

PRIMER PARCIAL

Juan David Gonzalez Hernandez - 20151025060

Raúl Camilo Martín Bernal - 20151025909

Junio 2020

1. Los datos que se muestran en la tabla a continuación se refieren a algunos datos económicos regionales en el Reino Unido. Informan el Valor Agregado Bruto (GVA, por sus siglas en inglés) regional como un porcentaje del total del país, la productividad laboral (reportado a un total de 100 países) y la tasa de natalidad empresarial.

Country	Region	GVA(% of UK)	Labor Productivity (UK) =100	Business Birth Rate (%)
Wales		3,6	81,5	9,3
Scotland Northern		8,3	96,9	10,9
Ireland		2,3	82,9	6,5
England	North of England	3,2	86,2	11,2
England	North West England	9,5	88,6	11,1
England	Yorkshire & Humberside	6,9	84,7	10,5
England	East Midlands	6,2	89,2	10,3
England	West Midlands	7,3	89,1	10,5
England	East Anglia	8,7	96,8	10,5
England	Greater London	21,6	139,7	14,6
England	South East England	14,7	108,3	10,8
England	South West England	7,7	89,8	9,6

Tabla 1: Valor Agregado Bruto (GVA) en el Reino unido

Dadas las 12 regiones del Reino Unido informadas en la Tabla anterior (archivo UK.N1.shp):

a) Cree el archivo *.gal y obtenga la matriz de pesos. Considere a Irlanda del Norte como adyacente a Escocia y Gales b) Calcule los valores rezagados espacialmente de la variable GVA (valor agregado bruto) para las 12 regiones del Reino Unido c) Dibuje el diagrama de dispersión de Moran para la variable GVA, ¿Qué tipo de conocimiento puedes derivar de él? d) Estimar el modelo GVA en función de la productividad laboral y la tasa de natalidad comercial y pruebe la presencia de autocorrelación

espacial entre los residuos usando el I de Moran. ¿Podemos aceptar la hipótesis de ausencia de autocorrelación espacial de los residuos?

Solución

a)

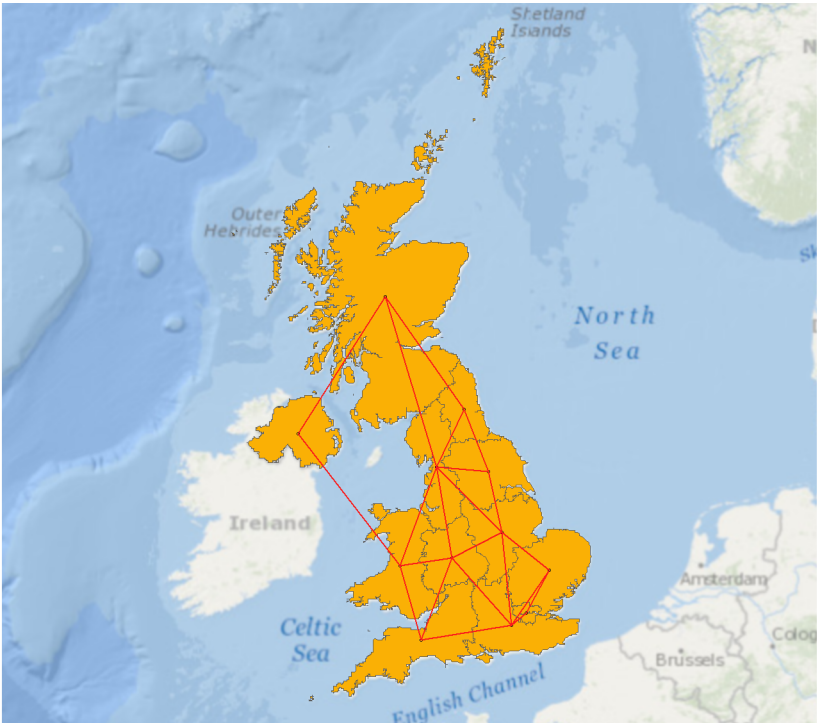


Figura 1: Conectividad UK

	N England	N W England	Yorkshire&Humberside	E Midlands	W Midlands	E Anglia	Greater London	S E England	S W England	Wales	Scotland Northern	Ireland
N England	0	1	1	0	0	0	0	0	0	0	1	0
N W England	1	0	1	1	1	0	0	0	0	1	1	0
Yorkshire&Humberside	1	1	0	1	0	0	0	0	0	0	0	0
E Midlands	0	1	1	0	1	1	0	1	0	0	0	0
W Midlands	0	1	0	1	0	0	0	1	1	1	0	0
E Anglia	0	0	0	1	0	0	1	1	0	0	0	0
Greater London	0	0	0	0	0	1	0	1	0	0	0	0
S E England	0	0	0	1	1	1	1	0	1	0	0	0
S W England	0	0	0	0	1	0	0	1	0	1	0	0
Wales	0	1	0	0	1	0	0	0	1	0	0	1
Scotland Northern	1	1	0	0	0	0	0	0	0	0	0	1
Ireland	0	0	0	0	0	0	0	0	0	1	1	0

Tabla 2: Matriz W para UK

b)

	Region	GVA	WGVA
0	Scotland	8.30	5.00
1	Northern Ireland	2.30	5.95
2	Wales	3.60	6.70
3	South West (England)	7.70	8.53
4	East Midlands (England)	6.20	9.42
5	Yorkshire and The Humber	6.90	6.30
6	East of England	8.70	14.17
7	West Midlands (England)	7.30	8.34
8	London	21.60	11.70
9	South East (England)	14.70	10.30
10	North East (England)	3.20	8.23
11	North West (England)	9.50	5.92

Tabla 3: Variables rezagadas espacialmente para UK

c)

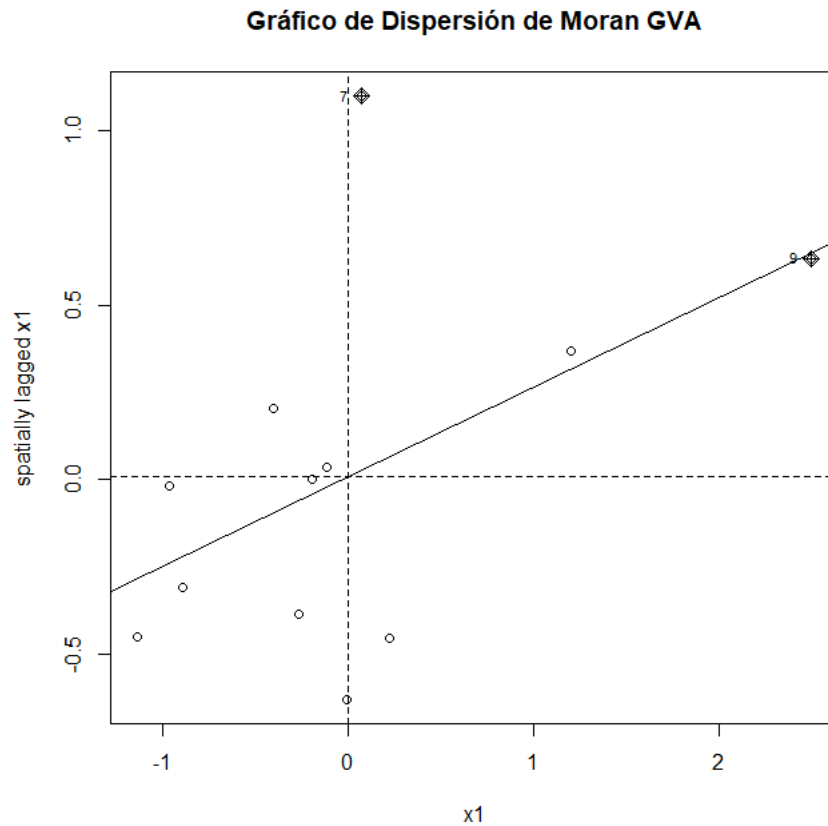


Figura 2: Moran UK

Se obtiene un I de moran de 0.256, con un $Z(I)=2,1918$ y un $P\text{-value}=0,02537$.

El I de moran en este caso corresponde a la pendiente de la recta ya que la variable GVA se encuentra estandarizada, además se concluye que existe una autocorrelacion espacial positiva esto quiere decir que el valor agregado bruto (GVA) para las diferentes regiones de reino unido se encuentra rodeado de valores similares de GVA.

Se observa que el mayor numero de datos se encuentran el tercer cuadrante y para estas regiones se concluyen valores agregados brutos bajos, rodeados de GVA bajos

d)

	Modelo
(Intercept)	-22,31*** (3,39)
Lbr_Prd	0,28*** (0,05)
Bsn_B_R	0,42 (0,47)
R ²	0.91
Adj. R ²	0.89
Num. obs.	12
RMSE	1.79
*** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$	
I de Moran de los residuos	
Indice	-0.174894
E(I)	-0.09090909
Z(I)	-0.4663
p-valor	0.641

Tabla 4: Resumen del modelo y resultados test I de Moran

Se acepta la hipótesis nula de una ausencia de autocorrelacion espacial para los residuos dado un P-value=0.641 >0.05

2. A partir de los datos del ejercicio anterior

- Realice la prueba de ausencia de autocorrelación espacial bivariada usando el I de Moran Bivariado, a partir de las variables GVA vs productividad laboral, y GVA vs tasa de natalidad comercial
- Construya correlogramas bivariados y dispersogramas bivariados de Moran.
- Obtenga para cada una de las variables mencionas los estadísticos LISA con sus probabilidades asociadas. Interprete los resultados obtenidos, usando la matriz de pesos espaciales de criterio K=3 vecinos más cercanos.

Solución

a) Podemos observar que en los resultados del I de moran bivariado no se rechaza la hipótesis de ausencia de autocorrelación para la Variable de tasa de natalidad comercial por lo tanto no hay autocorrelación en esa variable pero si se rechaza la hipótesis con la variable productividad laboral por lo tanto hay presencia de auto correlación espacial entre GVA y la productividad laboral.

	Observado (Y)	Esperado	p-value
GVA Labor Productivity	0,02741	-0,08619	0,037
GVA Business Birth Rate	0,09105	-0,07211	0,273

Tabla 5: Tasa de mortalidad infantil

b) Segun los correlogramas se observa que en los dos variables en todos los ordenes no hay autocorrelación espacial significativa, en los dispersogramas se pueden observar que no siguen mucho la tendencia aunque se puede observar un poco de tendencia en el sector de bajos rodeados de bajos

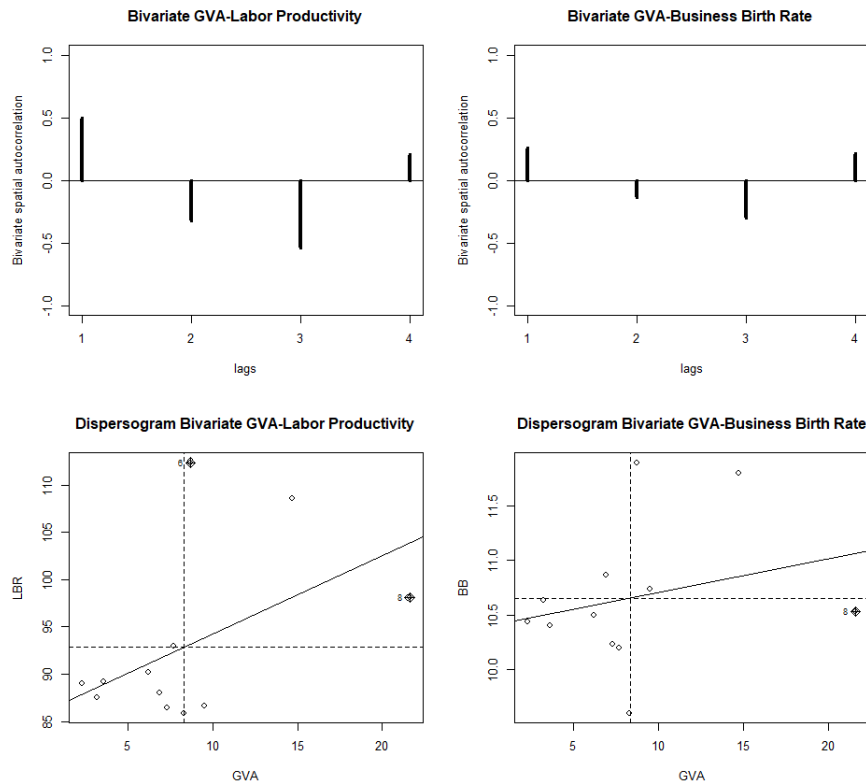


Figura 3: Bivariado

c) En los mapas de cluster se observa como solo en las variables de GVA y productividad laboral se tiene que hay una region con valores altos rodeados de altos pero no se presenta como un cluster, ademas no es tan significativo por lo que podemos interpretar que a nivel local tampoco se ven mucho las autocorrelaciones espaciales

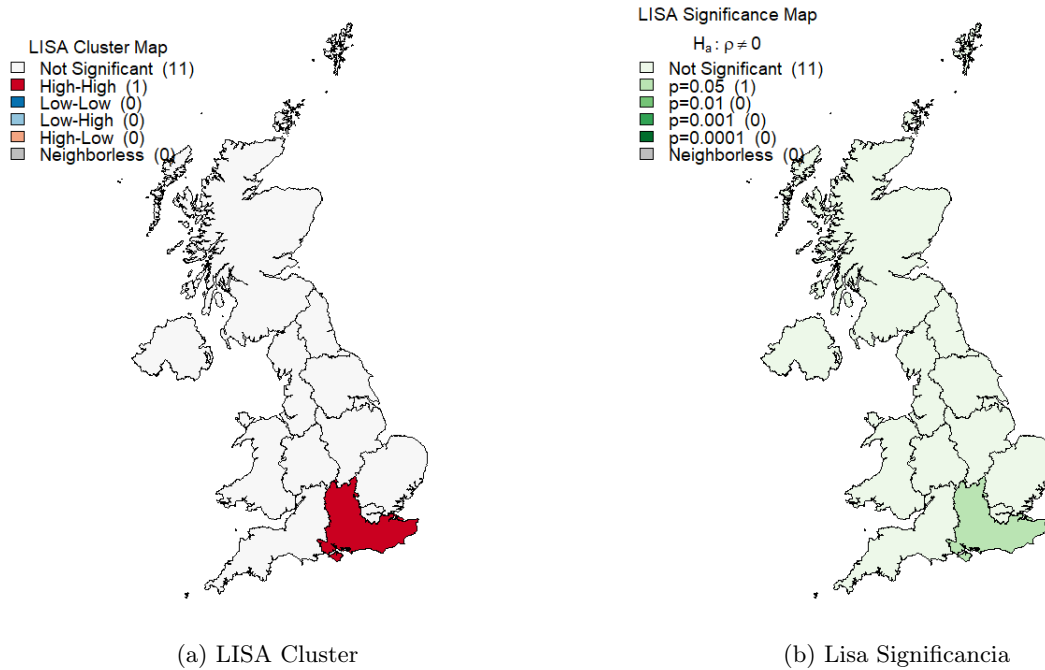


Figura 4: Resultado LISA GVA(% of UK)

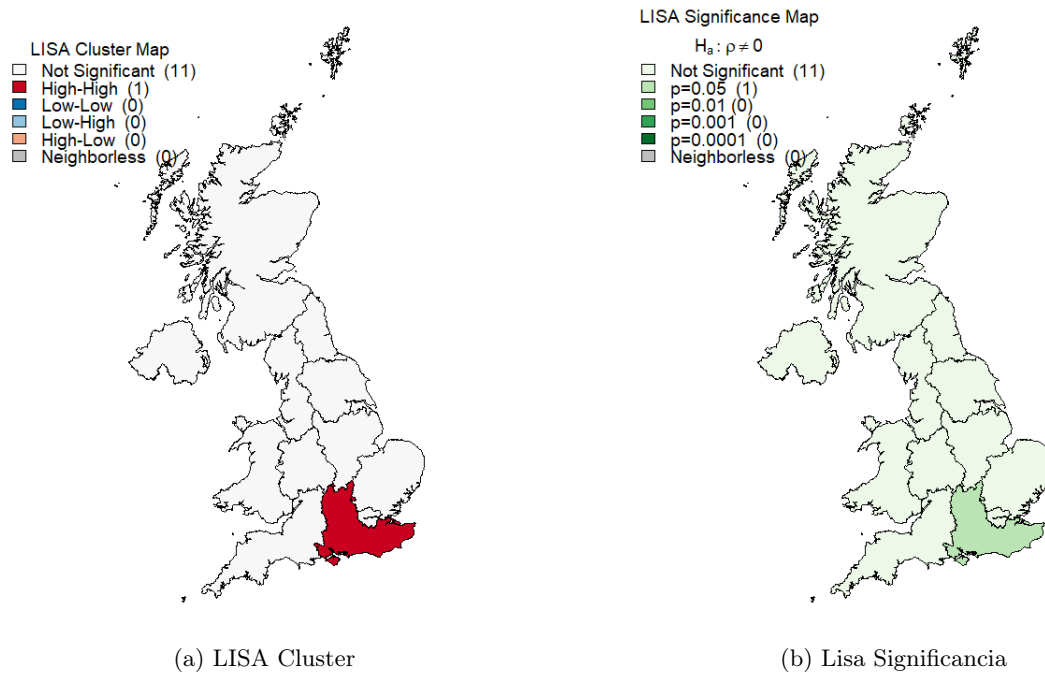


Figura 5: Resultado LISA Labor Productivity (UK) =100

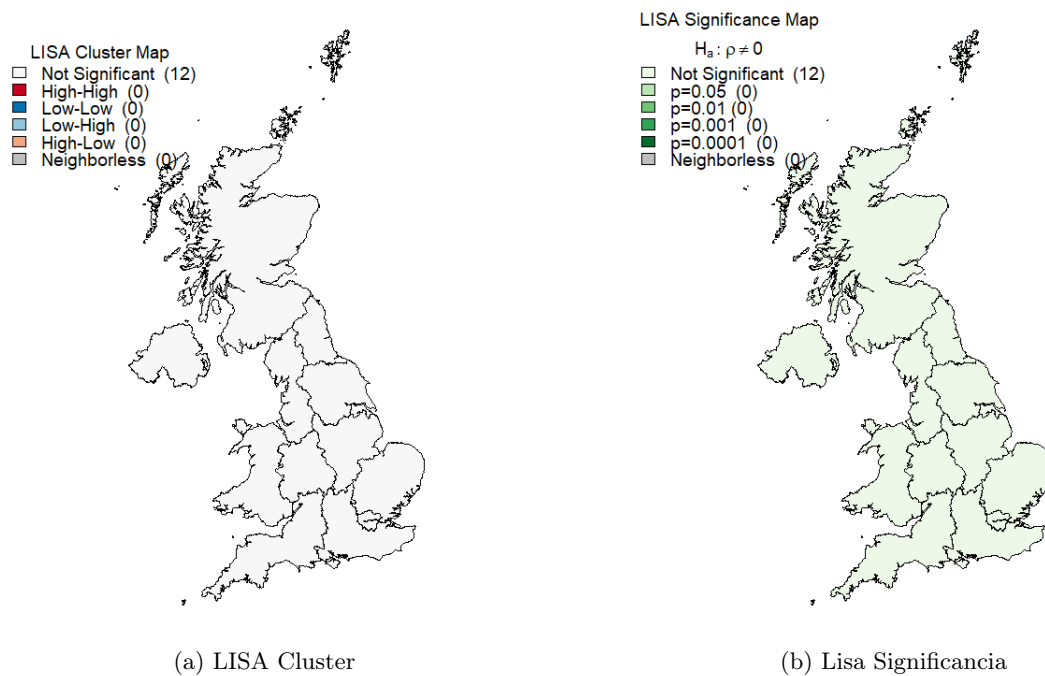


Figura 6: Resultado LISA Labor Business Birth Rate (%)

3) A partir de la base de datos “beveridge” suministrada, construya la matriz de pesos espaciales por

contigüidad reina orden 1 para dar respuesta a lo siguiente:

Matriz W contigüidad efecto reina orden 1

$$W = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix} \quad (1)$$

a) Estime por ML los modelos: clásico, SLX, SLM y SEM para la curva de Beveridge e interprete la significancia y ajuste de los diferentes modelos.

b) Especifique las ecuaciones e interprételas.

Modelo clásico:

$$u = 13,5 - 2,250v + \varepsilon \quad (2)$$

Modelo SLX:

$$u = 18,1538 - 2,0989v - 2,4176Wv + \varepsilon \quad (3)$$

Modelo SLM:

$$u = 10,40665 + 0,32533Wu - 2,16189v + \varepsilon \quad (4)$$

Modelo SEM:

$$u = 14,04108 - 2,49617v + \varepsilon \rightarrow \varepsilon = -1,1408W\varepsilon + u \quad (5)$$

Los modelos muestran como la variable Vacantes es significativa para explicar la variable de desempleo, ya que según teoría se puede entender que entre menos vacantes va a haber mas desempleo y si es significante para la estimación del desempleo.[2]

c) Construya una tabla de estadísticos , AIC y BIC para los modelos anteriores: Clásico, SLX, SLM y SEM, y explique el resumen de estadísticas para seleccionar el modelo.

El estadístico AIC (Criterio de Información de Akaike) puede aplicarse a varias clases de modelo ajustados según el criterio de máxima verosimilitud, el modelo se calcula con la formula:

$$AIC = \frac{1}{n\hat{\sigma}^2}(RSS + 2d\hat{\sigma}^2) \quad (6)$$

Para modelos ajustados por mínimos cuadrados, el estadístico BIC (Criterio de Información Bayesiano) aumenta la penalidad en modelos con muchas variables, el estadístico BIC se calcula con la ecuación:

$$BIC = \frac{1}{n}(RSS + \log(n)d\hat{\sigma}^2) \quad (7)$$

Para valores pequeños de BIC y AIC corresponde a un test error bajo, por lo que el mejor modelo será aquel con menor BIC Y AIC.[1]

Tabla 6: Tabla estadísticos AIC y BIC

	<i>Clasico</i>	<i>SLX</i>	<i>SLM</i>	<i>SEM</i>
AIC.	18.75097	17.82302	20.40991	19.2533
BIC.	17.57929	16.26077	18.84161	17.69105

d) ¿Qué modelo de regresión seleccionarías para modelar la curva de Beveridge?

Por Criterio de Información Bayesiano y Criterio de Información de Akaike se elije el modelo SLX a pesar de que la significancia en los parámetros es mucho mas alta (1 %) en los modelos SLM y SEM, además en comparación con el modelo clásico tomando como parámetro de comparación el R cuadrado ajustado concluimos que es mejor el modelo SLX.

Tabla 7: Resultados modelos

	<i>Dependent variable:</i>			
	<i>CLASICO</i>	<i>SLX</i>	<i>SLM</i>	<i>SEM</i>
v	-2.250**	-2.099*	-2.162***	-2.496***
Constant	13.500***	18.154**	10.407***	14.041***
ρ			0.32533	
λ				-1.1408
θ		-2.418		
R^2	0.844	0.913		
Adjusted R^2	0.792	0.826		
Nagelkerke pseudo- R^2			0.854	0.884
Log Likelihood			-6.205	-5.627
σ^2			0.675	0.383
Akaike Inf. Crit.	18.751	17.823	20.410	19.253
Bayesian Inf. Crit.	17.579	16.260	18.814	17.691
Residual Std. Error	1.118 (df = 3)	1.022 (df = 2)		
F Statistic	16.200** (df = 1; 3)	10.495* (df = 2; 2)		
Wald Test (df = 1)			0.727	6.842***
LR Test (df = 1)			0.341	1.498

Note:

*p<0.1; **p<0.05; ***p<0.01

Referencias

- [1] Cristina Gil Martínez. “TÉCNICAS DE REGULARIZACIÓN Y SELECCIÓN DEL MEJOR MODELO”. En: (2018). URL: https://rpubs.com/Cristina_Gil/Regularizacion_Seleccion.
- [2] Paula Nicole Roldán. “Curva de Beveridge”. En: (). URL: <https://economipedia.com/definiciones/curva-de-beveridge.html>.