

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

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

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

5 de junio 2019

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:

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

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 , 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 distribución Poisson no es el más adecuado.

Con el fin de corregir este problema, 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 de forma 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 es un requisito deseable en los modelos gravitacionales.

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 un enfoque de efectos fijos ya que si se incorporan las características de origen y de destino X_o y X_d simultáneamente, se inducen problemas de multicolinealidad. Es así que se contempla la estimación del modelo en el contexto de modelos mixtos, considerando efectos aleatorios en las regiones tanto de origen como de destino.

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 explicativas utilizadas, y segundo lugar, modelar la posible correlación entre observaciones de una misma región (Griffith y Fischer, 2016).

Modelos Mixtos

La formulación del modelo doblemente restringido sería la siguiente:

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

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

Como se señaló anteriormente, el hecho de incluir un efecto aleatorio en origen y otro en destino induce correlaciones entre los flujos que comparten uno de sus extremos.

No obstante, de existir correlación entre flujos con orígenes en regiones adyacentes, la misma no es capturada por estos modelos.

La matriz de adyacencia **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 adyacencia \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.*

Datos

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

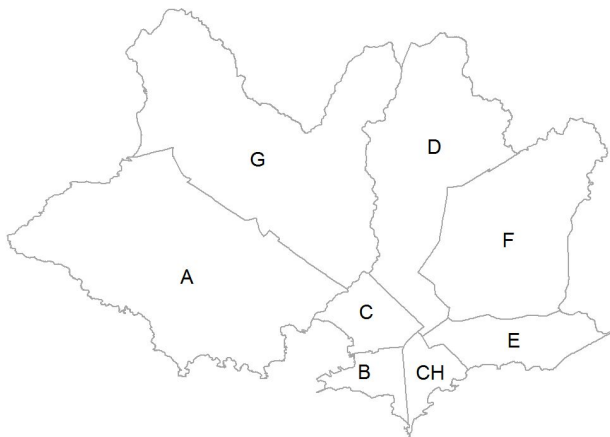


Figure 1: Unidades Geográficas de la matriz de Origen - Destino

Flujos OD

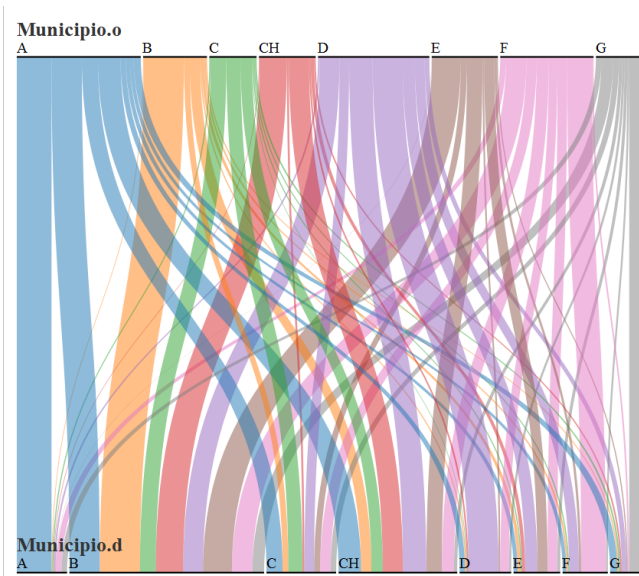
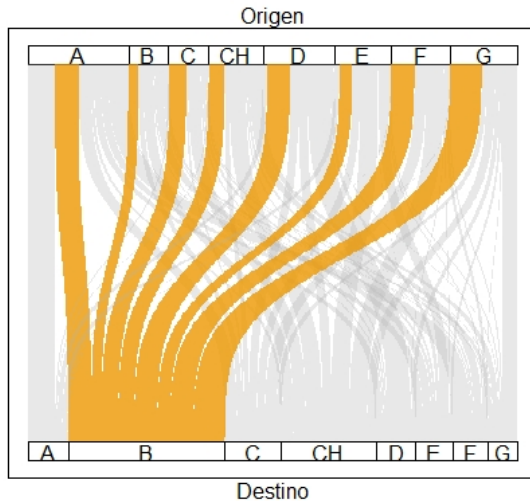
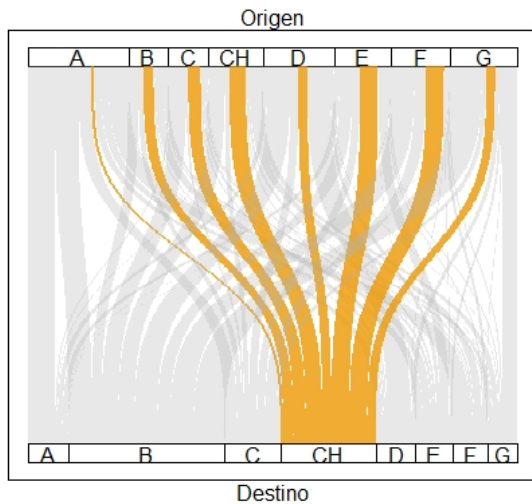


Figure 2: Cantidad de viajes según origen y destino

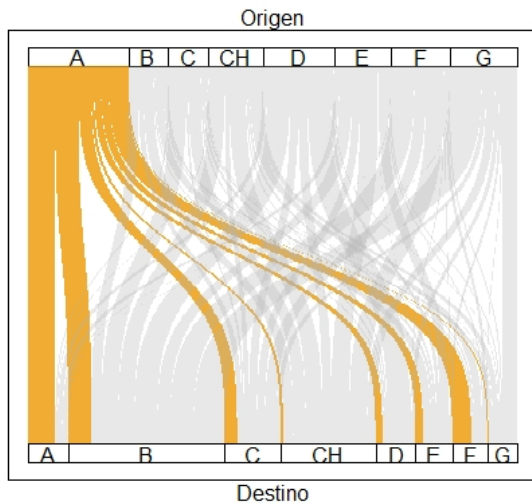
Flujos OD



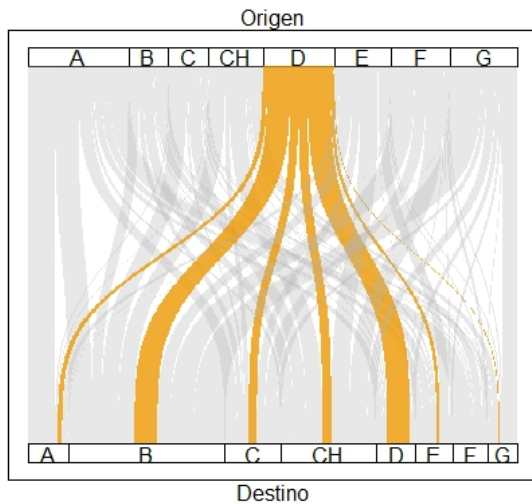
Flujos OD



Flujos OD



Flujos OD



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

	Empleos	Población	Empleo/Población
A	24032	214810	0.1118
B	166366	128517	1.2945
C	70716	163309	0.4330
CH	61870	177780	0.3480
D	29572	202431	0.1460
E	43194	183308	0.2356
F	31706	161044	0.1968
G	25814	148152	0.1742

Distancias

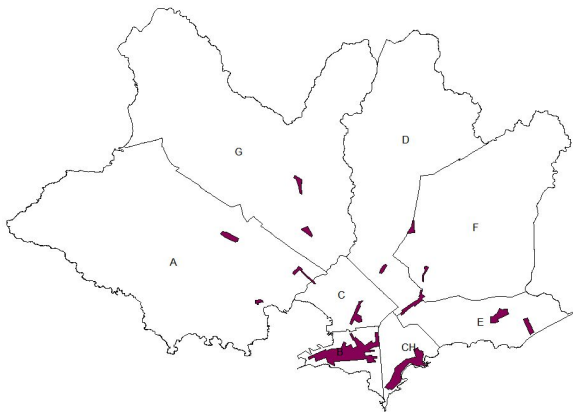


Figure 7: Centralidades de Montevideo

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

Metodología

La estrategia de estimación consistió de las siguientes etapas:

1. Se ajustaron GLM con distribución Poisson y Binomial Negativa sin variables explicativas.
2. Se calculó del Índice de Moran para las matrices W_o , W_d y W_{od} para determinar presencia de autocorrelación espacial.
3. Se incorporó el cociente entre empleos y población a los GLM del paso 1.
4. Se calculó del Índice de Moran para las matrices W_o , W_d y W_{od} para determinar presencia de autocorrelación espacial.
5. Estimación del modelo mixto doblemente restringido con variables explicativas.
6. Se volvió a calcular el Índice de Moran para las matrices W_o , W_d y W_{od} para determinar presencia de autocorrelación espacial.
7. Se contempló la posibilidad de incorporar un efecto de viajes dentro de los municipios.
8. Comparación de resultados.

GLM

En primer lugar se presentan los resultados correspondientes a los modelos de efectos fijos.

Table 1: Estimación de los modelos GLM

Coeficiente	Estimación	\exp^{β}	IC95%
Modelo Poisson			
tiempo	-0.035	0.965	(0.952 ; 0.969)
Modelo Binomial Negativo			
tiempo	-0.035	0.965	(0.949 ; 0.980)

Autocorrelación espacial (1)

Habiendo estimado los primeros modelos, se procedió a chequear el supuesto de incorrelación. Para eso se utilizó el test de Moran considerando las matrices W_o , W_d y W_{od} .

Table 2: Índice de Moran, modelos GLM

Matriz	Poisson	Binomial Negativa
W_o	≤ 0.001	≤ 0.001
W_d	0.086	0.123
W_{od}	0.002	0.001

Los modelos no superan esta etapa. Por lo cual se consideran modelos de efectos mixtos.

GLM + variable explicativa

Se presentan los resultados correspondientes a los modelos de efectos fijos.

Table 3: Estimación de los modelos GLM

Coeficiente	Estimación	\exp^{β}	IC95%
Modelo Poisson			
emp.pob _o	-0.831	0.992	(0.989 ; 0.994)
emp.pob _d	1.178	1.012	(1.010 ; 1.013)
tiempo	-0.039	0.962	(0.958 ; 0.966)
Modelo Binomial Negativo			
emp.pob _o	-1.006	0.990	(0.985 ; 0.995)
emp.pob _d	1.411	1.014	(1.009 ; 1.019)
tiempo	-0.038	0.963	(0.952 ; 0.974)

Autocorrelación espacial (2)

Habiendo estimado los primeros modelos, se procedió a chequear el supuesto de incorrelación. Para eso se utilizó el test de Moran considerando las matrices W_o , W_d y W_{od} .

Table 4: Índice de Moran, modelos GLM

Matriz	Poisson	Binomial Negativa
W_o	≤ 0.001	≤ 0.001
W_d	0.006	0.002
W_{od}	0.965	0.970

Pese a que la inclusión de la variable explicativa solucionó el problema de la falta de incorrelación en OD, se generó un problema de en destino. Por lo cual se consideran modelos de efectos mixtos.

Table 5: Estimación de los modelos GLMM

Coeficiente	Estimación	\exp^{β}	IC95%
Modelo Poisson			
emp.pob _o	-0.873	0.991	(0.983 ; 0.999)
emp.pob _d	1.330	1.013	(1.006 ; 1.021)
tiempo	-0.044	0.957	(0.953 ; 0.962)
σ_{b_o}	0.119		
σ_{b_d}	0.118		
Modelo Binomial Negativo			
emp.pob _o	-1.003	0.989	(0.981 ; 0.998)
emp.pob _d	1.500	1.015	(1.006 ; 1.024)
tiempo	-0.045	0.963	(0.947 ; 0.964)
σ_{b_o}	0.142		
σ_{b_d}	0.154		

Autocorrelación espacial (3)

Se procedió de la misma forma que con los GLM, se puso a prueba el supuesto de incorrelación mediante el test de Moran considerando las matrices W_o , W_d y W_{od} .

Table 6: Índice de Moran, modelos GLM

Matriz	Poisson mixto	Binomial Negativa mixta
W_o	0.287	0.337
W_d	0.833	0.883
W_{od}	0.748	0.747

Los modelos superan esta etapa. Por lo cual se comparan los resultados de cada caso.

Comparación

En ambos modelos se observa un efecto negativo en el tiempo de viaje, por cada minuto extra que dura un viaje, el flujo promedio se reduce alrededor de 4%. En cuanto al cociente entre número de empleos y población, en ambos caso se observa que cuanto mayor este cociente en origen, el número promedio de viajes se reduce. Por otro lado, cuando este cociente aumenta en destino, se da un fenómeno de atracción.

En términos de AIC el modelo mixto Poisson presenta un valor de 482.79 mientras que el Binomial Negativo presenta un valor de 424.35.

Efectos aleatorios

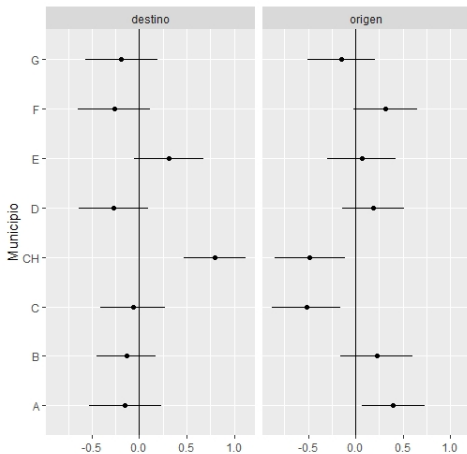


Figure 8: Efectos aleatorios, modelo Binomial Negativo Mixto

Viajes intra municipio

Se propuso la inclusión de un efecto (fijo) para modelar los viajes al interior de cada municipio.

Table 7: Estimación de los modelos GLMM

Coeficiente	Estimación	\exp^{β}	IC95%
Modelo Poisson			
emp.pob _o	-0.870	0.991	(0.984 ; 0.998)
emp.pob _d	1.393	1.014	(1.007 ; 1.022)
tiempo	-0.028	0.972	(0.963 ; 0.981)
intraM	0.560	1.751	(1.293 ; 2.371)
σ_{b_o}	0.110		
σ_{b_d}	0.107		
Modelo Binomial Negativo			
emp.pob _o	-0.979	0.990	(0.981 ; 0.999)
emp.pob _d	1.565	1.016	(1.007 ; 1.025)
tiempo	-0.032	0.968	(0.953 ; 0.983)
intraM	0.544	1.722	(0.996 ; 2.985)
σ_{b_o}	0.132		
σ_{b_d}	0.145		

Efecto del tiempo de viaje

El siguiente gráfico presenta el efecto multiplicativo estimado de la longitud de los viajes sobre el número promedio de viajes.

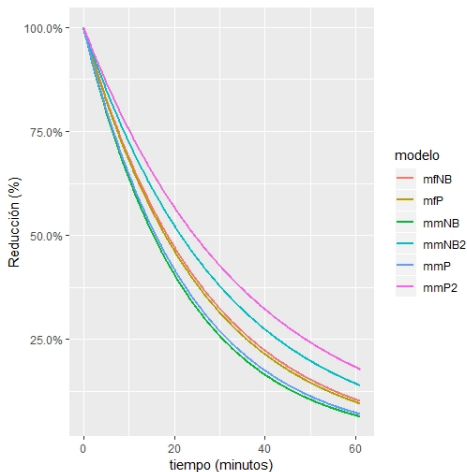


Figure 9: Efecto del tiempo

Conclusión y Lineamientos a futuro

- ▶ Al incorporar variables explicativas a los GLM se logró “explicar” la correlación en la interacción OD.
- ▶ El incorporar efectos aleatorios, se logró modelar la correlación en origen y destino.
- ▶ Se observó que los altos flujos a los municipios B y CH se dan por distintos motivos.
- ▶ Se cuantificó el efecto de la longitud del viaje en el recuento promedio de viajantes de un municipio a otro.

Conclusión y Lineamientos a futuro

- ▶ 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.
- ▶ Por último, los modelos estimados en este trabajo no tienen en cuenta los pesos muestrales. Se debe analizar a futuro cómo incorporarlos, 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>



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.