

COMPARACIÓN DE MODELOS ESPACIO-TEMPORALES DE DATOS DE PANEL CON R

Patricia Carracedo y Ana Debón

23 de noviembre de 2018

Introducción

Objetivo. Comparar y seleccionar el **mejor modelo de datos de panel** que tenga en cuenta la **dimensión espacial y temporal** de los datos utilizando el software R.

Motivación. Las compañías aseguradoras y los gobiernos disponen de información sobre la mortalidad de diferentes regiones a lo largo del tiempo pero sería interesante usar esta información para **identificar regiones** con mortalidad similar o diferente a lo largo del tiempo y saber que variables explican las **diferencias de mortalidad**.

Introducción

- ▶ La mortalidad es un fenómeno demográfico que se caracteriza por ser **irreversible, no repetible e inevitable** (Vinuesa Angulo 1997).
- ▶ La mortalidad depende de variables como el sexo, población, edad, raza y otras; a estas variables se les denomina variables **"confusoras"**.
- ▶ Para poder comparar tasas brutas de mortalidad entre diferentes áreas geográficas será necesario **estandarizar** dichas tasas (Rezaeian et al. 2007; Vinuesa Angulo 1997; Julious, Nicholl, and George 2001).
- ▶ La metodología estadística que propone este trabajo tiene en cuenta dos dimensiones de los datos: el **tiempo** y la **situación geográfica**.

Metodología espacio-temporal

Estadísticos

Modelos de datos de panel

Metodología espacio-temporal. Estadísticos

Datos

- ▶ Los datos de mortalidad que hemos tenido en cuenta en este estudio se corresponden con **26 países europeos** durante el periodo de temporal entre **1995 - 2012** (Human Mortality Database 2016) y rango de edad 65-110+.
- ▶ Se recopiló información de **4 covariables económicas y sociales** de la web de acceso libre The World Bank Database (2018): tasa de crecimiento Producto Interior Bruto (PIB), gasto sanitario, CO_2 y gasto en educación.

Metodología espacio-temporal. Estadísticos

Transformaciones de los datos

- ▶ Las covariables se han transformado con el fin de que éstas indiquen la cantidad destinada a los mayores de 65 años de cada sexo en el mismo país y periodo, para ello:

$$X_{i,t,s} = \frac{X_{i,t}}{\sum_s P_{i,t,s}} * P_{65,i,t,s}$$

- ▶ $X_{i,t}$ valor de la covariable para cada país i y año t .
- ▶ $P_{i,t}$ población de un determinado país i y año t .
- ▶ $P_{65,i,t,s}$ población de 65 años o más años en un determinado país i y año t .

Metodología espacio-temporal. Estadísticos

Multicolinealidad: Factor de Inflación de la Varianza (VIF). Masculino.

- ▶ Para evitar la multicolinealidad entre las covariables ($VIF < 2$),
- ▶ Se confirma que **ninguna variable** tiene problemas de multicolinealidad cuando se considera el sexo masculino.

```
##   var           vif
##   t_PIB_p       1.04952148776566
##   Gasto_san_pub_p 1.95012217057115
##   CO2_p         1.54517921483887
##   gasto_educ_p  1.38141632155552
## nAll variables have VIF < 2, max VIF 1.95 nn
```

Metodología espacio-temporal. Estadísticos

Multicolinealidad: Factor de Inflación de la Varianza (VIF). Femenino.

- ▶ Para evitar la multicolinealidad entre las covariables ($VIF < 2$),
- ▶ Se confirma que la variable **gasto en sanidad** produce multicolinealidad cuando se considera el sexo femenino.

```
##   var           vif
##   t_PIB_p      1.00921752577798
##   Gasto_san_pub_p 2.58351621589964
##   CO2_p        1.65426185567652
##   gasto_educ_p 1.7714196234945
## nremoved:  Gasto_san_pub_p 2.583516 nn
```

Metodología espacio-temporal. Estadísticos

Cifra Comparativa de Mortalidad (CMF)

El CMF es el ratio entre **número de muertes que se esperarían en la población estándar** si experimentara las tasas de mortalidad por edad de la población estudiada y el **número real de muertes en la población estándar** durante un período de tiempo

$$CMF_{i,t,s} = \frac{E_{i,t,s}}{O_{2012,m}} \quad i \in \{1, \dots, 26\}, t \in \{1995, \dots, 2012\} \quad \text{y} \quad s \in \{m, f\}$$

- ▶ $E_{i,t,s}$ número de muertes esperadas para cada país i , año t y sexo s .
- ▶ $O_{2012,m}$ número de muertes observadas en el conjunto de países europeos para el sexo masculino y año 2012.

Metodología espacio-temporal. Estadísticos

Cifra Comparativa de Mortalidad (CMF)

- ▶ $CMF > 1$.- Indica que se detectaron más muertes esperadas que las observadas; en este caso hay “exceso de muertes”.
- ▶ $CMF < 1$.- Indica que se detectaron más muertes observadas que esperadas; en este caso hay “déficit de muertes”.
- ▶ $CMF = 1$.- Indica que se detectaron el “mismo número de muertes” esperadas que observadas.

Índice de Moran Global

El Índice de Moran Global (GM) es una medida resumen que muestra la **intensidad de la dependencia espacial** de todos los países considerados en el estudio (Moran 1950a; Moran 1950b).

$$GM_{t,s} = \frac{N \sum_i \sum_j W_{ij} (CMF_{i,t,s} - \overline{CMF}_{g,t,s}) (CMF_{j,t,s} - \overline{CMF}_{t,s})}{\sum_i \sum_j w_{ij} \sum_i (CMF_{i,t,s} - \overline{CMF}_{t,s})^2}$$

donde $\overline{CMF}_{t,s}$ es la media del CMF de todos los países para cada año t y sexo s y $W = (\omega_{ij})$ es la matriz de pesos espaciales.

- ▶ $GM_{t,s} > 0$.- Autocorrelación espacial **positiva** entre los países.
- ▶ $GM_{t,s} < 0$.- Autocorrelación espacial **negativa** entre los países.
- ▶ $GM_{t,s} = 0$.- **Ausencia de autocorrelación** espacial entre los países.

Índice de Moran Local

El Índice de Moran Local (LM) proporciona información sobre la **contribución de cada país al GM** (Anselin 1995), el cual ofrece una idea del grado de similitud o diferencia entre el valor del CMF de un país con respecto al valor del CMF en los países vecinos.

$$LM_{i,t,s} = \frac{(CMF_{i,t,s} - \overline{CMF}_{t,s})}{S^2(CMF_{t,s})} \sum_i \sum_j w_{ij} (CMF_{j,t,s} - \overline{CMF}_{t,s})$$

donde $S^2(CMF_{t,s})$ es la varianza del CMF en todos los países para cada edad g , año t y sexo s .

Cuando el LM es significativo, se detectan dos tipos de clusters:

- ▶ $LM_{i,t,s} > 0$.- "Clusters Espaciales" HH o {LL}.
- ▶ $LM_{i,t,s} < 0$.- "Clusters Outlier" LH o {HL}.

Metodología espacio-temporal. Modelos de datos de panel

- ▶ Es un modelo de regresión que utiliza la **dimensión temporal y espacial** de los datos para la estimación de los parámetros de interés.
- ▶ Modelan la **heterogeneidad no observada** producida tanto por las unidades espaciales estudiadas como por el tiempo.
- ▶ Controlar la heterogeneidad **reduce problemas de multicolinealidad** entre las variables produciendo **estimaciones más eficientes** en los parámetros de los modelos de datos de panel (Kennedy 2003; Hsiao 2014).
- ▶ Los modelos de datos de panel utilizados con el R package `splm` (Millo and Piras 2012), consideran los efectos fijos del **espacio** y del **tiempo** además de incorporar la **vecindad**.

Metodología espacio-temporal. Modelos de datos de panel

Elhorst (2014) proporciona una revisión de los modelos de datos de panel más comúnmente utilizados en la investigación,

- ▶ Modelo de retardo espacial con efecto fijo espacial y temporal (SLMSTFE).
- ▶ Modelo de error espacial con efecto fijo espacial y temporal (SEMSTFE).
- ▶ Modelo de retardo espacial con efecto aleatorio espacial y efecto fijo temporal (SLMSTRE).
- ▶ Modelo de error espacial con efecto aleatorio espacial y efecto fijo temporal (SEMSTRE).

Metodología espacio-temporal. Modelos de datos de panel

Elhorst (2014) proporciona una revisión de los modelos de datos de panel más comúnmente utilizados en la investigación,

- ▶ Modelo de retardo espacial con efecto fijo espacial y temporal (SLMSTFE).
- ▶ Modelo de error espacial con efecto fijo espacial y temporal (SEMSTFE).
- ▶ Modelo de retardo espacial con efecto aleatorio espacial y efecto fijo temporal (SLMSTRE).
- ▶ Modelo de error espacial con efecto aleatorio espacial y efecto fijo temporal (SEMSTRE).

Metodología espacio-temporal. Modelos de datos de panel

Modelo de retardo espacial con efecto fijo espacial y temporal

La expresión matemática del modelo SLMSTFE es,

$$y_{it} = \alpha + \lambda \sum_{j=1}^N \omega_{ij} y_{jt} + x_{it} \beta + \mu_i + \nu_t + \epsilon_{it}$$

- ▶ y_{it} vector de observaciones de la variable dependiente de la unidad i en el año t ;
- ▶ α valor medio de la variable dependiente cuando todas las variables son cero;
- ▶ λ parámetro espacial asociado a la variable dependiente;
- ▶ $W = (\omega_{ij})$ matriz de pesos espaciales, donde i y j representan dos países cualquiera del total de N países;
- ▶ x_{it} matriz de observaciones de las variables independientes;
- ▶ β vector de coeficientes de las variables independientes;
- ▶ μ_i efecto fijo espacial ν_t es el efecto fijo temporal
- ▶ ϵ_{it} es el vector de términos de error independientes e idénticamente

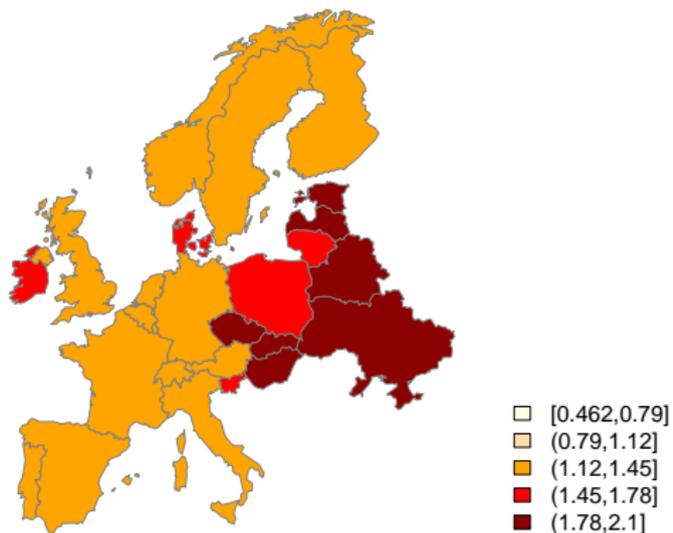
Resultados

Estadísticos

Modelos de datos de panel

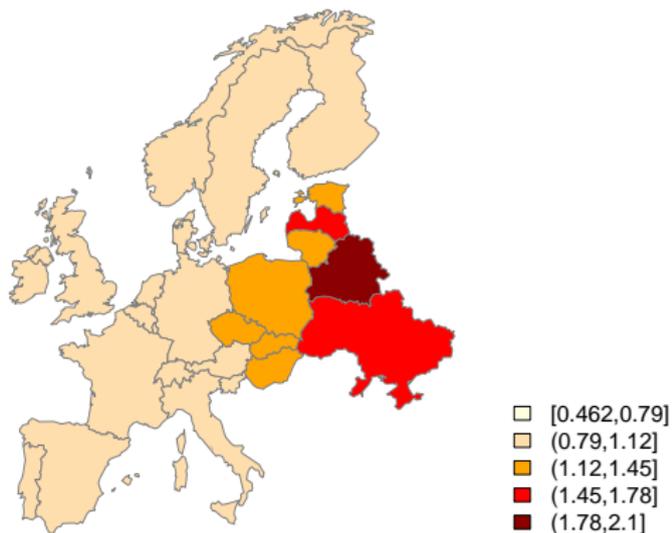
Estadísticos

Analisis exploratorio del CMF. Año 1995 sexo masculino.



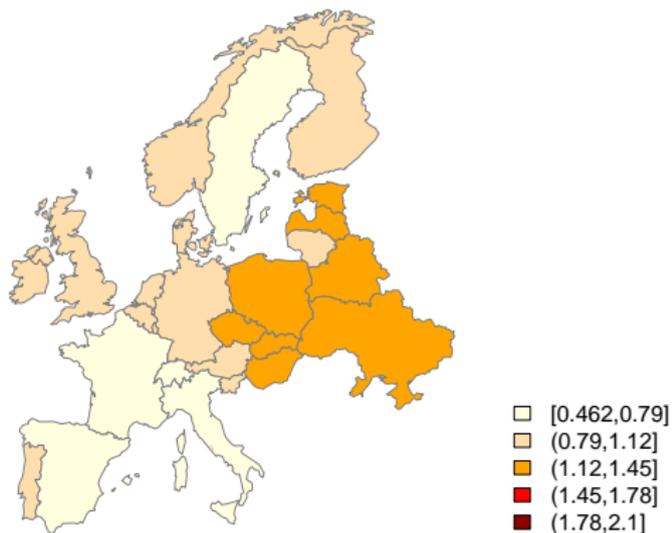
Resultados. Estadísticos

Análisis exploratorio del CMF. Año 2012 sexo masculino.



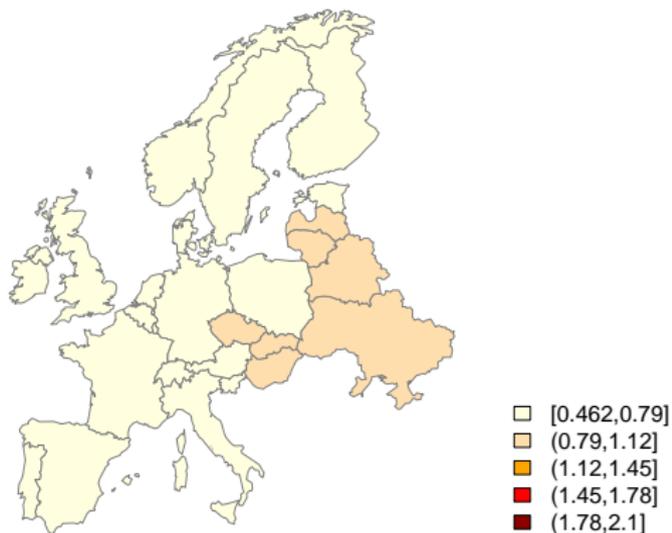
Resultados. Estadísticos

Análisis exploratorio del CMF. Año 1995 sexo femenino.



Resultados. Estadísticos

Análisis exploratorio del CMF. Año 2012 sexo femenino.



Resultados. Estadísticos

Valores del Índice de Moran Global y p-valores. Masculino

##		IMoran	p-value
##	N1995M	0.7181247	2.796950e-06
##	N1996M	0.7237027	2.305711e-06
##	N1998M	0.7215126	2.484818e-06
##	N1999M	0.7034038	3.620331e-06
##	N2001M	0.7346200	1.330597e-06
##	N2002M	0.7266863	1.416127e-06
##	N2004M	0.7556920	6.553703e-07
##	N2005M	0.7681835	4.488774e-07
##	N2007M	0.7883230	2.618044e-07
##	N2008M	0.7856997	2.554205e-07
##	N2010M	0.7617564	4.793523e-07
##	N2011M	0.7728145	3.322491e-07

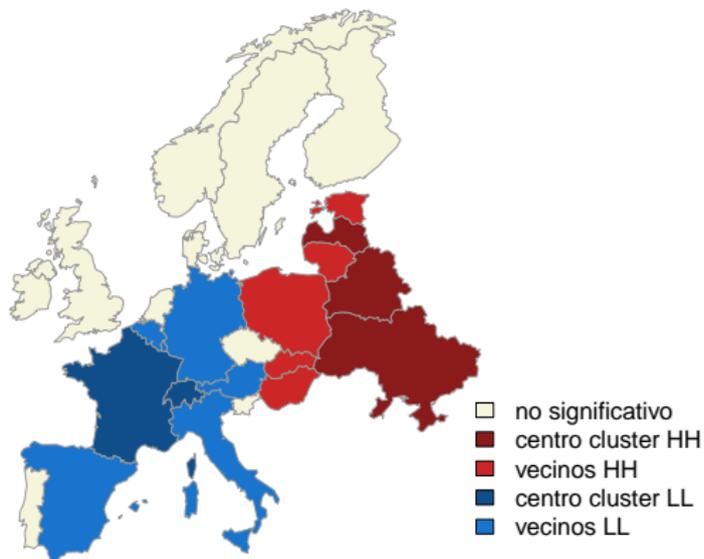
Resultados. Estadísticos

Valores del Índice de Moran Global y p-valores. Femenino

##		IMoran	p-value
##	N1995F	0.7114693	3.453038e-06
##	N1996F	0.7093587	3.490831e-06
##	N1998F	0.7397135	1.567133e-06
##	N1999F	0.7050235	3.644924e-06
##	N3002F	0.7007049	3.636570e-06
##	N2002F	0.6920223	4.113442e-06
##	N2004F	0.7090959	2.638732e-06
##	N2005F	0.7134444	2.191594e-06
##	N2007F	0.7186111	1.867723e-06
##	N2008F	0.7169178	1.658814e-06
##	N2010F	0.7017607	2.499654e-06
##	N2011F	0.6951033	3.083845e-06

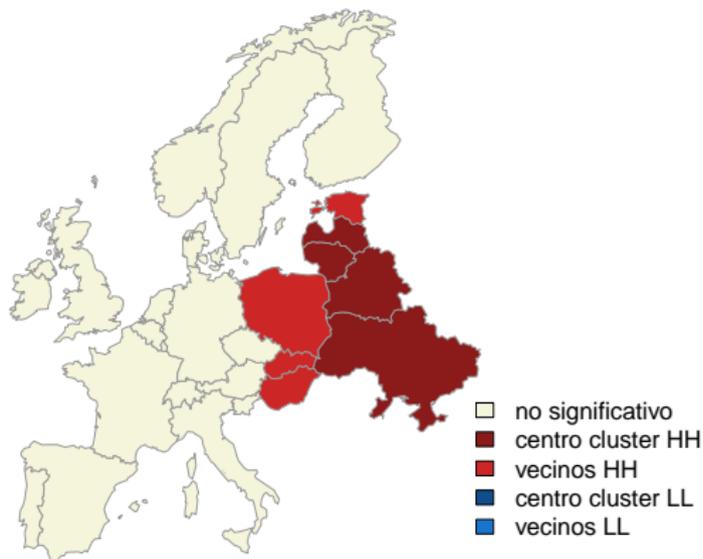
Resultados. Estadísticos

Índice de Moran Local en Europa. Masculino Año 1995.



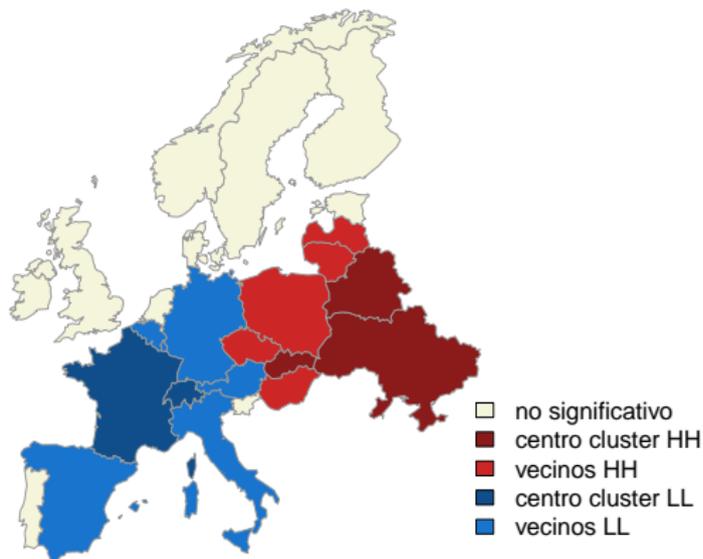
Resultados. Estadísticos

Índice de Moran Local en Europa. Masculino Año 2012.



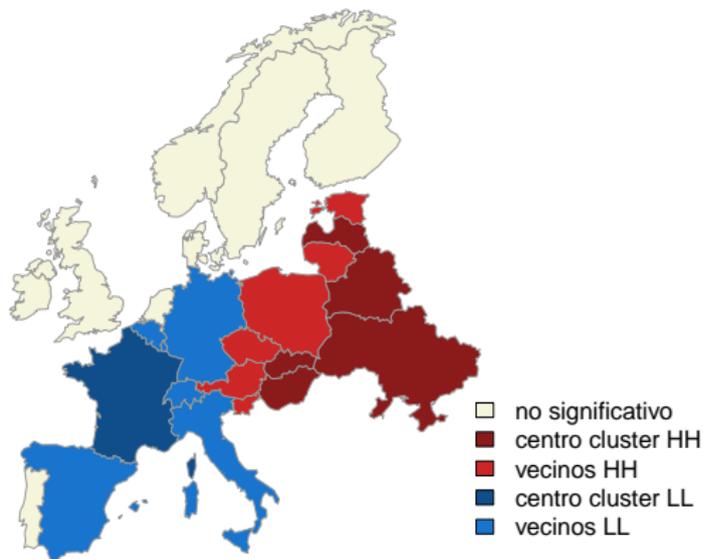
Resultados. Estadísticos

Índice de Moran Local en Europa. Femenino Año 1995.



Resultados. Estadísticos

Índice de Moran Local en Europa. Femenino Año 2012.



Resultados. Estadísticos

Test F para la inclusión efectos espaciales y/o temporales.
Masculino.

Prueba de **efectos fijos** espaciales y/o de tiempo basados en la comparación del modelo con efectos fijos y la regresión OLS (Baltagi 2008).

```
load("modelos.Rdata")  
library(plm)  
pFtest(gdm, gpm)
```

```
##  
## F test for twoways effects  
##  
## data: f15  
## F = 189.79, df1 = 42, df2 = 421, p-value < 2.2e-16  
## alternative hypothesis: significant effects
```

Resultados. Estadísticos

Test F para la inclusión efectos espaciales y/o temporales.
Femenino.

Prueba de **efectos fijos** espaciales y/o de tiempo basados en la comparación del modelo con efectos fijos y la regresión OLS (Baltagi 2008).

```
load("modelos.Rdata")  
library(plm)  
pFtest(gdf, gpf)
```

```
##  
## F test for twoways effects  
##  
## data: f5  
## F = 323.03, df1 = 42, df2 = 423, p-value < 2.2e-16  
## alternative hypothesis: significant effects
```

Resultados. Estadísticos

Prueba Multiplicador de Lagrange (LM) para la dependencia espacial

Analizar si el modelo de datos de panel seleccionado se puede convertir en un **modelo espacial** (Burridge 1980; Anselin 1988).

Masculino

```
##  
## LM = 198.15, df = 1, p-value < 2.2e-16
```

Femenino

```
##  
## #LM = 1.3138, df = 1, p-value = 0.2517
```

Modelos de datos de panel

Modelo SLMSTFE sexo masculino. Estimación de parámetros

Modelo	α	tPIB	gsan	CO ₂	gedu	λ	R ²	AIC	BIC
1	0.097*	0.056*				0.601*	0.986	-328.5	-133.5
2	0.192*		-0.019*			0.608*	0.986	-316.4	-121.4
3	0.142*			-0.086*		0.631*	0.986	-304.8	-109.8
4	0.076*				0.092*	0.603*	0.986	-318.7	-123.7
5	0.186*	0.055*	-0.019*			0.595*	0.987	-350.7	-151.6
6	0.137*	0.056*		-0.089*		0.620*	0.987	-339.1	-140.0
7	0.588*	0.058*			0.097*	0.588*	0.987	-356.3	-157.2
8	0.208*		-0.016*	-0.059*		0.621*	0.986	-120.4	-319.5
9	0.181*		-0.023*		0.112*	0.592*	0.987	-352.5	-153.4
10	0.115*			-0.085*	0.091*	0.621*	0.986	-328.1	-129.0
11	0.174*	0.058*	-0.024*		0.117*	0.576*	0.988	-393.4	-190.2
12	0.202*	0.055*	-0.016*	-0.061*		0.609*	0.987	-354.6	-151.3
13	0.194*		-0.021*	-0.049*	0.109*	0.604*	0.987	-354.4	-151.1
14	0.109*	0.058*		-0.087*	0.095*	0.608*	0.987	-367.3	-164.0
15	0.188*	0.058*	-0.021*	-0.051*	0.114*	0.589*	0.988	-395.9	-188.5

*p-valores < 0.05 son significativos

Modelos de datos de panel

Modelo SLMSTFE sexo masculino. Estimación de parámetros

Modelo	α	tPIB	gsan	CO ₂	gedu	λ	R ²	AIC	BIC
15	0.188*	0.058*	-0.021*	-0.051*	0.114*	0.589*	0.988	-395.95	-188.5
11	0.174*	0.058*	-0.024*		0.117*	0.576*	0.988	-393.4	-190.2
14	0.109*	0.058*		-0.087*	0.095*	0.608*	0.987	-367.3	-164.0
7	0.588*	0.058*			0.097*	0.588*	0.987	-356.3	-157.2
12	0.202*	0.055*	-0.016*	-0.061*		0.609*	0.987	-354.6	-151.3
13	0.194*		-0.021*	-0.049*	0.109*	0.604*	0.987	-354.4	-151.1
9	0.181*		-0.023*		0.112*	0.592*	0.987	-352.5	-153.4
5	0.186*	0.055*	-0.019*			0.595*	0.987	-350.7	-151.6
6	0.137*	0.056*		-0.089*		0.620*	0.987	-339.1	-140.0
1	0.097*	0.056*				0.601*	0.986	-328.5	-133.5
10	0.115*			-0.085*	0.091*	0.621*	0.986	-328.1	-129.0
4	0.076*				0.092*	0.603*	0.986	-318.7	-123.7
2	0.192*		-0.019*			0.608*	0.986	-316.4	-121.4
3	0.142*			-0.086*		0.631*	0.986	-304.8	-109.8
8	0.208*		-0.016*	-0.059*		0.621*	0.986	-120.4	-319.5

*p-valores <0.05 son significativos

Modelos de datos de panel

Modelo MSTFE sexo femenino. Estimación de parámetros

Modelo	α	tPIB	CO ₂	gedu	R ²	AIC	BIC
1	-0.1587*	0.0340*			0.9713	-1560.08	-1377.55
2	-0.1258*		-0.0348		0.9701	-1541.41	-1358.87
3	-0.1757*			0.0493*	0.9708	-1551.43	-1368.90
4	-0.1302*	0.034*	-0.0371		0.9714	-1560.60	-1373.92
5	-0.1837*	0.0355*		0.0526*	0.9722	-1572.59	-1385.90
6	-0.1446*		-0.0414.	0.0512*	0.9709	-1552.50	-1365.82
7	-0.1506 *	0.0358 *	-0.0442	0.0547 *	0.9724	-1574.26	-1383.43

Modelos de datos de panel

Modelo MSTFE sexo femenino. Estimación de parámetros

Modelo	α	tPIB	CO ₂	gedu	R ²	AIC	BIC
5	-0.1837*	0.0355*		0.0526*	0.9722	-1572.59	-1385.90
7	-0.1506*	0.0358*	-0.0442	0.0547*	0.9724	-1574.26	-1383.43
1	-0.1587*	0.0340*			0.9713	-1560.08	-1377.55
4	-0.1302*	0.034*	-0.0371		0.9714	-1560.60	-1373.92
3	-0.1757*			0.0493*	0.9708	-1551.43	-1368.90
6	-0.1446*		-0.0414.	0.0512*	0.9709	-1552.50	-1365.82
2	-0.1258*		-0.0348		0.9701	-1541.41	-1358.87

Modelos de datos de panel

Análisis residuos. Masculino

##		IMoran	p-value
##	H1995	-0.09161768	0.6222498
##	H1996	0.03919258	0.3129192
##	H1998	-0.14157265	0.7288618
##	H1999	-0.05227458	0.5301477
##	H2001	-0.22517731	0.8816863
##	H2002	-0.65277390	0.9999729
##	H2004	0.02699298	0.3390975
##	H2005	0.05172969	0.2678906
##	H2007	0.04152595	0.3053787
##	H2008	-0.03412481	0.4857147
##	H2010	-0.07949020	0.5944786
##	H2011	-0.23127243	0.8788402

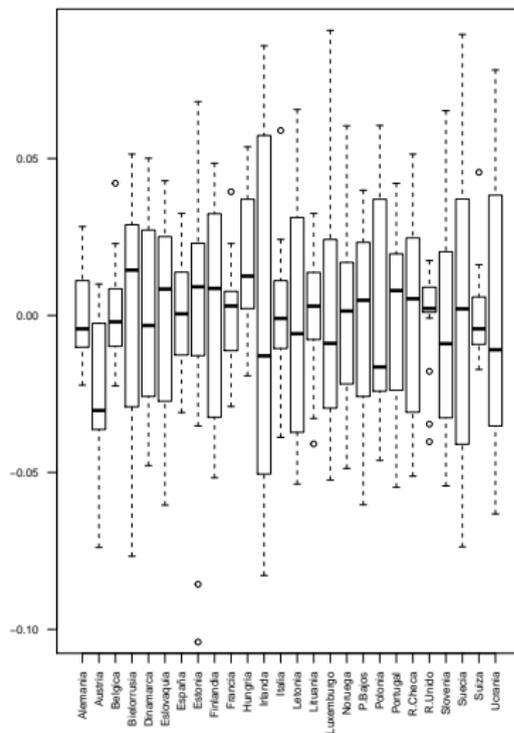
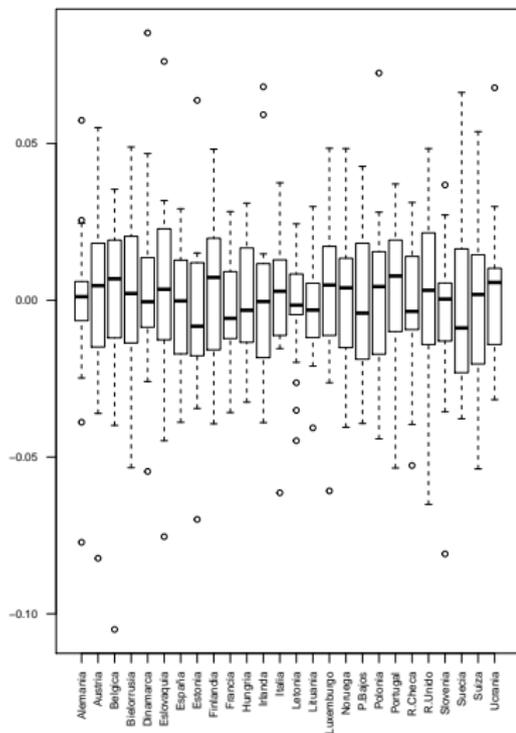
Modelos de datos de panel

Análisis residuos. Femenino

##		IMoran	p-value
##	M1995	-0.1306084667	0.70755748
##	M1996	-0.1257689341	0.69630042
##	M1998	0.0350007133	0.32504449
##	M1999	0.0359824951	0.31850401
##	M2001	0.0049516117	0.39163618
##	M2002	0.1161610954	0.17553104
##	M2004	0.0008335857	0.40274344
##	M2005	0.2125785849	0.06269785
##	M2007	-0.1186325812	0.68244346
##	M2008	-0.0387406603	0.49696196
##	M2010	0.1499913292	0.12729956
##	M2011	0.0174498385	0.23660190

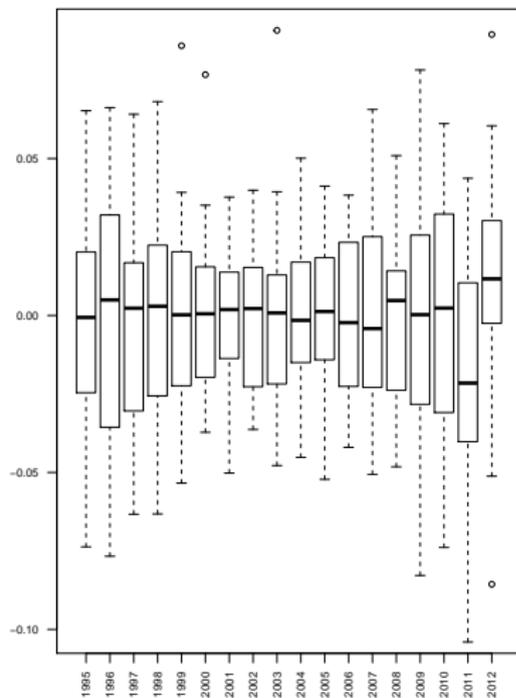
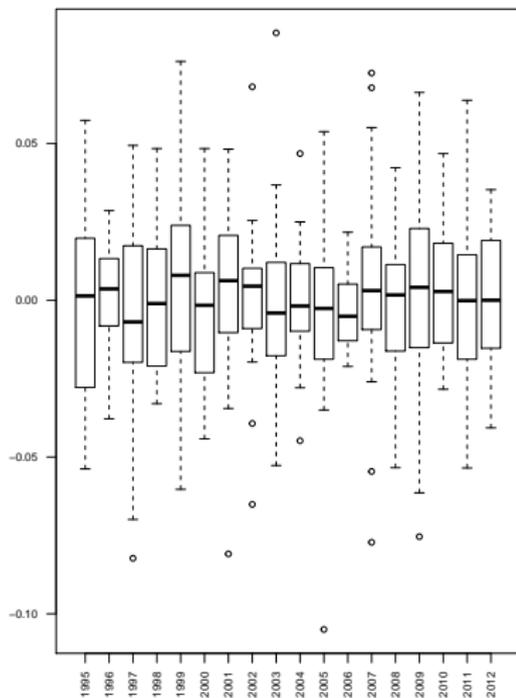
Modelos de datos de panel

Análisis residuos. Masculino vs Femenino



Modelos de datos de panel

Análisis residuos. Masculino vs Femenino



Conclusiones

Conclusiones

- ▶ El **CMF** en los países del **este de Europa es superior** al del oeste de Europa. En concreto, el CMF es mayor en los hombres que en las mujeres.
- ▶ Existe una **correlación espacial significativa** en el CMF de los 26 países europeos durante 1995-2012 en ambos sexos como indica el GM.
- ▶ LM ha detectado **2 clusters significativos de países europeos**. Un cluster **HH** formado por países del este de Europa y otro cluster **LL** compuesto por países del oeste de Europa. A partir del año 2007 el cluster de los países del este desaparece en el sexo masculino.

Conclusiones

- ▶ El modelo de retardo espacial con efecto fijo espacial y temporal en el sexo masculino es un buen modelo tal y como indican las medidas de bondad de ajuste (R^2 , AIC y BIC) y el análisis de los residuos.
- ▶ El nivel esperado del (CMF) de un país aumentaría un 60% si sus vecinos aumentaran un 1% el \overline{CMF} .
- ▶ Variaciones de las cuatro covariables provocarán variaciones en el $\log(CMF)$ de un país en un año.

Conclusiones

- ▶ El **modelo con efecto fijo espacial y temporal** en el sexo femenino es un buen modelo tal y como indican las medidas de bondad de ajuste (R^2 , AIC y BIC) y el análisis de los residuos.
- ▶ Variaciones de las **dos covariables** (tasa PIB y gasto educación) provocarán variaciones en el $\log(CMF)$ de un país en un año.

Agradecimientos

Esta investigación ha sido financiada por la convocatoria de proyectos de investigación Ignacio H. de Larramendi de la *Fundación Mapfre* de 2017.

References

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic.

———. 1995. “Local Indicators of Spatial Association—LISA.” *Geographical Analysis* 27 (2). Blackwell Publishing Ltd: 93–115. <http://dx.doi.org/10.1111/j.1538-4632.1995.tb00338.x>.

Baltagi, B. 2008. *Econometric Analysis of Panel Data*. John Wiley & Sons.

Burrige, P. 1980. “On the Cliff-Ord Test for Spatial Autocorrelation Among Regression Residuals.” *Geographical Analysis* 4: 267–84.

Elhorst, J.P. 2014. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer.

Hsiao, C. 2014. “Analysis of Panel Data (3rd Edition).” In. Cambridge: Cambridge University Press.

Human Mortality Database. 2016. *University of California, Berkeley (Usa), and Max Planck Institute for Demographic Research*