

ESTRUCTURA ESPACIAL DE LA TASA DE DESEMPLEO: UNA APROXIMACIÓN.

MONTERO LORENZO, JOSE M^a

LARRAZ IRIBAS, BEATRIZ

Universidad de Castilla – La Mancha

Jose.MLorenzo@uclm.es

Beatriz.Larraz@uclm.es

Palabras clave: I de Moran, semivariograma, tasa de paro, estadística espacial.

Resumen:

La literatura económica es rica en estudios relativos al mercado de trabajo, dada la indudable importancia económica y social de sus implicaciones. Por ello, la problemática del mercado laboral ha sido abordada desde muy diversas perspectivas, si bien son muy escasos los trabajos relativos a la estructura espacial de sus principales magnitudes.

En esta comunicación se aborda una primera tentativa de estimación de la estructura espacial de la tasa de paro mediante la combinación de semivariogramas. El objetivo que pretende dicha comunicación es determinar si existe una estructuración espacial de la tasa de desempleo en nuestro país.

1. Introducción.

Resulta evidente que existe abundante literatura, tanto teórica como empírica, acerca del mercado de trabajo, en todas y cada una de sus múltiples vertientes, pero raramente se han llevado a cabo análisis de las principales macromagnitudes laborales desde la perspectiva espacial y nunca desde la óptica de la geoestadística. Sin embargo, creemos que esta perspectiva puede resultar realmente interesante y por ello, en este trabajo, abordamos una tentativa del estudio de la tasa de paro, en el mercado de trabajo español, –una de las variables más significativas, más popular y con mayores connotaciones sociales- desde esta perspectiva. En concreto, el estudio se llevará a cabo sobre la tasa de paro procedente de la Encuesta de Población Activa, siendo el ámbito geográfico provincial la península, no considerándose las situaciones de Baleares, Canarias y Ceuta y Melilla. El ámbito temporal se refiere a los años naturales 2001 y 2002, ya que el cambio metodológico implantado en la Encuesta en el ejercicio 2001 impide el análisis para una referencia más larga.

El objetivo concreto que se persigue, en caso de evidencia de existencia de dependencia espacial entre las tasas de paro provinciales, es la identificación del semivariograma que ajusta la relación espacial existente entre las mismas. Creemos que éste sería un buen punto de partida para iniciar una fructífera línea de investigación. Lamentablemente, el cambio metodológico anteriormente aludido impide cualquier análisis de tipo evolutivo que, sin duda, hubiese sido de interés. No obstante, la determinación de la estructura espacial actual de la tasa de paro en el mercado de trabajo español ya constituye por sí misma un objetivo interesante.

Somos conscientes de los problemas que entraña el llevar a cabo un análisis de este tipo en el mercado laboral a nivel provincial, por cuanto la tasa de paro representa a toda la provincia y, sin embargo, puede ser diferente en cada una de las localizaciones que componen la misma. No obstante, la alternativa, dadas las disponibilidades estadísticas actuales, sería no realizar nunca análisis de este tipo.

El trabajo se estructura en las siguientes partes: En primer lugar, abordaremos la cuestión de la existencia de dependencia espacial entre las tasas de paro provinciales españolas, utilizando para ello el contraste I de Moran. Posteriormente, trataremos de

identificar la estructura de dicha dependencia mediante el ajuste del semivariograma empírico de la tasa de paro a uno de los semivariogramas teóricos recogidos en la literatura de geoestadística espacial. A continuación se procederá a establecer la validación del semivariograma propuesto, finalizando con las principales conclusiones del trabajo.

El programa informático utilizado en el análisis geoestadístico de la información ha sido ISATIS, elaborado por GEOVARIANCES & ECOLE DES MINES DE PARIS.

2. Contraste de autocorrelación espacial en la tasa de paro española.

La primera cuestión que debemos abordar en un estudio como el que pretendemos realizar es si existe o no correlación espacial entre las tasas de paro de las provincias españolas. Aunque existen varios procedimientos para contrastar la existencia de correlación espacial en la literatura estadística (entre los cuales los más populares son los proporcionados por Moran, Geary y Getis y Ord), nos apoyaremos en el propuesto por Moran (Moran, 1948), cuyo estadístico de contraste,

$$I = \frac{n \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i \sum_j w_{ij} \sum_i (x_i - \bar{x})^2}$$

para muestras grandes¹ y bajo la hipótesis de muestreo aleatorio, tiene como esperanza y varianza (Cliff y Ord,1981),

$$E(I) = \frac{-1}{n-1}$$

$$V(I) = \frac{n[(n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2] - k[n(n-1)S_1 - 2nS_2 + 6S_0^2]}{(n-1)(n-2)(n-3)S_0^2} - \frac{1}{(n-1)^2}$$

donde

$$k = \frac{m_4}{m_2^2} \text{ y } m_r = \frac{1}{n} \sum_i (x_i - \bar{x})^r$$

$$S_0 = \sum_i \sum_j w_{ij} \quad (i \neq j)$$

$$S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2 \quad (i \neq j)$$

¹ Para un análisis más detallado de esta cuestión véase, por ejemplo, Upton, G. Y Fingleton, B. (1985).

$$S_2 = \sum_i (w_{i.} + w_{.j})^2; \quad w_{i.} = \sum_j w_{ij}; \quad w_{.j} = \sum_i w_{ij};$$

siendo la distribución del estadístico estandarizado asintóticamente normal (0;1). En nuestro caso particular, el criterio de contigüidad física es el de vecindad si las provincias i,j comparten frontera, por lo que w_{ij} , tomará el valor 1 en caso de vecindad y 0 en caso contrario (Anexo 1).

En el caso que nos ocupa, y prefijando un nivel de significación del 5%, el estadístico I de Moran, sin estandarizar, para el ejercicio 2001 toma un valor de 0,44 mientras que para el año 2002 se eleva a 0,48. Dado que el valor esperado del estadístico bajo la hipótesis de no correlación espacial es $-0,02$ y su desviación típica es 0.01, el estadístico I de Moran estandarizado toma el valor 46.05 para el primer año y 50.01 para el segundo, lo que implica el rechazo de la hipótesis nula al nivel de significación prefijado del 5%.

Gráficamente, podemos apreciar la existencia de dicha correlación espacial en los scatterplot de Moran correspondientes a los ejercicios 2001 y 2002

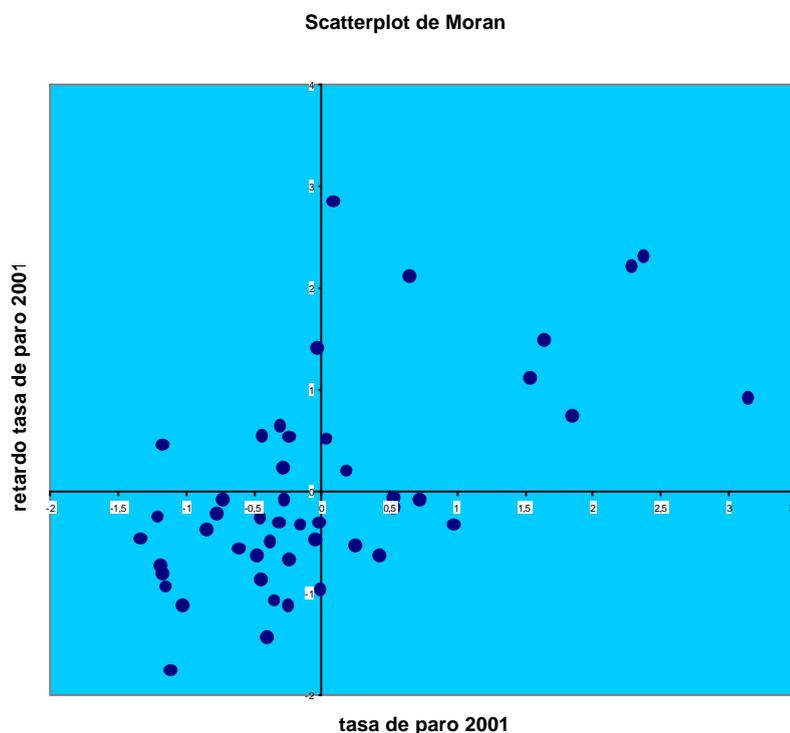


Gráfico 1: Scatterplot de Moran año 2001

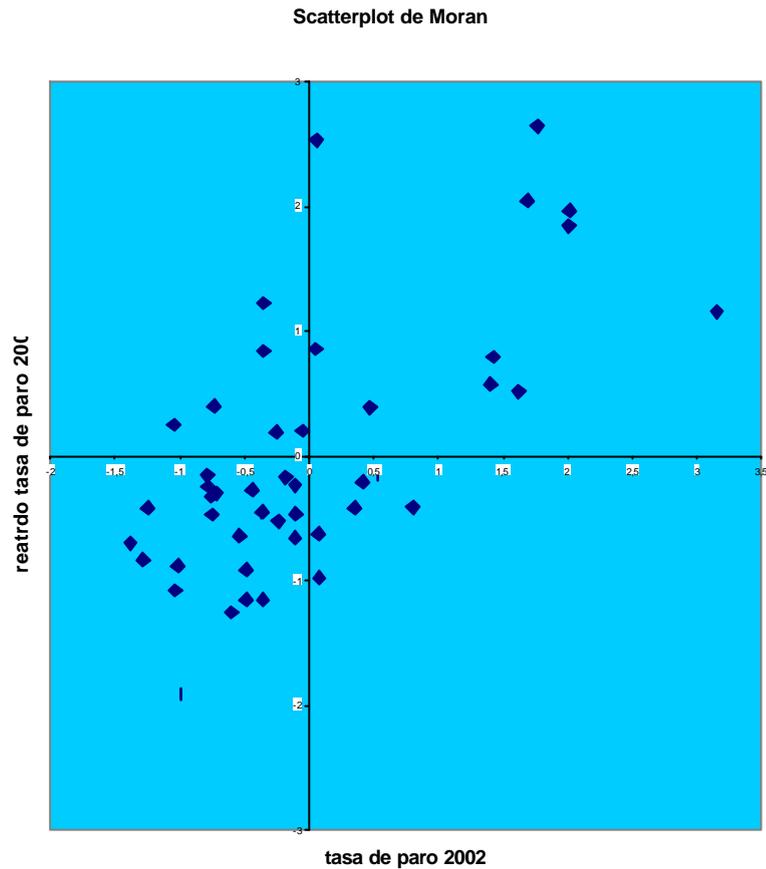


Gráfico 2: Scatterplot de Moran año 2002

3. Determinación de la estructura de dependencia espacial de la tasa de paro española.

Constatada la existencia de correlación espacial, en este epígrafe pasamos a determinar su estructuración mediante la estimación del semivariograma de la tasa de paro provincial española para los ejercicios 2001 y 2002.

Los mapas de base, que representan la magnitud de la tasa de paro mediante un símbolo “+” proporcional a la misma pueden verse a continuación en los Gráficos 3 y 4:

Gráfico 3: Mapa de base año 2001

Gráfico 4: Mapa de base año 2002

La nube variográfica, que representa la mitad de los cuadrados de las diferencias entre las tasas de paro de cada dos provincias que distan \mathbf{h} ,

$$\mathbf{g}^{**}(\mathbf{h}) = \frac{1}{2} (x(\mathbf{s}_a + \mathbf{h}) - x(\mathbf{s}_a))^2$$

y el semivariograma experimental omnidireccional,

$$\mathbf{g}^*(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{a=1}^{N(\mathbf{h})} (x(\mathbf{s}_a + \mathbf{h}) - x(\mathbf{s}_a))^2$$

que representa para cada distancia \mathbf{h} , el promedio de los puntos de la nube, estimado mediante el procedimiento clásico del método de los momentos, pueden verse en los Gráficos 5 y 6.

Gráfico 5: Nube de puntos y semivariograma experimental año 2001
Observación: 1cm = 60 km

Gráfico 6: Nube de puntos y semivariograma experimental año 2002
Observación: 1cm = 60 km

En ellos se puede apreciar que, a grandes distancias, los semivariogramas empíricos no se estabilizan en torno a la varianza estimada del proceso estocástico. Ello puede ser debido a la existencia de deriva o bien a que a grandes distancias la estimación del semivariograma pierde fiabilidad, debido a que, dada la configuración geográfica de nuestro país, el número de parejas de provincias susceptibles de utilización en la estimación del semivariograma a distancias superiores a 480 km es escaso. No obstante, como lo realmente importante es que el semivariograma esté estructurado a pequeñas y medianas distancias, se puede adoptar el supuesto de que el proceso generador de los datos es estacionario de segundo orden o, al menos, intrínsecamente estacionario.

A la vista de los semivariogramas experimentales para 2001 y 2002 se ha realizado un ajuste siguiendo las directrices del modelo lineal de regionalización, que consiste en la combinación lineal de semivariogramas básicos de meseta unidad. En nuestro caso particular se ha procedido a la combinación de tres semivariogramas básicos, tanto en el año 2001 como en el año 2002:

Tipo	Expresión	Meseta	Alcance
Efecto pepita puro	$g(h) = \begin{cases} 0 & h = 0 \\ m & h > 0 \end{cases}$	4	
Gaussiano	$g(h) = \begin{cases} 0 & \text{si } h = 0 \\ m \left[1 - e^{-\frac{h^2}{a^2}} \right] & \text{si } h > 0 \end{cases}$	15.9	570 km
Esférico	$g(h) = \begin{cases} 0 & \text{si } h = 0 \\ \frac{m}{2} \left[3\left(\frac{h}{a}\right) - \left(\frac{h}{a}\right)^3 \right] & \text{si } 0 \leq h \leq a \\ m & \text{si } h = a \end{cases}$	2	570 km

Tabla 1: Combinación de semivariogramas básicos correspondiente al año 2001, donde m representa la meseta y a el alcance.

Tipo	Expresión	Meseta	Alcance
Efecto pepita puro	$g(h) = \begin{cases} 0 & h = 0 \\ m & h > 0 \end{cases}$	3	
Gaussiano	$g(h) = \begin{cases} 0 & \text{si } h = 0 \\ m \left[1 - e^{-\frac{h^2}{a^2}} \right] & \text{si } h > 0 \end{cases}$	22	570 km
Esférico	$g(h) = \begin{cases} 0 & \text{si } h = 0 \\ \frac{m}{2} \left[3\left(\frac{h}{a}\right) - \left(\frac{h}{a}\right)^3 \right] & \text{si } 0 \leq h \leq a \\ m & \text{si } h = a \end{cases}$	2	480 km

Tabla 2: Combinación de semivariogramas básicos correspondiente al año 2002, donde m representa la meseta y a el alcance.

La combinación de semivariogramas resultante puede verse en los Gráficos 7 y 8.

Gráfico 7: Semivariograma ajustado correspondiente al año 2001
Observación: 1cm = 60 km

Gráfico 8: Semivariograma ajustado correspondiente al año 2002
Observación: 1cm = 60 km

4. Validación cruzada.

Llegados a este punto, y en aras de determinar la utilidad o no del semivariograma obtenido en cuanto a su capacidad para recoger la correlación espacial existente entre las tasas de paro de las provincias peninsulares españolas, procedemos a aplicar el método de la validación cruzada.

Para ello, dado el supuesto de que el proceso estocástico generador de los datos es estacionario de segundo orden o intrínsecamente estacionario, llevamos a cabo un krigado ordinario sobre todas los puntos del espacio peninsular excepto uno: aquél en el que se desea evaluar el error de estimación a través del semivariograma propuesto. Este procedimiento se aplicará a todas y cada una de las provincias peninsulares, utilizando el criterio de vecindad única y considerándose robusta la estimación obtenida si el error estandarizado asociado está situado en el intervalo $[-\mathbf{a}, \mathbf{a}]$, donde \mathbf{a} , en nuestro caso es 2.5. Evidentemente, el semivariograma propuesto será validado, y por tanto, útil para la estimación de la tasa de paro de una localización a partir de las demás, si el porcentaje de estimaciones exitosas es suficientemente elevado.

El estimador de krigado que utilizaremos es

$$X^*(\mathbf{s}_0) = \sum_{i=1}^n \mathbf{I}_i X(\mathbf{s}_i)$$

donde los ponderadores \mathbf{I}_i se obtienen imponiendo al error de estimación las condiciones de sesgo nulo y mínima varianza. La condición de sesgo nulo se traduce en que la suma de los ponderadores sea unitaria (Véase Wackernagel 1995) mientras que la minimización de la varianza del error de estimación, sujeta a la condición de insesgadez, nos llevaría a las ecuaciones de krigado ordinario

$$\sum_{j=1}^n \mathbf{I}_j \mathbf{g}(\mathbf{s}_i - \mathbf{s}_j) + \mathbf{a} = \mathbf{g}(\mathbf{s}_i - \mathbf{s}_0), \forall i = 1, \dots, n$$

$$\sum_{i=1}^n \mathbf{I}_i = 1$$

de las que se obtienen los ponderadores \mathbf{I}_i buscados resolviendo la ecuación matricial

$$\hat{\mathbf{U}} = \tilde{\mathbf{A}}^{-1} \cdot \tilde{\mathbf{A}}_0,$$

siendo

$$\mathbf{\hat{U}} = \begin{pmatrix} \mathbf{I}_1 \\ \vdots \\ \mathbf{I}_n \\ 1 \end{pmatrix}, \quad \tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{g}(\mathbf{0}) & \cdots & \mathbf{g}(\mathbf{s}_1 - \mathbf{s}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{g}(\mathbf{s}_n - \mathbf{s}_1) & \cdots & \mathbf{g}(\mathbf{0}) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \text{ y } \tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{g}(\mathbf{s}_1 - \mathbf{s}_0) \\ \vdots \\ \mathbf{g}(\mathbf{s}_n - \mathbf{s}_0) \\ 1 \end{pmatrix}$$

Los resultados obtenidos en el procedimiento de validación cruzada se muestran en los gráficos 9 y 10, habiéndose obtenido 44 y 45 estimaciones robustas en los respectivos años, de las 47 provincias analizadas.

Gráfico 9: Mapa básico en el que se muestran con puntos las estimaciones no robustas, nube de puntos de los datos observados frente a los valores estimados, histograma de los errores de estimación estandarizados y nube de puntos de los errores de estimación estandarizados frente a los valores estimados relativos al ejercicio 2001.

Gráfico 10: Mapa básico en el que se muestran con puntos las estimaciones no robustas, nube de puntos de los datos observados frente a los valores estimados, histograma de los errores de estimación estandarizados y nube de puntos de los errores de estimación estandarizados frente a los valores estimados relativos al ejercicio 2002.

5. Conclusiones.

Del análisis realizado se pueden extraer tres conclusiones relevantes en relación al comportamiento espacial de la tasa de paro en nuestro país.

La primera de ellas es que la tasa de paro provincial en nuestro país está correlacionada espacialmente, tal y como se desprende del valor calculado del estadístico I de Moran para los años 2001 y 2002.

En segundo lugar, es posible obtener la estructura espacial de dicha dependencia, si bien no es obvia y requiere la combinación de tres semivariogramas básicos.

En tercer lugar, al menos en los dos últimos años, la estructura de la dependencia espacial de la tasa de paro española es la misma, con ligeras variaciones en el rango y la meseta de los semivariogramas básicos combinados.

Por último, mediante el procedimiento de validación cruzada, se confirma la elección de los semivariogramas teóricos dado el alto porcentaje de estimaciones robustas.

Por consiguiente, la componente espacial debería ser tratada como un factor más a tener en cuenta en el estudio de las disparidades espaciales de la tasa de desempleo, por cuanto existe, puede ser modelizada y este modelo parece mantenerse en el tiempo. Evidentemente, de las conclusiones extraídas, se deduce que el componente espacial de la tasa de paro debería ser un factor importante a considerar en el análisis del mercado de trabajo, cosa que no ocurre en la actualidad.

Referencias

- Cliff, A. D., Ord, J. K., (1981)** “*Spatial and temporal analysis: autocorrelation in space and time*”. Quantitative Geography: A British View. Routledge and Kegan Paul: London.
- Cressie, N. (1993).** “*Statistics for spatial data*”. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc.
- Chica Olmo J., (1994)** “*Teoría de las Variables Regionalizadas. Aplicación en Economía Espacial y Valoración Inmobiliaria*”. Universidad de Granada.
- David, M. (1988)** “*Handbook of Applied Advanced Geostatistical Ore Reserve Estimation*”. Developments in Geomathematics 6. Elsevier Science Publishers B. V.
- Davis J.C., (1986)** “*Statistics and Data Analysis in Geology*”. John Wiley & Sons
- Gámez Martínez M., (1998)** “*Nuevas técnicas de Estadística Espacial para la Economía. Modelización del Precio de la Vivienda en la Ciudad de Albacete*” . Tesis Doctoral. Universidad de Castilla – La Mancha.
- Geary, R.C. (1954)** “*The contiguity ratio and statistical mapping*”. The incorporated Statistician, nº 5, pp 115-145
- Isaaks, E. H., Srivastava R. M., (1989)** “*An Introduction to Applied Geostatistics*”. Oxford University Press.
- Journel, A. G., (1989)** “*Fundamentals of Geostatistics in Five Lessons*”. Short Course in Geology: Volume 8. American Geophysical Union.
- Journel, A. G., Huijbregts, Ch. J., (1978)** “*Mining Geostatistics*”. Academic Press.
- Matheron, G. (1970)** “*La Théorie des variables régionalisées, et ses applications*”. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau.
- Moran, P. (1948)** “*The interpretation of statistical maps*”. Journal of the Royal Statistical Society, Series B, nº 10, pp 243-251
- Samper Calvete, F.J., Carrera Ramírez, J. (1996).** “*Geoestadística. Aplicaciones a la hidrología subterránea*”. CIMNE.
- Upton G., Fingleton B., (1985)** “*Spatial Data Analysis by Example*”. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons.
- Villaverde, J., Maza, A., (2003)** “*Desigualdades regionales y dependencia espacial en al Unión Europea*” CLM.Economía, nº 2. Asociación de Economía de Castilla-La Mancha.
- Wackernagel H., (1995)** “*Multivariate Geostatistics. An Introduction with Applications*”. Springer – Verlag.

