

SELECCIÓN DE MODELOS ESPACIO-TEMPORALES CON DATOS DE PANEL EN MATLAB Y R

PATRICIA CARRACEDO

patcarga@posgrado.upv.es
Universitat Politècnica de València, Centro de Gestión de la Calidad y del Cambio,
Camino de Vera, s/n 46022 Valencia
Valencian International University (VIU)

ANA DEBÓN

andeau@eio.upv.es
Universitat Politècnica de València, Centro de Gestión de la Calidad y del Cambio,
Camino de Vera, s/n 46022 Valencia

Recibido (05/06/2017)
Revisado (05/06/2017)
Aceptado (05/06/2017)

RESUMEN: El imparable interés en la econometría espacial, las novedosas técnicas estadísticas, la mayor disponibilidad de datos de panel y los avances en la tecnología de software han dado lugar al desarrollo de nuevos paquetes para modelar la dependencia espacial y temporal en un panel de datos. El objetivo de este trabajo es seleccionar el mejor modelo espacio-temporal con datos de panel utilizando dos softwares: MATLAB y R para su posterior comparación. La metodología estadística empleada tiene en cuenta las relaciones de vecindad entre las unidades espaciales a lo largo del tiempo. Los modelos finalmente seleccionados se validaron mediante dos medidas de bondad de ajuste: el coeficiente de determinación y la varianza residual. Además, sus coeficientes se interpretaron incluyendo los efectos directos e indirectos (spillover espacial). El panel de datos utilizado se corresponde con la mortalidad de 26 países europeos durante el periodo 1990-2009.

Palabras Clave: Modelo espacio-temporal con Datos de Panel; Mortalidad; Europa; MATLAB; R.

ABSTRACT: The unstoppable interest in spatial econometrics, new statistical techniques, increased availability of panel data and advances in software technology have led to development of new packages for modeling spatio-temporal dependence in a panel of data. The objective of this paper is to select the best spatio-temporal panel data model using two software packages: MATLAB and R for further comparison. The statistical methodology used takes into account the neighborhood relations between the spatial units over time. The models finally selected were validated by two measures of goodness of fit: the coefficient of determination and the residual variance. In addition, their coefficients were interpreted including direct and indirect effects (spatial spillover). The data panel used corresponds to the mortality of 26 European countries for the period 1990-2009.

Keywords: Spatio-temporal Panel Data Models; Mortality; Europe; MATLAB; R.

EXTENDED SUMMARY

Mortality can be defined as the number of people who die in a certain place over a period of time in relation to the total population. In particular, Europe is a continent with great variability in its mortality, especially in two large groups of countries: the countries of Eastern Europe, also known as East Europe, is the eastern part of the European continent and the countries of Western Europe, also known as West Europe, is the western part of the European continent. This difference between Eastern Europe and Western Europe mainly occurred after the collapse of the Soviet Union (20th century) and it is increasing with time (Vågerö, 2010). Therefore, mortality in Europe has not only varied over time, but has also varied according to the country, since not all of them have the same health and economic conditions (EUROSTAT, 2009).

On the other hand, socioeconomic variables such as sex, age, race and others, depending on the geographic unit considered, can influence the risk of disease and therefore mortality. The observed differences in mortality are confounded by these variables (Rezaeian et al., 2007); so that in order to compare crude mortality rates from different geographic areas, it is necessary to standardize these rates. The indirect standardization method produces the Standardized Mortality Ratio (SMR) (Fleiss et al., 2013). This ratio is calculated by dividing the observed number of deaths within the study population by the expected number of deaths in the study population, assuming that the age-specific rates in the reference population also applied to the population under study (Hinde, 1998; Julious et al., 2001; Rezaeian et al., 2007). Therefore, in this study, mortality was quantified using the SMR.

In addition, knowledge that mortality is concentrated in identifiable geographic areas is of vital importance. For example, in the field of public policies it helps to prevent, treat and improve mortality as well as indicating an efficient distribution of resources (Rezaeian et al., 2007). In the actuarial field, it helps actuaries to design pensions and life insurance. These spatial concentrations are produced by the existence of autocorrelation or spatial dependence, which implies that the mortality of geographically close areas is more related than that of geographically distant areas. This is because neighboring countries may have similar social, economic and cultural characteristics. The detection of spatial dependence is achieved using of spatial statistics.

Spatial statistics provides techniques for the study of spatial autocorrelation, which offer very useful information about the spatial distribution of data over time. Specifically, this paper selects and implements the best spatio-temporal model of panel data in order to model space-time autocorrelation using MATLAB and R software.

Thus, the basis of this paper lies in the ability to incorporate spatial dependence in the mortality of European countries over time by means of the model that best represents the geographical location between countries using MATLAB and R.

A panel data model is a regression model that uses the temporal and spatial dimension of the data for to estimate the parameters of interest. This model can control unobserved heterogeneity produced by both the spatial units studied and by time. Unobserved heterogeneity is produced by the variables that we do have not information for, but affect the variable in question. This unobservable heterogeneity can only be detected with panel data models (Hsiao, 2014; Elhorst, 2014). The importance of controlling this heterogeneity lies in the fact that it reduces problems of multicollinearity between variables by permitting parameter estimates in panel models to be more efficient (Kennedy, 2003; Baltagi, 2008).

Although the role of place in human health has been recognized historically, the focus in public health research has mostly been on people and time, with little consideration of the implications

of place over time (Rezaeian et al., 2007). This is mainly due to several reasons such as a dearth of appropriate databases and lack of adequate software to study data with spatial and temporal dimensions, that is, panel data (Rezaeian et al., 2007; Millo and Piras, 2012). So far, only the `splm` package in R (Millo and Piras, 2012), the code supplementing Kapoor et al. (2007) in Stata and the functions available from Elhorst (2011).

As methodological innovation is related to the development of new packages, the objective of this paper is to present the spatio-temporal methodology proposed by Elhorst (2014) to select the best space-time panel data model using the MATLAB statistical software. In addition, this methodology is implemented in R for further comparison, as it is in our interest to apply and show this methodology in the R statistical software, which is free, and therefore, within the reach of all.

Considering the defined objective, this work has been structured as follows: Section 2 starts by describing the data used in this study. Next, it shows the most used spatio-temporal panel data models in spatial econometrics. This section ends explaining the spatial methodology used to select the best spatio-temporal panel data model. Section 3 shows the results of applying the statistical methodology defined in the previous section using MATLAB and R software. This section ends interpreting the best space-time panel data model obtained with each software package including direct and indirect effects (spatial spillover). Finally, Section 4 presents the most important conclusions reached in this study.

Detailed statistical methodology at the theoretical level was applied to mortality data from 26 European Union countries for an age range of 0 to 110+ and a time range of 1990-2009. This data was downloaded from Human Mortality Database (2014). The logarithm of the SMR was modeled to correct the asymmetry and to approximate its distribution to the normal distribution. Thus, the logarithm of the SMR was the dependent variable in the spatio-temporal panel data model. Before fitting the model, a multicollinearity study was conducted, that showed that the independent variables to be considered in the model were the crude birth rate, activity rate, energy consumption of the road sector and GDP. The population growth variable caused multicollinearity between independent variables.

Finally, the best space-time panel data model obtained with R software was a panel data model with spatial and temporal fixed effects and spatial lag in the dependent variable (SLMSTFE). While the best space-time panel data model obtained with the MATLAB software was a spatio-temporal Durbin model (SDM). The difference between these models is that the latter also includes the spatial correlation in the covariables.

The selected models were validated using two measures of goodness of fit: the coefficient of determination R^2 and the residual variance σ^2 . These measures indicated that the SLMSTFE and SDM models are good, since the R^2 is close to 1 and σ^2 is small compared to the variance of the dependent variable, in both models.

In relation to the work of other authors, it should be pointed out that to our knowledge, although the differences in mortality have been studied, no spatial study about Standardized Mortality Ratio in Europe using panel data models has been carried out. In addition, a comparison of the MATLAB and R software has not been shown to select the best space-time panel data model. This is important because of the growing popularity of free R software in research compared to commercial software.

1. Introducción

Debido a los grandes avances informáticos, cada vez en mayor medida, se generan y almacenan un mayor número de datos, principalmente datos de panel. Un panel de datos está formado por varias unidades geográficas seguidas a lo largo del tiempo. Éstos contienen más información que los datos temporales y de corte transversal debido a que contienen más variabilidad, menos colinealidad, más grados de libertad y más eficiencia entre las variables (Baltagi, 2008). Así pues, el aumento de la disponibilidad de paneles de datos, ha estimulado un avance metodológico muy importante en el análisis de datos de panel espaciales. La econometría espacial ha considerado modelos con datos de panel donde la variable dependiente o el error están espacialmente autocorrelacionados, incluyendo además, efectos fijos o aleatorios. Estos modelos tienen en cuenta la distancia geográfica de las unidades espaciales, y por tanto, las relaciones de vecindad entre ellas a lo largo del tiempo.

Aunque existen paquetes para estimar modelos espaciales transversales en R, Stata y MATLAB, son escasos los paquetes que estiman modelos espacio-temporales de datos de panel (Millo and Piras, 2012). Hasta ahora, están disponibles el paquete `splm` (Millo and Piras, 2012) en R, el código que complementa a Kapoor et al. (2007) en Stata y por último el código de Elhorst (2011) en MATLAB.

Como la innovación metodológica está relacionada con el desarrollo de nuevos paquetes, este trabajo presenta la metodología estadística propuesta por Elhorst (2014) para seleccionar el mejor modelo espacio-temporal con datos de panel utilizando el software estadístico MATLAB. Adicionalmente, dicha metodología se implementa en R para su posterior comparación.

El conjunto de datos utilizado se corresponde con un panel de datos de mortalidad de 26 países europeos durante el periodo de tiempo 1990-2009. La mortalidad en Europa ha variado en función tanto del tiempo como entre países, ya que no todos disponen de las mismas condiciones sanitarias y económicas (EUROSTAT, 2009). Europa es un continente considerado desarrollado pero los países que lo componen han tenido un progreso diferente entre sí, lo cual ha dado lugar a la existencia de gran variabilidad entre sus tasas de mortalidad.

Así pues, la base de este trabajo radica en la capacidad de incorporar la dependencia espacial en la mortalidad de países europeos a lo largo del tiempo seleccionando el modelo que mejor represente la distancia geográfica entre los países utilizando los softwares MATLAB y R.

El trabajo se ha estructurado de la siguiente manera: La Sección 2 empieza detallando los datos que se han utilizado en este estudio. A continuación muestra los diferentes modelos espacio-temporales con datos de panel que más se utilizan en la econometría espacial. La sección finaliza explicando la metodología espacial utilizada para seleccionar el mejor modelo espacio-temporal con datos de panel. La Sección 3 muestra los resultados de aplicar la metodología estadística definida en la sección anterior utilizando los softwares MATLAB y R. La sección acaba interpretando el mejor modelo espacio-temporal que se ha podido conseguir con cada software. Finalmente, la Sección 4 presenta las conclusiones más importantes obtenidas en este trabajo.

2. Materiales y Métodos

2.1. Datos

En este estudio se han considerado datos de mortalidad correspondientes a 26 países europeos durante el periodo de tiempo comprendido entre 1990 y 2009. Estos datos, han sido descargados de Human Mortality Database (2014) para los siguientes países: Alemania, Austria, Bélgica,

Bielorrusia, Dinamarca, Eslovaquia, Eslovenia, España, Estonia, Finlandia, Francia, Hungría, Irlanda, Italia, Letonia, Lituania, Luxemburgo, Noruega, Países Bajos, Polonia, Portugal, Reino Unido, República Checa, Suecia, Suiza y Ucrania. Se han tomado estos 26 países porque, sólo éstos, tienen información común en la base de datos para el rango máximo de tiempo 1990-2009 y rango de edad 0-110+.

Cuantificar la mortalidad es importante para perfilar los niveles epidemiológicos, demográficos y de desarrollo de cualquier país. La mortalidad está influida por factores de carácter aleatorio como por ejemplo enfermedades o desastres naturales. Así pues, la mortalidad no permanece uniforme, es decir, varía a lo largo del tiempo. Los profesionales de la salud pública se enfrentan constantemente con la necesidad de comparar la mortalidad entre diferentes áreas geográficas. Si las poblaciones se distribuyeran de manera similar respecto a factores como la edad, raza, sexo o clase social no habría problema en comparar tasas de mortalidad, pero en realidad esto no sucede. Al comparar las tasas de mortalidad entre diferentes áreas, estas tasas están influenciadas por la proporción de individuos en cada grupo de edad de cada área geográfica. Para solucionar este problema, se desarrollaron los métodos de estandarización. La estandarización permite comparaciones de la mortalidad entre diferentes áreas geográficas eliminando los efectos que producen los diferentes tamaños en los subgrupos de la población (Julious et al., 2001). Hay dos métodos de estandarización: directo e indirecto (Fleiss et al., 2013). Para poder cuantificar la mortalidad, hemos utilizado el Ratio de Mortalidad Estandarizado (SMR) cuyas siglas provienen del inglés Standardized Mortality Ratio, debido a que es el índice más utilizado para comparar la mortalidad entre diferentes áreas (Hinde, 1998). El método indirecto de estandarización produce el SMR. Este ratio se ha obtenido para cada uno de los 26 países europeos considerados durante el periodo de tiempo 1990-2009. A continuación, se detalla.

El SMR se obtiene como el cociente entre el número real de muertes observadas en una población estudiada durante un período de tiempo considerado, y el número de muertes que se esperarían en esa población si tuviera la misma mortalidad que la población de referencia (Hinde, 1998). En este trabajo, las poblaciones estudiadas son los países y como población de referencia se toma el conjunto de países europeos. Su cálculo se expresa como

$$SMR_{i,t} = \frac{O_{i,t}}{E_{i,t}} \quad \text{para } i \in \{1, \dots, N\} \quad \text{y } t \in \{1, \dots, T\} \quad (1)$$

donde i indica el país y t es el periodo de tiempo, expresado en años. $O_{i,t}$ representa el número de muertes observadas en cada país cada país i en el año t y $E_{i,t}$ corresponde a las muertes esperadas en cada país i en el año t bajo la hipótesis de que cada país tiene la misma mortalidad que el conjunto de países de Europa. Si x es la edad de la muerte, $O_{i,t}$ se calcula como

$$O_{i,t} = \sum_{x=0}^{110+} m_{x,i,t} P_{x,i,t}, \quad (2)$$

donde $m_{x,i,t}$ representa la tasa de mortalidad y $P_{x,i,t}$ la población estudiada a la edad x , país i y año t . Las $E_{i,t}$ se pueden obtener como

$$E_{i,t} = \sum_{x=0}^{110+} E_{x,i,t}, \quad (3)$$

donde $E_{x,i,t}$ se define como

$$E_{x,i,t} = m_{x,t}P_{x,i,t}, \quad (4)$$

siendo $m_{x,t}$ la tasa de mortalidad a la edad x y año t en el conjunto de países europeos. Si $O_{x,i,t}$ son las muertes observadas a la edad x , país i y año t , $m_{x,t}$ se obtiene según la ecuación (5)

$$m_{x,t} = \frac{\sum_{i=1}^N O_{x,i,t}}{\sum_{i=1}^N P_{x,i,t}}. \quad (5)$$

Si el SMR es menor que 1, hay un menor número de muertes observadas que las esperadas. En este caso hay un “déficit de muertes” en ese país. Por contra, hablaremos de “exceso de muertes” si el SMR es mayor que 1, situación que ocurre cuando hay un mayor número de muertes observadas que esperadas.

Además, para estos 26 países y 20 años, se ha recopilado información correspondiente a 5 covariables: crecimiento de la población, producto interior bruto (PIB), tasa de natalidad, tasa de actividad y consumo de energía del sector vial de la base de datos The World Bank Database (2015). A continuación se detallan cada una de ellas:

- *Crecimiento de la población (% anual)*.- La tasa de crecimiento anual de la población para el año t es la tasa exponencial de crecimiento de la población a mediados del año $t - 1$ al t , expresada como porcentaje.
- *Producto Interior Bruto (% anual)*.- Tasa de crecimiento anual porcentual del PIB a precios de mercado basada en la moneda local constante. El PIB es la suma del valor añadido bruto de todos los productores residentes en la economía más los impuestos sobre los productos y menos las subvenciones no incluidas en el valor de los productos. Se calcula sin hacer deducciones por depreciación de activos manufacturados o por agotamiento y degradación de recursos naturales. Esta variable contiene datos faltantes a diferencia del resto. Para imputar dichos datos, se utilizó la función `knnImputation` del paquete `DMwR` de Torgo (2010). Esta función completa los valores faltantes con un promedio local ponderado.
- *Tasa bruta de natalidad (por 1.000 personas)*.- La tasa bruta de natalidad indica el número de nacidos vivos que se producen en un año por cada 1.000 habitantes.
- *Tasa de actividad (% de la población total de edades ≥ 15)*.- La tasa de actividad es la proporción de la población de 15 años o más que es económicamente activa. Esta población la forman todas las personas que suministran mano de obra para la producción de bienes y servicios durante un período determinado.
- *Consumo de energía del sector vial (% del consumo total de energía)*.- El consumo de energía en el sector vial es la proporción de la energía total utilizada en el sector vial, incluyendo productos petrolíferos, gas natural, electricidad y combustibles renovables y residuos.

Los análisis estadísticos se realizaron utilizando dos softwares: MATLAB (MathWorks, 2015) y R (R Core Team, 2015). Las rutinas de MATLAB para estimar los modelos espacio-temporales, se descargaron el 06/05/2015 en la siguiente web <http://www.regroningen.nl/>. Éstas fueron escritas por Elhorst (2011). Los paquetes utilizados en R para obtener los modelos espacio-temporales fueron: `spdep` (Bivand, 2012), `splm` (Millo and Piras, 2012) y `plm` (Croissant et al., 2008).

2.2. Modelos espacio-temporales con datos de panel

Antes de definir los modelos espacio-temporales con datos de panel, cabe detallar brevemente qué son datos de panel o un panel de datos. Los datos de panel son observaciones de un mismo corte seccional por ejemplo familias, empresas, países o provincias seguidas durante varios periodos de tiempo. Estos datos son una combinación de dos dimensiones: la espacial y la temporal (Wooldridge, 2010).

Dependiendo de la amplitud espacial y temporal de los datos, los paneles de datos pueden ser de tres tipos:

- *Paneles Micro*.- Paneles con más observaciones transversales que observaciones temporales.
- *Paneles Macro*.- Son paneles con más observaciones temporales que observaciones transversales.
- Panel Campo Aleatorio “Random Field”.- Son paneles con una dimensión temporal y transversal muy amplia.

Dependiendo de la existencia o ausencia de datos faltantes, los datos de panel puede ser de dos tipos:

- *Paneles Balanceados o Completos*.- Si todas las unidades estudiadas se observan a lo largo del periodo estudiado.
- *Paneles Desequilibrados o Incompletos*.- Si el rango temporal varía entre los individuos. En otras palabras, si existen datos faltantes en el panel considerado.

En concreto, los datos que utilizamos en este estudio forman un *Panel Micro*, ya que disponemos de más observaciones espaciales (26 países) que periodos de tiempo (20 años) y un *Panel Balanceado* ya que no hay datos faltantes, dado que se ha hecho una imputación de valores faltantes para el PIB.

Actualmente, la econometría espacial está centrando su interés en la especificación y estimación de modelos econométricos basados en la información que contienen los paneles de datos debido principalmente a dos razones. La primera es que cada vez con mayor frecuencia, se generan y se almacenan un gran número de datos de panel que pueden ser analizados. Y la segunda, es que un panel de datos contiene más información y más variabilidad que los datos de corte transversal o series temporales ofreciendo a los investigadores otras posibilidades de modelado (Elhorst, 2014).

Un modelo 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. Estos modelos, permiten modelar la heterogeneidad no observada producida tanto por las unidades espaciales estudiadas como por el tiempo. La heterogeneidad no observada la producen las variables de las cuales no tenemos información pero afectan a la variable de interés. Ejemplos de variables no observables que varían para cada unidad transversal pero permanecen constantes con el tiempo afectando a la variable dependiente del modelo podrían ser el sexo, la religión o la delincuencia. De manera similar, ejemplos de variables no observables que varían a lo largo del tiempo pero permanecen constantes entre las unidades transversales afectando a la variable endógena del modelo podrían ser una recesión/expansión económica, guerra mundial, cambios en la legislación o en la política gubernamental. Esta heterogeneidad no observable sólo se puede detectar con modelos de datos de panel (Hsiao, 2014). La importancia de controlar esta heterogeneidad radica en el hecho de que reduce problemas de multicolinealidad entre las variables permitiendo que las estimaciones de los parámetros sean más eficientes (Kennedy, 2003).

Los modelos espacio-temporales de datos de panel tienen en cuenta las relaciones de vecindad entre las unidades geográficas a lo largo del tiempo. Éstos pueden utilizarse para explicar el

comportamiento de unidades geográficas, siempre que se encuentren relacionadas entre sí.

El modelo más simple para analizar un panel de datos es calcular una regresión por mínimos cuadrados ordinarios. Este modelo no tiene en cuenta la dimensión espacial ni temporal de los datos y supone que la constante en el modelo es la misma para todas las unidades transversales y/o temporales.

Para poder estimar el carácter individual de cada unidad transversal y/o temporal nacen dos modelos: el modelo de efectos fijos y el modelo de efectos aleatorios. El modelo de efectos fijos, considera que el efecto tiene un valor fijo para cada unidad transversal y/o temporal, es decir, que las diferencias entre las unidades transversales y/o temporales son constantes. Mientras que para el modelo de efectos aleatorios considera que el efecto es una variable aleatoria con una media y una varianza no nula, es decir, que las diferencias entre las unidades transversales y/o temporales son aleatorias (Baltagi, 2008). Este último trata las diferencias del efecto aleatorio en la varianza de error y supone que las variables aleatorias y el error del modelo son independientes (Wooldridge, 2010).

Elhorst (2014) proporciona una revisión de los modelos espacio-temporales de datos de panel más comúnmente utilizados en la investigación:

- *Modelo de retardo espacial con efecto fijo espacial y temporal (SLMSTFE)*. Sus siglas provienen del inglés “Spatial Lag Model with Spatial and Time Fixed Effects”.
 - *Modelo de error espacial con efecto fijo espacial y temporal (SEMSTFE)*. Sus siglas provienen del inglés “Spatial Error Model with Spatial and Time Fixed Effects”.
- Cabe destacar que en estos modelos se puede considerar el efecto fijo espacial y/o temporal. En este trabajo se detalla el modelo con ambos efectos conocido como “two ways”, puesto que es el modelo general, ya que a partir de éste se obtiene el modelo de retardo espacial con efecto fijo espacial y el modelo de retardo espacial con efecto fijo temporal.
- *Modelo de retardo espacial con efecto aleatorio (SLMSTRE)*. Sus siglas provienen del inglés “Spatial Lag Model with Random Effects”. En este modelo, el efecto aleatorio es el efecto espacial, el temporal se considera fijo.
 - *Modelo de error espacial con efecto aleatorio (SEMRE)*. Sus siglas provienen del inglés “Spatial Error Model with Random Effects”.
 - *Modelo espacio-temporal Durbin (SDM)*. Sus siglas provienen del inglés “Spatial Durbin Model”. Este modelo incluye la correlación espacial en las covariables de los modelos SLMSTFE y SLMSTRE.
 - *Modelo de error espacio-temporal Durbin (SDEM)*. Sus siglas provienen del inglés “Spatial Durbin Error Model”. Este modelo incluye la correlación espacial de las covariables en los modelos SEMSTFE y SEMRE.

A continuación se proporcionan las ecuaciones matemáticas de los modelos que, finalmente, se implementarán.

- SLMSTFE

$$y_{it} = \alpha + \lambda \sum_{j=1}^N W y_{jt} + x_{it}\beta + \mu_i + \nu_t + \epsilon_{it} \quad (6)$$

- SLMSTRE

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

■ SDM

Incluyendo efectos fijos:

$$y_{it} = \alpha + \lambda \sum_{j=1}^N W y_{jt} + x_{it} \beta + \sum_{j=1}^N W x_{jt} \delta + \mu_i + \nu_t + \epsilon_{it} \quad (7)$$

Incluyendo efectos aleatorios espaciales y fijos temporales:

$$y_{it} = \alpha + \lambda \sum_{j=1}^N W y_{jt} + x_{it} \beta + \sum_{j=1}^N W x_{jt} \delta + \phi + \nu_t + \epsilon_{it}$$

donde y_{it} es un vector de dimensión $NT \times 1$ correspondiente a las observaciones de la variable dependiente para cada país i y año t , x_{it} es una matriz de dimensión $NT \times k$ de observaciones de las variables explicativas donde k el número de variables exógenas, W es la matriz de pesos espaciales o de contigüidad de dimensión $N \times N$ donde i y j representan dos países cualquiera de los N países totales y β es un vector de parámetros desconocidos asociados a las variables independientes de dimensión $k \times 1$. El resto son los parámetros que, dependiendo del modelo seleccionado, se estimarán:

- λ y δ son los parámetros espaciales. En concreto, λ es el parámetro autorregresivo espacial asociado a la variable dependiente y δ es un vector de dimensión $k \times 1$ de parámetros espaciales asociados a las variables independientes;
- μ_i es el efecto fijo espacial el cual recoge la heterogeneidad no observable producida por variables que cambian a través de los países, pero permanecen constantes en el tiempo. Es un vector de dimensión $N \times 1$;
- ν_t es efecto fijo temporal el cual recoge la heterogeneidad no observable producida por variables que cambian con el tiempo, pero permanecen constantes entre los países. Es un vector de dimensión $T \times 1$;
- α es el intercepto del modelo. Este modelo considera una variable dummy para cada unidad espacial y una variable dummy para cada unidad temporal (Elhorst, 2003). Los interceptos ($\alpha + \mu_i + \nu_t$) se estimarán bajo la restricción $\sum_{i=1}^N \mu_i = 0$ y $\sum_{t=1}^T \nu_t = 0$ (Baltagi, 2008; Hsiao, 2014). De esta forma el efecto espacial representa la desviación de la unidad espacial i respecto del promedio α y el efecto temporal representa la desviación de la unidad temporal t respecto del promedio α . Por ello, α es el efecto que recoge en términos medios la heterogeneidad no observable producida por variables que permanecen constantes entre países y con el tiempo.
- ϕ es el efecto aleatorio espacial independiente e idénticamente distribuido con media 0 y varianza $\sigma^2 \neq 0$;
- ϵ_{it} es un vector de términos del error independiente e idénticamente distribuido de dimensión $NT \times 1$, el cual recoge la heterogeneidad no observable producida por variables que cambian tanto a través de los países como a través del tiempo. Este vector no está espacialmente autocorrelacionado a diferencia de los modelos SEMSTFE, SEMRE y SDEM.

En este trabajo, se ha considerado estructura espacial de primer orden de forma que un país es vecino de otro cuando ambos comparten frontera. Con ello, $W = (\omega_{ij})$ puede tomar los siguientes valores:

$$\begin{aligned}
\omega_{ii} &= 0, \quad i = 1, \dots, N; \\
\omega_{ij} &= \frac{1}{n_i}, \quad \text{if } j \in V(i), \text{ with } n_i = \#V(i); \\
\omega_{ij} &= 0, \quad \text{if } j \notin V(i),
\end{aligned} \tag{8}$$

donde n_i es el número de vecinos de i y $V(i)$ el conjunto de vecinos del país i . De forma que si $\omega_{ij} = 0$ los países i y j no se consideran vecinos, mientras que si $\omega_{ij} \neq 0$ los países i y j sí que lo son con un peso $1/n_i$. Los valores de cada fila suman la unidad, porque los pesos ω_{ij} están estandarizados y un país no puede ser vecino de sí mismo dado que $\omega_{ii} = 0$.

A menudo, en muchas aplicaciones empíricas, los coeficientes de modelos espacio-temporales asociados a las covariables se interpretan incorrectamente como si fueran simples derivadas parciales. Esta interpretación no es válida para modelos con dependencia espacial en la variable dependiente o en las variables explicativas (Elhorst, 2014; Golgher and Voss, 2016) debido a que estos modelos pueden generar un proceso de spillover total o global. Este efecto spillover total indica que los cambios de una variable independiente en una determinada unidad espacial, afectarán directamente a la variable dependiente en esa unidad espacial, pero también afectarán indirectamente a los vecinos de sus vecinos desencadenando un efecto feedback que consigue el equilibrio final afectando a la unidad espacial original (Ward and Gleditsch, 2008). Así pues, el efecto spillover total se representa como la suma del efecto directo e indirecto. A continuación, la Tabla 1 muestra los efectos directos e indirectos correspondientes a los diferentes modelos espacio-temporales econométricos (Halleck Vega and Elhorst, 2012).

Tabla 1. Efectos directos e indirectos para los diferentes modelos espacio-temporales econométricos

Modelo	Efecto Directo	Efecto Indirecto
OLS y SEM	β	0
SLM	Elementos diagonal $(I_{it} - \lambda W)^{-1}\beta$	Elementos fuera diagonal $(I_{it} - \lambda W)^{-1}\beta$
SDEM	β	δ
SDM	Elementos diagonal $(I_{it} - \lambda W)^{-1}(\beta + W\delta)$	Elementos fuera diagonal $(I_{it} - \lambda W)^{-1}(\beta + W\delta)$

Nota:

OLS.- Regresión por mínimos cuadrados ordinarios

SEM.- Modelo de error espacial. Sus siglas provienen del inglés "Spatial Error Model"

SLM.- Modelo de retardo espacial. Sus siglas provienen del inglés "Spatial Lag Model"

SDEM.- Modelo de error espacio-temporal Durbin. Sus siglas provienen del inglés "Spatial Durbin Error Model"

SDM.- Modelo espacio-temporal Durbin. Sus siglas provienen del inglés "Spatial Durbin Model"

Para todos los modelos espacio-temporales de la Tabla 1, la matriz de contigüidad o de pesos espaciales se descompone de la siguiente forma:

$$(I_{it} - \lambda W)^{-1} = I_{it} + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 \dots \tag{9}$$

donde I_{it} la matriz identidad de dimensión $N \times T$, W es la matriz de contigüidad o de pesos espaciales. Cuando W se eleva a la potencia de 1 representa los vecinos de primer orden. La matriz W^2 refleja los vecinos de segundo orden, aquellos que son vecinos de los vecinos de primer orden. La matriz W^3 representa los vecinos de tercer orden y así sucesivamente (LeSage and Pace, 2009). De esta forma, la ecuación (9) refleja el efecto de feedback, es decir, los efectos que pasan a través de unidades vecinas vuelven a la unidad espacial original.

2.3. Metodología estadística para comparar y seleccionar modelos espacio-temporales con datos de panel

A continuación, se van definir los tests estadísticos y las medidas de bondad de ajuste que sugiere Elhorst (2014) para comparar y seleccionar los modelos espacio-temporales de datos de panel:

- La prueba clásica del Multiplicador de Lagrange (LM).- Sus singlas provienen del inglés “Lagrange Multiplier” test, el cual permite conocer si es significativo considerar en el modelo de datos de panel el retardo (dependencia) espacial en la variable dependiente o en el error. Para ello, estudia la interacción de los efectos espaciales en los datos de corte transversal. Esta prueba se basa en los valores de la función log-verosimilitud de diferentes modelos. La hipótesis nula es que la variable dependiente o el error no está espacialmente autocorrelacionado, y si se rechaza con un nivel de significación del 5% o 1% implica la inclusión en el modelo del retardo (dependencia) espacial en la variable dependiente o en el error. Este test fue propuesto por Burridge (1980) y Anselin (1988).
- El cociente de verosimilitud (LR).- Sus singlas provienen del inglés “Likelihood-ratio” test, el cual contrasta la hipótesis de si el efecto fijo espacial o temporal son significativos, y por tanto, se deben de considerar en el modelo. Esta prueba, igual que la anterior, se basa en los valores de la función log-verosimilitud de diferentes modelos. La hipótesis nula expresa que el efecto fijo espacial o temporal no es significativo, es decir, no es significativo añadir el efecto considerado al modelo. Shiba and Tsurumi (1988) proporcionaron este test para contrastar la hipótesis nula.
- Test de Wald. Esta prueba ayudará a confirmar si el modelo Durbin es apropiado para el análisis. Las hipótesis nula establece si el modelo SDM o el modelo SDEM puede ser simplificado a un modelo con retardo o error espacial. Si dicha hipótesis se rechaza, entonces el modelo SDM o el modelo SDEM es el que mejor describe los datos (Elhorst, 2014).
- Test de Hausman comprueba si el efecto del modelo ha de tratarse como fijo o aleatorio. Si la hipótesis nula no puede rechazarse, indica que se debe de utilizar el modelo de efectos aleatorios frente al modelo de efectos fijos (Baltagi et al., 2003).
- Coeficiente de determinación R^2 y la varianza residual σ^2 son las dos medidas de bondad de ajuste de los modelos. Cuanto mas próximo está el R^2 a 1 y menor es la σ^2 comparada con la varianza total σ_y^2 , mejor será el modelo.

3. Resultados

3.1. Selección del mejor modelo espacio-temporal con datos de panel en MATLAB y R.

Como la variable SMR tiene una fuerte asimetría positiva, la hemos transformado con el logaritmo y la utilizaremos como variable dependiente en el modelo espacio-temporal.

A continuación y con el objetivo de detectar la multicolinealidad entre las 5 variables explicativas disponibles, se utilizó el Factor de Inflación de la Varianza (VIF) medida es mayor para la variables causantes de la multicolinealidad. La Tabla 2 muestra el VIF para cada una de las variables de las que se seleccionaron aquellas con un VIF menor a 2: PIB, tasa de actividad, consumo de energía en el sector vial y tasa de natalidad.

Como ya mencionamos al final de la Sección 2.1, para seleccionar e implementar el mejor modelo espacio-temporal a los datos de mortalidad europea, hemos adaptado las rutinas que sugiere Elhorst (2011) en MATLAB y, a su vez, hemos implementado un código similar en R. A continuación se

Tabla 2. VIF de covariables

Variable	VIF
Crecimiento de la población	2.0419
Tasa bruta de natalidad	1.5860
Tasa de actividad	1.1738
Consumo de energía vial	1.5619
Producto Interior Bruto	1.0393

irán detallando los pasos a seguir para seleccionar el mejor modelo espacio-temporal con datos de panel junto con los resultados obtenidos utilizando ambos softwares para su comparación.

En primer lugar, se debe estudiar la elección de un modelo de datos de panel (no espacial), frente a una simple regresión por mínimos cuadrados ordinarios. Después, se analizará si el modelo de datos de panel seleccionado se puede convertir en un modelo espacio-temporal. Para ello, es necesario estimar los siguientes modelos con datos de panel no espaciales:

- Regresión por Mínimos Cuadrados Ordinarios (OLS).- Esta regresión no considera la dimensión espacial ni temporal de los datos.
- Modelo de datos en panel con Efecto Fijo Espacial (MSFE).- Su nombre proviene de las siglas en inglés “Model with Spatial Fixed Effects”. El efecto fijo del modelo controla la heterogeneidad no observable producida por variables de las cuales no tenemos información pero que afectan a la variable de interés. Estas variables varían con el tiempo pero permanecen constantes entre los países.
- Modelo de datos en panel con Efecto Fijo Temporal (MTFE).- Su nombre proviene de las siglas en inglés “Model with Time Fixed Effects”. El efecto fijo del modelo controla la heterogeneidad no observable producida por variables de las cuales no tenemos información pero que afectan a la variable de interés. Estas variables varían con el tiempo pero permanecen constantes entre los países.
- Modelo de datos en panel con Efecto Fijo Espacial y Temporal (MSTFE).- Su siglas en inglés provienen de “Model with Spatial and Time Fixed Effects”. Es un modelo de datos de panel con el efecto fijo tanto espacial como el temporal. Este modelo es también conocido como “two ways”, ya que, incluye ambos efectos fijos.

Cabe destacar que, a diferencia del modelo SLMSTFE, el modelo MSTFE no considera la dependencia o retardo espacial en la variable dependiente. En los modelos MSFE y MTFE sólo se considera un efecto y no ambos donde tampoco tienen en cuenta la dependencia espacial en la variable dependiente.

A continuación, se calcula la prueba clásica LM en cada uno de los 4 modelos con el objetivo de analizar si el modelo seleccionado puede convertirse en un modelo espacio-temporal. Para su obtención, se ha utilizado el fichero denominado *demoLMsarsem_panel* en MATLAB y en R la función `plm` del paquete `plm` (Croissant et al., 2008) para los modelos y la función `slmtest` del paquete `splm` (Millo and Piras, 2012) para la prueba clásica LM. Los resultados de estimar los 4 modelos definidos junto con la prueba clásica LM utilizando el software MATLAB y R se encuentran en la Tabla 3.

Las principales conclusiones que se pueden extraer de la Tabla 3 son las siguientes:

- Los valores obtenidos en las estimaciones de los parámetros de los modelos, la varianza residual σ^2 , el coeficiente de determinación R^2 sin la contribución de los efectos fijos y los resultados del test LM son las mismas con ambos softwares.

Tabla 3. Estimación de modelos no espaciales de datos de panel en MATLAB y R.

Parámetros	OLS	MSFE	MTFE	MSTFE
α MATLAB	0.7095*	0.4512*	0.7360*	0.4256*
$\beta_{tasa\ natalidad}$	-0.0291*	0.0016	-0.0332*	0.0010
$\beta_{tasa\ actividad}$	-0.0018	-0.0066*	-0.0016	-0.0058*
$\beta_{consumo\ energia\ vial}$	-0.0146*	-0.0020*	-0.0148*	-0.0033*
β_{PIB}	0.0037	0.0027*	0.0078*	0.0046*
σ^2	0.0402	0.0018	0.0392	0.0017
R^2 sin la contribución de los efectos fijos	0.2998	0.2101	0.3158	0.2575
Test clásico LM retardo espacial	301.4027*	42.6894*	290.4052*	40.1192*
Test clásico LM error espacial	197.2800*	37.6552*	177.1406*	26.9007*
Función Log-likelihood (MATLAB)	100.1503	909.5316	106.2504	928.1107
R^2 con la contribución de los efectos fijos (MATLAB)	-	0.9689	0.3160	0.9710
Suma de cuadrados total (R)	29.582	1.1661	29.571	1.1549
Suma de cuadrados residual (R)	20.713	0.92104	20.233	0.85753
F-estadístico (R)	55.1298*	32.5896*	57.2315*	40.8401*

*p-valores <0.05 son significativos

Al ser una tabla comparativa de todos los parámetros comunes a los distintos modelos, no mostramos la estimación de los efectos fijos de cada modelo.

La estimación de α en los modelos de datos de panel se obtiene utilizando MATLAB. El paquete `plm` en R no lo tiene implementado. Por contra el paquete `splm`, para modelos espacio-temporales, si que lo permite.

- La salida de los modelos en R no muestra directamente el valor de σ^2 . Para obtener la varianza residual del objeto modelo hay que acceder a la parte denominada `sigma2` en el output de la función `splm`.
- En cuanto al R^2 su valor en R se ha obtenido a partir del valor de la varianza residual σ^2 y de la varianza de la variable dependiente σ_y^2 de la siguiente manera:

$$R^2 = 1 - \frac{\sigma^2}{\sigma_y^2} \approx 1 - \frac{S_\epsilon^2}{S_y^2}$$

El software R muestra la suma de cuadrados total, la suma de cuadrados residual y el valor del estadístico F. El software MATLAB proporciona el valor del R^2 incluyendo además los efectos junto con el valor de la función Log-likelihood.

- Se observa que el modelo con datos de panel que incluye el efecto fijo espacial y temporal es el que obtiene mayor R^2 con la contribución de los efectos fijos y menor σ^2 . Esto indica, a priori, que ambos efectos son conjuntamente significativos y es importante incluirlos en el modelo. Así pues, se puede confirmar que se debe de desestimar una regresión OLS y por contra considerar un modelo con datos de panel MSTFE.
- La significatividad y, en algún caso el signo de variables como tasa bruta de natalidad, tasa de actividad y PIB difieren entre los modelos OLS y MSTFE. Esto es debido a que el modelo OLS no considera la dimensión espacial y temporal de los datos, a diferencia del MSTFE que si las considera. El modelo que consigue un mayor R^2 es el MSTFE. En él, todas las variables tienen el signo esperado excepto la variable PIB. Este resultado es acorde con el trabajo de EUROSTAT (2013). En él se muestra que, aunque las condiciones de salud están relacionadas con el PIB, éstas no dependen totalmente de la producción de riqueza en una economía determinada. Las diferencias entre los países también pueden atribuirse a otros factores, por ejemplo si los proveedores son organizaciones privadas o públicas, la calidad de los servicios de salud, los factores ambientales y las opciones culturales también afectan los resultados de la salud. La

única variable que no es significativa en el modelo es tasa bruta de natalidad, por lo que se decide excluirla en los posteriores análisis. Destacar que los coeficientes de todas las covariables son cercanos a cero. Esto se debe a que la variable dependiente toma valores muy pequeños en un rango de mínimo a máximo comparados con los de las variables independientes.

- La prueba clásica LM indica que considerar el retardo espacial de la variable dependiente en el modelo MSTFE es significativo, convirtiéndolo así en un modelo espacio-temporal SLMSTFE. Considerar la correlación espacial del error en el MSTFE es menos significativo, por lo que se selecciona el modelo SLMSTFE con efecto fijo espacial y temporal.
- La conclusión de este primer paso es que se debe de adoptar un modelo espacio-temporal con datos en panel, en concreto, un modelo SLMSTFE con efecto fijo espacial y temporal frente a un modelo no espacial.

El segundo paso es investigar si los efectos espaciales o temporales fijos se deben de incluir en el modelo. Para ello, se utilizó la prueba de Razón de Verosimilitud (LR) en MATLAB utilizando la función `LMsarsem_panel` del fichero denominado *demoLMsarsem_panel*. Dicha prueba compara el modelo MSTFE versus MSFE y MTFE cuya hipótesis nula es que el efecto fijo espacial o temporal no es significativo y por tanto, no se debe de incluir en el modelo. La Tabla 4 muestra los resultados del test LR en MATLAB, la cual justifica el MSTFE obtenido en el paso anterior dado que los p-valores de ambos efectos son significativos.

Tabla 4. Resultado de la prueba LR en MATLAB

Efecto	Test LR	Grados de libertad	p-valor
Espacial	1672.5700	26	0.0000
Temporal	37.7142	20	0.0096

La prueba LR no está implementada en los paquetes `plm` y `splm` de R. En su lugar, se utilizó la función `plmtest` del paquete `plm` (Croissant et al., 2008) en R. Dicha función obtiene el resultado de la prueba LM por Breusch and Pagan (1980). El objetivo de esta prueba es contrastar la incorporación de efecto espacial o efecto temporal o efectos espaciales y temporales en el modelo. Para ello, compara el modelo OLS versus MSFE, MTFE y MSTFE. Si el p-valor es menor que 0.05 la hipótesis nula será rechazada y por tanto será necesario incluir los efectos fijos considerados en el modelo. En dicha función, se debe de especificar el modelo de referencia (OLS), el efecto que se desea añadir al modelo y el tipo del test, en nuestro caso el de Breush-Pagan. Los resultados de la prueba LM se muestran en la Tabla 5, la cual incluye el valor del estadístico utilizado en el contraste, los grados de libertad y el p-valor asociado a cada una de las pruebas. La prueba LM concluye que el efecto espacial y temporal deben de considerarse en el modelo, ya que el p-valor obtenido es el más significativo. Por lo tanto, se confirma el modelo MSTFE obtenido en el paso anterior.

A partir de los resultados de los dos pasos anteriores se confirma que hemos de considerar un modelo SLMSTFE con efecto fijo espacial y temporal.

Si el modelo OLS se ha rechazado en favor de un modelo espacio-temporal en base a la prueba clásica LM, hay que estudiar con detenimiento el modelo espacio-temporal seleccionado (Elhorst, 2014). Por este motivo, LeSage and Pace (2009) recomiendan considerar también el modelo Durbin. En nuestro caso, vamos a considerar un modelo espacio-temporal Durbin (SDM) y no un modelo

Tabla 5. Salida de la Prueba del Multiplicador de Lagrange para el efecto espacial, temporal y ambos en R

Prueba LM	Chi ²	Grados de libertad	p-valor
Espacial	4168.5	1	0.0000*
Temporal	2.3718	1	0.1235
Espacial y Temporal	4170.9	2	0.0000*

de error espacio-temporal Durbin. Como ya se detalló en la Sección 2.2 el modelo SDM incluye la correlación espacial en las covariables de los modelos SLMSTFE y SLMSTRE. Así pues, *el tercer paso es estimar el modelo espacio-temporal SDM.*

Para obtener el modelo SDM en MATLAB se ha utilizado la función `sar_panel_FE` del fichero `demopanelcompare`.

Por contra, hasta la actualidad, el modelo SDM no se puede calcular con el paquete `splm` de Millo and Piras (2012). Por tanto, el mejor modelo espacio-temporal de datos en panel que se ha podido obtener con R ha sido el modelo SLMSTFE mientras que en MATLAB fue un modelo SDM. Destacar que la función `lagsarlm` del paquete `spdep` de Bivand (2012) ajusta un modelo SDM en R pero con datos de corte transversal, por lo que no podemos utilizar dicha función.

El cuarto paso es estudiar si los efectos que se deben de incluir en el modelo han de considerarse fijos o aleatorios. Para ello se calculado el test de Hausman. Recordar que la hipótesis nula de esta prueba, indica que se ha de adoptar un modelo de efectos aleatorios. La Tabla 6 muestra los resultados de aplicar dicho test a los modelos SDM y SLMSTFE. Para el modelo SDM se ha utilizando la función `hausman` del fichero `demopanelcompare` en MATLAB. Para el modelo SLMSTFE se ha utilizado la función `sphtest` del paquete `splm` donde es necesario especificar los dos modelos: el modelo SLMSTFE y el modelo SLMSTRE. Los resultados en MATLAB, indican que se ha de adoptar el SMD con efectos fijos mientras que en R se observa que se ha de adoptar el modelo de efectos fijos SLMSTFE, rechazando en ambos casos el modelo de efectos aleatorios.

Tabla 6. Resultado del test de Hausman en MATLAB y R

Software	χ^2	Grados de libertad	p-valor
MATLAB	490.4511	7	0.0000*
R	53.282	3	0.0000*

*p-valores <0.05 son significativos

Por todos los análisis anteriores, podemos concluir que finalmente el modelo seleccionado en MATLAB es un modelo espacio-temporal Durbin (SDM) con efecto fijo espacial y temporal. Mientras que el modelo seleccionado en R es un modelo con retardo espacial en la variable dependiente y efecto fijo espacial y temporal (SLMSTFE).

3.2. Estimación modelos finales: SDM en MATLAB y SLMSTFE en R.

Una vez seleccionados los modelos finales en ambos softwares, el siguiente paso es la estimación e interpretación de cada uno de ellos.

Tabla 7. Estimación del modelo SDM con efecto fijo espacial y temporal en MATLAB.

Parámetros	Estimación	p-valor
α	0.5275	0.0000*
λ	0.3522	0.0000*
$\beta_{tasa\ actividad}$	-0.0007	0.5424
$\beta_{consumo\ energia\ vial}$	-0.0048	0.0000*
β_{PIB}	0.0042	0.0000*
$\delta_{tasa\ actividad}$	0.0071	0.0000*
$\delta_{consumo\ energia\ vial}$	0.0026	0.0624
δ_{PIB}	-0.0028	0.0107*
Test de Wald retardo espacial	27.4798	0.0000*
σ^2	0.0015	
R^2 con la contribución de los efectos fijos	0.9763	
R^2 sin la contribución de los efectos fijos	0.3180	
LogL	968.4017	

*p-valores <0.05 son significativos

Modelo SDM estimado en MATLAB

La Tabla 7 muestra la estimación de parámetros del modelo SDM, dos medidas de bondad de ajuste (R^2 , σ^2) junto con el test de Wald en MATLAB. En ella, se observa que el parámetro espacial λ es positivo (0.3522) y estadísticamente significativo (p-valor <0.05). Este parámetro indica que, además de las contribuciones que las covariables realizan al logaritmo del SMR de un país, su valor aumenta en 35.22% cuando en promedio el logaritmo de los valores SMR correspondientes al entorno también aumentan. Las medidas de bondad de ajuste del modelo indican que es un buen modelo, ya que, el coeficiente de determinación es elevado (0.9763) y la varianza residual es pequeña (0.0015) comparada con la varianza de la variable dependiente (0.0570). El test de Wald confirma que hemos de considerar el modelo SDM ya que es el que mejor describe a los datos (se rechaza el modelo SLMSTFE en favor del modelo SDM).

Los resultados de la estimación de los efectos fijos espaciales μ_i del modelo SDM se muestran en la Tabla 8, la cual incluye el valor estimado de éstos, el valor del estadístico t y los p-valores asociados a cada uno de los parámetros. El valor de μ_i representa la desviación del país i respecto al intercepto α . En general, las estimaciones de efectos espaciales con un signo negativo corresponden a los países del oeste de Europa los cuales tienen un bajo SMR. Esto significa que las variables omitidas constantes en el tiempo afectan disminuyendo el logaritmo del SMR, en comparación con el promedio α . Por contra, las estimaciones de efectos espaciales con un signo positivo pertenecen a los países del este de Europa los cuales tienen un alto SMR. Esto significa que las variables omitidas constantes en el tiempo afectan aumentando el logaritmo del SMR, en comparación con el promedio α .

Los resultados de la estimación de los efectos fijos temporales ν_t del modelo SDM se muestran en la Tabla 9, la cual incluye el valor estimado de éstos, el valor del estadístico t y los p-valores asociados a cada uno de los parámetros. El valor de ν_t representa la desviación del año t respecto al intercepto α . En general, las estimaciones de los efectos temporales con un signo negativo significa que las variables omitidas constantes entre países afectan disminuyendo el logaritmo del SMR, en comparación con el promedio α . Por contra, las estimaciones de los efectos temporales con

Tabla 8. Estimación de los efectos espaciales del modelo SDM en MATLAB

País	Estimación	t-valor	p-valor
Austria	-0.1365	-1.5691	0.1166
Bélgica	-0.1241	-1.4490	0.1473
Bielorrusia	0.2796	3.2946	0.0009*
Suiza	-0.2331	-2.5504	0.0107*
República Checa	0.1161	1.3494	0.1772
Alemania	-0.1033	-1.1664	0.2434
Dinamarca	0.0167	0.1840	0.8539
Estonia	0.1362	1.5134	0.1301
España	-0.1909	-2.1856	0.0288*
Finlandia	-0.0289	-0.3184	0.7501
Francia	-0.2549	-2.9378	0.0033*
Hungría	0.2314	2.8166	0.0048*
Irlanda	0.0211	0.2291	0.8187
Italia	-0.1786	-2.0496	0.0403 *
Lituania	0.1255	1.4489	0.1473
Luxemburgo	0.0355	0.3969	0.6914
Letonia	0.2821	3.1939	0.0014*
Países Bajos	0.1543	-1.8230	0.0682
Noruega	-0.1039	-1.1421	0.2534
Polonia	0.0753	0.8952	0.3706
Portugal	-0.0095	-0.1073	0.9145
Suecia	-0.1421	-1.5628	0.1180
Eslovenia	-0.0227	-0.2711	0.7862
Eslovaquia	0.1141	1.3658	0.1719
Ucrania	0.3586	4.3979	0.0000*
Reino Unido	-0.109183	-1.1638	0.2444

*p-valores <0.05 son significativos

un signo positivo significa que las variables omitidas constantes entre países afectan aumentando el logaritmo del SMR, en comparación con el promedio α . Destacar que aunque ningún ν_t es significativo, no podemos excluirlos del modelo tal y como indicó la prueba LR en la Tabla 4.

Adicionalmente, estos ν_i se representaron gráficamente en la Figura 1. En la Figura 1, se puede observar que los efectos temporales siguen una tendencia creciente excepto para el período 1994-1997. Esta evolución desfavorable recoge el colapso del sistema soviético. Rusia tuvo la peor esperanza de vida en el año 1994, así como Estonia, Letonia y Lituania. Desde 1994 hasta 1998, la esperanza de vida en las Repúblicas Bálticas y Rusia se hizo más favorable. Este fue un signo de ajuste a las nuevas circunstancias (Vågerö, 2010).

Los resultados recogidos en la Tabla 10 muestran las estimaciones junto con el valor del estadístico t de los efectos directos, indirectos y totales basados en los parámetros estimados del modelo SDM. Destacar que estos coeficientes son cercanos a cero debido a que la variable dependiente se expresa en logaritmos.

Las estimaciones de los efectos directos asociados a las tres variables explicativas tienen los signos esperados. La tasa de actividad toma el valor de -0.0017, consumo de energía vial -0.0047

Tabla 9. Estimación de los efectos temporales del modelo SDM en MATLAB

Año	Estimación	t-valor	p-valor
1990	-0.0010	-0.0124	0.9901
1991	-0.0008	-0.0100	0.9919
1992	0.0025	0.0304	0.9757
1993	0.0037	0.0448	0.9642
1994	-0.0013	-0.0160	0.9872
1995	-0.0071	-0.0836	0.9333
1996	-0.0126	-0.1477	0.8825
1997	-0.0103	-0.1190	0.9052
1998	-0.0030	-0.0355	0.9716
1999	-0.0036	-0.0423	0.9662
2000	-0.0060	-0.0686	0.9452
2001	-0.0011	-0.0130	0.9895
2002	0.0007	0.0087	0.9930
2003	-0.0040	-0.0462	0.9630
2004	-0.0011	-0.0132	0.9894
2005	-0.0020	-0.0232	0.9814
2006	0.0064	0.0715	0.9429
2007	0.0097	0.1070	0.9147
2008	0.0093	0.1057	0.9157
2009	0.0220	0.2598	0.7950

*p-valores <0.05 son significativos

y el PIB 0.0040. Cabe mencionar que estas estimaciones son similares a las estimaciones de β en el modelo SDM (-0.0007, -0.0048 y 0.0042) recogidos en la Tabla 7. La diferencia entre estas estimaciones representan el feedback que surge como resultado de los efectos que pasan por los países vecinos y regresan al propio país (Fischer et al., 2009; Elhorst, 2012). Las estimaciones de los efectos directos en las variables consumo de energía vial y PIB son significativas indicando que variaciones en esas variables explicativas en un determinado país producen variaciones sobre el logaritmo de SMR en ese mismo país.

Respecto a los efectos indirectos, sólo la variable tasa de actividad es significativa y negativa, provocando que el efecto total sea significativo y negativo. Esto significa que el aumento de la tasa de actividad en un determinado país no sólo disminuirá el logaritmo del SMR en el propio país, sino también en el resto de países. En este modelo, el ratio entre el efecto indirecto y el efecto directo depende de los parámetros espaciales λ , δ y de la especificación de la matriz de pesos espaciales W lo que hace dicho modelo sea más atractivo que otros modelos de regresión espacial en un estudio empírico (Elhorst, 2010; Golgher and Voss, 2016).

Por todos los análisis anteriores, concluimos que el modelo finalmente seleccionado en MATLAB es un modelo espacio-temporal Durbin con efecto fijo espacial y temporal. Es el modelo que mejor representa la dependencia espacial de los datos de mortalidad, tal y como indican las medidas de bondad de ajuste R^2 y σ^2 . El parámetro espacial λ es positivo y estadísticamente significativo indicando que, además de las contribuciones que las covariables realizan al logaritmo del SMR de un país su valor se incrementa en 35.22% cuando en promedio el logaritmo de los valores SMR correspondientes al entorno también aumentan. En cuanto al efecto spillover total, sólo la variable tasa de actividad es significativa y negativa. Esto significa que el aumento de la tasa de actividad

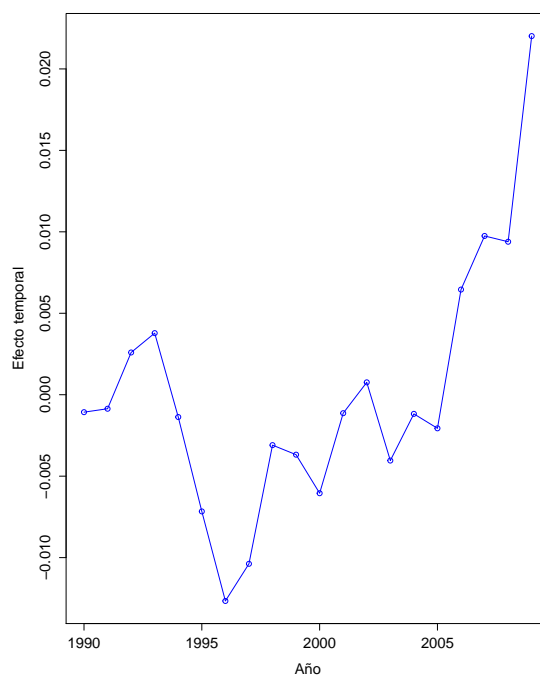


Figura 1. Representación gráfica de los efectos temporales del modelo SDM.

Tabla 10. Estimaciones de efectos directos e indirectos basados en los coeficientes del modelo SDM

Efecto	Variable	Coefficientes	t valores	p-valor
Directo	Tasa de actividad	-0.0017	-1.5308	0.1380
Indirecto	Tasa de actividad	-0.0103	-6.0031	0.0000*
Total	Tasa de actividad	-0.0120	-5.7794	0.0000*
Directo	Consumo de energía vial	-0.0047	-5.2799	0.0000*
Indirecto	Consumo de energía vial	0.0013	0.6507	0.5210
Total	Consumo de energía vial	-0.0034	-1.3795	0.1795
Directo	PIB	0.0040	6.1058	0.0000*
Indirecto	PIB	-0.0018	-1.2676	0.2162
Total	PIB	0.0022	1.3567	0.1866

*p-valores <0.05 son significativos

en el propio país disminuirá el logaritmo del SMR no sólo en su propio país, sino también en el resto de países.

Modelo SLMSTFE estimado en R

La Tabla 11 muestra la estimación de parámetros del modelo SLMSTFE con efecto fijo espacial y temporal, el error estándar, el estadístico t y los p-valores. En ella, se observa que tanto los parámetros (β) asociados a las tres variables como el parámetro espacial λ son significativos (p-valor <0.05). El λ es positivo (0.3465) y estadísticamente significativo indicando que, además de las contribuciones que las covariables realizan al logaritmo del SMR de un país su valor se

incrementa en 34.65% cuando en promedio el logaritmo de los valores SMR correspondientes al entorno también aumentan. Las medidas de bondad de ajuste del modelo son la varianza residual y el coeficiente de determinación. Ambas medidas no se obtienen directamente en el output en R del modelo. La σ^2 tiene un valor aparentemente muy pequeño 0.0014, el cual depende de las unidades de medida. El R^2 se ha obtenido a partir del valor del σ^2 y de la varianza total el cual obtiene un valor de 0.97. Ambas medidas indican que es un buen modelo.

Tabla 11. Output del modelo SLMSTFE en R

Parámetros	Estimación	Error Estándar	t-valor	p-valor
α	0.2896	0.0523	5.5337	0.0000*
λ	0.3465	0.0391	8.8624	0.0000*
β_{PIB}	0.0041	0.0006	6.9201	0.0000*
$\beta_{tasa\ actividad}$	-0.0033	0.0009	-3.7814	0.0002*
$\beta_{consumo\ energia\ vial}$	-0.0041	0.0008	-5.3165	0.0000*

*p-valores <0.05 son significativos

Los resultados de la estimación de los efectos fijos espaciales μ_i del modelo SLMSTFE se muestran en la Tabla 12, la cual incluye el valor estimado de éstos, el error estándar, el valor del estadístico t y su correspondiente p-valor. El valor μ_i representa la desviación del país i respecto al intercepto α . En general, las estimaciones de efectos espaciales con signo positivo representan a los países del este de Europa, los cuales tienen un alto SMR. Esto significa que las variables omitidas constantes en el tiempo afectan positivamente al logaritmo del SMR, en comparación con el promedio α . Por contra, las estimaciones de efectos espaciales con signo negativo corresponden a los países del oeste de Europa, los cuales tienen un bajo SMR. Esto significa que las variables omitidas constantes en el tiempo afectan negativamente al logaritmo del SMR, en comparación con el promedio α .

Los resultados de la estimación de los efectos fijos temporales ν_i del modelo SLMSTFE se muestran en la Tabla 13, la cual incluye el valor estimado de éstos, el error estándar, el valor del estadístico t y los p-valores correspondientes. El valor ν_i representa la desviación del año t respecto al intercepto α . Igual que pasaba en el modelo SDM, aunque ningún ν_t es significativo, no podemos excluirlos del modelo tal y como nos indicó la prueba Multiplicador de Lagrange en la Tabla 5.

Adicionalmente, estos ν_i se representaron gráficamente en la Figura 2. En ella, se puede observar que los efectos temporales siguen una tendencia creciente excepto para el período 1994-1997. Esta evolución negativa coincide con la obtenida en la Figura 1, la cual recoge el colapso de la Unión Soviética (Vågerö, 2010).

Para poder interpretar correctamente los coeficientes del modelo SLMSTFE se han obtenido los efectos directos, indirectos y totales mediante la función `impacts` del paquete `splm` de Millo and Piras (2012). En dicha función se debe de especificar el modelo SLMSTFE y la matriz de pesos espaciales. Los resultados de esta función, se recogen en la Tabla 14, la cual muestra las estimaciones junto con el valor del estadístico t de los tres efectos basados en los parámetros estimados del modelo SLMSTFE.

Las estimaciones de efectos directos de las tres variables explicativas tienen los signos esperados. La tasa de actividad toma el valor de -0.0034, consumo de energía vial -0.0042 y el PIB 0.0043.

Tabla 12. Estimación de los efectos espaciales del modelo SLMSTFE en R

País	Estimación	Error Estándar	t-valor	p-valor
Austria	-0.1368	0.0534	-2.5632	0.0104*
Bélgica	-0.1182	0.0462	-2.5564	0.0106*
Bielorrusia	0.2724	0.0529	5.1495	0.0000*
Suiza	-0.1821	0.0608	-2.9974	0.0027*
República Checa	0.1167	0.0532	2.1946	0.0282*
Alemania	-0.1025	0.0523	-1.9614	0.0498*
Dinamarca	0.0349	0.0592	0.5886	0.5561
Estonia	0.1350	0.0551	2.4527	0.0142*
España	-0.1978	0.0496	-3.9840	0.0000*
Finlandia	-0.0673	0.0545	-1.2360	0.2164
Francia	-0.2256	0.0500	-4.5154	0.0000*
Hungría	0.2044	0.0451	4.5344	0.0000*
Irlanda	-0.0074	0.0541	-0.1371	0.8909
Italia	-0.2123	0.0452	-4.6915	0.0000*
Lituania	0.1161	0.0544	2.1342	0.0328*
Luxemburgo	0.0224	0.0564	0.3972	0.6912
Letonia	0.2544	0.0545	4.6665	0.0000*
Países Bajos	-0.1217	0.0550	-2.2123	0.0269*
Noruega	-0.1231	0.0575	-2.1433	0.0321*
Polonia	0.0522	0.0511	1.0211	0.3072
Portugal	0.0381	0.0554	0.6889	0.4909
Suecia	-0.1727	0.0563	-3.0659	0.0022*
Eslovenia	0.0213	0.0528	0.4042	0.6860
Eslovaquia	0.1271	0.0541	2.3502	0.0188*
Ucrania	0.3612	0.0513	7.0437	0.0000*
Reino Unido	-0.0887	0.0557	-1.5932	0.1111

*p-valores <0.05 son significativos

*p-valores <0.05 son significativos

Igual que pasaba en el modelo SDM el valor de estos coeficientes son similares a los β estimados en el modelo SDM recogidos en la Tabla 11 (-0.0033, -0.0041 y 0.0041). La diferencia entre estas estimaciones representa los efectos de feedback que surgen como resultado de los efectos que pasan por los países vecinos y regresan al propio país (Fischer et al., 2009; Elhorst, 2012). Las estimaciones de los efectos directos de las tres variables son significativas indicando que variaciones en las tres variables explicativas en un determinado país producen variaciones sobre la variable dependiente en ese mismo país. Recordar que en el modelo SDM sólo la estimación del efecto directo de la variable tasa de actividad era significativa.

Respecto a las estimaciones de efectos indirectos de las tres variables explicativas todas son significativas. El signo de estos efectos en las variables tasa de actividad y consumo de energía vial son negativos. Esto significa que aumentos de estas variables en un país disminuirán el el logaritmo del SMR no sólo en su propio país, sino también en el resto de países. Por contra, un aumento del PIB tiene un efecto positivo y significativo sobre el logaritmo del SMR en el propio país (efecto directo 0.0043) y también lo tiene en en el resto de países (efecto indirecto 0.0020). En este modelo, el ratio entre el efecto indirecto y el efecto directo tienen valores similares para las tres covariables tasa de actividad 0.4705, consumo de energía vial 0.4762 y PIB 0.4651. Esto

Tabla 13. Estimación de los efectos temporales del modelo SLMSTFE en R

Año	Estimación	Error Estándar	t-valor	p-valor
1990	-0.0074	0.0532	-0.1392	0.8893
1991	-0.0042	0.0528	-0.0800	0.9363
1992	0.0030	0.0525	0.0565	0.9549
1993	0.0081	0.0519	0.1551	0.8767
1994	-0.0010	0.0522	-0.0190	0.9849
1995	-0.0081	0.0522	-0.1541	0.8775
1996	-0.0134	0.0522	-0.2567	0.7974
1997	-0.0153	0.0528	-0.2900	0.7718
1998	-0.0060	0.0527	-0.1135	0.9096
1999	-0.0038	0.0526	-0.0720	0.9426
2000	-0.0089	0.0532	-0.1673	0.8672
2001	0.0003	0.0526	0.0054	0.9957
2002	0.0033	0.0527	0.0618	0.9507
2003	-0.0025	0.0529	-0.0464	0.9630
2004	-0.0032	0.0534	-0.0591	0.9529
2005	-0.0036	0.0535	-0.0676	0.9461
2006	0.0010	0.0541	0.0178	0.9858
2007	0.0042	0.0544	0.0770	0.9386
2008	0.0130	0.0535	0.2427	0.8083
2009	0.0446	0.0518	0.8615	0.3889

*p-valores <0.05 son significativos

es debido a que en el modelo SLMSTFE dicho ratio depende del parámetro espacial λ y de la especificación de la matriz de pesos espaciales W . Esta limitación, no se traslada al modelo SDM, la cual es una de las características interesantes de este modelo debido a que dicho ratio depende además del parámetro espacial δ . En consecuencia, el modelo SLMSTFE es demasiado rígido para modelar adecuadamente los efectos indirectos (Elhorst, 2014; Golgher and Voss, 2016) de forma en que las estimaciones de los efectos indirectos asociados a todas las variables son significativos a diferencia del modelo SDM dónde sólo el asociado a la variable tasa de actividad es significativo. Esto no ocurre en el modelo SDM dónde sólo el efecto total de la variable tasa de actividad es significativo.

Por todos los análisis anteriores, concluimos que el modelo finalmente seleccionado en R es un modelo de efecto fijo espacial y temporal con retardo espacial. Es el modelo que mejor representa la dependencia espacial de los datos, tal y como indican las medidas de bondad de ajuste R^2 y σ^2 . El parámetro espacial λ es positivo y estadísticamente significativo indicando que, además de las contribuciones que las covariables realizan al logaritmo del SMR de un país su valor se incrementa en 34.65% cuando en promedio el logaritmo de los valores SMR correspondientes al entorno también aumentan. En cuanto al efecto spillover total, en las tres covariables es significativo. Esto significa que variaciones de estas covariables en un determinado país variará el logaritmo de SMR no sólo en el propio país, sino también en el resto de países.

4. Conclusiones

En este trabajo, se presenta una metodología estadística propuesta por Elhorst (2014) para seleccionar el mejor modelo espacio-temporal con datos de panel utilizando el software estadístico

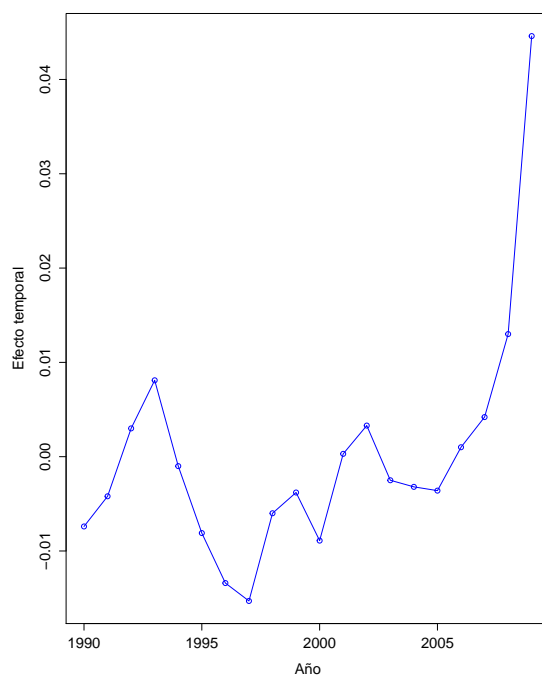


Figura 2. Representación gráfica de los efectos temporales del modelo SLMSTFE.

Tabla 14. Estimaciones de efectos directos e indirectos basados en los coeficientes del modelo SLMSTFE

Efecto	Variable	Coefficientes	z-valor	p-valor
Directo	Tasa de actividad	-0.0034	-3.8633	0.0001*
Indirecto	Tasa de actividad	-0.0016	-3.3874	0.0007*
Total	Tasa de actividad	-0.0050	-3.8430	0.0001*
Directo	Consumo de energía vial	-0.0042	-5.9235	0.0000*
Indirecto	Consumo de energía vial	-0.0020	-4.2171	0.0000*
Total	Consumo de energía vial	-0.0062	-3.8430	0.0000*
Directo	PIB	0.0043	6.7740	0.0000*
Indirecto	PIB	0.0020	4.2896	0.0000*
Total	PIB	0.0063	6.1523	0.0000*

*p-valores <0.05 son significativos

MATLAB. Posteriormente, esta metodología se implementa en R para su comparación con la anterior. A partir de ambos procedimientos puede deducirse que el modelo espacio-temporal seleccionado incluye la correlación debida a la localización geográfica, debido a que tiene en cuenta la vecindad entre los países a lo largo del tiempo. Estos modelos aprovechan la estructura del panel de datos para plantear el análisis de regresión, permitiendo a su vez, otras opciones de modelado de las que las series temporales y los estudios de corte transversal carecen (Hsiao, 2014; Elhorst, 2014).

La aplicación llevada a cabo en este estudio parte de datos de mortalidad correspondientes a 26 países europeos durante el período 1990-2009. Para cuantificar la mortalidad en Europa se

ha propuesto el ratio de mortalidad estandarizado (SMR). Este ratio es ampliamente utilizado en áreas como la epidemiología, demografía y la econometría. Se modeliza su logaritmo para corregir la asimetría y aproximar su distribución a la Normal. Así pues, el logaritmo del SMR es la variable dependiente en el modelo espacio-temporal. Antes de ajustar el modelo, se realizó un estudio de multicolinealidad, el cual reveló que las variables independientes que se debían considerar en el modelo eran la tasa bruta de natalidad, la tasa de actividad, el consumo de energía del sector vial y el PIB pues la causante de multicolinealidad era la variable crecimiento de la población.

En cuanto a las principales conclusiones del ajuste y selección del modelo espacio-temporal con datos en panel son las siguientes:

Las dos medidas de bondad de ajuste R^2 y σ^2 confirmaron que se debía de desestimar una regresión OLS y por contra, estimar un modelo de datos de panel con efecto fijo espacial y temporal. Las covariables que fueron significativas para explicar el logaritmo del SMR fueron la tasa de actividad, consumo de energía del sector vial y el PIB. Una vez confirmado el modelo de datos de panel se estudió su conversión a un modelo espacio-temporal. La prueba clásica LM concluyó que era significativo considerar la dependencia espacial de la variable dependiente en el modelo de datos de panel con efecto fijo espacial y temporal, convirtiéndolo así en un modelo espacio-temporal. Estos resultados fueron los mismos utilizando ambos softwares.

La prueba LR obtenida en MATLAB concluyó que era significativo considerar el efecto espacial y temporal en el modelo de datos de panel. Dicha prueba justificó la estimación de un modelo de datos de panel con efecto fijo espacial y temporal a los datos. La prueba LR no está implementada en R, por lo que en su lugar se utilizó la prueba del Multiplicador de Lagrange de Breusch and Pagan (1980). El resultado de la prueba confirmó nuevamente que el efecto espacial y temporal no podían ser excluidos del modelo.

A partir de los resultados de los dos pasos anteriores se confirmó que se debía de seleccionar un modelo de datos de panel con efecto fijo espacial y temporal con retardo espacial en la variable dependiente (SLMSTFE).

A continuación, se consideró el modelo espacio-temporal Durbin (SDM) para analizar si es significativo incluir la dependencia espacial en las covariables. El modelo SDM sólo se estimó en MATLAB utilizando las rutinas de Elhorst (2014) pues dicho modelo no se puede obtener en R.

El siguiente paso fue estudiar si los efectos que se deben de incluir en el modelo debían de considerarse como fijos o aleatorios. El test de Hausman aplicado al modelo SDM en MATLAB y al modelo SLMSTFE en R indicó que los efectos espaciales y temporales debían de considerarse fijos y no aleatorios.

Por todos los análisis anteriores, el mejor modelo espacio-temporal con datos en panel que se ha podido obtener con el software R fue un modelo SLMSTFE mientras que en MATLAB fue un modelo SDM.

Los modelos seleccionados se validaron utilizando las medidas de bondad de ajuste R^2 y σ^2 . Estas medidas, indicaron que los modelos SLMSTFE y SDM son buenos, ya que, el coeficiente de determinación es cercano a 1 y la varianza residual es pequeña, comparada con la varianza de la variable dependiente, en ambos modelos.

En relación con el trabajo de otros autores, hay que destacar que hasta donde sabemos, aunque se ha estudiado las diferencias en mortalidad, ningún estudio espacial de modelización del ratio de mortalidad estandarizada en Europa mediante modelos datos de panel se ha llevado a cabo. Además, tampoco se ha establecido una comparativa de los softwares MATLAB y R para

seleccionar el mejor modelo espacio-temporal con datos de panel. Esto es importante debido a la creciente popularidad del software libre R en la investigación en relación con softwares comerciales.

Por otra parte, en la mayoría de los países del centro y oeste de Europa, la esperanza de vida está aumentando sobre todo gracias a un gran crecimiento económico y descenso de las enfermedades cardiovasculares. Muchos estudios como Meslé (2004) y Vågerö (2010) abordan la conocida esperanza de vida, pero ninguno de ellos aplica una metodología espacial para modelizar la dependencia espacial en los países a través del tiempo.

Agradecimientos

Este trabajo fue financiado por la subvención del Ministerio de Economía y Competitividad, proyecto MTM2013-45381-P.

Referencias

1. Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer, Dordrecht.
2. Baltagi, B. (2008). *Econometric Analysis of Panel Data*. John Wiley & Sons.
3. Baltagi, B. H., Bresson, G., and Pirotte, A. (2003). Fixed effects, random effects or Hausman–Taylor? A pretest estimator. *Economics letters*, 79(3): 361–369.
4. Bivand, R. (2012). *spdep: Spatial Dependence: Weighting Schemes, Statistics and Models*. R package version 0.5-53.
5. Breusch, T. S. and Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies*, 47(1): 239–253.
6. Burridge, P. (1980). On the Cliff-Ord test for spatial autocorrelation among regression residuals. *Geographical Analysis*, 4: 267–284.
7. Croissant, Y., Millo, G., et al. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27(2): 1–43.
8. Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International regional science review*, 26(3): 244–268.
9. Elhorst, J. P. (2010). Applied spatial econometrics: raising the bar. *Spatial Economic Analysis*, 5(1): 9–28.
10. Elhorst, J. P. (2011). Matlab software to estimate spatial panels. Version 2011-04-11, <http://www.regrooningen.nl/elhorst/software.shtml>, rutinas descargadas el 06/05/2015.
11. Elhorst, J. P. (2012). Dynamic spatial panels: models, methods, and inferences. *Journal of geographical systems*, 14(1): 5–28.
12. Elhorst, J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels*. Springer.
13. EUROSTAT (2009). Health statistics-atlas on mortality in the European Union. *European Communities*. Luxembourg.
14. EUROSTAT (2013). Quality of life indicators. *Eurostat*. http://ec.europa.eu/eurostat/statistics-explained/index.php/Quality_of_life_indicators_-_health.
15. Fischer, M. M., Bartkowska, M., Riedl, A., Sardadvar, S., and Kunnert, A. (2009). The impact of human capital on regional labor productivity in Europe. *Letters in Spatial and Resource Sciences*, 2(2-3): 97.
16. Fleiss, J., Levin, B., and Paik, M. (2013). *Statistical Methods for Rates and Proportions*. Wiley Series in Probability and Statistics. Wiley.
17. Golgher, A. B. and Voss, P. R. (2016). How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects. *Spatial Demography*, 4(3): 175–205.
18. Halleck Vega, S. and Elhorst, J. (2012). On spatial econometric models, spillover effects, and W. *University of Groningen*. Working paper.

19. Hinde, A. (1998). *Demographic methods*. Routledge.
20. Hsiao, C. (2014). *Analysis of Panel Data (3rd edition)*. Cambridge University Press, Cambridge.
21. Human Mortality Database (2014). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). <http://www.mortality.org> or www.humanmortality.de, datos descargados el 17/04/2014.
22. Julious, S., Nicholl, J., and George, S. (2001). Why do we continue to use standardized mortality ratios for small area comparisons? *Journal of Public Health*, 23(1): 40–46.
23. Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1): 97–130. TSP and Stata software available at http://econweb.umd.edu/prucha/Research_Prog3.htm.
24. Kennedy, P. (2003). *A guide to econometrics*. MIT press.
25. LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*, chapter Model Comparison. Statistics: A Series of Textbooks and Monographs. CRC Press Taylor & Francis Group.
26. MathWorks, T. (2015). *MATLAB - The Language of Technical Computing, Version 8.5.0.197613*. Natick, Massachusetts. <http://www.mathworks.com/products/matlab/>.
27. Meslé, F. (2004). Mortality in Central and Eastern Europe: Long-term trends and recent upturns. *Demographic Research*, S2: 45–70.
28. Millo, G. and Piras, G. (2012). splm: Spatial panel data models in R. *Journal of Statistical Software*, 47(1) :1–38.
29. R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
30. Rezaeian, M., Dunn, G., St Leger, S., and Appleby, L. (2007). Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *Journal of epidemiology and community health*, 61(2): 98–102.
31. Shiba, T. and Tsurumi, H. (1988). Bayesian and non-bayesian tests of independence in seemingly unrelated regressions. *International Economic Review*, (29): 377–395.
32. The World Bank Database (2015). *World Development Indicators*. <http://data.worldbank.org/>, datos descargados el 9/01/2015.
33. Torgo, L. (2010). *Data Mining with R, learning with case studies*. Chapman and Hall/CRC.
34. Vågerö, D. (2010). The East–West health divide in Europe: Growing and shifting eastwards. *European Review*, 18(1): 23–34.
35. Ward, M. and Gleditsch, K. (2008). *Spatial Regression Models*. SAGE Publications, Thousand Oaks, CA.
36. Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.