ÍNDICE

ÍNDICE DE TABELAS

LISTA DE ABREVIATURAS

SE – erro padrão Standard error

GWR- Geographically Weighted Regression

OLS- Ordinary Least Squares

AIC- Critério de Informação de Akaike

AIC$_c$- Critério de Informação de Akaike corrigido

Resumo

O presente trabalho aborda a temática dos estudos sobre a análise espacial dos resultados eleitorais em Portugal.

Os resultados eleitorais quer ao nível das eleições legislativas quer das autárquicas exibem tendências geográficas onde a influência de vários factores parece ser evidente.

A abordagem analítica dos resultados eleitorais de um país sugere que se tenham em conta variados efeitos contextuais, nomeadamente, as diferenças etárias nas diversas regiões, diferenças sócio-culturais da população, assim como o isolamento das populações.

Torna-se, deste modo, importante ter em consideração as variações espaciais na modelação de fenómenos de cariz eleitoral.

Este estudo questiona a existência de variações espaciais na relação entre os resultados eleitorais (eleições legislativas referentes ao ano de 2005) dos cinco principais partidos políticos Portugueses, (PSD, CDS, BE,PCP,PS) centrando-se nas diferenças entre Municípios de Portugal Continental e apresenta, como principais objectivos:

1. Demonstrar a importância da utilização de técnicas estatísticas espaciais (Modelos Geographically Weighted Regression) como melhoria dos Modelos de Regressão tradicionais (OLS) na modelação de fenómenos baseados em resultados eleitorais.
2. Identificar tendências espaciais partidárias, ao nível dos municípios de Portugal Continental, relacionando, para cada partido, os resultados num determinado município e o total de votantes nesse mesmo município.
3. Reconhecer a importância da variação etária, do número de votos em branco e da constituição de cada município (relativamente ao número de freguesias que o compõem) como determinantes dos resultados eleitorais dos diferentes partidos políticos nas zonas geográficas em estudo.

1- Introdução

Em variados processos sociais observam-se relações que variam de acordo com a sua representação geográfica. São múltiplas as razões para esta variação, nomeadamente o uso de diferentes amostras de dados e problemas com a especificação do modelo adoptado, uma vez que no processo de modelação omitem-se muitas vezes variáveis relevantes.

Algumas relações são intrinsecamente diferentes ao longo do espaço geográfico, os chamados efeitos contextuais.

Quando os valores registados por uma determinada variável reflectem uma dimensão espacial, o uso de técnicas estatísticas espaciais que especifiquem correctamente essa dimensão é aconselhável, na medida em que pode dever-se a “efeitos de contágio”, isto é, a uma dependência espacial sistemática entre fenómenos geograficamente vizinhas.

Esta realidade estatística não pode ser ignorada aquando da análise empírica de um ciclo eleitoral, baseado naquela variável de decisão, ao nível daquelas unidades geográficas.

Deste modo, essa dependência relaciona-se com a existência de autocorrelação espacial, isto é, “clusters” espaciais de valores semelhantes para a variável a ser explicada. Neste contexto tal acontecerá quando, por exemplo, os resultados eleitorais obtidos pelos diversos partidos nos diferentes locais dependam dos valores daqueles resultados em localizações vizinhas. Como exemplo podemos pensar na proximidade geográfica enquanto elemento permissivo da ocorrência de um fenómeno associado à votação por comparação.

Tendo em conta a realidade portuguesa, não se pode excluir que uma explicação parcial dos resultados eleitorais em cada município seja resultante da comparação que os eleitores fazem do desempenho do seu município face a municípios vizinhos (sobretudo quando estão sob a tutela de um partido diferente). Este facto alerta-nos para a importância do número de freguesias que compõem cada município, uma vez que, quanto menor for esse número, maior homogeneidade se observa e consequentemente menor será esse efeito de comparação.

Por outro lado, devemos ter em conta a heterogeneidade espacial, isto é, um padrão espacial nos dados que resulta em instabilidade dos parâmetros nas relações entre as variáveis ao longo do espaço em estudo. Neste caso a relação entre a variável explicada e as variáveis explicativas é "localização específica", dado que pode variar significativamente ao longo do espaço.

A abordagem analítica dos resultados eleitorais de um país sugere que se tenham em conta variados efeitos contextuais, nomeadamente, as diferenças etárias nas diversas regiões, diferenças sócio-culturais da população, assim como o isolamento das populações.

Em suma, a dependência espacial pode assumir duas formas: dependência espacial na variável explicada e dependência espacial nos erros/resíduos. A primeira pode ser explicada como equivalente ao contágio espacial, enquanto que a segunda conduz a uma forma de modelo espacial de erros.

Neste contexto surge a metodologia Geographical Weighted Regression (GWR), como alternativa mais abrangente para a modelação espacial.

Torna-se importante ter em consideração as variações espaciais na modelação de fenómenos de cariz eleitoral, justificando-se a aplicação dos modelos GWR quando:

1. A capacidade explicativa da regressão GWR for significativamente maior que a de uma regressão OLS;

2. Se considere como melhor modelo o modelo GWR, apoiando-nos nos resultados do Critério de Informação Akaike;

3. A autocorrelação espacial, detectada ao nível de “clusters” espaciais, seja confirmada através de testes estatísticos de correlação espaciais, nomeadamente a estatística-T;

4. A variabilidade espacial das variáveis explicativas seja confirmada pelo teste de significância de Monte Carlo.

No entanto, a questão central continua em aberto: ao observarmos variações espaciais em relações, estas devem-se simplesmente a uma má especificação do modelo ou estão relacionadas com comportamentos locais intrinsecamente diferentes?

O interesse por este tipo de modelação não é recente. Johnston (1973) apresentou um exemplo de análise local num contexto de comportamento eleitoral. Contudo, como observa Fotheringham (1977), o aumento do interesse em formas locais torna emergente uma série de técnicas para a modelação local.

Jones e Hanham (1995) advertem contra as generalizações das abordagens quantitativas que revelam pouco interesse na identificação de excepções locais.

As formas locais de análise espacial fornecem uma ponte entre os outputs das técnicas espaciais e o poder das capacidades visuais dos softwares estatísticos gráficos. Talvez mais importante ainda, embora forneçam mais informação das relações espaciais, como ajuda ao desenvolvimento do modelo e a um melhor conhecimento dos processos espaciais.

As estatísticas e os modelos locais fornecem-nos o equivalente a um microscópio ou telescópio; são ferramentas com as quais podemos ver com maior detalhe e acuidade. Sem estas, o cenário apresentado pelas estatísticas globais é de uniformidade e falta de variações espaciais; com elas é possível observar padrões espaciais de relações que são frequentemente mascarados.

O GWR baseia-se no quadro da regressão tradicional, que nos é familiar e incorpora relações locais espaciais no quadro de regressão de modo intuitivo e explícito.

Permite-nos, então, detectar tendências geográficas dos dados em estudo, de um modo relativamente simples.

É de notar, no entanto, que nem todos os modelos justificam a aplicação desta metodologia, o que salienta a importância dos testes diagnóstico.

No presente trabalho focamos a nossa atenção numa técnica local geral, desenvolvida para dados espaciais denominada por GWR aplicada ao estudo dos resultados eleitorais de Portugal continental no ano de 2005.

2 – Metodologia Estatística Espacial: Geographically Weighted Regression

2.1-Geographically Weighted Regression

Consideremos um modelo de regressão global, em que os parâmetros são considerados invariantes no espaço:

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \qquad \textbf{(1)}$$

A extensão do modelo de regressão tradicional ao GWR é dada por :

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \qquad \textbf{(2)}$$

onde (u_i,v_i) são as coordenadas do i-ésimo ponto no espaço e $\beta_k(u_i, v_i)$ é a realização da função contínua $\beta_k(u\ ,v\ )$ no ponto i. (a existência desta função contínua permite a generalização da estimação dos parâmetros).

Assume-se que os coeficientes são funções determinantes de outras variáveis, neste caso localizações no espaço (não se assume que sejam aleatórios).

Neste sentido, o *enviesamento* é resultante da localização espacial, isto é, o output de um processo não estacionário na localização i, através de dados recolhidos noutras localizações diferentes de i.

O processo de calibração do GWR deve ser uma solução de “compromisso” entre enviesamentos locais e erros globais.

Ao estimar um parâmetro com determinada localização i, aproxima-se o modelo(2) na região i por um modelo (1) e aplica-se uma regressão usando um subconjunto de pontos próximos de i.

Cada uma das i estimações acarreta enviesamentos, pois os coeficientes $\beta_0(u_i, v_i)$ variam ao longo do subconjunto da calibração local.

Se a amostra local for suficientemente grande, reduzem-se os erros gerais de estimação dos coeficientes. Por outro lado, quanto maior for a amostra maiores são as possíveis variações dos coeficientes o que induz a um maior enviesamento.

Por isso deve-se fazer um ajustamento final ao modelo, assumindo que os pontos do subconjunto calibrado que se situem a maior distância de i tenham maior probabilidade de ter coeficientes diferentes, isto é, assume-se implicitamente que os dados mais próximos da localização i tenham maior influência na estimação da função $\beta_k(u_i, v_i)$.

A estimativa dos parâmetros é dada por:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \qquad \textbf{(3)}$$

onde $\hat{\beta}, X, W$ e y são matrizes, sendo W uma matriz quadrada nxn com todos os elementos nulos excepto os da diagonal principal. Os elementos da diagonal principal w_{ii}, representam os pesos geográficos a atribuir à localização i.

Sendo o modelo de regressão clássica, sob forma matricial, dado por $y = X\beta + \varepsilon$, com β vector dos parâmetros (invariante no espaço), o modelo GWR equivalente é $y = (\beta \otimes X)1 + \varepsilon$ **(4)**

em que $\otimes$ representa a multiplicação matricial, β é uma matriz nx(k+1) e 1 é o vector unitário (k+1)x1.

Portanto, num modelo com k+1 variáveis explicativas, as estimativas dos parâmetros são dadas pela seguinte matriz:

$$\beta = \begin{bmatrix} \beta_0(u_1,v_1)\beta_1(u_1,v_1)................\beta_k(u_1,v_1) \\ \beta_0(u_2,v_2)\beta_1(u_2,v_2)................\beta_k(u_2,v_2) \\ .... \\ \beta_0(u_n,v_n)\beta_1(u_n,v_n)................\beta_k(u_n,v_n) \end{bmatrix} \quad \textbf{(5)}$$

Os parâmetros de cada linha (i=1,…..n) são estimados por:

$\hat{\beta}(i) = (X^T W(i) X)^{-1} X^T W(i) y$, **(6)**

sendo W a matriz referida em (3).

No que se refere aos erros gerais, consideremos C_y os parâmetros estimados $\hat{\beta}(u_i,v_i)$.

Então, $C_y = \hat{\beta}(u_i,v_i)$ e, consequentemente, $C = (X^T W(u_i,v_i) X)^{-1} X^T W(u_i,v_i)$.

Deste modo a variância para os C_y é dada por $\mathrm{var}(\hat{\beta}(u_i,v_i) = CC^T\sigma^2$, onde σ^2 representa o resíduo normalizado da soma dos quadrados da regressão local.

Logo, $\sigma^2 = \sum_i \frac{y_i - \hat{y}_i}{n - 2v_1 v_2}$,**(7)**

onde $y_i - \hat{y}_i$ é o desvio do valor real ao valor estimado, S é a matriz que caracteriza o desvio de modo que $\hat{y} = Sy$, v_1 é o traço da matriz S, v_2 é o traço da matriz S^TS.

Cada linha i de S é dada por: $r_i = X_i (X^T W(u_i,v_i) X)^{-1} X^T W(u_i,v_i)$.**(8)**

O número de graus de liberdade do resíduo é dado por n-2v_1+v_2, e o número efectivo de parâmetros do modelo GWR local é 2v_1-v_2.

Como Tr(S)$\approx$Tr(S^TS), podemos aproximar para v_1 o número efectivo de parâmetros na regressão local.

Concluindo obtemos a equação para os erros gerais SE ($\hat{\beta}_i$)=$\sqrt{\mathrm{var}(\hat{\beta}_i)}$ **(9)**, (sendo SE a abreviatura para desvio-padrão).

2.2-Testes estatísticos nos modelos GWR

De modo semelhante aos modelos de estimação OLS, o modelo GWR sofre os efeitos de outliers. Mais precisamente, os problemas de não-estacionaridade espacial e dependência espacial desafiam a análise de dados espacial (Fotheringham *et al.*2002).

A não-estacionaridade espacial refere-se a relações que variam no espaço e não se detecta nas estatísticas globais. Por exemplo, uma variável explicativa pode ser bastante relevante numa região e ser irrelevante noutra zona.

Se não considerarmos a não-estacionaridade espacial o modelo em estudo pode produzir enviesamentos nos resultados locais.

As variações espaciais podem simplesmente dever-se a incorrectas especificações do modelo, por exemplo nos casos em que variáveis importantes, do ponto de vista de variações espaciais, não são incluídas no modelo. Por conseguinte, torna-se importante identificar as verdadeiras relações.

A dependência espacial ou associação espacial refere-se ao facto de que o valor de um determinado atributo numa localização depende dos valores do mesmo atributo em localizações vizinhas. A dependência espacial é frequentemente medida pelas estatísticas de autocorrelação espacial, que descrevem as semelhanças de observações vizinhas.

A autocorrelação positiva significa que áreas adjacentes partilham os mesmos atributos e a autocorrelação negativa indica que áreas adjacentes exibem os atributos opostos. Na prática, a dependência espacial é muitas vezes ignorada, ou descrita simplesmente por uma medida estatística global para o total de graus de dependência.

Uma vez que pode existir, nos dados, autocorrelação espacial positiva ou negativa, uma estatística global falha na identificação dos diferentes graus de dependência espacial dos dados.

Acresce que a validação do modelo pode não ser possível, uma vez que as estimativas dos parâmetros não são eficientes e os testes de significância geram resultados pouco fiáveis na presença de autocorrelação espacial.

Logo, é importante abordar e ter sempre em atenção os problemas de existência de outliers, problemas de heterogeneidade e autocorrelação espacial.

2.2.1 Outliers

Numa amostra aleatória, chamamos outlier às observações que admitem um valor de distância discrepante (muito maior ou muito menor) relativamente aos valores das outras observações, isto é parecem muito diferentes dos valores das outras observações.

Os outliers podem enviesar os modelos de regressão, os quais supostamente descrevem o comportamento “normal” de um sistema.

Da equação (3), o vector dos resíduos $e_i\,'s$ na notação vectorial é dado por:

$$e = y - Sy = (I-S)y \quad \textbf{(10)}$$

Deste modo, sendo $Q=(I-S)(I-S)^T$, vem que:

$$\text{var(e)} = \text{(I-S)(I-S)}^{T}\sigma^{2} = \text{Q}\sigma^{2} \quad \textbf{(11)}$$

Os resíduos internos studentizados, como a seguir se escrevem na equação (12), podem ser usados como uma ferramenta diagnóstica útil na detecção de outliers.

$\mathrm{r_i} = \dfrac{e_i}{\hat{\sigma}\sqrt{q_{ii}}}$ **(12)**, onde q_{ii} é o i-ésimo elemento da diagonal de Q.

Alternativamente, os resíduos studentizados externos podem ser usados para detectar outliers, eliminando o efeito dos outliers para o cálculo de σ^2.

Os resíduos studentizados externos definem-se do seguinte modo:

$\mathrm{r_i^*} = \dfrac{e_i}{\hat{\sigma}_{-i}\sqrt{q_{ii}}}$ **(13)**, onde $\hat{\sigma}^2_{-i}$ é a estimativa de σ^2 quando se exclui a i-ésima observação.

Devem-se considerar como outliers as observações em que $|\mathrm{r_i^*}| > 3$ (Fotheringham *et al.* 2002).

Uma vez os outliers identificados e excluidos do conjunto de dados, o modelo GWR deve ser recalibrado e os resultados devem ser comparados com os conjuntos de dados antes e depois da exclusão dos outliers para assegurar que os dados excluídos eram mesmo outliers.

2.2.2.Heterogeneidade Espacial

A heterogeneidade espacial existe quando uma relação não é estacionária no espaço.

Neste caso, uma estatística global não é suficiente para representar a relação que varia ao longo da área em estudo.

A abordagem GWR tem sido utilizada para resolver problemas vulgarmente associados com severa não-estacionaridade espacial e autocorrelação.

Por exemplo, Calvo e Escolar (2003) aplicaram o GWR para obter estimativas locais não enviesadas e consistentes para os dois mais populares métodos de inferência ecológica a partir de dados que exibiam extrema heterogeneidade espacial.

No modelo básico GWR, assume-se que a variância do termo de erro, ε_i segue uma distribuição normal com média zero e variância σ^2.

Esta hipótese pode ser flexível, de modo a permitir que o termo de erro, em vez de ser considerado constante, seja uma função contínua das localizações das observações ao longo do espaço, isto é, $N(0,\sigma^2,(\mu_i,v_i))$. É o que se entende por heterocedasticidade, uma vez que a variância do termo de erro varia de localização em localização.

Uma abordagem prática para calibrar um modelo GWR composto por erros com heterocedasticidade é primeiro calibrar um modelo GWR standard como uma estimação experimental, adaptando-se o modelo a um esquema de reponderação em que as ponderações sejam determinadas pelos resíduos, isto é, $\frac{1}{e_i^2}$. Este procedimento pode ser repetido várias vezes, resultando, em cada repetição, um novo conjunto de resíduos que implica um novo conjunto de estimativas de e_i^2. Este processo iterativo é, de um modo geral, na prática, convergente (Fotheringham *et al.* 2002).

Paéz *et al.*(2002) apresentam outro ponto de vista na abordagem da heterogeneidade espacial. Eles definem um termo de erro mais geral tal que:

$\varepsilon \sim N(0,\Omega)$ **(14),** onde Ω é a matriz nxn das covariâncias, constituída pelos seguintes elementos:

$w_{ii} = \sigma^2 g_i(\gamma, z_i)$ e $w_{ii} = 0, \forall i, j \; i \neq j$ **(14)**

Por outras palavras, assume-se que a variância associada à observação i seja uma função de uma variável vectorial conhecida, z_i ,e do vector parâmetro γ . Assume-se que a função g seja uma função exponencial

$g_i = \exp(\gamma_0 d_{0i}^2)$, **(15)**

onde d_{0i} é a distância entre um ponto focal *o* e a localização i, para cada amostra gravada. O modelo GWR básico é um caso particular da equação (15), onde γ_0 é constante em cada ponto do espaço.

Páez *et al* (2002) adoptou a abordagem da log-probabilidade para estimar um modelo GWR com termos de erro hetegonéneos. A correspondente log-probabilidade, isto é, L, é dada por:

$$L = -\frac{n}{2}\ln\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\ln|G| - \frac{1}{2\sigma^2}(Y - XB)^T G^{-1}(Y - XB) \quad \textbf{(16)}$$

Onde G representa a matriz com os elementos diagonais w_{ii} da equação (14).

A seguinte equação obtém-se ajustando as equações resultantes da igualdade a zero das derivadas parciais da log-probabilidade, relativamente aos parâmetros β, σ e γ_0:

$$L = -\frac{n}{2}\ln[\frac{1}{n}(Y - X\hat{B})^T G^{-1}(Y - X\hat{B})] - \frac{1}{2}\sum_{i=1}^{n}\gamma_0 d_{0i}^2 \quad \textbf{(17)}$$

Maximiza-se a equação (17) de modo a obter a localização específica dos parâmetros. A estimativa de máxima verosimilhança para $\hat{\beta}$ e $\hat{\sigma}$ pode ser então calculada.

Testar a não-estacionaridade espacial paramétrica é equivalente a testar se γ_0 é significativamente diferente de zero. O teste estatístico é dado por Páez *et al* (2002) como:

$$\xi = s_r^2 F_{rr}^{-1} \quad \textbf{(18)}$$

Onde:

$$s_r = \frac{1}{2}\left[\frac{1}{\sigma_{OLS}^2}(Y - X\hat{B}_{OLS})^T D(Y - X\hat{B}_{OLS}) - trD\right]$$

$$F_{rr} = 2\left[\wedge trD^2 - \frac{1}{n}(trD)^2\right]^{-1} \qquad \text{e} \qquad D = \begin{bmatrix} d_{o1}^2 & 0 & 0 \\ 0 & .... & 0 \\ 0 & 0 & d_{on}^2 \end{bmatrix}$$

Sob a hipótese nula, isto é γ_0=0, a estatística teste para a não estacionaridade espacial segue uma distribuição χ^2 com um grau de liberdade.

Os parâmetros $\hat{\beta}$ e $\hat{\sigma}$ são as estimativas dos mínimos quadrados globais indexadas a uma regressão OLS. Para cada conjunto de n pontos devem-se efectuar n testes para detectar a não-estacionaridade paramétrica espacial.

De modo a implementar todos os testes de não-estacionaridade simultaneamente, deve-se ter em conta as seguintes etapas:

Etapa 1: Determinar a probabilidade de cometer um erro do tipo I para cada ponto de dados, isto é determinar o p-value (p_i):

$$p_i = 1 - F(\zeta_i \mid v) \quad \textbf{(19)}$$

Onde F é a distibuição cumulativa χ^2 avaliada em ζ_1 com v (=1) graus de liberdade.

Etapa 2: Determinar o valor crítico l_i para cada hipótese:

$$l_i = \frac{\alpha}{n-i+1} \quad \textbf{(20)}$$

Em que é o nível de significância nominal, usualmente com valor 0,05.

Etapa 3: Considerar $p^{(1)} \leq ... \leq p^{(n)}$ os p-values ordenados e $H^{(1)} \leq ... \leq H^{(n)}$ as correspondentes hipóteses.

Etapa 4: Iniciar com o menor p-value. Se $p^{(1)} > l_1$ pára-se o teste e conclui-se que nenhuma das hipóteses pode ser rejeitada. Caso contrário rejeita-se $H^{(1)}$ e passa-se à seguinte etapa.

Etapa 5: Se o teste continuar até a etapa i e se $p^{(i)} > l_i$ pára-se o teste e não se rejeita nenhuma das restantes hipóteses. Caso contrário rejeita-se $H^{(i)}$ e testa-se a seguinte hipótese.

2.2.3 Teste para não estacionaridade espacial

A abordagem GWR assume que o efeito de uma dada variável relativamente à variável dependente varia geograficamente.

Este efeito de não-estacionaridade deve ser verificado de modo a justificar a aplicação de um modelo GWR. O método para testar a não-estacionaridade das estimativas dos parâmetros num modelo GWR envolve o cálculo da variância dos coeficientes para a k-ésima variável a partir de um modelo GWR, que se escreve como V_k (Fotheringham *et al.*2002):

$$V_k = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta} - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{ik})^2 \quad \textbf{(21)}$$

Intuitivamente, V_k deveria ser significativamente diferente de zero, de modo a rejeitar-se a hipótese de que o k-ésimo parâmetro é globalmente fixo. No entanto, uma vez que as distribuições teóricas subjacentes às estimativas locais dos parâmetros são desconhecidas, a variabilidade de V_k dificulta a verificação analítica.

Pode-se utilizar como alternativa o teste de significância de Monte Carlo, que recorre a simulação, e é indicado para verificar a não-estacionaridade espacial. Neste teste, a variância observada das estimativas locais dos parâmetros do conjunto de dados original é inicialmente calculada e armazenada. Seguem-se um dado número de processos aleatórios para obter variâncias simuladas para cada variável. Estas variâncias simuladas são posteriormente escolhidas para constituir o rank das variâncias observadas. O correspondente p-value determina-se subtraindo o rácio do rank sobre o número total de variâncias da unidade (Fotheringham *et al.*2002).

2.2.4 Autocorrelação Espacial

Estamos perante uma autocorrelação espacial quando o valor de uma determinada variável, com determinada localização geográfica, depende do valor dessa mesma variável em localizações vizinhas. Ora os termos de erro num modelo GWR devem ser identicamente e independentemente distribuídos. Se o termo de erro estiver autocorrelacionado, as hipóteses subjacentes aos testes estatísticos gerais são violadas. Discute-se, assim, de modo breve, a independência dos termos de erro.

As hipóteses teste para a autocorrelação espacial no termos de erro $\varepsilon_1,...,\varepsilon_n$ de um modelo GWR podem ser formuladas do seguinte modo:

H_0: Não existe autocorrelação espacial no ruído, isto é, $\mathrm{var}(\varepsilon) = \mathrm{var}(\varepsilon\varepsilon^T) = \sigma^2 I$, onde $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^T$ é o vector que causa o ruído.

H_1: Existe autocorrelação espacioal positiva ou negativa nos ruídos, relativamente à matriz de ponderação espacial W.

A estatística teste para detectar a autocorrelação é o teste de Moran I_0 .Estas estatística é frequentemente usada para quantificar o grau de autocorrelação espacial. Os valores I do teste de Moran variam entre 1 e -1, onde 1 indica um alto grau de autocorrelação espacial positiva e -1 indica um alto grau de autocorrelação espacial negativa. O valor esperado de I_0, sob a hipótese de não existência de autocorrelação é aproximadamente zero.

Teste de Moran I_0

$$I_0 = \frac{\hat{\varepsilon}^T N^T W N \hat{\varepsilon}}{\hat{\varepsilon}^T N^T N \hat{\varepsilon}} \quad \textbf{(22)}$$

Onde W representa uma matriz específica simétrica das ponderações espaciais em ordem a n e a matriz N é dada por:

$$N = I - L = I - \begin{bmatrix} \mathrm{x}_1^{\mathrm{T}}[\mathrm{X}^{\mathrm{T}} W(1) \mathrm{X}]^{-1} X^T W(1) \\ \mathrm{x}_2^{\mathrm{T}}[\mathrm{X}^{\mathrm{T}} W(2) \mathrm{X}]^{-1} X^T W(2) \\ .... \\ \mathrm{x}_n^{\mathrm{T}}[\mathrm{X}^{\mathrm{T}} W(n) \mathrm{X}]^{-1} X^T W(n) \end{bmatrix}$$

Sob a hipótese nula, a probabilidade de I_0 ser menor que um dado valor r calcula-se do seguinte modo:

$$P(I_0 \le r) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\sin(\theta(t))}{t\rho(t)} dt \quad \textbf{(23)}$$

Onde:

$$\theta(t) = \frac{1}{2}\sum_{i=1}^{n} \arctan(\lambda_i t)$$

$$\rho(t) = \prod_{i=1}^{n} (1+\lambda_i^2 t^2)^{\frac{1}{4}}$$

e λ_i são os valores próprios de $N^T(W - rI)N$.

2.3 -Escolha da função espacial ponderada

Numa regressão OLS tradicional, o esquema ponderado implícito consiste em considerar, para os elementos w_{ij} iguais a 1, sendo j o ponto específico no espaço, onde se situa a observação e i o ponto, no espaço, para o qual o parâmetro é estimado, isto é, no modelo global cada observação tem o peso unitário.

O primeiro passo para a abordagem da localização consiste em considerar uma função de ponderação definida por:

$$w_{i,j} = \begin{cases} 1 \; se \; \mathrm{d}_{ij} < d \\ 0 \text{ caso contrário} \end{cases}$$

Esta função exclui da calibração do modelo observações que se situam a uma distância (do ponto de regressão) maior que uma determinada distância d, no entanto, por se tratar de uma função descontínua acarreta problemas.

De modo a combater o problema desta descontinuidade, especifica-se w_{ij} como uma função contínua da distância entre i e j, como por exemplo $w_{ij} = \exp[-\frac{1}{2}(\frac{d_{ij}}{b})^2]$, sendo b a largura da faixa (banda) considerada.

Se i=j, $d_{ij} = 0$, o que implica que $w_{ij} = \exp(0) = 1$, o que significa que a ponderação dos outros dados decresce como a curva gaussiana, enquanto a distância entre i e j aumenta.

Podemos adoptar uma função alternativa

$$w_{i,j} = \begin{cases} [1-(\frac{d_{ij}}{b})^2]^2 \ \textit{se}\ \mathrm{d}_{ij} < d \\ 0 \text{ caso contrário} \end{cases}$$

Esta função alternativa é particularmente útil porque fornece uma função contínua próxima da Gaussiana.

No entanto o uso de Kernels apresenta fragilidades, devido à variação espacial dos núcleos, dependentes da densidade das regiões.

Como forma de minimizar o problema pode-se aplicar o GWR com núcleos espaciais variáveis, existindo para o efeito três métodos.

O primeiro método consiste em dispor, em rank, os pontos de acordo com a sua distância a cada ponto i, de modo que , R_{ij} é a posição (no rank) do j-ésimo ponto a partir de i, em termos da distância de j para i. O ponto mais próximo de i tem o peso igual a 1 e o peso decresce à medida que a posição, no rank, aumenta. Por exemplo $\mathrm{w}_{\mathrm{ij}} = \exp(-\frac{\mathrm{R}_{\mathrm{ij}}}{b})$.

Este método reduz automaticamente a faixa de núcleos localizados em regiões com grande quantidade de dados, pois a distância, por exemplo, ao décimo ponto mais próximo do ponto de regressão será menor do que nos casos em que o ponto de regressão se situa numa região com poucas observações.

2.4-Testes diagnóstico

Os testes diagnósticos permitem-nos decidir sobre o processo de selecção do modelo, na medida em que nos permitem verificar se uma tendência observada é devida à variação aleatória do modelo global ou reflecte uma verdadeira tendência geográfica no modelo.

Esta questão é uma questão da inferência estatística e podem-se adoptar duas abordagens: a abordagem clássica, baseada em intervalos de confiança e uma abordagem baseada no critério de Informação Akaike (AIC) para selecção do modelo.

A inferência estatística preocupa-se com o processo de inferir a partir da análise de conjuntos de dados estatísticos.

O processo de inferência estatística assenta na resposta a uma das três questões seguintes: existe algum facto verdadeiro na base dos dados?, em que intervalo de confiança pertencem os coeficientes do modelo?, que modelo matemático é o melhor?

Existe sempre algum grau de aleatoriedade na variação da colecção de dados, no entanto para decidir se a tendência observada é uma tendência geográfica, devemos responder a, pelo menos, uma das três questões anteriores.

A primeira questão aborda a inferência denominada por teste de hipóteses ou teste de significância. Trata-se de questionar quão similares são os resultados observados (hipótese nula). Se a hipótese nula for verdadeira, então podemos concluir que os dados são gerados por um modelo global, no caso contrário trabalhamos com valores do p-value (probabilidade de se obter um ou mais padrões) que tornem a hipótese nula verdadeira.

A segunda questão baseia-se nos intervalos de confiança, isto é, na premissa de que o modelo escolhido é o correcto. Numa regressão standard é possível estimar os erros dos coeficientes de regressão, uma vez que existe o desvio-padrão amostral dos coeficientes estimados, que nos permite atribuir intervalos de confiança às estimativas dos coeficientes. Por exemplo, em amostras suficientemente grandes, o intervalo de confiança definido pelas estimativas dos parâmetros é aproximadamente 1,96 vezes o erro standard, o que significa que conterá o valor verdadeiro do coeficiente 95% das vezes (grau de confiança de 95%).

Na metodologia do GWR estimam-se coeficientes de superfícies, ou valores de coeficientes regressão para um conjunto de localizações geográficas. Pode-se considerar a diferença entre a estimativa de um coeficiente em duas localizações distintas. Trabalha-se com o desvio-padrão, (e, consequentemente, com um intervalo de confiança) para essa diferença e decide-se se o valor estimado da diferença tem importância suficiente que justifique o uso de um modelo geográfico não estacionário.

No que respeita à terceira questão, que modelo matemático é o melhor, devemos ter sempre em conta que todos os modelos apresentam lacunas, no entanto uns são mais úteis que outros. É neste contexto que devemos perguntar se um modelo de regressão tradicional é mais útil, ou não, que um modelo GWR.

Apoiamo-nos numa medida de utilidade denominada por AIC (Critério de Informação Akaike) ou no critério de informação akaike corrigido (AIC_c). Este critério baseia-se na ideia de que um modelo verdadeiro pode existir, mas não é directamente verificável. No entanto, é possível estimar a proximidade do modelo a um modelo verdadeiro. O AIC é a medida dessa proximidade e o modelo mais próximo é considerado o melhor modelo. Observe-se que o AIC não é simplesmente uma medida de ajustamento, tal como a soma dos erros quadrados, mas tem em conta a complexidade do modelo.

3 – Apresentação do problema em estudo

Este trabalho centra-se na análise dos resultados eleitorais dos cinco principais partidos, PS, PCP, BE, CDS/PP e PSD relativos às eleições legislativas do ano de 2005.

O nível de desagregação dos dados são os municípios e o universo espacial é Portugal Continental.

Consideraram-se cinco modelos, um para cada partido político em estudo, tendo cada um dos modelos como variável de interesse o resultado obtido pelo respectivo partido em cada município.

No que diz respeito às variáveis explicativas consideraram-se:

VotBRan – o número de votos em branco em cada município;
TotalVot – o número total de votantes de cada município;
TotFreg – o número total de freguesias de cada município;
Ind – o índice de envelhecimento, isto é, o quociente entre o número de idosos (população com idade superior ou igual a 65 anos) e o número de jovens (população com idades entre0-14 anos)

Os dados relacionados com os resultados das votações foram obtidos no site www.stape.pt.
As coordenadas cartesianas foram disponibilizadas pelo Instituto Geográfico Português e os dados etários foram disponibilizados pelo Instituto Nacional de Estatística.

4 – Modelo Espacial Explicativo dos Resultados Eleitorais em Portugal

4.1-Especificação do modelo GWR

Neste trabalho adopta-se uma regressão GWR3 com termo de erro Gaussiano. Optou-se por escolher um núcleo adaptativo, seleccionando-se a faixa por utilização do critério de minimização de Akaike corrigido. Escolheu-se o para teste de significância o teste de Monte Carlo.

Para cada partido organizou-se uma folha de Excel guardada com a extensão .csv. Em cada folha de Excel, na primeira linha, indicam-se os nomes das variáveis, tendo uma extensão máxima de oito caracteres.

As duas primeiras variáveis explicativas descrevem a localização geográfica das freguesias, seguindo um esquema de ponderação geográfica baseada na especificação do Kernel (núcleo), a qual se encontra seleccionada por defeito no software.

Construíram-se, deste modo, os modelos que se especificam em seguida:

Modelos de Regressão Global, em que os parâmetros são considerados invariantes no espaço:

$ResPS_i = Constante + \beta_{1i}\ VotBRan_i + \beta_{2i}\ TotalVot_i + \beta_{3i}\ TotFreg_i + \beta_{4i}\ Ind_i + \varepsilon_i$

$ResBE_i = Constante + \beta_{1i}\ VotBRan_i + \beta_{2i}\ TotalVot_i + \beta_{3i}\ TotFreg_i + \beta_{4i}\ Ind_i + \varepsilon_i$

$ResPCP_i = Constante + \beta_{1i}\ VotBRan_i + \beta_{2i}\ TotalVot_i + \beta_{3i}\ TotFreg_i + \beta_{4i}\ Ind_i + \varepsilon_i$

$ResPSD_i = Constante + \beta_{1i}\ VotBRan_i + \beta_{2i}\ TotalVot_i + \beta_{3i}\ TotFreg_i + \beta_{4i}\ Ind_i + \varepsilon_i$

$ResCDS_i = Constante + \beta_{1i}\ VotBRan_i + \beta_{2i}\ TotalVot_i + \beta_{3i}\ TotFreg_i + \beta_{4i}\ Ind_i + \varepsilon_i$

Extensões dos modelos anteriores aos modelos GWR:

$ResPS_i = Constante(u_i,v_i) + \beta_{1i}(u_i,v_i)\ VotBRan_i + \beta_{2i}\ (u_i,v_i)TotalVot_i + \beta_{3i}(u_i,v_i)\ TotFreg_i + \beta_{4i}\ (u_i,v_i)Ind_i + \varepsilon_i$

$ResBE_i = Constante(u_i,v_i) + \beta_{1i}(u_i,v_i)\ VotBRan_i + \beta_{2i}\ (u_i,v_i)TotalVot_i + \beta_{3i}(u_i,v_i)\ TotFreg_i + \beta_{4i}\ (u_i,v_i)Ind_i + \varepsilon_i$

$ResPCP_i = Constante(u_i,v_i) + \beta_{1i}(u_i,v_i)\ VotBRan_i + \beta_{2i}\ (u_i,v_i)TotalVot_i + \beta_{3i}(u_i,v_i)\ TotFreg_i + \beta_{4i}\ (u_i,v_i)Ind_i + \varepsilon_i$

$ResPSD_i = Constante(u_i,v_i) + \beta_{1i}(u_i,v_i)\ VotBRan_i + \beta_{2i}\ (u_i,v_i)TotalVot_i + \beta_{3i}(u_i,v_i)\ TotFreg_i + \beta_{4i}\ (u_i,v_i)Ind_i + \varepsilon_i$

$ResCDS_i = Constante(u_i,v_i) + \beta_{1i}(u_i,v_i)\ VotBRan_i + \beta_{2i}\ (u_i,v_i)TotalVot_i + \beta_{3i}(u_i,v_i)\ TotFreg_i + \beta_{4i}\ (u_i,v_i)Ind_i + \varepsilon_i$

Onde:

(u_i,v_i) são as coordenadas cartesianas do i-ésimo município;

$ResPS_i$ são os resultados do partido político PS no i-ésimo município;

$ResBE_i$ são os resultados do partido político BE no i-ésimo município;

$ResPCP_i$ são os resultados do partido político PCP no i-ésimo município;

$ResPSD_i$ são os resultados do partido político PSD no i-ésimo município;

$ResCDS_i$ são os resultados do partido político CDS no i-ésimo município;

$\beta_{ki}\ (u_i,v_i)$ com k=1,…,4 é a realização da função contínua $\beta_k\ (u,v)$ no ponto i.

ou, de modo equivalente, mas utilizando a notação matricial:

$ResPS_i = Constante + \beta_{1i}\ VotBRan_i + \beta_{2i}\ TotalVot_i + \beta_{3i}\ TotFreg_i + \beta_{4i}\ Ind_i + \rho_1 W_1 + u$

$ResBE_i = Constante + \beta_{1i}\ VotBRan_i + \beta_{2i}\ TotalVot_i + \beta_{3i}\ TotFreg_i + \beta_{4i}\ Ind_i + \rho_2 W_2 + u$

$ResPCP_i = Constante + \beta_{1i}\ VotBRan_i + \beta_{2i}\ TotalVot_i + \beta_{3i}\ TotFreg_i + \beta_{4i}\ Ind_i + \rho_3 W_3 + u$

$ResPSD_i = Constante + \beta_{1i}\ VotBRan_i + \beta_{2i}\ TotalVot_i + \beta_{3i}\ TotFreg_i + \beta_{4i}\ Ind_i + \rho_4 W_4 + u$

$ResCDS_i = Constante + \beta_{1i}\ VotBRan_i + \beta_{2i}\ TotalVot_i + \beta_{3i}\ TotFreg_i + \beta_{4i}\ Ind_i + \rho_5 W_5 + u$

Onde:

ρ_1 representa o coeficiente autoregressivo espacial do modelo do PS

W_1representa a matriz de vizinhanças do modelo do PS

P_2 representa o coeficiente autoregressivo espacial do modelo do BE

W_2representa a matriz de vizinhanças do modelo do BE

ρ_3 representa o coeficiente autoregressivo espacial do modelo do PCP

W_3representa a matriz de vizinhanças do modelo do PCP

ρ_4 representa o coeficiente autoregressivo espacial do modelo do PSD

W_4representa a matriz de vizinhanças do modelo do PSD

ρ_5 representa o coeficiente autoregressivo espacial do modelo do CDS

W_5representa a matriz de vizinhanças do modelo do CDS

4.2-Resultados das estimações estatísticas

4.2.1 Resultados referentes ao modelo do Partido Socialista

A partir do ficheiro de resultados do " output" do modelo de regressão dos resultados do Partido Socialista, podemos concluir, por comparação que existe uma melhoria na adopção do modelo GWR, relativamente ao modelo global, uma vez que os valores do AIC são menores e valor de R^2 é maior.

O que confere ao modelo GWR, uma maior capacidade explicativa e um modelo mais ajustado como se pode observar na tabela 1.

Tabela 1 – Resultados da Estimação Global e estimação GWR para o modelo do PS

	Global Regression Parametres	GWR Estimation
R^2	0,9898	0.991367
AIC	4824	4788

Os resultados da ANOVA, exibem para valor do teste F, o valor de : 10,6425.

No que se refere à correlação entre as variáveis observamos que as variáveis VotBranc, TotFreg, Ind, estão negativamente relacionadas com a variável dependente. A variável TotVot, está positivamente relacionada com a variável dependente, conforme os resultados da tabela 2.

Tabela 2 – Resultados do Teste-T para os parâmetros do modelo do PS

Parameter	Estimate	Std Err	T
---------	------------	------------	------------
Intercept	503.651059396720	180.233081427710	2.794442892075
VotBRan	-6.333536601832	0.722183899773	-8.769977569580
TotalVot	0.584122857798	0.015834230576	36.889881134033
TotFreg	-34.736142267707	8.345511476813	-4.162254333496
Ind	-243.594499731847	95.564224338056	-2.549013614655

No que se refere à análise dos erros padrão em comparação com a amplitude interquartil, a tabela 3, mostra-nos que as variáveis VotBranc, TotFreg, são não estacionárias espacialmente.

Esta tabela serve de comparação entre os valores da amplitude interquartil, e os erros padrão, sendo consideradas não estacionárias espacialmente as variáveis em que o valor da amplitude interquartil é maior que duas vezes SE, devido ao facto de estarmos na presença de uma estimação Gausiana.

Tabela 3- Dados referentes ao SE e ao sumário dos parâmetros dos 5 números do modelo PS

VARIÁVEIS	SE	2XSE	Lwr Quartile	Upr Quartile	Amplitude Interquartil
VotBranc	0.7222	1.4444	- 6.520041	-4.989512	1.530529
TotVot	0.0158	0.0316	0.571258	0.582985	0.011727
TotFreg	8.3455	16.691	-40.03505	-18.43539	21.599659
Ind	95.5642	191.1284	-188.5771	-145.8676	42.709562

Relativamente à variabilidade espacial dos parâmetros, o teste de significância de Monte Carlo, não assinala nenhuma variável como significativa.

4.2.2 Resultados referentes ao modelo do Partido Comunista Português

A partir do ficheiro de resultados do “ output” do modelo de regressão dos resultados do Partido Comunista Português podemos concluir, por comparação que existe uma significativa melhoria na adopção do modelo GWR, relativamente ao modelo global, uma vez que os valores do AIC são menores e valor de R^2 é maior.

O que confere ao modelo GWR, uma maior capacidade explicativa e um modelo mais ajustado como se pode observar na tabela 4.

Tabela 4– Resultados da Estimação Global e estimação GWR para o modelo do PCP

	Global Regression Parametres	GWR Estimation
R^2	0,829493	0,969838
AIC	4751,83	4306,16

Os resultados da ANOVA, exibem para valor do teste F, o valor de : 49,5509.

No que se refere à correlação entre as variáveis, observamos que as variáveis VotBranc, TotFreg, Ind, estão negativamente relacionadas com a variável dependente. As variáveis TotVot, e Ind, estão positivamente relacionadas com a variável dependente, conforme os resultados da tabela 5.

Tabela 5 – Resultados do Teste-T para os parâmetros do modelo do PCP

Parameter	Estimate	Std Err	T
Intercept	-51.964318057372	157.859261395951	-0.329181313515
VotBRan	-2.610784518190	0.632533251427	-4.127505779266
TotalVot	0.142368316014	0.013868596840	10.265517234802
TotFreg	-57.215640635454	7.309514253794	-7.827557086945
Ind	448.085409650571	83.701048389018	5.353402614594

No que se refere à análise dos erros padrão em comparação com a amplitude interquartil, a tabela 6, mostra-nos que as variáveis VotBranc, TotFreg, e TotVot, são não estacionárias espacialmente.

Esta tabela serve de comparação entre os valores da amplitude interquartil, e o valor de duas vezes o erro padrão, sendo consideradas não estacionárias espacialmente as variáveis em que o primeiro valor é maior que o segundo.

Tabela 6- Dados referentes ao SE e ao sumário dos parâmetros dos 5 números do modelo PCP

VARIÁVEIS	SE	2XSE	Lwr Quartile	Upr Quartile	Amplitude Interquartil
VotBranc	0,632533	1,2650665	-14,05065	0.590931	13,459715
TotVot	0,0138686	0,0277372	0,055967	0,421134	0,365167
TotFreg	7,3095143	14,6190286	-70,43355	-11,0676	59,36599
Ind	83,701048	167,402096	-223,7326	74,21904	149,5135

Relativamente à variabilidade espacial dos parâmetros, o teste de significância de Monte Carlo, apresenta como variáveis significativas VotBran, TotVot, TotFreg, a um nível de significância de 0,1 %.

4.2.3 Resultados referentes ao modelo do Bloco de Esquerda

A partir do ficheiro de resultados do “ output” do modelo de regressão dos resultados do Bloco de Esquerda, podemos concluir, que a regressão GWR, apresenta mais vantagens.

Por um lado observa-se um acréscimo da capacidade explicativa do modelo GWR, relativamente ao modelo de regressão global (maior valor de R^2).

Por outro lado o valor do AIC, é bastante menor na regressão GWR do que no modelo de regressão global, o que, independentemente da diferença entre o número de graus de liberdade de cada um dos modelos, faz com que o modelo GWR se afigure como melhor modelo. Ver tabela 7.

Tabela 7– Resultados da Estimação Global e estimação GWR para o modelo do BE

	Global Regression Parametres	GWR Estimation
R^2	0,976	0,989
AIC	4151,84	3955,39

Os resultados da ANOVA, exibem para valor do teste F, o valor de : 28,9807. No que se refere à correlação entre as variáveis, observamos que as variáveis VotBranc, TotVot, Ind, estão positivamente relacionadas com a variável dependente. A variável TotFreg, encontra-se negativamente relacionadas com a variável dependente, conforme os resultados da tabela 8.

Tabela 8– Resultados do Teste-T para os parâmetros do modelo do BE

Parameter	Estimate	Std Err	T
Intercept	-125.280203175642	53.237611378768	-2.353227376938
VotBRan	1.010930077505	0.213320137988	4.739027976990
TotalVot	0.067732611380	0.004677146988	14.481608390808
TotFreg	-29.452928136310	2.465114024796	-11.947896957397
Ind	110.093207518267	28.227953474032	3.900148391724

No que se refere à análise dos erros padrão em comparação com a amplitude interquartil, a tabela 9 mostra-nos que todas as variáveis são não estacionárias espacialmente.

Tabela 9- Dados referentes ao SE e ao sumário dos parâmetros dos 5 números do modelo BE

VARIÁVEIS	SE	2XSE	Lwr Quartile	Upr Quartile	Amplitude Interquartil
VotBranc	0,21332	0,42664	- 0,550541	1,468696	0,918155
TotVot	0,00467	0,00934	0,050657	0,106594	0,055937
TotFreg	2,46511	4,93022	-49,5256	-19,9856	29,54
Ind	28,22795	56,4559	-196,9134	93,87303	103,0403

Relativamente à variabilidade espacial dos parâmetros, o teste de significância de Monte Carlo, apresenta como variáveis significativas TotVot, TotFreg, e Ind. Como se pode observar na figura 1.

```
*************************************************
*                                               *
*   Test for spatial variability of parameters  *
*                                               *
*************************************************

Tests based on the Monte Carlo significance test
procedure due to Hope [1968,JRSB,30(3),582-598]

Parameter          P-value
----------      ------------------
Intercept         0.01000   **
VotBRan           0.01000   **
TotalVot          0.00000   ***
TotFreg           0.00000   ***
Ind               0.00000   ***

*** = significant at .1% level
**  = significant at 1% level
*   = significant at 5% level
```

Figura 1

4.2.4 Resultados referentes ao modelo do Partido Social Democrata

A partir do ficheiro de resultados do “ output” do modelo de regressão dos resultados do Partido Social Democrata podemos concluir, por comparação que existe uma significativa melhoria na adopção do modelo GWR, relativamente ao modelo global, uma vez que os valores do AIC são menores e valor de R^2 é maior.

O que confere ao modelo GWR, uma maior capacidade explicativa e um modelo mais ajustado como se pode observar na tabela 10.

Tabela 10– Resultados da Estimação Global e estimação GWR para o modelo do PSD

	Global Regression Parametres	GWR Estimation
R^2	0,957110	0,976915
AIC	4905,76	4747,34

Os resultados da ANOVA, exibem para valor do teste F, o valor de : 22,1992

No que se refere à correlação entre as variáveis, observamos que a variável, Ind, está negativamente relacionada com a variável dependente. As variáveis VotBran, TotVot, e TotFreg, estão positivamente relacionadas com a variável dependente, conforme os resultados da tabela 11.

Tabela 11– Resultados do Teste-T para os parâmetros do modelo do PSD

Parameter	Estimate	Std Err	T
Intercept	-127.558756668129	208.631367190840	-0.611407399178
VotBRan	3.940908790159	0.835974246122	4.714150905609
TotalVot	0.152692425602	0.018329138842	8.330583572388
TotFreg	104.040835001728	9.660465523432	10.769753456116
Ind	-230.276911335535	110.621727266965	-2.081660747528

No que se refere à análise dos erros padrão em comparação com a amplitude interquartil, a tabela 12, mostra-nos que as variáveis VotBranc, TotFreg, e Ind, são não estacionárias espacialmente.

TABELA 12- Dados referentes ao SE e ao sumário dos parâmetros dos 5 números do modelo PSD

VARIÁVEIS	SE	2XSE	Lwr Quartile	Upr Quartile	Amplitude Interquartil
VotBranc	0,83597	1,67194	0,820467	12,62131	11,800764
TotVot	0,01833	0,3666	-0,05951	0,232362	0,172852
TotFreg	9,66047	19,32094	66,434665	120,5779	54,14327
Ind	110,6217	221,2434	-39,80461	415,1852	375,3806

Relativamente à variabilidade espacial dos parâmetros, o teste de significância de Monte Carlo, apresenta todas as variáveis como significativas, destacando-se as variáveis VotBran, TotVot, e Ind.

4.2.5 Resultados referentes ao modelo do Centro Democrático Social

A partir do ficheiro de resultados do " output" do modelo de regressão dos resultados do Centro Democrático Social podemos concluir, por comparação que existe uma melhoria na adopção do modelo GWR, relativamente ao modelo global, uma vez que os valores do AIC são menores, e o valor de R^2 é maior.

O que confere ao modelo GWR, uma maior capacidade explicativa e um modelo mais ajustado como se pode observar na tabela 13.

Tabela 13– Resultados da Estimação Global e estimação GWR para o modelo do CDS

	Global Regression Parametres	GWR Estimation
R^2	0,923611	0,940435
AIC	4492,85	4427,717

Os resultados da ANOVA, exibem para valor do teste F, o valor de : 24,7672

No que se refere à correlação entre as variáveis, observamos que a variável, Ind, está negativamente relacionada com a variável dependente. As variáveis

VotBran, TotVot, e TotFreg, estão positivamente relacionadas com a variável dependente, conforme os resultados da tabela 14.

Tabela 14– Resultados do Teste-T para os parâmetros do modelo do CDS

Parameter	Estimate	Std Err	T
Intercept	-186.683336884485	98.745277123890	-1.890554547310
VotBRan	2.947645966196	0.395666815174	7.449818611145
TotalVot	0.023979681380	0.008675185898	2.764169216156
TotFreg	16.847718896867	4.572300695247	3.684735536575
Ind	-120.688002152222	52.357290574183	-2.305084943771

No que se refere à análise dos erros padrão em comparação com a amplitude interquartil, a tabela 15, mostra-nos que as variáveis VotBranc, TotVot apresenta não estacionáridade espacial, como se pode observar na tabela 15.

TABELA 15 - Dados referentes ao SE e ao sumário dos parâmetros dos 5 números do modelo CDS

VARIÁVEIS	SE	2XSE	Lwr Quartile	Upr Quartile	Amplitude Interquartil
VotBranc	0,39567	0,79133	1,392558	3,267796	1,875238
TotVot	0,0086752	0,0173504	0,015635	0,048689	0,033054
TotFreg	4,5723007	9,1446014	11,2981	14,528029	3,229903
Ind	52,35729	104,71458	-117,7078	-76,60843	41,09936

Relativamente à variabilidade espacial dos parâmetros, o teste de significância de Monte Carlo, apresenta como variáveis significativas, as variáveis VotBran, e TotVot, com um nível de significância de 0,1 %, e 5% respectivamente.

4.3-Discussão dos resultados

Observamos em todos os modelos, uma melhoria significativa na adopção de um modelo de regressão GWR, em detrimento de uma regressão OLS, uma vez que os valores dos respectivos coeficientes de determinação são significativamente maiores, o que proporciona maior capacidade explicativa à regressão GWR. Por comparação dos resultados obtidos pelo critério de informação AIC, considera-se igualmente melhor os modelos GWR, para os quais os valores do AIC são comparativamente menores que os de uma regressão OLS.

Por outro lado a variável VotBranc, encontra-se positivamente relacionada com os resultados eleitorais do CDS, PSD, e Bloco de Esquerda. Ao contrário do que acontece com os resultados eleitorais do PS e do PCP, nos quais a relação com esta variável é negativa.

Da população que vota em branco, verifica-se serem pessoas que habitualmente votam no PS e no PCP, uma vez que quanto maior for o número de votantes em branco, menor será o resultado destes dois partidos, nas diversas regiões.

Este facto é reforçado pela constatação de que o aumento dos votos em branco produz igualmente um aumento dos resultados eleitorais do Bloco de Esquerda, PSD, e CDS.

Por outras palavras os votos em branco funcionam como “ um outro partido “, que vai buscar votos ao PS e ao PCP, mas não aos partidos CDS, PSD, e Bloco de Esquerda.
No que se refere à variável Ind, que representa o índice de envelhecimento, por definição, quando o seu valor é maior que um estamos perante uma população envelhecida, e quando for menor que um, a população é mais jovem.

No caso em que a variável Ind é igual a um, nada se pode concluir, no entanto neste estudo nunca se observou um valor desta variável igual a um.

Dos resultados obtidos observa-se uma tendência partidária, isto é enquanto que nos partidos de “ esquerda “ (BE, PCP), a variável Ind, encontra-se positivamente relacionada com os resultados eleitorais destes partidos no partidos de “ cento – direita “ (PS, PSD, e CDS), esta encontra-se negativamente relacionada com os resultados eleitorais destes partidos.

Por conseguinte quanto mais envelhecida for a população, maiores serão os resultados do Bloco de Esquerda e do Partido Comunista Português.

Observa-se claramente uma tendência de votação nos partidos de “ direita “, por parte dos estractos etários mais jovens.

Relativamente ao total de freguesias de cada município, a variável TotFreg, encontra-se positivamente relacionada com os partidos de “ direita “ (CDS, e PSD), e negativamente relacionada com os partidos de “ esquerda “ (Bloco de Esquerda, PCP, e PS).

Isto significa que, em cada município, quanto maior for o numero de freguesias que o compõem, menor serão os resultados eleitorais dos partidos de “ esquerda “.

Se tivermos em conta que quanto maior for o numero de freguesias, maior será a probabilidade de se observarem efeitos de “ contágio “, podemos concluir que os resultados eleitorais obtidos pelos partidos de “esquerda” sofrem poucos efeitos de contágio.

Em regiões mais homogéneas (com menos freguesias e territórios menos divididos), há menor probabilidade de os resultados eleitorais obtidos pela esquerda dependerem dos valores das localizações vizinhas, isto é, a proximidade geográfica enquanto elemento permissivo da ocorrência de um fenómeno associado à votação por comparação. Este facto reforça a hipótese de que uma explicação parcial dos resultados eleitorais em cada município seja resultante da comparação que os eleitores fazem do desempenho do seu município, face a municípios vizinhos (sobretudo quando estão sob a tutela de um partido diferente).

Confirma-se deste modo a autocorrelação espacial, ao nível dos “ clusters “ espaciais, como nos mostram os valores da estatística – T, de todos os partidos.

4.4-Conclusões

Podemos concluir que, relativamente ao modelo de estudo dos resultados eleitorais, observam-se variações espaciais nas relações entre os resultados dos partidos, que podem ser influenciadas pelas diferencias etárias, pela composição de municípios no que respeita às suas freguesias, pelo número de votos em branco, e ainda no que toca à participação da população no acto eleitoral.

Reforça-se, deste modo a utilidade da metodologia GWR no estudo de fenómenos eleitorais, uma vez que se ganha em precisão e capacidade explicativa.

Seria pertinente, em futuros trabalhos incorporar mais variáveis aos modelos, nomeadamente o grau de desertificação das diversas zonas geográficas, e outros relevantes, como por exemplo o grau de instrução.

4.4-Referências Bibliográficas

Caleiro, António (2008). “Para uma visão espacial dos resultados eleitorais em Portugal,” *Documento de Trabalho Nº 2008/01*, Universidade de Évora, Departamento de Economia.

Calvo, E. and M. Escolar (2003). “The Local Voter: A Geografically Weighted Approach to Ecological Inference,” *American Journal of Political Science*, Vol. 47, Nº 1, pp. 189-204.

Fotheringham, A. S., C. Brunsdon, and Charlton (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons Ltd., West Sussetx, England.

Fotheringham, A. S., Charlton and C. Brunsdon (1997). “ Two Techniques for Exploring Non-Sationarity in Geographical Data,” *Geographical Systems*, Vol. 4, pp. 59-82.

Fotheringham, A. S., Charlton and C. Brunsdon (1997). “Measuring Sapatial Variations in Relationships with Geographically Weighted Regression,” *Recent Developments in Saptial Analysis: Spatial Satistics, Behavioral Modeling, and Computational Intelligence*, New York, Springer, pp. 60-82.

Páez A, T. Uchida, and K. Miyamoto (2002). “A General Framework for Estimation and Inference of Geographically Weighted Regression Models: 1. Location-Specific Kernel Bandwidths and a Test fot Location Heterogeneity,” *Environment and Planining A*, Vol. 34, pp. 733-754.

Páez A, T. Uchida, and K. Miyamoto (2002). “A General Framework for Estimation and Inference of Geographically Weighted Regression Models: 2. Spatial Association and Model Specifications Tests,” *Enviroment and Planning A*, Vol. 34, pp. 883-904.

Anexos

Anexo I – Ficheiro de Resultados do modelo referente ao partido político PS

```
**********************************************
*      Geographically Weighted Regression      *
*              Release 3.0.1                 *
*           Dated: 06-vii-2003              *
*                                          *
*      Martin Charlton, Chris Brunsdon        *
*           Stewart Fotheringham             *
*    (c) University of Newcastle upon Tyne    *
**********************************************
Program starts at: Tue Jun 29 21:34:04 2010

** Program limits:
** Maximum number of variables.....     52
** Maximum number of observations..  80000
** Maximum number of fit locations.  80000

ps
** Observed data file:        C:\Documents and Settings\Utilizador\Amb
** Prediction location file:  Estimation at sample point locations
** Result output file:        C:\Documents and Settings\Utilizador\Amb

** Variables in the data file...
 ResPS    Latitude Longitud VotBRan  TotalVot TotFreg  Ind

** Dependent (y) variable..........ResPS
** Easting  (x-coord) variable.....Latitude
** Northing (y-coord) variable.....Longitud
** No weight variable specified
** Independent variables in your model...
 VotBRan  TotalVot TotFreg  Ind
** Kernel type:  Fixed
** Kernel shape: Gaussian
** Bandwidth selection by AICc minimisation
** Use all regression points
** Calibration history requested
** No prediction report requested
** Output estimates to be written to .txt file
** Monte Carlo significance tests for spatial variation
** No casewise diagnostics requested

*** Analysis method ***
*** Geographically weighted multiple regression
** Cartesian coordinates: Euclidean Distance
```

```
*****************************************************************
*                                             *
*       GEOGRAPHICALLY WEIGHTED GAUSSIAN REGRESSION         *
*                                             *
*****************************************************************
 Number of data cases read: 276
 Observation points read...

 Dependent mean=  8820.33008
 Number of observations, nobs= 276
 Number of predictors,   nvar= 4
 ** Observation Easting extent:   5.53449011
 ** Observation Northing extent:  3.77543998
*Finding bandwidth...
  ... using all regression points
 This can take some time...
 *Calibration will be based on  276 cases
 *Fixed kernel bandwidth search limits:   0.188772  3.77544
 *AICc minimisation begins...
          Bandwidth              AICc
         1.297113367423       4840.651796509893
         1.982106000000       4844.936248850719
         0.873764637590       4848.343829880011
         1.558757270168       4841.095522931159
         1.135408542250       4841.782742536687
         1.397052444995       4840.553561109022
         1.458818191864       4840.668653447037
         1.358879114292       4840.546800918605
         1.335286698405       4840.569240949213
** Convergence after    9 function calls
** Convergence: Bandwidth=      1.35888

************************************************************
*          GLOBAL REGRESSION PARAMETERS                *
************************************************************
Diagnostic information...
Residual sum of squares.........    604415092.904616
Effective number of parameters..           5.000000
Sigma..........................        1493.423582
Akaike Information Criterion....        4824.992705
Coefficient of Determination....           0.989800
Adjusted r-square...............          0.989611

Parameter             Estimate             Std Err                T
---------         ------------         ------------         ------------
Intercept      503.651059396720      180.233081427710       2.794442892075
 VotBRan         -6.333536601832       0.722183899773      -8.769977569580
 TotalVot         0.584122857798       0.015834230576      36.889881134033
 TotFreg        -34.736142267707       8.345511476813       -4.162254333496
 Ind         -243.594499731847      95.564224338056      -2.549013614655
```

****
*   GWR ESTIMATION   *
****

Fitting Geographically Weighted Regression Model...
Number of observations............ 276
Number of independent variables... 5
(Intercept is variable 1)
Bandwidth (in data units)......... 1.35887911
Number of locations to fit model.. 276

Diagnostic information...
Residual sum of squares......... 511549756.901938
Effective number of parameters.. 9.545134
Sigma........................... 1385.581619
Akaike Information Criterion.... 4788.650118
Coefficient of Determination.... 0.991367
Adjusted r-square............... 0.991057
** Results written to .txt file

****
*   ANOVA   *
****

Source	SS	DF	MS	F
OLS Residuals	604415092.9	5.00		
GWR Improvement	92865336.0	4.55	20431815.6365	
GWR Residuals	511549756.9	266.45	1919836.4219	10.6425

****
*   PARAMETER 5-NUMBER SUMMARIES   *
****

Label	Minimum	Lwr Quartile	Median	Upr Quartile	Maximum
Intrcept	132.773469	257.123381	315.226109	355.278502	507.870304
VotBRan	-6.970763	-6.520041	-5.855587	-4.989512	-3.439373
TotalVot	0.543935	0.571258	0.578418	0.582985	0.591216
TotFreg	-52.845037	-40.035049	-29.830045	-18.435390	5.051082
Ind	-227.821701	-188.577135	-175.946778	-145.867573	-40.444813

<----------------- LOWER ----------------><----------------- UPPER ---------------->

Label	Far Out	Outer Fence	Outside	Inner Fence	Inner Fence	Outside	Outer Fence	Far Out
Intrcept	0	-37.341980	0	109.890701	502.511182	1	649.743862	0

VotBRan 0 -11.111626 0 -8.815834 -2.693719 0 -
0.397926 0
TotalVot 0 0.536076 8 0.553667 0.600576 0 0.618167
0
TotFreg 0 -104.834026 0 -72.434537 13.964099 0
46.363588 0
Ind 0 -316.705820 0 -252.641478 -81.803231 10 -
17.738888 0

```
**************************************************
*                                    *
*   Test for spatial variability of parameters  *
*                                    *
**************************************************
```

Tests based on the Monte Carlo significance test
procedure due to Hope [1968,JRSB,30(3),582-598]

Parameter	P-value	
Intercept	0.84000	n/s
VotBRan	0.55000	n/s
TotalVot	0.89000	n/s
TotFreg	0.14000	n/s
Ind	0.92000	n/s

*** = significant at .1% level
** = significant at 1% level
* = significant at 5% level

Program terminates normally at: Tue Jun 29 21:34:12 2010

Anexo II – Ficheiro de Resultados do modelo referente ao partido político PCP

```
*****************************************************
*      Geographically Weighted Regression      *
*              Release 3.0.1                 *
*           Dated: 06-vii-2003               *
*                                            *
*      Martin Charlton, Chris Brunsdon        *
*           Stewart Fotheringham              *
*    (c) University of Newcastle upon Tyne     *
*****************************************************
Program starts at: Tue Jun 29 21:30:49 2010

** Program limits:
** Maximum number of variables.....     52
** Maximum number of observations..  80000
** Maximum number of fit locations.  80000

pcp
** Observed data file:        C:\Documents and Settings\Utilizador\Amb
** Prediction location file:  Estimation at sample point locations
** Result output file:        C:\Documents and Settings\Utilizador\Amb

** Variables in the data file...
 ResPCP   Latitude Longitud VotBRan  TotalVot TotFreg  Ind

** Dependent (y) variable..........ResPCP
** Easting  (x-coord) variable.....Latitude
** Northing (y-coord) variable.....Longitud
** No weight variable specified
** Independent variables in your model...
 VotBRan  TotalVot TotFreg  Ind
** Kernel type:  Fixed
** Kernel shape: Gaussian
** Bandwidth selection by AICc minimisation
** Use all regression points
** No calibration history requested
** No prediction report requested
** Output estimates to be written to .txt file
** Monte Carlo significance tests for spatial variation
** No casewise diagnostics requested

*** Analysis method ***
*** Geographically weighted multiple regression
** Cartesian coordinates: Euclidean Distance
```

```
*************************************************************
*                                            *
*       GEOGRAPHICALLY WEIGHTED GAUSSIAN REGRESSION        *
*                                            *
*************************************************************
 Number of data cases read: 276
 Observation points read...

 Dependent mean=  1513.3623
 Number of observations, nobs= 276
 Number of predictors,   nvar= 4
 ** Observation Easting extent:   5.53449011
 ** Observation Northing extent:  3.77543998
*Finding bandwidth...
  ... using all regression points
 This can take some time...
 *Calibration will be based on  276 cases
 *Fixed kernel bandwidth search limits:   0.188772  3.77544
 *AICc minimisation begins...
         1.297113367423       4542.417215427875
         1.982106000000       4635.506540743399
** Convergence after    11 function calls
** Convergence: Bandwidth=      0.55035

********************************************************
*          GLOBAL REGRESSION PARAMETERS               *
********************************************************
Diagnostic information...
Residual sum of squares.........    463667245.826179
Effective number of parameters..           5.000000
Sigma..........................      1308.032586
Akaike Information Criterion....        4751.826747
Coefficient of Determination....           0.832593
Adjusted r-square...............          0.829493

Parameter              Estimate              Std Err                T
---------         ------------         ------------         ------------
Intercept      -51.964318057372      157.859261395951       -0.329181313515
 VotBRan        -2.610784518190        0.632533251427       -4.127505779266
 TotalVot        0.142368316014        0.013868596840       10.265517234802
 TotFreg       -57.215640635454        7.309514253794       -7.827557086945
 Ind          448.085409650571       83.701048389018        5.353402614594


********************************************************
*              GWR ESTIMATION                    *
********************************************************
 Fitting Geographically Weighted Regression Model...
 Number of observations............ 276
 Number of independent variables... 5
```

(Intercept is variable 1)
Bandwidth (in data units)......... 0.550354986
Number of locations to fit model.. 276

Diagnostic information...
Residual sum of squares......... 74128364.246105
Effective number of parameters.. 30.984176
Sigma.......................... 550.041100
Akaike Information Criterion.... 4306.155341
Coefficient of Determination.... 0.973236
Adjusted r-square............... 0.969838
** Results written to .txt file

****
* ANOVA *
****

Source	SS	DF	MS	F
OLS Residuals	463667245.8	5.00		
GWR Improvement	389538880.0	25.98	14991388.6517	
GWR Residuals	74128364.2	245.02	302545.2112	49.5509

****
* PARAMETER 5-NUMBER SUMMARIES *
****

Label	Minimum	Lwr Quartile	Median	Upr Quartile	Maximum
Intrcept	-391.223644	-25.086766	113.858469	226.255069	1315.183065
VotBRan	-18.782269	-14.050646	-0.072169	0.590931	1.238295
TotalVot	0.036452	0.055967	0.063303	0.421134	0.484953
TotFreg	-232.613262	-70.433554	-14.800644	-11.067560	46.139328
Ind	-484.873883	-223.732575	-77.317127	74.219038	156.522621

<------------------ LOWER -----------------><------------------ UPPER ----------------->

Label	Far Out	Outer Fence	Outside	Inner Fence	Inner Fence	Outside	Outer Fence	Far Out
Intrcept	0	-779.112271	0	-402.099519	603.267821	7	980.280573	7
VotBRan	0	-57.975378	0	-36.013012	22.553297	0	44.515663	0
TotalVot	0	-1.039533	0	-0.491783	0.968884	0	1.516634	0
TotFreg	0	-248.531536	13	-159.482545	77.981431	0	167.030422	0
Ind	0	-1117.587414	0	-670.659994	521.146458	0	968.073878	0

```
*************************************************
*                                               *
*   Test for spatial variability of parameters  *
*                                               *
*************************************************

Tests based on the Monte Carlo significance test
procedure due to Hope [1968,JRSB,30(3),582-598]

Parameter              P-value
----------      ------------------
Intercept          0.49000   n/s
VotBRan            0.00000   ***
TotalVot           0.00000   ***
TotFreg            0.00000   ***
Ind                0.94000   n/s

*** = significant at .1% level
**  = significant at 1% level
*   = significant at 5% level

Program terminates normally at: Tue Jun 29 21:30:58 2010
```

Anexo III – Ficheiro de Resultados do modelo referente ao partido político BE

```
*************************************************
*      Geographically Weighted Regression      *
*              Release 3.0.1                   *
*           Dated: 06-vii-2003                 *
*                                              *
*      Martin Charlton, Chris Brunsdon         *
*          Stewart Fotheringham                *
*    (c) University of Newcastle upon Tyne     *
*************************************************
Program starts at: Tue Jun 29 21:40:01 2010

** Program limits:
** Maximum number of variables.....     52
** Maximum number of observations..  80000
** Maximum number of fit locations.  80000

be
** Observed data file:        C:\Documents and Settings\Utilizador\Amb
** Prediction location file:  Estimation at sample point locations
** Result output file:        C:\Documents and Settings\Utilizador\Amb

** Variables in the data file...
 ResBE    Latitude Longitud VotBRan  TotalVot TotFreg  Ind

** Dependent (y) variable..........ResBE
** Easting  (x-coord) variable.....Latitude
** Northing (y-coord) variable.....Longitud
** No weight variable specified
** Independent variables in your model...
 VotBRan  TotalVot TotFreg  Ind
** Kernel type:  Fixed
** Kernel shape: Gaussian
** Bandwidth selection by AICc minimisation
** Use all regression points
** Calibration history requested
** No prediction report requested
** Output estimates to be written to .txt file
** Monte Carlo significance tests for spatial variation
** No casewise diagnostics requested

*** Analysis method ***
*** Geographically weighted multiple regression
** Cartesian coordinates: Euclidean Distance
****************************************************************
*                                                              *
*       GEOGRAPHICALLY WEIGHTED GAUSSIAN REGRESSION            *
*                                                              *
****************************************************************
 Number of data cases read: 276
```

Observation points read...

Dependent mean= 1264.23193
Number of observations, nobs= 276
Number of predictors,   nvar= 4
** Observation Easting extent: 5.53449011
** Observation Northing extent: 3.77543998
*Finding bandwidth...
... using all regression points
This can take some time...
*Calibration will be based on 276 cases
*Fixed kernel bandwidth search limits: 0.188772 3.77544
*AICc minimisation begins...

Bandwidth	AICc
1.297113367423	4053.106891757345
1.982106000000	4090.000520751196
0.873764637590	4034.784627580246
0.612120732931	4051.838509810157
1.035469462763	4039.358799232596
0.773825559287	4036.507817354056
0.935530384459	4035.738103667296
0.835591306608	4034.895971112672
0.897357053476	4034.999549715452
0.859183722666	4034.756288191715

** Convergence after   10 function calls
** Convergence: Bandwidth=   0.85918

****
*   GLOBAL REGRESSION PARAMETERS   *
****

Diagnostic information...
Residual sum of squares.........   52735540.704288
Effective number of parameters..   5.000000
Sigma..........................   441.130470
Akaike Information Criterion....   4151.836550
Coefficient of Determination....   0.976702
Adjusted r-square...............   0.976270

Parameter	Estimate	Std Err	T
Intercept	-125.280203175642	53.237611378768	-2.353227376938
VotBRan	1.010930077505	0.213320137988	4.739027976990
TotalVot	0.067732611380	0.004677146988	14.481608390808
TotFreg	-29.452928136310	2.465114024796	-11.947896957397
Ind	110.093207518267	28.227953474032	3.900148391724

```
*********************************************************
*              GWR ESTIMATION                    *
*********************************************************
 Fitting Geographically Weighted Regression Model...
 Number of observations............ 276
 Number of independent variables... 5
  (Intercept is variable 1)
 Bandwidth (in data units).........  0.859183723
 Number of locations to fit model.. 276

Diagnostic information...
Residual sum of squares.........     23726379.611569
Effective number of parameters..          15.970296
Sigma..........................          302.067678
Akaike Information Criterion....        3955.386854
Coefficient of Determination....           0.989518
Adjusted r-square...............           0.988872
 ** Results written to .txt file

*********************************************************
*                  ANOVA                        *
*********************************************************
      Source                   SS        DF             MS            F
OLS Residuals            52735540.7      5.00
GWR Improvement             29009162.0     10.97     2644337.2014
GWR Residuals             23726379.6    260.03       91244.8818       28.9807


*********************************************************
*        PARAMETER 5-NUMBER SUMMARIES              *
*********************************************************
  Label      Minimum  Lwr Quartile       Median  Upr Quartile      Maximum
-------- ------------- ------------- ------------- ------------- -------------
Intrcept   -119.985761    -88.174078    -48.472967    112.218742    291.059850
VotBRan      -1.260677     -0.550541      1.040909      1.468696      1.949110
TotalVot      0.045777      0.050657      0.057754      0.106594      0.126376
TotFreg     -75.361411    -49.525625    -24.675462    -19.985597    -14.866989
Ind        -415.814250   -196.913352      9.894763     93.873032    121.131353


      <----------------- LOWER ---------------->< ----------------- UPPER ----------------
->
  Label Far Out   Outer Fence Outside   Inner Fence   Inner Fence Outside
Outer Fence Far Out
-------- ------- ------------- ------- ------------- ------------ ------- ------------ -------
Intrcept       0   -689.352537      0   -388.763307    412.807972      0
713.397201      0
VotBRan        0     -6.608254      0     -3.579397      4.497552      0      7.526408
0
```

TotalVot 0 -0.117152 0 -0.033247 0.190498 0 0.274403
0
TotFreg 0 -138.145708 0 -93.835666 24.324444 0
68.634486 0
Ind 0 -1069.272504 0 -633.092928 530.052608 0
966.232183 0

```
*************************************************
*                                               *
*   Test for spatial variability of parameters  *
*                                               *
*************************************************
```

Tests based on the Monte Carlo significance test
procedure due to Hope [1968,JRSB,30(3),582-598]

Parameter	P-value	
Intercept	0.01000	**
VotBRan	0.01000	**
TotalVot	0.00000	***
TotFreg	0.00000	***
Ind	0.00000	***

*** = significant at .1% level
** = significant at 1% level
* = significant at 5% level

Program terminates normally at: Tue Jun 29 21:40:09 2010

Anexo IV – Ficheiro de Resultados do modelo referente ao partido político PSD

```
*********************************************************
*      Geographically Weighted Regression      *
*             Release 3.0.1                  *
*           Dated: 06-vii-2003               *
*                                     *
*      Martin Charlton, Chris Brunsdon        *
*           Stewart Fotheringham              *
*    (c) University of Newcastle upon Tyne     *
*********************************************************
Program starts at: Tue Jun 29 21:36:52 2010

** Program limits:
** Maximum number of variables.....     52
** Maximum number of observations..  80000
** Maximum number of fit locations.  80000

psd
** Observed data file:        C:\Documents and Settings\Utilizador\Amb
** Prediction location file:  Estimation at sample point locations
** Result output file:        C:\Documents and Settings\Utilizador\Amb

** Variables in the data file...
 ResPSD   Latitude Longitud VotBRan  TotalVot TotFreg  Ind

** Dependent (y) variable..........ResPSD
** Easting  (x-coord) variable.....Latitude
** Northing (y-coord) variable.....Longitud
** No weight variable specified
** Independent variables in your model...
 VotBRan  TotalVot TotFreg  Ind
** Kernel type:  Fixed
** Kernel shape: Gaussian
** Bandwidth selection by AICc minimisation
** Use all regression points
** Calibration history requested
** No prediction report requested
** Output estimates to be written to .txt file
** Monte Carlo significance tests for spatial variation
** No casewise diagnostics requested

*** Analysis method ***
*** Geographically weighted multiple regression
** Cartesian coordinates: Euclidean Distance
*************************************************************************
*                                                *
*       GEOGRAPHICALLY WEIGHTED GAUSSIAN REGRESSION          *
*                                                *
*************************************************************************
 Number of data cases read: 276
```

```
 Observation points read...

 Dependent mean=  5488.48926
 Number of observations, nobs= 276
 Number of predictors,   nvar= 4
 ** Observation Easting extent:   5.53449011
 ** Observation Northing extent:  3.77543998
*Finding bandwidth...
  ... using all regression points
 This can take some time...
 *Calibration will be based on  276 cases
 *Fixed kernel bandwidth search limits:   0.188772  3.77544
 *AICc minimisation begins...
           Bandwidth             AICc
         1.297113367423       4819.937954493751
         1.982106000000       4856.484765786355
         0.873764637590       4805.366115007428
         0.612120732931       4814.192205877753
         1.035469462763       4808.340589202436
         0.773825559287       4806.373668633992
         0.935530384459       4805.942432704764
         0.835591306608       4805.437559032426
         0.897357053476       4805.492026339154
         0.859183722666       4805.351568867179
** Convergence after    10 function calls
** Convergence: Bandwidth=      0.85918

*********************************************************
*          GLOBAL REGRESSION PARAMETERS                  *
*********************************************************
Diagnostic information...
Residual sum of squares.........    809888978.199876
Effective number of parameters..           5.000000
Sigma..........................       1728.733703
Akaike Information Criterion....        4905.760233
Coefficient of Determination....           0.957890
Adjusted r-square...............         0.957110

Parameter              Estimate              Std Err                 T
---------          ------------          ------------         ------------
Intercept      -127.558756668129      208.631367190840       -0.611407399178
 VotBRan          3.940908790159        0.835974246122        4.714150905609
 TotalVot         0.152692425602        0.018329138842        8.330583572388
 TotFreg        104.040835001728        9.660465523432       10.769753456116
 Ind           -230.276911335535      110.621727266965       -2.081660747528
```

```
*********************************************************
*            GWR ESTIMATION                  *
*********************************************************
 Fitting Geographically Weighted Regression Model...
 Number of observations............ 276
 Number of independent variables... 5
  (Intercept is variable 1)
 Bandwidth (in data units).........  0.859183723
 Number of locations to fit model.. 276

Diagnostic information...
Residual sum of squares.........    418211552.951388
Effective number of parameters..          15.970296
Sigma..........................        1268.196447
Akaike Information Criterion....        4747.341232
Coefficient of Determination....           0.978255
Adjusted r-square...............          0.976915
 ** Results written to .txt file

*********************************************************
*                ANOVA                       *
*********************************************************
      Source                 SS       DF           MS          F
OLS Residuals           809888978.2      5.00
GWR Improvement            391677440.0     10.97    35703452.0865
GWR Residuals            418211553.0    260.03     1608322.2287       22.1992


*********************************************************
*        PARAMETER 5-NUMBER SUMMARIES              *
*********************************************************
  Label      Minimum  Lwr Quartile      Median  Upr Quartile      Maximum
-------- ------------- ------------- ------------- ------------- -------------
Intrcept  -1175.138131   -434.013514   -153.462889    -71.340866
304.775451
VotBRan      -0.599554     0.820467      2.497411     12.621231     15.288442
TotalVot     -0.110521    -0.059510      0.199062      0.232362      0.270388
TotFreg      31.560226     66.434665     78.974232    120.577935    211.890414
Ind        -222.361086    -39.804613     52.879839    415.185234   1047.430277


      <----------------- LOWER ---------------><----------------- UPPER ---------------
->
  Label Far Out   Outer Fence Outside   Inner Fence   Inner Fence Outside
Outer Fence Far Out
-------- ------- ------------- ------- ------------- ------------ ------- ------------ -------
Intrcept       0  -1522.031456       3   -978.022485   472.668106       0
1016.677077       0
```

VotBRan 0 -34.581824 0 -16.880678 30.322377 0
48.023522 0
TotalVot 0 -0.935127 0 -0.497319 0.670171 0 1.107980
0
TotFreg 0 -95.995146 0 -14.780241 201.792841 8
283.007746 0
Ind 0 -1404.774153 0 -722.289383 1097.670004 0
1780.154775 0

```
*************************************************
*                                  *
*   Test for spatial variability of parameters  *
*                                  *
*************************************************
```

Tests based on the Monte Carlo significance test
procedure due to Hope [1968,JRSB,30(3),582-598]

Parameter	P-value	
Intercept	0.48000	n/s
VotBRan	0.00000	***
TotalVot	0.00000	***
TotFreg	0.01000	**
Ind	0.00000	***

*** = significant at .1% level
** = significant at 1% level
* = significant at 5% level

Program terminates normally at: Tue Jun 29 21:37:01 2010

Anexo V – Ficheiro de Resultados do modelo referente ao partido político CDS

```
*****************************************************
*      Geographically Weighted Regression      *
*             Release 3.0.1                 *
*           Dated: 06-vii-2003               *
*                                          *
*      Martin Charlton, Chris Brunsdon        *
*           Stewart Fotheringham              *
*    (c) University of Newcastle upon Tyne     *
*****************************************************
Program starts at: Tue Jun 29 21:42:35 2010

** Program limits:
** Maximum number of variables.....     52
** Maximum number of observations..  80000
** Maximum number of fit locations.  80000

cds
** Observed data file:        C:\Documents and Settings\Utilizador\Amb
** Prediction location file:  Estimation at sample point locations
** Result output file:        C:\Documents and Settings\Utilizador\Amb

** Variables in the data file...
 ResCDS   Latitude Longitud VotBRan  TotalVot TotFreg  Ind

** Dependent (y) variable..........ResCDS
** Easting  (x-coord) variable.....Latitude
** Northing (y-coord) variable.....Longitud
** No weight variable specified
** Independent variables in your model...
 VotBRan  TotalVot TotFreg  Ind
** Kernel type:  Fixed
** Kernel shape: Gaussian
** Bandwidth selection by AICc minimisation
** Use all regression points
** Calibration history requested
** No prediction report requested
** Output estimates to be written to .txt file
** Monte Carlo significance tests for spatial variation
** No casewise diagnostics requested

*** Analysis method ***
*** Geographically weighted multiple regression
** Cartesian coordinates: Euclidean Distance
```

```
*************************************************************
*                                            *
*       GEOGRAPHICALLY WEIGHTED GAUSSIAN REGRESSION          *
*                                            *
*************************************************************
 Number of data cases read: 276
 Observation points read...

 Dependent mean=  1430.1449
 Number of observations, nobs= 276
 Number of predictors,   nvar= 4
 ** Observation Easting extent:   5.53449011
 ** Observation Northing extent:  3.77543998
*Finding bandwidth...
  ... using all regression points
 This can take some time...
 *Calibration will be based on  276 cases
 *Fixed kernel bandwidth search limits:   0.188772  3.77544
 *AICc minimisation begins...
          Bandwidth              AICc
         1.297113367423       4577.808541508000
         1.982106000000       4577.219844200743
         2.667098632577       4580.286560537834
         1.720462097255       4576.420854833991
         1.558757272082       4576.345696881056
         1.458818193779       4576.579585903940
         1.620523018951       4576.320516371567
         1.658696349934       4576.340440521167
** Convergence after     8 function calls
** Convergence: Bandwidth=      1.62052

********************************************************
*          GLOBAL REGRESSION PARAMETERS                 *
********************************************************
Diagnostic information...
Residual sum of squares.........    181425766.791527
Effective number of parameters..            5.000000
Sigma..........................          818.210104
Akaike Information Criterion....         4492.850267
Coefficient of Determination....            0.925000
Adjusted r-square...............            0.923611

Parameter             Estimate              Std Err                T
---------          ------------          ------------         ------------
Intercept     -186.683336884485       98.745277123890       -1.890554547310
 VotBRan         2.947645966196        0.395666815174        7.449818611145
 TotalVot        0.023979681380        0.008675185898        2.764169216156
 TotFreg        16.847718896867        4.572300695247        3.684735536575
 Ind          -120.688002152222       52.357290574183       -2.305084943771
```

```
*********************************************************
*              GWR ESTIMATION                    *
*********************************************************
 Fitting Geographically Weighted Regression Model...
 Number of observations............ 276
 Number of independent variables... 5
  (Intercept is variable 1)
 Bandwidth (in data units).........  1.62052302
 Number of locations to fit model.. 276

Diagnostic information...
Residual sum of squares.........    139781032.615145
Effective number of parameters..           8.221153
Sigma...........................        722.496901
Akaike Information Criterion....        4427.717186
Coefficient of Determination....           0.942216
Adjusted r-square...............         0.940435
 ** Results written to .txt file

*********************************************************
*                   ANOVA                        *
*********************************************************
       Source                SS        DF            MS            F
OLS Residuals           181425766.8      5.00
GWR Improvement            41644736.0     3.22    12928517.5673
GWR Residuals            139781032.6    267.78      522001.7723       24.7672


*********************************************************
*         PARAMETER 5-NUMBER SUMMARIES              *
*********************************************************
   Label      Minimum  Lwr Quartile       Median  Upr Quartile      Maximum
-------- ------------- ------------- ------------- ------------- -------------
Intrcept  -224.297730  -144.914395  -119.268325    -93.834183    -53.161090
VotBRan       0.615451     1.392558     2.094353     3.267796     5.123676
TotalVot    -0.020185     0.015635     0.037133     0.048689     0.061861
TotFreg      8.631590    11.298126    13.146011    14.528029    23.813475
Ind       -152.496695  -117.707790   -94.553310   -76.608433   -59.404859


       <------------------ LOWER ----------------><------------------ UPPER ----------------
->
   Label Far Out   Outer Fence Outside   Inner Fence   Inner Fence Outside
Outer Fence Far Out
-------- ------- ------------- ------- ------------- ------------ ------- ------------ -------
Intrcept       0   -298.155029      1   -221.534712    -17.213866       0
59.406451       0
```

VotBRan 0 -4.233158 0 -1.420300 6.080654 0 8.893511
0
TotalVot 0 -0.083530 0 -0.033948 0.098272 0 0.147854
0
TotFreg 0 1.608416 0 6.453271 19.372884 15
24.217739 0
Ind 0 -241.005861 0 -179.356825 -14.959397 0
46.689639 0

****
* *
* Test for spatial variability of parameters *
* *
****

Tests based on the Monte Carlo significance test
procedure due to Hope [1968,JRSB,30(3),582-598]

Parameter	P-value	
Intercept	0.51000	n/s
VotBRan	0.00000	***
TotalVot	0.02000	*
TotFreg	0.48000	n/s
Ind	0.53000	n/s

*** = significant at .1% level
** = significant at 1% level
* = significant at 5% level

Program terminates normally at: Tue Jun 29 21:42:44 2010