

Modelos de interacción espacial con efectos aleatorios: una aplicación a los viajes por trabajo en Montevideo

María Eugenia Riaño, Fernando Massa, Antonio Rey y Martín Hansz

Grupo de Investigación de Estadística Espacial Aplicada, Instituto de Estadística,
Departamento de Métodos Cuantitativos

21 de mayo 2019

Modelo de Transporte de Cuatro Etapas

La estructura general de este modelo se basa en cuatro grandes etapas que representan la decisión de las personas de trasladarse de un lugar a otro:

- ▶ Generación: número de viajes producido y atraído por las zonas.
- ▶ Distribución: desde dónde y hacia dónde se dirigen los viajes.
- ▶ Elección de modo: en qué se realiza el viaje, por ejemplo transporte público o privado.
- ▶ Asignación: elección de la ruta de viaje.

El desarrollo de este trabajo se enmarca en la etapa de distribución, en donde el objetivo es la predicción de los flujos en una matriz de Origen - Destino (OD).

Modelos Gravitacionales

Modelo Gravitacional sin interacciones

Tradicionalmente, el análisis de las matrices OD se realiza con los denominados modelos gravitacionales.

$$y_{ij} = A_i B_j X_{o_i}^{\beta_o} X_{d_j}^{\beta_d} c_{ij}^{\phi} \quad (1)$$

- ▶ y_{ij} es la cantidad de viajes del origen i al destino j .
- ▶ A_i y B_j son factores de balance. Son coeficientes que ajustan las estimaciones a los totales de las marginales de la matriz OD.
- ▶ X_{o_i} y X_{d_j} son características asociadas al origen y al destino respectivamente.
- ▶ c_{ij}^{γ} es la función de impedancia o disuasión, donde c_{ij} es una medida del costo (distancia, tiempo) entre el origen i y el destino j .
- ▶ β_o , β_d y ϕ son los parámetros a estimar.

Modelo Gravitacional sin interacciones

Dadas n regiones, tendremos $N = n^2$ combinaciones de posibles flujos entre estas regiones. La matriz OD es:

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{pmatrix}$$

Se asume que el flujo en el par i, j es independiente del flujo en el par k, l .

Vector de Flujos

Existen dos formas de ordenar los valores de la matriz en un vector: centrado en el origen o centrado en el destino. El vector centrado en el origen tiene la siguiente forma:

$l(o)$	$o(o)$	$d(o)$
1	1	1
\vdots	\vdots	\vdots
n	1	n
\vdots	\vdots	\vdots
$N - n + 1$	n	1
\vdots	\vdots	\vdots
N	n	n

Variables Regresoras

Sean X_1, X_2, \dots, X_k variables sobre n regiones que pretenden explicar los flujos entre las regiones, y $\mathbf{X} = (X_1, X_2, \dots, X_k) \in \mathbb{R}^{n \times k}$ la matriz que recoge estas características.

La matriz $\mathbf{X}_o \in \mathbb{R}^{N \times k}$ se construye de la siguiente manera:

$$\mathbf{X}_o = \mathbf{X} \otimes \mathbf{I}_n = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nk} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

La matriz $\mathbf{X}_d \in \mathbb{R}^{N \times k}$ que representa las características de los destinos se define como:

$$\mathbf{X}_d = \mathbf{I}_n \otimes \mathbf{X} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \otimes \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nk} \end{pmatrix}$$

Distancia

Sea $\mathbf{C} \in \mathbb{R}^{n \times n}$ la matriz simétrica de distancias entre las regiones i y j . Las distancias se incorporan al modelo definiendo $\mathbf{c} = \text{vec}(\mathbf{C})$ al vector de \mathbb{R}^N que resulta de apilar las filas de la matriz \mathbf{C} .

Modelo Log- Normal

Aplicando logaritmos, el modelo resultante es:

$$\log(\mathbf{y}) = \alpha \mathbf{1}_N + \log(\mathbf{X}_o)\beta_o + \log(\mathbf{X}_d)\beta_d + \phi \log(\mathbf{c}) + \varepsilon, \quad (2)$$

con $\varepsilon \sim N(\vec{0}, \sigma^2 \mathbf{I})$ (LeSage y Pace, 2008).

En la gran mayoría de las aplicaciones los flujos OD no cumplen con los supuestos necesarios para que esta especificación sea adecuada.

Modelos de Conteo

Aplicando la especificación Poisson al caso del flujos de viajes, se establece que (Burger, van Oort y Linders 2009) el volumen observado de viajes entre las regiones i y j , y_{ij} tiene una distribución Poisson, tal que

$$Pr(y_{ij}) = \frac{\exp(-\mu_{ij})\mu_{ij}^{y_{ij}}}{y_{ij}!} \quad (3)$$

con

$$\mu_{ij} = \exp(\lambda + \beta_o X_{o_i} + \beta_d X_{d_j} + \phi c_{ij}) \quad (4)$$

donde λ es una constante, X_{o_i} y X_{d_j} son los valores de las variables explicativas en el origen i y el destino j , y β_o y β_d los vectores de parámetros asociados. El modelo se estima por máxima verosimilitud.

Modelos de Conteo

En el modelo de Poisson, $E(y_{ij}) = V(y_{ij}) = \mu_{ij}$. Cuando esto no se cumple y $V(y_{ij}) > \mu_{ij}$ se dice que los datos presentan sobredispersión, y el ajuste con la Poisson no es el más adecuado.

Con el fin de corregir el problema de la sobredispersión surge la regresión con la especificación Binomial Negativa. La varianza se especifica como una función de la media condicional μ_{ij} y un parámetro de dispersión α tal que $V(y_{ij}|X) = \mu_{ij} + \alpha\mu_{ij}^2$.

$$Pr(y_{ij}) = \frac{\Gamma(y_{ij} + \alpha^{-1})}{y_{ij}! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ij}} \right)^{\alpha^{-1}} \left(\frac{\mu_{ij}}{\alpha^{-1} + \mu_{ij}} \right)^{y_{ij}} \quad (5)$$

Modelos Restringidos

Los modelos antes considerados no reproducen los totales de las marginales de la matriz OD. Este requisito es deseable, de acuerdo al modelo de cuatro etapas.

En el modelo sin restricciones, se cumple que el total de flujos estimados iguala a la cantidad total observada de flujos, es decir

$$Y_{..} = \sum_{i=1}^I \sum_{j=1}^J y_{ij} = \sum_{i=1}^I \sum_{j=1}^J \hat{y}_{ij} \quad (6)$$

En el caso individualmente restringido (restringido en origen o en destino), las marginales de las predicciones deben coincidir con las salidas $Y_{i.}$ o entradas $Y_{.j}$ observadas totales, en donde

$$Y_{i.} = \sum_{j=1}^J y_{ij} = \sum_{j=1}^J \hat{y}_{ij} \quad y \quad Y_{.j} = \sum_{i=1}^I y_{ij} = \sum_{i=1}^I \hat{y}_{ij} \quad (7)$$

Modelos Restringidos

En el contexto de los modelos de conteo, al incorporar las regiones de origen o de destino, según el caso, como efectos fijos (con especificación Poisson o Binomial Negativa), se cumple con las restricciones de la ecuación 7.

Para el caso restringido en origen, la media de la ecuación 4 para los modelos de conteo es:

$$\mu_{ij} = \exp(\lambda + \lambda_i^o + \beta_d X_d + \phi c_{ij})$$

donde λ_i^o es el efecto fijo asociado a las regiones como origen. Para el modelo restringido en destino,

$$\mu_{ij} = \exp(\lambda + \lambda_j^d + \beta_o X_o + \phi c_{ij})$$

donde λ_j^d es el efecto fijo asociado a las regiones como destino.

Modelos Mixtos

El modelo doblemente restringido, no puede ser estimado bajo este enfoque si se incorporan las características de origen y de destino X_o y X_d a la vez. Es así que se contempla la estimación del modelo en el contexto de modelos mixtos, tratando a las regiones como efectos aleatorios.

El rol de los efectos aleatorios puede tener un doble propósito. En primer lugar, fortalecer las inferencias sobre la predicción de los flujos más allá de las variables regresoras utilizadas, y, segundo modelizar la posible correlación entre observaciones de una misma región (Griffith y Fischer, 2016).

Para el modelo doble restringido,

$$\mu_{ij} = \exp(\lambda + b_i^o + b_j^d + \beta_o X_o + \beta_d X_d + \phi c_{ij})$$

donde $(b_i, b_j) \sim N(\vec{0}, V)$.

Filtrado Espacial

La matriz de pesos **W**

Se utiliza una matriz de pesos binaria, $\mathbf{W} = ((w_{ij}))_{i,j=1,\dots,n}$ donde w_{ij}

$$w_{ij} = \begin{cases} 1 & \text{si } i \text{ y } j \text{ son contiguos} \\ 0 & \text{si } i \text{ y } j \text{ no son contiguos} \end{cases}$$

Esta matriz refleja la interacción entre regiones geográficas, que deberá ser llevada a la estructura de flujos.

La matriz de pesos \mathbf{W}

Se definen tres tipos de interacción entre los flujos: sólo en origen, sólo en destino, y en origen - destino.

$$\mathbf{W}_o = \mathbf{W} \otimes \mathbf{I}_n$$

$$\mathbf{W}_d = \mathbf{I}_n \otimes \mathbf{W}$$

$$\mathbf{W}_{od} = \mathbf{W} \otimes \mathbf{W}$$

Cada una de estas especificaciones se encuentra asociada a un modelo sobre el comportamiento de los “viajantes” (Chun, 2008).

- ▶ *Intervening Opportunities.*
- ▶ *Competing Destinations Model.*
- ▶ *Competing Destinations Model y Intervening Opportunities en simultáneo.*

Filtrado Espacial

La autocorrelación espacial implica que no se cumple el supuesto de independencia entre las observaciones. No tenerla en cuenta lleva a tener sesgo en los errores y en la estimación de los parámetros.

El filtrado espacial es una técnica no paramétrica que se basa en los vectores propios de la matriz de pesos espaciales (en el presente caso W_o , W_d y/o W_{od}).

La incorporación de subconjuntos de vectores propios en el modelo como variables regresoras “filtra” los residuos, haciendo que éstos no presenten correlación espacial.

Filtrado Espacial

Desde una perspectiva de una especificación errónea del modelo (Tiefelsdorf y Griffith, 2007), se asume que el modelo

$$y = X\beta + \varepsilon^* \quad (8)$$

tiene errores autocorrelacionados espacialmente ε^* , que pueden ser descompuestos en un componente de ruido blanco, ε , y en un conjunto de variables exógenas no especificadas, E . La estructura del modelo con una especificación errónea sería:

$$y = X\beta + E\gamma + \varepsilon \quad (9)$$

Filtrado Espacial

Los vectores propios se extraen de la matriz de pesos espaciales transformada. Se consideran dos matrices de proyección posibles:

$$M_{(1)} = I - 1(1^T 1)^{-1} 1^T$$

y

$$M_{(x)} = I - X(X^T X)^{-1} X^T$$

con dos transformaciones

$$T_1 = M_{(1)} 1/2(W + W^T) M_{(1)}$$

y

$$T_X = M_{(x)} 1/2(W + W^T) M_{(x)}$$

Filtrado Espacial

En el presente caso se tienen tres matrices de pesos espaciales: W_o , W_d y W_{od} .

Optando por la matriz de proyección tipo $M_{(1)}$, se obtienen sus respectivas matrices transformadas, T_o , T_d y T_{od} .

La selección de los vectores propios se realiza con un procedimiento similar al *forward*, midiendo en cada iteración el Índice de Moran de los residuos, hasta que no presenten correlación espacial.

En el contexto del modelo doble restringido,

$$\mu_{ij} = \exp \left(\lambda + b_i^O + b_j^D + \beta_o X_o + \beta_d X_d + \phi c_{ij} + \sum_{q=1}^Q \gamma_q E_q \right)$$

siendo Q la cantidad de vectores propios seleccionados de las matrices transformadas T_o , T_d y/o T_{od} .

Datos

Los datos provienen de la Encuesta de Movilidad en el Área Metropolitana de Montevideo. Las unidades geográficas consideradas son los Municipios. La matriz OD es:

	A	B	C	CH	D	E	F	G
A	60	51	27	39	9	7	6	12
B	1	69	10	19	2	5	3	2
C	2	28	25	20	1	2	2	3
CH	1	50	4	36	3	6	0	3
D	3	35	17	39	50	21	17	8
E	0	51	11	27	3	18	4	2
F	11	37	18	22	15	17	41	3
G	10	22	10	6	4	5	3	31

Datos

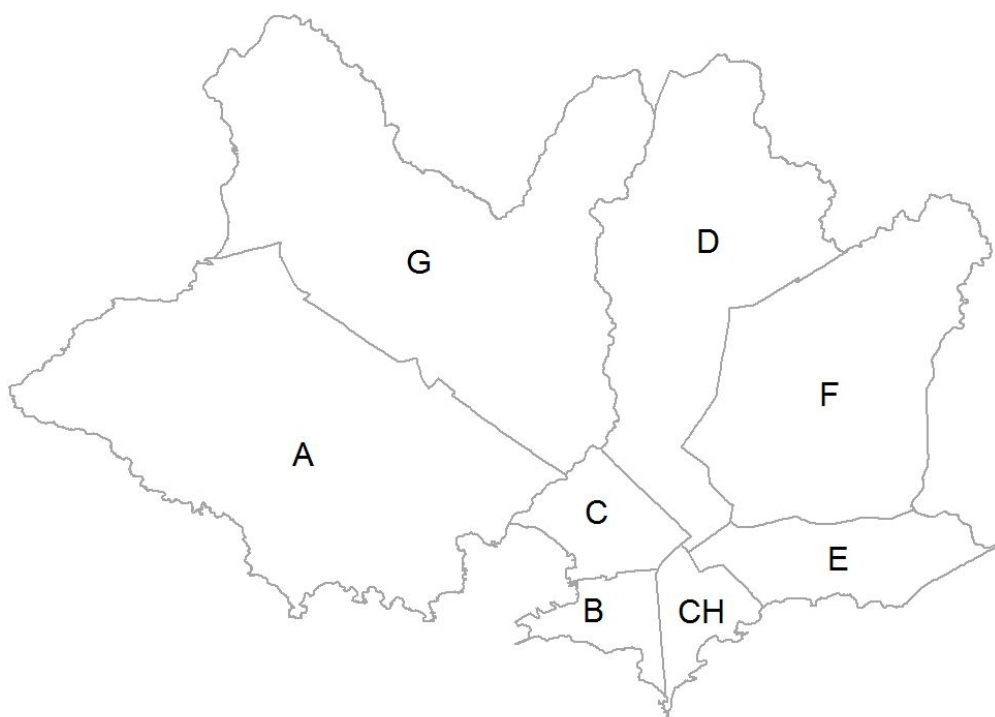


Figure 1: Unidades Geográficas de la matriz de Origen - Destino

Cantidad de empleos aproximada por Municipios según Directorio de Empresas del INE.

	Empleos	Porcentaje (%)
A	24032	5.3
B	166366	36.7
C	70716	15.6
CH	61870	13.6
D	29572	6.5
E	43194	9.5
F	31706	7.0
G	25814	5.0
Total	453270	100

Distancias

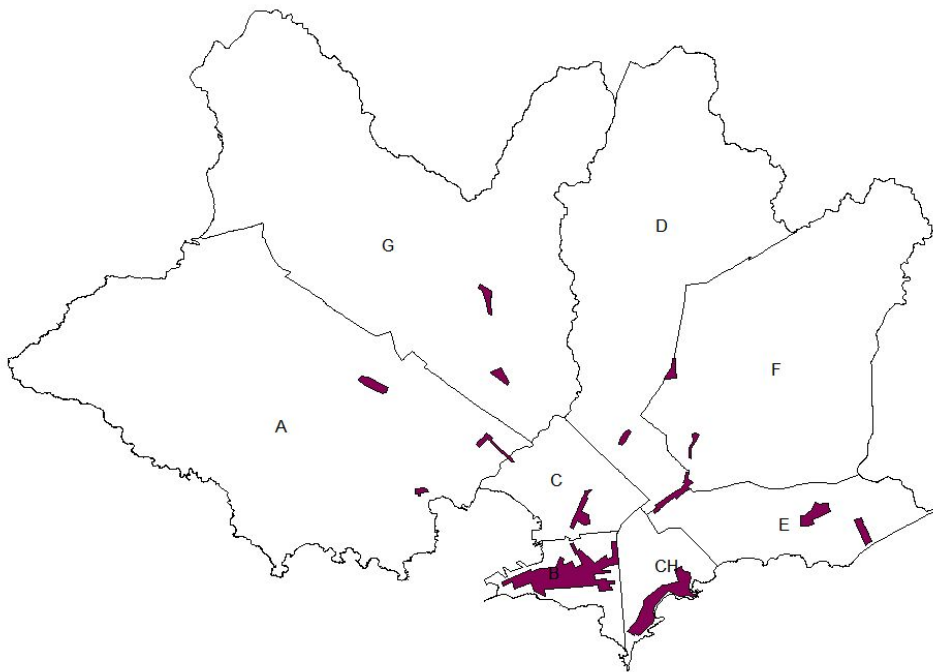


Figure 2: Centralidades de Montevideo

La distancia se mide como el promedio entre el tiempo de transporte público y privado entre los centros de los Municipios.

Metodología

1. Cálculo del Índice de Moran para la matriz W_{od} .
2. Estimación del modelo sin restricciones.
3. Estimación del modelo sin restricciones, con variables.
4. Estimación del modelo mixto doblemente restringido.
5. Estimación del modelo mixto doblemente restringido con variables explicativas.
6. Comparación de resultados.

Resultados

Índices de Moran

Se analiza la existencia de correlación espacial en los residuos del modelo (8) utilizando el Índice de Moran.

Table 1: Índice de Moran para los residuos del modelo (8)

Matriz	<i>I</i> de Moran	<i>p</i> - valor
<i>W_{od}</i>	0.263357258	0.002046

Se rechaza la hipótesis de no correlación, por lo que se concluye que los flujos muestran correlación espacial en la matriz de origen - destino.

Resultados

Estimación de modelos

En todos los casos la especificación de los modelos con una Binomial Negativa presentaron menores valores de AIC frente a los modelos ajustados con una distribución Poisson.

Table 2: Correlación espacial de los residuos en los modelos estimados

Modelo	Residuos	
	Incorrelacionados	AIC
Sin Restricciones	✗	400.2
Sin Restricciones, con variables	✗	412.5
Doble Restringido	✓	432.7
Doble Restringido, con variables	✓	421.8

Resultados

Estimación del modelo Sin Restricciones

Table 3: Resultados del ajuste del Modelo Gravitacional sin restricciones

Efecto	Parámetro	Estimación	p-valor
Intercept	λ	3.588	$< 2e-16$
E_3	γ_3	6.521	$< 2e-16$
E_{23}	γ_{23}	-1.820	0.000607
E_{15}	γ_{15}	1.666	0.001797
E_4	γ_4	-1.209	0.011890
Distancia	ϕ	-0.0456	$< 2e-16$

Resultados

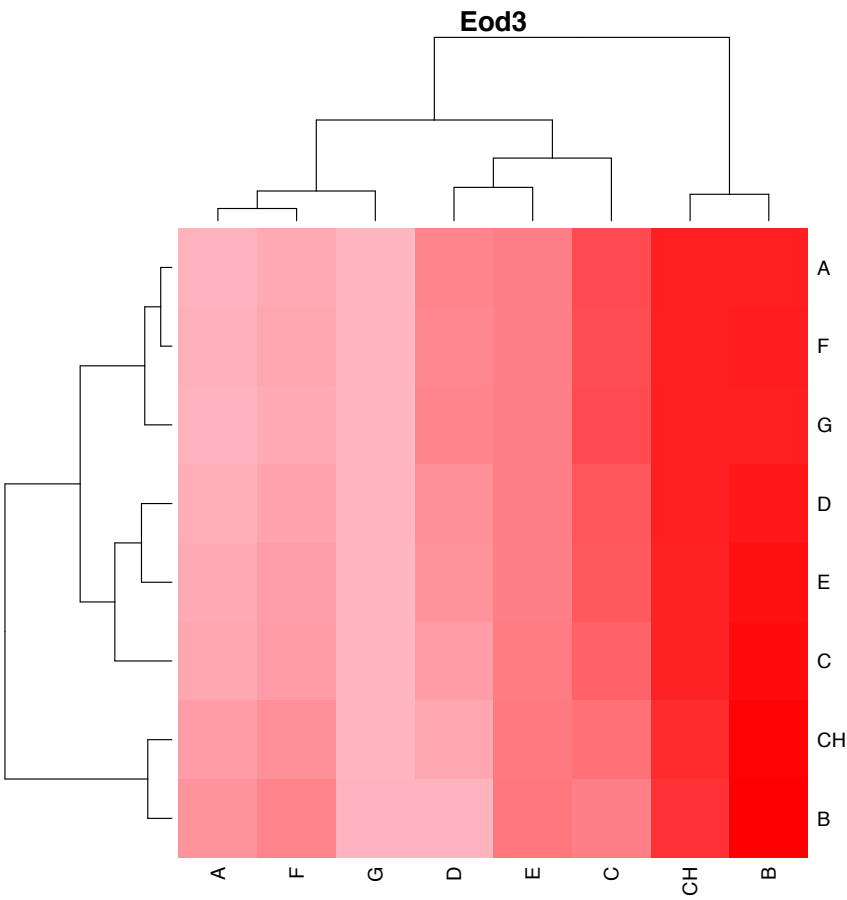


Figure 3: Vector Propio (número 3) seleccionado en el filtrado espacial.

Resultados

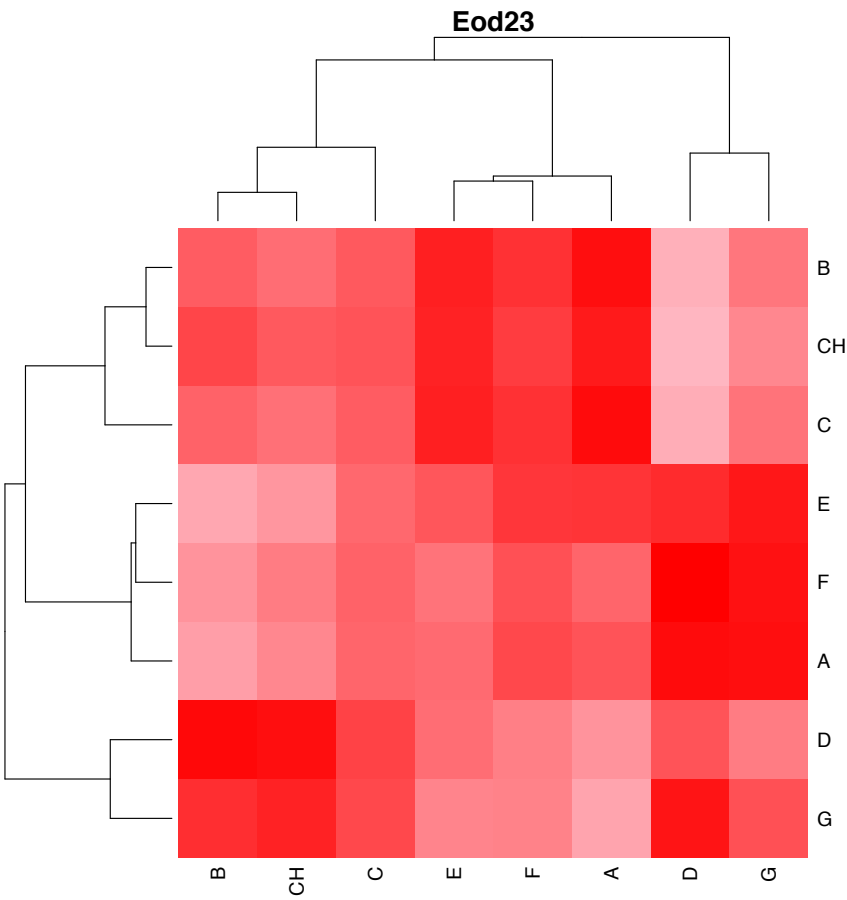


Figure 4: Vector Propio (número 23) seleccionado en el filtrado espacial.

Resultados

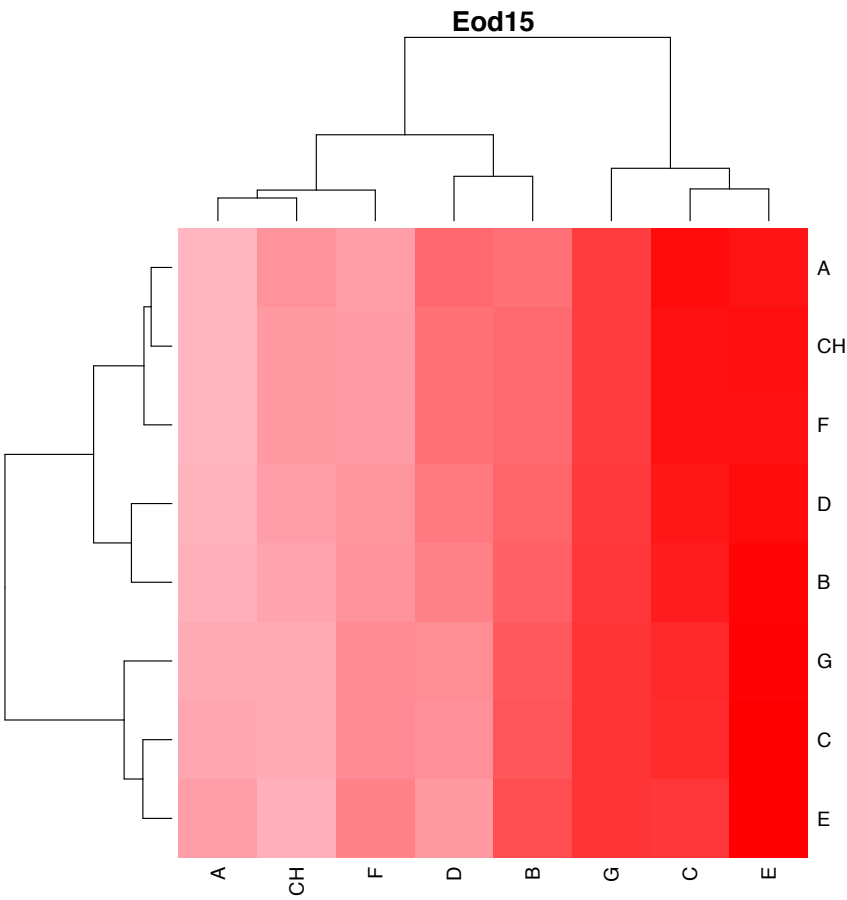


Figure 5: Vector Propio (número 15) seleccionado en el filtrado espacial.

Resultados

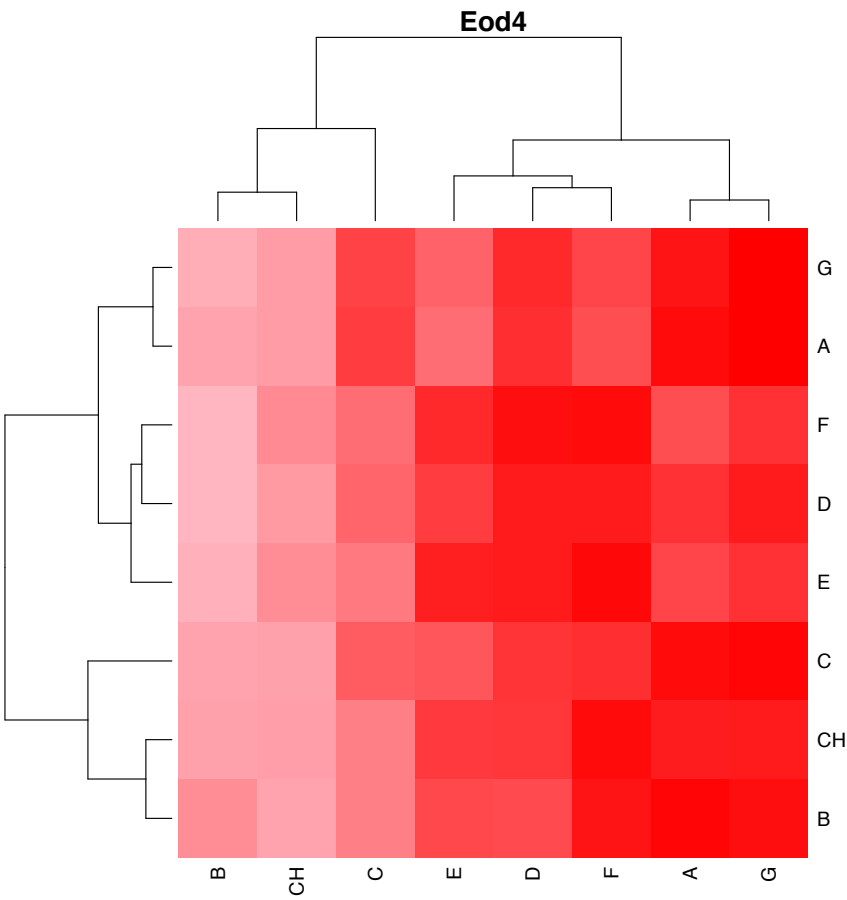


Figure 6: Vector Propio (número 4) seleccionado en el filtrado espacial.

Resultados

Estimación del modelo Sin restricciones, con variables exógenas

Table 4: Resultados del ajuste del Modelo Gravitacional sin restricciones

Efecto	Parámetro	Estimación	$e^{(\cdot)}$	p-valor
Intercept	λ	3.291	26.888	$< 2e-16$
Emp. Origen	β_o	-0.0414	0.9594	1.43e-07
Emp. Destino	β_d	0.0592	1.0609	$< 2e-16$
E_7	γ_7	3.293	26.932	1.72e-08
Distancia	ϕ	-0.041	0.959	$< 2e-16$

Resultados

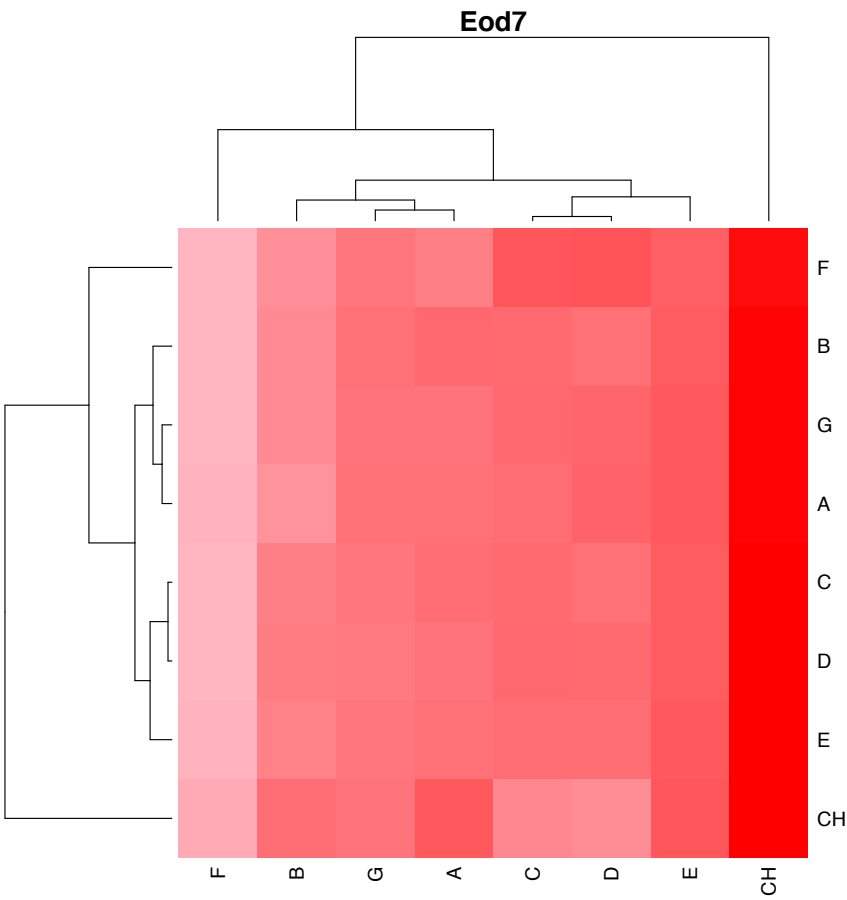


Figure 7: Vector Propio (número 7) seleccionado en el filtrado espacial.

Resultados

Estimación del modelo Doble Restringido

Table 5: Resultados del ajuste del Modelo Gravitacional con restricciones

Efecto	Parámetro	Estimación	$e^{(.)}$	p-valor
Intercept	λ	3.589160	36.20365	$< 2e-16$
Distancia	γ	-0.045785	0.9552473	$< 2e-16$
Covarianza de b_i		Varianza		
$var(b_o)$	v_{11}	0.2958		
$var(b_d)$	v_{22}	0.4742		

Resultados

Estimación del modelo Doble Restringido

Table 6: Resultados del ajuste del Modelo Gravitacional sin restricciones

Efecto	Parámetro	Estimación	$e^{(\cdot)}$	p-valor
Intercept	λ	3.347	28.433	$< 2e-16$
Emp. Origen	β_o	-0.0411	0.9597	0.00394
Emp. Destino	β_d	0.0587	1.0604	.11e-05
Distancia	ϕ	-0.0449	0.956	$< 2e-16$
Covarianza de b_i		Varianza		
$var(b_o)$	v_{11}	0.12		
$var(b_d)$	v_{22}	0.1209		





Discusión y Lineamientos a futuro

- ▶ Considerando los modelos sin restricciones, la incorporación del empleo como variable explicativa elimina parte de la correlación espacial de los residuos.
- ▶ Cuando se incorporan las restricciones por medio de los efectos aleatorios los residuos no presentan correlación espacial.
- ▶ Los municipios son unidades geográficas muy grandes y puede llevar a que las correlaciones existentes sean débiles comparadas con las que pueden estar presentes en unidades geográficas menores. Se estimarán los modelos con Centros Comunes Zonales como unidad de análisis.

Discusión y Lineamientos a futuro

- ▶ Una forma alternativa de que se cumplan las restricciones del modelo es mediante el cálculo explícito de los factores de balance A_i y B_j . Griffith y Fischer (Griffith y Fischer, 2016) proponen partir de las estimaciones de las marginales del modelo sin restricciones, y luego mediante el algoritmo iterativo de Deming - Stephan ajustar las marginales a los totales de la matriz OD.
- ▶ Por último, los modelos estimados en este trabajo no tienen en cuenta los pesos muestrales. Se debe analizar a futuro cómo incorporar los pesos muestrales, de forma que se tenga una estimación para toda la población de los flujos de la matriz OD.

Bibliografía

-  Burger, M. J., Van Oort, F. G. y Linders, G. (2009). On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-Inflated Estimation. *ERIM Report Series* Reference No. ERS-2009-003-ORG. Available at SSRN: <https://ssrn.com/abstract=1376148>
-  Chun, Y. (2008). Modeling Network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems*. Vol. 10, pp. 317- 344.
-  Griffith, D. A. (2003) *Spatial Autocorrelation and Spatial Filtering*, Advances in Spatial Science, Springer - Verlag.
-  Griffith, D. A. y Fischer, M. (2016). *Constrained Variants of the Gravity Model and Spatial Dependence: Model Specification and Estimation Issues*. Capítulo 3, Patuelli, R. y Arbia, G. editores, Spatial Econometric Interaction Modelling, Springer.

Bibliografía



LeSage, J. y Pace, R. K. 2008. Spatial Econometric modeling of origin – destination flows. *Journal of Regional Science*, Vol. 48, No. 5, pp. 941- 967.



Tiefelsdorf, M., Griffith, D.A. (2007). Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning*, vol. 39, 1193 - 1221.