

ESTIMACIÓN DO GASTO EN CONSUMO FINAL DOS FOGARES NAS COMARCAS GALEGAS

Suárez Morao, Marta¹, López Vizcaíno, M.E¹, Jácome Rodríguez, R¹

¹ Instituto Galego de Estatística (IGE)

RESUMO

O obxectivo deste traballo é estimar o gasto en consumo final dos fogares galegos por comarcas para o período 2006 a 2008, empregando técnicas de estimación en áreas pequenas. En concreto, empregárase a teoría baseada en modelos e utilizarase un modelo de área lineal mixto con efectos aleatorios de área e de tempo.

Palabras e frases clave áreas pequenas, modelos lineais mixtos, gasto en consumo final

1. INTRODUCCIÓN

Na operación estatística “Contas económicas anuais” elaborada polo Instituto Galego de Estatística (IGE) estímase o agregado Gasto en consumo final dos fogares da Comunidade Autónoma de Galicia. Esta estimación realízase seguindo a metodoloxía e os criterios establecidos polo Sistema Europeo de Contas 1995 (SEC 95). Este regulamento europeo define o Gasto en Consumo Final como aquel gasto realizado polas unidades institucionais residentes en bens e servizos que se empregan para satisfacer directamente as necesidades ou carencias individuais, ou as necesidades colectivas dos membros da comunidade (SEC 95 3.75). A fonte de información principal para obter esta estimación é a “Encuesta de Presupuestos Familiares” (EPF)¹, elaborada polo Instituto Nacional de Estadística (INE) e que ofrece información a nivel de comunidade autónoma.

Por outra banda, o IGE tamén elabora a operación estatística “Contas de distribución da renda dos fogares” a cal ofrece información sobre a Renda Dispoñible Bruta (RDB) dos fogares para as comarcas e os concellos galegos. O obxectivo que se plantea dende o IGE é poder chegar a ter o Aforro dos fogares a nivel inferior á comunidade autónoma para o cal sería necesario ter unha estimación do gasto en consumo final dos fogares por comarcas e concellos.

O deseño mostrai da principal fonte de información, a EPF, non está preparado para dar información a nivel inferior a comunidade autónoma. De feito, a maioría dos concellos non están representados na mostra, e os que si o están teñen un tamaño mostrai moi reducido, o que fai que as estimacións a nivel municipal ou comarcal teñan unha precisión inaceptable. Por outra parte, hai que considerar que a EPF é unha enquisa anual levada a cabo por un instituto de estatística oficial. A dispoñibilidade de series de tempo, se se usan dunha maneira apropiada, permite un incremento significativo da mostra total do dominio en estudo. Isto debería levar a unha mellora no poder predictivo dos estimadores.

Ante esta situación xorde a posibilidade do emprego de técnicas de estimación en áreas pequenas. A estimación en áreas pequenas é unha parte da estatística que combina información mostrai e a inferencia en poboacións finitas con modelos estatísticos. Pódese ver a descrición desta teoría na monografía de Rao(2003) ou máis recentemente no artigo de Jiang(2006). Neste traballo propónse o uso de estimadores EBLUP baseados nun modelo lineal mixto con efectos de tempo correlacionados. Estes modelos xa foron introducidos na literatura por Rao e Yu (1994) e Saei e Chambers (2003). Co obxectivo de avaliar a acuracidade das estimacións tamén se calcularán os coeficientes de variación das estimacións obtidas empregando para elo a aproximación analítica de Prasad e Rao (1990).

¹ A EPF é a fonte estatística fundamental para obter a estimación do agregado do gasto en consumo final dos fogares cunha excepción importante: o gasto en alugueiro imputado que non se obtén desta enquisa. O gasto en alugueiro imputado escapa do alcance deste traballo, dado que xa se dispón dunha estimación interna a nivel municipal.

2. DESCRICIÓN DO MÉTODO

Co obxectivo de obter unha estimación do gasto medio por fogar para cada unha das 53 comarcas galegas axustouse o seguinte modelo lineal mixto de área con efectos de tempo correlacionados:

$$Y_{dt} = X_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1, \dots, D, \quad t = 1, \dots, T$$

Onde Y_{dt} é o estimador directo do gasto medio na área d e no ano t obtido da EPF, X_{dt} é un vector coas variables auxiliares na área d e no ano t , u_{dt} é o efecto aleatorio da área d no ano t . Supoñemos que os efectos aleatorios de área seguen procesos AR(1) independentes: $u_{dt} = \rho u_{dt-1} + v_t$, sendo $v_t \sim N(0, \sigma_u^2)$; os erros do modelo e_{dt} supóñense independentes con distribución $N(0, \sigma_{edt}^2)$ e independentes dos efectos aleatorios de área, as varianzas σ_{edt}^2 supóñense coñecidas e estimanse de acordo coa aproximación de Horvitz-Thomson:

$$\sigma_{edt}^2 = \frac{1}{(\sum_{j \in S_d} w_{jt})^2} \sum_{j \in S_d} w_{jt} (w_{jt} - 1) y_{jt}^2 ; \text{ onde } S_d \text{ é a mostra}$$

pertencente á comarca d ; y_{jt} é o gasto do fogar j no ano t e w_{jt} é o peso mostral asignado ao fogar j no ano t

En notación matricial o modelo é:

$$Y = X\beta + Zu + e$$

Sendo:

D =número de áreas; T =número de períodos; $M=D \times T$; p =nº de variables auxiliares

Y un vector de tamaño M , que contén os estimadores directos para cada área e período temporal

X unha matriz $M \times p$, cuxas columnas conteñen as variables auxiliares para cada área e período temporal

Z é a matriz identidade

Asúmese que u ten distribución $N(0, V_u)$, e ten distribución $N(0, V_e)$, son independentes entre si con matrices de varianzas-covarianzas:

$$V_e = \text{diag}(\sigma_{edt}^2)$$

$$V_u = \sigma_u^2 \Omega(\rho) \quad \text{con } \Omega(\rho) = \text{diag} \left(\frac{1}{1-\rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-3} & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{T-2} & \rho^{T-3} & \dots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \dots & \rho & 1 \end{pmatrix} \right)$$

Por tanto, $V = \text{Var}(Y) = V_e + V_u$

As compoñentes da varianza descoñecidas, σ_u^2 e ρ estimanse polo método de máxima verosimilitude restrinxida usando un algoritmo de Fisher-scoring.

Unha vez obtidas as compoñentes da varianza, os efectos fixos e aleatorios obtéñense a partir das expresións:

$$\hat{\beta} = (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} Y$$

$$\hat{u} = \hat{V}_u Z' \hat{V}^{-1} (Y - X \hat{\beta})$$

A partir destas estimación obtense o EBLUP do gasto medio por fogar $\hat{Y}_{dt} = X_{dt} \hat{\beta} + \hat{u}_{dt}$.

Como medida da acuracidade destas estimacións, empregouse a conocida aproximación ao error cuadrático medio introducida por Prasad e Rao no ano 1990 e que ten unha forma explícita do tipo $G_1 + G_2 + 2G_3$.

3. DESCRICIÓN DA INFORMACIÓN EMPREGADA

A EPF é unha operación estatística de carácter anual realizada polo Instituto Nacional de Estatística (INE). Esta enquisa emprega unha mostraxe bietápica con estratificación das unidades de primeira etapa, seleccionando unha mostra independente dentro de cada comunidade autónoma (CCAA). As unidades de primeira etapa son as seccións censuais nas que se atopa dividido o territorio nacional no momento da enquisa e as unidades de segunda etapa as vivendas familiares principais. En cada CCAA as unidades de primeira etapa estratifícanse seguindo un criterio xeográfico, que asigna o estrato segundo o tamaño do municipio ao que pertence a sección. As seccións selecciónanse dentro de cada estrato con probabilidade proporcional ao tamaño poboacional. As vivendas selecciónanse, dentro de cada sección, con igual probabilidade mediante mostraxe sistemático con arranque aleatorio.

Dado que as seccións se seleccionan dentro de cada estrato non aleatoriamente, senón con probabilidade proporcional ao tamaño poboacional, a mostra distribúese en Galicia sempre, nos catro anos dos cales dispoñemos de información (2006 a 2009) nos mesmos concellos. O tamaño mostral na nosa comunidade rolda os 1300 fogares. Polo tanto, os tamaños mostrais por comarca son moi reducidos. En máis da metade das mesmas, non acadan os 20 fogares e, unicamente, nas comarcas da Coruña e Vigo superan os 100 fogares. Existen ademais oito comarcas das cales non se dispón de dato e, que por tanto quedan fóra do estudo.

Os estimadores directos proporcionados pola EPF constitúen a variable resposta do modelo axustado e como información auxiliar empregouse información da “Enquisa de condicións de vida das familias”, da operación “Contas de distribución da renda dos fogares por concello”, da matriculación de vehículos rexistrada pola Dirección General de Tráfico e dos datos de alumnado matriculado nos distintos centros de ensino.

A Enquisa de condicións de vida das familias é unha actividade estatística anual que realiza o IGE dende 1999 e que estuda diferentes características socioeconómicas dos fogares galegos. No referente ao presente estudo, son de interese as variables referentes a certos gastos correntes dos fogares, en concreto, o gasto en telefonía, luz, auga, gas, servizo doméstico e persoal adicado ao coidado de maiores e/ou nenos. As categorías de gastos investigados na ECV non cobren a totalidade do gasto en consumo final dos fogares, pero si representan un bo indicador dos mesmos. A partir de esta información obtemos a primeira variable auxiliar do modelo: o gasto medio ECV por fogar e comarca.

Outra variable auxiliar do modelo é a renda dispoñible bruta dos fogares obtida da operación “Contas de distribución da renda dos fogares por concellos”, elaborada polo IGE. Esta macromagnitude mide a renda da que dispoñen os fogares ao longo dun ano para destinar ao consumo ou aforro e reflicte a capacidade adquisitiva dos fogares como resultado dos procesos de produción, asignación e distribución da renda.

Por último, tamén se utilizaron como variables auxiliares os datos referentes ao número de vehículos matriculados en cada comarca galega, información que proporciona mensualmente a *Dirección General de Tráfico (DGT)*, e os datos de matriculación de alumnos nos distintos centros de ensino proporcionados pola *Consellería de Educación e Ordenación Universitaria*.

4. RESULTADOS

Nas Figuras 1,2 e 3 represéntase as estimacións basadas no modelo do gasto medio mensual por fogar (esquerda) e o índice do gasto medio de cada comarca sobre o gasto medio de Galicia (dereita) para o período 2006 a 2008.

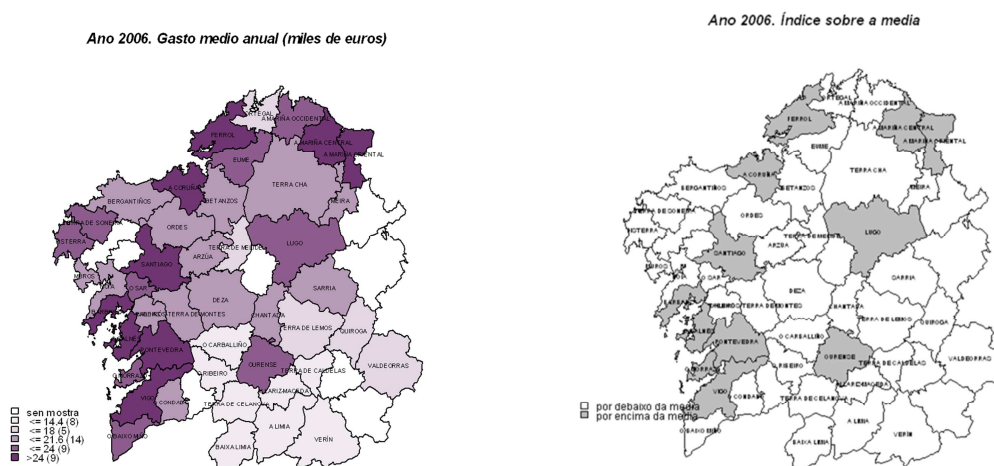
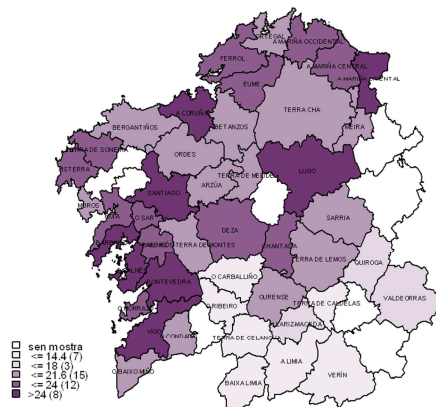


Figura 1: Ano 2006. Gasto medio mensual por comarca

Ano 2007. Gasto medio anual (miles de euros)



Ano 2007. Índice sobre a media

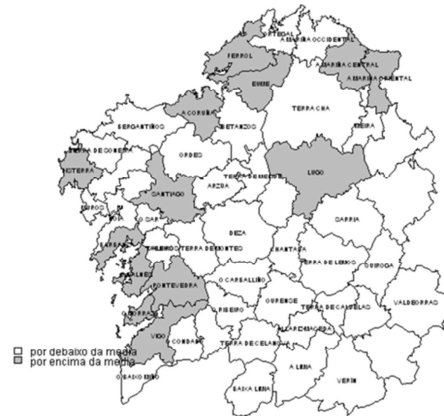
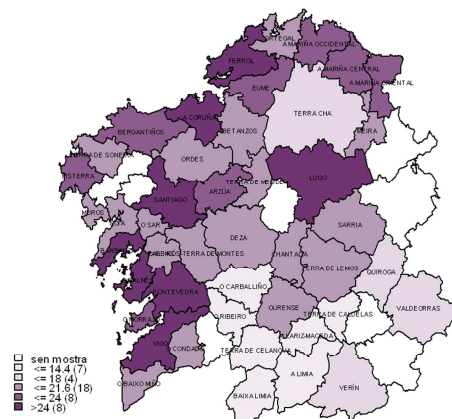


Figura 2: Ano 2007. Gasto medio mensual por comarca

Ano 2008. Gasto medio anual (miles de euros)



Ano 2008. Índice sobre a media

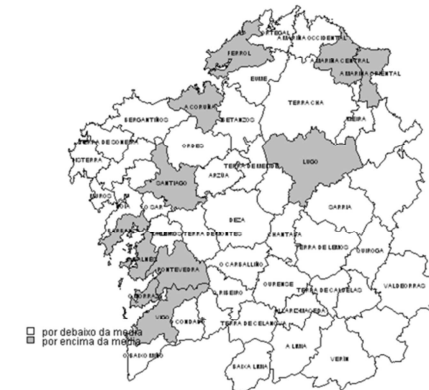


Figura 3: Ano 2008. Gasto medio mensual por comarca

Das catro provincias galegas os valores máis baixos do gasto medio atópanse na provincia de Ourense e os valores máis altos nas provincias da Coruña, Pontevedra e no norte da provincia de Lugo. No referente ao índice sobre a media obsérvase que, en termos xerais, as comarcas das 7 grandes cidades galegas, a Mariña central e Oriental e as comarcas da Barbanza, O Salnés e O Morrazo teñen un gasto superior á media galega.

No referente á acuracidade dos resultados, atopamos unha mellora significativa nos coeficientes de variación dos estimadores do gasto con respecto aos correspondentes estimadores directos, así como unha diminución da variabilidade do gasto por comarca e unha maior estabilidade nas estimacións anuais de cada área, como pode observarse nos gráficos da Figura 4. Todos eles están ordenados por tamaño mostral crecente das comarcas. Como é natural, a medida que o tamaño mostral é maior, os estimadores obtidos empregando o modelo e os estimadores directos son máis similares.

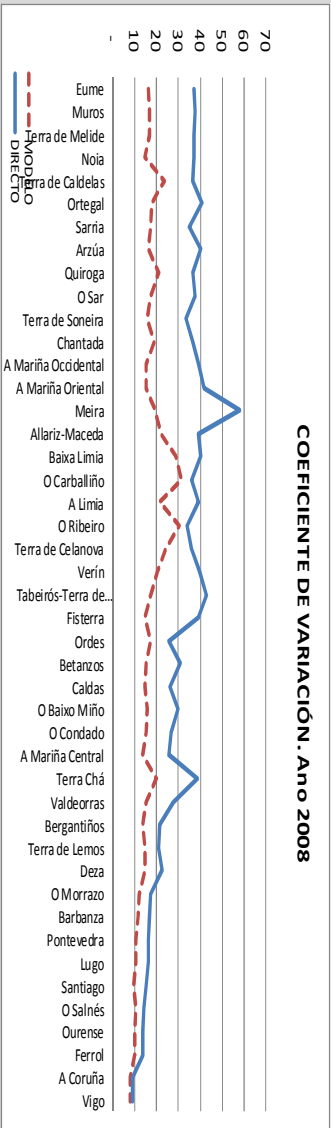
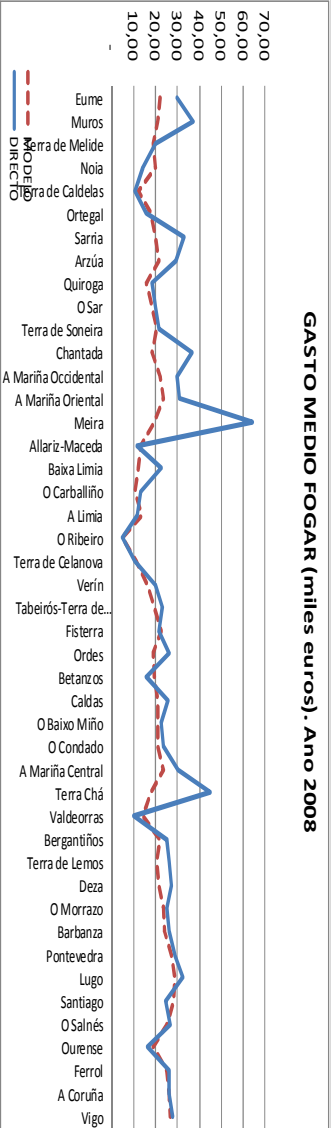
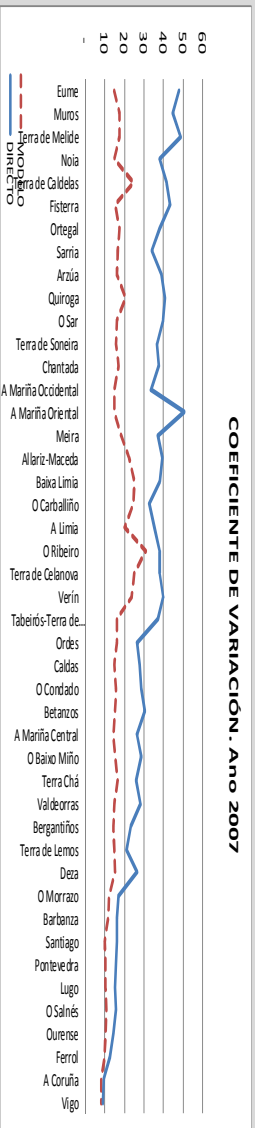
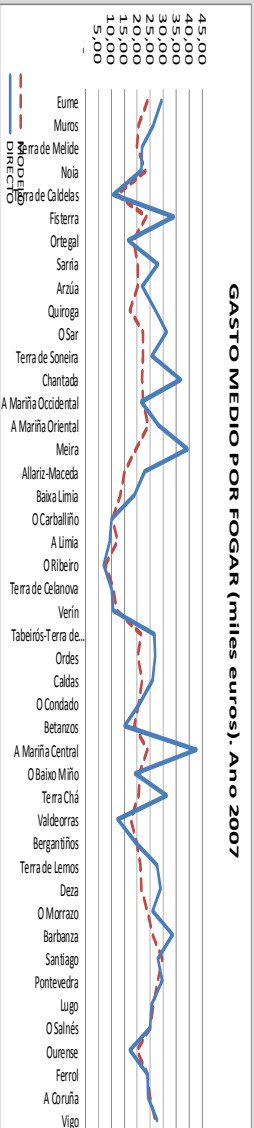
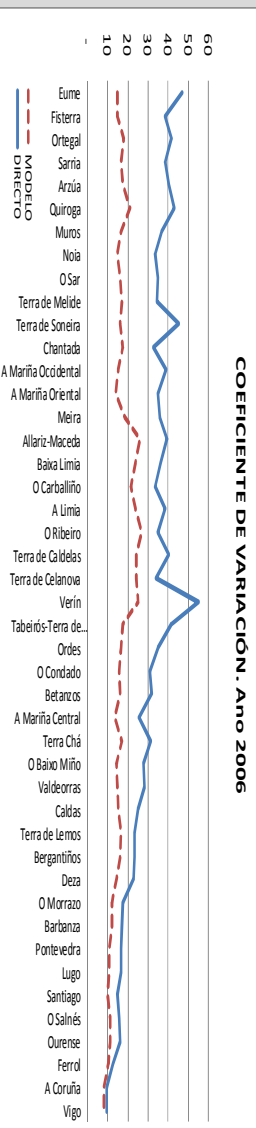
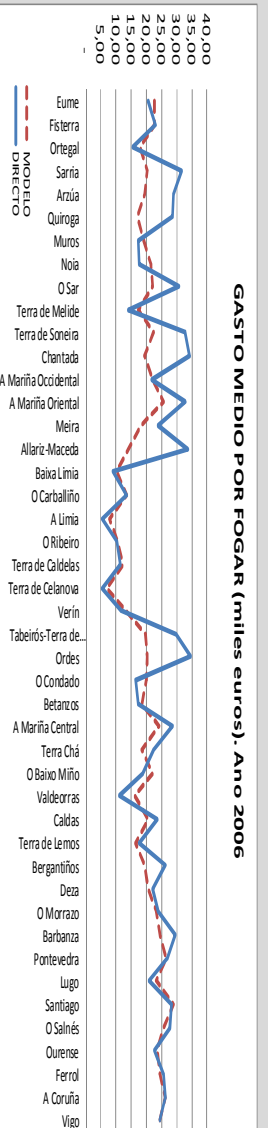


Figura 4: Ano 2006 a 2008. Gastos medios mensuais por fogar e coeficientes de variación

3. CONCLUSIONES

Neste estudo puidemos comprobar as vantaxes do uso das técnicas de estimación en áreas pequenas para a estimación do gasto en consumo final dos fogares nas comarcas galegas. Os estimadores obtidos teñen un erro cuadrático medio moi inferior aos correspondentes estimadores directos que se obteñen a partir da EPF. É máis, a maior parte dos coeficientes de variación son inferiores ao 20%, cota establecida pola Office for National Statistics (ONS) do Reino Unido para considerar un dato publicable.

No que respecta aos resultados obtidos do gasto nas comarcas, e como en moitos outros estudos que se fan sobre Galicia, existe unha diferenza salientable entre a Galicia costeira e a Galicia interior. Neste caso e en termos xerais, obsérvase un maior gasto por fogar nas comarcas das provincias de Coruña e Pontevedra e o norte de Lugo e un menor gasto no sur de Lugo e na provincia de Ourense.

REFERENCIAS

Eurostat (1996). Sistema Europeo de Cuentas. SEC 1995. Oficina de Publicaciones Oficiales de las Comunidade Europeas.

González-Manteiga, W., Lombardía, M.J., Molina, I., Morales, D. and Santamaría, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model.

Jiang, J. and Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, 15, 1-96.

ONS (2004). Office for national statistics. labour force survey user guide, vol 6. Disponible en http://www.statistics.gov.uk/downloads/theme_labour/Vol6.pdf.

Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85, 163-171.

Rao, J.N.K. (2003). *Small area estimation*. John Wiley and Sons, New-York.

Rao, J.N.K. and Yu, M. (1994). Small Area Estimation by Combining Time Series and Cross-sectional Data, *Canadian Journal of Statistics*, 22, 511-528.

Saei, A. and Chambers, R. (2003). Small area estimation under linear and generalized linear mixed models with time and area effects. S3RI Methodology Working Paper M03/15, Southampton Statistical Sciences Research Institute.