

Autor
Roberto Domínguez Gómez

Directores
María José Lombardía Cortiña
Wenceslao González Manteiga

Introducción
Un área es considerada como área pequeña cuando la muestra en el área no es suficientemente grande para conseguir estimaciones directas fiables. Ejemplos de áreas pequeñas pueden ser áreas geográficas (estados, provincias, municipios, comarcas, distritos escolares), grupos socio-demográficos (grupos específicos por edad-sexo-raza) y otras subpoblaciones como un conjunto de firmas de empresas.

Un modo de afrontar el problema en áreas pequeñas es aumentar el tamaño muestral. Sin embargo, un aumento de tamaño de muestra lleva, entre otras cosas, a un aumento en el coste del estudio, una mayor carga de respuesta a los informantes y mayores errores ajenos al muestreo, lo que, en general, se trata de evitar. Otra alternativa es utilizar técnicas más complejas, asistidas y basadas en modelos. Las estimaciones basadas en el modelo están siendo de gran interés en los últimos años por sus buenos resultados. La inclusión de efectos aleatorios de área en el modelo es común en la estimación para áreas pequeñas. Estos efectos recogen la variación en las áreas que no está explicada por las variables auxiliares.

El objetivo en este trabajo es la estimación del ingreso medio mensual por hogar en las 53 comarcas de Galicia, en este caso las áreas pequeñas son las comarcas. La muestra disponible proviene de la Encuesta de Condiciones de Vida de las familias (ECV) del año 2005. El tamaño muestral varía desde 16 hogares en la comarca de Muros hasta 864 hogares en la comarca de Vigo con un total de 6304 hogares muestrales. Para compensar la falta de información muestral se va a utilizar información auxiliar disponible de la Agencia Estatal de la Administración Tributaria (AEAT).

Estimación basada en el diseño

• **Estimador Horvitz-Thompson:**

$$\hat{Y}_d^{HT} = \frac{\sum s_d \omega_j y_j}{\sum s_d \omega_j}, \quad d = 1, \dots, D$$

donde s_d es la muestra perteneciente a la comarca d , y_j el ingreso mensual del hogar j y ω_j el peso muestral asignado al hogar j .

• **Estimador postestratificado sintético:**

$$\hat{Y}_d^{POST,S} = \frac{1}{M_d} \sum_{h=1}^H M_{dh} \hat{Y}_h^{HT}, \quad d = 1, \dots, D$$

denotando por h los postestratos relevantes, M_d el total poblacional de las personas en la comarca d y M_{dh} el total poblacional de las personas que pertenecen a la intersección de la comarca d con el estrato h . Además definimos \hat{Y}_h^{HT} como el estimador de Horvitz-Thompson de la media mensual en el estrato h .

• **Estimador compuesto:**

$$\hat{Y}_d^{COMP,P} = \gamma_d \hat{Y}_d^{HT} + (1 - \gamma_d) \hat{Y}_d^{POST,S}, \quad d = 1, \dots, D$$

donde γ_d son pesos dependientes del tamaño muestral en la comarca d , definidos como:

$$\gamma_d = \begin{cases} 1 & \text{si } \hat{M}_d^{HT} \geq \delta M_d \\ \frac{\hat{M}_d^{HT}}{\delta M_d} & \text{en caso contrario} \end{cases}$$

con \hat{M}_d^{HT} la estimación de Horvitz-Thompson del total de las personas M_d en la comarca d . Para el parámetro δ se pueden tomar distintos valores, $\delta \in \{1, 3/2, 2\}$ (Eustat, 2008).

Estimación basada en el modelo

• **Modelo Fay-Herriot:**
El modelo se construye a partir de la información auxiliar disponible de la AEAT. La información está disponible solamente a nivel de comarca y por lo tanto la estimación se basa en el siguiente modelo de área (Fay and Herriot, 1979):

$$\hat{Y}_d^{HT} = \beta_0 + X_d \beta_1 + u_d + \varepsilon_d, \quad d = 1, \dots, D$$

Este modelo asume que los efectos aleatorios de área u_d son independientes e idénticamente distribuidos según $u_d \sim N(0, \sigma_u^2)$ y los errores de muestreo ε_d son independientes y distribuidos según $\varepsilon_d \sim N(0, \sigma_d^2)$, con u_d independientes de ε_d .

También se asume que la varianza σ_u^2 de los efectos aleatorios es desconocida y que las varianzas $\sigma_d^2 = s_d^2/n_d$ de los errores muestrales son conocidas, siendo s_d^2 y n_d la cuasivarianza muestral de la variable objetivo y el tamaño muestral en la comarca d , respectivamente.

Además, como covariable X_d se toma la renta media anual declarada por individuo en la comarca d (datos del IRPF).

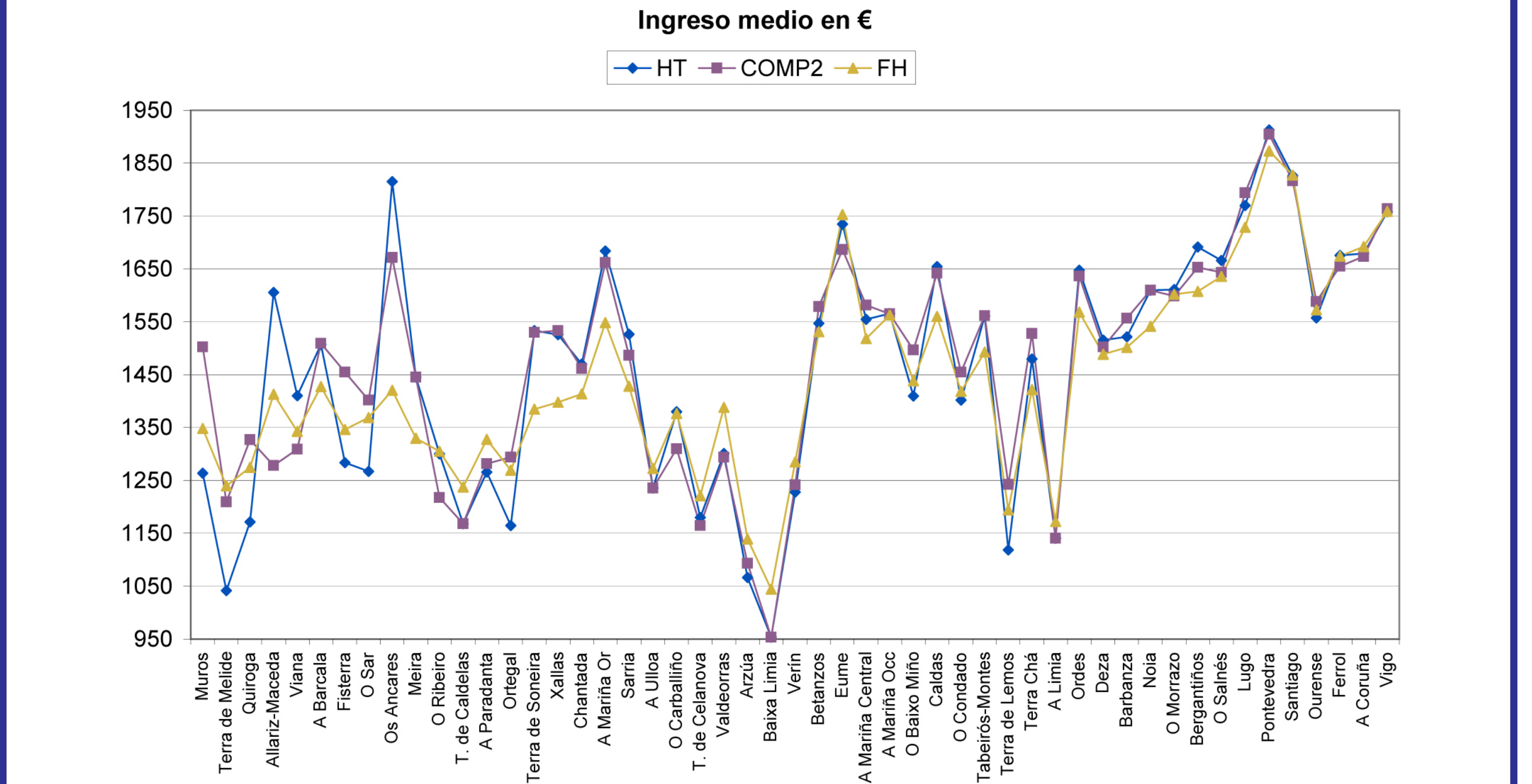
Basándose en este modelo, se define el **estimador Fay-Herriot**:

$$\hat{Y}_d^{FH} = \hat{\beta}_0 + X_d \hat{\beta}_1 + \hat{u}_d, \quad d = 1, \dots, D$$

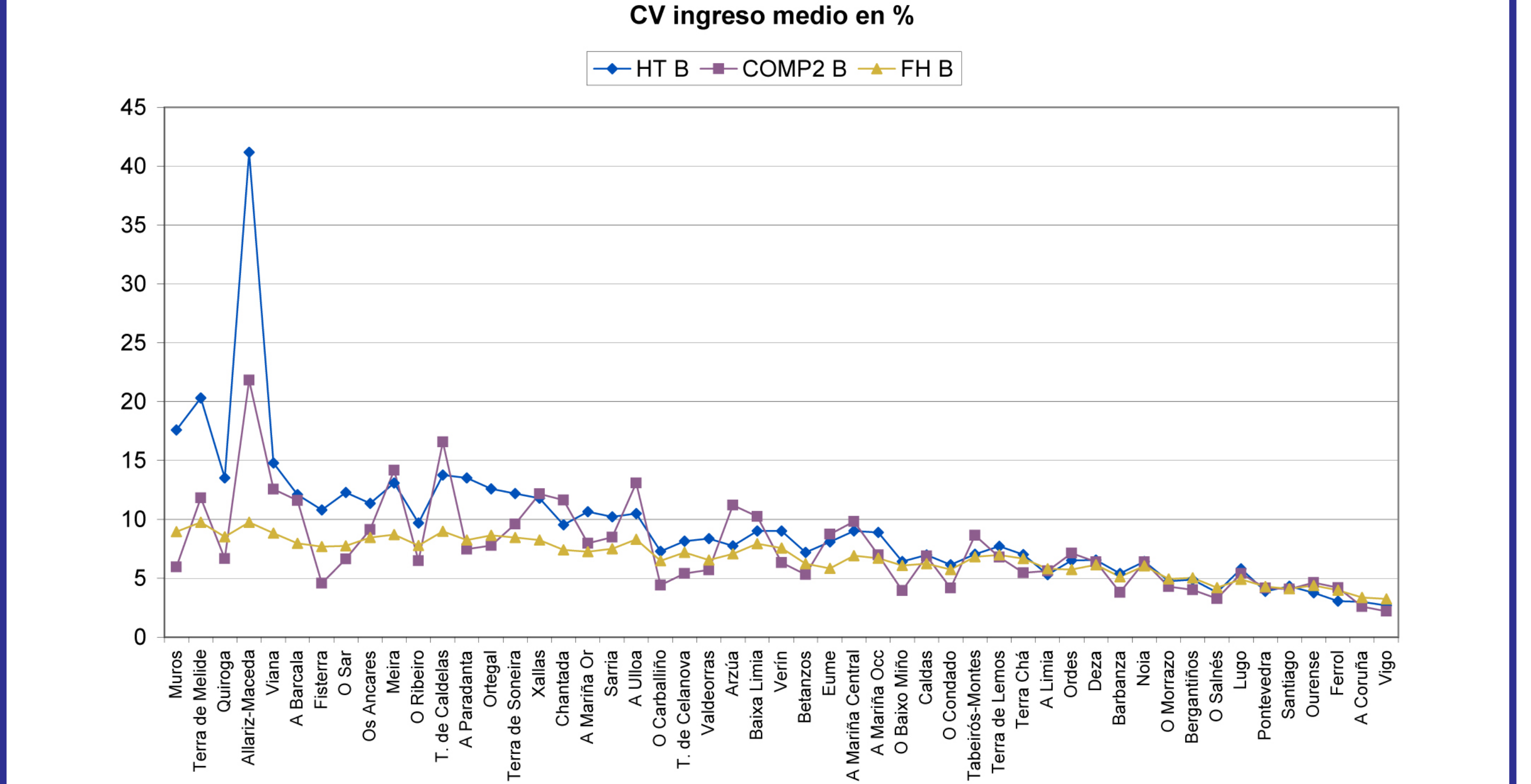
Los parámetros se estiman por máxima verosimilitud restringida (REML). Para más detalle, véase Rao (2003).

Aplicación a datos reales

Se aplican los estimadores Horvitz-Thompson (HT), compuesto (COMP2) y Fay-Herriot (FH) con su coeficiente de variación estimado (CV) para cada comarca. Es necesario destacar que los resultados presentados no se pueden considerar como estadística oficial.

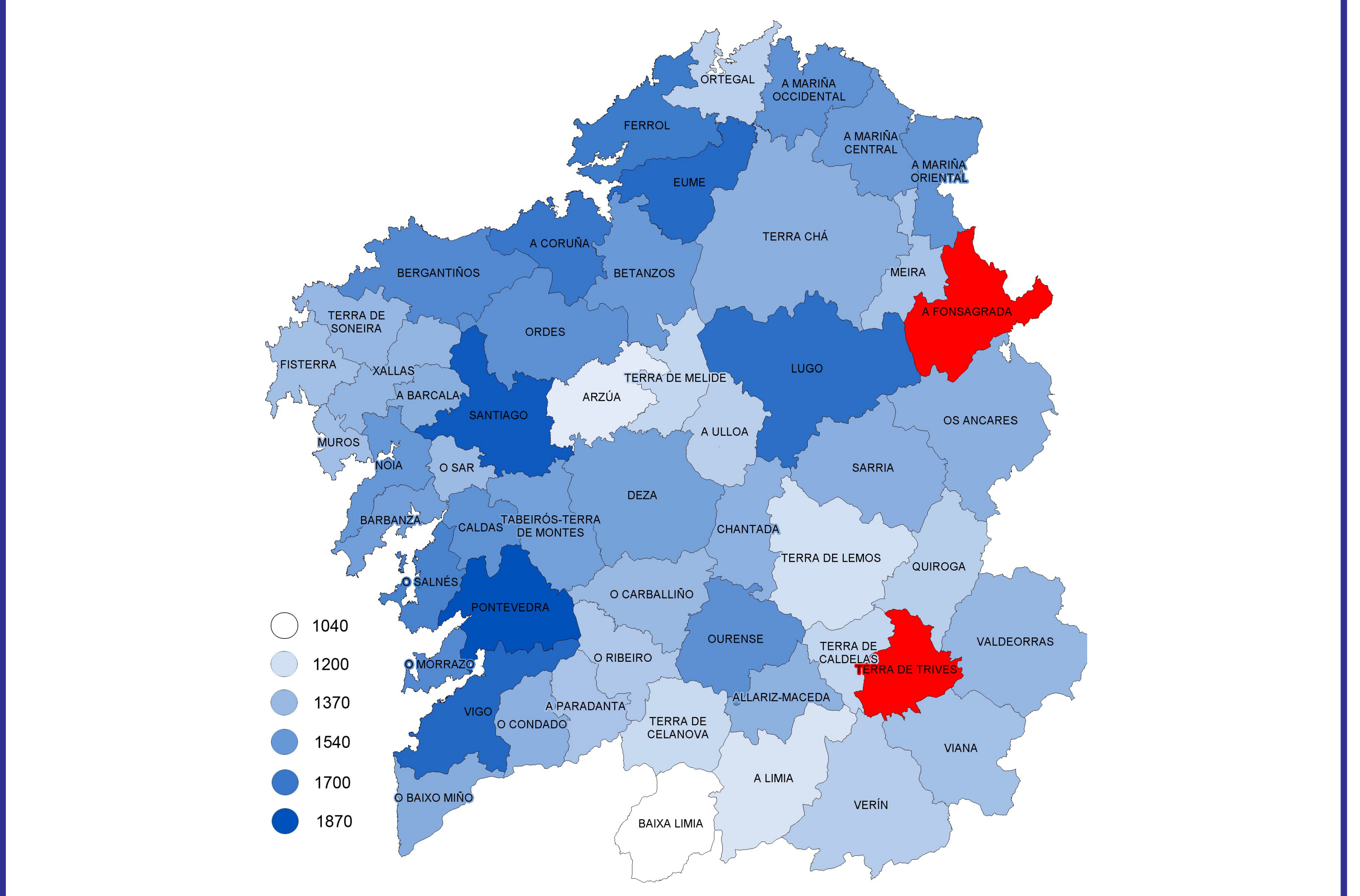


Las comarcas están ordenadas según su tamaño muestral. Existen diferencias entre los distintos estimadores del ingreso medio mensual. Por ejemplo, se observa una diferencia de 400 euros en la comarca de Os Ancares. Para estas comarcas se toma el estimador con menor CV.



En cuanto al CV se ve que el estimador Fay-Herriot obtiene los mejores resultados, sobre todo para las comarcas con menor tamaño muestral. Por ejemplo, se observa una diferencia de hasta 30% en el CV para la comarca de Allariz-Maceda a favor del estimador Fay-Herriot.

A continuación se representa geográficamente la estimación Fay-Herriot del ingreso medio mensual por hogar:



No se dispone de datos muestrales para las comarcas pintadas en color rojo.

Conclusión

Analizando los resultados obtenidos se puede concluir que el estimador Fay-Herriot basado en el modelo es un serio competidor de los estimadores clásicos basados en el diseño.

Referencias

- Eustat (2008). Estimación de áreas pequeñas en la encuesta de población en relación con la actividad de la C.A. de Euskadi. *Eustat*.
- Fay, R. E. and 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*, 74, 269-277.
- Rao, J. N. K. (2003). *Small Area Estimation*. Wiley, New York.