

MESTRADO
ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO
DISSERTAÇÃO

**MIGRAÇÃO INTERNA EM PORTUGAL EM TEMPOS DE CRISE – UMA
ANÁLISE DE ECONOMETRIA ESPACIAL**

KEVIN FRANK ROQUE PAIS

NOVEMBRO – 2020

MESTRADO EM ECONOMETRIA APLICADA E PREVISÃO

TRABALHO FINAL DE MESTRADO DISSERTAÇÃO

**MIGRAÇÃO INTERNA EM PORTUGAL EM TEMPOS DE CRISE – UMA
ANÁLISE DE ECONOMETRIA ESPACIAL**

KEVIN FRANK ROQUE PAIS

ORIENTAÇÃO:

PROFESSORA DOUTORA ISABEL PROENÇA

NOVEMBRO – 2020

*A quem acreditou em mim nos momentos
em que nem eu próprio acreditei.*

ABSTRACT

The 2008 economic crisis had a great impact in the Portuguese economy in several different ways. The high unemployment rate and the low GDP are some examples of how this crisis impacted Portugal. The purpose of this study is to analyze the impact in terms of its migration at the municipal level between 2008 and 2015. As it is also the aim to determine which factors influence it. Since we are working with spatial data, it is necessary to apply spatial econometrics techniques in order to model it.

The dependent variable in this study represents the population growth in each municipality during the referenced period. The population growth is controlled so that the only alterations can come from migration. Based on previous studies from different authors, there were selected a group of several variables considered relevant to explain the population growth.

With the use of appropriate software to work the data, it is concluded that tools to capture spatial dependence are required. Hence, it is necessary to estimate models with the capability to do so. The models SLX, Spatial Lag Model, Spatial Error Model, SARAR, Spatial Durbin Model and Spatial Durbin Error Model are estimated, concluding that the SLX and the Spatial Lag Model models are the ones who better capture the spatial dependence. Both models deliver similar results. The unemployment rate, the average age of the population, the municipality being or not the capital of a *distrito*, the number of non-financial companies and the percentage of working population that works in the primary and secondary sectors of the region decrease the population growth due to migration in that region while the trade balance of a particular region increases it. Additionally, in the SLX model, the population growth due to migration of the location increases also with the average monthly wage of the neighboring regions and decreases with the percentage of the population with a college degree and the percentage of working population that works in the primary and secondary sectors of the neighboring locations. Furthermore, in the Spatial Lag Model, the population growth due to migration of the neighboring regions also determines the population growth due to migration of a specific region.

KEYWORDS: spatial econometrics, spatial dependence, internal migration in Portugal, 2008 economic crisis

RESUMO

A crise económica de 2008 afetou gravemente a economia portuguesa de diversas formas, sendo o aumento da taxa de desemprego e a diminuição do PIB alguns exemplos disso. Este trabalho tem como objetivo analisar o impacto desta crise na migração ao nível municipal, entre os anos de 2008 e 2015, assim como determinar quais os fatores que influenciam a dita migração. Para esta análise utilizam-se dados espaciais, pelo que se recorre ao uso de ferramentas de econometria espacial para os trabalhar.

A variável dependente em estudo representa o crescimento populacional, no período em análise, de cada município. Este crescimento encontra-se controlado de modo a que as únicas alterações populacionais provenham da migração. Selecionaram-se, com base em estudos anteriormente realizados no mesmo âmbito, um grupo de variáveis consideradas relevantes para explicar o crescimento populacional de cada município.

Com recurso a software apropriado, trabalharam-se os dados mostrando ser justificável a utilização de mecanismos que captem a dependência espacial. Assim, estimaram-se modelos próprios para o efeito: SLX, Spatial Lag Model, Spatial Error Model, SARAR e modelos espaciais de Durbin clássico e com erros espacialmente desfasados. Conclui-se que os modelos que melhor captam a dependência espacial são os modelos SLX e Spatial Lag Model, sendo que ambos produzem resultados semelhantes. Em ambos se verifica que a taxa de desemprego, a média de idades da população, o facto do município ser ou não capital de distrito, o número de empresas não financeiras e a percentagem da população ativa empregada nos setores primário e secundário da região, influenciam negativamente o crescimento populacional devido a migração da mesma, enquanto que a balança comercial da região o influencia positivamente. No caso do SLX, verifica-se ainda que o salário mensal médio das regiões vizinhas influencia positivamente o crescimento populacional devido a migração da região, sendo este influenciado negativamente pela percentagem da população com o ensino superior concluído e a percentagem da população ativa empregada nos setores primário e secundário das regiões vizinhas. Adicionalmente, no Spatial Lag Model, o crescimento da população devido a migração de uma dada localização é determinado pelo crescimento populacional devido a migração das regiões vizinhas.

PALAVRAS-CHAVE: econometria espacial, dependência espacial, migração interna em Portugal, crise económica de 2008

AGRADECIMENTOS

Gostaria de agradecer em primeiro lugar à Professora Isabel Proença por todas as palavras de apoio, por toda a paciência e pela sua incrível orientação.

Aos docentes e colaboradores do Instituto Superior de Economia e Gestão, que fizeram parte do meu percurso tanto na licenciatura em Economia como, especialmente, no Mestrado de Econometria Aplicada e Previsão.

Aos meus pais e à minha irmã pelo apoio incondicional e pela motivação que me deram para fazer mais e melhor em todos os aspetos da minha vida. A todos os membros da minha família por me terem passado um pouco de si mesmos contribuindo para a pessoa que sou hoje.

Aos meus amigos e colegas, em especial, à Carolina, à Clara, ao Fábio, ao Leonardo e ao Ludgero, à Margarida e ao Rafael pelo apoio, pela paciência e pela ajuda sempre que precisei.

E por último, a todos os docentes que me acompanharam ao longo do meu percurso académico pela transmissão dos seus conhecimentos e por me despertarem curiosidade em querer procurar saber sempre mais.

TABELA DE CONTEÚDOS

Abstract.....	i
Resumo.....	ii
Agradecimentos.....	iii
Tabela de conteúdos.....	iv
Índice de Figuras.....	vi
Índice de Tabelas.....	vi
Índice de Anexos.....	vi
Glossário.....	viii
1.Introdução.....	1
2.Revisão da Literatura.....	2
3.Metodologia.....	4
3.1.Abordagem Espacial.....	4
3.2.Matriz W e Desfasamento Espacial.....	4
3.2.1.Matriz W.....	4
3.2.2.Desfasamento Espacial.....	5
3.3.Modelos Econométricos Espaciais.....	5
3.3.1.Modelo de Regressão Espacial Clássico com Regressores Espacialmente Desfasados.....	6
3.3.2.Spatial Lag Model.....	6
3.3.3.Spatial Error Model.....	7
3.3.4.Modelo SARAR.....	8
3.3.5.Modelo de Regressão Espacial Clássico de Durbin.....	8
3.3.6.Modelo de Regressão Espacial de Durbin com Erro Espacialmente Desfasado.....	9

3.4.Efeitos Parciais.....	9
3.5.Testes de Dependência Espacial.....	11
3.5.1.Teste I de Moran.....	11
3.5.2.Testes de Multiplicador de Lagrange.....	12
4.Análise Exploratória dos Dados.....	13
4.1.Caraterização dos Dados.....	13
4.2.Caraterização das Variáveis.....	13
4.3.Análise Exploratória dos dados Espaciais.....	16
4.4.Dependência Espacial.....	18
5.Modelação e Resultados.....	19
6.Conclusões e Desenvolvimentos Futuros.....	27
Referências Bibliográficas.....	30
Anexos.....	32

ÍNDICE DE FIGURAS

Figura 1: Dispersão geográfica do crescimento populacional entre 2008 e 2015.....	17
Figura 2: Análise de clusters k-means das variáveis cres, desemp, setor_prim, setor_sec e setor_ter.....	18

ÍNDICE DE TABELAS

Tabela I: Descrição das variáveis utilizadas.....	14
Tabela II: Teste de Moran à variável dependente.....	19
Tabela III: Modelo de Regressão Clássico e SLX não restrito para o crescimento populacional.....	20
Tabela IV: Testes I de Moran aos resíduos dos modelos (1) e (2).....	21
Tabela V: Teste BP ao modelo (2).....	21
Tabela VI: Testes LM ao modelo SLX restrito.....	21
Tabela VII: Modelos SLX restrito, SDEM e SDM para o crescimento populacional com erros padrão robustos.....	22
Tabela VIII: Testes LM ao Modelo de Regressão Clássico.....	23
Tabela IX: Testes de Wald às variáveis omitidas nos modelos SLX restrito e SLM.....	24
Tabela X: Modelos SLM, SEM e SARAR para o crescimento populacional com erros padrão robustos.....	26
Tabela XI: Efeitos parciais do modelo SLX restrito.....	27
Tabela XII: Efeitos parciais do modelo SLM.....	27

ÍNDICE DE ANEXOS

Anexo 1: Variáveis não utilizadas.....	32
Anexo 2: Matriz de correlação das variáveis utilizadas.....	32
Anexo 3: Estatísticas descritivas das variáveis utilizadas.....	33
Anexo 4: Gráfico do Teste I de Moran à variável cres.....	33

Anexo 5: SLM com a variável dependente espacialmente desfasada como único regressor.....	33
Anexo 6: Teste I de Moran aos resíduos do modelo SLX restrito.....	34
Anexo 7: SEM não restrito para o crescimento populacional com erros padrão robustos.....	34
Anexo 8: SLM não restrito para o crescimento populacional com erros padrão robustos.....	35
Anexo 9: SARAR não restrito do crescimento populacional com erros padrão robustos.....	35
Anexo 10: Teste de Wald à significância conjunta dos coeficientes de rho e lambda do modelo SARAR.....	36
Anexo 11: Regressão Geograficamente Ponderada do crescimento populacional devido a migração com as variáveis do modelo (6).....	36

GLOSSÁRIO

2SLS – Mínimos Quadrados em dois Passos – *Two-Stage Least Squares*

FGLS – Mínimos Quadrados Generalizável Praticável – *Feasible General Least Squares*

GMM – Método dos Momentos Genaralizado – *Generalized Method of Moments*

GS2SLS – versão espacial do 2SLS – *Generalized Spatial Two-Stage Least Squares*

GWR – Regressão Geograficamente Ponderada – *Geographically Weighted Regression*

i.i.d – independente e identicamente distribuído

INE – Instituto Nacional de Estatística

LM – Multiplicador de Lagrange

OLS – Mínimos Quadrados Ordinários – *Ordinary Least Squares*

PIB – Produto Interno Bruto

SDEM – Modelo de regressão espacial de Durbin com erro espacialmente desfasado

SDM – Modelo de regressão espacial clássico de Durbin – *Spatial Durbin Model*

SEM – Modelo de regressão especial clássico com erro espacialmente desfasado - *Spatial Error Model*

SLM – Modelo de regressão especial clássico com variável dependente espacialmente desfasada - *Spatial Lag Model*

SLX – Modelo de regressão espacial clássico com regressores espacialmente desfasados

TFM – Trabalho Final de Mestrado

1. INTRODUÇÃO

Entre os anos de 2008 a 2012, como retratado por Maria (2016), viveu-se em Portugal um período difícil na economia. Houve um aumento drástico do desemprego, uma diminuição abrupta do PIB e, em 2011, Portugal chegou mesmo a recorrer a ajuda financeira internacional junto do Fundo Monetário Internacional.

Assim, optou-se por, neste Trabalho Final de Mestrado, analisar o efeito que esta crise económica teve na migração em Portugal ao nível municipal, assim como quais os fatores que influenciam a dita migração. A análise realizada reporta-se ao período entre 31 de dezembro de 2008 e 31 de dezembro de 2015, onde a escolha deste último ano se deve ao facto de não existir uma data específica para o final da crise assim como ao facto de os seus efeitos se terem prolongado no tempo tornando, deste modo, o estudo mais abrangente.

De modo a poder-se realizar este estudo, o crescimento populacional tem de ser estabilizado de forma a que a variação da população do município seja expurgada do aumento/diminuição devido ao número de nascimentos ou mortes. É ainda de referir que este estudo será realizado recorrendo a ferramentas de econometria espacial, assim como vários autores, por exemplo Bigotte *et. al* (2014), as utilizam quando estudam a migração interna de um país, uma vez que a dependência espacial tem vindo a ganhar um enorme foco quando se analisa este tipo de variáveis. É também necessário assinalar que o uso desta referida metodologia é justificável quando se tratam dados observados no espaço.

Este Trabalho Final de Mestrado divide-se em seis partes. A presente parte (primeira) introduz o tema em estudo. A segunda parte refere obras e estudos de outros autores no âmbito da migração interna de um determinado país, com especial foco em estudos que aplicam uma abordagem econométrica espacial aos seus dados. A terceira parte apresenta a metodologia que será utilizada na análise destes dados, uma breve introdução à econometria espacial, assim como os modelos a utilizar na comparação de dependência espacial. A quarta parte apresenta, descreve e analisa as variáveis a utilizar com especial ênfase na análise espacial da variável dependente em estudo. A quinta parte apresenta a modelação dos dados assim como a análise dos resultados obtidos. Na sexta e última parte, apresentam-se tanto as conclusões a retirar da análise realizada como também as limitações encontradas aquando feito este estudo e propostas para análises futuras no âmbito deste tema.

2. REVISÃO DA LITERATURA

Nesta secção serão referenciados diversos estudos empíricos realizados no âmbito da migração dentro do próprio país, os quais aplicam maioritariamente uma abordagem de econometria espacial.

Mulhern & Watson (2009), no seu estudo realizado acerca da migração interna entre províncias de Espanha, utilizando dados de 1999 a 2006 e através da estimação de modelos SEM e SLM, concluíram que a diferença entre os salários, a diferença entre a taxa de desemprego, a distância entre as províncias e a diferença entre o preço das habitações entre províncias, eram fatores determinantes para explicar a deslocação de pessoas dentro do país. Contudo, é de referenciar que a variável dependente no referido estudo é uma variável de fluxo entre províncias, ou seja, tem em consideração de onde e para onde os indivíduos se deslocaram, o que não será o caso no presente trabalho. A ideia de que o emprego é um fator determinante para a explicação da movimentação de pessoas dentro do próprio país é reforçada nos estudos de Chun & Griffith (2011), Etzo (2011) e Ubarevičiene *et. al* (2017).

Etzo (2011), realizou uma análise do papel das determinantes macroeconómicas na migração inter-regional em Itália no período de 1996 a 2002, utilizando dados de painel. Para esta análise, o autor estimou um modelo gravitacional estendido através do *Fixed Effects Regression with Vector Decomposition* (FEVD) e ainda um modelo dinâmico para verificar a presença de efeitos de rede recorrendo à estimação FEVD e à estimação GMM. Conclui no seu estudo que variáveis como a dimensão da população da região, a distância entre as cidades principais de duas regiões, o PIB per capita das regiões, a taxa de desemprego das regiões, como referido anteriormente, o número de crimes violentos por cada 10000 habitantes e um índice de infraestruturas (infraestruturas estas que se referem a redes de estradas, aeroportos, redes de comunicação, bancos, centros de saúde, centros educacionais e culturais, caminhos de ferro, centrais elétricas, entre outras) são relevantes na explicação da migração inter-regional.

Para estudar a variação da população na Lituânia, Ubarevičiene *et. al* (2017), utilizando dados dos censos de 2001 e 2011, estimaram cinco modelos lineares. Começam com um primeiro modelo que continha apenas características geográficas ao qual foram adicionando progressivamente novas variáveis até ao modelo final que contém para além das características geográficas, características sociodemográficas e socioeconómicas.

Assim, de acordo com o modelo final concluem que a variação da população se pode explicar através das seguintes variáveis: 3 maiores centros urbanos, áreas a 15km desses 3 centros, percentagem de pessoas com ensino superior concluído em 2001, diferença na percentagem de pessoas com ensino superior concluído entre 2001 e 2011, percentagem da população empregada em 2001, percentagem da população empregada no setor dos serviços em 2001, diferença da percentagem da população empregada no setor dos serviços entre 2001 e 2011 e percentagem de *high-ranking occupations* em 2001. De notar ainda que quando os autores estimam um modelo com as características geográficas e adicionam a elas uma variável com a percentagem da população em idade ativa em 2001, esta é estatisticamente significativa, contudo perde significância quando adicionadas novas variáveis ao modelo.

Já referidos anteriormente, Chun & Griffith (2011) utilizaram a técnica do *eigenvector spatial filtering* de modo a modelar a autocorrelação em rede existente nos fluxos de migração interestatal nos Estados Unidos da América no período de 1994 a 2006. Para procederem a esta análise linear, estimaram quatro tipos de modelos diferentes: modelo linear misto com efeitos aleatórios não estruturados, modelo linear misto com termo temporal autorregressivo de ordem 1, modelo linear misto generalizado de Poisson com efeitos aleatórios não estruturados e modelo linear misto generalizado de Poisson com termo temporal autorregressivo de ordem 1. Nestes modelos estimados pelos autores, verifica-se que a temperatura média de janeiro, o rendimento per capita, a população total, a distância entre estados e a taxa de desemprego são fatores determinantes na explicação da migração interestatal.

Na análise da relação entre as hierarquias urbanas e as dinâmicas da população em Portugal, Bigotte *et. al* (2014) utilizam como variáveis explicativas para o crescimento da população ao nível municipal, 5 níveis hierárquicos. Esses níveis são os seguintes: nível 1 município, nível 2 cidade principal de grupo de municípios, nível 3 capital de distrito, nível 4 cidade principal da região e nível 5 área metropolitana. Estimam por fim uma regressão espacial utilizando como fatores explicativos para a diferença de população entre 1991 e 2011 apenas os centros urbanos de nível 2 e 4, uma vez que o nível 3 não se revela estatisticamente significativo devido ao facto do seu efeito ser absorvido pelo nível 2 e o nível 5 não ser utilizado dado que para este nível só estão disponíveis duas observações. Concluem, ainda, que os valores mais altos para as taxas de crescimento demográficos se verificam nos municípios em centros urbanos de

hierarquia alta assim como nos municípios vizinhos e que esta taxa de crescimento é baixa em municípios perto da fronteira espanhola.

3.METODOLOGIA

3.1. *Abordagem Espacial*

Com o aumento da disponibilidade de dados com referência à localização geográfica, nasce a necessidade de incorporação desta informação na análise dos mesmos, surgindo assim entre finais dos anos 60 e início dos anos 70, a Econometria Espacial. Segundo LeSage (1999), a Econometria Espacial distingue-se da Econometria tradicional no sentido em que tem novos desafios a incorporar na sua abordagem, nomeadamente, a dependência espacial entre as observações e a heterogeneidade espacial nas relações que se modelam.

A dependência espacial significa que as observações localizadas em i dependem de observações localizadas em j

$$y_i = f(y_j) \quad j = 1, \dots, n; \quad j \neq i \quad (1)$$

A dependência espacial existe maioritariamente por duas razões: erros de medição na recolha dos dados e a importância da inter-relação espacial na modelação de dados sociodemográficos e socioeconómicos.

A heterogeneidade espacial é referente à variação das relações no espaço geográfico que pode ser traduzida em

$$y_i = X_i\beta_i + \varepsilon_i \quad i = 1, \dots, n \quad (2)$$

Onde y_i representa a variável dependente no local i , X_i representa o vetor ($1 \times k$) das variáveis independentes ao qual está associado um vetor de parâmetros β_i e ε_i representa o erro estocástico desta relação linear.

3.2. *Matriz W e Desfasamento Espacial*

3.2.1. *Matriz W*

Segundo Arbia (2014), uma das maiores diferenças entre a Econometria tradicional e a Econometria Espacial é o facto de serem necessários dois tipos diferentes de informação para a modelação dos dados espaciais: dados relativos às variáveis económicas e dados relativos à localização dessas mesmas variáveis e, conseqüentemente, informação de

proximidade entre as observações. Caso os dados fossem observados numa grelha, a definição de proximidade ou vizinhança seria intuitiva, escolhendo-se um critério desde o *rook*, em que duas unidades são consideradas vizinhas caso partilhem uma aresta, ao critério *queen*, em que duas unidades são consideradas vizinhas caso partilhem um vértice ou uma aresta. Contudo, estes critérios de vizinhança não podem ser aplicados diretamente na prática uma vez que se trabalha com unidades irregulares, sendo então necessário o recurso à matriz de ponderação espacial, ou matriz W

$$W = \begin{bmatrix} w_{11} & \cdots & w_{n1} \\ \vdots & w_{ij} & \vdots \\ w_{1n} & \cdots & w_{nn} \end{bmatrix}, \quad w_{ij} = \begin{cases} 1, & \text{se } j \in N(i) \text{ com } j \neq i \\ 0, & \text{caso contrário} \end{cases} \quad (3)$$

em que, $N(i)$ representa o conjunto de localizações vizinhas à localização i e $w_{ii} = 0$ por definição. Os critérios de vizinhança para $N(i)$ podem variar desde contiguidade, distância máxima até *nearest neighbor criterion*.

3.2.2. Desfasamento Espacial

De forma análoga ao conceito de desfasamento temporal utilizado em séries temporais, existe também o conceito de desfasamento espacial. Enquanto em séries temporais o desfasamento implica uma deslocação no tempo, neste âmbito um desfasamento implica uma deslocação no espaço, embora que de uma forma restrita. Na prática, o conceito de desfasamento espacial está relacionado com a vizinhança de uma dada localização sendo que neste contexto funciona para criar uma média ponderada das observações vizinhas.

3.3. Modelos Econométricos Espaciais

A Econometria tradicional tem vindo a ignorar os desafios implicados pelos dados espaciais, apesar de, segundo LeSage (1999), estes violarem as hipóteses de Gauss-Markov na modelação. Assim, nesta secção referem-se alguns modelos mais utilizados no âmbito da análise de dados *cross-section* (ou dados em corte transversal) que se podem estimar quando a hipótese de ausência de autocorrelação espacial dos erros é violada de modo a remover a dita autocorrelação. De modo a captar a correlação espacial, opta-se pela introdução de variáveis desfasadas espacialmente. Estes desfasamentos podem estar presentes na variável dependente, nas variáveis explicativas ou no termo de erro. Neste TFM apenas serão apresentados modelos que incorporam a dependência espacial não havendo foco na abordagem da heterogeneidade espacial.

3.3.1. Modelo de Regressão Espacial Clássico com Regressores Espacialmente Desfasados

Neste modelo, Modelo de Regressão Espacial Clássico com Regressores Espacialmente Desfasados (SLX), assume-se que as variáveis explicativas de uma dada localização afetam a variável dependente das localizações vizinhas e, conseqüentemente, as variáveis explicativas das localizações vizinhas explicam, também, a variação da variável dependente numa dada localização, segundo Almeida (2012). Em termos práticos esta influência pode ser representada pela introdução no Modelo de Regressão Clássico de uma componente que inclui as variáveis independentes espacialmente desfasadas

$$y = X\beta + WX\tau + \varepsilon \quad (4)$$

onde, X representa a matriz ($n \times k$) das variáveis independentes ao qual está associado um vetor de parâmetros β ($k \times 1$), W representa a matriz de ponderação espacial ($n \times n$) e τ representa um vetor de parâmetros ($k-1 \times 1$) associado à matriz WX , sendo que alguns elementos deste vetor podem assumir o valor zero, implicando assim que não é obrigatório introduzir no modelo os desfasamentos espaciais de todas as variáveis incluídas em X . Assume-se que o termo de erro verifica as hipóteses clássicas, nomeadamente ausência de autocorrelação espacial e homocedasticidade.

De acordo com Arbia (2014), quando a matriz de todos os regressores tem característica completa e o erro não é autocorrelacionado e é homocedástico este modelo pode ser simplesmente estimado pelo método OLS sem que assim surjam quaisquer problemas de estimação.

3.3.2. Spatial Lag Model

Este modelo, *Spatial Lag Model* (SLM), assume a hipótese de que a variável explicada de uma dada região tem impacto na variável explicada das regiões vizinhas, e vice-versa. Na sua forma mais pura assume a forma

$$y = \rho Wy + \varepsilon \quad (5)$$

onde, Wy representa um vetor ($n \times 1$) de desfasamentos espaciais da variável y e ρ representa o coeficiente de autoregressão espacial que assume valores entre -1 e 1 nunca atingindo os valores extremos (Almeida, 2012). Neste modelo, a variação da variável dependente depende exclusivamente do valor da mesma nas regiões vizinhas e no caso em que o coeficiente ρ não é estatisticamente significativo assume-se a não existência de

autocorrelação espacial da variável dependente. Assume-se também neste modelo que o termo de erro verifica as hipóteses clássicas de ausência de autocorrelação espacial e homocedasticidade.

Para enriquecer a capacidade explicativa da variação da variável dependente, incluem-se no modelo (5) outras variáveis exógenas, assumindo a forma do Modelo de Regressão Clássico adicionado da componente que inclui o desfasamento espacial da variável dependente

$$y = \rho W y + X\beta + \varepsilon \quad (6)$$

A equação (6) pode escrever-se na forma reduzida

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon \quad (7)$$

Uma vez que neste modelo a variável dependente espacialmente desfasada é um dos regressores, surge então um problema de endogeneidade, não podendo este modelo ser estimado pelo método OLS sendo necessário recorrer ao método de estimação da Máxima Verossimilhança ou ao método dos Mínimos Quadrados em dois Passos (*Two-Stage Least Squares*, 2SLS).

3.3.3. Spatial Error Model

No modelo SEM, assume-se que a dependência espacial afeta a variável explicada ainda que de uma forma residual. Esta dependência espacial assume a forma de um desfasamento espacial no termo de erro, podendo assim ser representado por um Modelo de Regressão Clássico adicionado de uma componente de desfasamento espacial no erro

$$y = X\beta + \varepsilon \text{ com } \varepsilon = \lambda W\varepsilon + u \quad (8)$$

onde, λ representa o coeficiente de desfasamento espacial do termo de erro. Assim, no modelo (8), de acordo com Almeida (2012), os erros de uma qualquer localização representam a média ponderada dos erros das localizações vizinhas adicionada de uma componente de erro aleatória $u \sim N(0, \Omega)$.

Uma vez que, devido à presença de autocorrelação dos erros, estes não são i.i.d., a estimação OLS perde as suas propriedades ótimas, sendo necessário recorrer ao método de estimação da Máxima Verossimilhança ou ao método dos Mínimos Quadrados Generalizável Praticável (*Feasible General Least Squares*, FGLS).

3.3.4. Modelo SARAR

Neste tipo de modelo assumem-se ambas as hipóteses referidas nos dois modelos anteriores, isto é, de uma forma muito sucinta, assume-se que a dependência espacial se verifica tanto na variável dependente como no termo de erro. Ou seja, existe um efeito espacial na variável dependente tanto de forma direta, assim como de forma residual. Em termos práticos, este modelo assume a forma de um Modelo de Regressão Clássico adicionado de um desfasamento espacial da variável dependente e de um desfasamento espacial no termo de erro

$$y = \rho W_1 y + X\beta + \varepsilon \quad \text{com} \quad \varepsilon = \lambda W_2 \varepsilon + u \quad e \quad u \sim N(0, \Omega) \quad (9)$$

onde, W_1 e W_2 representam duas matrizes de ponderação espaciais ($n \times n$) que podem possuir diferentes estruturas espaciais. É ainda necessário assumir as restrições de estabilidade $|\rho| < 1$ e $|\lambda| < 1$.

Assim como nos dois modelos anteriores, surgem neste modelo problemas de estimação. A variável dependente espacialmente desfasada é endógena e existe autocorrelação no termo de erro, assim de modo a estimar este modelo temos de aplicar o método da Máxima Verosimilhança, uma versão espacial do 2SLS (*Generalized Spatial Two-Stage Least Squares*, GS2SLS) ou então o estimador das variáveis instrumentais de Lee (LIV) (Lee, 2003).

3.3.5. Modelo de Regressão Espacial Clássico de Durbin

Neste modelo, abreviadamente SDM, a variável explicada depende das variáveis explicativas espacialmente desfasadas assim como dela própria espacialmente desfasada, assumindo a forma de desfasamentos espaciais da variável y tal como dos regressores. Adicionados estes termos ao Modelo de Regressão Clássico obtemos a equação geral deste modelo

$$y = \rho W_1 y + X\beta + W_2 X\tau + \varepsilon \quad (10)$$

Assim como nos modelos SLX e SLM, também neste modelo se assume que o termo de erro verifica as hipóteses clássicas.

Mais uma vez, devido à presença de endogeneidade provocado pela presença da variável explicada espacialmente desfasada, este modelo não pode ser estimado pelo método OLS.

É então necessário o recurso aos métodos da Máxima Verossimilhança ou ao método dos Mínimos Quadrados em dois Passos (*Two-Stage Least Squares, 2SLS*)

3.3.6. Modelo de Regressão Espacial de Durbin com Erro Espacialmente Desfasado

No Modelo de Regressão Espacial de Durbin com Erro Espacialmente Desfasado, assumem-se as hipóteses de que a dependência espacial é captada tanto através das variáveis explicativas assim como do termo de erro espacialmente desfasados. A equação geral deste modelo é a seguinte

$$y = X\beta + W_1X\tau + \varepsilon \text{ com } \varepsilon = \lambda W_2\varepsilon + u \text{ e } u \sim N(0, \Omega) \quad (11)$$

Segundo Almeida (2012), impõe-se ainda uma restrição sobre parâmetro λ , $|\lambda| < 1$. Este modelo é estimado através do método da Máxima Verossimilhança ou do método dos Mínimos Quadrados Generalizável Praticável (*Feasible General Least Squares, FGLS*)

3.4. Efeitos Parciais

Quando se trabalha com um Modelo de Regressão Clássico a interpretação dos coeficientes é direta sendo que representa a derivada parcial da variável explicada na localização i em ordem à variável explicativa nessa mesma localização (Almeida, 2012). Quando se entra no âmbito da Econometria Espacial e se trabalha com os modelos referidos anteriormente, a interpretação dos coeficientes deixa de ser direta uma vez que tem de ser tido em conta a dependência espacial e o associado impacto das variáveis referentes às localizações vizinhas na variável dependente de determinada região. Assim, associado a cada variável explicativa, temos um efeito direto, um efeito indireto e um efeito total na variável explicada.

O efeito direto de uma determinada variável, k , é dado pela derivada parcial da variável dependente em i em ordem à variável explicativa k em i

$$\frac{\partial y_i}{\partial X_{ki}} \quad (12)$$

O efeito indireto de uma variável, k , fornece informação acerca do impacto que uma variação nessa variável na localização vizinha j tem na variável explicada da localização i , é dado pela derivada parcial da variável dependente em i em ordem à variável k em j

$$\frac{\partial y_i}{\partial X_{kj}} \quad (13)$$

O efeito total de uma variável, k , é a soma dos efeitos diretos e indiretos, assumindo diversas formas dependendo do modelo em questão.

Para o modelo SLX, o efeito total de uma variável k é dado por

$$\beta_k + \tau_k \quad (14)$$

Para o modelo SLM, o efeito direto irá variar de acordo com a localização i atendendo à equação (7). Assim é preferível apresentar os efeitos parciais na sua expressão matricial,

$$(I_n - \rho W_n)^{-1} I_n \beta_k \quad (15)$$

A equação (15) origina uma matriz ($n \times n$) de efeitos parciais para a variável k . Os elementos da diagonal principal desta matriz representam os efeitos diretos provocados na variável dependente observada numa dada região, pela variação da variável k nessa mesma região. O efeito direto médio da variável k é apenas a média dos valores observados na diagonal principal. O efeito indireto da variável k na região i representa a soma dos efeitos que a variação dos valores observados da variável k nas regiões j vizinhas de i , com $j = 1, \dots, n$, e $j \neq i$, têm na variável dependente localizada em i . Este efeito indireto é assim obtido pela soma de todos os elementos da linha i , excetuando os elementos da diagonal principal. O efeito indireto médio da variável k é a média das somas dos elementos de cada linha com a exceção dos elementos da diagonal principal. O efeito total médio da variável k é a soma do efeito direto médio com o efeito indireto médio.

Para o modelo SEM, o efeito total de uma variável não se altera relativamente ao Modelo de Regressão Clássico, sendo apenas β_k .

Para o modelo SARAR, o efeito total é dado de forma semelhante à do modelo SLM.

Para o modelo SDM, pelos motivos indicados relativamente ao modelo SLM, será também mais conveniente analisar a expressão matricial

$$(I_n - \rho W_n)^{-1} I_n \beta_k + (I_n - \rho W_n)^{-1} I_n \tau_k \quad (16)$$

Semelhantemente à equação (15), também a equação (16) dá origem a uma matriz de efeitos parciais para a variável k , onde os efeitos médios (direto, indireto e total) são obtidos de forma semelhante à descrita para o modelo SLM mas tendo como base a matriz (16).

Para o modelo SDEM, o efeito total é dado de forma semelhante à do modelo SLX.

3.5. Testes de dependência espacial

3.5.1. Teste I de Moran

Introduzido por Moran (1950) e proposto por Cliff & Ord (1972), o teste I de Moran tem sido o mais utilizado para testar a presença de autocorrelação espacial nos resíduos dos modelos, assumindo na hipótese nula a ausência desta. Mais uma vez, comparando ao que se verifica em séries temporais, este teste pode-se considerar análogo ao teste de Durbin-Watson que testa a presença de autocorrelação dos resíduos dos modelos de séries temporais.

A estatística de Moran, na sua forma mais simples assume a forma do coeficiente de correlação entre os resíduos do modelo e os resíduos espacialmente desfasados

$$\text{Corr}(\varepsilon, L\varepsilon) = \frac{\text{Cov}(\varepsilon, L\varepsilon)}{\sqrt{\text{Var}(\varepsilon)\text{Var}(L\varepsilon)}} \quad (17)$$

em que, ε representa os resíduos do modelo e L representa o operador de desfasamento espacial.

Ao contrário das séries temporais, a variância dos resíduos do modelo e a variância dos resíduos do modelo desfasados não são iguais, neste contexto $\text{Var}(\varepsilon) \geq \text{Var}(L\varepsilon)$. Assim, a estatística de Moran tem em atenção este enviesamento assim como um fator de normalização que é igual à soma das ponderações dadas pelos elementos da matriz W ,

$$I = \frac{n\varepsilon^T W\varepsilon}{\varepsilon^T \varepsilon \sum_i \sum_j w_{ij}} \quad (18)$$

Quando a matriz W é estandardizada por linhas $\sum_i \sum_j w_{ij} = n$ (onde n representa o número de localizações no modelo), podendo a expressão (18) ser simplificada para

$$I = \frac{\varepsilon^T W\varepsilon}{\varepsilon^T \varepsilon} \quad (19)$$

A distribuição da estatística I de Moran foi deduzida por Cliff & Ord (1972) sob a hipótese de normalidade e também para a situação de não normalidade dos erros. Os autores ainda mostram que a distribuição assintótica é normal nas duas situações e o valor esperado da

estatística, ao contrário da variância¹, não depende da hipótese de distribuição dos erros assumida e é dado por

$$E(I) = \frac{n \operatorname{tr}(M_x W)}{S_0(n - k)} \quad (20)$$

onde, $S_0 = \sum_i \sum_j w_{ij}$, $M_x = I - P_x$ e $P_x = X(X^T X)^{-1} X^T$.

O teste de I de Moran pode ainda ser aplicado a vetores de informação espacial seguindo os mesmos princípios referidos anteriormente, substituindo o vetor dos resíduos do modelo pelo vetor de dados a testar. Apesar de tudo o referido anteriormente, o teste I de Moran possui uma limitação, que é o facto de não se basear numa hipótese alternativa explícita.

3.5.2. Testes de Multiplicador de Lagrange

Combatendo a limitação do teste I de Moran, surgem os testes de Multiplicador de Lagrange (LM) para a dependência espacial, onde a hipótese alternativa assume a forma de um dos diversos modelos espaciais (referidos na secção anterior). Segundo Anselin (1988), os testes LM são testes bastante convenientes, uma vez que se baseiam na estimação do modelo segundo a hipótese nula (H0).

Considere-se um modelo espacial geral (SARAR):

$$y = \rho W_1 y + X\beta + \varepsilon \quad \text{com} \quad \varepsilon = \lambda W_2 \varepsilon + u \quad \text{e} \quad u \sim N(0, \Omega) \quad (21)$$

onde, X representa a matriz ($n \times k$) das variáveis independentes ao qual está associado um vetor de parâmetros β ($k \times 1$), ρ representa o coeficiente da variável dependente espacialmente desfasada, λ representa o coeficiente de desfasamento espacial do erro do modelo e W_1 e W_2 representam duas matrizes de ponderação espaciais ($n \times n$) que podem possuir diferentes estruturas espaciais.

No caso em que $\rho = 0$ e $\lambda = 0$, obtém-se um Modelo de Regressão Clássico

$$y = X\beta + \varepsilon \quad (22)$$

No caso em que $\lambda = 0$, obtém-se um modelo conhecido como *Spatial Lag Model*

$$y = \rho W_1 y + X\beta + \varepsilon \quad (23)$$

¹ Para as fórmulas da variância consultar Cliff & Ord (1972).

Quando $\rho = 0$, fica-se com um modelo designado por *Spatial Error Model*

$$y = X\beta + \lambda W_2 \varepsilon + \mu \quad (24)$$

O modelo restrito, Modelo de Regressão Clássico (22), pode assim ser visto como a hipótese nula do teste LM. Esta será contraposta contra a hipótese alternativa (H1) que pode assumir a forma de um dos modelos mais gerais (21), (23) ou (24) tornando assim o teste de má especificação do modelo num mero problema de variáveis omitidas.

4. ANÁLISE EXPLORATÓRIA DOS DADOS

4.1. Caracterização dos dados

A base de dados utilizada para a realização deste estudo foi criada utilizando maioritariamente dados provenientes do PORDATA e do INE, com a exceção das variáveis índice de desenvolvimento humano e percentagem da população com acesso a sistemas de esgotos que foram retiradas do estudo de Silva & Ferreira-Lopes (2014). A unidade de observação é o município, reportando-se ao período entre 2008 e 2015, inclusive, e a amostra utilizada para esta análise é constituída por 278 observações.

O tratamento dos dados foi realizado com recurso aos softwares R, Microsoft Excel, Q-Gis e GeoDa. O Microsoft Excel foi utilizado para a realização de vários cálculos com os dados recolhidos de modo a obter-se as variáveis finais utilizadas no modelo (cálculo de médias e variações, por exemplo). O software Q-Gis foi utilizado como modo de agregação de dados enquanto que o GeoDa foi utilizado para a análise exploratória espacial inicial dos dados, nomeadamente a análise de clusters *k-means*. Por fim, no software R procedeu-se a toda a modelação apresentada nesta dissertação.

4.2. Caracterização das variáveis

De modo a clarificar o estudo realizado, é conveniente expor a metodologia de como a variável dependente (*cres*) foi criada. Utilizando dados obtidos através do PORDATA subtraiu-se a estimativa da população residente em cada município no ano de 2008 à estimativa da população para o ano de 2015 (essas estimativas referem-se a 31 de dezembro do respetivo ano). Antes de se efetuar a subtração, retirou-se à estimativa da população de 2015 a soma do total do número de nados vivos de 2009 a 2015. Analogamente, retirou-se à população residente do ano de 2008 a soma do total do número de óbitos de 2009 a 2015. Para efeitos de simplificação assume-se que todos os óbitos registados no município no período referido se referem a pessoas que já se

encontravam no município em 2008 e assume-se, também, que as crianças que nasceram no município no período em causa se mantiveram no mesmo até 31 de dezembro de 2015. Por fim, dividiu-se o resultado da subtração final pela estimativa da população estabilizada do ano de 2008. O método utilizado para estabilização das estimativas da população tanto em 2008 como em 2015, é o método utilizado no artigo de Jacinto & Ramos (2010).

De modo a selecionar-se quais as variáveis a incluir no modelo² teve-se em conta as variáveis já utilizadas em estudos semelhantes e apresentados na secção da revisão da literatura, a interpretação económica das variáveis e, por fim, a matriz de correlação das variáveis de modo a que não existam casos de multicolinearidade (anexo 2). Após passarem nestes três critérios referidos, selecionaram-se as variáveis expostas na tabela I de modo a explicar o crescimento da população nos municípios portugueses entre 2008 e 2015 devido a migração.

Tabela I: Descrição das variáveis utilizadas

cres	Crescimento da população entre 2008 e 2015 devido a migração
desemp	Taxa de desemprego média entre 2009 e 2015
hab	Média do preço médio anual por m ² das habitações entre 2009 e 2015
sal	Média do ganho/salário médio mensal entre 2011 e 2015
metro	Municípios pertencentes às áreas metropolitanas de Lisboa e Porto
int	Municípios pertencentes ao interior português
dist	Municípios que são capitais de distrito
idd	Média de idades da população em 2011
nfin	Média do número de empresas não financeiras entre 2009 e 2015 por cada 100 habitantes
pescs	Pessoal ao serviço nos centros de saúde por cada mil habitantes em 2009
sup	Percentagem da população com o ensino superior completo em 2011
setor_prim	Percentagem da população ativa empregada no setor primário em 2011
setor_sec	Percentagem da população ativa empregada no setor secundário em 2011
bc	Média da balança comercial entre 2009 e 2015
idh	Índice de desenvolvimento humano em 2008
esg	Percentagem da população com acesso a sistemas de esgotos em 2008

² As variáveis criadas, mas não utilizadas encontram-se descritas no anexo 1.

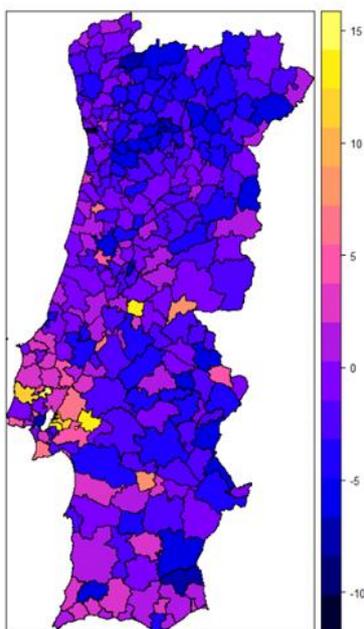
As variáveis selecionadas podem ser divididas em três categorias: variáveis geográficas (área metropolitana, interior e capital de distrito), variáveis sociodemográficas (média de idades, pessoal ao serviço em centros de saúde e percentagem da população com o ensino superior concluído) e variáveis socioeconómicas (taxa de desemprego, preço médio da habitação, salário mensal médio, número de empresas não financeiras, percentagem da população ativa empregada no setor primário, percentagem da população ativa empregada no setor secundário, balança comercial, índice de desenvolvimento humano e percentagem da população com acesso a sistemas de esgotos). As variáveis índice de desenvolvimento humano e percentagem da população com acesso a sistemas de esgotos são utilizadas como variáveis proxy do nível de desenvolvimento do município, representando respetivamente a percentagem da população com acesso a sistemas de esgotos em 2008 na região onde o município se insere e o índice de desenvolvimento humano em 2008 também na região onde o município está inserido. Estas variáveis são utilizadas juntamente com as variáveis pessoal ao serviço em centros de saúde e número de empresas não financeiras uma vez que no caso de Portugal não existe um índice de infraestruturas para medir o grau de desenvolvimento de um município como é o caso no estudo de Etzo (2011). A variável pessoal ao serviço em centros de saúde representa o número do pessoal ao serviço nos centros de saúde por cada mil habitantes em 2009, já a variável número de empresas não financeiras representa a média do número de empresas não financeiras entre 2009 e 2015 por cada 100 habitantes. As variáveis área metropolitana, interior e capital de distrito são variáveis *dummy* que assumem o valor 1 quando um município pertence à categoria referente e o valor zero caso contrário. Elas representam, respetivamente, os municípios pertencentes às áreas metropolitanas de Lisboa e Porto (sendo que os municípios de Lisboa e Porto não estão incluídos nesta caracterização), os municípios pertencentes ao interior português, de acordo com a caracterização feita por Bigote et al. (2014) que assumem como litoral português “*the stretch of land approximately 50 km wide located along the coast between Lisbon and Porto plus the southern region of Algarve*” sendo então as restantes regiões pertencentes ao interior e, por fim, os municípios que são capitais de distrito. As variáveis sociodemográficas ainda não referidas anteriormente, *idd* e *sup*, representam a média de idades da população do concelho em 2011 e a percentagem da população com o ensino superior completo em 2011, respetivamente. A variável *desemp* refere-se à taxa de desemprego média entre 2009 e 2015, *hab* representa a média do preço médio anual por m^2 das habitações entre 2009 e 2015, *sal* representa a média do ganho/salário médio

mensal entre 2011 e 2015 que serve como *proxy* do nível de rendimento do município, *setor_prim* representa a percentagem da população ativa empregada no setor primário em 2011, *setor_sec* representa a percentagem da população ativa empregada no setor secundário em 2011 e, por último, *bc* representa a média da balança comercial entre 2009 e 2015 que funciona como indicador do dinamismo económico do município.

Na figura presente no anexo 3, estão patentes as estatísticas descritivas das variáveis onde são apresentados os valores mínimos, máximos, médios e os quantis de todas as variáveis. É de notar que a média do crescimento da população devido a migração é negativa (-1,004) o que nos indica que Portugal perdeu população para o exterior. Os municípios possuem em média uma taxa de desemprego de 8,259%, a idade média da população por concelho é de 44,33 anos, cada município tem, em média, 10,40 empresas não financeiras por cada 100 habitantes, o valor médio por metro quadrado das habitações é de 900,5 euros e o ganho médio mensal é de 885,5 euros. Cada município tem, em média, 3,653 pessoas ao serviço nos centros de saúde por cada mil habitantes e 9,34% da população com o ensino superior completo. Em cada concelho, 7,488% da população ativa trabalha no setor primário, em média, e 27,59% trabalha no setor secundário. A balança comercial média por município tem um valor médio de -44,842 milhares de euros enquanto que cada município tem em média um índice de desenvolvimento humano de 0,6847 e em média 79,74% da população tem acesso a sistemas de esgotos.

4.3. *Análise Exploratória de Dados Espaciais*

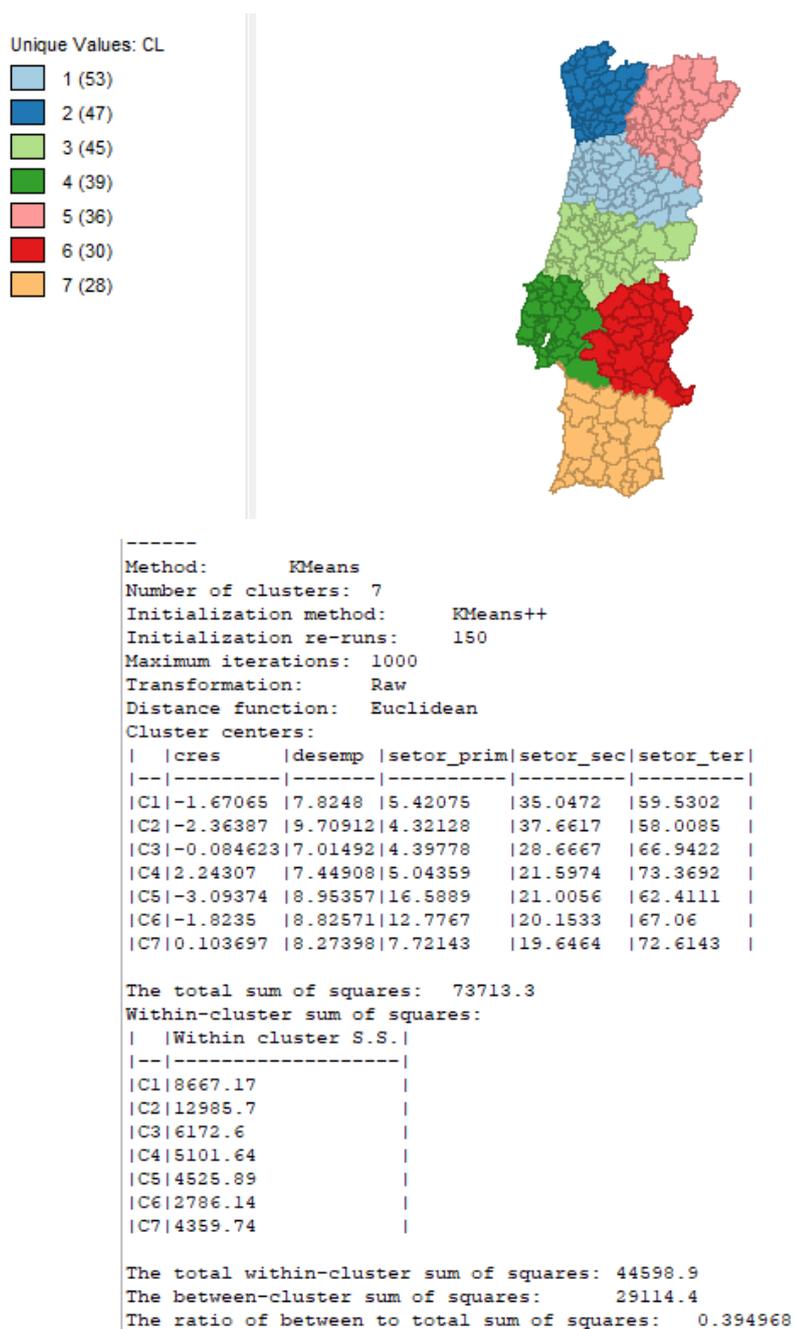
A figura 1 permite analisar a dispersão geográfica da variável dependente, ou seja, do crescimento da população devido a migração. Como se pode verificar, existe uma concentração de valores próximos da variável dependente em regiões vizinhas. Com esta análise preliminar, não parece correto assumir que a distribuição do crescimento da população por município entre 2008 e 2015 é aleatória, existindo evidência de correlação espacial entre os diversos concelhos. Sendo que os valores mais baixos de crescimento estão representados a azul escuro e os valores mais altos de crescimento a amarelo, pode-se observar um menor crescimento na região Norte de Portugal enquanto na Área Metropolitana de Lisboa parece ter havido um maior crescimento da população.

Figura 1: Dispersão geográfica do crescimento populacional entre 2008 e 2015

Fonte: Elaboração própria

Realizou-se ainda uma análise de clusters *k-means* de modo a poder verificar-se a dispersão geográfica de algumas variáveis, assim como quais as características que caracterizam os grupos de municípios de cada cluster. Para esta análise, recorreu-se mais uma vez à variável *cres*, assim como às variáveis *desemp*, *setor_prim*, *setor_sec* e a variável *setor_ter* que representa a percentagem da população ativa empregada no setor terciário. Esta análise está representada na figura 2. Aqui pode verificar-se, mais uma vez, a evidência da existência de correlação espacial visto que os clusters agregam municípios vizinhos. A título de curiosidade pode verificar-se com esta análise que o cluster que teve o crescimento da população mais baixo foi o cluster 5, este é caracterizado por ser o segundo cluster com a maior taxa de desemprego, o cluster com a maior percentagem da população ativa empregada no setor primário e o cluster com as terceiras mais baixas percentagens da população ativa empregada no setor secundário e terciário. O cluster em que se verificou o maior crescimento da população foi o cluster 4 e este caracteriza-se por ter a segunda menor taxa de desemprego, a terceira menor percentagem da população ativa empregada no setor primário, a quarta maior percentagem da população ativa empregada no setor secundário e a maior percentagem da população ativa empregada no setor terciário.

Figura 2: Análise de *clusters k-means* das variáveis *cres*, *desemp*, *setor_prim*, *setor_sec* e *setor_ter*



Fonte: Elaboração própria

4.4. Dependência Espacial

De modo a avaliar a existência de dependência espacial no crescimento da população devido a migração por município, procedeu-se à realização do teste I de Moran. Neste teste assume-se na hipótese nula a ausência de correlação espacial, logo se esta hipótese for rejeitada, confirma-se a presença de correlação espacial na variável dependente. Para a realização deste teste é necessária a escolha de uma matriz W de modo a definir o critério de ponderação espacial a utilizar. Neste estudo, sempre que necessário o uso da

matriz W , tanto em testes como em modelação, utilizar-se-á a uma matriz de contiguidade *Queen*, uma vez ter sido este o critério selecionado. Esta matriz é estandardizada por linhas o que significa que a soma dos elementos de cada linha é 1. Ao se realizar o teste I de Moran obteve-se um valor-p que permite rejeitar a hipótese nula a 1% (tabela II), provando assim necessária a utilização de ferramentas de econometria espacial para o estudo do crescimento da população devido a migração.

Tabela II: Teste de Moran à variável dependente

Estadística de Moran	Valor-p
0,3082	0,000

Esta dependência espacial é também confirmada pelo diagrama de dispersão de Moran apresentado no anexo 4, uma vez que a reta da regressão possui um declive positivo o que confirma uma dependência espacial positiva na variável *cres*.

Ainda para confirmar a existência de correlação espacial, estimou-se um modelo SLM puro apenas com a variável dependente espacialmente desfasada como variável explicativa (anexo 5). O valor-p obtido para o teste de significância individual da estimativa do coeficiente da variável dependente espacialmente desfasada permite rejeitar a hipótese nula de não significância estatística a 1%, sendo assim, mais uma vez, confirmada a necessidade da inclusão de ferramentas econométricas espaciais no estudo do crescimento da população devido a migração.

5. MODELAÇÃO E RESULTADOS

De modo a iniciar a análise sobre os fatores que explicam a migração nos concelhos portugueses, começou-se por se estimar um Modelo de Regressão Clássico com todas as variáveis referidas na secção 4 (tabela III). Após a estimação deste modelo (modelo 1), realizou-se um teste I de Moran de modo a verificar a existência de correlação espacial nos resíduos do modelo (tabela IV). Uma vez confirmada a existência de correção espacial nos resíduos do modelo, procedeu-se à criação das variáveis utilizadas no modelo anterior espacialmente desfasadas numa tentativa de eliminar a correlação espacial dos erros, estimando-se em seguida um modelo com todas as variáveis utilizadas no modelo 1 e com essas variáveis espacialmente desfasadas, o modelo SLX. Após a estimação deste modelo, realizou-se novamente o teste I de Moran verificando-se que a correlação espacial dos resíduos havia sido eliminada (tabela IV). De modo a validar a inferência

efetuada, realizou-se um teste de Breusch-Pagan para verificar a presença de heterocedasticidade no modelo 2 (tabela V), confirmando-se a existência desta ao nível de 10%. Em seguida, voltou-se a estimar o modelo com erros padrão robustos, apresentando-se este modelo na tabela III.

Tabela III: Modelo de Regressão Clássico e SLX não restrito para o crescimento populacional

	OLS (1)		SLX (2)	
	<u>Coeficiente</u>	<u>Desvio Padrão</u>	<u>Coeficiente</u>	<u>Desvio Padrão</u>
constante	27,2662***	4,5752	17,7808*	10,7045
desemp	-0,7350***	0,0852	-0,7181***	0,1381
int	-0,2410	0,4618	0,5916	0,7887
dist	-4,0088***	0,9128	-4,0816***	0,9785
metro	-0,4231	0,6274	0,3002	1,2039
idd	-0,3340***	0,0747	-0,4039***	0,1125
nfin	-0,2289*	0,1183	-0,3168**	0,1282
hab	0,1826	1,0093	-0,0639	1,1667
sal	1,3002	1,4784	-0,4527	2,2238
pescs	-0,2054	0,1679	-0,1734	0,1616
sup	-0,0275	0,0894	0,0788	0,1009
setor_prim	-0,1059**	0,0419	-0,0399	0,0493
bc	0,0006**	0,0003	0,0004*	0,0002
idh	-1,1051	2,1642	-9,3147**	4,2458
esg	0,8849	1,5700	1,8945	1,9882
setor_sec	-0,1462***	0,0271	-0,0812**	0,0381
desemp_lag	-	-	0,1364	0,1829
int_lag	-	-	-0,5813	1,0831
dist_lag	-	-	-0,8131	2,1949
metro_lag	-	-	-0,8875	1,6591
idd_lag	-	-	0,2375	0,1667
nfin_lag	-	-	0,1491	0,2403
hab_lag	-	-	-0,9253	1,9048
sal_lag	-	-	8,1115*	4,4170
pescs_lag	-	-	-0,3864	0,3081
sup_lag	-	-	-0,2872	0,1834
setor_prim_lag	-	-	-0,1420	0,0884
bc_lag	-	-	0,0009**	0,0004
idh_lag	-	-	11,7191*	6,5299
esg_lag	-	-	-3,1616	3,3429
setor_sec_lag	-	-	-0,1205**	0,0583
Dimensão da amostra	278		278	
Estatística de Wald	14,157		8,2498	
Valor-p do teste de Wald	0,0000		0,0000	
R²	0,4477		0,5005	

Níveis de significância: ***1%, **5% e *10%.

Tabela IV: Testes I de Moran aos resíduos dos modelos (1) e (2)

Teste de Moran	(1)	(2)
Estatística	2,6178	1,1107
Valor-p	0,0089	0,2667

Tabela V: Teste BP ao modelo (2)

Teste BP	Modelo (2)
Estatística	40,859
Valor-p	0,0893

Seguidamente, estimou-se o modelo SLX restrito com erros padrão robustos sem as variáveis estatisticamente não significativas ao nível de 10%. Neste modelo verificou-se que a migração nos concelhos portugueses é explicada pelas variáveis *desemp*, *dist*, *idd*, *nfin*, *setor_prim*, *bc*, *setor_sec*, *sal_lag*, *sup_lag*, *setor_prim_lag* e *setor_sec_lag* (tabela VII). Realizou-se novamente um teste I de Moran ao modelo e verificou-se que não existe correlação espacial dos resíduos ao nível de 5% (Anexo VII). Realizou-se também um teste F robusto à significância conjunta das variáveis retiradas do modelo 2, concluindo-se que estas não eram conjuntamente estatisticamente significativas (tabela IX). De modo a confirmar que o modelo SLX restrito era o que melhor captava a dependência espacial realizaram-se os diversos testes LM, sendo que em todos eles a hipótese nula não era rejeitada (tabela VI). Nos Testes LM o modelo SLX é contraposto contra o modelo SLM, SEM e SARAR, visto que temos presentes variáveis explicativas espacialmente desfasadas os dois primeiros assumem a forma de modelos SDM e SDEM, respetivamente.

Tabela VI: Testes LM ao modelo SLX restrito

Testes LM ao modelo 3	
	Valor-p
SEM	0,1272
SLM	0,1128
SARAR	0,2708

Apesar dos resultados obtidos nos Testes LM, ainda assim, estimaram-se os modelos SDEM e SDM pelos métodos FGLS e 2SLS robusto, respetivamente (tabela VII). Como se pode verificar na tabela VII, as estimativas para o lambda, no modelo SDEM, e para o rho, no modelo SDM, não são estatisticamente significativas.

Tabela VII: Modelos SLX restrito, SDEM e SDM para o crescimento populacional com erros padrão robustos

	SLX (3)		SDEM (4)		SDM (5)	
	<u>coeficiente</u>	<u>desvio padrão</u>	<u>coeficiente</u>	<u>desvio padrão</u>	<u>coeficiente</u>	<u>desvio padrão</u>
constante	28,8887***	4,6204	29,1509***	4,6961	29,132***	5,1431
desemp	-0,7145***	0,0848	-0,7128***	0,0848	-0,7036***	0,0892
dist	-3,5271***	0,6966	-3,5144***	0,7264	-3,5446***	0,7143
idd	-0,3984***	0,0790	-0,4035***	0,0594	-0,3932***	0,0783
nfin	-0,2030**	0,0985	-0,2171**	0,1055	-0,2128**	0,1049
setor_prim	-0,0730**	0,0349	-0,0716*	0,0406	-0,0687*	0,0393
bc	0,0006***	0,0002	0,0005**	0,0002	0,0006***	0,0002
setor_sec	-0,0868***	0,0304	-0,0871***	0,0315	-0,0863**	0,0356
sal_lag	6,9089**	2,9123	6,7069**	2,9794	5,8878	3,6024
sup_lag	-0,3400***	0,1161	-0,3268***	0,1244	-0,3282***	0,1227
setor_prim_lag	-0,1238**	0,0561	-0,1150*	0,0640	-0,1230**	0,0602
setor_sec_lag	-0,1088***	0,0347	-0,1057**	0,0426	-0,0631	0,0849
lambda	-	-	0,1309	0,2422	-	-
rho	-	-	-	-	5,0819e-36	8,2391e-36
Dimensão da amostra	278		278		278	
Estatística de Wald	16,932		211,8		176,9	
Valor-p do teste de Wald	0,0000		0,0000		0,0000	
R²	0,4679		0,4601		-	

Níveis de significância: ***1%, **5% e *10%.

Em alternativa ao modelo SLX apresentado na tabela III, decidiu-se, de modo a resolver a dependência espacial, aplicar os métodos de estimação SLM, SEM e SARAR ao modelo original, não utilizando as variáveis explicativas espacialmente desfasadas. Iniciou-se este processo pela realização dos testes LM ao Modelo de Regressão Clássico, sendo este a hipótese nula e os modelos SLM, SEM e SARAR as hipóteses alternativas respetivamente em cada um dos testes (tabela VIII). Pode-se verificar que em todos os testes o valor-p obtido permite rejeitar a hipótese nula a 5% e a 10% para a alternativa SEM, sendo que o teste que contrapõe o modelo SLM ao Modelo de Regressão Clássico regista o valor-p mais baixo.

Tabela VIII: Testes LM ao Modelo de Regressão Clássico

Testes LM ao modelo 1	
	<u>Valor-p</u>
SEM	0,0652
SLM	0,0119
SARAR	0,0350

Em seguida, procedeu-se à estimação do modelo SLM pelo método 2SLS com erros padrão robustos utilizando todas as variáveis referidas na secção 4³. Posteriormente, estimou-se sucessivamente o mesmo modelo retirando, uma a uma, as variáveis estatisticamente não significativas ao nível de 10%, resultando no modelo exposto na tabela X. Realizou-se, ainda, o teste de *Wald* à significância conjunta das variáveis retiradas ao modelo obtendo-se um valor-p de 0,98 (tabela IX). É de referir que a estimativa para o coeficiente do rho é estatisticamente significativa ao nível de 1%.

Tomou-se a mesma abordagem utilizada anteriormente na estimação dos modelos SEM e SARAR, isto é, começou-se por se estimar os modelos com todas as variáveis⁴ e, seguidamente, retiraram-se sucessivamente as variáveis estatisticamente não significativas até se atingir um modelo restrito em que todas as variáveis eram estatisticamente significativas ao nível de 10% (tabela X). O modelo SEM foi estimado pelo método FGLS e o modelo SARAR foi estimado pelo método GS2SLS. É de notar que em todos os três modelos restritos as variáveis estatisticamente significativas são as mesmas: *desemp*, *dist*, *idd*, *nfin*, *setor_prim*, *bc* e *setor_sec*. Como se pode verificar na

³ O modelo SLM não restrito encontra-se estimado no anexo 8.

⁴ Os modelos SEM e SARAR não restritos encontram-se estimados nos anexos 7 e 9, respetivamente.

tabela X o coeficiente para o lambda, no modelo SEM, não é estatisticamente significativo e no modelo SARAR, apesar dos coeficientes para o lambda e para o rho serem estatisticamente significativos ao nível de 10%, o valor-p para o teste de *Wald* da significância conjunta de ambas as estimativas é de 0,52932 (anexo 10). Assim, no seguimento dos resultados dos testes LM, pode-se afirmar que dentro dos modelos SLM, SEM e SARAR, o que melhor resolve a dependência espacial é o modelo SLM.

Tabela IX: Testes de Wald às variáveis omitidas nos modelos SLX restrito e SLM

Teste de Wald às variáveis omitidas nos modelos	(3)	(6)
Estatística	1,2867	2,1
Valor-p	0,1923	0,98

Ainda de modo a enriquecer o estudo realizado neste TFM, realizou-se uma análise preliminar da heterogeneidade espacial dos coeficientes. Estimou-se uma Regressão Geograficamente Ponderada (*Geographically Weighted Regression*, GWR) com as variáveis estatisticamente significativas comuns aos modelos (3) e (6), uma vez que este processo não permite a utilização de uma grande quantidade de variáveis, os resultados encontram-se expostos no anexo 11. Nesta análise, é possível verificar a disparidade dos coeficientes estimados, levando a concluir a existência de indícios de heterogeneidade espacial. Contudo, a extensão deste estudo não permitiu o desenvolvimento desta análise no limite temporal disponível para a realização de um TFM, sendo, no entanto, um tópico que merece atenção em trabalhos futuros.

Pelo estudo realizado, conclui-se que os modelos SLX e SLM são os que melhor servem para modelar o crescimento da população devido a migração nos municípios portugueses entre 2008 e 2015. Em ambos os modelos, acontecimentos ocorrentes nas regiões vizinhas determinam alterações da variável dependente de determinada região. No primeiro, o salário dos municípios vizinhos influencia positivamente o crescimento da população do município, enquanto que a percentagem de indivíduos com o ensino superior concluído, a percentagem da população ativa empregada no setor primário e a percentagem da população ativa empregada no setor secundário dos concelhos vizinhos influenciam negativamente o crescimento da população do concelho. No segundo modelo, o crescimento da população devido a migração das regiões vizinhas determina positivamente o crescimento da população na região.

É de salientar que tanto no modelo SLX restrito como no modelo SLM as variáveis do próprio concelho que influenciam o crescimento da população são as mesmas, influenciando a variável dependente no mesmo sentido, tendo coeficientes estimados muito próximos: a percentagem de população desempregada, o facto da região ser capital de distrito, a idade média dos habitantes, o número de empresas não financeiras por cada cem habitantes, a percentagem da população empregada no setor primário e a percentagem da população empregada no setor secundário influenciam o crescimento da população de forma negativa enquanto que a média da balança comercial de cada município influencia o crescimento da população de forma positiva.

Por fim, de modo a saber quanto cada variável determina o crescimento da população do concelho, estimaram-se os efeitos parciais para ambos os modelos, estando os resultados expostos nas tabelas XI e XII.

Os efeitos parciais verificados em ambos os modelos são maioritariamente negativos com a exceção da balança comercial média, nos modelos SLX e SLM, e do salário médio mensal, no modelo SLX. Nas variáveis em que é permitida realizar tal comparação, é possível verificar que os efeitos parciais são superiores em módulo no modelo SLM relativamente ao modelo SLX. No modelo SLX, o aumento de um ponto percentual da taxa de desemprego leva a uma diminuição média do crescimento da população devido a migração de 0,7145 pontos percentuais, mantendo-se todos os restantes fatores constantes. Já no modelo SLM, o aumento da taxa de desemprego em um ponto percentual leva a uma diminuição média de 0,9447 pontos percentuais do crescimento da população devido a migração, caso os restantes fatores se mantenham constantes. A variável cujo efeito total é mais próximo em ambos os modelos é a variável da balança comercial média, onde no modelo SLX um aumento da balança comercial média em um milhar de euros conduz a um aumento médio do crescimento populacional por migração em 0,006 pontos percentuais, com todos os restantes fatores mantendo-se constantes enquanto, no modelo SLM, um aumento da balança comercial média em um milhar de euros leva a um aumento do crescimento populacional por migração de, em média, 0,008 pontos percentuais, mantendo-se todos os restantes fatores constantes.

Tabela X: Modelos SLM, SEM e SARAR para o crescimento populacional com erros padrão robustos

	SLM (6)		SEM (7)		SARAR (8)	
	<u>coeficiente</u>	<u>desvio padrão</u>	<u>coeficiente</u>	<u>desvio padrão</u>	<u>coeficiente</u>	<u>desvio padrão</u>
constante	25,8647***	3,4431	30,0318***	3,3714	23,8961***	3,4395
desemp	-0,6250***	0,0905	-0,7413***	0,0858	-0,5790***	0,0817
dist	-3,6021***	0,7012	-3,8087***	0,6794	-3,5555***	0,6127
idd	-0,3387***	0,0554	-0,3984***	0,0573	-0,3070***	0,0633
nfin	-0,2295**	0,1025	-0,2321**	0,1089	-0,2160**	0,0897
setor_prim	-0,0666*	0,0344	-0,0985***	0,0336	-0,0660**	0,0270
bc	0,0005**	0,0002	0,0004*	0,0002	0,0006***	0,0001
setor_sec	-0,1157***	0,0210	-0,1388***	0,0201	-0,1116***	0,0161
rho	0,3384***	0,1121	-	-	-0,3224*	0,1841
lambda	-	-	0,1912	0,2269	0,4068***	0,1036
Dimensão da amostra	278		278		278	
Estatística de Wald	134,9		180,4		123,7	
Valor-p da estatística de Wald	0,0000		0,0000		0,0000	
R²	-		0,4122		-	

Níveis de significância: ***1%, **5% e *10%.

Tabela XI: Efeitos parciais do modelo SLX restrito

	direto	indireto	total
desemp	-0,7145	-	-0,7145
dist	-3,5271	-	-3,5271
idd	-0,3984	-	-0,3984
nfin	-0,2030	-	-0,2030
setor_prim	-0,0730	-0,1238	-0,1968
bc	0,0006	-	0,0006
setor_sec	-0,0868	-0,1088	-0,1956
sal	-	6,9089	6,9089
sup	-	-0,3400	-0,3400

Tabela XII: Efeitos parciais do modelo SLM

	direto	indireto	total
desemp	-0,6410	-0,3038	-0,9447
dist	-3,6941	-1,7508	-5,4449
idd	-0,3473	-0,1646	-0,5119
nfin	-0,2354	-0,1116	-0,3469
setor_prim	-0,0683	-0,0324	-0,1006
bc	0,0005	0,0003	0,0008
setor_sec	-0,1187	-0,0563	-0,1750

6. CONCLUSÕES E DESENVOLVIMENTOS FUTUROS

O crescimento da população devido a migração pode ser influenciado por diversas causas, como é demonstrado pelos diversos estudos apresentados na revisão da literatura. Neste estudo, analisam-se quais as causas que afetaram o crescimento populacional dos municípios portugueses durante a crise de 2008, sendo o período de análise entre 31 de dezembro de 2008 e 31 de dezembro de 2015, utilizando para tal dados espaciais.

A análise espacial realizada ao longo deste estudo é executada com recurso à utilização da matriz *W*, definida pelo critério *Queen*, de modo a captar a dependência espacial existente entre os municípios portugueses. Ao longo da referida análise, definem-se as variáveis consideradas pertinentes à migração interna com base em estudos anteriores realizados por diversos autores, procede-se à definição e à análise dos dados selecionados incluindo a realização de estatísticas descritivas dos mesmos e análise espacial da variável dependente, realiza-se a modelação e a respetiva interpretação dos outputs obtidos assim como dos efeitos parciais das variáveis para os modelos que melhor se ajustam a estes dados.

Quando se analisa o mapa de dispersão geográfica do crescimento da população, é notável que municípios próximos tendem a ter um comportamento semelhante. Ainda no que respeita à análise gráfica, é possível, através da análise de clusters *k-means*, verificar que os clusters com comportamentos semelhantes se localizam em áreas vizinhas. Os municípios que possuem uma maior migração caracterizam-se por uma baixa taxa de desemprego, uma baixa percentagem da população ativa empregada no setor primário e uma alta percentagem da população ativa empregada no setor terciário, levando à suspeita da existência de dependência espacial. A suspeita da existência de dependência espacial é confirmada e a utilização de ferramentas de Econometria Espacial justificada pela estatística do teste de Moran realizado à variável dependente do crescimento populacional de cada município, *cres*, uma vez que este permite rejeitar a hipótese de ausência de autocorrelação espacial. Esta utilização é ainda corroborada através da estimação de um modelo SLM puro com a variável dependente espacialmente desfasada como único regressor, pois o valor-p obtido do teste de significância individual do coeficiente associado a este regressor não rejeita a significância estatística. Relativamente à variável dependente, através das estatísticas descritivas pode-se verificar que a média desta tem um valor negativo, indicando que Portugal perdeu população para o estrangeiro no período em análise.

Visto que a estimação de um Modelo de Regressão Clássico não remove a dependência espacial procede-se à estimação de diversos modelos com o intuito de remover a referida dependência espacial. Pelos resultados obtidos, os modelos SLX e SLM são os que melhor se ajustam a estes dados. No modelo SLX, o resultado do valor-p do teste I de Moran aos resíduos não permite rejeitar a hipótese de ausência de autocorrelação espacial nos erros e ainda nos testes LM, quando este modelo é contraposto com modelos SDEM, SDM e SARAR com regressores espacialmente desfasados, a hipótese nula não é rejeitada. No caso do Modelo de Regressão Clássico, quando contraposto com os modelos SLM, SEM e SARAR, a hipótese nula é sempre rejeitada, sendo que, quando a hipótese alternativa é o SLM, o valor-p deste teste é o inferior, rejeitando com maior evidência o modelo clássico. Quando estimados os modelos SLM, SEM e SARAR, o único que apresenta um coeficiente associado à componente espacial estatisticamente significativo é o modelo SLM.

Retirando a ambos os modelos SLX e SLM as variáveis estatisticamente não significativas, obtemos resultados relativamente semelhantes. Ambos os modelos

indicam que as variações na migração de uma dada localização advêm de variações da taxa de desemprego, do facto do município ser ou não capital de distrito, da idade média dos habitantes, do número de empresas não financeiras por cada 100 habitantes, da percentagem de população ativa empregada no setor primário, da percentagem de população ativa empregada no setor secundário e da balança comercial da própria região. No caso do modelo SLX acrescem ainda, como fatores determinantes para a variação da migração de uma dada região, o salário mensal médio, a percentagem da população com o ensino superior concluído, a percentagem da população ativa empregada no setor primário e a percentagem da população ativa empregada no setor secundário das regiões vizinhas. No modelo SLX todas as variáveis têm um efeito negativo na migração de um dado concelho com a exceção da balança comercial do próprio concelho e do salário médio mensal dos concelhos vizinhos, caso no qual o crescimento populacional por migração aumenta com o aumento destas variáveis. No modelo SLM, tirando a variável da balança comercial, todas as variáveis têm um efeito negativo na migração de uma dada região, portanto, a migração aumenta com o aumento da balança comercial média e diminui com o aumento das restantes variáveis. Ainda no modelo SLM, acresce-se como variável explicativa a variável dependente espacialmente desfasada, implicando que a média do crescimento populacional devido a migração das regiões vizinhas afeta positivamente o crescimento populacional devido a migração de dada região.

A análise preliminar da heterogeneidade espacial dos coeficientes das variáveis utilizadas no modelo SLM, realizada através da estimação de um GWR, permite concluir que existem indícios de presença de heterogeneidade dos coeficientes. Uma vez que não coube no âmbito deste trabalho o desenvolvimento deste tema, um possível desenvolvimento futuro seria a realização de uma análise mais cuidada deste tópico. Pelo facto de não existir uma base de dados que permita obter informação relativa ao município de origem dos migrantes e para qual município esses migrantes se deslocaram, não foi possível neste TFM modelar os fluxos migratórios entre concelhos, o que enriqueceria a análise deste fenómeno. Sendo assim, espera-se que no futuro seja possível, aplicando uma base de dados que contenha esta informação, realizar uma análise de fluxos migratórios assim como Mulhern & Watson (2009), por exemplo, realizam.

REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida, E. (2012). *Econometria Espacial Aplicada*, 1ª Ed. Guanabara: Editora Alínea.
- Anselin, L. (1988). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, 20(1), 1–17.
- Arbia, G. (2014). *A Primer for Spatial Econometrics: With Applications in R*. Palgrave Macmillan.
- Bigotte, J. F., Antunes, A. P., Krass, D., & Berman, O. (2014). The Relationship between Population Dynamics and Urban Hierarchy: Evidence from Portugal. *International Regional Science Review*, 37(2), 149–171.
- Chun, Y. & Griffith, D. (2011) Modeling Network Autocorrelation in Space–Time Migration Flow Data: An Eigenvector Spatial Filtering Approach, *Annals of the Association of American Geographers*, 101:3, 523-536.
- Cliff, A.D. & Ord J.K. (1972). *Spatial Autocorrelation*, London: Pion
- Etzo, I. (2011). Determinants of Interregional Migration in Italy: A Panel Data Analysis. *SSRN Electronic Journal*, 1–28.
- Gomes, M. C. S., Moreira, M. J. G., & Pinto, M. L. R. (2017). Movimentos migratórios internos em Portugal (1995-2011): Caracterização exploratória considerando grupos de idades, níveis de habilitações e atividade. *Sociologia, Problemas e Praticas*, 83, 55–74.
- Jacinto, A. R. & Ramos, P. N. (2010). Movimentos migratórios regionais do interior português: quem saiu e quem entrou?. *Revista Portuguesa de Estudos Regionais*, (24),71-85
- Lee, L.F. (2003) Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances, *Econometric Reviews*, 22, 307–35.

Lesage, J. P. (2015). Theory and Practice of Spatial Econometrics. *Spatial Economic Analysis*, 10(3), 400–400.

Maria., J. R. (2016). Output and unemployment, Portugal, 2008–2012. *Working Papers Banco de Portugal*, No. w201603.

Moran, P.A.P. (1950). Notes on Continuous Stochastic Phenomena, *Biometrika* 37, 1, 17–23.

Mulhern, A. & Watson, J. (2009) Spanish Internal Migration: Is there Anything New to Say?, *Spatial Economic Analysis*, 4:1, 103-120.

Silva, R., & Ferreira-Lopes, A. (2014). A Regional Development Index for Portugal. *Social Indicators Research*, 118(3), 1055–1085.

Ubarevičiene, R., Van Ham, M., & Burneika, D. (2017). Shrinking regions in a shrinking country: The geography of population decline in Lithuania 2001-2011. *A+BE Architecture and the Built Environment*, 9, 91–125.

ANEXOS

Anexo 1: Variáveis não utilizadas

Variáveis criadas não utilizadas	
crimes	Média entre 2009 e 2015 de crimes por cada mil habitantes.
pdpc	Média entre 2009 e 2015 do poder de compra per capita.
fronteira	Variável dummy que assume o valor 1 caso o município faça fronteira com Espanha
região	Variável dummy que assume o valor 1 caso o município seja capital de região
ensino	Número total de estabelecimentos do ensino superior na sub-região NUTS 3 em 2015
hosp	Número de hospitais e parcerias público privadas em 2013
cs	Número de centros de saúde em 2012
dens	Número médio de 100 habitantes por km^2 em 2009
isus	Índice de sustentabilidade potencial em 2011

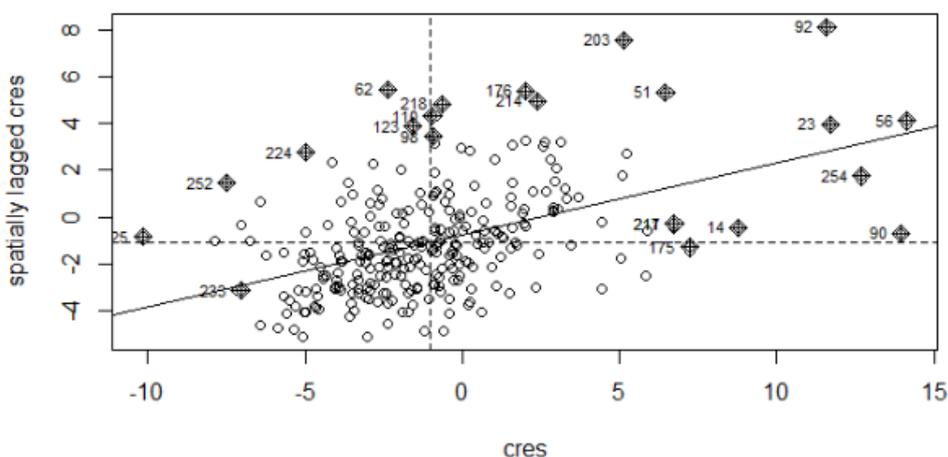
Anexo 2: Matriz de correlações das variáveis utilizadas

	cres	desemp	int	dist	metro	idd	nfin	hab
cres	1.00000000	-0.37780232	-0.25126662	-0.17459166	0.17304589	-0.23223538	-0.05730345	0.23406437
desemp	-0.37780232	1.00000000	0.04888762	0.01530820	0.09858935	-0.24169400	-0.15294157	-0.09389987
int	-0.25126662	0.04888762	1.00000000	0.02471644	-0.33414688	0.51599271	0.12750602	-0.26659783
dist	-0.17459166	0.01530820	0.02471644	1.00000000	-0.05137172	-0.11393366	0.23407203	0.03520099
metro	0.17304589	0.09858935	-0.33414688	-0.05137172	1.00000000	-0.38897808	-0.07931295	0.35637910
idd	-0.23223538	-0.24169400	0.51599271	-0.11393366	-0.38897808	1.00000000	0.07135804	-0.20154767
nfin	-0.05730345	-0.15294157	0.12750602	0.23407203	-0.07931295	0.07135804	1.00000000	0.26214275
hab	0.23406437	-0.09389987	-0.26659783	0.03520099	0.35637910	-0.20154767	0.26214275	1.00000000
sal	0.19300183	-0.07889274	-0.23474989	0.25876121	0.39351427	-0.32448941	0.18715588	0.32944742
pescs	-0.23165881	-0.03536080	0.45088286	-0.17327445	-0.36980584	0.70740454	0.01162924	-0.20741733
sup	0.15859848	-0.04272696	-0.30299131	0.57990336	0.38811682	-0.41740421	0.33290021	0.36435744
setor_prim	-0.24455340	0.09054800	0.57816522	-0.19497784	-0.31962401	0.46100341	0.28939128	-0.21377512
bc	0.08075821	0.02227111	0.07395366	-0.18025790	-0.08321486	-0.01307933	-0.28039465	-0.17677368
idh	0.25221087	-0.20151115	-0.18866402	0.08078583	0.25385943	-0.13411209	0.35882569	0.42710660
esg	0.09449345	0.01612296	0.35003502	0.08387871	0.05367010	0.19394034	0.32161902	0.21827242
setor_sec	-0.12051425	0.02047166	-0.29244818	-0.17590141	0.02059292	-0.31417491	-0.41000011	-0.26478621
	sal	pescs	sup	setor_prim	bc	idh	esg	setor_sec
cres	0.19300183	-0.23165881	0.15859848	-0.24455340	0.08075821	0.25221087	0.09449345	-0.12051425
desemp	-0.07889274	-0.03536080	-0.04272696	0.09054800	0.02227111	-0.20151115	0.01612296	0.02047166
int	-0.23474989	0.45088286	-0.30299131	0.57816522	0.07395366	-0.18866402	0.35003502	-0.29244818
dist	0.25876121	-0.17327445	0.57990336	-0.19497784	-0.18025790	0.08078583	0.08387871	-0.17590141
metro	0.39351427	-0.36980584	0.38811682	-0.31962401	-0.08321486	0.25385943	0.05367010	0.02059292
idd	-0.32448941	0.70740454	-0.41740421	0.46100341	-0.01307933	-0.13411209	0.19394034	-0.31417491
nfin	0.18715588	0.01162924	0.33290021	0.28939128	-0.28039465	0.35882569	0.32161902	-0.41000011
hab	0.32944742	-0.20741733	0.36435744	-0.21377512	-0.17677368	0.42710660	0.21827242	-0.26478621
sal	1.00000000	-0.35735561	0.67332499	-0.33350756	-0.37495153	0.48069439	0.21024383	-0.14339255
pescs	-0.35735561	1.00000000	-0.45758617	0.47493281	0.06676282	-0.25664999	0.14613769	-0.27491353
sup	0.67332499	-0.45758617	1.00000000	-0.46736033	-0.41852799	0.46948608	0.22957347	-0.26926555
setor_prim	-0.33350756	0.47493281	-0.46736033	1.00000000	0.08033003	-0.13319898	0.31276930	-0.38231990
bc	-0.37495153	0.06676282	-0.41852799	0.08033003	1.00000000	-0.24321080	-0.15606932	0.20819878
idh	0.48069439	-0.25664999	0.46948608	-0.13319898	-0.24321080	1.00000000	0.49135356	-0.36890881
esg	0.21024383	0.14613769	0.22957347	0.31276930	-0.15606932	0.49135356	1.00000000	-0.70683255
setor_sec	-0.14339255	-0.27491353	-0.26926555	-0.38231990	0.20819878	-0.36890881	-0.70683255	1.00000000

Anexo 3: Estatísticas descritivas das variáveis utilizadas

cres		desemp		int		dist		metro			
Min.	:-10.151	Min.	: 3.629	Min.	:0.0000	Min.	:0.00000	Min.	:0.0000		
1st Qu.	:-3.108	1st Qu.	: 6.718	1st Qu.	:0.0000	1st Qu.	:0.00000	1st Qu.	:0.0000		
Median	:-1.355	Median	: 7.986	Median	:0.0000	Median	:0.00000	Median	:0.0000		
Mean	:-1.004	Mean	: 8.259	Mean	:0.4532	Mean	:0.06475	Mean	:0.1187		
3rd Qu.	: 0.422	3rd Qu.	: 9.486	3rd Qu.	:1.0000	3rd Qu.	:0.00000	3rd Qu.	:0.0000		
Max.	: 14.143	Max.	:15.471	Max.	:1.0000	Max.	:1.00000	Max.	:1.0000		
idd		nfin		hab		sal		pescs		sup	
Min.	:36.69	Min.	: 6.10	Min.	:0.0000	Min.	:0.6997	Min.	:1.025	Min.	: 2.80
1st Qu.	:41.52	1st Qu.	: 9.10	1st Qu.	:0.7821	1st Qu.	:0.7869	1st Qu.	:2.538	1st Qu.	: 6.60
Median	:44.18	Median	:10.21	Median	:0.8418	Median	:0.8477	Median	:3.364	Median	: 8.30
Mean	:44.33	Mean	:10.40	Mean	:0.9005	Mean	:0.8855	Mean	:3.653	Mean	: 9.34
3rd Qu.	:46.49	3rd Qu.	:11.48	3rd Qu.	:0.9471	3rd Qu.	:0.9368	3rd Qu.	:4.509	3rd Qu.	:11.10
Max.	:55.92	Max.	:18.43	Max.	:1.8806	Max.	:1.7560	Max.	:8.703	Max.	:31.10
setor_prim		bc		idh		esg		setor_sec			
Min.	: 0.100	Min.	:-10739.538	Min.	:0.4760	Min.	:0.2960	Min.	:10.50		
1st Qu.	: 2.525	1st Qu.	: -2.161	1st Qu.	:0.6170	1st Qu.	:0.6940	1st Qu.	:19.80		
Median	: 5.500	Median	: 0.647	Median	:0.7160	Median	:0.8570	Median	:25.30		
Mean	: 7.488	Mean	: -44.842	Mean	:0.6847	Mean	:0.7974	Mean	:27.59		
3rd Qu.	:10.475	3rd Qu.	: 14.553	3rd Qu.	:0.7370	3rd Qu.	:0.9390	3rd Qu.	:32.15		
Max.	:40.800	Max.	: 709.405	Max.	:0.9490	Max.	:0.9900	Max.	:63.00		

Anexo 4: Gráfico do Teste I de Moran à variável cres



Fonte: Elaboração própria

Anexo 5: SLM com a variável dependente espacialmente desfasada como único regressor

Call: `spautolm(formula = cres ~ 1, data = dataset, listw = data_w_queen_nb)`

Residuals:

Min	1Q	Median	3Q	Max
-9.25314	-1.95734	-0.28654	1.55042	14.77182

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.95061	0.35932	-2.6456	0.008155

Lambda: 0.48489 LR test value: 46.325 p-value: 1.0019e-11

Numerical Hessian standard error of lambda: 0.064077

Log likelihood: -715.0351

ML residual variance (sigma squared): 9.5238, (sigma: 3.0861)

Number of observations: 278

Number of parameters estimated: 3

AIC: 1436.1

Anexo 6: Teste I de Moran aos resíduos do modelo SLX restrito

Teste I de Moran aos resíduos do modelo SLX restrito	
Estatística	1,649
Valor-p	0,0992

Anexo 7: SEM não restrito para o crescimento populacional com erros padrão robustos

```
Call:spatialreg::GMerrorsar(formula = formula, data = data, listw = listw,
  na.action = na.action, zero.policy = zero.policy, method = method,
  arnoldwied = arnoldwied, control = control, pars = pars,
  scaleU = scaleU, verbose = verbose, legacy = legacy, se.lambda = se.lambda,
  returnHcov = returnHcov, pWOrder = pWOrder, tol.Hcov = tol.Hcov)
```

```
Residuals:
  Min      1Q  Median      3Q      Max
-6.56072 -1.54157 -0.20819  1.42747 14.19592
```

```
Type: GM SAR estimator
Coefficients: (GM standard errors)
      Estimate Std. Error z value Pr(>|z|)
(Intercept) 28.06190218  4.58275341  6.1234 9.162e-10
desemp      -0.73832544  0.08748071 -8.4399 < 2.2e-16
int         -0.18758170  0.48792916 -0.3844  0.70065
dist        -4.01229341  0.87316103 -4.5951 4.325e-06
metro       -0.35405777  0.64535724 -0.5486  0.58326
idd         -0.35116083  0.07547399 -4.6527 3.276e-06
nfin        -0.23678426  0.11970876 -1.9780  0.04793
hab          0.15214217  0.98664512  0.1542  0.87745
sal          0.71612418  1.43601695  0.4987  0.61800
pescs       -0.17286215  0.16552172 -1.0443  0.29632
sup          0.00169506  0.08828956  0.0192  0.98468
setor_prim  -0.09465587  0.04247871 -2.2283  0.02586
bc           0.00050529  0.00025465  1.9842  0.04723
idh         -1.57556186  2.25572864 -0.6985  0.48488
esg          1.09498080  1.60965483  0.6803  0.49634
setor_sec   -0.13650518  0.02783904 -4.9034 9.421e-07
```

```
Lambda: 0.15842 (standard error): 0.23712 (z-value): 0.6681
Residual variance (sigma squared): 6.4416, (sigma: 2.538)
GM argmin sigma squared: 6.4341
Number of observations: 278
Number of parameters estimated: 18
```

Anexo 8: SLM não restrito para o crescimento populacional com erros padrão robustos

```
Call:stsls(formula = cres ~ desemp + int + dist + metro + idd + nfin +
  hab + sal + pescs + sup + setor_prim + bc + idh + esg + setor_sec,
  data = dataset, listw = data_w_queen_nb, robust = TRUE)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-6.24451 -1.60524 -0.27631  1.44554 14.43976
```

```
Coefficients:
      Estimate HCO std. Error z value Pr(>|z|)
Rho          0.37523212  0.14154702  2.6509  0.008027
(Intercept) 25.52315674  5.00791507  5.0966  3.459e-07
desemp      -0.61717982  0.09262551 -6.6632  2.680e-11
int          0.26294434  0.44288727  0.5937  0.552710
dist        -4.13377651  0.79348161 -5.2097  1.892e-07
metro       -0.33178366  0.66668065 -0.4977  0.618720
idd         -0.31903220  0.09164119 -3.4813  0.000499
nfin        -0.23192171  0.10122422 -2.2912  0.021954
hab         -0.10110629  0.99868204 -0.1012  0.919360
sal          0.00987181  1.87508753  0.0053  0.995799
pescs       -0.11232782  0.15748132 -0.7133  0.475674
sup          0.04674415  0.09969602  0.4689  0.639165
setor_prim  -0.06681544  0.03963816 -1.6856  0.091866
bc           0.00054331  0.00019940  2.7247  0.006436
idh         -1.51002960  2.33980634 -0.6454  0.518691
esg          0.48337770  1.38287285  0.3495  0.726679
setor_sec   -0.11090118  0.02678537 -4.1404  3.468e-05
```

Residual variance (sigma squared): 6.7929, (sigma: 2.6063)

Anexo 9: SARAR não restrito do crescimento populacional com erros padrão robustos

Generalized stsls

```
Call:
gstslshet(formula = cres ~ desemp + int + dist + metro + idd +
  nfin + hab + sal + pescs + sup + setor_prim + bc + idh +
  esg + setor_sec, data = dataset, listw = data_w_queen_nb)
```

```
Residuals:
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-6.1117 -1.5568 -0.2074  0.0009  1.3856 14.4219
```

```
Coefficients:
      Estimate Std. Error t-value Pr(>|t|)
(Intercept) 22.95602338  4.58249655  5.0095 5.457e-07 ***
desemp      -0.54662916  0.08482205 -6.4444 1.160e-10 ***
int          0.28931485  0.37780287  0.7658 0.4438057
dist        -4.05477538  0.77248772 -5.2490 1.529e-07 ***
metro       -0.35141517  0.57068649 -0.6158 0.5380422
idd         -0.28484357  0.08251884 -3.4519 0.0005567 ***
nfin        -0.21133702  0.09010928 -2.3453 0.0190097 *
hab         -0.20992566  0.97117537 -0.2162 0.8288659
sal          0.31455437  1.87499265  0.1678 0.8667697
pescs       -0.09866066  0.14887642 -0.6627 0.5075216
sup          0.03907604  0.09389349  0.4162 0.6772826
setor_prim  -0.06597728  0.03635679 -1.8147 0.0695675 .
bc           0.00060431  0.00018088  3.3410 0.0008348 ***
idh         -0.95535346  2.10037346 -0.4548 0.6492176
esg          0.10090670  1.22209139  0.0826 0.9341944
setor_sec   -0.10691175  0.02463448 -4.3399 1.425e-05 ***
lambda      0.46893861  0.12850156  3.6493 0.0002630 ***
rho         -0.40248286  0.19525833 -2.0613 0.0392760 *
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Wald test that rho and lambda are both zero:
Statistics: 0.23646 p-val: 0.62678

Anexo 10: Teste de Wald à significância conjunta dos coeficientes de rho e lambda do modelo SARAR

Teste de Wald à significância conjunta dos coeficientes de rho e lambda do modelo (8)	
Estatística	0,3957
Valor-p	0,5293

Anexo 11: Regressão Geograficamente Ponderada do crescimento populacional devido a migração com as variáveis do modelo (6)

```
Call:
gwr(formula = cres ~ desemp + dist + idd + nfin + setor_prim +
     bc + setor_sec, data = data_poly, bandwidth = bwg, gweight = gwr.Gauss,
     hatmatrix = TRUE)
Kernel function: gwr.Gauss
Fixed bandwidth: 44084.14
Summary of GWR coefficient estimates at data points:
      Min.      1st Qu.      Median      3rd Qu.      Max.  Global
X.Intercept. -0.73202745  9.54925080  27.90432707  29.85412575  39.69024180  30.0181
desemp       -0.69768402 -0.41651300 -0.38845982 -0.33989244 -0.15275487 -0.7537
dist1       -7.45790687 -6.84186035 -6.46883139 -1.02376125  0.47119614 -3.9914
idd         -0.76913302 -0.68378543 -0.66660944 -0.02288155  0.04814088 -0.3890
nfin        -0.58618950 -0.03226584  0.22077403  0.31380925  0.65076976 -0.2269
setor_prim  -0.19602529 -0.01941820  0.01229686  0.07391034  0.14590971 -0.1110
bc          -0.00331889  0.00045267  0.00257259  0.00353746  0.01706924  0.0005
setor_sec   0.04224524  0.30754477  0.40072301  0.50797335  0.57640399 -0.1478
Number of data points: 278
Effective number of parameters (residual: 2*traceS - traceS'S): 78.54465
Effective degrees of freedom (residual: 2*traceS - traceS'S): 199.4554
Sigma (residual: 2*traceS - traceS'S): 2.326656
Effective number of parameters (model: traceS): 59.86861
Effective degrees of freedom (model: traceS): 218.1314
Sigma (model: traceS): 2.224825
Sigma (ML): 1.970755
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 1322.714
AIC (GWR p. 96, eq. 4.22): 1225.998
Residual sum of squares: 1079.717
Quasi-global R2: 0.6724332
```