INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

150 anos CENSOS EM PORTUGAL 1864-2014

Estimação do desemprego ao nível NUTS III

Soraia Pereira, Luís Correia, Pedro Campos

Departamento de Metodologias e Sistemas de Informação
Serviço de Métodos Estatísticos

10 de Abril de 2014

Problema

Estimação em Pequenos Domínios

Quadro 1: Taxas de desemprego por região NUTS II (NUTS-2002)

Unidade: %

	4ºT-2012	3ºT-2013	4ºT-2013	2012	2013
Portugal	**16,9**	**15,6**	**15,3**	**15,7**	**16,3**
Norte	17,8	16,6	16,4	16,1	17,2
Centro	12,7	11,2	10,7	12,0	11,7
Lisboa	18,7	17,9	17,2	17,6	18,5
Alentejo	17,2	16,1	15,5	15,9	16,8
Algarve	19,7	13,8	17,5	17,9	17,1
R. A. Açores	16,2	17,7	17,3	15,3	17,0
R. A. Madeira	19,7	17,3	17,2	17,5	18,3

Fonte: INE, Estatísticas do Emprego - 4º trimestre de 2013.

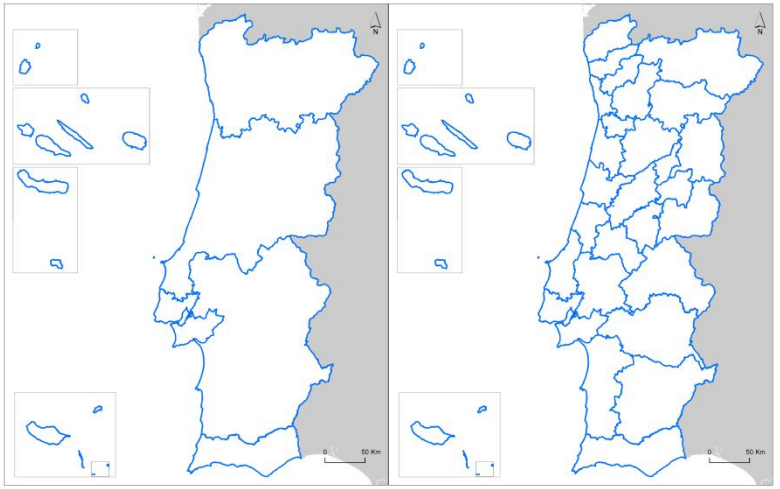


Figura 1: **Taxa de Desemprego do 4º trimestre de 2013**

Objetivos

- Estimação trimestral da taxa de desemprego e do total de desempregados por NUTS III de Portugal Continental.
- Coerência com as estimativas oficiais ao nível NUTS II.

Impacto na Sociedade

- Planeamento regional
- Combate ao desemprego
- Melhoria das condições de vida
- Crescimento económico
- Decisões de negócio



Figura 2: **Desemprego local**

Inferência

Design-based A estimação é feita com base no desenho de amostragem.

Model-based A estimação é feita com base em modelos probabilísticos.

Notação

Considere-se uma população finita U constituída por N indivíduos, que se subdivide em 28 regiões NUTS III (referentes a Portugal Continental).
Seja y_{ij} a variável que indica se o j-ésimo indivíduo ($j = 1, ..., N_i$) da i-ésima região ($i = 1, ..., 28$) está desempregado:

$$y_{ij} = \begin{cases} 1 & \text{se o j-ésimo indivíduo da i-ésima região está desempregado} \\ 0 & \text{caso contrário} \end{cases}$$

Variáveis de interesse

- Total de desempregados na região i: $y_{i.} = \sum_j y_{ij}$
- Proporção de desemprego na região i: $p_i = y_{i.}/N_i$
- Probabilidade do indivíduo j da região i estar desempregado: p_{ij}
- Total de desempregados, empregados e inativos na região i: $z_i = (z_{i1}, z_{i2}, z_{i3})$

Variáveis auxiliares

Considere-se que estão disponíveis q variáveis auxiliares $x = (x_1, ..., x_q)$ e x_{ij} representa o vetor auxiliar para o j-ésimo indivíduo da NUTS III i.

Design-based

Considere-se uma amostra $s = (s_1, ..., s_{28})$ extraída da população U a partir de um esquema de seleção probabilístico.

O estimador direto, também conhecido como estimador Horvitz-Thompson, para o total de desempregados na i-ésima região é dado por:

$$\hat{y}_{i.}^{HT} = \sum_{j \in s_i} w_j y_{ij}$$

onde w_j é o inverso da probabilidade de inclusão do indivíduo j na amostra s_i segundo o esquema de amostragem usado.

O estimador direto para a proporção de desempregados na região i é dado por:

$$\hat{p}_i^{HT} = \sum_{j \in s_i} w_j y_{ij} / N_i$$

Model-based

Total de desempregados na região i

- Modelo Fay-Herriot
- Modelo Fay-Herriot espacial
- Modelo Fay-Herriot espaço-temporal

Proporção de desemprego na região i

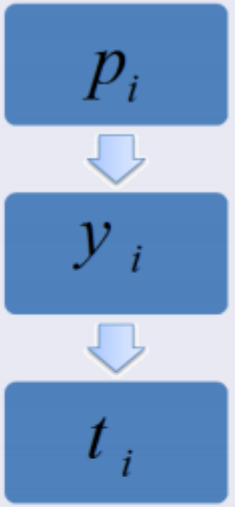
- Modelo de regressão logística (nível área)
- Modelo Beta

Probabilidade do indivíduo j da região i estar desempregado

- Modelo de regressão logística (nível indivíduo)

Total de desempregados, empregados e inativos na região i

- Modelo Multinomial



Modelo Fay-Herriot

Supõe-se que se pretende modelar o total de desempregados na região i.
O modelo de Fay-Herriot (1979) estabelece a seguinte relação

$$y_{i.} = x'_{i.}\beta + u_i, \ i = 1, ..., 28 \qquad (1)$$

onde $\beta = (\beta_1, ..., \beta_p)'$ é o vetor $p \times 1$ dos coeficientes de regressão e $u'_i s$ são efeitos aleatórios específicos da área assumidos i.i.d. com valor médio 0 e variância σ^2_u.

Assume-se que o estimador direto de $y_{i.}$, $\hat{y}^{HT}_{i.}$ pode ser expresso como

$$\hat{y}^{HT}_{i.} = y_{i.} + \epsilon_i, \ i = 1, ..., 28 \qquad (2)$$

onde os erros amostrais ϵ_i são independentes com $E(\epsilon_i|y_{i.}) = 0$ e $V(\epsilon_i|y_{i.}) = \phi_i$ com variâncias amostrais ϕ_i conhecidas.

Combinando (1) e (2), o modelo Fay-Herriot pode ser expresso como

$$\hat{y}^{HT}_{i.} = x'_{i.}\beta + u_i + \epsilon_i, \ i = 1, ..., 28 \qquad (3)$$

em que u_i e ϵ_i são independentes.

Modelo Fay-Herriot Espacial

Supõe-se que se pretende modelar o total de desempregados na região i.

Singh et al (2005) consideraram uma extensão do modelo Fay-Herriot:

$$\hat{y}_{i.}^{HT} = x_{i.}'\beta + u_i + \epsilon_i,\ i = 1, ..., 28, \tag{4}$$

onde o vetor dos efeitos área $(u_1, ..., u_{28})'$ segue um processo SAR(1)

$$u_i = \rho_1 \sum_{l \neq i} w_{i,l} u_l + \epsilon_i,\ |\rho_1| < 1,\ \epsilon_i \sim N(0, \sigma_1^2),\ i = 1, ..., 28 \tag{5}$$

e $W = (w_{i,l})$ é a matriz de proximidade standardizada por linha.

Modelo Fay-Herriot Espaço-Temporal

Supõe-se que se pretende modelar o total de desempregados na região i.

Marhuenda et al (2013) introduziram uma extensão do modelo Fay-Herriot:

$$\hat{y}_{i.t}^{HT} = x'_{i.t}\beta + u_{1i} + u_{2it} + \epsilon_{it},\ i = 1, ..., 28,\ t = 1, ..., T \qquad (6)$$

Os vetores dos efeitos aleatórios área-tempo $(u_{2i1}, ..., u_{2iT})'$ são i.i.d. para cada área i e seguem um processo AR(1)

$$u_{2it} = \rho_2 u_{2i,t-1} + \epsilon_{2it},\ |\rho_2| < 1,\ \epsilon_{2it} \sim N(0, \sigma_2^2) \qquad (7)$$

O vetor dos efeitos área $(u_{1,1}, ..., u_{1,28})'$ segue um processo SAR(1)

$$u_{1i} = \rho_1 \sum_{l \neq i} w_{i,l} u_{1l} + \epsilon_{1i},\ |\rho_1| < 1,\ \epsilon_{1i} \sim N(0, \sigma_1^2),\ i = 1, ..., 28 \qquad (8)$$

onde $W = (w_{i,l})$ é a matriz de proximidade standardizada por linha.

Modelo de regressão logística

Suponha-se agora que se pretende modelar a proporção de desemprego por região i. Os parâmetros de interesse são as proporções $p_i = \sum_j y_{ij}/N_i$ nas pequenas áreas.

Assume-se que $(y_{ij}|p_{ij}) \overset{iid}{\sim} \text{Bernoulli}(p_{ij})$, onde p_{ij} representa a probabilidade do indivíduo j da região i estar desempregado.

MacGibbon and Tomberlin (1989) usaram um modelo de regressão logística com efeitos aleatórios específicos da área para estimar p_i:

$$\text{logit}(p_{ij}) = \log\frac{p_{ij}}{1-p_{ij}} = x_{ij}^T\beta + v_i \tag{9}$$

onde $v_i \sim N(0, \sigma_v^2)$ e x_{ij} são covariáveis específicas da unidade.

Um modelo mais simples assume que $(y_{ij}|p_i) \overset{iid}{\sim} \text{Bernoulli}(p_i)$ onde

$$\text{logit}(p_i) = \log\frac{p_i}{1-p_i} = x_i^T\beta + v_i \tag{10}$$

Alternativamente, pode assumir-se que p_i segue uma distribuição beta. Da-Silva e Migon (2011) sugerem um modelo beta temporal.

Coerência ao nível NUTS II

Quando agregadas ao nível NUTS II, as estimativas obtidas devem coincidir com as estimativas oficiais.

Coerência

Seja $\hat{y}_k^{INE}$ o estimador do total y_k na NUTS II k. Assuma-se que a NUTS II k é uma partição de D_k NUTS III.
O estimador

$$\hat{y}_i^c = \frac{\hat{y}_k^{INE}}{\sum_{i=1}^{D_k} \hat{y}_i} \hat{y}_i$$

satisfaz $\hat{y}_k^{INE} = \sum_{i=1}^{D_k} \hat{y}_i^c$.

Resultados

Estimação do total de desempregados por NUTS III para o 1º trimestre de 2012

Aplicação dos seguintes métodos:

- Direto
- FH (modelo Fay-Herriot)
- STFH1 (modelo Fay-Herriot espacial)
- STFH2 (modelo Fay-Herriot espaço-temporal)

Informação auxiliar: número de desempregados registados nos Centros de Emprego.

Package do software R utilizado: sae.

Resultados

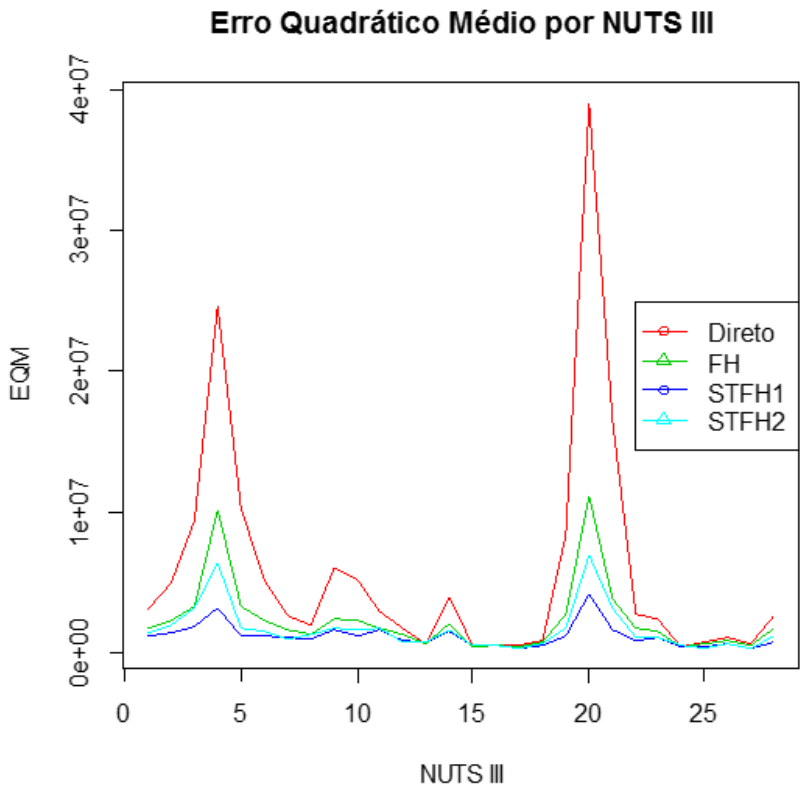


Figura 3: Erro Quadrático Médio das estimativas do total de desempregados por NUTS III para o 1º trimestre de 2012

Resultados

Estimação da proporção de desemprego por NUTS III para o 1º trimestre de 2012

Aplicação dos seguintes métodos:

- Modelo de regressão logística
- Modelo beta

Informação auxiliar: número de desempregados registados nos Centros de Emprego.

Packages do software R utilizados: lme, betareg.

Resultados

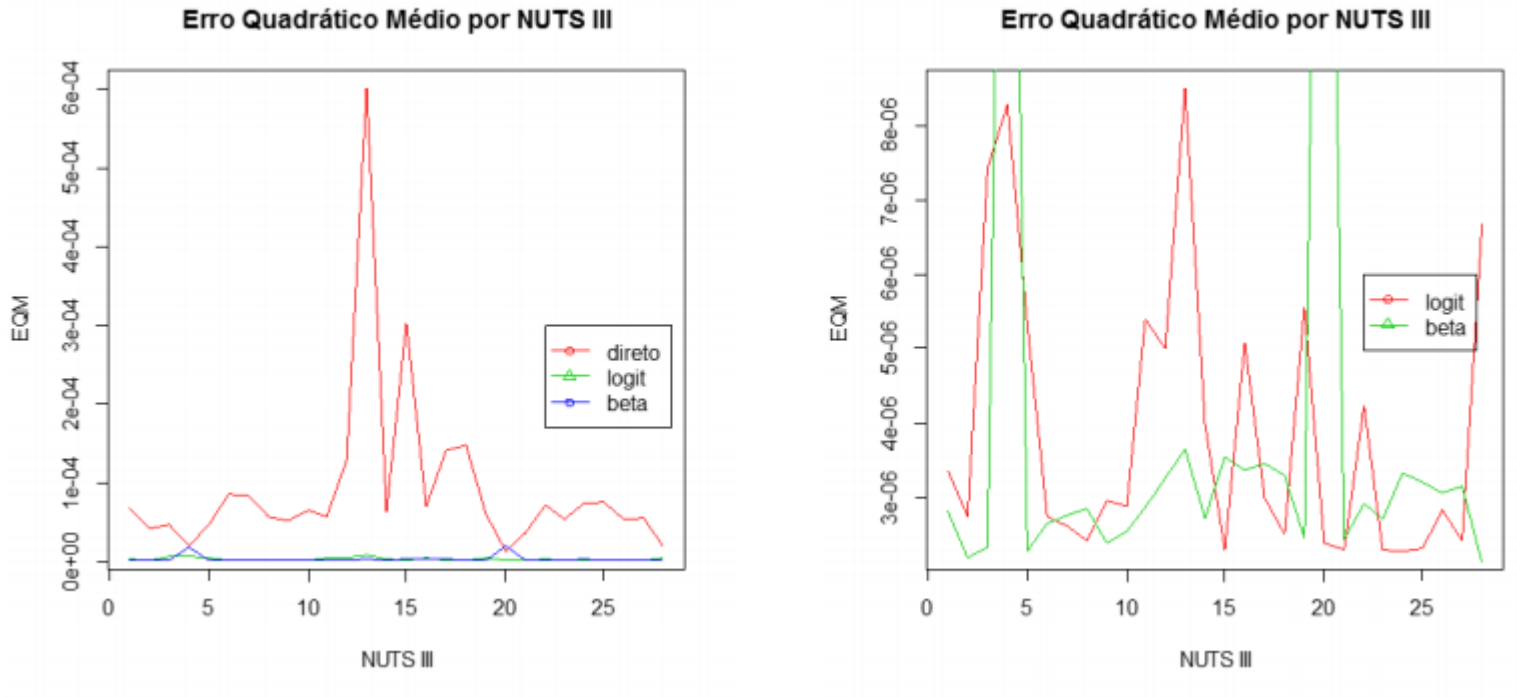


Figura 4: Erro Quadrático Médio das estimativas da proporção de desemprego por NUTS III para o 1º trimestre de 2012

Referências

- Rao, J. N. K., *Small Area Estimation*, Wiley, 2003.
- Molina, I., Saei, A., Lombardía, M. J. (2007). *Small area estimates of labour force participation under a multinomial logit mixed model*, Journal of the Royal Statistical Society, Ser. A, 170, Part 4, pp. 975-1000.
- Fernández, F. S., Sánchez-Romero, M., Grijalba, J. V. (2007). *Employment and Unemployment Transitions in Spain from 1996 to 2005*, Economic Analysis - Working Paper Series.
- Fay, R. E., Herriot, R. A. (1979). *Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data*, Journal of the American Statistical Association, Vol. 74, No. 366, pp. 269-277.
- Marhuenda, Y., Molina, I., Morales, D. (2013). *Small area estimation with spatio-temporal Fay-Herriot models*, Computational Statistics and Data Analysis, 58, pp. 308-325.
- Lopez-Vizcaíno, E., Lombardía, M. J., Morales, D. (2013). *Multinomial-based small area estimation of labour force indicators*, Statistical Modelling, Vol. 13, pp. 153-178.
- Da-Silva, C. Q., Migon, H. S. (2011). *Hierarchical dynamic Beta model*, Artigo submetido para publicação.