tasa de desempleo en el año 2001, con valores departamentales que alcanzan un máximo de 51 % de la población económicamente activa. Sin entrar en mayores detalles sobre las posibles deficiencias a la hora de la medición,<sup>1</sup> lo cierto es que los niveles de desempleo reflejaban una situación extrema. La consecuente crisis de Argentina, y su notoria recuperación, nos ubican en el año 2010 con unas condiciones laborales muy diferentes, con un promedio del desempleo

<sup>1</sup>El censo se desarrolló en un momento de elevado descontento poblacional por la situación socio-económica que atravesaba el país y esta situación puede haber afectado la calidad de la encuesta censal.

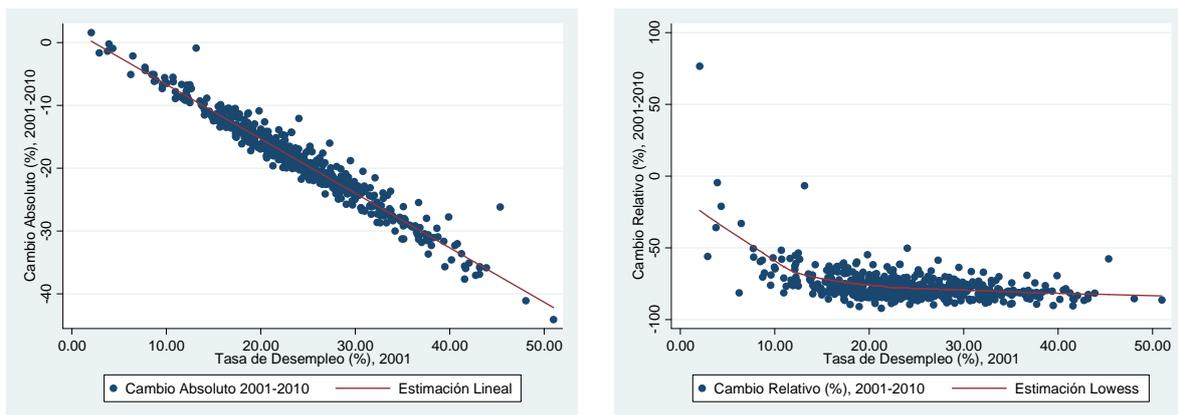
departamental de 5.2 %. La Figura 1 resume el drástico cambio en la distribución del desempleo entre ambos censos. Las líneas verticales en cada distribución muestran el valor promedio de cada censo.

Figura 1: Distribución del Desempleo en los Censos 2001 y 2010.



Por su parte, una simple inspección de los datos disponibles entre censos muestra que los departamentos con mayores tasas de desempleo en el 2001 son los que registran una mayor caída absoluta y también porcentual en el año 2010 (Figura 2). Una gran cantidad de departamentos registran reducciones entre los dos periodos de más del 50% respecto al valor del 2001. La mejora en la tasa de desempleo es global, exceptuando un único departamento que muestra un incremento en el desempleo entre censos.

Figura 2: Cambio Absoluto y Relativo del Desempleo.



Estos hechos tienden a confirmar los resultados de la baja persistencia del desempleo y la existencia de mecanismos de mercado tendientes a la convergencia a un valor promedio. Sin embargo, el tipo de información hasta aquí analizada no difiere del tipo tradicional en donde no se considera la ubicación relativa de cada departamento en el espacio geográfico.

Una forma de comenzar a explorar la distribución espacial del desempleo es mediante mapas como los presentados en la Figura 3. En cada uno de los mapas puede observarse a los departamentos y sus respectivos centroides. Cada centroe es representado por un círculo: los departamentos con valores de tasa de desempleo por encima del promedio son representados por círculos sólidos y los



### 3. Estadísticos Espaciales para Comparación Intertemporal

Los estadísticos que se presentan a continuación brindan diferentes medidas de asociación para datos espaciales. Al hablar de datos espaciales hacemos referencia a un conjunto de mediciones tomadas en localizaciones específicas dentro de un espacio geográfico o región de estudio.

Los estadísticos pueden clasificarse como medidas de dependencia globales o locales. Las medidas globales de dependencia espacial usan la información completa del conjunto de datos con el fin de obtener un valor individual para todo el espacio geográfico. Sin embargo, estos estadísticos ocultan información sobre la heterogeneidad dentro de pequeñas áreas locales. En tales circunstancias, la evaluación global de la dependencia espacial puede ser engañosa debido a que estos estadísticos ofrecen información del valor promedio pero no ofrecen información sobre la variabilidad de la intensidad de la dependencia ni la localización exacta de estos patrones. Las limitaciones de estos estadísticos globales han sido reconocidas en diferentes trabajos como Getis y Ord (1992), Anselin (1995) y Boots (2003), por citar algunos.

En contraposición a los estadísticos globales, las medidas locales de asociación examinan la dependencia espacial en un subconjunto de datos definidos con respecto al valor asumido en cada localización y su entorno local. Estas medidas se enfocan en la identificación de variaciones de la asociación espacial sin buscar regularidades generales. De esta forma, los enfoques locales son capaces de generar un valor particular para cada localización del conjunto de observaciones. Este valor particular puede ser contrastado estadísticamente y clasificado en diferentes grupos denominados clusters espaciales. Al hablar de cluster espacial nos referimos a un grupo de observaciones geográficamente limitado, con un tamaño y concentración suficiente, tal que es poco probable que haya ocurrido por casualidad o aleatoriamente.

Un elemento fundamental para los estadísticos espaciales, ya sean globales o locales, es la determinación del “vecindario” relevante, esto es, aquellas unidades que influyen a la observación considerada. El conjunto de estas unidades vecinas se consideran que interactúan significativamente. Convencionalmente la estructura de vecinos para todo el conjunto de datos se resume en una matriz de contacto o de pesos espaciales  $W$ . Esta matriz tiene dimensión  $n \times n$ , donde cada fila y columna corresponde a un par de observaciones. Los elementos  $w_{ij}$  toman valores diferentes de 0 cuando las observaciones  $i$  y  $j$  son vecinas, en caso contrario toman un valor igual a 0. Por convención, una observación no puede ser vecina de si misma implicando que los elementos de la diagonal principal,  $w_{ii}$ , son todos iguales a 0. Es usual que la matriz de pesos sea estandarizada por filas, tal que la suma de cada fila sea igual a la unidad. Los vecinos pueden ser definidos por diferentes criterios geográficos y/o económicos. Entre los criterios geográficos más comunes podemos mencionar al de distancia y de contigüidad. En este trabajo utilizaremos el criterio de distancia, más específicamente, el criterio de  $k$ -vecinos más cercanos que permite asegurar un número constante de vecinos para cada localidad, evitando de esta forma que alguna unidad no contenga vecinos.

#### 3.1. Estadísticos Globales

Entre los estadísticos más habituales en análisis espacial se encuentra el coeficiente de autocorrelación  $I$  de Moran. Este estadístico cuantifica la similaridad de una variable entre áreas que son definidas como espacialmente relacionadas (Moran, 1950) y viene dado por:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S_0 \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (1)$$

siendo  $S_0 = \sum_i \sum_j w_{ij}$ ,  $n$  el tamaño muestral,  $x_i$  el valor de la variable en el área  $i$  y  $\bar{x}$  el valor promedio de la variable  $x$  en toda la muestra. El peso  $w_{ij}$  es un elemento de la matriz de pesos espaciales  $W$ . Para el caso de una matriz  $W$  estandarizada se tendrá  $S_0 = n$ .

Un valor positivo del I de Moran indica autocorrelación espacial positiva (es decir, predominancia de observaciones con vecinos que asumen valores similares), mientras que un coeficiente negativo indica autocorrelación espacial negativa (es decir, predominancia de observaciones con vecinos que poseen valores disímiles).

Bajo la hipótesis nula de aleatoriedad, la distribución del I de Moran se aproxima a la normal cuando  $n \rightarrow \infty$ , por lo que puede plantearse el estadístico estandarizado:

$$z[I] = \frac{I - E(I)}{\sqrt{V(I)}} \underset{as.}{\sim} \mathcal{N}(0, 1). \quad (2)$$

Otro estadístico global es el test c de Geary (Geary, 1954) que considera la similaridad entre pares de regiones. El rango de valores que puede asumir este test va de 0 a 2, con 0 indicando autocorrelación espacial positiva perfecta y 2 autocorrelación espacial perfecta negativa.

$$c = \frac{(n-1) \sum_i \sum_j w_{ij} (x_i - x_j)^2}{2S_0 \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (3)$$

Tal como sucedía con el test I de Moran, el estadístico c de Geary puede aproximarse en su distribución a la normal:

$$z[c] = \frac{c - E(c)}{\sqrt{V(c)}} \underset{as.}{\sim} \mathcal{N}(0, 1). \quad (4)$$

El estadístico global  $G$  propuesto por Getis y Ord (1992) viene dado por:

$$G = \frac{\sum_i \sum_j w_{ij} x_i x_j}{\sum_i \sum_j x_i x_j}, \quad \forall j \neq i, \quad (5)$$

donde cada elemento ya ha sido definido previamente.

Nuevamente este estadístico posee una distribución compleja y la aproximación asintótica a la normal puede ser obtenida como:

$$z[G] = \frac{G - E(G)}{\sqrt{V(G)}} \underset{as.}{\sim} \mathcal{N}(0, 1). \quad (6)$$

### 3.2. Estadísticos Locales

Anselin (1991) define un conjunto de indicadores locales de asociación espacial comúnmente denominados LISA. Los estadísticos LISA permiten la descomposición de indicadores globales. Un test LISA es definido por Anselin como todo estadístico que cumple dos requisitos:

1. El valor LISA de cada observación indica el alcance de la significancia espacial alrededor de la observación.
2. La suma de todos los LISA es proporcional al indicador global de asociación espacial.

Un estadístico que cumple con las condiciones LISA es la forma local del test  $I$  de Moran, tal que para la observación  $i$  –ésima viene dado como:

$$I_i = \frac{n(x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \sum_{j=1}^n w_{ij} (x_j - \bar{x}), \quad (7)$$

donde  $w_{ij}$  es un elemento de la matriz de pesos espaciales  $W$ .

Bajo el supuesto de aleatorización completa, el valor esperado de  $I_i$  es:

$$E [I_i] = -\frac{1}{n-1} \sum_{j=1}^n w_{ij}. \quad (8)$$

La significancia del  $I_i$  puede ser contrastada sobre la base de la distribución normal donde el Moran Local es estandarizado como:

$$z [I_i] = \frac{[I_i - E [I_i]]}{\sqrt{\text{Var} [I_i]}}. \quad (9)$$

La varianza de  $I_i$ , asumiendo completa aleatorización, es

$$\text{Var} [I_i] = \frac{w_{i(2)} (n - b_2)}{n - 1} + \frac{2w_{i(kh)} (2b_2 - n)}{(n - 1) (n - 2)} - \left( \frac{\sum_{j=1}^n w_{ij}}{n - 1} \right)^2, \quad (10)$$

donde  $b_2 = m_4/m_2^2$  y  $m_r = \sum_{i=1}^n z^r/n$ , con  $z = (x_i - \bar{x})$ ,  $w_{i(2)} = \sum_{j=1, j \neq i}^n w_{ij}^2$  y  $2w_{i(kh)} = \sum_{k=1, k \neq i}^n \sum_{h=1, h \neq i}^n w_{ik}w_{ih}$ .

Tal como sucede con el test global  $I$  de Moran, el test  $I_i$  puede ser computado como desviaciones respecto al valor medio de la variable  $x$ .

$$I_i = z_i \sum_{j=1, j \neq i}^n w_{ij} z_j, \quad (11)$$

donde  $z_i$  es la desviación de la  $i$  -ésima observación respecto a la media,  $(x_i - \bar{x})$ .

La suma del test  $I$  local de Moran viene dada por:

$$\sum_{i=1}^n I_i = \sum_{i=1}^n z_i \sum_{j=1}^n w_{ij} z_j \quad (12)$$

y el test global  $I$  de Moran viene dado por:

$$I = \left( \frac{n}{S_0} \right) \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2}, \quad (13)$$

o

$$I = \sum_{i=1}^n I_i \left[ S_0 \left( \frac{\sum_{i=1}^n z_i^2}{n} \right) \right], \quad (14)$$

donde  $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$ .

El estadístico  $I$  de Moran permite la clasificación de las observaciones en 4 grupos diferentes considerando si el valor de  $x_i$  esta por encima de la media (H) o por debajo de la media (L) y si el valor de  $Wx_i$  esta por encima (H) o por debajo (L) de la media de  $Wx_i$ . Los grupos generalmente son representados por: (1) H-H, valores altos de  $x_i$  y altos de  $Wx_i$ ; (2) L-L, valores bajos de  $x_i$  y bajos de  $Wx_i$ ; (3) L-H, valores bajos de  $x_i$  y altos de  $Wx_i$ ; y (4) H-L, valores altos de  $x_i$  y altos de  $Wx_i$ . El diagrama de dispersión entre  $x_i$  y  $Wx_i$ , denominado diagrama de Moran, permite la visualización de los patrones espaciales representados por los 4 grupos definidos.

Alternativamente, Getis y Ord (1992) proponen dos nuevos estadísticos espaciales de dependencia local: el estadístico  $G_i$ , donde el valor de la localización  $i$  es excluido en el cómputo del estadístico, y el estadístico  $G_i^*$  donde se considera el valor de localización  $i$  – *ésima*.

La expresión del test  $G_i$  es igual a:

$$G_i = \frac{\sum_{j=1}^n w_{ij} x_j}{\sum_{j=1}^n x_j}, \quad j \neq i \quad (15)$$

donde el valor esperado, bajo el supuesto de completa aleatoriedad, depende del número de vecinos locales:

$$E(G_i) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n w_{ij}. \quad (16)$$

De manera similar, el estadístico local  $G_i^*$  es computado como:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j}{\sum_{j=1}^n x_i}, \quad (17)$$

donde el valor esperado, nuevamente bajo el supuesto de completa aleatoriedad, es:

$$E(G_i^*) = \frac{1}{n} \sum_{j=1}^n w_{ij}. \quad (18)$$

Getis y Ord (1992) y Ord y Getis (1995) proveen los valores esperados y varianzas de ambos estadísticos. Sus distribuciones se aproximan a la normal. Para la detección de clusters, es apropiado utilizar el test  $G^*$  ya que brinda la usual definición de cluster entre la relación que existe de la observación y su vecindario local.

El estadístico  $G^*$  clasifica a las observaciones en clusters de altos valores (hotspots) y de bajos valores (coldspots), difiriendo de la clasificación brindada por el estadístico local de Moran.

### 3.3. Dinámica Espacio-Temporal con Estadísticos Locales

Nuestro interés se centra en la comparación de la distribución de una variable en diferentes periodos temporales. La dificultad proviene en que el tiempo juega un rol relevante y los estadísticos presentados en las secciones 3.1 y 3.2 son de naturaleza estática. Es decir, no permiten comparaciones de distribuciones en el mismo espacio geográfico a través de diferentes periodos temporales.

Una forma pragmática para analizar el cambio espacio-temporal de una variable puede ser la siguiente: usar un estadístico local obtener el mapa de clusters en el tiempo  $t$  y en el tiempo  $t + h$ ; luego, construir una matriz de tabulación cruzada para identificar las transiciones más importantes; finalmente, aplicar algún test de independencia entre ambas clasificaciones. La matriz de tabulación cruzada (o de transición) sigue un formato como el presentado en el Cuadro 1, donde las filas muestran las categorías del tiempo  $t$  y en el tiempo  $t + h$ . La notación  $n_{ij}$  denota la número de observaciones que experimenta una transición desde la categoría  $i$  a la categoría  $j$  donde el número de categorías es igual a  $J$ . Las entradas en la diagonal indican persistencia, así  $n_{JJ}$  denota la cantidad de observaciones que muestran persistencia en la categoría  $J$ . Las entradas fuera de la diagonal indican una transición de la categoría  $i$  a la categoría diferente  $j$ . En la columna Total  $t$ , la notación  $n_{i+}$  denota la cantidad de observaciones en la categoría  $i$  en el tiempo  $t$ , siendo igual a la suma de todas las observaciones de cada categoría  $j$ :  $n_{i+} = \sum_j n_{ij}$ . Lo mismo puede decirse para la fila Total  $t + h$ .

Cuadro 2: Matriz de Tabulación Cruzada

Test LISA		Tiempo $t + h$					
		Categoría 1	...	Categoría $j$	...	Categoría $J$	Total $t$
Tiempo $t$	Categoría 1	$n_{11}$	...	$n_{1j}$	...	$n_{1J}$	$n_{1+}$
	⋮	⋮		⋮		⋮	⋮
	Categoría $i$	$n_{i1}$	...	$n_{ij}$	...	$n_{iJ}$	$n_{i+}$
	⋮	⋮		⋮		⋮	⋮
	Categoría $J$	$n_{J1}$	...	$n_{Jj}$	...	$n_{JJ}$	$n_{J+}$
	Total $t + h$	$n_{+1}$	...	$n_{+j}$	...	$n_{+J}$	$n$

En una matriz de transición como la propuesta, la hipótesis nula de independencia estadística de dos respuestas tiene la forma de:

$$H_0: p_{ij} = p_{i+}p_{+j}$$

para todo  $i$  y  $j$ . La probabilidad conjunta,  $p_{ij}$ , es igual al producto de las probabilidades marginales,  $p_{i+}$  y  $p_{+j}$

Para contrastar  $H_0$ , identificamos a  $n_{ij}^e = np_{ij} = np_{i+}p_{+j}$  como la frecuencia esperada. Entonces,  $n_{ij}^e$  es el valor esperado de  $n_{ij}$  asumiendo independencia. Dado que las frecuencias esperadas son usualmente desconocidas, estimamos las mismas utilizando las proporciones muestrales:

$$\hat{n}_{ij}^e = np_{i+}p_{+j} = n \frac{n_{i+}}{n} \frac{n_{+j}}{n} = \frac{n_{i+} \cdot n_{+j}}{n}.$$

Una vez computada la matriz de transición, es posible analizar la matriz mediante un test tipo Chi-cuadrado. El enfoque Chi-cuadrado compara la matriz observada de valores a una matriz de valores esperados que son generados por aleatoriedad. Asumiendo que el número de observaciones  $n_{i+}$  y  $n_{+j}$  viene dado a priori, la frecuencia esperada estimada para cada categoría es igual a  $n_{ij}^e = (n_{i+}n_{+j})/n$ . La fórmula del estadístico propuesto por Pearson es:

$$X^2 = \sum_{i=1}^J \sum_{j=1}^J \left\{ \frac{[n_{ij} - \hat{n}_{ij}^e]^2}{\hat{n}_{ij}^e} \right\}. \quad (19)$$

Otro estadístico alternativo es el  $G^2$ , llamado estadístico de razón de verosimilitudes Chi-cuadrado:

$$G^2 = 2 \sum_{i=1}^J \sum_{j=1}^J n_{ij} \ln \left( \frac{n_{ij}}{\hat{n}_{ij}^e} \right). \quad (20)$$

Tal como el test de Pearson,  $G^2$  alcanza un valor mínimo de 0 cuando  $n_{ij} = n_{ij}^e$ , y asume grandes valores cuando existe fuerte evidencia en contra de  $H_0$ .

#### 4. Análisis Simbólico: Estadístico $\Psi_2$ .

En esta sección se presenta un estadístico de reciente desarrollo que utiliza el análisis simbólico para detectar dependencia espacial entre variables. El test  $\Psi_2$  puede verse como una extensión del test univariante  $SG$  (López *et al.*, 2010).

Asumamos que existen dos procesos espaciales,  $\{x_s\}_{s \in S}$  e  $\{y_s\}_{s \in S}$ , donde  $S$  es un conjunto de puntos o localizaciones sobre el espacio. Con el objetivo de simbolizar la serie, tenemos que definir un conjunto finito no vacío de símbolos que sean capaces de recoger la información necesaria de cada proceso espacial. Denotaremos a este conjunto por  $\Gamma_k = \{\sigma_1, \sigma_2, \dots, \sigma_k\}$ ; y a cada uno de sus elementos  $\sigma_i$  lo llamaremos *símbolo* para  $i = 1, 2, \dots, k$ .

Simbolizar una serie es definir una función

$$f : \{x_s\}_{s \in S} \rightarrow \Gamma_k, \quad (21)$$

tal que a cada elemento  $x_s$  se le asocia un único símbolo  $f(x_s) = \sigma_{i_s}$  con  $i_s \in \{1, 2, \dots, k\}$ . Diremos que la localización  $s \in S$  es de tipo  $\sigma_i$ , con respecto a la serie  $\{x_s\}_{s \in S}$ , si y solo si  $f(x_s) = \sigma_{i_s}$ . Llamaremos a  $f$  *función de simbolización*. El mismo proceso se puede repetir para la serie  $y_s$ .

A continuación, introducimos el proceso bivalente  $\{Z_s\}_{s \in S}$  como:

$$Z_s = \{x_s, y_s\}, \quad (22)$$

donde  $x_s$  e  $y_s$  son los procesos espaciales univariantes antes definidos. Para este proceso bivalente definimos el conjunto de símbolos  $\Omega_k$  como el producto directo de los dos conjuntos  $\Gamma_k$ , es decir,  $\Omega_k^2 = \Gamma_k \times \Gamma_k$  y sus elementos son de la forma  $\eta_{ij} = (\sigma_i^x, \sigma_j^y)$ . La función de simbolización del proceso bivalente será

$$g : \{Z_s\}_{s \in S} \rightarrow \Omega_k^2 = \Gamma_k \times \Gamma_k, \quad (23)$$

definida por

$$g(Z_s = (x_s, y_s)) = (f(x_s), f(y_s)) = \eta_{ij} = (\sigma_i^x, \sigma_j^y). \quad (24)$$

Diremos que  $s$  es de tipo  $\eta_{ij}$  para  $Z = (x, y)$  o simplemente que  $s$  es de tipo  $\eta_{ij}$ , si y solo si  $s$  es tipo  $\sigma_i^x$  para  $x$  y de tipo  $\sigma_j^y$  para  $y$ .

Dependiendo del problema, pueden definirse diferentes funciones de simbolización. Ejemplos de diferentes funciones de simbolización pueden ser consultados en Matilla y Ruiz (2008, 2009), López *et al.* (2010) y Ruiz *et al.* (2009). En nuestro caso, definiremos una función de simbolización  $f$  utilizando la mediana,  $M_e^x$ , del proceso espacial univariante  $\{x_s\}_{s \in S}$ . Sea la función indicadora:

$$\tau_s = \begin{cases} 1, & \text{si } x_s \geq M_e^x, \\ 0, & \text{en cualquier otro caso.} \end{cases} \quad (25)$$

Para cada  $s \in S$ , sea  $n_s$  el conjunto formado por los  $(m - 1)$  vecinos de  $s$ . Denominaremos  $m - \text{entorno}$  al conjunto formado por cada  $s$  y el conjunto  $n_s$ , tal que el  $m - \text{entorno}$   $x_m(s) = (x_s, x_{s_1}, \dots, x_{s_{m-1}}) = (x_s, x_{n_s})$ . Con  $m \geq 2$  nos referiremos a la *dimensión de encaje*. Definimos para cada  $s_i$  con  $i = 1, 2, \dots, m - 1$  la función indicadora:

$$l_{ss_i} = \begin{cases} 0, & \text{si } \tau_s \neq \tau_{s_i}, \\ 1, & \text{en cualquier otro caso.} \end{cases} \quad (26)$$

Finalmente, podemos establecer una función de simbolización para el proceso espacial  $\{x_s\}_{s \in S}$  como  $f : \{x_s\}_{s \in S} \rightarrow \Gamma_m$ , definida por:

$$f(x_s) = \sum_{i=1}^{m-1} l_{ss_i}, \quad (27)$$

donde  $\Gamma_m = \{0, 1, \dots, m - 1\}$ .

El proceso de simbolización consiste en comparar, para cada localización  $s$ , el valor  $\tau_s$  con  $\tau_{s_i}$  recorriendo  $s_i$  el conjunto de los  $m - 1$  vecinos más próximos a la localización  $s$ . Esta simbolización nos permite capturar la información relevante del vecindario de la observación  $s$ .

Nótese que si cada proceso espacial es independiente sobre el espacio (en el sentido que su distribución espacial es aleatoria), para la función de simbolización propuesta, la probabilidad de ocurrencia de cada símbolo viene dada por  $p(\sigma) = C_{\sigma}^{m-1}/2^{m-1}$ , donde  $C_{\sigma}^{m-1} = (m-1)!/[(m-1-\sigma)!\sigma!]$  denota las combinaciones de  $m-1$  elementos tomados de  $\sigma$  en  $\sigma$  para todo símbolo  $\sigma \in \{0, \dots, m-1\}$ . Es decir, para  $m=4$ , bajo la hipótesis de independencia espacial del proceso respectivo, las frecuencias relativas esperadas para cada símbolo son:  $p(\sigma=0) = 1/8$ ,  $p(\sigma=1) = 3/8$ ,  $p(\sigma=2) = 3/8$ ,  $p(\sigma=3) = 1/8$ .

#### 4.1. Entropía: Definiciones y Conceptos

En esta sección brindamos algunos conceptos básicos de la Teoría de Información. Un tratamiento exhaustivo puede encontrarse en Cover y Thomas (1991).

El núcleo central de la Teoría de la Información es el concepto de entropía, como medida de la incertidumbre de un proceso estocástico. Sea  $x$  una variable aleatoria discreta que toma los valores  $\{x_1, x_2, \dots, x_n\}$  con probabilidades  $p(x_i)$  para cada  $i = 1, 2, \dots, n$ , respectivamente.

**Definición 1:** La entropía de Shannon,  $h(x)$ , de una variable aleatoria discreta  $x$  se define como:

$$h(x) = -\sum_{i=1}^n p(x_i) \ln(p(x_i)).$$

Usualmente, cuando la base del logaritmo es igual a 2, las unidades de medida se expresan en *bits*. Nosotros trabajaremos con la base neperiana, por lo que las unidades se expresan en *nats*. Se asume, por convención, que  $0 \ln 0 = 0$ , es decir añadir términos iguales a cero no modifica la entropía.

Sobre la base de la definición de la entropía individual, podemos considerar la entropía conjunta de un par de variables aleatorias.

**Definición 2:** La entropía  $h(x, y)$  de un par de variables aleatorias discretas  $(x, y)$  con distribución conjunta  $p(x, y)$  es:

$$h(x, y) = -\sum_x \sum_y p(x, y) \ln(p(x, y)).$$

Es necesario destacar que las medidas de entropía son funciones de la distribución de probabilidad de las variables aleatorias. Es decir, no dependen del valor que dichas variables toman en un caso particular, solo de su probabilidad. En contraste, la varianza depende de los valores que asumen las variables y es sensible a las unidades de medida.

Estos términos, asociados al concepto de entropía pueden ser adaptados al caso de la distribución de probabilidad de los símbolos computados en la sección previa. Para ello necesitamos una serie de definiciones adicionales.

Obtenida la simbolización de la serie para una dimensión de encaje  $m \geq 2$ , fácilmente se puede calcular la frecuencia absoluta y relativa de las diferentes colecciones de símbolos  $\sigma_{i_s}^x \in \Gamma_n$  y  $\sigma_{j_s}^y \in \Gamma_n$ .

Definiendo la frecuencia absoluta del símbolo  $\sigma_i^x$  como  $n_{\sigma_i^x}$  y la frecuencia absoluta del símbolo  $\sigma_j^y$  como  $n_{\sigma_j^y}$ , es posible calcular las frecuencias relativas:

$$p(\sigma_i^x) \equiv p_{\sigma_i^x} = \frac{n_{\sigma_i^x}}{|S|}, \quad (28)$$

$$p(\sigma_j^y) \equiv p_{\sigma_j^y} = \frac{n_{\sigma_j^y}}{|S|}, \quad (29)$$

donde  $|S|$  denota el cardinal del conjunto  $S$ ; en general  $|S| = n$ .

De manera similar, puede calcularse la frecuencia relativa para  $\eta_{ij} \in \Omega_n^2$ :

$$p(\eta_{ij}) \equiv p_{\eta_{ij}} = \frac{n_{\eta_{ij}}}{|S|}. \quad (30)$$

Usando estas definiciones, se desarrolla el concepto de *entropía simbólica* para una serie espacial *bidimensional*  $\{Z_s\}_{s \in S}$ . Esta entropía es la entropía de Shannon para los  $m^2$  símbolos distintos

$$h_Z(m) = - \sum_{\eta \in \Omega_m^2} p(\eta) \ln(p(\eta)). \quad (31)$$

La entropía simbólica es un indicador de la información contenida en los  $m^2$  símbolos utilizados en la simbolización.

De manera similar se pueden definir las entropías simbólicas marginales como

$$h_x(m) = - \sum_{\sigma^x \in \Gamma_m} p(\sigma^x) \ln(p(\sigma^x)), \quad (32)$$

$$h_y(m) = - \sum_{\sigma^y \in \Gamma_m} p(\sigma^y) \ln(p(\sigma^y)). \quad (33)$$

Obsérvese que las entropías marginales y la conjunta satisfacen que  $0 \leq h(m) \leq \ln(n)$ . La cota inferior se alcanza cuando sólo aparece un único símbolo y la cota superior cuando todos los símbolos tienen igual probabilidad de ocurrencia.

#### 4.2. Estadístico $\Psi_2$

Si las series son independientes, es fácil comprobar que la entropía conjunta es igual a la sumas de las marginales:

$$h_Z(m) = h_x(m) + h_y(m), \quad (34)$$

por lo que puede plantearse el siguiente estadístico  $\Psi_2 = h_x(m) + h_y(m) - h_Z(m)$  siendo la hipótesis nula:

$$H_0 : \{x_s\}_{s \in S} \text{ e } \{y_s\}_{s \in S} \text{ son independientes.} \quad (35)$$

La obtención de la distribución del estadístico se desarrolla mediante técnicas de bootstrap (pocos resultados han sido establecidos para datos espaciales; ver Cliff y Ord, 1981, o Hall, 1985, para una revisión). El enfoque sigue las líneas generales del bootstrap en bloques no traslapado de Carlstein (1986), aplicado a series temporales. La nueva propuesta, bootstrap en bloques espacial (SBB), quiebra la dependencia espacial de las series manteniendo (en parte) la estructura espacial propia de cada una de ellas. El procedimiento SBB es como sigue:

1. Dividir la serie espacial en  $b = n/l$  bloques observacionales contiguos de longitud  $l$ . Los  $b$  bloques no pueden traslaparse y agotan exhaustivamente el espacio.
2. Hay  $b$  diferentes sub-muestras de longitud  $l$  que son aleatoriamente re-muestreadas, con reemplazo, para construir una nueva serie de longitud  $n$ .
3. Cada bloques es añadido en la correspondiente área de acuerdo al ordenamiento del bootstrap. El re-muestreo ha sido completado.

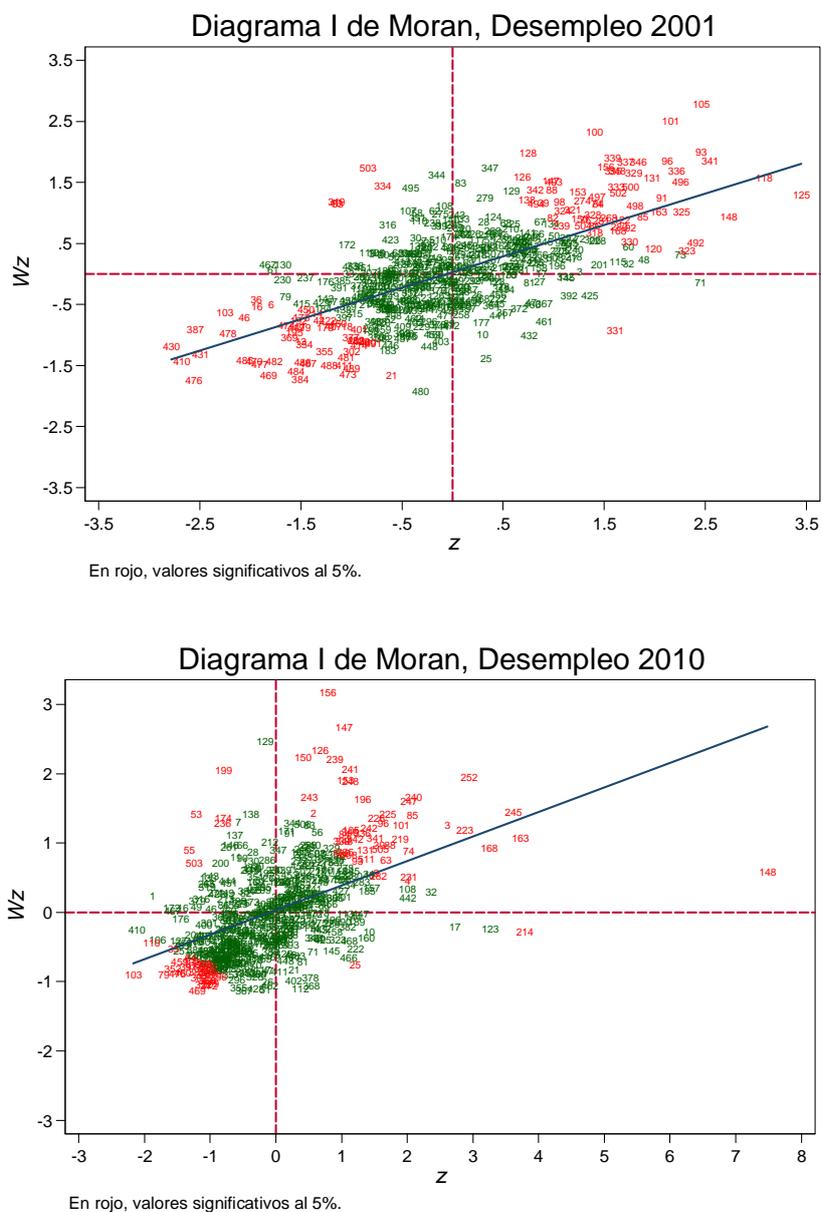
Es importante destacar que la estructura espacial interna de cada bloque, en términos de conectividad entre regiones, no cambia durante el re-muestreo. La construcción de los bloques se realiza con el criterio de distancia a  $b$  puntos definidos previamente. Estos puntos sirven de anclaje para determinar asignar las unidades al punto más próximo para formar el correspondiente bloque. Una vez realizada la partición del espacio, se procede al muestreo con reemplazo de los  $b$  bloques. Un desarrollo pormenorizado puede verse en Herrera *et al.* (2013).

## 5. Estructuras Persistentes en el Desempleo Departamental

Utilizando los estadísticos presentados en secciones anteriores, analizaremos en primera instancia la existencia o no de dependencia espacial global en cada censo. El criterio de vecindad elegido ha sido el de 4 vecinos más cercanos en todos los casos.

Los gráficos de dispersión I de Moran se presentan en la Figura 4: en el eje de las abscisas se muestra la variable tasa de desempleo estandarizada respecto a su media y desviación estándar. En el eje de las ordenadas se muestra el valor de la tasa de desempleo de los 4 vecinos más cercanos, en versión estandarizada. La pendiente de la regresión entre ambas variables nos brinda el valor de estadístico I de Moran. En ambos casos esta pendiente es positiva indicando que los departamentos tienden a comportarse de forma similar a la de sus vecinos.

Figura 4: Gráficos de Dispersión I de Moran



Cuadro 3: Test de Dependencia Espacial Globales

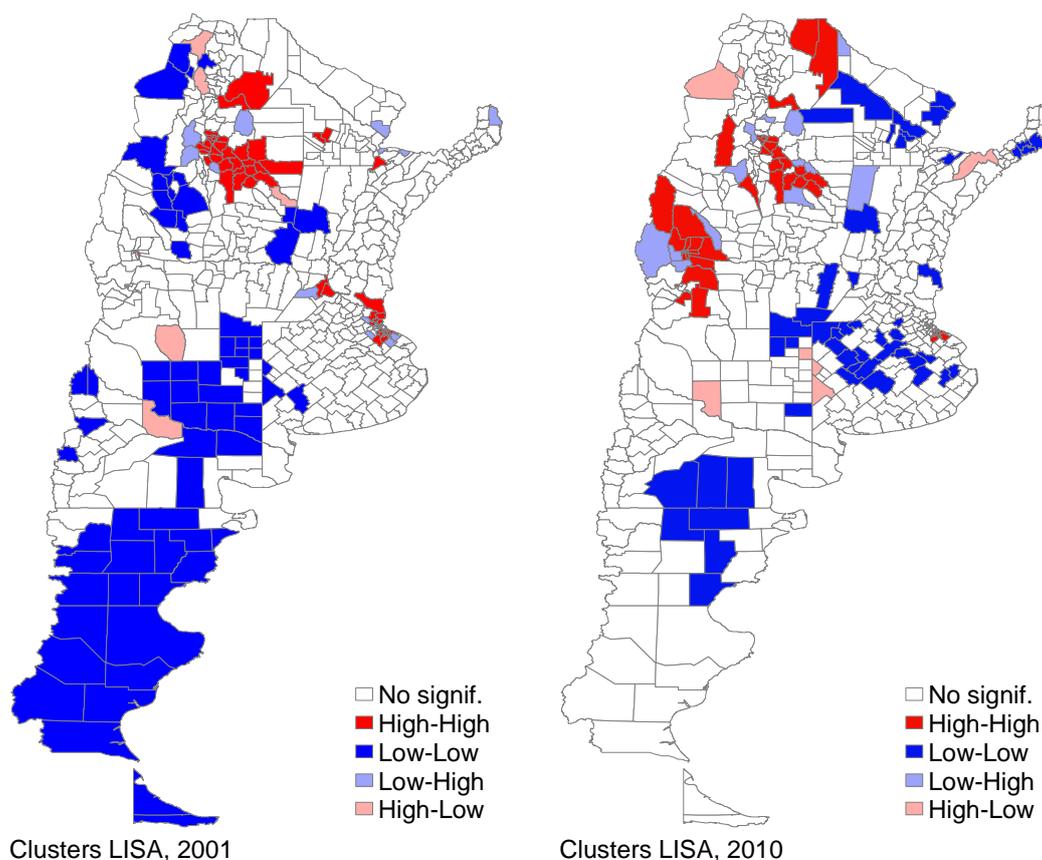
Variables	Desempleo, 2001		Desempleo, 2010	
Test	Valor	p-valor	Valor	p-valor
I de Moran	0.51	0.00	0.35	0.00
c de Geary	0.50	0.00	0.65	0.00
G de Getis-Ord	0.01	0.00	0.01	0.00

El valor del estadístico I de Moran se presenta en el Cuadro 3. Adicionalmente, el cuadro muestra cada uno de los estadísticos globales que se ha desarrollado. Para cada censo, los tres estadísticos globales rechazan la hipótesis nula de aleatoriedad indicando que existe una dependencia espacial (positiva) a nivel departamental.

Los diagramas de dispersión I de Moran incluyen la identificación de la observación por su número y se distinguen por colores. Los valores en rojo identifican los departamentos con valores significativamente diferentes de cero y sirven para construir los clusters espaciales. En verde se muestran los departamentos que no son significativos. Esta clasificación ha sido realizada mediante el estadístico I de Moran local.

Una fácil visualización de los cuatro grupos en que el estadístico local divide a las observaciones es mediante mapas (Figura 5).

Figura 5: Mapas de Clusters de Desempleo. LISA

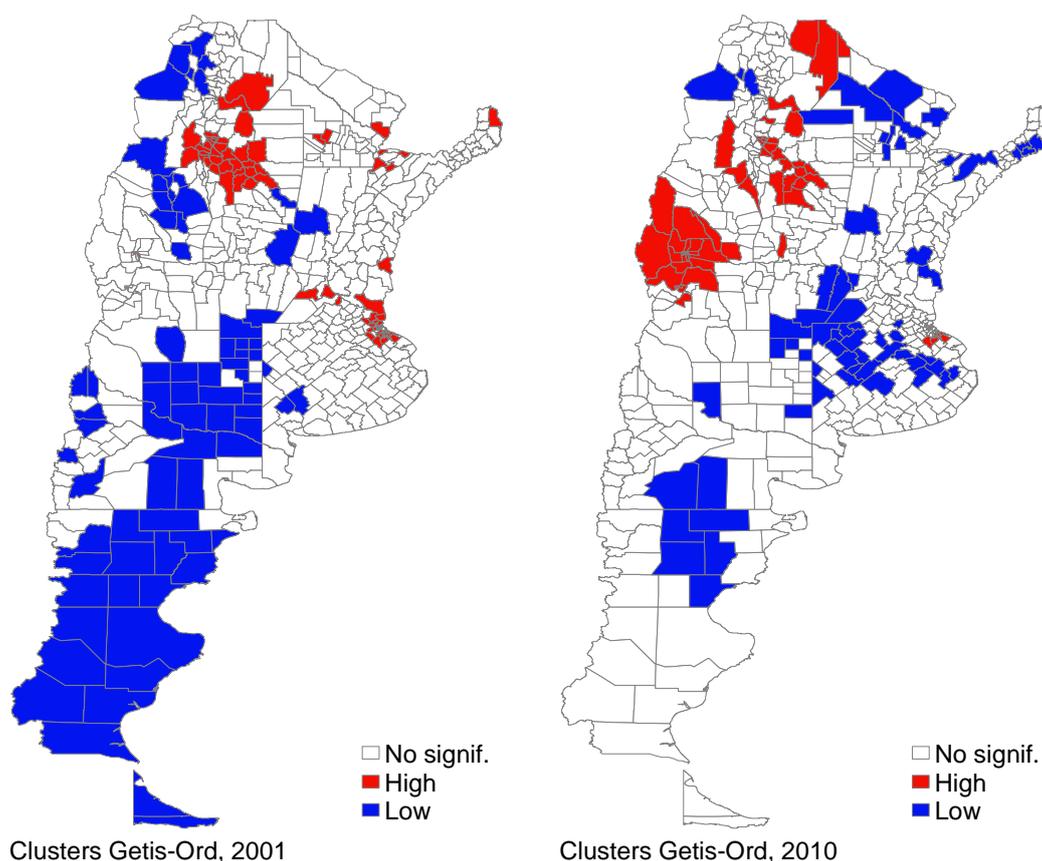


En el censo 2001, los clusters de alto valor (High-High) se ubican principalmente en el norte del país, entre Salta, Tucumán y Santiago del Estero, y en algunas zonas cercanas al conurbano bonaerense. Para el censo 2010, la distribución de altos valores no se encuentra tan concentrada en

los regiones mencionadas y surge un cluster importante en la región de Cuyo. Los clusters de bajo valor, Low-Low, se distribuyen principalmente en el sur y centro del país. Para el censo 2010, la cantidad de clusters bajos se incrementa en la zona pampeana y se reduce en el sur.

Mapas con similar información pueden ser construidos utilizando el estadístico  $G^*$  de Getis-Ord. En la Figura 6, se presentan los mapas para cada censo. En este caso se tienen solamente dos tipos de clusters.

Figura 6: Mapas de Clusters de Desempleo.  $G^*$  de Getis-Ord



La información que arroja el estadístico local de Getis-Ord se encuentra en consonancia con el I de Moran local. En el censo 2001, los valores de elevado desempleo se concentran principalmente en el norte del país y los clusters de bajo desempleo en el sur y centro de Argentina. Para el censo 2010, surge con notoriedad un cluster de alto valor en la zona cuyana.

Nuestra finalidad es detectar si existe persistencia en los clusters entre ambos censos. Para ello procedemos a realizar la matriz de transición considerando la clasificación realizada por los estadísticos locales. El Cuadro 4 muestra la frecuencia absoluta de cada categoría según el I de Moran.

El Cuadro 4 permite ver claramente la transición entre clusters. La cantidad de departamentos H-H disminuye entre censos, pasando de 59 a 49 en el 2010. La cantidad de observaciones L-L también disminuye entre ambos censos. Sin embargo, 36 departamentos muestran persistencia en los clusters H-H (26) y L-L (10). Ninguna región con desempleo alto y vecinos con desempleo alto (H-H), en el censo 2001, ha pasado a la categoría de desempleo bajo con vecinos de desempleo bajo (L-L). Solo dos departamentos del censo 2001 logran disminuir significativamente su nivel de desempleo a pesar de que su entorno es de alto desempleo (ver la entrada H-H (2001) y L-H (2010)).

Cuadro 4: Comparación de los Resultados Censales

<i>I de Moran Local</i>		Censo 2010					Total
		No sig.	H-H	L-L	L-H	H-L	
Censo 2001	No sig.	292	22	40	10	1	365
	H-H	31	26	0	2	0	59
	L-L	50	0	10	0	5	65
	L-H	14	1	0	1	0	16
	H-L	5	0	0	0	0	5
	Total	392	49	50	13	6	510

La aplicación de los estadísticos de independencia intertemporal arroja los siguientes resultados,  $X^2(16) = 130,5$  y  $G^2(16) = 102,1$ , con los grados de libertad entre paréntesis. En ambos casos se rechaza la independencia de las distribuciones de clusters.

Construyendo la matriz de transición con la clasificación de Getis-Ord, podemos obtener el Cuadro 5. Los departamentos con altos valores han tendido de disminuir entre censos, pasando de 76 a 62. Lo mismo puede afirmarse sobre la cantidad de departamentos con bajos valores (se pasa de 75 a 63). Pero, nuevamente, existe un núcleo duro de departamentos que persisten con valores elevados en ambos censos (29) y otro núcleo que permanece en la categoría de bajo valor (17). Hay un único caso que cambia su posición entre las categorías alto y bajo. Este caso es un departamento de La Rioja que deja de pertenecer al cluster Low para ubicarse en un entorno de elevado desempleo (High).

Cuadro 5: Comparación de los Resultados Censales

<i>Test G* Getis-Ord</i>		Censo 2010			
		No sig.	High	Low	Total
Censo 2001	No sig.	281	32	46	359
	High	47	29	0	76
	Low	57	1	17	75
	Total	385	62	63	510

Los resultados de los tests de independencia entre ambos censos son  $X^2(4) = 70,8$  y  $G^2(4) = 68,4$ . En ambos casos se rechaza la independencia de las distribuciones de clusters.

El último enfoque que aplicaremos será el de entropía simbólica que permite comparar la dependencia de la distribución espacial del desempleo en ambos censos, considerando la existencia de la dependencia espacial en cada uno de ellos. Para ofrecer un mayor entendimiento de lo que el test  $\Psi_2$  detecta, se ha procedido a obtener una distribución alternativa del desempleo departamental en el censo 2010.

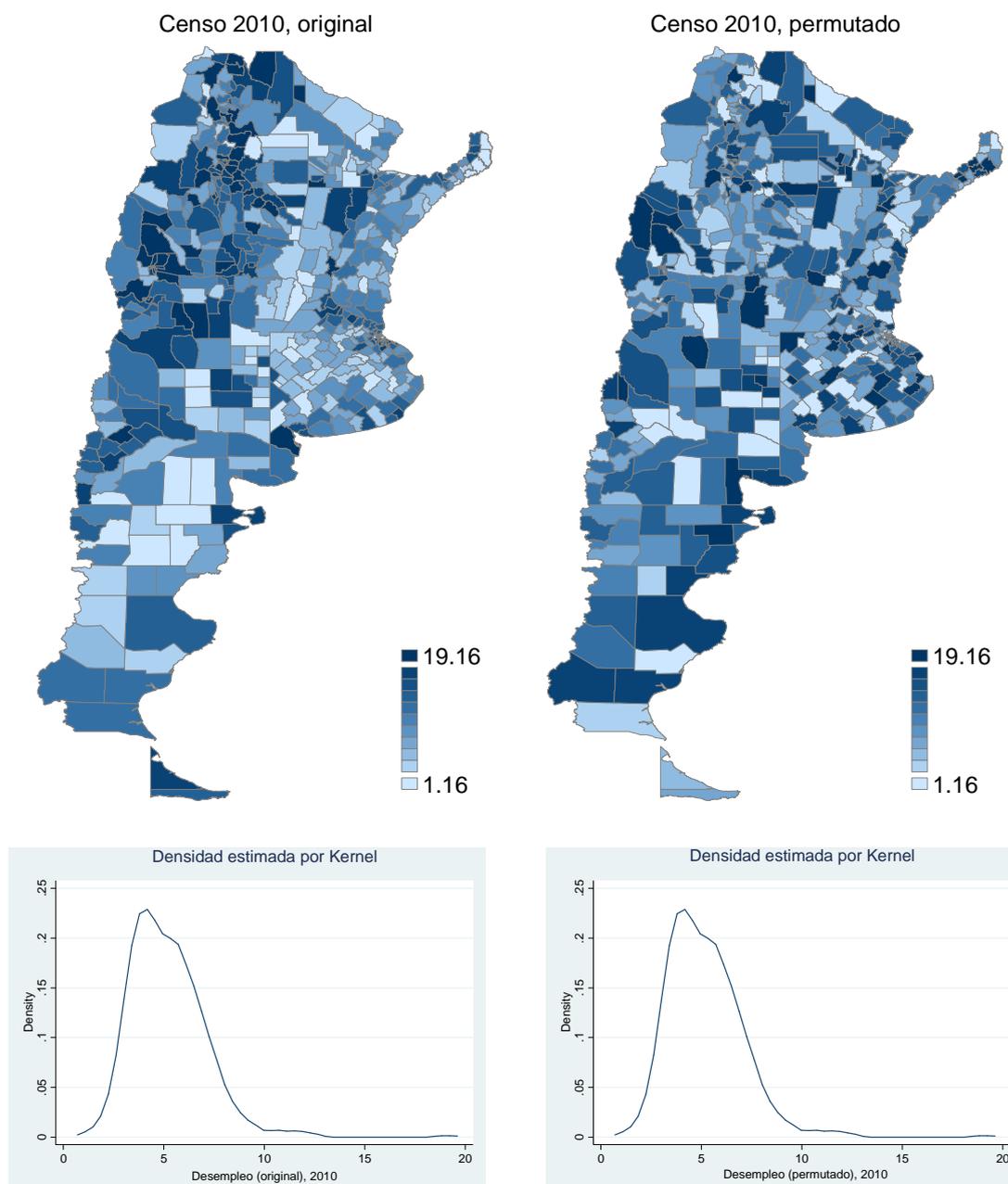
En la Figura 7, se muestra la distribución observada del desempleo departamental y una distribución alternativa obtenida por permutación de los valores departamentales. La permutación fue realizada tal que permanezca cierto grado de dependencia espacial entre los departamentos pero, sin que la distribución no georeferenciada se modifique (estimaciones de la densidad al final de la Figura 7).

El estadístico  $\Psi_2$  contrasta la siguiente hipótesis nula:

$$H_0 : \{y_s\}_t \text{ e } \{y_s\}_{t+j} \text{ son independientes,} \quad (36)$$

donde  $\{y_s\}$  representa a la tasa de desempleo en el periodo  $t = 2001$  y  $t + j = 2010$ .

Figura 7: Distribución del Desempleo Original y Permutado, Censo 2010



Aplicando el test entre las distribuciones alternativas se obtienen los resultados del Cuadro 6.

Cuadro 6: Comparación de los Resultados Censales

Test de Independencia entre Mapas			
Test $\Psi_2$	Valor	p-valor	Conclusión
Desemp. 2010 (original) - Desemp. 2001	0,068	0,002	Mapas dependientes
Desemp. 2010 (permut.) - Desemp. 2001	0,011	0,835	Mapas independientes

Nota: W-4nn, 10 bloques, 399 boots.

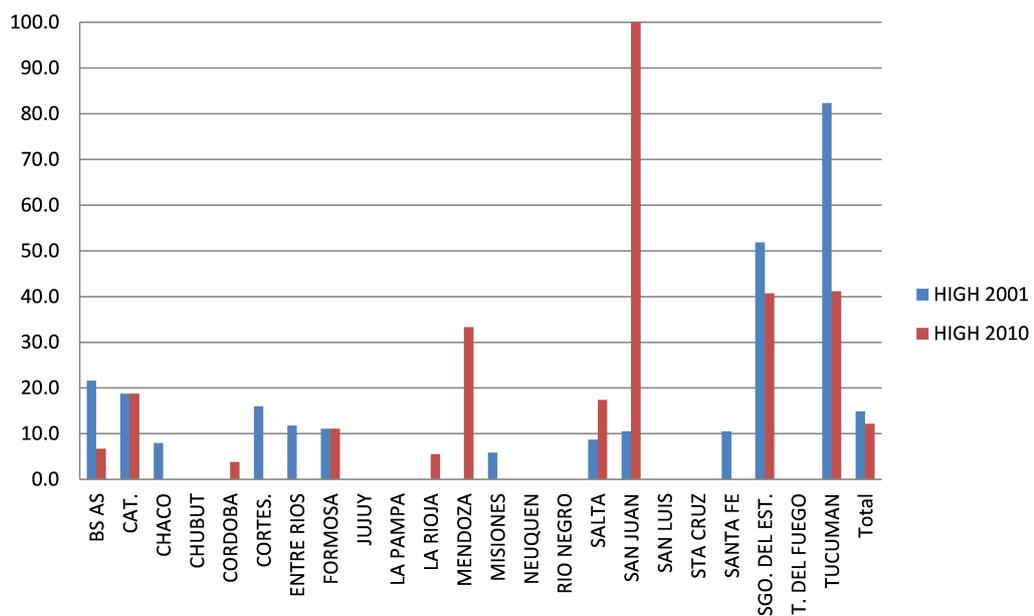
La conclusión a la que se arriba es la existencia de dependencia entre ambos censos y por consiguiente estructuras espaciales que persisten. Adicionalmente, el estadístico  $\Psi_2$  no puede rechazar

la distribución alternativa (permutada) siendo este el resultado esperado si el desempleo no presentase persistencia entre periodos temporales.

Finalmente, puede ser de interés aproximarse al nivel de agregación de los anteriores estudios. Es decir, ver el comportamiento que se obtiene a nivel provincial. Es posible que exista alguna provincia o conjunto de las mismas con un alto porcentaje de departamentos con alta o baja incidencia en ambos censos.

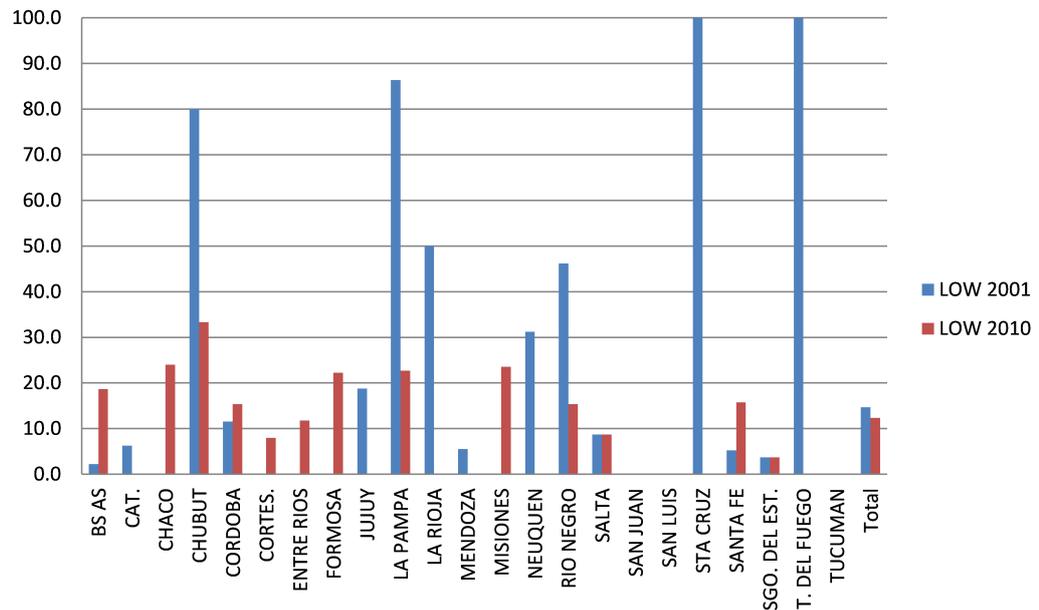
La Figura 8 muestra la proporción de departamentos de cada provincia que pertenece a clusters de alto valor según Getis-Ord. Las barras azules representan el porcentaje para el censo 2001 y las de marrón claro el porcentaje para el censo 2010. Para 8 provincias, el porcentaje de departamentos con alto valor decrece entre censos. Otras 10 provincias permanecen sin cambios entre los años 2001 y 2010. Solo unas pocas provincias pasan a tener un mayor porcentaje de departamentos con alto desempleo (Córdoba, La Rioja, Mendoza, Salta y San Juan). El caso más llamativo es el de San Juan con la totalidad de sus departamentos con altos valores. Es decir, a pesar del descenso generalizado en todo el país, la mejora relativa de los departamentos de San Juan no ha sido tal y presentan valores relativamente elevados respecto a sus vecinos más próximos.

Figura 8: Provincias según (%) de Departamentos con Clusters High Getis-Ord.



La Figura 9 presenta similar información que la Figura 8 pero para los clusters de bajo valor. En 11 provincias se experimenta una disminución de departamentos de bajo desempleo. Esto puede verse como un reflejo de la convergencia a nivel nacional, en donde las bajas tasas no se ajustan tan fuertemente. Las provincias que han ganado más departamentos con bajo desempleo son Buenos Aires, Chaco, Córdoba, Corrientes, Entre Ríos, Formosa, Misiones y Santa Fe. San Luis es un caso particular, ya que ninguno de sus departamentos puede ser clasificado como de alto o bajo desempleo en ambos censos. Las provincias que han visto empeorado sus situación son principalmente Santa Cruz y Tierra del Fuego que tenían el 100 % de sus departamentos como de bajo desempleo y en el 2010 no tienen ninguno dentro de este cluster.

Figura 9: Provincias según (%) de Departamentos con Clusters Low Getis-Ord.



## 6. Conclusiones

Este trabajo ha pretendido analizar la evolución espacio-temporal del desempleo departamental. Este nivel de desagregación no es habitualmente investigado. A su vez, el análisis se basa en herramientas de econometría espacial que han sido poco utilizadas para examinar la persistencia del desempleo regional.

Nuestro país ha experimentado una notoria recuperación en términos de desempleo. Pero, a pesar de la marcada disminución en prácticamente todos los departamentos, se detecta dependencia espacial global en cada censo. Esto implica que el desempleo posee una naturaleza geográfica significativa. A su vez, los estadísticos locales detectan clusters espaciales en cada censo y cierta persistencia de clusters de alto y bajo valor. El uso de matrices de transición intertemporales permite afirmar que la dependencia espacial es persistente. Los estadísticos de Pearson y de Razón de Verosimilitudes rechazan la hipótesis nula de independencia entre categorías de clusters. El estadístico  $\Psi_2$ , basado en análisis simbólico, confirma que ambas distribuciones son dependientes: la dependencia espacial persiste en el tiempo.

El comportamiento a nivel provincial arroja cierta evidencia de una mayor variabilidad del desempleo entre provincias. Esta conclusión avala los resultados obtenidos por estudios que trabajan a nivel provincial. Esto no contradice que existan estructuras sub-provinciales persistentes.

Es importante mencionar que los resultados presentados se basan en un análisis espacial univariante y deben tomarse como preliminares. En la medida de que se cuente con mejor información departamental pueden intentarse líneas de investigación alternativas. De todas formas, la intención del presente trabajo ha sido llamar la atención de estructuras persistentes y presentar herramientas poco difundidas en economía que permiten manejar el problema de una forma alternativa. La línea de investigación se encuentra abierta y en desarrollo.

## Referencias

- [1] Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.
- [2] Anselin, L. (1995). "Local Indicators of Spatial association-LISA," *Geographical Analysis*, 27, pp. 93-115.
- [3] Anselin, L. y A. Bera (1998). "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," en A. Ullah y D. Giles (eds.), *Handbook of Applied Economic Statistics*. New York: Marcel Dekker.
- [4] Bertola G. e I. Ichino (1995). "Wage Inequality and Unemployment: United States versus Europe," NBER Macroeconomics Annual 1995, 10, pp. 13-66. National Bureau of Economic Research, Inc.
- [5] Blanchard, O. y L. Katz (1992). "Regional Evolutions," *Brookings Papers on Economic Activity*, 1, pp.1-75.
- [6] Boots, B. (2003). "Developing Local Measures of Spatial Association for Categorical Data," *Journal of Geographical Systems*, 5, pp. 139-160.
- [7] Carlstein, E. (1986). "The Use of Subseries Methods for Estimating the Variance of a General Statistic from a Stationary Time Series," *The Annals of Statistics*, 14, pp. 1171-1179.
- [8] Cliff, A. D. y J. K. Ord (1969). "The Problem of Spatial Autocorrelation," en A. J. Scott (ed.), *London Papers in Regional Science 1*, Studies in Regional Science. London: Pion.
- [9] Cliff, A. D. y J. K. Ord (1973). *Spatial Autocorrelation*. London: Pion.
- [10] Cliff, A. D. y J. K. Ord (1981). *Spatial Processes, Models and Applications*. London: Pion.
- [11] Cressie, N. (1993). *Statistics for Spatial Data*. Revised edition. New York: Wiley.
- [12] Decressin, J. y A. Fatás (1995). "Regional Labor Market Dynamics in Europe," *European Economic Review*, 39, pp. 1627-1655.
- [13] Díaz Cafferata, A. y A. Figueras (1996). "Dimensión Espacial de la Desocupación en Argentina. Alternativas de Regionalización," Documento de Trabajo, AAEP.
- [14] Elhorst, P. (2003). "The Mystery of Regional Unemployment Differentials: Theoretical and Empirical Explanations," *Journal of Economic Surveys*, 17, pp. 709-748.
- [15] Faini, R., Galli, P. y F. Rossi (1997). "An Empirical Puzzle: Falling Migration and Growing Unemployment Differentials Among Italian Regions," *European Economic Review*, 41, pp. 571-579.
- [16] Galiana, S., Lamarche, C., Porto, C. y W. Sosa Escudero (2004). "Persistence and Regional Disparities in Unemployment (Argentina 1980–1997)," *Regional Science and Urban Economics*, 35, pp. 375-394.
- [17] Geary, R. (1954). "The Contiguity Ratio and Statistical Mapping," *The Incorporated Statistician*, 5 (3), pp. 115-145.
- [18] Getis, A. (1995). "Spatial Filtering in a Regression Framework: Experiments on Regional Inequality, Government Expenditures, and Urban Crime," pp. 172-188, en Anselin, L. y R. Florax (eds). *New Directions in Spatial Econometrics*. Springer, Berlin, Germany
- [19] Getis, A. y J. Ord (1992). "The Analysis of Spatial Association by Use of Distance Statistics," *Geographical Analysis*, 24, pp. 189-206.

- [20] Getis, A. y J. Ord (1996). "Local Spatial Statistics: An Overview," en Longley, P. y M. Batty (eds), *Spatial Analysis: Modelling in a GIS Environment*. Cambridge: GeoInformation International.
- [21] Griffith, D. (2000). "A Linear Regression Solution to the Spatial Autocorrelation Problem," *Journal of Geographical Systems*, 2, pp. 141-156.
- [22] Hall, P. (1985). "Resampling a Coverage Pattern," *Stochastic Processes and Their Applications*, 20, pp. 231-246.
- [23] Herrera, M., Ruiz, M. y J. Mur (2013). "Detecting Dependence between Spatial Processes," *Spatial Economic Analysis*, forthcoming.
- [24] Jimeno, F. y S Bentolila (1998). "Regional Unemployment Persistence (Spain, 1976-1994)," *Labour Economics*, 5, pp. 25-51.
- [25] Lamarche, C., Porto, A. y W. Sosa Escudero (1998). "Aspectos Regionales del Desempleo en la Argentina", Documento de Trabajo N° 8, Depto. de Economía, Facultad de Cs. Es., Universidad Nacional de La Plata.
- [26] Lesage, J. y K. Pace (2009). *Introduction to Spatial Econometrics*. Boca Raton: CRC Press.
- [27] Molho, I. (1995). "Spatial Autocorrelation in British Unemployment," *Journal of Regional Science*, 35, pp. 641-658.
- [28] Moran, P. (1948). "The Interpretation of Statistical Maps," *Journal of the Royal Statistical Society B*, 10, pp. 243-251.
- [29] Moran, P. (1950). "Notes on Continuous Stochastic Phenomena," *Biometrika*, 37 (1), pp. 17-23.
- [30] Ord, J. y A. Getis (2001). "Testing for Local Spatial Autocorrelation in the Presence of Global Autocorrelation," *Journal of Regional Science*, 41, pp. 411-432.
- [31] Overman, H. y D. Puga (2002). "Unemployment Clusters across Europe's Regions and Countries," *Economic Policy*, Abril, pp. 117-147.
- [32] Ruiz, M., López, F y A. Páez (2012). "Comparison of Thematic Maps using Symbolic Entropy," *International Journal of Geographical Information*, 26(3), pp. 413-439.
- [33] Taylor, J. y S. Bradley (1997). "Unemployment in Europe: A Comparative Analysis of Regional Disparities in Germany, Italy and the U.K.," *Kyklos*, 50, pp. 221-245.