



UNIVERSIDADE DA CORUÑA *Universidade de Vigo*

# Estimación en áreas pequenas de medidas de pobreza

MÁSTER EN TÉCNICAS ESTADÍSTICAS  
TRABALLO DE FIN DE MÁSTER. CURSO 2011/2012

AUTORA: Marta Suárez Morao

DIRECTORA: María José Lombardía Cortiña



## Grazas!!

A María José Lombardía, a titora deste proxecto, por toda a axuda e o tempo que me prestou durante a elaboración do mesmo.

Ao Instituto Galego de Estatística (IGE) pola cesión dos microdatos da "Enquisa de condicións de vida das familias". Cumprindo co compromiso adquirido na solicitude dos datos, fago constar que o IGE se limitou unicamente a ceder os datos primarios, sen participar de ningunha outra maneira neste estudo, polo que non é responsable en ningún caso da utilización feita dos microdatos nin dos resultados amosados neste traballo.

A María Ginzo e a Ana Andión, grazas polos bos momentos e polo voso soporte nos menos bos.

Aos meus compañeiros, Raúl, Carmen, Sara, Trini, Eva, Jose, Sol, Estefanía, Lola, Ángeles, Pablo, Begoña e Esther, polo seu apoio e por facer que empece os días en in Mellorable compañía.

Ás miñas amigas de sempre, Natalia e Cruz, polo seu apoio incondicional e sincero en todo momento.

A todos, grazas!!

*Adicado á miña irmá Tania*

---



## Resume

O obxectivo deste traballo é profundizar nas técnicas de estimación en áreas pequenas incidindo na súa aplicación no estudo da pobreza.

Á hora de obter indicadores de pobreza a un nivel de desagregación xeográfica inferior ao nacional ou rexional, os estimadores directos obtidos de enquisas presentan importantes eivas. Isto é debido a que as enquisas das cales se obtén a información de base non están deseñadas para ser explotadas a ese nivel de desagregación, consecuentemente, os estimadores directos presentan elevados erros de mostraxe. É nestes casos, cando se debe recorrer ás técnicas de estimación en áreas pequenas e aos métodos de estimación indirectos. Hai que ter en conta, ademais, que a maior parte das medidas de pobreza son funcións non lineais de variables relativas ao ingreso ou o gasto dos fogares. Por este motivo, moitos dos métodos tradicionais no ámbito da estimación en áreas pequenas non son aplicables. Recentes artigos propoñen a aplicación de métodos de estimación baseados en modelos. Podendo empregarse modelos a nivel de área ou de individuo. Neste traballo farase unha revisión dos estudos feitos ata o momento, así como unha análise das posibilidades de aplicar estes modelos ou quizais propoñer algún novo para a súa aplicación ao estudo dos niveis de pobreza nas comarcas galegas, empregando como fonte primaria de información a "Enquisa de condicións de vida das familias" elaborada polo Instituto Galego de Estatística (IGE).

---



# Índice xeral

<b>1. Introducción</b>	<b>1</b>
1.1. Introducción . . . . .	1
<b>2. Medidas de pobreza</b>	<b>3</b>
2.1. A pobreza e a súa medición . . . . .	3
2.2. Elección do limiar de pobreza . . . . .	4
2.3. Medidas de pobreza. A familia de Foster Greer e Thorbecke . . . . .	5
<b>3. Medidas de pobreza en Galicia</b>	<b>9</b>
3.1. A Enquisa de condicións de vida das familias . . . . .	9
3.2. Definicións empregadas na ECV . . . . .	9
3.3. Ámbito da enquisa e unidades de análise empregadas . . . . .	11
3.4. Deseño mostral da ECV . . . . .	12
3.5. Información publicada . . . . .	14
<b>4. Estimadores</b>	<b>17</b>
4.1. Notación . . . . .	17
4.2. Estimador directo . . . . .	18
4.3. Estimador indirecto baseado no modelo de Fay-Herriot . . . . .	18
4.4. Estimador indirecto baseado nun modelo de individuo: EB-Method	22
4.5. Estimador indirecto baseado nun modelo de área con efecto tempo	25

<b>5. Estudio de simulación</b>	<b>31</b>
5.1. Simulación 1 . . . . .	31
5.1.1. Poboación simulada . . . . .	32
5.1.2. Avaliación dos estimadores propostos . . . . .	35
5.1.3. Avaliación dos estimadores do erro cuadrático medio . . . . .	38
5.2. Simulación 2 . . . . .	40
5.2.1. Poboación simulada . . . . .	40
5.2.2. Avaliación dos estimadores propostos . . . . .	41
5.2.3. Avaliación dos estimadores do erro cuadrático medio dos estimadores propostos . . . . .	44
5.2.4. Comparativa dos modelos con efecto tempo e o modelo de Fay-Herriot . . . . .	46
5.3. Conclusións xerais . . . . .	46
<b>6. Estimación de medidas de pobreza para as comarcas galegas</b>	<b>49</b>
6.1. Obxectivos . . . . .	49
6.2. Análise descritivo da información de base . . . . .	50
6.3. Estimadores indirectos baseados nun modelo a nivel de área . . . . .	54
6.3.1. Variables auxiliares . . . . .	54
6.3.2. Obtención do estimador de Fay-Herriot . . . . .	55
6.3.3. Obtención dos estimadores baseados nun modelo de área con efecto tempo (Modelo 0 e Modelo 1) . . . . .	59
6.4. Estimadores indirectos baseados nun modelo a nivel de individuo: estimador EB . . . . .	61
<b>7. Conclusións</b>	<b>75</b>
<b>Anexos</b>	<b>77</b>
<b>A. Tamaños mostrais e coeficientes de variación do estimador</b>	

---



<b>directo</b>	<b>79</b>
<b>B. Resultados para o estimador de Fay-Herriot</b>	<b>81</b>
<b>C. Resultados para o estimador do modelo 1</b>	<b>83</b>
<b>D. Resultados para o estimador do modelo 0</b>	<b>85</b>
<b>E. Resultados para o estimador EB</b>	<b>87</b>
<b>F. Software</b>	<b>89</b>
<b>Bibliografía</b>	<b>91</b>

---



# Índice de figuras

3.1. $FGT_0$ por grupos de idade para Galicia . . . . .	15
3.2. $FGT_0$ por sexo para Galicia . . . . .	16
3.3. $FGT_0$ por nacionalidade para Galicia . . . . .	16
5.1. Simulación 1: Nesgo relativo dos tres estimadores propostos . . . . .	36
5.2. Simulación 1: Raíz do erro cuadrático medio dos tres estimadores propostos . . . . .	36
5.3. Simulación 1: Erro cuadrático medio ( $ECM_d$ ) dos estimadores propostos . . . . .	37
5.4. Simulación 1: Valores poboacionais vs. estimadores indirectos propostos . . . . .	37
5.5. Simulación 1: Comparativa entre o estimador do erro cuadrático medio do estimador directo e o verdadeiro valor do erro cuadrático medio . . . . .	38
5.6. Simulación 1: Comparativa entre os estimadores do erro cuadrático medio do estimador Fay-Herriot e o verdadeiro valor do erro cuadrático medio . . . . .	39
5.7. Simulación 1: Comparativa entre os estimadores bootstrap e analítico do ECM do estimador de Fay-Herriot . . . . .	39
5.8. Simulación 1: Comparativa entre o estimador do erro cuadrático medio do estimador EB e o verdadeiro valor do erro cuadrático medio . . . . .	40
5.9. Simulación 2: Boxplot do $ECM_d$ dos modelos 0 e 1 con T=3 períodos para distintos valores de $\rho$ . . . . .	42
5.10. Simulación 2: Boxplot do $ECM_d$ dos modelos 0 e 1 con T=6 períodos para distintos valores de $\rho$ . . . . .	43

5.11. Simulación 2: Boxplot do $ECM_d$ dos modelos 0 e 1 con $T=10$ períodos para distintos valores de $\rho$ . . . . .	43
5.12. Simulación 2: $ECM_d$ dos modelos 0 e 1 con $T=3$ períodos para distintos valores de $\rho$ . . . . .	43
5.13. Simulación 2: $ECM_d$ dos modelos 0 e 1 con $T=6$ períodos para distintos valores de $\rho$ . . . . .	44
5.14. Simulación 2: $ECM_d$ dos modelos 0 e 1 con $T=10$ períodos para distintos valores de $\rho$ . . . . .	44
5.15. Simulación 2: ECM dos estimadores analítico e bootstrap do erro cuadrático medio do estimador do Modelo 1. $T = 3$ e $\rho = 0,5$ . . . .	45
5.16. Simulación 2: ECM dos estimadores analítico e bootstrap do erro cuadrático medio do estimador do Modelo 0. $T = 3$ e $\rho = 0,5$ . . . .	45
5.17. Simulación 2: ECM dos estimadores obtidos co Modelo 0, Modelo 1 e modelos Fay-Herriot independentes . . . . .	46
6.1. Ano 2007. Tamaños mostrais (esquerda) e coeficientes de variación (dereita) do estimador directo. . . . .	53
6.2. Ano 2008. Tamaños mostrais (esquerda) e coeficientes de variación (dereita) do estimador directo. . . . .	53
6.3. Ano 2009. Tamaños mostrais (esquerda) e coeficientes de variación (dereita) do estimador directo. . . . .	53
6.4. Ano 2007. Coeficientes de variación do estimador directo (esquerda) e do estimador de Fay-Herriot (dereita) . . . . .	58
6.5. Ano 2008. Coeficientes de variación do estimador directo (esquerda) e do estimador de Fay-Herriot (dereita) . . . . .	58
6.6. Ano 2009. Coeficientes de variación do estimador directo (esquerda) e do estimador de Fay-Herriot(dereita) . . . . .	58
6.7. Coeficientes de variación dos estimadores Fay-Herriot, Modelo 1 e Modelo 0 . . . . .	60
6.8. Ano 2009. Histograma (esquerda) e gráfico Q-Q (dereita) dos ingresos equivalentes . . . . .	62
6.9. Ano 2009. Histograma (esquerda) e gráfico Q-Q (dereita) da variable transformada . . . . .	63
6.10. Ano 2007. CV dos estimadores directo, Fay-Herriot e EB . . . . .	65

---

6.11. Ano 2008. CV dos estimadores directo, Fay-Herriot e EB . . . . .	66
6.12. Ano 2009. CV dos estimadores directo, Fay-Herriot e EB . . . . .	67
6.13. Ano 2007. Coeficientes de variación do estimador directo (esquerda) e do estimador EB (dereita) . . . . .	68
6.14. Ano 2008. Coeficientes de variación do estimador directo (esquerda) e do estimador EB (dereita) . . . . .	68
6.15. Ano 2009. Coeficientes de variación do estimador directo (esquerda) e do estimador EB (dereita) . . . . .	68
6.16. Ano 2007. Taxa de risco da pobreza obtida do estimador EB . . . .	70
6.17. Ano 2008. Taxa de risco da pobreza obtida do estimador EB . . . .	71
6.18. Ano 2009. Taxa de risco da pobreza obtida do estimador EB . . . .	72
6.19. Ano 2008. Renda bruta dispoñible por habitante (Euros/habitante)	73

---



# Índice de cadros

3.1. $FGT_0$ para Galicia e as súas provincias . . . . .	15
5.1. Simulación 1: Tamaños mostrais e e poboacionais . . . . .	33
5.2. Simulación 1: Parámetros das variables de Bernoilli empregadas para simular as variables auxiliares . . . . .	34
5.3. Distribución da raíz do erro cuadrático medio relativo . . . . .	37
6.1. Continuación da táboa . . . . .	51
6.1. Tamaños mostrais por comarca . . . . .	51
6.2. Variables relativas ás persoas . . . . .	51
6.3. Variables relativas ao fogar . . . . .	52
6.4. Limiar e taxa de risco de pobreza para Galicia . . . . .	52
6.5. Coeficientes de correlación coa taxa de risco de pobreza . . . . .	54
6.6. Coeficientes estimados e p-valores dos modelos de Fay-Herriot axustados . . . . .	56
6.7. Ano 2007. CV dos estimadores directo e Fay-Herriot nas comarcas de maior e menor tamaño de mostra . . . . .	56
6.8. Ano 2008. CV dos estimadores directo e Fay-Herriot nas comarcas de maior e menor tamaño de mostra . . . . .	57
6.9. Ano 2009. CV dos estimadores directo e Fay-Herriot nas comarcas de maior e menor tamaño de mostra . . . . .	57
6.10. Coeficientes estimados e p-valor dos modelos axustados . . . . .	59

---

6.11. Coeficientes estimados e p-valores dos modelos empregados para obter o estimador EB. . . . .	64
6.12. Ano 2007. CV dos estimadores directo, Fay-Herriot e EB nas comarcas de maior e menor tamaño de mostra . . . . .	65
6.13. Ano 2008. CV dos estimadores directo, Fay-Herriot e EB nas comarcas de maior e menor tamaño de mostra . . . . .	66
6.14. Ano 2009. CV dos estimadores directo, Fay-Herriot e EB nas comarcas de maior e menor tamaño de mostra . . . . .	67
1. Resultados co estimador directo . . . . .	79
1. Continuación da táboa . . . . .	80
2. Resultados co estimador de Fay-Herriot . . . . .	81
2. Continuación da táboa . . . . .	82
3. Resultados co estimador baseado no modelo 1 . . . . .	83
3. Continuación da táboa . . . . .	84
4. Resultados co estimador baseado no modelo 0 . . . . .	85
4. Continuación da táboa . . . . .	86
5. Resultados co estimador EB . . . . .	87
5. Continuación da táboa . . . . .	88

---



# Capítulo 1

## Introducción

### 1.1. Introducción

A demanda de información estatística a niveis de desagregación cada vez máis pequenos, en concreto, a nivel comarcal e municipal é crecente nos últimos anos. Sen embargo, as enquisas raramente están diseñadas para dar información a ese nivel. Máis aínda é difícil que en algún momento os deseños mostrais inclúan este nivel de desagregación, debido principalmente a razóns de tipo económico.

O problema que se plantexa é o da insuficiencia do número de puntos de mostra para obter unha estimación fiable en determinados ámbitos poboacionais non considerados nos deseños das enquisas. Unha posibilidade para solventar este problema é incrementar o tamaño mostral, sen embargo existen condicionantes como o económico que fan que esta vía non sexa factible na práctica. Unha opción alternativa é o emprego de métodos específicos de estimación en áreas pequenas. Moitos de estos métodos empregan ademais da información contida na propia enquisa, información adicional de outras fontes estatísticas e baséanse na existencia dun modelo de superpoboación que permite relacionalas. No contexto da estimación en áreas pequenas os subconxuntos poboacionais de tamaño inferior ao considerado no deseño da enquisa denomínanse pequenos dominios. No caso de dominios xeográficos utilízase alternativamente a denominación de pequeno dominio ou área pequena.

No presente traballo abordarase a estimación en áreas pequenas da familia de indicadores de pobreza de Foster, Greer e Thorbecke. Esta familia de indicadores e, en particular, a taxa de risco de pobreza, son unhas das medidas máis utilizadas para estudar a incidencia da pobreza nunha determinada poboación.

O traballo organízase da seguinte maneira:

No Capítulo 2 introdúcense os conceptos de pobreza e a súa medición, así como a familia de indicadores de Foster, Greer e Thorbecke.

No Capítulo 3 introdúcese as fontes de información que se empregan actualmente en Galicia para obter indicadores de pobreza. En particular, describirase a Enquisa de Condicións de Vida das familias (ECV), operación estatística da cal se obteñen os datos oficiais da taxa de risco de pobreza para a Comunidade Autónoma de Galicia. De esta maneira, situaremos o obxectivo final do presente estudo na obtención de estimacións precisas de medidas de pobreza para as comarcas galegas (nivel de desagregación non considerado no deseño da ECV). Con este fin, fíxose unha petición ao Instituto Galego de Estatística (IGE)<sup>1</sup>, organismo elaborador da enquisa, dos microdatos correspondentes aos últimos anos dispoñibles. Esta información será o punto de partida para obter estimacións comarcais empregando técnicas de estimación en áreas pequenas.

No Capítulo 4 propóñense estimadores alternativos para obter indicadores de pobreza en áreas pequenas. Propóñense, en primeiro lugar, dous estimadores: un baseado nun modelo a nivel de área e outro nun modelo a nivel de individuo. Nesta primeira proposta, non se considera a posibilidade de introducir o factor tempo nos correspondentes modelos. En segundo lugar, introdúcese un novo estimador alternativo baseado nun modelo de área con factor tempo.

No Capítulo 5 farase un estudo de simulación para avaliar os estimadores propostos.

Por último, no Capítulo 6 aplicaranse os estimadores propostos ao caso real galego, para o cal se empregará como fonte principal de información os microdatos da ECV referidos aos anos 2007, 2008 e 2009.

---

<sup>1</sup>O IGE non participou na elaboración deste traballo e, por tanto, non é responsable en ningún caso da utilización feita dos microdatos nin dos resultados amosados no mesmo

---

# Capítulo 2

## Medidas de pobreza

### 2.1. A pobreza e a súa medición

A pobreza é un fenómeno complexo que pode ser estudado dende diferentes perspectivas. En primeiro lugar, poderíamos diferenciar entre pobreza obxectiva e pobreza subxectiva atendendo ao tipo de información de base que se utilice. Os estudos de pobreza obxectiva empregan información recollida mediante variables cuxa medición provén da observación directa do investigador, o que lles confire un alto grao de obxectividade. En xeral, as variables empregadas son os ingresos ou os gastos dos fogares. Os estudos de pobreza subxectiva baséanse na propia percepción dos fogares ou individuos a cerca da súa situación.

Aplicando o enfoque obxectivo realízanse tanto análises de pobreza absoluta como relativa. A pobreza absoluta defínese como a situación na cal non están cubertas as necesidades básicas do individuo, é dicir, existe carencia de bens e servizos básicos (vivenda, alimentación, roupa...). Este concepto de pobreza debería poder aplicarse por igual en tódolos países e rexións. É dicir, unha persoa considerada pobre seguindo este criterio calificaríase da mesma maneira en todo o mundo.

A pobreza relativa, nembargantes, sitúa o fenómeno de pobreza dentro da sociedade de referencia na cal se fai o estudo. Dende esta perspectiva, consideraríase pobre unha persoa que ten determinadas carencias en relación ao resto de persoas do seu entorno. Este concepto de pobreza está ligado ao concepto de desigualdade e é o enfoque no cal nos centraremos neste traballo. Neste caso, a clasificación entre pobres e non pobres depende do grao de desenvolvemento da sociedade concreta estudada e non se pode trasladar a outra sociedade diferente.

Un aspecto importante a ter en conta cando se trata de medir a pobreza é que na maioría dos casos as medidas se obteñen a partir de datos procedentes de enquisas a fogares. Neste tipo de enquisas non se recolle información de persoas que viven en institucións ou indixentes, polo que estes individuos, que están

afectados pola pobreza en maior medida, non forman parte das medicións que habitualmente se realizan da mesma.

## 2.2. Elección do limiar de pobreza

O enfoque tradicional da medición da pobreza baséase na existencia dun limiar de pobreza que permite clasificar ás persoas da sociedade obxecto de estudo en dous grupos, aquelas máis desfavorecidas (consideradas pobres) e o resto. Estes limiares obtéñense a partir de variables monetarias como son o ingreso ou o gasto. En ambos casos, fíxase un nivel mínimo da variable por debaixo do cal as persoas se clasifican en pobres e por encima, como non pobres.

Debido á necesidade de obter datos altamente comparables na Unión Europea, creouse en cada estado membro a "Statistics on Income and Living Conditions (EU-SILC)", para asegurar un maior nivel de harmonización. En Galicia, esta operación toma o nome de "Enquisa de Condicións de Vida das familias"(ECV). Na actualidade, as medicións oficiais de pobreza en Galicia baséanse na ECV e seguen as recomendacións de Eurostat. A variable monetaria que se emprega para a súa obtención son os datos de ingresos proporcionados por esta enquisa.

O procedemento seguido nas estatísticas oficiais da Unión Europea e os seus estados membros, ante as recomendacións de Eurostat, para obter o limiar de pobreza é o seguinte:

En primeiro lugar calcúlase o ingreso total do fogar. As rendas que habitualmente se utilizan para construír este ingreso son: as rendas de traballo por conta propia ou allea, as rendas de capital e as prestacións sociais incluídas as de desemprego.

Unha das decisións máis relevantes nos estudos de pobreza relativa é a unidade de análise empregada: o fogar ou o individuo. Inicialmente empregábase o fogar, pero nos últimos anos recoméndase empregar o individuo, xa que son as persoas as que realmente están afectadas ou non pola pobreza, sendo o fogar un concepto teórico. De todas maneiras, suponse que as situacións persoais dependen dos ingresos totais do fogar e non unicamente dos ingresos persoais. Así, a tódalas persoas dun mesmo fogar se lles asigna un mesmo ingreso que depende do ingreso total do fogar. O ingreso asignado ao individuo é o ingreso equivalente por unidade de consumo, que se obtén dividindo o ingreso do fogar entre o número de unidades de consumo de dito fogar. Este concepto corrixe o efecto do número de persoas no volume de ingresos do fogar, facendo comparables os ingresos de fogares con distinto número de membros.

Existen diferentes alternativas para obter as unidades de consumo do fogar. En xeral, as unidades de consumo calcúlanse dacordo cunha certa escala de equivalencia. Estas escalas de equivalencia tratan de reflexar a realidade dos fogares, baseándose nas teorías de existencia de economías de escala. A existencia de eco-

---

nomías de escala nos fogares implica que o aumento do número de membros dun fogar non debe ir acompañado dun aumento proporcional de ingresos para seguir disfrutando o mesmo nivel de benestar. A escala empregada xeralmente por Eurostat é a "Escala da OCDE modificada". Neste caso, as unidades de consumo do fogar calcúlanse como a suma do peso que se lle adxudica a cada membro. Os pesos asígnanse da seguinte forma:

- Primeiro adulto: 1
- Segundo adulto e seguintes: 0.5
- Menores de 14 anos: 0.3

É dicir, o número de unidades de consumo ( $UC$ ) calcularíase así:

$$UC = 1 + 0,5 * (a - 1) + 0,3 * b$$

sendo  $a$  o número de adultos (persoas de 14 anos ou máis) e  $b$  o número de nenos (menores de 14 anos).

Defínese o ingreso equivalente do fogar como o ingreso do fogar dividido polas unidade de consumo do mesmo, e o ingreso equivalente do individuo como o ingreso equivalente do seu fogar.

O limiar de pobreza defínese como unha porcentaxe (actualmente Eurostat fíxaa no 60 por cento) da mediana dos ingresos equivalentes dos individuos. Defíníndose como pobre a toda persoa que teña un ingreso equivalente inferior a esta cifra.

### 2.3. Medidas de pobreza. A familia de Foster Greer e Thorbecke

Ao abordar o estudo da pobreza nunha sociedade debe facerse uso de medidas de incidencia, de distribución e de intensidade da mesma.

As **medidas de incidencia** da pobreza informan da extensión do problema, é dicir, proporcionan datos sobre a cantidade de persoas que están afectados. Estas medidas poden calcularse en toda a poboación e en todos os subgrupos da mesma que se desexe. De este modo, pode verse que grupos son máis vulnerables á pobreza. As **medidas de distribución** da pobreza indican cómo se distribúen os pobres e que características teñen. As **medidas de intensidade** permiten saber ata que punto a pobreza afecta á poboación. É dicir, céntranse no grao de pobreza que sufren as persoas, máis que no número de individuos considerados pobres.

---

Mediante o emprego conxunto das medidas de incidencia e intensidade da pobreza pode describirse con máis detalle o comportamento dunha sociedade. Pode existir, por exemplo, unha sociedade cunha porcentaxe elevada de pobres na cal todos os pobres estén próximos ao limiar de pobreza ou unha sociedade na que exista un número pequeno de pobres, pero moi alonxados do limiar, etc.

Existen certas propiedades que é desexable que cumpran as medidas de pobreza:

- **Invarianza de escala:** as medidas de pobreza deben permanecer inalteradas se, por exemplo, a poboación se duplica en tamaño mentras que todo o demais se mantén na mesma proporción
- **Monotocidade:** manténdose todo o demais constante, cando se reduce o ingreso dun fogar pobre, as medidas de pobreza deben aumentar ou manterse constantes.
- **Axioma de transferencia (principio de Pigou-Dalton):** manténdose todo o demais constante, se transferimos diñeiro dunha persoa pobre a outra persoa menos pobre, as medidas de pobreza deben incrementarse.
- **Axioma de transferencia sensible:** se unha transferencia,  $t > 0$ , de ingreso ten lugar dende un fogar pobre con ingreso  $y_i$  a un fogar pobre con ingreso  $y_i + d$  ( $d > 0$ ), entón a magnitude do incremento na pobreza debe ser menor para un maior  $y_i$ .
- **Axioma de descomponibilidade:** as medidas de pobreza deben poder descompoñerse en subpoboacións.

Foster, Greer e Thorbecke (Foster *et al.*, 1984) propuxeron unha familia de indicadores de pobreza que cumpren os axiomas citados, polo que permiten identificar a heteroxeneidade da pobreza. Se denotamos por  $q$  o número de pobres, por  $n$  o total da poboación,  $z$  o limiar de pobreza e  $y_j$  o ingreso equivalente para a persoa pobre  $j$ , os indicadores de Foster Greer e Thorbecke virían dados por:

$$FGT_{\alpha} = \frac{1}{n} \sum_{j=1}^q \left( \frac{z - y_j}{z} \right)^{\alpha} I_{\{y_j < z\}}, \alpha > 0 \quad (2.1)$$

sendo  $I_{\{y_j < z\}}$  a función indicadora, que toma o valor 1 se  $y_j < z$  e 0 noutro caso.

Algúns casos particulares de especial relevancia son os seguintes:

- Para  $\alpha = 0$ ;  $FGT_0 = \left( \frac{q}{n} \right)$  temos a proporción da poboación que é considerada pobre. Denomínase habitualmente *taxa de risco de pobreza* (ou *headcount ratio*) e é probablemente o indicador de incidencia de pobreza máis empregado. A taxa de risco de pobreza é sencilla de calcular e de entender, pero ten certas carencias. En particular, incumple o principio de transferencia de

Pigou-Dalton. Este ratio non indica como de pobres son as persoas consideradas pobres, por tanto, non sofre modificacións se as persoas pobres se tornan máis pobres.

- Para  $\alpha = 1$ ;  $FGT_1$  denomínase *brecha relativa de pobreza* (ou *poverty gap index*) e indica canto faría falta ser transferido aos pobres para facer que os seus ingresos superasen o limiar de pobreza.
  - Para  $\alpha = 2$ ;  $FGT_2$  denomínase *índice de pobreza severa* (*severity poverty index*) e capta as diferenzas na severidade da pobreza entre os pobres. Sen embargo, é máis difícil de interpretar que as anteriores.
-





# Capítulo 3

## Medidas de pobreza en Galicia

### 3.1. A Enquisa de condicións de vida das familias

En Galicia, o Instituto Galego de Estatística (IGE) publica anualmente diversos indicadores de cohesión social entre os que se atopan o limiar de pobreza e a taxa de risco de pobreza. A fonte de información da cal se obteñen estas medidas é a "Enquisa de condicións de vida das familias"(ECV).

O obxectivo que nos marcamos neste traballo é obter estimacións dos indicadores de Foster, Greer e Thorbecke, en particular da taxa de risco de pobreza, a un nivel inferior ao do deseño desta enquisa. En concreto, establécese o obxectivo de obter indicadores de pobreza a nivel comarcal. Para este fin, empregaranse os microdatos da ECV das edicións 2008, 2009 e 2010.

Nas próximas seccións descríbense as principais características da ECV, que pode consultarse con máis detalle na metodoloxía desta operación ([www.ige.eu/estatico/pdfs/s3/metodoloxias/met\\_ecv\\_2009\\_ga.pdf](http://www.ige.eu/estatico/pdfs/s3/metodoloxias/met_ecv_2009_ga.pdf)).

### 3.2. Definicións empregadas na ECV

A ECV é unha actividade estatística anual que dende 1999 realiza o IGE. Trátase dunha enquisa dirixida aos fogares galegos co obxecto de obter información sobre as súas características socioeconómicas.

Entre os obxectivos principais desta enquisa atópanse os seguintes:

- Analizar e describir as características principais dos fogares galegos e as diferenzas existentes entre as distintas áreas territoriais galegas.

- Proporcionar información sobre as características socioeconómicas da poboación e dos fogares galegos.
- Coñecer a contía dos ingresos dos fogares galegos e a súa tipoloxía, proporcionando información significativa sobre aqueles fogares cuns ingresos monetarios menores.
- Coñecer a contía dos gastos comúns dos fogares galegos e a súa relación con diversas variables socioeconómicas.

Aínda que esta enquisa se leva realizando dende o ano 1999, a partir da edición do ano 2008 introducíronse modificacións no deseño da mesma, que fan que os resultados non sexan directamente comparables coa serie anterior. No presente traballo empregáronse os microdatos das edicións 2008, 2009 e 2010.

A continuación introdúcense certas definicións empregadas nesta enquisa e que se poden consultar con maior detalle na metodoloxía da mesma:

**Vivenda familiar:** considérase vivenda familiar todo cuarto ou conxunto de cuartos e as súas dependencias, que ocupan un edificio ou unha parte estruturalmente separada del e que, pola forma en que foron construídos, reconstruídos ou transformados, están destinados a ser habitados por unha ou varias persoas, e na data da entrevista non se utilizan totalmente para outros fins.

**Vivenda familiar principal:** considérase vivenda familiar principal toda vivenda familiar que é utilizada como residencia habitual (a maior parte do ano) por un ou máis fogares.

**Fogar privado:** defínese o fogar privado como a persoa ou conxunto de persoas que ocupan en común unha vivenda principal ou parte dela, e consomen e/ou comparten alimentos ou outros bens con cargo a un mesmo orzamento.

**Fogar principal:** considérase fogar principal aquel no que un dos seus membros ten a titularidade da vivenda que ocupa, xa sexa en propiedade ou alugamento. No caso de que a vivenda estea cedida, o fogar principal será aquel no que estea a persoa a favor da que se fixo a cesión.

**Ingreso medio mensual do fogar:** o ingreso medio mensual do fogar é a media mensual dos ingresos netos monetarios de todos os membros do fogar no ano natural anterior, sexan estes regulares ou non. Con referencia aos ingresos percibidos polo fogar e os seus membros, estes poden proceder de:

- Traballo: fogares que perciben ingresos debidos ao traballo, xa sexa por conta propia ou por conta allea.
  - Prestacións: fogares que perciben ingresos debidos a pensións contributivas ou non contributivas, subsidios e prestacións por desemprego ou outros subsidios e prestacións sociais.
-

- Outros ingresos: fogares que perciben ingresos debidos ás rendas da propiedade ou do capital, rendas de compañías de seguros, transferencias doutros fogares ou outro tipo de ingresos que non sexan debidos ao traballo ou a pensións ou subsidios.

Nótese que a variable ingresos proporcionada pola enquisa de edición  $t$  fai referencia aos ingresos do ano  $t - 1$ .

**Mediana de ingresos dos fogares:** a mediana de ingresos é o valor que, ordenando todos os fogares da poboación de menor a maior ingreso, deixa a metade dos mesmos por debaixo de dito valor e a outra metade por enriba.

**Ingreso equivalente:** este concepto corrixe o efecto do número de persoas no volume de ingresos do fogar, facendo comparables os ingresos de fogares con distinto número de membros. Para iso calcúlanse as unidades de consumo do fogar empregando a escala de equivalencia da OCDE modificada e defínese o ingreso equivalente do fogar como o ingreso medio mensual do fogar dividido polas unidades de consumo do mesmo. Como ingreso equivalente de cada persoa tómase o ingreso equivalente do fogar ao que pertence.

**Limiar de risco de pobreza:** correspóndese co 60% da mediana dos ingresos equivalentes de todas as persoas galegas.

**Taxa de risco de pobreza:** porcentaxe de persoas cun ingreso equivalente inferior ao limiar de risco de pobreza.

### 3.3. **Ámbito da enquisa e unidades de análise empregadas**

**Ámbito poboacional:** As poboacións obxecto de estudo son: o conxunto de fogares privados que residen habitualmente en vivendas familiares principais e o conxunto das persoas que conforman estes fogares.

**Ámbito xeográfico:** O ámbito xeográfico da enquisa abarca todo o territorio galego. Dentro de cada provincia os concellos agrúpanse en áreas, cada unha delas é unha comarca ou unha agregación de comarcas veciñas.

**Período de referencia:** As preguntas referidas aos ingresos percibidos e gastos efectuados polo fogar refírense ao ano natural anterior á realización da enquisa. A partires da edición do 2008 a información referida aos ingresos, gastos e composición do fogar vai referida á poboación e fogares existentes a 31 de decembro do ano natural anterior á realización da enquisa.

**Unidades de análise:** Distínguense dous tipos de unidades de análise: o fogar e a persoa membro do fogar.

O fogar é a unidade básica de análise. Cando se analizan os datos dos fogares téñense en conta as persoas que conforman cada un deles.

**Unidades de mostraxe:** A mostraxe é bietápica. A unidade primaria de mostraxe é a sección censual e a unidade de segunda etapa é a vivenda familiar principal. Inclúense na mostra todos os fogares residentes nas vivendas seleccionadas.

### 3.4. Deseño mostral da ECV

O territorio galego divídese nas seguinte áreas<sup>1</sup>:

#### Provincia da Coruña

- Ferrol-Eume-Ortegal: comprende as comarcas de Ferrol, Eume e Ortegal.
- Área da Coruña: comprende as comarcas da Coruña e Betanzos
- Área da Costa da morte: comprende as comarcas de Bergantiños, Fisterra, Muros, Soneira e Xallas
- A Coruña suroriental: comprende as comarcas de Arzúa, Ordes e Terra de Melide.
- Área de Santiago: comprende as comarcas da Barcala, O Sar e Santiago.
- A Barbanza-Noia: comprende as comarcas da Barbanza e Noia.

#### Provincia de Lugo

- Lugo oriental: comprende as comarcas da Fonsagrada, Os Ancares e Sarria.
- Lugo sur: comprende as comarcas de Chantada, Quiroga e Terra de Lemos.
- Lugo central: comprende as comarcas da Ulloa, Lugo, Meira e A Terra Chá.
- A Mariña: comprende as comarcas da Mariña Central, a Mariña Oriental e a Mariña Occidental.

#### Provincia de Ourense

- O Carballiño-O Ribeiro: comprende as comarcas do Carballiño e O Ribeiro.

---

<sup>1</sup>Empregamos o término *área* para manter os conceptos empregados na metodoloxía da enquisa, sen embargo non debe confundirse coas *áreas pequenas* consideradas neste traballo, que son as comarcas.

---

- Ourense central: comprende as comarcas de Allariz e Maceda, Terra de Caldelas, Terra de Trives e Valdeorras.
- Ourense sur: comprende as comarcas da Limia, a Baixa Limia, Terra de Celanova, Verín e Viana.
- Área de Ourense: comprende a comarca de Ourense.

#### Provincia de Pontevedra

- Pontevedra nororiental: comprende as comarcas de Deza e Tabeirós-Terra de montes.
- Pontevedra sur: comprende as comarcas da Paradanta, O Baixo Miño e O Condado.
- Área de Pontevedra: comprende a comarca de Pontevedra.
- Caldas-O Salnés: comprende as comarcas de Caldas e O Salnés.
- Área de Vigo: comprende a comarca de Vigo.
- O Morrazo: comprende a comarca do Morrazo.

Dentro de cada área na que se divide cada unha das provincias galegas, a mostra é bietápica con estratificación previa das unidades de primeira etapa (as seccións censais). Cada área está dividida en estratos de acordo coa seguinte clasificación:

**Estrato 0:** concellos autorrepresentados

**Estrato 1:** concellos de máis de 20.000 habitantes

**Estrato 2:** concellos de 15.000 a 20.000 habitantes

**Estrato 3:** concellos de 10.000 a 15.000 habitantes

**Estrato 4:** concellos de 5.000 a 10.000 habitantes

**Estrato 5:** concellos de menos de 5.000 habitantes

Nalgunha área é necesario unir estratos contiguos para evitar a existencia de estratos de escasa representatividade.

Os concellos autorrepresentados son: A Coruña, Lugo, Ourense, Pontevedra, Santiago, Ferrol e Vigo.

**Marco da mostra:** O marco que se emprega para a extracción da mostra é o Padrón de habitantes.

---

**Tamaño teórico da mostra:** Na edición de 2010, a mostra consta de 512 seccións coa seguinte repartición por provincias: A Coruña 180, Lugo 90, Ourense 91 e Pontevedra 151. En cada sección entrevistáronse como mínimo a 18 vivendas, co que resulta unha mostra mínima de 9216 vivendas.

**Selección da mostra:** Nos concellos correspondentes ao estrato 0 a mostraxe é unietápica: ordénanse os fogares segundo características sociodemográficas e a continuación escóllese a mostra mediante mostraxe sistemática con arrinque aleatorio.

Para o resto dos estratos das distintas áreas a mostraxe é bietápica: en primeiro lugar ordénanse as seccións dese estrato segundo características sociodemográficas e logo escóllese a mostra de seccións mediante mostraxe sistemática con arrinque aleatorio. Nunha segunda etapa, para a mostra de seccións xa escollidas, ordénanse os fogares desas seccións mediante as ditas características sociodemográficas e a continuación escóllese a mostra de fogares mediante mostraxe sistemática con arrinque aleatorio.

### 3.5. Información publicada

A ECV proporciona anualmente resultados referentes ao limiar de pobreza e a taxa de risco de pobreza para Galicia, as súas provincias e as áreas nas que divide o territorio galego, que poden consultarse en [www.ige.eu/web/mostrar\\_actividade\\_estadistica.jsp?idioma=gl&codigo=0205002](http://www.ige.eu/web/mostrar_actividade_estadistica.jsp?idioma=gl&codigo=0205002). Para o cálculo da taxa de risco de pobreza emprega un estimador que utiliza unicamente a información proporcionada pola enquisa na subpoboación considerada e que será definido en 4.2. Os estimadores deste tipo, denomínanse estimadores directos. Tódolos datos desta sección foron extraídos da páxina web citada.

O limiar do risco de pobreza para os anos 2007, 2008 e 2009<sup>2</sup> foi respectivamente: 603.11, 609.48 e 605.02 euros. Estes son os valores de ingresos equivalentes mensuais que se toman como referencia para clasificar un individuo como pobre ou non pobre. Tomando estos limiares de pobreza, obtense a taxa de risco de pobreza nos anos citados en Galicia e nas súas provincias que pode verse no Cadro 3.1.

Vemos que a porcentaxe de persoas que se consideran pobres cos criterios definidos ronda o 14% en Galicia, e que, nos anos citados as porcentaxes menores se atopan na provincia da Coruña e os maiores, na provincia de Ourense, en termos xerais.

Ademais, publícase a taxa de risco de pobreza por sexo, idade e nacionalidade. Nas Figuras 3.1, 3.2 e 3.3 represéntase a taxa de risco de pobreza en función

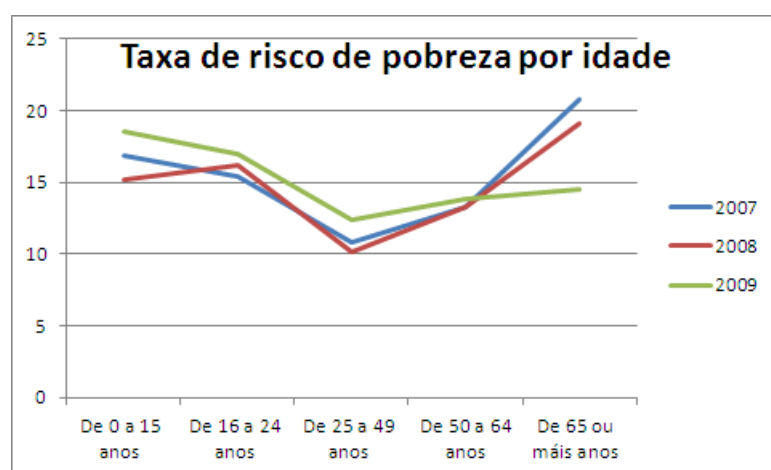
---

<sup>2</sup>Datos correspondentes ás edicións 2008, 2009 e 2010 da ECV respectivamente

	2007	2008	2009
<b>Galicia</b>	0,1463	0,1388	0,1432
<b>A Coruña</b>	0,1171	0,1200	0,1291
<b>Lugo</b>	0,1579	0,1401	0,1323
<b>Ourense</b>	0,1908	0,2171	0,1831
<b>Pontevedra</b>	0,1615	0,1334	0,1503

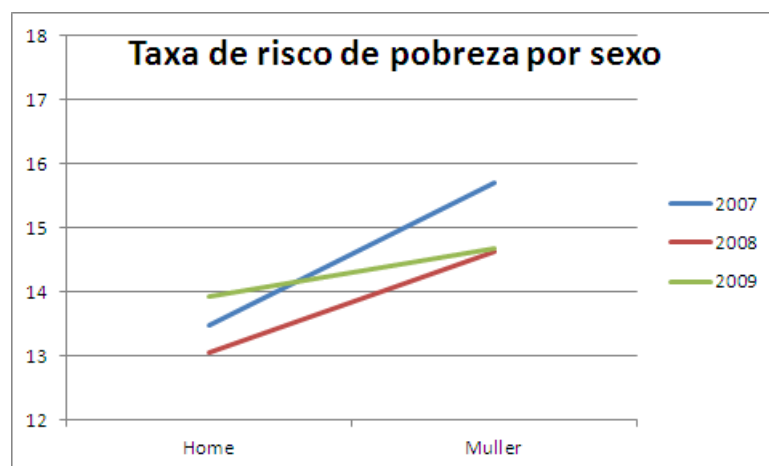
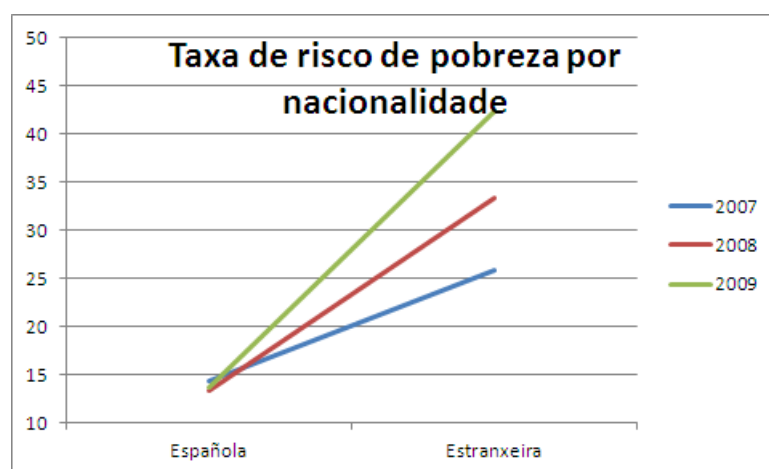
Cadro 3.1:  $FGT_0$  para Galicia e as súas provincias

das citadas categorías. Observándose que existen diferencias significativas nas diferentes categorías. Así, a proporción de pobres é maior nas mulleres que nos homes, é maior no colectivo de persoas con nacionalidade estranxeira e é menor no grupo de persoas que teñen idades comprendidas entre os 16 e os 64 anos. Outras variables que habitualmente se empregan en estudos de pobreza son as relativas ao nivel educativo ou á situación laboral dos individuos, sen embargo, a ECV non proporciona información sobre estas características.

Figura 3.1:  $FGT_0$  por grupos de idade para Galicia

O maior nivel de desagregación xeográfica para o cal se poden dar estimacións precisas empregando o estimador directo é o das áreas definidas no deseño da enquisa. Se se quere obter información a un maior nivel de detalle, o estimador baseado no deseño mostral presentará uns elevados erros de mostraxe. É por isto, que se fai necesario a utilización de técnicas de estimación en áreas pequenas cando se desexa obter información para outros niveis de desagregación como poden ser as comarcas ou os municipios galegos.

Neste traballo, marcámonos como obxectivo a obtención de estimacións precisas da taxa de risco de pobreza para as comarcas galegas (que serán as nosas áreas pequenas).

Figura 3.2:  $FGT_0$  por sexo para GaliciaFigura 3.3:  $FGT_0$  por nacionalidade para Galicia



# Capítulo 4

## Estimadores

Neste capítulo propóñense dous estimadores indirectos como alternativa ao estimador directo baseado no deseño da enquisa de condicións de vida amosado no capítulo anterior (o seu cálculo detallado verase na Sección 4.2). Ambos estimadores indirectos se basean na utilización de modelos lineais mixtos, nun caso trátase dun modelo a nivel de área (as variables refírense a datos agregados) e no outro dun modelo a nivel de individuo (as variables coñécense para cada individuo da poboación).

A continuación introdúcense outros dous estimadores indirectos baseados nun modelo a nivel de área no cal se inclúe o efecto tempo para obter estimacións precisas cando se dispón de información para varios períodos.

Así mesmo, introdúcese para cada estimador proposto o correspondente estimador do erro cuadrático medio.

### 4.1. Notación

No que segue empregaremos a seguinte notación:

Sexa  $P$  unha poboación finita de tamaño  $N$ . Supoñemos que a poboación  $P$  está dividida en  $D$  dominios  $P_d$  de tamaños  $N_d$ ,  $\sum_{d=0}^D N_d = N$ . No contexto da estimación en áreas pequenas, denomínase dominio ou área pequena aos subconxuntos poboacionais de tamaño inferior aos considerados no deseño da enquisa.

Sexa  $S$  unha mostra de tamaño  $n$  da poboación anterior, e  $S_d = S \cap P_d$  a submostra no dominio  $d$  de tamaño  $n_d$  ( $\sum_{d=0}^D n_d = n$ ). Sexa  $R = P - S$  o subconxunto da poboación non mostreado.

Denotamos por  $z_{dj}$  o ingreso mensual equivalente do individuo  $j$  ( $j = 1 \dots N_d$ ) do dominio  $d$  ( $d = 1 \dots D$ ) e por  $z$  o limiar de pobreza da poboación.

Con esta notación, o indicador de Foster, Greer e Thorbecke de orde  $\alpha$  ( $FGT_\alpha$ ) para a comarca  $d$  viría dado por:

$$\bar{Y}_{\alpha,d} = \frac{1}{N_d} \sum_{j=1}^{N_d} \left( \frac{z - z_{dj}}{z} \right)^\alpha I_{\{z_{dj} < z\}}, \alpha = 0, 1, 2 \quad (4.1)$$

sendo  $I_A$  a función indicadora do conxunto  $A$ .

## 4.2. Estimador directo

No contexto da estimación en áreas pequenas, un estimador directo é aquel estimador que se basea unicamente nos datos específicos da mostra no dominio correspondente, sen empregar ningún outro tipo de información auxiliar. Os estimadores directos son tipicamente estimadores baseados no deseño mostral. No caso dos índices de pobreza da familia de Foster, Greer e Thorbecke, os correspondentes estimadores directos defínense como segue:

$$\bar{y}_{\alpha,d} = \frac{1}{\hat{N}_d} \sum_{j \in S_d} \omega_{dj} \left( \frac{z - z_{dj}}{z} \right)^\alpha I_{\{z_{dj} < z\}}, \alpha = 0, 1, 2 \quad (4.2)$$

sendo  $\hat{N}_d = \sum_{j \in S_d} \omega_{dj}$  e  $\omega_{dj}$  os factores de elevación do deseño mostral.

Este é o estimador empregado na publicación da ECV.

Como medida do seu erro cuadrático medio<sup>1</sup> empregamos o seguinte estimador (ver [Molina & Morales, 2009](#)):

$$ecm(\bar{y}_{\alpha,d}) = \frac{1}{\hat{N}_d^2} \sum_{j \in S_d} \omega_{dj} (\omega_{dj} - 1) \left( \left( \frac{z - z_{dj}}{z} \right)^\alpha I_{\{z_{dj} < z\}} - \bar{y}_{\alpha,d} \right)^2; \alpha = 0, 1, 2 \quad (4.3)$$

## 4.3. Estimador indirecto baseado no modelo de Fay-Herriot

O estimador proposto baséase nun modelo lineal mixto a nivel de área introducido por Fay e Herriot (1979) que emprega estimadores directos como variable resposta. De esta maneira, o modelo incorpora información do deseño mostral. Este é un dos estimadores en áreas pequenas baseado en modelos máis amplamente estudado e utilizado ata a actualidade.

<sup>1</sup>Na ECV emprégase un procedemento de tipo Jackknife para obter os erros de mostraxe, sen embargo non foi posible replicar dita estimación, dado que non se facilitou a variable "Estrato" na base de datos primarios

Dado un  $\alpha$ , o modelo de Fay-Herriot asume que a verdadeira media poboacional da área pequena (neste caso, o indicador  $FGT_\alpha$ ),  $\bar{Y}_{\alpha,d}$ , está linealmente relacionada cun conxunto de variables auxiliares dos dominios:

$$\bar{Y}_{\alpha,d} = x_d\beta + u_d, \quad d = 1, \dots, D \quad (4.4)$$

onde:

- $x_d$ ,  $d = 1, \dots, D$ , é un vector  $1 \times p$  contendo os valores de  $p$  variables auxiliares no dominio  $d$ .
- $\beta$  é un vector  $p \times 1$  contendo os efectos fixos
- $u_d$ ,  $d = 1, \dots, D$  é o efecto aleatorio na área  $d$ , sendo  $u_1, u_2, \dots, u_D$  independentes e idénticamente distribuídos  $N(0, \sigma_u^2)$ , con  $\sigma_u^2$  descoñecida e constante.

O modelo (4.4) denomínase a veces *modelo de conexión*. Nel, o efecto aleatorio da área  $u_d$  emprégase para relacionar a verdadeiro valor do indicador  $FGT_\alpha$  na comarca  $d$  coas variables auxiliares ( $x_d$ ). Isto conséguese introducindo unha compoñente da varianza descoñecida,  $\sigma_u^2$ , que deberá ser estimada a partir dos datos.

Ademais, asúmese que existe un estimador directo relacionado co verdadeiro valor do indicador  $FGT_\alpha$  a través do seguinte modelo (*modelo mostral*):

$$\bar{y}_{\alpha,d} = \bar{Y}_{\alpha,d} + e_d \quad d = 1, \dots, D \quad (4.5)$$

onde:

- $e_d$ ,  $d = 1, \dots, D$ , son os erros mostrais que se asumen independentes, con distribución Normal de media cero e varianza coñecida  $\sigma_{e_d}^2$ .

Os modelos (4.4) e (4.5) poden expresarse como un modelo lineal mixto :

$$\bar{y}_{\alpha,d} = x_d\beta + u_d + e_d, \quad d = 1, \dots, D \quad (4.6)$$

con matriz de varianzas-covarianzas,

$$\mathbf{V} = \mathbf{V}_u + \mathbf{V}_e$$

sendo:

$$\mathbf{V}_u = \sigma_u^2 I_D \quad e \quad \mathbf{V}_e = \text{diag}(\sigma_{e_1}^2, \sigma_{e_2}^2, \dots, \sigma_{e_D}^2)$$

Alternativamente, podemos expresar o modelo con notación matricial como segue:

$$\bar{\mathbf{y}}_\alpha = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

con:

- $\bar{y}_\alpha = \text{col}_{1 \leq d \leq D}(\bar{y}_{\alpha d});$
- $\mathbf{u} = \text{col}_{1 \leq d \leq D}(u_d);$
- $\mathbf{e} = \text{col}_{1 \leq d \leq D}(e_d);$
- $\mathbf{X} = \text{col}_{1 \leq d \leq D}(x_d);$
- $\mathbf{Z} = \mathbf{I}$

Tomando como referencia o libro de Rao (Rao, 2003), temos os seguintes resultados:

Os mellores predictores lineais inesgado (BLUP) de  $\beta$  e  $\mathbf{u}$ , asumindo que a matriz de varianzas-covarianzas,  $\mathbf{V}$ , é coñecida son:

$$\begin{aligned}\tilde{\beta} &= (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \bar{y}_\alpha, \\ \tilde{\mathbf{u}} &= \mathbf{V}_u \mathbf{Z}^t \mathbf{V}^{-1} (\bar{y}_\alpha - \mathbf{X} \tilde{\beta})\end{aligned}$$

As compoñentes de  $\tilde{\mathbf{u}}$  son:

$$\tilde{u}_d = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{ed}^2} (\bar{y}_{\alpha,d} - x_d \tilde{\beta})$$

O BLUP do indicador  $FGT_\alpha$  na comarca  $d$ ,  $\bar{Y}_{\alpha,d}$ , obtense como :

$$\hat{Y}_{\alpha,d}^{BLUP} = x_d \tilde{\beta} + \tilde{u}_d = x_d \tilde{\beta} + \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{ed}^2} (\bar{y}_{\alpha,d} - x_d \tilde{\beta}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_{ed}^2} \bar{y}_{\alpha,d} + \frac{\sigma_{ed}^2}{\sigma_u^2 + \sigma_{ed}^2} x_d \tilde{\beta}$$

O estimador de Fay-Herriot é o estimador EBLUP (mellor predictor lineal inesgado empírico) de  $\bar{Y}_{\alpha,d}$  que se obtén substituíndo  $\sigma_u^2$  polo seu estimador  $\hat{\sigma}_u^2$  na expresión anterior. Dado que, na práctica,  $\sigma_{ed}^2$  serán tamén descoñecidas aproxímanse por  $\hat{\sigma}_{ed}^2 = \text{ecm}(\bar{y}_{\alpha,d})$ ,  $d = 1, \dots, D$ .

$$\hat{Y}_{\alpha,d}^{FH} = \gamma_d \bar{y}_{\alpha,d} + (1 - \gamma_d) x_d \hat{\beta}, \quad d = 1, \dots, D \quad (4.7)$$

$$\gamma_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \sigma_{ed}^2}$$

$$\hat{\beta} = (\mathbf{X}^t \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \hat{\mathbf{V}}^{-1} \bar{y}_\alpha$$

#### Estimación da varianza dos efectos aleatorios de área $\sigma_u^2$

Empregouse o método de máxima verosimilitude restrinxida (REML) para obter unha estimación de  $\sigma_u^2$ . O correspondente estimador obtense empregando o seguinte algoritmo de Fisher-Scoring (ver Rao, 2003, pax. 119):

1. Definir  $\hat{\sigma}_u^{2(0)} = 1$
2. Para  $k = 1, 2, \dots$

$$\hat{\sigma}_u^{2(k)} = \hat{\sigma}_u^{2(k-1)} + \left( F(\hat{\sigma}_u^{2(k-1)}) \right)^{-1} S(\hat{\sigma}_u^{2(k-1)}),$$

sendo:

$$F(\hat{\sigma}_u^2) = \frac{1}{2} \text{tr}(PP)$$

$$S(\hat{\sigma}_u^2) = -\frac{1}{2} \text{tr}(P) + \frac{1}{2} \bar{y}_\alpha^t P P \bar{y}_\alpha$$

$$P = V^{-1} - V^{-1} X (X^t V^{-1} X)^{-1} X^t V^{-1}$$

$$V^{-1} = \text{diag} \left( \frac{1}{\sigma_{ed}^2 + \hat{\sigma}_u^{2(k-1)}} \right)$$

3. Continuar ata que:

$$\left| \frac{\hat{\sigma}_u^{2(k)} - \hat{\sigma}_u^{2(k-1)}}{\hat{\sigma}_u^{2(k-1)}} \right| \leq \epsilon$$

4. O estimador buscado é:

$$\hat{\sigma}_u^2 = \hat{\sigma}_u^{2(k)}$$

#### Estimación do erro cuadrático medio do estimador $\hat{Y}_{\alpha,d}^{FH}$

Para estimar o erro cuadrático medio do estimador  $\hat{Y}_{\alpha,d}^{FH}$  empregáronse dous estimadores: o primeiro deles baseado na aproximación de Prasad-Rao (ver [Prasad & Rao, 1990](#)) do erro cuadrático medio:

$$ecm_1 \left( \hat{Y}_{\alpha,d}^{FH} \right) = g_1(\hat{\sigma}_u^2) + g_2(\hat{\sigma}_u^2) + 2g_3(\hat{\sigma}_u^2) \quad (4.8)$$

$$g_1(\hat{\sigma}_u^2) = \frac{\hat{\sigma}_u^2 \hat{\sigma}_{ed}^2}{\hat{\sigma}_u^2 + \hat{\sigma}_{ed}^2}, \quad g_2(\hat{\sigma}_u^2) = \frac{\hat{\sigma}_{ed}^4}{(\hat{\sigma}_u^2 + \hat{\sigma}_{ed}^2)^2} x_d (X^t V^{-1} X)^{-1} x_d^t$$

$$g_3(\hat{\sigma}_u^2) \approx \frac{\hat{\sigma}_{ed}^4}{(\hat{\sigma}_u^2 + \hat{\sigma}_{ed}^2)^3} \left( \sum_{d=1}^D \frac{2}{(\hat{\sigma}_u^2 + \hat{\sigma}_{ed}^2)^2} \right)^{-1}$$

e o segundo empregando un bootstrap paramétrico (ver [González Manteiga et al., 2008](#)). O procedemento é o seguinte:

1. Obter estimacións dos parámetros descoñecidos do modelo:  $\hat{\sigma}_u^2$  e  $\hat{\beta}$
2. Repetir para cada réplica bootstrap,  $b = 1, 2, \dots, B$ ;

- Xerar  $D$  variables  $u_d^{*b}$  independentes e idénticamente distribuídas con distribución Normal de media cero e varianza  $\hat{\sigma}_u^2$ . Sexa  $\mathbf{u}^{*b} = (u_1^{*b}, u_2^{*b}, \dots, u_D^{*b})$
- Xerar  $D$  variables  $e_d^{*b}$  independentes con distribución Normal de media cero e varianza  $\sigma_{e_d}^2$ . Sexa  $\mathbf{e}^{*b} = (e_1^{*b}, e_2^{*b}, \dots, e_D^{*b})$
- Xerar o modelo bootstrap:  $\bar{y}_\alpha^{*b} = \mathbf{X}\hat{\beta} + \mathbf{u}^{*b} + \mathbf{e}^{*b}$
- Obter o verdadeiro valor do parámetro da poboación bootstrap, que ven dado por:  $\bar{Y}_\alpha^{*b} = \mathbf{X}\hat{\beta} + \mathbf{u}^{*b}$
- Obter o estimador de Fay-Herriot:  $\hat{Y}^{\text{FH}*b}$ , tal e como se describe nas seccións previas.

3. Obter o estimador bootstrap do erro cuadrático medio:

$$ecm_2\left(\hat{Y}_{\alpha,d}^{\text{FH}}\right) = \frac{1}{(B-1)} \sum_{b=1}^B \left( \hat{Y}_{\alpha,d}^{\text{FH}*b} - \bar{Y}_{\alpha,d}^{*b} \right)^2 \quad (4.9)$$

## 4.4. Estimador indirecto baseado nun modelo de individuo: EB-Method

Máis recentemente, Molina e Rao ([Molina & Rao, 2010](#)) propoñen un método xeral para estimar parámetros non lineais en áreas pequenas empregando *Empirical Best* (EB) estimadores baseados nun modelo de erro anidado. Estes estimadores EB obtéñense por aproximación de tipo Monte Carlo. Aínda que neste traballo nos imos centrar na utilización de estos estimadores para obter os indicadores de Foster, Greer e Thorbecke, esta técnica é tamén válida para obter outros indicadores non lineais de desigualdade e pobreza, en particular o índice de Gini ou o indicador de Theil.

Sexa  $y = (y_1, y_2, \dots, y_N)$  un vector contendo os valores dunha variable nos  $N$  individuos da poboación. Denotamos por  $y_s$  o subvector de  $y$  que contén os elementos mostreados e por  $y_r$  o subvector de  $y$  que contén os elementos que non están na mostra. Denotamos por  $y_d$  o subvector de  $y$  que contén os elementos do dominio  $d$ .

O cálculo do estimador EB do indicador de Foster, Greer e Thorbecke de orde  $\alpha$ ,  $\hat{Y}_{\alpha,d}^{\text{EB}}$ , empregando este método require poder expresar  $\bar{Y}_{\alpha,d}$  en función dun certo vector  $y_d$ , para o cal se coñece a distribución condicionada do vector non mostreado  $y_r$ , coñecida a mostra  $y_s$ . A variable coa que se traballa á hora de obter medidas de pobreza, os ingresos equivalentes, non ten unha distribución coñecida a priori. Sen embargo, este feito non supón unha gran dificultade, xa que as variables de tipo económico como os ingresos equivalentes,  $z_{dj}$ , acostuman a presentar unha distribución asimétrica á dereita que se pode transformar en Normal

mediante algunha transformación de tipo Box-Cox. En particular, a toma de logaritmos usualmente convirte variables de este tipo en variables con distribución Normal. Así pois, este método presupón a existencia dunha aplicación bixectiva  $T$  tal que  $y_{dj} = T(z_{dj})$  ten distribución Normal. Entón, o parámetro obxectivo pode expresarse como:

$$\bar{Y}_{\alpha,d} = \frac{1}{N_d} \sum_{j=1}^{N_d} \left( \frac{z - T^{-1}(y_{dj})}{z} \right)^\alpha I_{\{T^{-1}(y_{dj}) < z\}} \equiv h_\alpha(y_d)$$

É dicir, o parámetro obxectivo exprésase como unha función non lineal  $h_\alpha$  da variable  $y_d$ . O mellor predictor ven dado por:

$$\hat{Y}_{\alpha,d}^{BP} = E_{y_r}(\bar{Y}_{\alpha,d}/\mathbf{y}_s)$$

A esperanza condicionada anterior non pode ser calculada explícitamente debido á complexidade da función  $h_\alpha$ . Sen embargo, debido a que  $\mathbf{y} = (\mathbf{y}_s, \mathbf{y}_r)$  se asume normalmente distribuído con vector de medias  $\boldsymbol{\mu} = (\boldsymbol{\mu}_s, \boldsymbol{\mu}_r)$  e matriz de varianzas covarianzas:

$$\mathbf{V} = \begin{pmatrix} V_{ss} & V_{sr} \\ V_{rs} & V_{rr} \end{pmatrix}$$

A distribución condicionada de  $\mathbf{y}_r$  dado  $\mathbf{y}_s$  é

$$\mathbf{y}_r/\mathbf{y}_s \sim N(\boldsymbol{\mu}_{r/s}, V_{r/s}) \quad (4.10)$$

onde:

$$\begin{aligned} \boldsymbol{\mu}_{r/s} &= \boldsymbol{\mu}_r + V_{rr} V_s^{-1} (\mathbf{y}_s - \boldsymbol{\mu}_s), \\ V_{r/s} &= V_r - V_{rs} V_s^{-1} V_{sr} \end{aligned}$$

Os autores proponen empregar unha aproximación empírica, do mellor predictor, baseada nunha simulación de tipo Monte Carlo de un número  $L$  de vectores  $\mathbf{y}_r$  xerados a partir de (4.10). Denotamos por  $y_{dj}^{(l)}$  o valor da observación non mostreada  $y_{dj}$ ,  $j \in r_d$  obtida na  $l$ -ésima simulación  $l = 1, \dots, L$  e por  $\mathbf{y}_r^{(l)}$  o correspondente vector de observacións non mostreadas. Unindo estes datos simulados por Monte Carlo ás observacións mostrais, obtemos  $L$  poboacións  $\mathbf{y}^{(l)} = (\mathbf{y}_s, \mathbf{y}_r^{(l)})$  e aproxímase o mellor predictor como:

$$\hat{Y}_{\alpha,d}^{BP} \approx \frac{1}{L} \sum_{h=1}^L h_\alpha(y_{dj}^{(l)})$$

Na práctica os vectores de medias  $\boldsymbol{\mu}$  e a matriz de varianzas covarianzas  $\mathbf{V}$  dependerán dun vector de parámetros descoñecidos  $\boldsymbol{\theta}$ . O procedemento a seguir será, neste caso, obter en primeiro lugar un estimador  $\hat{\boldsymbol{\theta}}$  de  $\boldsymbol{\theta}$  e logo proceder como

se acaba de describir. O estimador resultante é chamado estimador EB ou mellor predictor empírico (*empirical best predictor*).

Os autores asumen un modelo de superpoboación que relaciona as variables transformadas con  $p$  variables auxiliares e que inclúe un efecto aleatorio de área ( $u_d$ ). Así,

$$y_{dj} = x_{dj}\beta + u_d + e_{dj}, j = 1, \dots, N_d, d = 1, \dots, D \quad (4.11)$$

$$u_d \quad iid \sim N(0, \sigma_u^2)$$

$$e_{dj} \quad iid \sim N(0, \sigma_e^2)$$

onde:

- $x_{dj}$  é un vector  $1 \times p$  contendo os valores de  $p$  variables auxiliares no dominio  $d$  e no individuo  $j$
- $\beta$  é un vector  $p \times 1$  contendo os efectos fixos
- $u_d$  é o efecto aleatorio na área  $d$ , sendo  $u_1, u_2, \dots, u_D$  independentes e idénticamente distribuídos  $N(0, \sigma_u^2)$ , con  $\sigma_u^2$  descoñecida pero constante.
- $e_{dj}$  son os erros do modelo, que se supoñen independentes e idénticamente distribuídos  $N(0, \sigma_e^2)$  con  $\sigma_e^2$  descoñecida pero constante. Os erros asúmense independentes de  $u_d$ .

O feito de asumir este modelo de superpoboación simplifica a distribución condicionada de  $y_r/y_s$  que se emprega para as simulacións de Monte Carlo. Os pasos a seguir para obter o estimador EB poden resumirse como segue:

1. Axustar o modelo (4.11) empregando os datos iniciais (transformados)  $y_s$  e obter os estimadores  $\hat{\theta} = (\hat{\beta}, \hat{\sigma}_u^2, \hat{\sigma}_e^2)$  e  $\hat{u}$ . Empregouse o método de máxima verosimilitude restrinxida, e a librería *nlme* de R.
2. Xerar  $L$  vectores non mostreados  $y_r^{(l)}, l = 1, \dots, L$  empregando a distribución de  $y_r/y_s$ . Na práctica (ver [Molina & Rao, 2010](#)) esto faise xerando  $y_{dj}^{(l)}, j \in r_d$  como segue:

$$y_{dj}^{(l)} = x_{dj}\hat{\beta} + \hat{u}_d + v_d + \epsilon_{dj}$$

$$v_d \sim N(0, \hat{\sigma}_u^2(1 - \hat{\gamma}_d))$$

sendo

$$\hat{\gamma}_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2/n_d}$$

$$\epsilon_{dj} \sim N(0, \hat{\sigma}_e^2)$$



3. Cos  $L$  vectores xerados  $\mathbf{y}_r^{(l)}$  e os datos mostrais  $\mathbf{y}_s$ , obter o estimador EB

$$\hat{Y}_{\alpha,d}^{EB} \approx \frac{1}{L} \sum_{h=1}^L h_{\alpha}(y_{dj}^{(l)})$$

#### Estimación do erro cuadrático medio do estimador $\hat{Y}_{\alpha,d}^{EB}$

O erro cuadrático medio de este estimador obtívose empregando un bootstrap paramétrico proposto por Molina, I. e Rao, J.N.K.(ver [Molina & Rao, 2010](#)). O procedemento foi o seguinte:

1. Axustar o modelo (4.11) empregando os datos iniciais (transformados)  $\mathbf{y}_s$  e obter os estimadores  $\hat{\theta} = (\hat{\beta}, \hat{\sigma}_u^2, \hat{\sigma}_e^2)$  e  $\hat{u}$ .
2. Para  $b = 1, 2, \dots, B$ ;
  - Xerar  $D$  variables  $u_d^{*b}$  independentes e idénticamente distribuídas con distribución Normal de media cero e varianza  $\hat{\sigma}_u^2$ .
  - Xerar  $N$  erros aleatorios  $e_{dj}^{*b}$ ,  $d = 1, \dots, D$ ,  $j = 1, \dots, N_d$  independentes e idénticamente distribuídos con distribución Normal de media cero e varianza  $\hat{\sigma}_e^2$
  - Xerar o modelo bootstrap:  $y_{dj}^{*b} = x_{dj} \hat{\beta} + u_d^{*b} + e_{dj}^{*b}$ ,  $d = 1, \dots, D$ ,  $j = 1, \dots, N_d$
  - Obter para esta poboación o verdadeiro valor do parámetro obxectivo  $\bar{Y}_{\alpha,d}^{*b}$
  - Seleccionar unha mostra  $s$  da poboación simulada:  $y_{dj}^{*b}$  e obter o estimador EB ( $\hat{Y}_{\alpha,d}^{EB*b}$ ) tal e como se describe na sección anterior.
3. Obter o estimador bootstrap do erro cuadrático medio:

$$ecm\left(\hat{Y}_{\alpha,d}^{EB}\right) = \frac{1}{(B-1)} \sum_{b=1}^B \left(\hat{Y}_{\alpha,d}^{EB*b} - \bar{Y}_{\alpha,d}^{*b}\right)^2 \quad (4.12)$$

## 4.5. Estimador indirecto baseado nun modelo de área con efecto tempo

Nas seccións previas introducíamos estimadores dos indicadores de Foster, Greer e Thorbecke para un ano ou período temporal concreto. Na presente sección introducimos dous novos estimadores baseados en sendos modelos lineais mixtos a nivel de área con efecto tempo. Os modelos que se propoñen poden considerarse

unha extensión do modelo de área de Fay-Herriot, non cal se introduce a existencia dunha certa dependencia temporal e baséanse en diversos estudos de recente aparición, como (Esteban *et al.*, 2011b) e (Esteban *et al.*, 2011a).

Denotamos os períodos temporais por  $t$ , supoñendo que  $t = 1, \dots, T$ .

Sexa  $P_t$  unha poboación finita de tamaño  $N_t$  no período  $t$ . Supoñemos que a poboación  $P_t$  está dividida en  $D$  dominios  $P_{dt}$  de tamaños  $N_{dt}$ ,  $\sum_{d=0}^D N_{dt} = N_t$ .

Sexa  $S_t$  unha mostra de tamaño  $n_t$  da poboación anterior no período  $t$ , e  $S_{dt} = S_t \cap P_{dt}$  a submostra no dominio  $d$  de tamaño  $n_{dt}$  ( $\sum_{d=0}^D n_{dt} = n_t$ ).

Denotamos por  $z_{dtj}$  o ingreso anual equivalente do individuo  $j$  ( $j = 1 \dots N_d$ ) do dominio  $d$  ( $d = 1 \dots D$ ) no período  $t$  e por  $z_t$  o limiar de pobreza da poboación no período  $t$ .

Con esta notación, o indicador de Foster, Greer e Thorbecke de orde  $\alpha$  para o período  $t$  viría dado por:

$$\bar{Y}_{\alpha,dt} = \frac{1}{N_{dt}} \sum_{j=1}^{N_{dt}} \left( \frac{z_t - z_{dtj}}{z_t} \right)^\alpha I_{\{z_{dtj} < z_t\}}, \alpha = 0, 1, 2 \quad (4.13)$$

e o estimador directo viría dado por:

$$\bar{y}_{\alpha,dt} = \frac{1}{\hat{N}_{dt}} \sum_{j \in S_{dt}} \omega_{dtj} \left( \frac{z_t - z_{dtj}}{z_t} \right)^\alpha I_{\{z_{dtj} < z_t\}}, \alpha = 0, 1, 2 \quad (4.14)$$

sendo  $\hat{N}_{dt} = \sum_{j \in S_{dt}} \omega_{dtj}$  e  $\omega_{dtj}$  os factores do deseño no período  $t$ .

Consideramos o seguinte modelo que denotamos por **Modelo 1**:

$$\bar{y}_{\alpha,dt} = x_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1, \dots, D; \quad t = 1, \dots, T \quad (4.15)$$

sendo:

- $\bar{y}_{\alpha,dt}$  o estimador directo no período  $t$  e na área  $d$ .
- $x_{dt}$  un vector  $1 \times p$  contendo os valores de  $p$  variables auxiliares no dominio  $d$  e no período  $t$ .
- $\beta$  un vector  $p \times 1$  contendo os efectos fixos.
- $u_{dt}$  o efecto aleatorio na área  $d$  e no período  $t$ , con  $(u_{d1}, u_{d2}, \dots, u_{dT})$  independentes e idénticamente distribuídas AR(1);

$$u_{dt} = \rho u_{dt-1} + v_t; \quad \text{con} \quad v_t \quad iid \sim N(0, \sigma_u^2)$$

e  $\sigma_u^2$  descoñecida e constante.

- $e_{dt}$  os erros mostrais que se asumen independentes, con distribución Normal de media cero e varianza coñecida  $\sigma_{edt}^2$

Consideramos tamén un modelo máis simple, que denotamos por **Modelo 0**, obtido restrinxindo o modelo (4.15) ao caso  $\rho = 0$ .

Os modelos anteriores pódense expresar en modo matricial como segue:

$$\bar{\mathbf{y}}_\alpha = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (4.16)$$

con:

- $\bar{\mathbf{y}}_\alpha = \text{col}_{1 \leq d \leq D}(\bar{\mathbf{y}}_{\alpha d}); \bar{\mathbf{y}}_{\alpha d} = \text{col}_{1 \leq t \leq T}(\bar{\mathbf{y}}_{\alpha dt})$
- $\mathbf{u} = \text{col}_{1 \leq d \leq D}(\mathbf{u}_d); \mathbf{u}_d = \text{col}_{1 \leq t \leq T}(u_{dt})$
- $\mathbf{e} = \text{col}_{1 \leq d \leq D}(\mathbf{e}_d); \mathbf{e}_d = \text{col}_{1 \leq t \leq T}(e_{dt})$
- $\mathbf{X} = \text{col}_{1 \leq d \leq D}(\mathbf{X}_d); \mathbf{X}_d = \text{col}_{1 \leq t \leq T}(x_{dt})$
- $\mathbf{Z} = \mathbf{I}$

Con esta notación,  $\mathbf{e} \sim N(0, \mathbf{V}_e)$  e  $\mathbf{u} \sim N(0, \mathbf{V}_u)$  son independentes, con matrices de varianzas-covarianzas:

$$\mathbf{V}_e = \text{diag}_{1 \leq d \leq D}(\mathbf{V}_{ed}); \mathbf{V}_{ed} = \text{col}_{1 \leq t \leq T}(\sigma_{edt})$$

e

$$\mathbf{V}_u = \sigma_u^2 \Omega(\rho); \Omega(\rho) = \text{diag}_{1 \leq d \leq D}(\Omega_d(\rho))$$

con

$$\Omega_d(\rho) = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-2} & \rho^{T-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{T-2} & \dots & \dots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \dots & \rho & 1 \end{pmatrix}$$

Asumindo coñecida a matriz de varianzas-covarianzas,  $\mathbf{V} = \mathbf{V}_u + \mathbf{V}_e$ , o mellor predictor lineal inesgado (BLUP) viría dado por: (Esteban *et al.*, 2011b)

$$\begin{aligned} \tilde{\beta} &= (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \bar{\mathbf{y}}_\alpha, \\ \tilde{\mathbf{u}} &= \mathbf{V}_u \mathbf{Z}^t \mathbf{V}^{-1} (\bar{\mathbf{y}}_\alpha - \mathbf{X} \tilde{\beta}) \end{aligned}$$

Dado que existen compoñentes da varianza descoñecidas,  $\theta = (\sigma_u^2, \rho)$  deben estimarse en primeiro lugar. Posteriormente substitúense polos seus estimadores na expresión anterior, e obtense o correspondente estimador do parámetro obxectivo como:

$$\hat{Y}_{\alpha, dt} = x_{dt} \hat{\beta} + \hat{u}_{dt}$$

sendo:

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}^t \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \hat{\mathbf{V}}^{-1} \bar{y}_\alpha, \\ \hat{u} &= \hat{\mathbf{V}}_u \mathbf{Z}^t \hat{\mathbf{V}}^{-1} (\bar{y}_\alpha - \mathbf{X} \hat{\beta}). \end{aligned}$$

Na práctica,  $\sigma_{edt}^2$  será tamén descoñecida e aproximarase por  $\hat{\sigma}_{edt}^2 = ecm(\bar{y}_{\alpha, dt})$ ,  $d = 1, \dots, D$ ;  $t = 1, \dots, T$ .

### Estimación das compoñentes descoñecidas da matriz de varianzas-covarianzas

Para estimar  $\theta = (\sigma_u^2, \rho)$  empregamos o método de máxima verosimilitude restrinxida (REML). O correspondente estimador obtense empregando o seguinte algoritmo de Fisher-Scoring (Esteban *et al.*, 2011b):

1. Definir  $\theta^{(0)} = (\sigma_u^{2(0)}, \rho^{(0)}) = (1, 0)$

2. Para  $k = 1, 2, \dots$

$$\theta^{(k)} = \theta^{(k-1)} + (F(\theta^{(k-1)}))^{-1} S(\theta^{(k-1)})$$

sendo  $S$  e  $F$  o vector de puntuacións e a matriz de información de Fisher respectivamente

$$S_a = -\frac{1}{2} \text{tr}(P V_a) + \frac{1}{2} \bar{y}_\alpha P V_a P \bar{y}_\alpha; \quad a, b = 1, 2$$

$$F_{ab} = \frac{1}{2} \text{tr}(P V_a P V_b); \quad a, b = 1, 2$$

sendo:

$$P = V^{-1} - V^{-1} X (X^t V^{-1} X)^{-1} X^t V^{-1}$$

$$V_1 = \frac{\partial V}{\partial \sigma_u^2} = \Omega(\rho)$$

$$V_2 = \frac{\partial V}{\partial \rho} = \sigma_u^2 \text{diag}_{1 \leq d \leq D} \left( \frac{\partial \Omega_d(\rho)}{\partial \rho} \right) = \sigma_u^2 \text{diag}_{1 \leq d \leq D} (\dot{\Omega}_d(\rho))$$

e

$$\dot{\Omega}_d(\rho) = \frac{1}{1 - \rho^2} \begin{pmatrix} 0 & 1 & \dots & (T-2)\rho^{T-3} & (T-1)\rho^{T-2} \\ 1 & 0 & \dots & \dots & (T-2)\rho^{T-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ (T-2)\rho^{T-3} & \dots & \dots & 0 & 1 \\ (T-1)\rho^{T-2} & (T-2)\rho^{T-3} & \dots & 1 & 0 \end{pmatrix} + \frac{2\rho\Omega_d(\rho)}{(1 - \rho^2)^2}$$

3. Fixado un certo nivel de tolerancia  $\epsilon$ , continuar mentras que se cumpla:

$$\left| \frac{\hat{\sigma}_u^{2(k)} - \hat{\sigma}_u^{2(k-1)}}{\hat{\sigma}_u^{2(k-1)}} \right| \leq \epsilon; \left| \frac{\hat{\rho}^{2(k)} - \hat{\rho}^{2(k-1)}}{\hat{\rho}^{2(k-1)}} \right| \leq \epsilon$$

4. Os estimadores buscados son:

$$\hat{\sigma}_u^2 = \hat{\sigma}_u^{2(k)}$$

$$\hat{\rho}^2 = \hat{\rho}^{2(k)}$$

### Estimación do erro cuadrático medio

Para estimar o erro cuadrático medio do estimador  $\hat{Y}_{\alpha, dt}$  propóñense dous estimadores: o primeiro deles está baseado na aproximación de Prasad-Rao (1990) e foi tomado de Esteban et al. (2011):

$$ecm_1(\hat{Y}_{\alpha, dt}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}) \quad (4.17)$$

$$\begin{aligned} g_1(\hat{\theta}) &= \sigma_u^2 a_d^t \Omega_d a_d - \sigma_u^4 a_d^t \Omega_d V_d^{-1} \Omega_d a_d \\ g_2(\hat{\theta}) &= \left[ a_d^t X_d - \sigma_u^2 a_d^t \Omega_d V_{ed}^{-1} X_d + \sigma_u^4 a_d^t \Omega_d V_d^{-1} \Omega_d V_{ed}^{-1} X_d \right] Q \\ &\quad \left[ X_d^t a_d - \sigma_u^2 X_d^t V_{ed}^{-1} \Omega_d a_d + \sigma_u^4 X_d^t V_{ed}^{-1} \omega_d V_d^{-1} \Omega_d a_d \right] \\ g_3(\hat{\theta}) &\approx tr \left\{ \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}^{-1} \right\} \end{aligned}$$

Sendo  $F_{ab}$  os elementos da matriz de información de Fisher,  $a_d = col_{1 \leq k \leq T}(\delta_{tk})$ , e  $\delta_{ij}$  a función delta de Kronecker que toma o valor 1 cando  $i = j$  e 0 noutro caso.

$$q_{11} = a_d^t \Omega_d V_d^{-1} \Omega_d a_d - 2\sigma_u^2 a_d^t \Omega_d V_d^{-1} \Omega_d V_d^{-1} \Omega_d a_d + \sigma_u^4 a_d^t \Omega_d V_d^{-1} \Omega_d V_d^{-1} \Omega_d V_d^{-1} \Omega_d a_d,$$

$$\begin{aligned} q_{12} &= \sigma_u^2 a_d^t \Omega_d V_d^{-1} \dot{\Omega}_d a_d - \sigma_u^4 a_d^t \Omega_d V_d^{-1} \dot{\Omega}_d V_d^{-1} \Omega_d a_d - \\ &\quad \sigma_u^4 a_d^t \Omega_d V_d^{-1} \Omega_d V_d^{-1} \dot{\Omega}_d a_d + \sigma_u^6 a_d^t \Omega_d V_d^{-1} \Omega_d V_d^{-1} \dot{\Omega}_d V_d^{-1} \Omega_d a_d \end{aligned}$$

$$q_{22} = \sigma_u^4 a_d^t \dot{\Omega}_d V_d^{-1} \dot{\Omega}_d a_d - 2\sigma_u^6 a_d^t \Omega_d V_d^{-1} \dot{\Omega}_d V_d^{-1} \dot{\Omega}_d a_d + \sigma_u^8 a_d^t \Omega_d V_d^{-1} \dot{\Omega}_d V_d^{-1} \dot{\Omega}_d V_d^{-1} \Omega_d a_d$$

Alternativamente propoñemos outro estimador para o erro cuadrático medio empregando un bootstrap paramétrico que se detalla a continuación:

1. Obter estimacións dos parámetros do modelo:  $\hat{\theta} = (\hat{\sigma}_u^2, \hat{\rho})$  e  $\hat{\beta}$
2. Repetir para  $b = 1, 2, \dots, B$

- Xerar  $D \cdot T$  variables  $v_{dt}^{*b}$  independentes e idénticamente distribuídas con distribución Normal de media cero e varianza  $\hat{\sigma}_u^2$ .
- Xerar a variable

$$u_{d1}^{*b} = \frac{1}{\sqrt{1 - \hat{\rho}^2}} v_{d1}^{*b}$$

$$u_{dt}^{*b} = \hat{\rho} u_{d(t-1)}^{*b} + v_{dt}^{*b}, \text{ con } d = 1, 2, \dots, D; t = 1, 2, \dots, T$$

- Xerar  $D \cdot T$  variables  $e_{dt}^{*b}$  independentes e idénticamente distribuídas con distribución Normal de media cero e varianza  $\sigma_{e_{dt}}^2$ .
- Xerar o modelo bootstrap:  $\bar{y}_{\alpha, dt}^{*b} = x_{dt} \hat{\beta} + u_{dt}^{*b} + e_{dt}^{*b}$ ,  $d = 1, \dots, D; t = 1, \dots, T$
- Definimos o verdadeiro parámetro da poboación bootstrap:

$$\bar{Y}_{\alpha, dt}^{*b} = x_{dt} \hat{\beta} + u_{dt}^{*b}$$

- Obtemos o estimador baseado no modelo,  $\hat{Y}_{\alpha, dt}^{*b}$  tal e como se describe nas seccións anteriores.

3. Obter o estimador bootstrap do erro cuadrático medio de  $\hat{Y}_{\alpha, dt}$ :

$$ecm_2(\hat{Y}_{\alpha, dt}) = \frac{1}{(B-1)} \sum_{b=1}^B \left( \hat{Y}_{\alpha, dt}^{*b} - \bar{Y}_{\alpha, dt}^{*b} \right)^2 \quad (4.18)$$

# Capítulo 5

## Estudio de simulación

Co obxectivo de avaliar o rendemento dos estimadores propostos antes da súa aplicación práctica, leváronse a cabo dous experimentos de simulación. O obxectivo é avaliar a mellora ou non en termos do erro cometido ao empregar os estimadores indirectos fronte a utilización dos convencionais estimadores directos. Hai que ter en conta, que tódolos estimadores indirectos propostos son estimadores baseados nun certo modelo. Un punto fundamental a ter en conta á hora de empregar estes métodos é a dispoñibilidade de boas variables auxiliares.

Realizáronse dous estudos:

- Simulación 1: prescínlese do efecto tempo e lévase a cabo un estudo comparativo entre os estimadores indirectos de Fay-Herriot e o estimador EB, fronte ao estimador directo.
- Simulación 2: Introdúcese a compoñente temporal e realizamos un estudo comparativo entre o Modelo 0 (efecto tempo independente) e o Modelo 1 (efecto tempo con autocorrelación de orde 1), co obxectivo de ver a gañancia ou non de acuracidade co modelo máis complexo que ten parámetro máis.

O motivo para realizar dúas simulacións é principalmente, a imposibilidade de ter un control adecuado sobre o valor que toma  $\rho$  (empregado nos modelos con efecto tempo) xerando unha poboación a nivel unidade (necesario para obter o estimador EB).

### 5.1. Simulación 1

Para un período temporal concreto, quérese avaliar os beneficios de empregar algún dos dous estimadores indirectos propostos para estimar os indicadores de

pobreza de Foster, Greer e Thorbecke. Hai que ter en conta que ambos estimadores se basean na existencia dun certo modelo de superpoboación que seguen os datos dispoñibles. Un inconveniente á hora de simular a poboación é que un dos estimadores se basea nun modelo a nivel de área e outro a nivel de individuo. Este problema solvéntase xerando unha poboación a nivel de unidade.

Esta primeira simulación pode dividirse en dúas partes, en primeiro lugar comparamos o estimador directo, o estimador de Fay-Herriot e o estimador EB. A continuación avaliamos os estimadores propostos do erro cuadrático medio de estos estimadores. En particular, comparamos os estimadores analítico e bootstrap do estimador do erro cuadrático medio do estimador de Fay-Herriot.

### 5.1.1. Poboación simulada

Co obxecto de avaliar o rendemento dos estimadores propostos, xenerouse unha poboación simulada a nivel de unidade que recolle información sobre a variable obxectivo (os ingresos equivalentes mensuais) e tamén sobre tres variables auxiliares. Escolléronse as variables auxiliares seguintes por ser variables que están correlacionadas cos índices de *FGT* a nivel comarcal na poboación verdadeira, ademais de ser variables explicativas da incidencia da pobreza. As variables que forman a poboación simulada son as seguintes:

- Código de comarca: *Cod\_comarca*
- Ingresos equivalentes mensuais: *ingresos\_equiv*
- Variable indicadora de pobreza: *pobre*.
- Variable indicadora de nacionalidade española:  $X_1$
- Variable indicadora de ter 65 ou máis anos:  $X_2$
- Variable indicadora de estar en situación de paro:  $X_3$

En esta poboación simulada consideramos  $D = 53$  áreas pequenas que identificamos coas comarcas galegas, cuxos tamaños poboacionais ( $N_d$ ) se obteñen en base ás cifras oficiais de poboación de Galicia subministradas polo Padrón Municipal de Habitantes do ano 2008. Para reducir o tempo de execución dos algoritmos de simulación, os tamaños poboacionais fixémoslos no 25 % dos tamaños poboacionais reais de cada comarca. Así, a poboación simulada consta de  $N = 696.042$  individuos distribuídos en 53 comarcas, cuxos tamaños oscilan dende os 976 a os 105.140 individuos. Os tamaños poden consultarse no Cadro 5.1.

O procedemento para xerar as variables foi o seguinte:



Comarca	N_d	n_d	Comarca	N_d	n_d	Comarca	N_d	n_d
1	976	8	19	4.159	46	37	10.606	100
2	1.494	12	20	6.915	53	38	7.448	113
3	4.263	15	21	7.139	64	39	11.261	117
4	3.735	20	22	5.368	66	40	7.642	137
5	2.190	20	23	3.072	67	41	8.446	140
6	1.279	22	24	5.968	78	42	11.406	141
7	3.695	23	25	9.652	78	43	17.572	142
8	2.567	24	26	3.789	81	44	16.918	149
9	1.762	24	27	8.992	82	45	20.625	197
10	6.756	27	28	5.950	86	46	27.558	227
11	1.584	29	29	6.273	87	47	30.534	338
12	3.390	35	30	7.135	87	48	29.658	357
13	4.365	36	31	8.820	87	49	39.719	389
14	2.775	41	32	4.470	88	50	40.790	391
15	4.992	42	33	7.143	89	51	36.092	465
16	1.625	43	34	9.865	92	52	97.322	709
17	3.907	45	35	12.503	92	53	105.140	803
18	3.841	46	36	4.896	98	Total	696.042	6.848

Cadro 5.1: Simulación 1: Tamaños mostrais e e poboacionais

- Xeramos tres variables indicadoras ( $X_i, i = 1, 2, 3$ ) mediante distribucións de Bernouilli

$$X_{idj} \sim Ber(p_{id}), j = 1 \dots N_d, d = 1, \dots D$$

con parámetros ( $p_{id}$ ) dados polos valores das proporcións de españois, a proporción de maiores de 64 anos e a proporción de parados<sup>1</sup>, respectivamente, na comarca  $d$ , amosados no Cadro 5.2

- Xeramos as variables:

$$u_d \sim N(0, \sigma_u) \quad \text{con} \quad \sigma_u = 0,2$$

$$e_{dj} \sim N(0, \sigma_e) \quad \text{con} \quad \sigma_e = 0,5$$

- Xeramos a variable resposta:

$$y_{dj} = X_{dj}^t \beta + u_d + e_{dj}, j = 1 \dots N_d, d = 1, \dots D$$

sendo  $X_{dj}^t = (1, X_{1dj}, X_{2dj}, X_{3dj})$  e  $\beta = (6.9, 0.01, -0.005, -0.002)$

Esta variable podemos identificala co logaritmo dos ingresos equivalentes (variable a empregar á hora de obter o estimador EB)

<sup>1</sup>Obtida como o cociente entre parados rexistrados(Consellería de Asuntos Sociais, Emprego e Relacións Laborais) e poboación entre 16 e 64 anos (Padron Municipal de Habitantes)

Comarca	p1_d	p2_d	p3_d	Comarca	p1_d	p2_d	p3_d
1	0,990	0,451	0,055	28	0,969	0,352	0,103
2	0,980	0,337	0,062	29	0,976	0,315	0,064
3	0,989	0,217	0,099	30	0,921	0,304	0,147
4	0,980	0,331	0,069	31	0,984	0,208	0,103
5	0,955	0,396	0,096	32	0,988	0,276	0,068
6	0,989	0,404	0,086	33	0,986	0,283	0,073
7	0,989	0,283	0,091	34	0,975	0,264	0,067
8	0,985	0,354	0,067	35	0,963	0,180	0,106
9	0,940	0,358	0,114	36	0,973	0,327	0,090
10	0,981	0,214	0,077	37	0,955	0,190	0,115
11	0,981	0,363	0,082	38	0,965	0,312	0,094
12	0,988	0,290	0,079	39	0,960	0,255	0,053
13	0,965	0,284	0,058	40	0,952	0,250	0,052
14	0,980	0,276	0,103	41	0,965	0,348	0,098
15	0,984	0,244	0,105	42	0,984	0,317	0,062
16	0,989	0,390	0,055	43	0,982	0,227	0,107
17	0,988	0,251	0,093	44	0,977	0,184	0,082
18	0,972	0,339	0,066	45	0,980	0,157	0,102
19	0,969	0,308	0,081	46	0,971	0,175	0,098
20	0,956	0,251	0,067	47	0,965	0,179	0,088
21	0,934	0,260	0,083	48	0,958	0,221	0,075
22	0,970	0,382	0,113	49	0,964	0,165	0,083
23	0,986	0,378	0,056	50	0,980	0,223	0,099
24	0,991	0,222	0,108	51	0,960	0,227	0,091
25	0,976	0,235	0,076	52	0,956	0,185	0,079
26	0,978	0,339	0,074	53	0,954	0,164	0,098
27	0,989	0,229	0,087				

Cadro 5.2: Simulación 1: Parámetros das variables de Bernoilli empregadas para simular as variables auxiliares

- Obtemos os ingresos equivalentes da poboación simulada:

$$ingresos\_equiv_{dj} = \exp(y_{dj}), j = 1 \dots N_d, d = 1, \dots D$$

A mediana de ingresos na poboación simulada é 1038,25 euros, e o limiar de pobreza 622,95 euros. É dicir, as persoas cuxos ingresos equivalentes sexan inferiores a esta cantidade, consideraríanse pobres. Con este limiar de pobreza, a taxa de risco de pobreza, ou proporción de persoas pobres, ( $FGT_0$ ) sitúase en 0,1714.

### 5.1.2. Avaliación dos estimadores propostos

A partir da poboación simulada, extraemos  $K = 1000$  mostras empregando mostraxe aleatoria simple sen reemplazamento dentro de cada comarca. O tamaño mostral ( $n_d$ ) en cada comarca obtívose aplicando a fracción de mostraxe ( $n_d/N_d$ ) da ECV aos tamaños poboacionais. Os tamaños mostrais e poboacionais poden consultarse no Cadro 5.1. As áreas preséntanse ordenadas en función de tamaño mostral crecente. O procedemento seguido fai que existan algunhas áreas cun tamaño mostral moi reducido, polo que é pertinente a aplicación da metodoloxía explicada na sección previa.

Empregando cada unha destas mostras calculamos os tres estimadores propostos: o estimador directo,  $\hat{Y}_d^{DIR}$ , e dous estimadores indirectos: un baseado nun modelo de área (estimador de Fay-Herriot),  $\hat{Y}_d^{FH}$ , e outro baseado nun modelo de individuo (estimador EB),  $\hat{Y}_d^{EB}$ . Para obter o estimador EB, tomouse  $L = 50$ . Este valor de  $L$  fíxose por motivos operativos relacionados co elevado tempo que se necesita para obter o citado estimador. No artigo tomado de referencia, (ver Molina & Rao, 2010), díse que a elección de  $L = 50$  proporciona resultados bastante precisos, aínda que recomentan que nas aplicacións prácticas, cando se traballe con mostras reais, se empreguen valores superiores tales como  $L \geq 200$ .

Sexa  $\bar{Y}_d$  o parámetro obxectivo na comarca  $d$  e sexa  $\hat{Y}_d^k$  o correspondente estimador na comarca  $d$  e na réplica  $k$ . Calculamos as seguintes medidas de eficiencia de cada un dos tres estimadores propostos:

**Avaliación do nesgo :**

- Nesgo:  $NESGO_d = \frac{1}{K} \sum_{k=1}^K \left( \hat{Y}_d^k - \bar{Y}_d \right)$
- Nesgo relativo:  $NESGOR_d = \frac{NESGO_d}{\bar{Y}_d} \times 100$

**Avaliación do erro cuadrático medio :**

- Erro cuadrático medio :  $ECM_d = \frac{1}{K} \sum_{k=1}^K \left( \hat{Y}_d^k - \bar{Y}_d \right)^2$
- Raíz do erro cuadrático medio relativo :  $RECMR_d = \frac{1}{\bar{Y}_d} \sqrt{ECM_d} \times 100$

Obtemos tamén os estimadores promedio no total das  $K$  mostras:

$$\hat{\bar{Y}}_d = \frac{1}{K} \sum_{k=1}^K \hat{Y}_d^k$$

Na Figura 5.1 represéntase o boxplot do  $NESGOR_d$  de cada un dos tres estimadores propostos, e na Figura 5.2 os correspondentes boxplot do  $RECMR_d$ . Vemos

que o estimador directo é aproximadamente inesgado, mentres que os dous estimadores indirectos presentan unha maior variabilidade no que se refire ao seu nesgo. Sendo ademais, a distribución do nesgo relativo do estimador EB máis dispersa e asimétrica. Por outra banda, en termos do erro cuadrático medio relativo, ámbolos dous estimadores indirectos presentan valores moito mellores que o estimador directo. Isto pode verse mellor no Cadro 5.3 onde se amosa un resumo da distribución da raíz do erro cuadrático medio relativo ( $RECMR_d$ ). A mediana do erro relativo (medido pola raíz do erro cuadrático medio) é superior a 23 empregando o estimador directo, sendo máis de cinco puntos inferior cando se emprega calquera dos dous estimadores indirectos. Así mesmo a distribución do erro relativo dos dous estimadores indirectos é máis homoxénea que a correspondente do estimador directo. É dicir, en termos xerais, podemos dicir que a pesar de que os estimadores indirectos presentan un maior nesgo, en termo do erro cuadrático medio son notablemente mellores que o estimador directo.

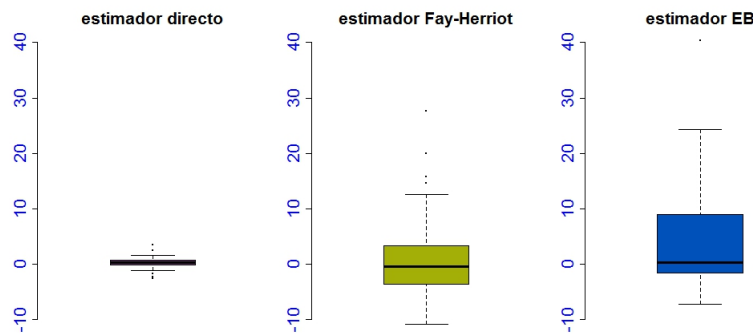


Figura 5.1: Simulación 1: Nesgo relativo dos tres estimadores propostos

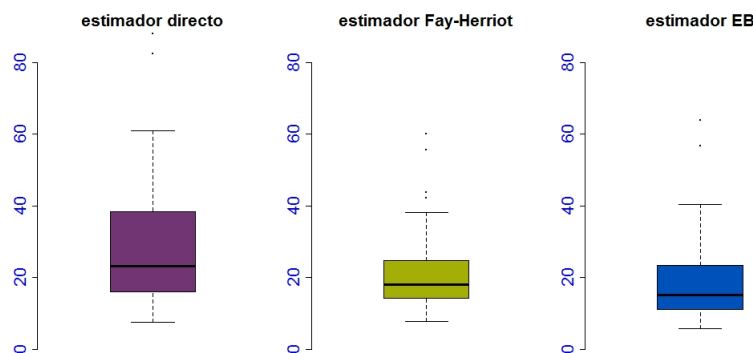


Figura 5.2: Simulación 1: Raíz do erro cuadrático medio dos tres estimadores propostos

Na Figura 5.3 detállase o erro cuadrático medio ( $ECM_d$ ) de cada estimador para cada comarca da poboación simulada. Tamén no detalle por comarca se observa

que os estimadores indirectos presentan un menor erro. A ganancia de acuracidade é maior naquelas comarcas onde os tamaños mostrais eran máis pequenos.

ESTIMADOR	Mínimo	D1	Q1	Mediana	Media	Q3	D9	Máximo
Est. directo	7,54	13,19	16,05	23,15	29,70	38,44	51,71	92,39
Est. Fay-Herriot	7,64	11,46	14,26	17,93	21,71	24,76	37,77	60,03
Est. EB-Method	5,83	9,27	11,10	15,21	18,82	23,32	31,91	63,89

Cadro 5.3: Distribución da raíz do erro cuadrático medio relativo

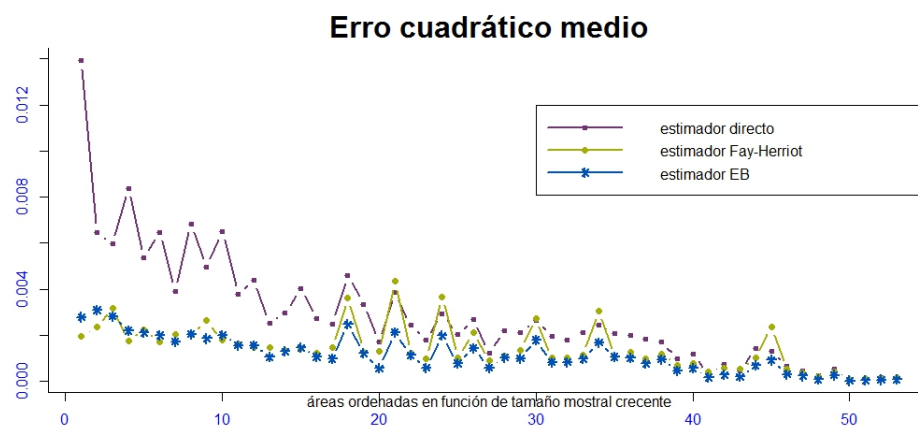


Figura 5.3: Simulación 1: Erro cuadrático medio ( $ECM_d$ ) dos estimadores propostos

Por último, na Figura 5.4 representábase o valor da taxa de risco de pobreza en cada comarca da poboación simulada fronte aos valores promedio ( $\hat{Y}_d$ ) dos dous estimadores indirectos propostos.

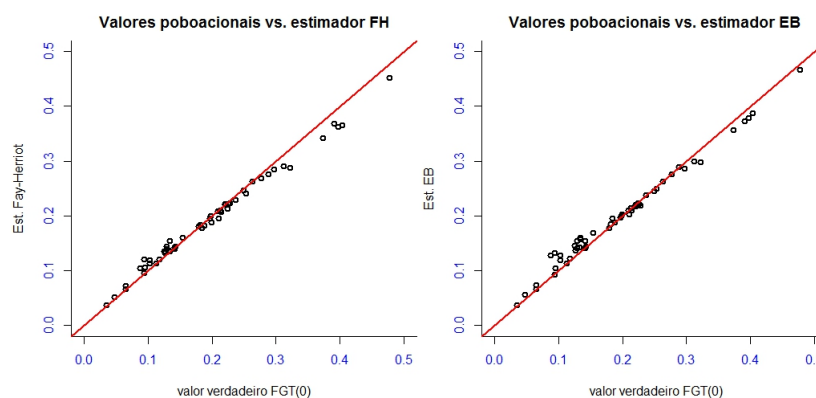


Figura 5.4: Simulación 1: Valores poboacionais vs. estimadores indirectos propostos

### 5.1.3. Avaliación dos estimadores do erro cuadrático medio

Sexa  $ecm_d^k$  o estimador de  $ECM_d$  na comarca  $d$  e na réplica  $k$ . Co obxecto de avaliar o comportamento dos estimadores propostos para o erro cuadrático medio, obtemos as seguintes medidas de eficiencia:

- Erro cuadrático medio do estimador do erro cuadrático medio:

$$ECM(ecm_d) = \frac{1}{K} \sum_{k=1}^K (ecm_d^k - ECM_d)^2$$

- Raíz do erro cuadrático medio do estimador do erro cuadrático medio:

$$RECMR(ecm_d) = \frac{1}{ECM_d} \sqrt{ECM(ecm_d)} \times 100$$

Obtemos tamén os estimadores promedio no total das  $K$  mostras:

$$\overline{ecm}_d = \frac{1}{K} \sum_{k=1}^K ecm_d^k$$

Avaliamos en primeiro lugar o estimador proposto para o erro cuadrático medio do estimador directo (4.3). Na Figura 5.5 represéntase o valor real do erro cuadrático medio do estimador directo ( $ECM_d$ ) fronte ao seu estimador  $\overline{ecm}_d$ . Observamos que o estimador proposto está moi próximo ao valor real, incluso nas comarcas de menor tamaño mostral.

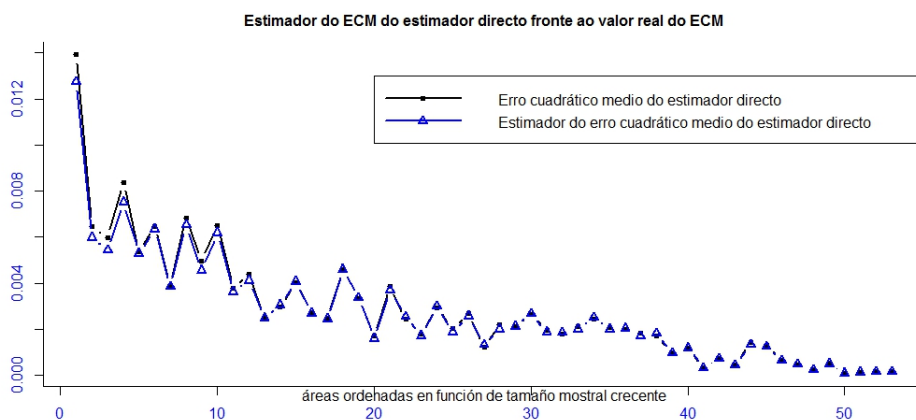


Figura 5.5: Simulación 1: Comparativa entre o estimador do erro cuadrático medio do estimador directo e o verdadeiro valor do erro cuadrático medio

Para estimar o erro cuadrático medio do estimador de Fay-Herriot propoñíanse dúas alternativas: un estimador analítico (4.8) e un estimador bootstrap (4.9).

Ambos estimadores compáranse co verdadeiro valor do erro cuadrático medio ( $ECM_d$ ) na Figura 5.6. Obsérvase nesta gráfica que ámbolos dous estimadores toman valores moi próximos entre si, e que ambos estiman bastante ben o comportamento do verdadeiro erro cuadrático medio. Se ben, obsérvase que nas comarcas de menor tamaño, a tendencia de ambos estimadores é a sobreestimar o verdadeiro valor do erro cuadrático medio. Na Figura 5.7 preséntase o erro cuadrático medio de cada un dos estimadores do erro cuadrático medio do estimador de Fay-Herriot. Observamos que o erro cometido ao empregar ambos estimadores é moi semellante, sendo o estimador Bootstrap máis preciso en algunha das comarcas. Neste exercizo de simulación, para obter o estimador bootstrap do erro cuadrático medio do estimador de Fay-Herriot, empregouse  $B = 3000$ .

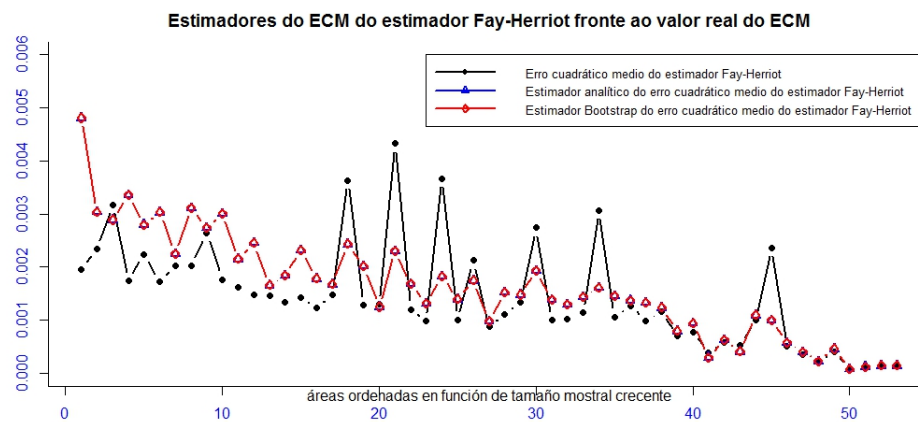


Figura 5.6: Simulación 1: Comparativa entre os estimadores do erro cuadrático medio do estimador Fay-Herriot e o verdadeiro valor do erro cuadrático medio

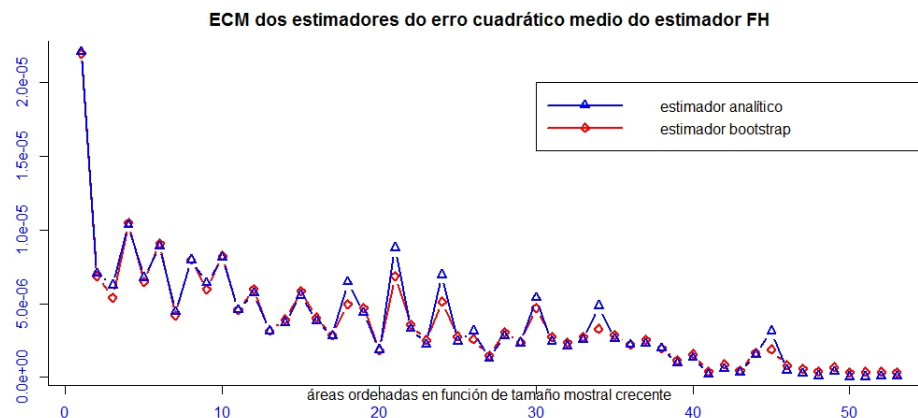


Figura 5.7: Simulación 1: Comparativa entre os estimadores bootstrap e analítico do ECM do estimador de Fay-Herriot

Por último, avalíase o estimador proposto para o erro cuadrático medio do estimador EB (4.12). Para obter este estimador tomouse  $B = 500$ . Na Figura 5.8 represéntase o valor real do erro cuadrático medio do estimador EB fronte ao seu estimador  $\overline{ecm}_d$ . Obsérvase tamén neste caso que o estimador proposto ten en xeral un bo comportamento. Nótase ademais que o estimador do erro cuadrático medio ten tendencia a sobreestimar o verdadeiro valor do erro cuadrático medio nas comarcas de menor tamaño.

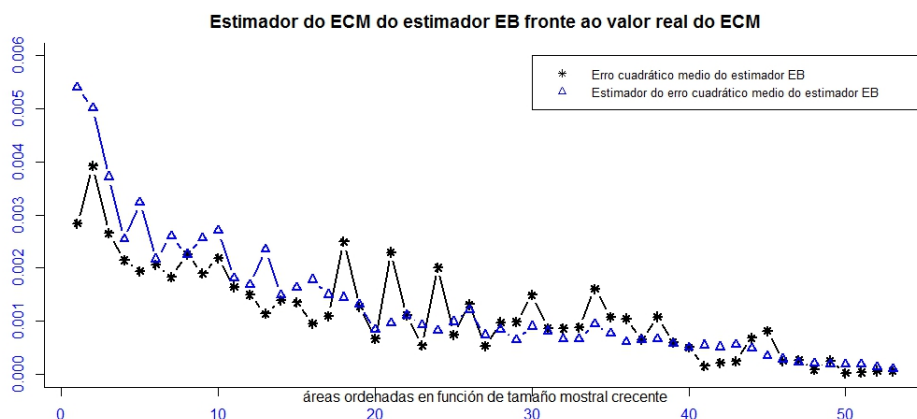


Figura 5.8: Simulación 1: Comparativa entre o estimador do erro cuadrático medio do estimador EB e o verdadeiro valor do erro cuadrático medio

## 5.2. Simulación 2

Este segundo experimento de simulación, no cal se inclúe o efecto tempo no modelo, tamén se pode dividir en dúas partes. En primeiro lugar, avaliamos as ventaxas de considerar o modelo máis complexo (Modelo 1) fronte ao modelo máis simple (Modelo 0). A continuación, comparamos os dous estimadores do erro cuadrático medio propostos: o estimador analítico baseado na aproximación de Prasad-Rao e o estimador bootstrap.

### 5.2.1. Poboación simulada

Neste segundo experimento de simulación consideramos tamén  $D = 53$  áreas pequenas, que identificamos coas comarcas. Consideramos tres horizontes temporais:  $T = 3$ ,  $T = 6$  e  $T = 10$  e catro posibilidades no que se refire ao valor de  $\rho$  ( $\rho = 0$ ,  $\rho = 0,2$ ,  $\rho = 0,5$ ,  $\rho = 0,7$ ). Empregamos tres variables auxiliares ( $X_{idt}$ ,  $i = 1, 2, 3$ ;  $d = 1, 2, \dots, D$ ;  $t = 1, 2, \dots, T$ ), que identificamos respectivamente coa proporción de españois na comarca  $d$  e no período  $t$ , a proporción de perso-



as con 65 ou máis anos na comarca  $d$  e no período  $t$  e a proporción de parados<sup>2</sup> na comarca  $d$  e no período  $t$ . Empregáronse os valores poboacionais referidos a estas variables obtidos do Padrón e da Consellería de Asuntos Sociais, Emprego e Relacións Laborais para os anos 2000 a 2009.

Partindo de estas premisas, o procedemento foi o seguinte:

- Xeramos a variable:

$$e_{dt} \sim N(0, \sigma_{edt}) \quad \text{sendo} \quad \sigma_{edt}^2 = \frac{d}{D} + \frac{t}{T}; \quad \text{con} \quad d = 1, 2, \dots, D; t = 1, 2, \dots, T$$

- Xeramos a variable:

$$u_{d1} = \frac{1}{\sqrt{1-\rho^2}} v_{d1};$$

$$u_{dt} = \rho u_{d(t-1)} + v_{dt}, \quad d = 1, 2, \dots, D; t = 1, 2, \dots, T;$$

$$v_{dt} \sim N(0, 1)$$

- Xeramos a variable resposta:

$$y_{dt} = \beta_1 X_{1dt} + \beta_2 X_{2dt} + \beta_3 X_{3dt} + u_{dt} + e_{dt} \quad d = 1, 2, \dots, D; t = 1, 2, \dots, T.$$

$$\text{Sendo } \beta = (\beta_1, \beta_2, \beta_3) = (-4, 3, 7)$$

### 5.2.2. Avaliación dos estimadores propostos

Para cada  $T$  e  $\rho$ , extraemos  $K = 1000$  mostras co procedemento descrito na sección anterior. Para cada unha de estas mostras ( $k = 1, \dots, 1000$ ), obtemos o valor poboacional do parámetro,  $\bar{Y}_{dt}^k$ , o estimador obtido de aplicar o Modelo 0 e o estimador obtido de aplicar o Modelo 1. Denotamos por  $\hat{Y}_{dt}^k$  os estimadores. Calculamos as seguintes medidas de eficiencia:

$$ECM_{dt} = \frac{1}{K} \sum_{k=1}^K \left( \hat{Y}_{dt}^k - \bar{Y}_{dt}^k \right)^2$$

$$ECM_d = \frac{1}{T} \sum_{t=1}^T ECM_{dt}$$

Os resultados en termos do erro cuadrático medio dos distintos estimadores, amósanse nas Figuras 5.9, 5.10 e 5.11 onde se representa o boxplot da variable  $ECM_d$

<sup>2</sup>Obtida como o cociente entre parados rexistrados(Consellería de Asuntos Sociais, Emprego e Relacións Laborais) e poboación entre 16 e 64 anos (Padron Municipal de Habitantes)

para cada combinación de  $T$  e  $\rho$ . Obsérvase que o estimador do Modelo 1 presenta unha mellora significativa no que respecta á acuracidade do estimador cando a poboación de partida presenta dependencia temporal con  $\rho > 0,5$ . Cando a magnitude de  $\rho$  é menor, a gañancia de acuracidade co estimador obtido do Modelo 1 non é tan elevada, se ben, faise máis notable a medida que o número de períodos temporais ( $T$ ) considerados se incrementa. Só no caso de que  $\rho = 0$  o estimador obtido do Modelo 0 é mellor que o obtido empregando o Modelo 1. Sen embargo, a perda de precisión ao empregar o modelo equivocado é pequena, como se pode ver nas gráficas anteriores.

Nas Figuras 5.12, 5.13 e 5.14 represéntanse os valores do erro cuadrático medio obtido para cada comarca ( $ECM_d$ ) para os distintos valores de  $T$  e  $\rho$  da poboación e confírmase o comportamento descrito polos boxplots anteriores. Cando o parámetro de autocorrelación toma un valor elevado ( $\rho > 0,5$ ), o Modelo 1 que ten en conta este tipo concreto de dependencia temporal é claramente mellor que o Modelo 0. A gañancia de precisión é maior canto maior sexa o número de períodos temporais ( $T$ ) dispoñibles. Pola contra, se a autocorrelación é débil, apenas hai diferenzas entre o emprego do estimador baseado no modelo máis complexo e no máis simple.

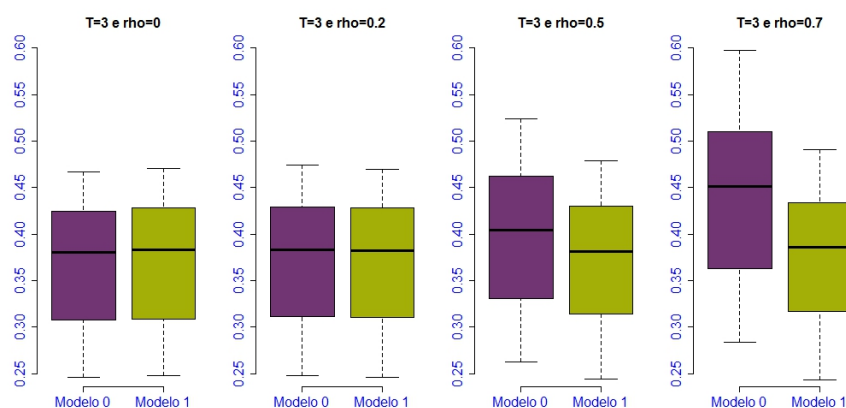


Figura 5.9: Simulación 2: Boxplot do  $ECM_d$  dos modelos 0 e 1 con  $T=3$  períodos para distintos valores de  $\rho$

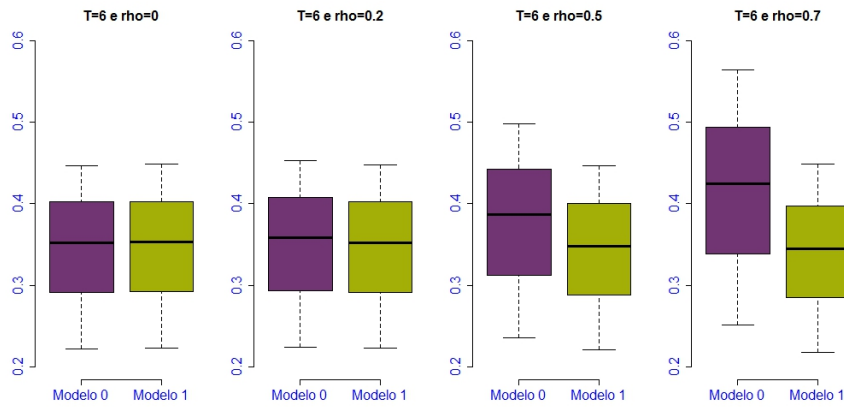


Figura 5.10: Simulación 2: Boxplot do  $ECM_d$  dos modelos 0 e 1 con  $T=6$  períodos para distintos valores de  $\rho$

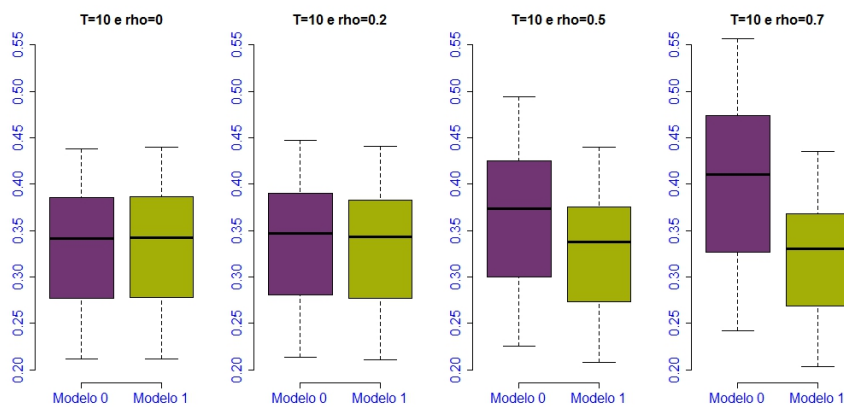


Figura 5.11: Simulación 2: Boxplot do  $ECM_d$  dos modelos 0 e 1 con  $T=10$  períodos para distintos valores de  $\rho$

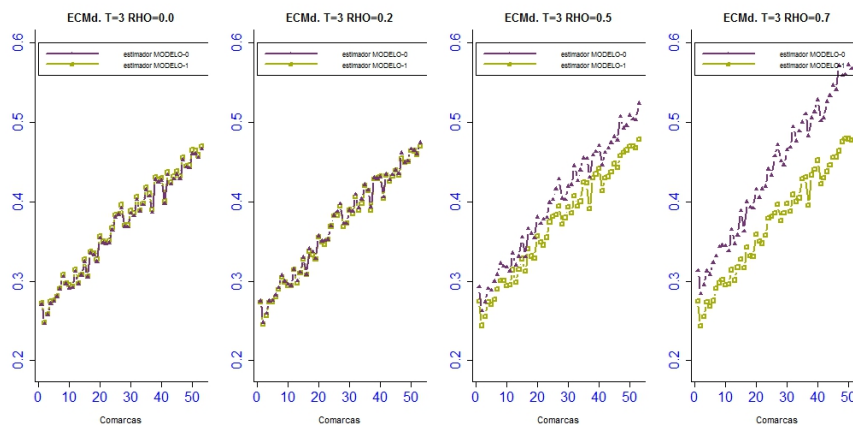


Figura 5.12: Simulación 2:  $ECM_d$  dos modelos 0 e 1 con  $T=3$  períodos para distintos valores de  $\rho$

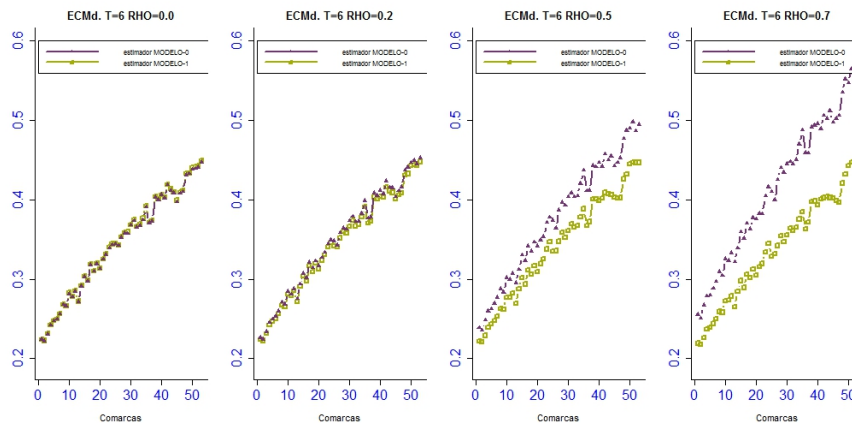


Figura 5.13: Simulación 2:  $ECM_d$  dos modelos 0 e 1 con  $T=6$  períodos para distintos valores de  $\rho$

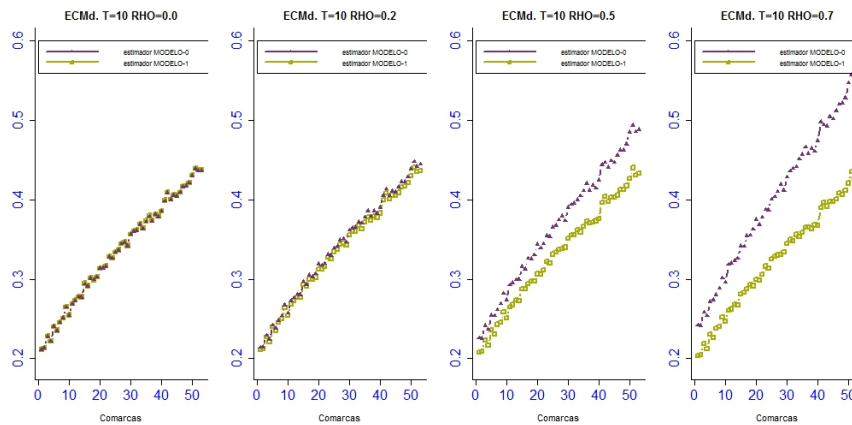


Figura 5.14: Simulación 2:  $ECM_d$  dos modelos 0 e 1 con  $T=10$  períodos para distintos valores de  $\rho$

### 5.2.3. Avaliación dos estimadores do erro cuadrático medio dos estimadores propostos

Sexa  $ecm_{dt}^k$  o estimador de  $ECM_{dt}$  na comarca  $d$ , período  $t$  e na réplica  $k$ . Co obxecto de avaliar o comportamento dos estimadores propostos para o erro cuadrático medio, obtemos as seguintes medidas de eficiencia:

#### Avaliación do erro cuadrático medio

$$ECM(ecm_{dt}) = \frac{1}{K} \sum_{k=1}^K (ecm_{dt}^k - ECM_{dt})^2$$

$$ECM(ecm_d) = \frac{1}{T} \sum_{t=1}^T ECM(ecm_{dt})$$

Para obter o estimador bootstrap empregouse  $B=500$ . Nos gráficos 5.15 e 5.16 representase  $ECM(ecm_d)$  para os estimadores analítico e bootstrap propostos. Na primeira gráfica representase os resultados para o Modelo 1 e na segunda para o Modelo 0. Obsérvase en ambos casos que o estimador analítico proporciona uns valores notablemente mellores que o estimador bootstrap proposto. O estudo foi repetido para os outros valores de  $T$  sendo os resultados obtidos semellantes. Recentemente foi presentado un traballo (Morales & Esteban Lefler, 2012) como póster na SEIO 2012 no que se estudan métodos bootstrap en modelos lineais mixtos a nivel de área con efecto tempo e os resultados obtidos eran tamén semellantes aos obtidos nestes traballo. En función dos resultados obtidos, o estimador Bootstrap que se propoñía queda descartado, para a aplicación práctica.

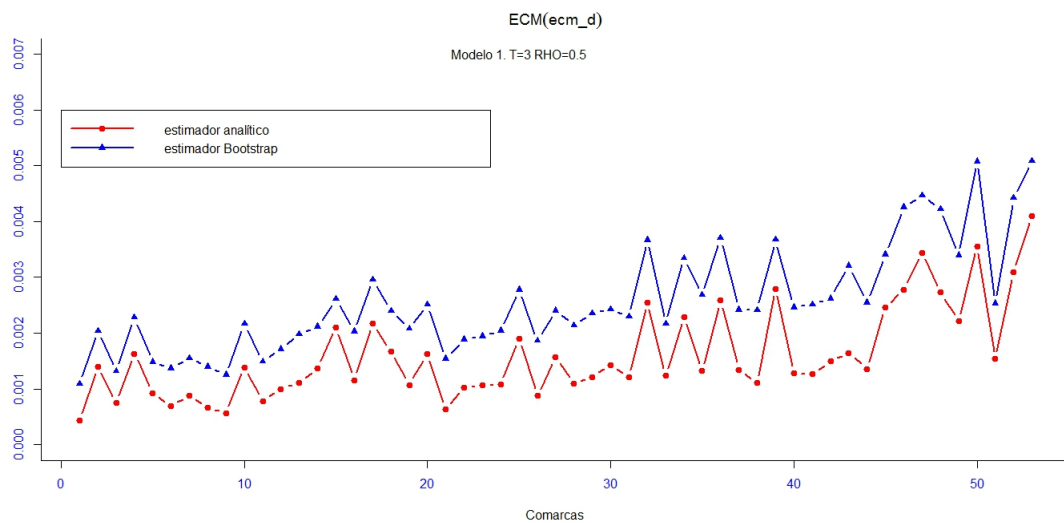


Figura 5.15: Simulación 2: ECM dos estimadores analítico e bootstrap do erro cuadrático medio do estimador do Modelo 1.  $T = 3$  e  $\rho = 0,5$

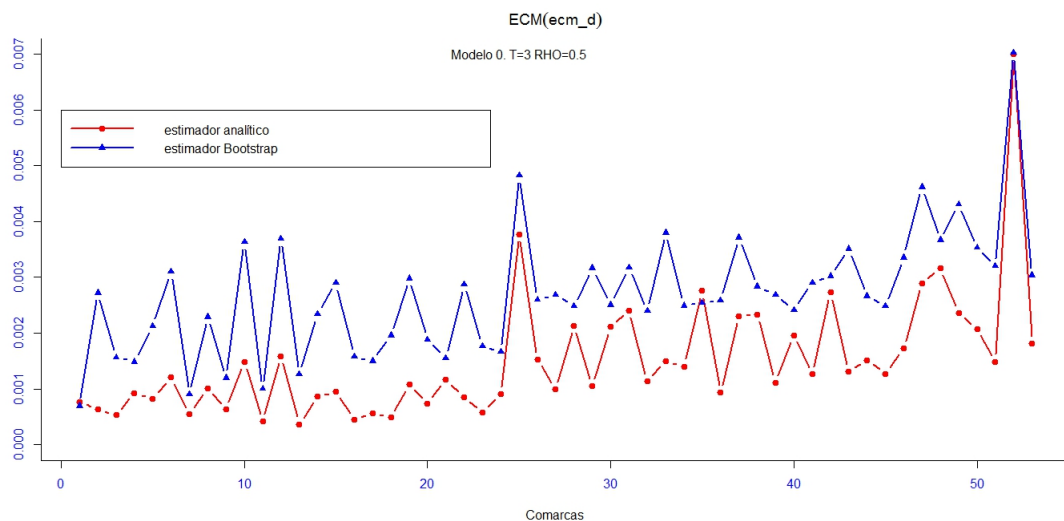


Figura 5.16: Simulación 2: ECM dos estimadores analítico e bootstrap do erro cuadrático medio do estimador do Modelo 0.  $T = 3$  e  $\rho = 0,5$

### 5.2.4. Comparativa dos modelos con efecto tempo e o modelo de Fay-Herriot

Complementariamente, obtivéronse (para cada unha das mostras  $k = 1, \dots, 1000$ ) os correspondentes estimadores de Fay-Herriot empregando modelos de Fay-Herriot independentes para cada período. Desta maneira, vemos a influencia de incluír ou non o efecto tempo dentro dun modelo a nivel de área, fronte a un modelo máis simple, onde se prescinde do tempo. Na Figura 5.17 represéntanse o erro cuadrático medio do estimador obtido do Modelo 0, do Modelo 1 e de Fay-Herriot. Amósase só a gráfica correspondente a  $T = 3$ , dado que os resultados eran practicamente idénticos para os outros valores de  $T$ . Obsérvase que o erro cometido ao empregar modelos de Fay-Herriot independentes ano a ano, é un pouco superior ao cometido empregando os modelos que inclúen a compoñente temporal. En concreto os resultados empregando modelos de Fay-Herriot independentes son moi semellantes aos obtidos empregando o chamado Modelo 0. A gañancia de precisión do Modelo 1 é notable cando o valor de  $\rho$  é elevado, sendo recomendable empregar un modelo que inclúa esta autocorrelación existente. Cando o valor de  $\rho$  é pequeno, non hai grandes diferenzas nos resultados, non compensando excesivamente a consideración de modelos máis complexos.

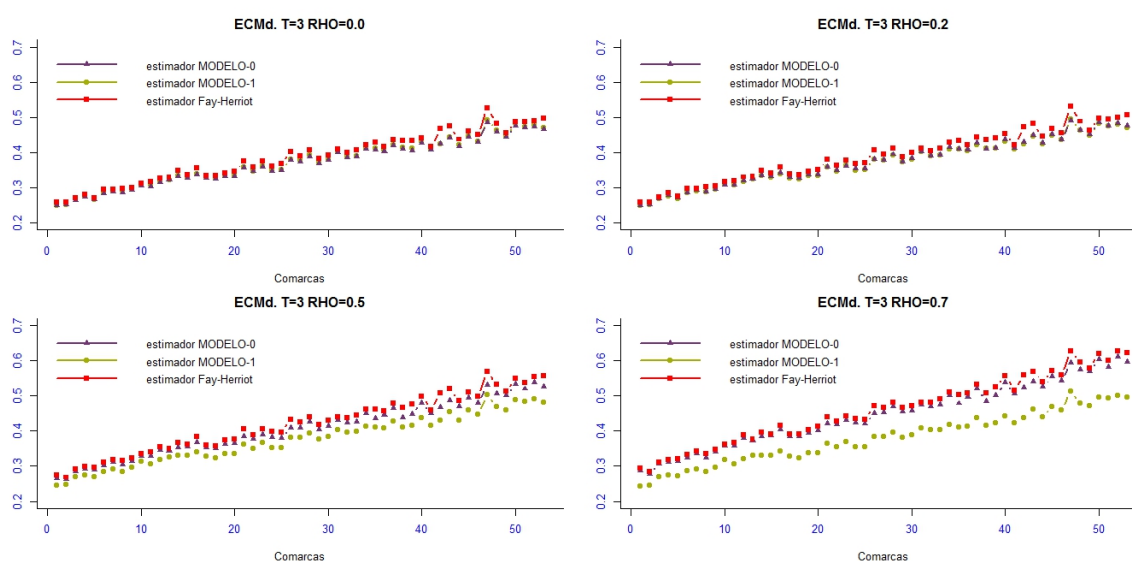


Figura 5.17: Simulación 2: ECM dos estimadores obtidos co Modelo 0, Modelo 1 e modelos Fay-Herriot independentes

## 5.3. Conclusións xerais

Dos citados estudos de simulación podemos sacar como conclusións xerais as seguintes:

- Da Simulación 1, dedúcese que tanto o estimador Fay-Herriot como o estimador EB son competitivos á hora de obter estimacións en pequenos dominios da taxa de risco da pobreza. Ámbolos dous proporcionan resultados máis acurados que o estimador directo, sendo a gañancia especialmente grande no caso das comarcas de menor tamaño.

En canto aos estimadores propostos para o erro cuadrático medio, todos eles proporcionan bos resultados. No caso dos dous estimadores propostos para o erro cuadrático medio do estimador de Fay-Herriot (un estimador analítico e un estimador bootstrap) apenas se atopan diferenzas entre eles.

- Da Simulación 2, dedúcese que, cando existe unha forte autocorrelación, a utilización do Modelo 1, máis complexo, proporciona unha notable gañancia de acuracidade. Cando o grao de autocorrelación é débil, non se observan grandes diferenzas entre a utilización do Modelo 0 ou o Modelo 1. Cando  $\rho = 0$ , a perda de precisión ao empregar o modelo erróneo é moi pequena. Na comparación destes dous modelos que teñen en conta a existencia dunha dependencia temporal (Modelo 0 e Modelo 1) coa utilización de modelos de Fay-Herriot independentes ano a ano, vemos que o erro cuadrático medio do estimador de Fay-Herriot e do estimador do Modelo 0 son moi semellantes.

En canto aos estimadores propostos para o erro cuadrático medio, o estimador bootstrap proposto é claramente peor que o estimador analítico, polo que queda descartado para a súa aplicación práctica.

---





# Capítulo 6

## Estimación de medidas de pobreza para as comarcas galegas

### 6.1. Obxectivos

Nos Capítulos 2 e 3 introducíronse os indicadores de Foster, Greer e Thorbecke e presentouse a "Enquisa de Condicións de Vida" (ECV). A partir desta enquisa obtéñense anualmente estimacións da taxa de risco de pobreza ( $FGT_0$ ), entre outros indicadores, para Galicia, as súas provincias e unhas certas áreas definidas nesta enquisa e que se poden consultar no Capítulo 3. O obxectivo que nos marcamos neste capítulo é obter estimacións a un nivel xeográfico máis desagregado do que o deseño da ECV permite facelo empregando estimadores directos. En concreto, as comarcas serán as nosas "áreas pequenas" de interese e o parámetro obxectivo será a taxa de risco de pobreza.

Empregaremos como fonte básica de información os microdatos da ECV das edicións 2008, 2009 e 2010<sup>1</sup>.

Para cumprir este obxectivo, utilizaremos os estimadores indirectos propostos no Capítulo 4. Para estimar o erro cuadrático medio do estimador de Fay-Herriot e tamén dos estimadores baseados nos modelos con efecto tempo empregaranse os estimadores analíticos propostos. No caso do modelo con efecto tempo, porque o estimador bootstrap proposto era claramente peor que o analítico e no caso do estimador de Fay-Herriot porque ambos estimadores (analítico e bootstrap) proporcionaban resultados practicamente iguais e o estimador analítico precisa de menos tempo computacional.

---

<sup>1</sup>O IGE non participou na elaboración deste traballo e, por tanto, non é responsable en ningún caso da utilización feita dos microdatos nin dos resultados amosados no mesmo

## 6.2. Análise descritivo da información de base

A mostra dispoñible é a correspondente ás edicións 2008, 2009 e 2010 da ECV (datos referidos aos anos naturais 2007, 2008 e 2009 respectivamente). Os tamaños mostrais efectivos foron: 9228 vivendas e 24453 persoas (ano 2009), 10245 vivendas e 27395 persoas (ano 2008) e 10246 vivendas e 27527 persoas (ano 2007). No Cadro 6.1 poden consultarse os tamaños mostrais (persoas) nas diferentes comarcas. Neste cadro (e en xeral, en todo o capítulo), *codcoma* denota o código da comarca e  $n_d$  denota o tamaño mostral.

codcoma	COMARCA	$n_d$		
		2.007	2.008	2.009
1501	Arzúa	229	351	236
1502	Barbanza	630	595	571
1503	A Barcala	94	166	64
1504	Bergantiños	528	570	720
1505	Betanzos	367	370	250
1506	A Coruña	2.863	2.836	2.425
1507	Eume	235	109	150
1508	Ferrol	1.511	1.565	1.290
1509	Fisterra	254	310	161
1510	Muros	139	92	148
1511	Noia	335	327	297
1512	Ordes	432	314	463
1513	Ortegal	143	80	84
1514	Santiago	1.456	1.557	1.456
1515	O Sar	189	61	162
1516	Terra de Melide	125	139	91
1517	Terra de Soneira	173	169	52
1518	Xallas	191	178	60
2701	Os Ancares	209	269	81
2702	Chantada	232	184	187
2703	A Fonsagrada	126	172	117
2704	Lugo	1.410	1.428	1.475
2705	A Mariña Central	391	547	323
2706	A Mariña Occidental	326	212	327
2707	A Mariña Oriental	214	144	87
2708	Meira	-	46	-
2709	Quiroga	71	117	80
2710	Sarria	474	348	539
2711	Terra Chá	623	565	335
2712	Terra de Lemos	486	560	451
2713	A Ulloa	112	95	174
3201	Allariz-Maceda	195	323	234
3202	Baixa Limia	139	81	78
3203	O Carballiño	455	451	474
3204	A Limia	210	342	275
3205	Ourense	1.867	1.861	1.825

(segue na páxina seguinte)

Cadro 6.1: Continuación da táboa

		$n_d$		
3206	O Ribeiro	368	393	187
3207	Terra de Caldelas	60	33	-
3208	Terra de Celanova	291	264	210
3209	Terra de Trives	79	87	88
3210	Valdeorras	367	258	347
3211	Verín	415	347	268
3212	Viana	74	95	95
3601	O Baixo Miño	430	366	464
3602	Caldas	308	349	234
3603	O Condado	416	402	235
3604	Deza	454	467	496
3605	O Morrazo	785	787	812
3606	A Paradanta	109	185	114
3607	Pontevedra	1.362	1.351	1.278
3608	O Salnés	917	908	857
3609	Tabeirós-Terra de Montes	369	356	288
3610	Vigo	3.289	3.213	2.738
		<b>27.527</b>	<b>27.395</b>	<b>24.453</b>

Cadro 6.1: Tamaños mostrais por comarca

A comarca de Meira nos anos 2007 e 2009 e a comarca de Terra de Caldelas no ano 2009 carecen de información mostral, polo que quedan fóra deste estudo.

A base de datos, para cada ano, organízase en dous ficheiros, un relativo aos fogares e outro relativo ás persoas. Nos Cadros 6.2 e 6.3 descríbense as variables máis importantes para o estudo que nos ocupa.

VARIABLE	DESCRIPCIÓN DA VARIABLE
CODVIV	Código do fogar
IDADE	Idade (en anos cumpridos)
SEXO	Sexo
NAC	Nacionalidade
FTOTA1	Ingreso total no ano por traballo por conta allea
FTOTA2	Ingreso total no ano por traballo por conta propia
FTOTBC	Ingreso total no ano por prestacións (non se inclúen as prestacións por desemprego)
FTOTD	Ingreso total no ano por prestacións ou subsidios por desemprego
FTOTE	Ingreso total no ano por rendas
FTOTF	Ingreso total no ano por outra fonte de ingresos
FREP	Factor de elevación

Cadro 6.2: Variables relativas ás persoas

O cuestionario da citada enquisa pode consultarse en [www.ige.eu](http://www.ige.eu)

VARIABLE	DESCRIPCIÓN DA VARIABLE
Comarca	Código da comarca do fogar
CODVIV	Código do fogar
UC	Número de unidades de consumo. Escala OCDE modificada

Cadro 6.3: Variables relativas ao fogar

Os datos de ingresos aparecen no ficheiro de persoas. En primeiro lugar, deben obter-se os ingresos totais para cada individuo da mostra como suma das variables FTOTA1, FTOTA2, FTOTBC, FTOTD, FTOTE, FTOTF. A continuación, e seguindo a metodoloxía da enquisa e as recomendacións de Eurostat para o cálculo das medidas de pobreza, defínese o *ingreso do fogar* como suma dos ingresos dos seus membros. A continuación obtense o *ingreso medio mensual equivalente dos fogares* dividindo o ingreso mensual do fogar entre o número de unidades de consumo (variable UC). Finalmente, defínese o *ingreso medio mensual equivalente de cada persoa* ( $z_{dj}$  con  $d = 1, 2, \dots, D$  e  $j = 1, 2, \dots, n_d$ ) como o ingreso medio mensual equivalente do seu fogar. Calcúlase a mediana dos ingresos medios mensuais equivalentes dos individuos e obtense o limiar de pobreza ( $z$ ) como o 60 % da mediana. Unha persoa considérase pobre se os seus ingresos medios mensuais equivalentes son inferiores a este limiar.

Nos citados anos, obtéñense os seguintes limiares de pobreza e a correspondente taxa de risco de pobreza (para Galicia):

ANO	Limiar de pobreza	Taxa de risco de pobreza
2007	603,11	0,1463
2008	609,48	0,1388
2009	605,02	0,1432

Cadro 6.4: Limiar e taxa de risco de pobreza para Galicia

Empregando o estimador directo (4.2), obtemos unha primeira estimación da taxa de risco de pobreza para cada unha das comarcas. Calcúlase ademais o coeficiente de variación (CV), definido como o cociente entre a raíz cuadrada do erro cuadrático medio e o estimador, e que expresaremos en porcentaxe. O erro cuadrático medio foi calculado dacordo coa expresión definida en (4.3). Nas Figuras 6.1, 6.2 e 6.3 represéntanse os coeficientes de variación obtidos, xunto cos correspondentes tamaños mostrais. Vemos, que nas comarcas con menor tamaño mostral, os erros relativos cometidos ao empregar o estimador directo son elevados. Obsérvase tamén, en termos xerais, que os coeficientes de variación do estimador directo no ano 2009 son maiores aos dos anos anteriores, quizais polo menor tamaño mostral dispoñible neste ano. Nalgunhas comarcas como a de Pontevedra ou Ourense, os coeficientes de variación son pequenos dado que estas comarcas coinciden coas áreas definidas no deseño da enquisa. Pola contra, atopamos coeficientes de variación moi elevados na maior parte das comarcas, poñendo de manifesto a imposibilidade de empregar o estimador directo para dar estimacións precisas nestes casos. No Apéndice A proporciónanse os coeficientes de variación obtidos en cada unha das comarcas e anos empregando o estimador directo.

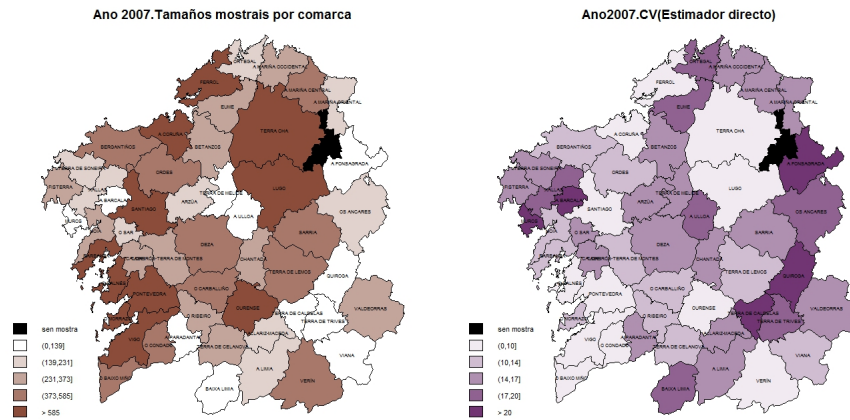


Figura 6.1: Ano 2007. Tamaños mostrais (esquerda) e coeficientes de variación (dereita) do estimador directo.

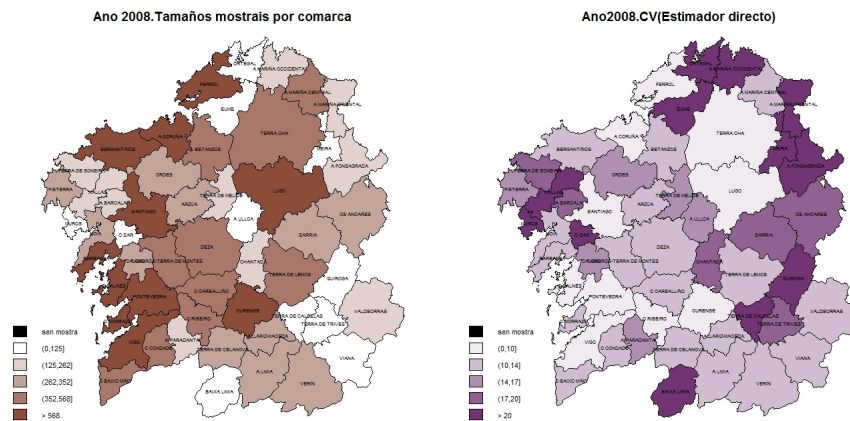


Figura 6.2: Ano 2008. Tamaños mostrais (esquerda) e coeficientes de variación (dereita) do estimador directo.

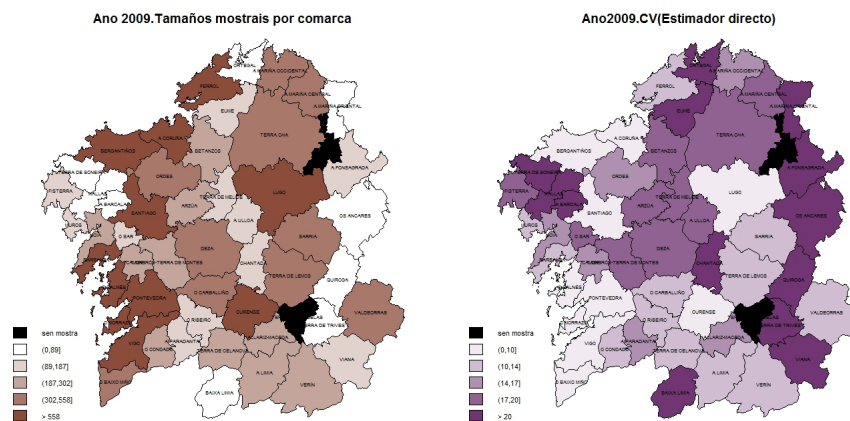


Figura 6.3: Ano 2009. Tamaños mostrais (esquerda) e coeficientes de variación (dereita) do estimador directo.

A continuación calculamos de novo a taxa de risco de pobreza empregando os estimadores indirectos propostos nos capítulos anteriores. Consideramos separadamente os estimadores baseados nun modelo a nivel de área e os baseados nun modelo a nivel de individuo. En cada caso descríbese a información auxiliar empregada.

## 6.3. Estimadores indirectos baseados nun modelo a nivel de área

Incluimos aquí o estimador de Fay-Herriot e os estimadores propostos que se baseaban en modelos que incluían un certo efecto temporal (Modelo 0 e Modelo 1).

### 6.3.1. Variables auxiliares

Os modelos a nivel de área só necesitan os totais das variables auxiliares por dominios, información que soe estar dispoñible nos rexistros. Esta é unha ventaxa á hora da súa utilización fronte aos modelos a nivel de individuo.

Nestes modelos a variable resposta é o estimador directo do parámetro obxectivo. No noso caso, para estimar a taxa de risco de pobreza, necesitábase variables explicativas da mesma a nivel de comarca. Empregouse información procedente do Padrón Municipal de habitantes, como as proporcións comarcais por sexo, nacionalidade e grupos de idade. Escolléronse estas variables dado que a información dispoñible da ECV parece indicar a existencia dunha relación entre a taxa de risco de pobreza e estas variables. Ademais, empregouse información procedente da Consellería de Asuntos Sociais, Emprego e Relacións Laborais, relativa ao paro rexistrado por comarca. Concretamente, definiuse un indicador da proporción de parados en cada comarca, obtido como o cociente entre os parados rexistrados nesa comarca e a poboación con idade entre 16 e 64 anos (datos do Padrón).

No Cadro 6.5 recóllense os coeficientes de correlación entre os estimadores directos e as variables auxiliares consideradas, denotando por:  $P\_MULLERES$  a proporción de mulleres na comarca,  $P\_ESPANOIS$  a proporción de persoas con nacionalidade española,  $P\_MAIS64$  a proporción de persoas con 65 ou máis anos e  $Paro\_Rex$  recolle o indicador da proporción de parados na comarca ao que se fixo referencia.

VARIABLE	coeficiente de correlacion		
	ano 2007	ano 2008	ano 2009
$P\_MULLERES$	-0,314	-0,212	-0,050
$P\_ESPANOIS$	-0,210	-0,114	-0,178
$P\_MAIS64$	0,427	0,475	0,261
$Paro\_Rex$	0,268	0,186	0,334

Cadro 6.5: Coeficientes de correlación coa taxa de risco de pobreza

A selección de variables fíxose empregando o método *Backward*: comezouse considerando como explicativas todas as variables auxiliares dispoñibles e foron eliminándose en etapas sucesivas en función da súa capacidade explicativa. En concreto, vaise eliminando a variable que presenta un menor coeficiente de correlación parcial coa variable dependente (ou equivalentemente un menor valor do estatístico  $t$ ) e así sucesivamente ata chegar a unha situación na cal a eliminación dunha variable máis supoña un descenso acusado no coeficiente de determinación. Neste caso, foi descartada a variable de proporción de mulleres que resultaba non significativa.

### 6.3.2. Obtención do estimador de Fay-Herriot

Para obter o estimador de Fay-Herriot axustáronse tres modelos de Fay-Herriot independentes, un para cada ano, empregando o estimador directo como variable resposta e as seguintes variables auxiliares:

$$\bar{y}_{0,d} = x_d\beta + u_d + e_d, \quad d = 1, \dots, D \quad (6.1)$$

sendo:

- $\bar{y}_{0,d}$ : estimador directo da taxa de risco de pobreza na comarca  $d$
- $X_{1d}$ : proporción de persoas con nacionalidade española na comarca  $d$
- $X_{2d}$ : proporción de persoas con 65 ou máis anos na comarca  $d$
- $X_{3d}$ : proporción de persoas en paro (variable obtida como cociente entre o número de parados rexistrados na comarca e a poboación con idade comprendida entre 16 e 65 anos) na comarca  $d$
- $u_d$ ,  $d = 1, \dots, D$  é o efecto aleatorio na área  $d$ , sendo  $u_1, u_2, \dots, u_D$  independentes e idénticamente distribuídos  $N(0, \sigma_u^2)$ , e  $\sigma_u^2$  descoñecida e constante.
- $e_d$ ,  $d = 1, \dots, D$ , son os erros mostrais que se asumen independentes entre sí e independentes dos efectos aleatorios, que seguen unha distribución Normal de media cero e varianza coñecida  $\sigma_{ed}^2$ .

A varianza dos erros mostrais, que se asume coñecida, estimouse polo erro cuadrático medio do estimador directo.

Os coeficientes estimados, e os p-valores asociados aos contrastes de significatividade individual resúmense no Cadro 6.6

Os coeficientes estimados indican que a taxa de risco de pobreza é menor naquelas comarcas onde é maior a proporción de persoas con nacionalidade española, e menor a proporción de persoas anciáns e de poboación parada.

O erro cuadrático medio dos estimadores de Fay-Herriot obtívose segundo a expresión definida en (4.8). Unicamente se obtivo o erro cuadrático medio analítico dado que como se viu no Capítulo 5 apenas existían diferenzas entre os resultados proporcionados por

	Ano 2007		Ano 2008		Ano 2009	
	$\hat{\beta}$	<i>p-valor</i>	$\hat{\beta}$	<i>p-valor</i>	$\hat{\beta}$	<i>p-valor</i>
$X_1$	-0,05	0,0703	-0,14	0,0293	-0,14	0,0226
$X_2$	0,42	0,0000	0,65	0,0000	0,44	0,0000
$X_3$	1,42	0,0006	1,51	0,0028	1,68	0,0000
$\hat{\sigma}_u$	0,0016		0,0038		0,0019	

Cadro 6.6: Coeficientes estimados e p-valores dos modelos de Fay-Herriot axustados

ambos estimadores. A continuación obtense o coeficiente de variación que se amosa nos Cadros 6.7, 6.8 e 6.9 para as 5 comarcas de menor e maior tamaño mostral en cada ano. Obsérvase na comparativa cos resultados do estimador directo, unha mellora xeral en termos do coeficiente de variación nas comarcas de tamaño mostral máis reducido. Pola contra, naquelas comarcas onde o tamaño mostral é maior e o correspondente coeficiente de variación do estimador directo é máis pequeno, apenas hai diferenzas entre a estimación obtida empregando o método Fay-Herriot e o estimador directo. Isto é lóxico pola propia definición do estimador Fay-Herriot, que ten en conta o deseño mostral e os estimadores directos na súa formulación, e que en última instancia se relaciona co estimador directo da seguinte forma :

$$\hat{Y}_{\alpha,d}^{FH} = \gamma_d \bar{y}_{\alpha,d} + (1 - \gamma_d) x_d \hat{\beta}, \quad d = 1, \dots, D \quad (6.2)$$

$$\gamma_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_{ed}^2}$$

É dicir, o estimador Fay-Herriot corrixe en certa maneira o estimador directo en base ao coeficiente  $\gamma_d$ , sendo a corrección maior canto menor sexa o valor de  $\gamma_d$ , é dicir, cando o erro cuadrático medio do estimador directo ( $\hat{\sigma}_{ed}^2$ ) sexa maior.

Ano 2007				
<i>cod_coma</i>	<i>comarca</i>	$n_d$	<i>cv(directo)</i>	<i>cv(FH)</i>
3207	Terra de Caldelas	60	23,0	15,7
2709	Quiroga	71	27,7	15,8
3212	Viana	74	11,6	11,3
3209	Terra de Trives	79	18,9	13,6
1503	A Barcala	94	27,0	17,4
...	...	...	...	...
1514	Santiago	1.456	8,0	7,8
1508	Ferrol	1.511	6,7	6,5
3205	Ourense	1.867	5,8	5,6
1506	A Coruña	2.863	6,3	6,2
3610	Vigo	3.289	4,3	4,2

Cadro 6.7: Ano 2007. CV dos estimadores directo e Fay-Herriot nas comarcas de maior e menor tamaño de mostra



Ano 2008				
<i>cod_coma</i>	<i>comarca</i>	$n_d$	$cv(directo)$	$cv(FH)$
3207	Terra de Caldelas	33	25,4	19,3
2708	Meira	46	69,4	44,6
1515	O Sar	61	56,6	40,9
1513	Ortegal	80	24,0	20,9
3202	Baixa Limia	81	30,6	19,8
...	...	...	...	...
1514	Santiago	1.557	7,5	7,4
1508	Ferrol	1.565	7,4	7,3
3205	Ourense	1.861	5,5	5,5
1506	A Coruña	2.836	6,9	6,9
3610	Vigo	3.213	4,9	4,9

Cadro 6.8: Ano 2008. CV dos estimadores directo e Fay-Herriot nas comarcas de maior e menor tamaño de mostra

Ano 2009				
<i>cod_coma</i>	<i>comarca</i>	$n_d$	$cv(directo)$	$cv(FH)$
1517	Terra de Soneira	52	22,4	17,3
1518	Xallas	60	25,1	18,2
1503	A Barcala	64	39,1	20,6
3202	Baixa Limia	78	25,7	17,0
2709	Quiroga	80	25,4	15,9
...	...	...	...	...
1514	Santiago	1.456	7,7	7,6
2704	Lugo	1.475	8,5	8,3
3205	Ourense	1.825	6,3	6,2
1506	A Coruña	2.425	6,3	6,2
3610	Vigo	2.738	5,2	5,2

Cadro 6.9: Ano 2009. CV dos estimadores directo e Fay-Herriot nas comarcas de maior e menor tamaño de mostra

Nas Figuras 6.4, 6.5 e 6.6 representáanse os coeficientes de variación dos estimadores directo e Fay-Herriot da taxa de risco de pobreza para cada comarca galega nos anos 2007, 2008 e 2009. Os intervalos considerados son os mesmos para ambos estimadores, para que sexan máis facilmente comparables. No gráfico obsérvase unha gañancia xeralizada en termos do erro relativo ao empregar o estimador indirecto, polo que debemos consideralo unha boa alternativa ao estimador directo. No Apéndice B poden consultarse todos os resultados (estimadores e coeficientes de variación) do estimador de Fay-Herriot.

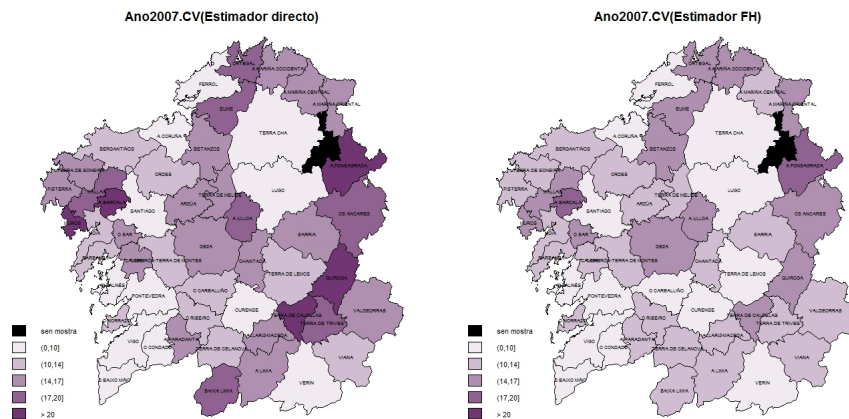


Figura 6.4: Ano 2007. Coeficientes de variación do estimador directo (esquerda) e do estimador de Fay-Herriot (dereita)

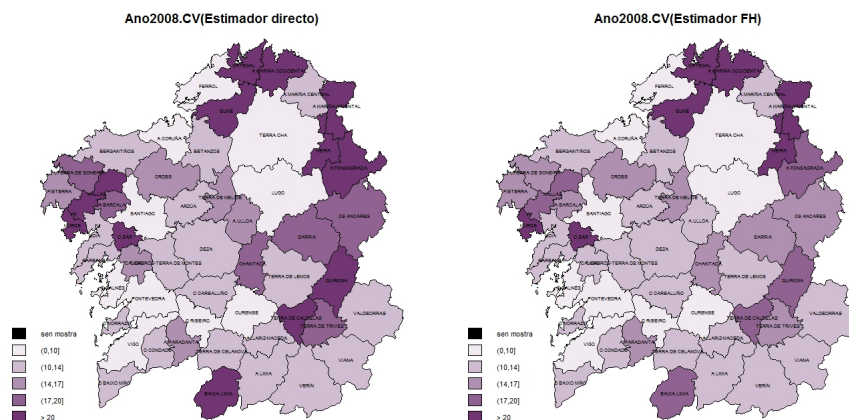


Figura 6.5: Ano 2008. Coeficientes de variación do estimador directo (esquerda) e do estimador de Fay-Herriot (dereita)

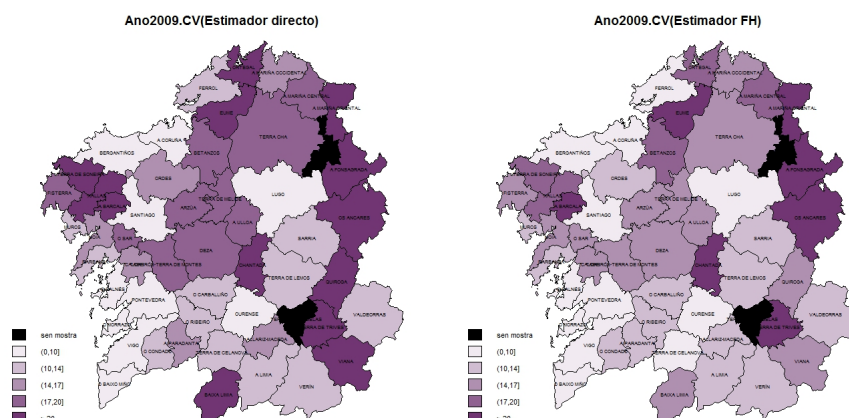


Figura 6.6: Ano 2009. Coeficientes de variación do estimador directo (esquerda) e do estimador de Fay-Herriot(dereita)

### 6.3.3. Obtención dos estimadores baseados nun modelo de área con efecto tempo (Modelo 0 e Modelo 1)

Alternativamente, empregando as mesmas variables auxiliares que no caso do modelo de Fay-Herriot, axustouse un único modelo para tódolos anos (**Modelo 1**) que leva implícito o posible efecto temporal.

$$\bar{y}_{0,dt} = x_{dt}\beta + u_{dt} + e_{dt}, \quad d = 1, \dots, D; \quad t = 1, \dots, T \quad (6.3)$$

onde:

- $\bar{y}_{0,dt}$  é o estimador directo no período  $t$  e na área  $d$  da taxa de risco de pobreza
- $x_{dt}$ , é un vector  $1 \times 3$  contendo os valores das 3 variables auxiliares no dominio  $d$  e no período  $t$ .
- $\beta$  é un vector  $3 \times 1$  contendo os efectos fixos
- $u_{dt}$ , é o efecto aleatorio na área  $d$  e no período  $t$ , sendo  $(u_{d1}, u_{d2}, \dots, u_{dT})$ ,  $d = 1, \dots, D$  independentes e idénticamente distribuídas  $AR(1)$ ;

$$u_{dt} = \rho u_{dt-1} + v_t; \quad \text{con } v_t \text{ iid} \sim N(0, \sigma_u^2)$$

sendo  $\sigma_u^2$  descoñecida e constante.

- $e_{dt}$ , son os erros mostrais que se asumen independentes, con distribución Normal de media cero e varianza  $\sigma_{edt}^2$  que se aproxima polos erros cuadráticos medios do estimador directo.

Consideramos tamén un modelo máis simple, que denotamos por **Modelo 0**, obtido restrinxindo o modelo anterior ao caso  $\rho = 0$ .

Os coeficientes estimados, e os p-valores resúmense no Cadro 6.10

	Modelo 1		Modelo 0	
	$\hat{\beta}$	p-valor	$\hat{\beta}$	p-valor
$X_1$	-0,03	0,0602	-0,05	0,0548
$X_2$	0,44	0,0000	0,46	0,0000
$X_3$	0,88	0,0002	1,02	0,0000
$\hat{\sigma}_u^2$	0,0025		0,0027	
$\hat{\rho}$	0,337			

Cadro 6.10: Coeficientes estimados e p-valor dos modelos axustados

En base ao exercizo de simulación realizado no capítulo previo e á vista do número de períodos dispoñible ( $T = 3$ ) e do baixo valor estimado de  $\rho$  ( $\hat{\rho} = 0,337$ ), todo parece indicar que os estimadores obtidos empregando o modelo máis complexo (Modelo1) e o

modelo máis simple (Modelo0) non van a presentar diferenzas significativas no que respecta á súa acuracidade. Isto pode verse con claridade na Figura 6.7 onde se representan para cada ano os coeficientes de variación dos estimadores obtidos empregando o Modelo 0 e o Modelo 1. Tamén se comparan cos correspondentes coeficientes de variación do estimador obtido empregando modelos de Fay-Herriot independentes para cada ano.

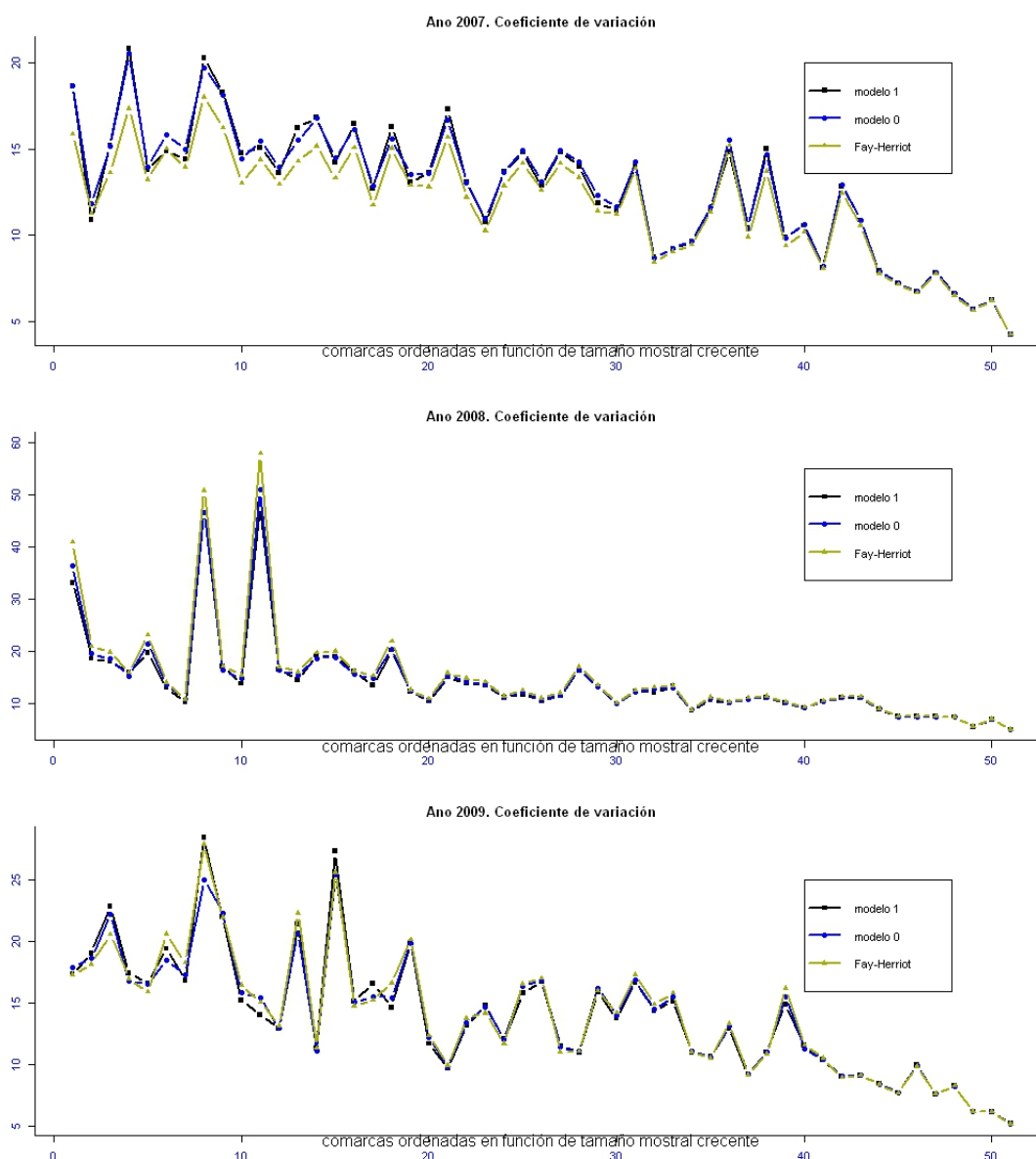


Figura 6.7: Coeficientes de variación dos estimadores Fay-Herriot, Modelo 1 e Modelo 0

Observamos, efectivamente, que apenas hai diferenza entre os erros relativos dos estimadores obtidos empregando os modelos chamados Modelo 1 e Modelo 0. Por outra banda, non se atopa mellora significativa ao comparar estes estimadores cos obtidos empregando modelos de Fay-Herriot independentes ano a ano. Por este motivo e, dado que os cálculos necesarios para obter os estimadores baseados nos modelos con efecto tempo e os seus

correspondentes erros cuadráticos medios son maiores que os necesarios para obter o estimador de Fay-Herriot e, ademais, a súa interpretación non é tan sinxela, quedámonos cos estimadores de Fay-Herriot como mellores estimadores (ata o momento) alternativos ao estimador directo.

#### 6.4. Estimadores indirectos baseados nun modelo a nivel de individuo: estimador EB

O estimador EB ten a ventaxa de que permite obter estimacións en áreas pequenas de calquera parámetro non lineal. Neste sentido, a súa xeralidade é unha das súas principais características positivas. Pola contra, ten o inconveniente de que se basea na existencia dun modelo de superpoboación a nivel de individuo e, polo tanto para a súa aplicación práctica necesítase información de cada unha das unidades de tódalas variables involucradas no modelo. Isto pode ser un problema, xa que son escasas as fontes de información a nivel de individuo dispoñibles. Outra desvantaxa deste método é o elevado tempo computacional necesario para obter os estimadores e os correspondentes estimadores do erro cuadrático medio.

Para aplicar este método, debemos axustar, para cada ano, o modelo (4.11):

$$y_{dj} = x_{dj}\beta + u_d + e_{dj}, \quad j = 1, \dots, N_d, \quad d = 1, \dots, D$$

$$u_d \quad iid \sim N(0, \sigma_u^2)$$

$$e_{dj} \quad iid \sim N(0, \sigma_e^2)$$

onde:

- $x_{dj}$ , é un vector  $1 \times p$  contendo os valores de  $p$  variables auxiliares no dominio  $d$  e no individuo  $j$
- $\beta$  é un vector  $p \times 1$  contendo os efectos fixos
- $u_d$ , é o efecto aleatorio na área  $d$ , sendo  $u_1, u_2, \dots, u_D$  independentes e idénticamente distribuídos  $N(0, \sigma_u^2)$ , sendo  $\sigma_u^2$  descoñecida pero constante.
- $e_{dj}$ , son os erros do modelo, que se supoñen independentes e idénticamente distribuídos  $N(0, \sigma_e^2)$ , e tamén independentes de  $u_d$ .

Este modelo ten como variable resposta ( $y_{dj}$ ) unha variable transformada dos ingresos equivalentes dos individuos ( $z_{dj}$ ) que se asume ten distribución normal. Analizouse a distribución na mostra dos ingresos equivalentes das persoas para cada un dos tres anos. Como é habitual neste tipo de variables económicas, os datos non teñen distribución normal, senón que presentan unha notable asimetría. Para transformar esta variable nunha variable con distribución normal empregouse unha transformación de tipo Box-Cox. Se desexamos transformar unha variable  $z_{dj}$ , a transformación de Box-Cox depende dun parámetro  $\lambda$  e ven dada por:

$$y_{dj}(\lambda) = \begin{cases} \frac{z_{dj}^\lambda - 1}{z_{dj}} & \text{se } \lambda \neq 0 \\ \text{Log}(z_{dj}) & \text{se } \lambda = 0 \end{cases} \quad (6.4)$$

O parámetro  $\lambda$  estimouse empregando o método de máxima verosimilitude empregando a librería *forecast* de R. Os valores estimados foron  $\hat{\lambda} = 0,1$  para o ano 2007 e  $\hat{\lambda} = 0,2$  para os anos 2008 e 2009. As correspondentes variables transformadas dos ingresos equivalentes mensuais dos individuos empregáronse como variables resposta nos modelos.

Nas Figuras 6.8 e 6.9 represéntase o histograma e o gráfico Q-Q Normal dos ingresos equivalentes e da variable transformada para o ano 2009. Os gráficos para os anos 2007 e 2008 non se presentan por ser semellantes a este.

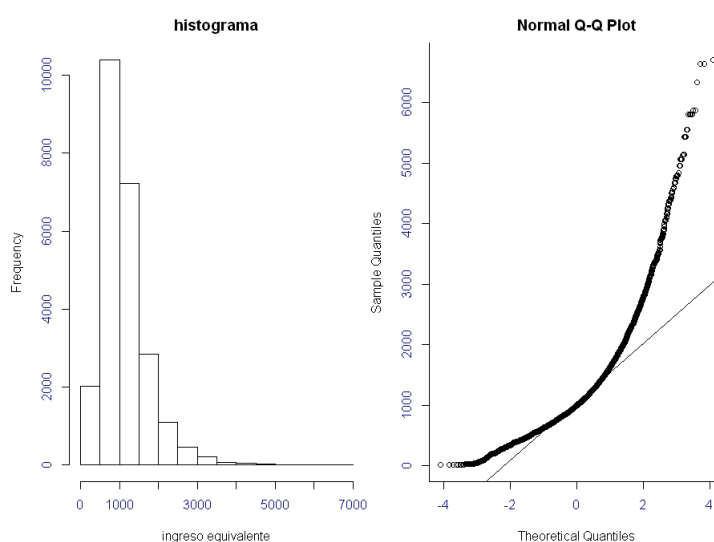


Figura 6.8: Ano 2009. Histograma (esquerda) e gráfico Q-Q (dereita) dos ingresos equivalentes

En canto ás variables explicativas, empregáronse variables da propia Enquisa de Condicións de Vida. Tal e como se comentaba, non é posible obter información individual de cada unha das persoas incluídas na mostra procedentes de outras fontes. As variables consideradas foron as variables indicadoras de sexo, de 5 grupos de idade, nacionalidade e principal fonte de renda do fogar. No caso da principal fonte da renda, definíronse tres categorías: traballo (por conta propia e por conta allea), prestacións (inclúense prestacións por desemprego) e resto. Estas variables indicadoras foron construídas para cada individuo, a partir das variables do Cadro 6.2. No caso da principal fonte de renda, clasifícanse como rendas do traballo as recollidas nas variables FTOTA1 e FTOTA2, como prestacións as recollidas nas variables FTOTBC e FTOTD; e como outras as recollidas polas variables FTOTE e FTOTF.

Para cada variable auxiliar, unha das categorías foi considerada de referencia, omitindo o correspondente indicador e incluíndo un termo independente no modelo. As variables incluídas no modelo son as seguintes:

- $X_1$ : Variable indicadora de sexo=varón

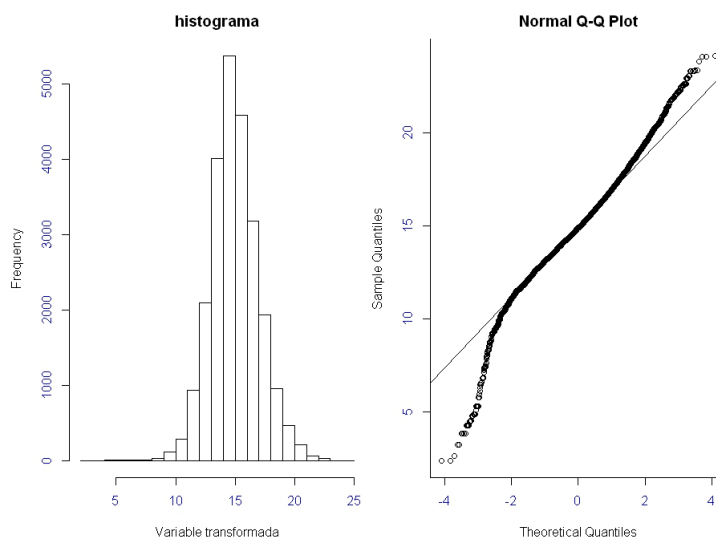


Figura 6.9: Ano 2009. Histograma (esquerda) e gráfico Q-Q (dereita) da variable transformada

- $X_2$ : Variable indicadora de idade de 0 a 14 anos
- $X_3$ : Variable indicadora de idade de 15 a 29 anos
- $X_4$ : Variable indicadora de idade de 30 a 44 anos
- $X_5$ : Variable indicadora de idade de 45 a 64 anos
- $X_6$ : Variable indicadora de nacionalidade española
- $X_7$ : Variable indicadora de principal fonte de renda=rendas do traballo
- $X_8$ : Variable indicadora de principal fonte de renda=prestacións

É dicir, a categoría tomada como referencia é muller, de 65 ou máis anos, con nacionalidade estranxeira e que vive nun fogar onde a principal fonte de renda non procede do traballo nin das prestacións.

Os coeficientes estimados, e os p-valores resúmense no Cadro 6.11

Recórdese que a variable resposta de estos modelos axustados é unha variable transformada dos ingresos mensuais equivalentes dos individuos da mostra (empregándose os valores  $\lambda = 0,1$  para o ano 2007 e  $\lambda = 0,2$  para os anos 2008 e 2009). O coeficiente estimado  $\beta_0$  reflexa o valor medio da variable transformada dun individuo con ( $X_i = 0$ ,  $i = 1, \dots, 8$ ). Se desfacemos a transformación de Box-Cox para que sexa máis facilmente interpretable, estaríamos dicindo (segundo os modelos axustados) que os ingresos equivalentes mensuais medios dunha muller, de nacionalidade estranxeira, de 65 ou máis anos e cuxo fogar ten como principal fonte de renda outras distintas do traballo e das prestacións foi de 547,7, 516,4 e 480,2 euros respectivamente, nos anos 2007, 2008 e 2009. Ademais, os coeficientes estimados indicarían que o feito de ser varón, incrementa con respecto ao individuo tomado como referencia os ingresos medios, así como o feito de ter nacionalidade española, ou idades superiores a 30 anos. Por outra banda, o feito

	Ano 2007		Ano 2008		Ano 2009	
	$\hat{\beta}$	<i>p</i> -valor	$\hat{\beta}$	<i>p</i> -valor	$\hat{\beta}$	<i>p</i> -valor
1	8,787	0,0000	12,441	0,0000	12,189	0,0000
$X_1$	0,057	0,0000	0,081	0,0003	0,049	0,0448
$X_2$	-0,228	0,0000	-0,384	0,0000	-0,476	0,0000
$X_3$	-0,075	0,0000	-0,172	0,0000	-0,328	0,0000
$X_4$	0,043	0,0026	0,053	0,0701	-0,008	0,7789
$X_5$	0,067	0,0000	0,124	0,0000	0,044	0,1052
$X_6$	0,445	0,0000	1,231	0,0000	1,423	0,0000
$X_7$	0,688	0,0000	1,28	0,0000	1,416	0,0000
$X_8$	0,106	0,0000	0,266	0,0000	0,357	0,0000
$\hat{\sigma}_u^2$	0,0386		0,1922		0,1637	
$\hat{\sigma}_e^2$	0,8457		3,4682		3,6349	

Cadro 6.11: Coeficientes estimados e p-valores dos modelos empregados para obter o estimador EB.

de ter idades inferiores a 30 anos, manténdose o resto das condicións constantes, fai que diminúan esos ingresos medios.

Para obter o estimador EB, necesítase coñecer o valor que toman as variables explicativas nos individuos da poboación. Estas variables non son coñecidas para a poboación non mostreada, pero o estimador EB só precisa coñecer o total de persoas na poboación que ten os mesmos valores nas variables auxiliares. Estes totais son aproximados polos factores de elevación das unidades mostreadas.

Con estos antecedentes, e empregando  $L = 200$  no procedemento de Monte Carlo, obtéñense os estimadores EB da taxa de risco de pobreza e os estimadores do erro cuadrático medio tal e como se describiu no Capítulo 4.

Se comparamos o estimador directo e o estimador de Fay-Herriot co estimador EB en termos do coeficiente de variación, observamos que o erro cometido con este último estimador é sensiblemente menor.

Nos Cadros 6.12, 6.13 e 6.14 represéntase o coeficiente de variación do estimador EB para as comarcas de menor e maior tamaño mostral, xunto cos coeficientes de variación obtidos co estimador directo e co estimador de Fay-Herriot. Obsérvase que, en termos do erro relativo, o estimador EB obtén moi bos resultados. Obténdose valores do coeficiente de variación sensiblemente menores aos correspondentes do estimador directo e tamén aos do estimador de Fay-Herriot. Tamén neste caso, a mellora na acuracidade do estimador é notable nas comarcas que teñen un tamaño mostral máis reducido.

Os coeficientes de variación dos tres estimadores e para todas as comarcas represéntanse nas Figuras 6.10, 6.11 e 6.12. Nelas vemos que o erro relativo do estimador EB é inferior ao do estimador FH e tamén ao estimador directo na totalidade das comarcas e para os tres anos considerados.



Ano 2007					
<i>cod_coma</i>	<i>comarca</i>	$n_d$	$cv(directo)$	$cv(FH)$	$cv(EB)$
3207	Terra de Caldelas	60	23,0	15,7	13,2
2709	Quiroga	71	27,7	15,8	16,9
3212	Viana	74	11,6	11,3	8,9
3209	Terra de Trives	79	18,9	13,6	9,1
1503	A Barcala	94	27,0	17,4	14,7
...	...	...	...	...	...
1514	Santiago	1.456	8,0	7,8	8,4
1508	Ferrol	1.511	6,7	6,5	4,7
3205	Ourense	1.867	5,8	5,6	4,7
1506	A Coruña	2.863	6,3	6,2	5,9
3610	Vigo	3.289	4,3	4,2	3,3

Cadro 6.12: Ano 2007. CV dos estimadores directo, Fay-Herriot e EB nas comarcas de maior e menor tamaño de mostra

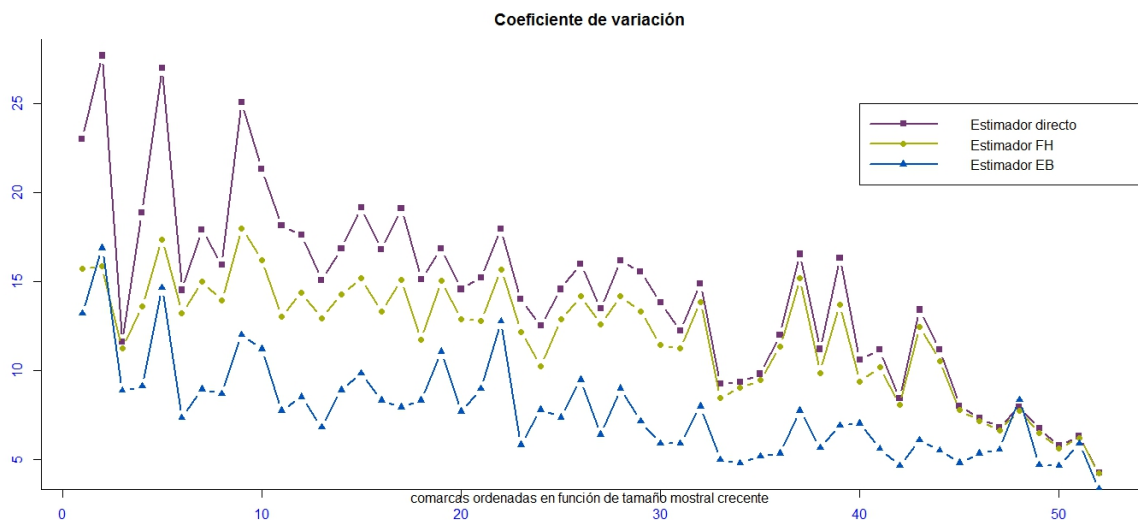


Figura 6.10: Ano 2007. CV dos estimadores directo, Fay-Herriot e EB

Ano 2008					
<i>cod_coma</i>	<i>comarca</i>	$n_d$	$cv(directo)$	$cv(FH)$	$cv(EB)$
3207	Terra de Caldelas	33	25,4	19,3	14,1
2708	Meira	46	69,4	44,6	21,5
1515	O Sar	61	56,6	40,9	17,7
1513	Ortegal	80	24,0	20,9	11,2
3202	Baixa Limia	81	30,6	19,8	13,5
...	...	...	...	...	...
1514	Santiago	1.557	7,5	7,4	6,8
1508	Ferrol	1.565	7,4	7,3	4,8
3205	Ourense	1.861	5,5	5,5	4,0
1506	A Coruña	2.836	6,9	6,9	6,6
3610	Vigo	3.213	4,9	4,9	3,8

Cadro 6.13: Ano 2008. CV dos estimadores directo, Fay-Herriot e EB nas comarcas de maior e menor tamaño de mostra

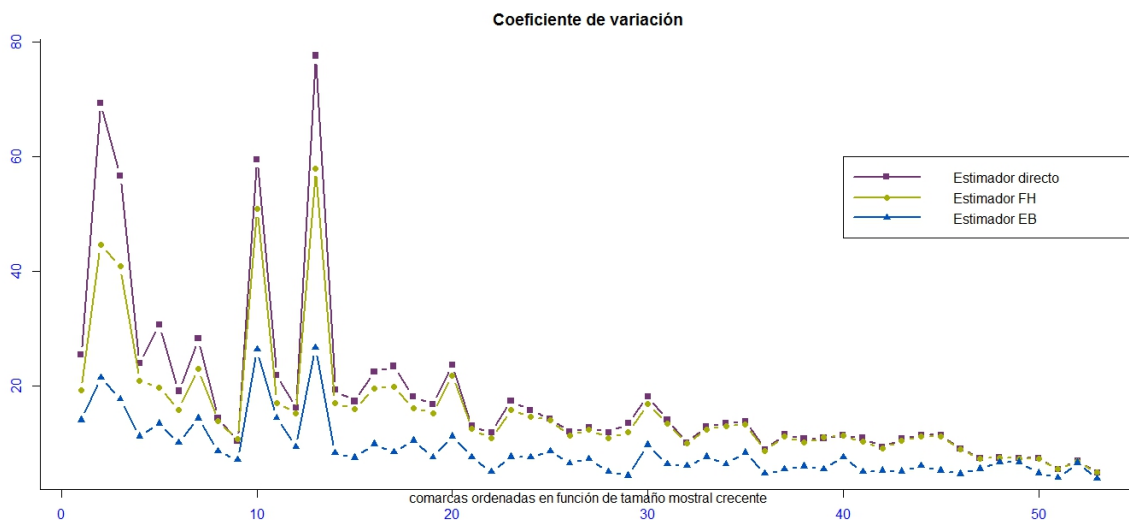


Figura 6.11: Ano 2008. CV dos estimadores directo, Fay-Herriot e EB

Ano 2009					
<i>cod_coma</i>	<i>comarca</i>	$n_d$	$cv(directo)$	$cv(FH)$	$cv(EB)$
1517	Terra de Soneira	52	22,4	17,3	9,4
1518	Xallas	60	25,1	18,2	12,8
1503	A Barcala	64	39,1	20,6	12,1
3202	Baixa Limia	78	25,7	17,0	11,1
2709	Quiroga	80	25,4	15,9	13,4
...	...	...	...	...	...
1514	Santiago	1.456	7,7	7,6	6,6
2704	Lugo	1.475	8,5	8,3	6,3
3205	Ourense	1.825	6,3	6,2	4,5
1506	A Coruña	2.425	6,3	6,2	5,9
3610	Vigo	2.738	5,2	5,2	3,9

Cadro 6.14: Ano 2009. CV dos estimadores directo, Fay-Herriot e EB nas comarcas de maior e menor tamaño de mostra

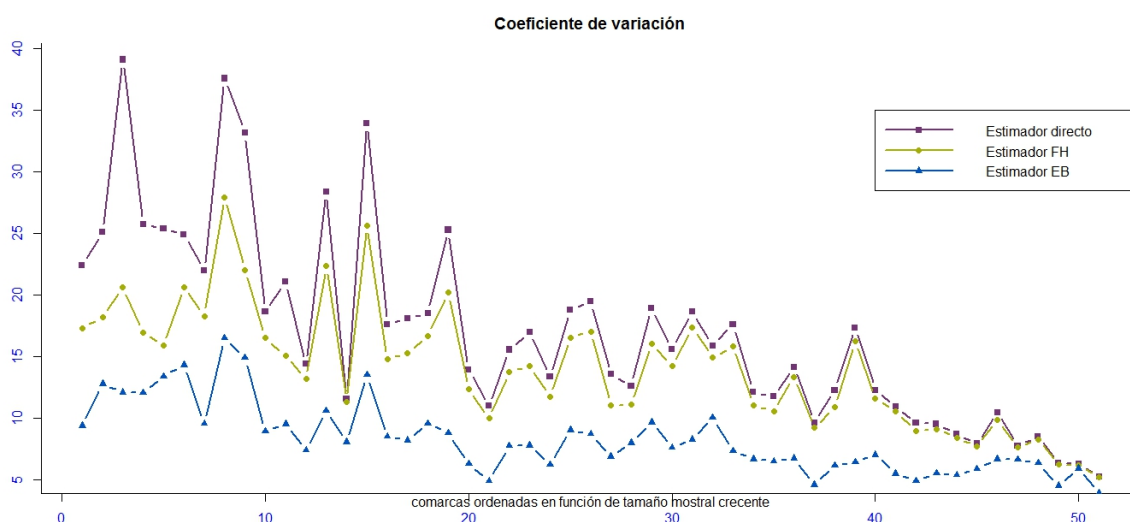


Figura 6.12: Ano 2009. CV dos estimadores directo, Fay-Herriot e EB

Ao igual que se fixo co estimador Fay-Herriot, nas Figuras 6.13, 6.14 e 6.15 representáanse os coeficientes de variación dos estimadores directo e EB da taxa de risco de pobreza para cada comarca galega nos anos 2007, 2008 e 2009. Os intervalos considerados son os mesmos para ambos estimadores, para que sexan máis facilmente comparables. Obsérvase unha clara redución dos coeficientes de variación ao empregar este estimador indirecto. De feito, obtéñense coeficientes de variación inferiores ao 20% en tódolos casos, agás nas comarcas de Meira, a Mariña Oriental e do Eume no ano 2008. Nestas tres comarcas, obsérvase un notable descenso do coeficiente de variación con respecto ao obtido empregando calquera dos outros dous estimadores. No apéndice E poden consultarse a totalidade dos valores estimados e os seus coeficientes de variación empregando este método.

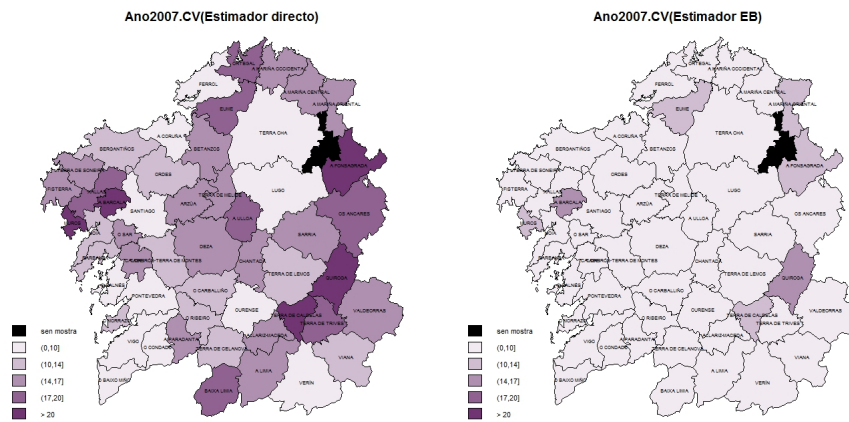


Figura 6.13: Ano 2007. Coeficientes de variación do estimador directo (esquerda) e do estimador EB (dereita)

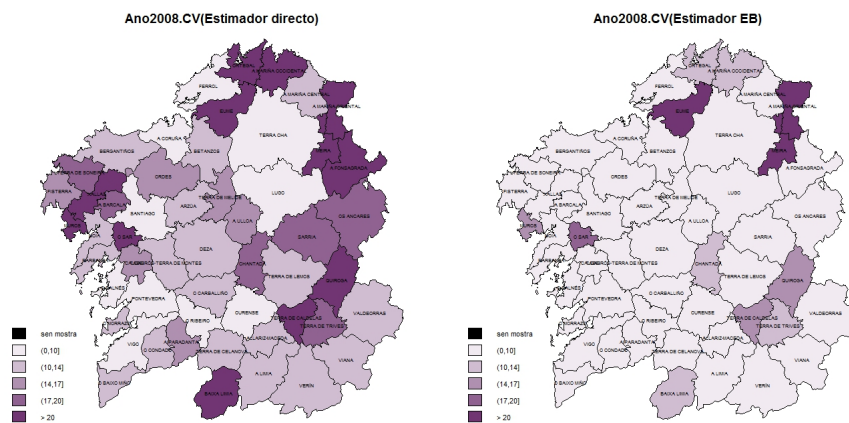


Figura 6.14: Ano 2008. Coeficientes de variación do estimador directo (esquerda) e do estimador EB (dereita)

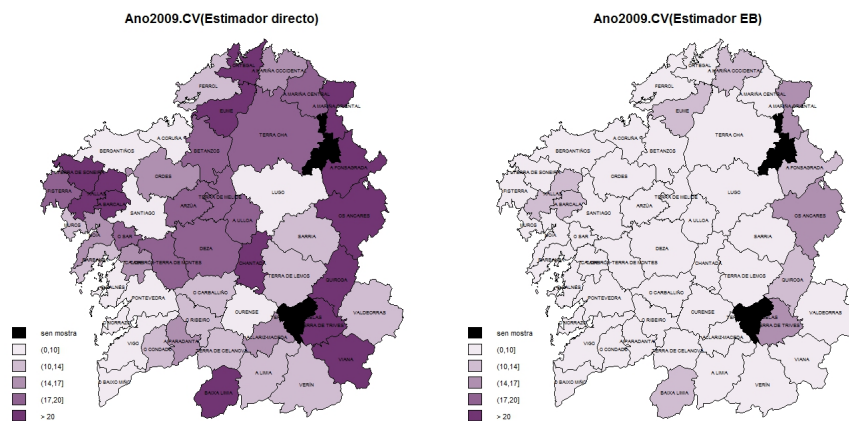


Figura 6.15: Ano 2009. Coeficientes de variación do estimador directo (esquerda) e do estimador EB (dereita)

Á vista do anterior, concluímos que o estimador EB é o mellor candidato para obter a taxa de risco de pobreza a nivel comarcal. Os coeficientes de variación obtidos son case todos inferiores ao 20 %, cota establecida pola Office for National Statistics (ONS) do Reino Unido para considerar un dato publicable.

Nos mapas seguintes represéntase graficamente os valores estimados para cada comarca e para os tres anos en estudo. Os intervalos definidos correspóndense cos quintiles da distribución da taxa de risco de pobreza para cada ano.

Analizando os resultados, obsérvase que os niveis máis baixos na taxa de risco de pobreza os atopamos nas comarcas das sete grandes cidades galegas (A Coruña, Lugo, Ourense, Pontevedra, Vigo, Santiago e Ferrol) e nas comarcas do norte de Lugo. Desta maneira dedúcese que nas zonas rurais o indicador  $FGT_0$  presenta valores máis elevados. Porén, debe terse en conta que este indicador se basea puramente nos ingresos monetarios dos individuos. Non tendo en conta factores alleos aos ingresos, como o aforro en alimentos procedente do autoconsumo que moitas familias do rural teñen ou o diferente nivel de prezos que poden facer, que cun nivel menor de ingresos estas persoas poidan ter un mesmo nivel de benestar que familias que residen nunha gran cidade.

Continuando coa descrición dos resultados, a maior parte das comarcas da provincia de Ourense presentan unha proporción de persoas por debaixo do limiar de pobreza superior ao 20 %. Tamén se observa un diferente comportamento no que se pode chamar a "Galicia interior" e a "Galicia costeira", atopándonos unha maior taxa de risco de pobreza no interior en termos xerais.

Os resultados obtidos son coherentes con outros indicadores como a Renda dispoñible bruta por habitante (RDBpc). A renda dispoñible bruta é un agregado macroeconómico definido no Sistema Europeo de Contas (SEC 95). Obtense como o saldo contable da conta de distribución secundaria da renda e reflexa os recursos dos que dispoñen os fogares para destinar ao consumo final ou ao aforro. A Figura 6.19 é o mapa comarcal desta variable. Obsérvase que existe unha elevada correlación negativa entre a proporción de persoas por debaixo do limiar da pobreza nunha comarca e a RDBpc de dita comarca.

### Ano 2007. Estimador EB da taxa de risco de pobreza

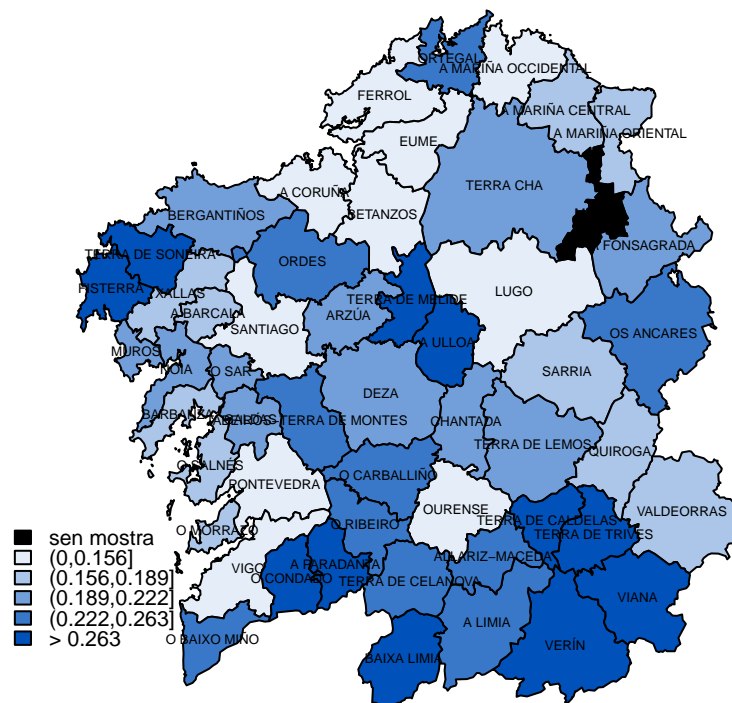


Figura 6.16: Ano 2007. Taxa de risco da pobreza obtida do estimador EB

### Ano 2008. Estimador EB da taxa de risco de pobreza

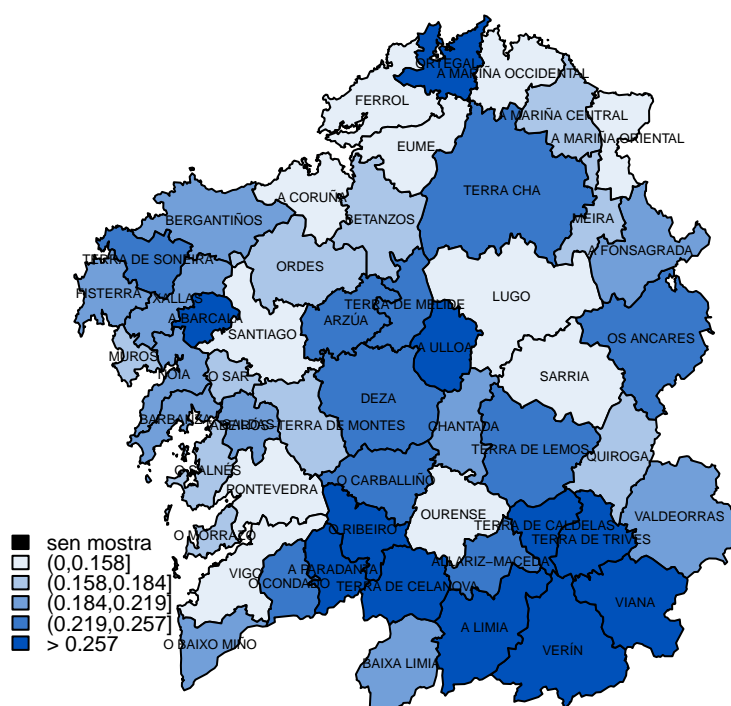


Figura 6.17: Ano 2008. Taxa de risco da pobreza obtida do estimador EB

### Ano 2009. Estimador EB da taxa de risco de pobreza

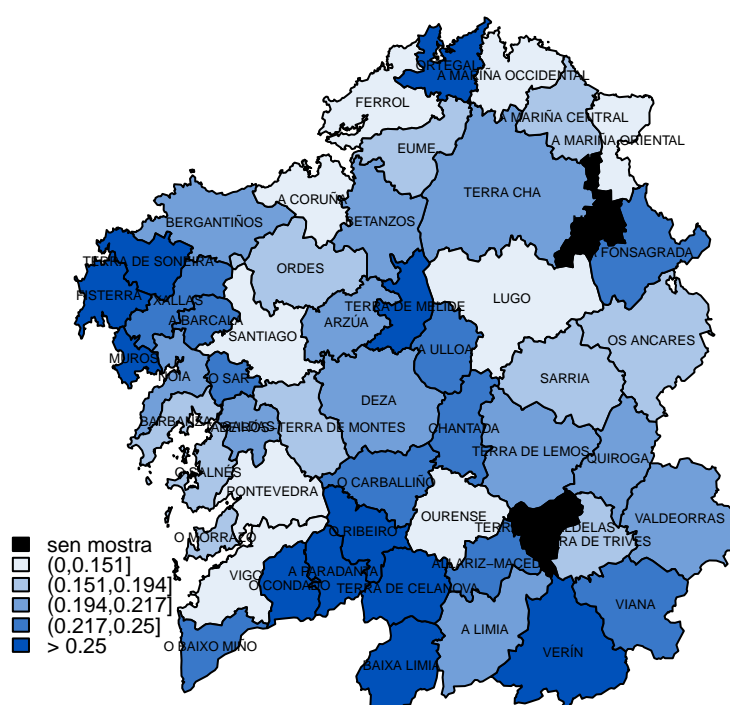
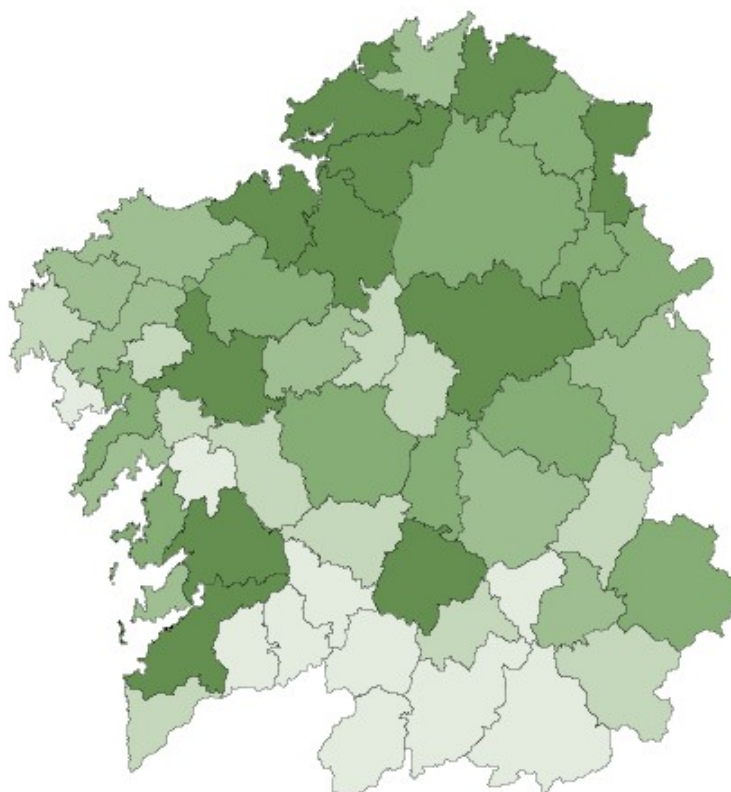


Figura 6.18: Ano 2009. Taxa de risco da pobreza obtida do estimador EB



### Renda bruta dispoñible por habitante



Fonte: IGE. Contas de distribución da renda dos fogares por concellos

Figura 6.19: Ano 2008. Renda bruta dispoñible por habitante (Euros/habitante)



# Capítulo 7

## Conclusiones

Os estimadores indirectos baseados en modelos constitúen unha boa alternativa aos tradicionais estimadores baseados no deseño cando o obxectivo é obter estimacións en pequenos dominios.

Estos estimadores indirectos proporcionan moi bos resultados cando o modelo está ben especificado. Quizais a maior dificultade para a súa posta en práctica sexa a dispoñibilidade de información auxiliar de calidade para a definición e axuste dos diferentes modelos.

Neste sentido, na práctica é máis fácil dispoñer de información auxiliar a nivel agregado, polo que a utilización de modelos de área parece máis factible. En particular, o estimador de Fay-Herriot, cuxo procedemento de cálculo é bastante sinxelo e ademais permite introducir información do deseño mostral da enquisa que emprega como fonte principal sitúase como un bo candidato en moitas situacións. Outra ventaxa deste estimador é que é facilmente interpretable, dado que pode considerarse unha combinación lineal do estimador directo, que corrixe este a medida que o erro de mostraxe deste estimador é maior.

Os estimadores baseados en modelos de área con efecto tempo constitúen tamén unha boa alternativa sempre e cando se sexa capaz de especificar correctamente a relación temporal. Un inconveniente para a súa aplicación pode ser que, na maior parte das situacións reais, o número de períodos dispoñibles acostuma a ser moi baixo, polo que a detección da relación temporal é difícil.

Por outra banda, neste traballo presentouse o estimador EB, estimador indirecto que utiliza auxiliarmente un modelo a nivel de individuo. Como xa se comentou, un problema de estos estimadores é que existen moi poucas fontes de información a nivel de individuo para a especificación e axuste deste tipo de modelos. Porén, vimos que cando se dispón da información axeitada e se define correctamente o modelo, os resultados obtidos son realmente bos. Hai que dicir, deste método en particular, que outro inconveniente que detectamos foi o elevado tempo necesario para realizar o procedemento de cálculo do estimador e tamén do seu erro cuadrático medio. Neste sentido, sería interesante a búsqueda de alternativas que permitisen optimizar este método neste sentido.

No que respecta aos estimadores do erro cuadrático medio, non hai apenas diferenzas entre os estimadores analítico e bootstrap do erro cuadrático medio do estimador de Fay-Herriot. Por outra banda, o estimador bootstrap que se propuxo para o erro cuadrático medio do estimador baseado no modelo de área con efecto temporal (Modelo1), non proporcionaba resultados competitivos co estimador analítico. Sería tamén interesante estudar outro estimador alternativo de tipo bootstrap diferente ao proposto.

---

# **Anexos**



## A. Tamaños mostrais e coeficientes de variación do estimador directo

Cadro 1: Resultados co estimador directo

codcoma	Ano 2007		Ano 2008		Ano 2009	
	$n_d$	CV(directo)	$n_d$	CV(directo)	$n_d$	CV(directo)
1501	229	14,58	351	10,12	236	18,75
1502	630	13,42	595	11,44	571	10,90
1503	94	27,01	166	19,32	64	39,08
1504	528	11,18	570	10,78	720	9,60
1505	367	16,18	370	13,85	250	19,46
1506	2.863	6,31	2.836	6,93	2.425	6,25
1507	235	17,96	109	59,46	150	33,93
1508	1.511	6,75	1.565	7,43	1.290	10,45
1509	254	14,01	310	15,76	161	17,59
1510	139	21,32	92	28,31	148	11,55
1511	335	13,49	327	12,76	297	15,59
1512	432	12,02	314	14,30	463	14,13
1513	143	17,62	80	24,00	84	21,95
1514	1.456	7,96	1.557	7,46	1.456	7,74
1515	189	16,84	61	56,64	162	18,05
1516	125	15,94	139	16,18	91	18,63
1517	173	15,09	169	17,35	52	22,38
1518	191	19,17	178	23,48	60	25,08
2701	209	19,10	269	17,45	81	24,89
2702	232	15,23	184	18,16	187	25,26
2703	126	25,08	172	22,52	117	28,33
2704	1.410	6,82	1.428	7,49	1.475	8,47
2705	391	14,90	547	11,45	323	18,65
2706	326	16,00	212	23,73	327	15,88
2707	214	16,86	144	77,65	87	37,55
2708	-	-	46	69,36	-	-
2709	71	27,69	117	21,94	80	25,35
2710	474	16,33	348	18,16	539	12,27
2711	623	8,43	565	9,37	335	17,60
2712	486	10,62	560	10,94	451	11,73
2713	112	17,90	95	14,43	174	18,49
3201	195	16,81	323	12,03	234	15,55
3202	139	18,14	81	30,65	78	25,74

(segue na páxina seguinte)

Cadro 1: Continuación da táboa

codcoma	Ano 2007		Ano 2008		Ano 2009	
	$n_d$	CV(directo)	$n_d$	CV(directo)	$n_d$	CV(directo)
3203	455	11,21	451	10,77	474	12,27
3204	210	15,13	342	11,91	275	12,57
3205	1.867	5,78	1.861	5,52	1.825	6,34
3206	368	13,82	393	8,96	187	13,89
3207	60	23,00	33	25,43	-	-
3208	291	12,51	264	11,88	210	10,97
3209	79	18,89	87	19,12	88	33,15
3210	367	15,56	258	13,04	347	12,09
3211	415	9,27	347	13,51	268	13,54
3212	74	11,63	95	10,43	95	21,05
3601	430	9,82	366	13,53	464	9,61
3602	308	14,58	349	14,12	234	16,96
3603	416	9,35	402	11,54	235	13,33
3604	454	16,54	467	10,92	496	17,32
3605	785	11,18	787	11,47	812	9,50
3606	109	14,53	185	16,86	114	14,37
3607	1.362	7,34	1.351	7,47	1.278	7,92
3608	917	8,03	908	9,10	857	8,70
3609	369	12,23	356	12,87	288	18,92
3610	3.289	4,28	3.213	4,93	2.738	5,25

(Fin da táboa)



## B. Resultados para o estimador de Fay-Herriot

Cadro 2: Resultados co estimador de Fay-Herriot

codcoma	Ano 2007		Ano 2008		Ano 2009	
	FH	CV(FH)	FH	CV(FH)	FH	CV(FH)
1501	0,170	12,89	0,232	10,01	0,104	16,54
1502	0,114	12,45	0,127	11,27	0,148	10,52
1503	0,169	17,35	0,148	16,99	0,163	20,58
1504	0,134	10,20	0,191	10,43	0,161	8,95
1505	0,113	14,18	0,144	13,33	0,109	17,02
1506	0,081	6,21	0,077	6,89	0,105	6,17
1507	0,126	15,67	0,021	50,93	0,092	25,64
1508	0,131	6,51	0,117	7,32	0,107	9,84
1509	0,190	12,19	0,147	14,70	0,201	14,77
1510	0,148	16,22	0,125	23,08	0,273	11,32
1511	0,200	12,62	0,169	12,45	0,182	14,20
1512	0,158	11,36	0,159	14,02	0,122	13,33
1513	0,181	14,38	0,211	20,89	0,184	18,28
1514	0,094	7,76	0,115	7,42	0,113	7,62
1515	0,176	14,29	0,049	40,92	0,207	15,24
1516	0,221	13,95	0,215	15,31	0,196	16,48
1517	0,198	12,95	0,209	16,01	0,224	17,31
1518	0,119	15,18	0,146	19,95	0,189	18,18
2701	0,165	15,09	0,120	15,81	0,191	20,64
2702	0,157	12,81	0,166	16,18	0,097	20,18
2703	0,128	18,00	0,101	19,65	0,118	22,34
2704	0,134	6,66	0,110	7,43	0,107	8,28
2705	0,153	13,87	0,138	11,32	0,093	17,32
2706	0,122	14,17	0,135	21,86	0,140	14,89
2707	0,154	15,05	0,025	58,01	0,131	27,91
2708	-	-	0,051	44,64	-	-
2709	0,192	15,85	0,241	17,09	0,202	15,90
2710	0,130	13,72	0,076	16,93	0,174	11,58
2711	0,187	8,10	0,216	9,21	0,102	15,81
2712	0,215	9,39	0,177	10,26	0,185	10,52
2713	0,218	15,01	0,285	13,91	0,174	16,65
3201	0,172	13,33	0,213	11,39	0,190	13,75
3202	0,241	13,02	0,200	19,79	0,223	16,96
3203	0,182	9,89	0,173	10,21	0,193	10,89

(segue na páxina seguinte)

Cadro 2: Continuación da táboa

codcoma	Ano 2007		Ano 2008		Ano 2009	
	FH	CV(FH)	FH	CV(FH)	FH	CV(FH)
3204	0,230	11,74	0,244	10,95	0,259	11,07
3205	0,148	5,65	0,166	5,49	0,132	6,21
3206	0,189	11,43	0,293	8,71	0,234	12,35
3207	0,238	15,73	0,267	19,25	-	-
3208	0,224	10,27	0,313	10,88	0,289	9,96
3209	0,250	13,63	0,237	15,90	0,096	21,99
3210	0,141	13,35	0,194	12,52	0,178	11,03
3211	0,276	8,47	0,232	11,98	0,240	11,02
3212	0,324	11,26	0,423	10,76	0,210	15,07
3601	0,212	9,46	0,162	13,05	0,235	9,19
3602	0,134	12,87	0,123	13,48	0,154	14,20
3603	0,227	9,05	0,157	11,17	0,206	11,72
3604	0,142	15,19	0,216	11,04	0,128	16,21
3605	0,145	10,56	0,118	11,26	0,155	9,11
3606	0,237	13,25	0,182	15,26	0,257	13,16
3607	0,132	7,17	0,119	7,42	0,114	7,70
3608	0,161	7,80	0,120	8,98	0,164	8,39
3609	0,180	11,24	0,201	12,46	0,110	16,00
3610	0,152	4,25	0,121	4,91	0,136	5,18

(Fin da táboa)

## C. Resultados para o estimador do modelo 1

Cadro 3: Resultados co estimador baseado no modelo 1

codcoma	Ano 2007		Ano 2008		Ano 2009	
	Estimador	CV	Estimador	CV	Estimador	CV
1501	0,175	13,06	0,224	9,92	0,110	15,80
1502	0,112	12,83	0,128	10,99	0,150	10,41
1503	0,154	20,84	0,146	16,48	0,155	22,87
1504	0,131	10,57	0,186	10,31	0,161	9,03
1505	0,111	14,83	0,141	13,16	0,112	16,74
1506	0,080	6,26	0,078	6,86	0,105	6,18
1507	0,117	17,31	0,022	46,63	0,088	27,36
1508	0,129	6,64	0,116	7,30	0,106	9,98
1509	0,185	13,06	0,151	13,77	0,203	15,16
1510	0,139	18,28	0,138	19,61	0,286	11,33
1511	0,208	12,86	0,173	11,69	0,193	13,76
1512	0,161	11,50	0,159	13,38	0,127	12,96
1513	0,184	15,08	0,208	18,53	0,212	16,87
1514	0,093	7,85	0,115	7,37	0,113	7,59
1515	0,165	16,23	0,059	33,15	0,199	16,57
1516	0,239	14,39	0,221	13,75	0,223	15,23
1517	0,201	13,62	0,213	14,49	0,245	17,36
1518	0,110	16,83	0,144	18,95	0,192	19,07
2701	0,157	16,44	0,123	14,95	0,214	19,44
2702	0,152	13,63	0,157	16,06	0,099	19,86
2703	0,117	20,29	0,099	19,17	0,123	21,41
2704	0,134	6,73	0,110	7,37	0,107	8,25
2705	0,157	14,03	0,138	11,04	0,097	16,65
2706	0,120	14,79	0,136	20,15	0,146	14,39
2707	0,148	16,30	0,029	48,11	0,135	28,44
2708	-	-	-	-	-	-
2709	0,178	18,64	0,213	17,05	0,199	16,64
2710	0,122	14,99	0,078	16,31	0,176	11,57
2711	0,189	8,11	0,211	9,09	0,107	15,15
2712	0,210	9,89	0,175	10,02	0,185	10,62
2713	0,244	14,86	0,274	12,91	0,203	14,65
3201	0,169	14,21	0,208	11,10	0,202	13,22
3202	0,230	14,79	0,195	18,07	0,228	17,45
3203	0,177	10,42	0,171	10,00	0,193	11,03

(segue na páxina seguinte)

Cadro 3: Continuación da táboa

codcoma	Ano 2007		Ano 2008		Ano 2009	
	Estimador	CV	Estimador	CV	Estimador	CV
3204	0,222	12,69	0,240	10,48	0,267	11,01
3205	0,148	5,70	0,166	5,46	0,132	6,20
3206	0,189	11,87	0,283	8,53	0,255	11,71
3207	-	-	-	-	-	-
3208	0,220	10,76	0,296	10,38	0,302	9,67
3209	0,250	15,19	0,213	15,85	0,097	21,99
3210	0,139	13,99	0,189	12,21	0,181	11,01
3211	0,269	8,70	0,226	11,31	0,228	11,49
3212	0,381	10,92	0,389	10,11	0,234	14,02
3601	0,214	9,62	0,168	12,03	0,237	9,21
3602	0,129	13,69	0,123	13,14	0,149	14,81
3603	0,230	9,22	0,160	10,60	0,203	12,07
3604	0,150	14,85	0,211	10,73	0,140	14,91
3605	0,143	10,87	0,120	10,94	0,155	9,15
3606	0,254	13,78	0,194	13,49	0,277	12,92
3607	0,132	7,24	0,119	7,35	0,114	7,72
3608	0,160	7,92	0,121	8,79	0,164	8,42
3609	0,183	11,43	0,195	12,21	0,113	15,86
3610	0,152	4,26	0,121	4,88	0,136	5,19

(Fin da táboa)

## D. Resultados para o estimador do modelo 0

Cadro 4: Resultados co estimador baseado no modelo 0

codcoma	Ano 2007		Ano 2008		Ano 2009	
	Estimador	CV	Estimador	CV	Estimador	CV
1501	0,170	13,53	0,229	9,85	0,106	16,34
1502	0,111	12,91	0,128	11,06	0,150	10,44
1503	0,158	20,52	0,149	16,26	0,161	22,27
1504	0,131	10,65	0,190	10,24	0,160	9,07
1505	0,110	14,93	0,145	12,96	0,111	16,85
1506	0,080	6,26	0,078	6,85	0,105	6,17
1507	0,123	16,65	0,022	46,71	0,095	25,43
1508	0,129	6,63	0,117	7,27	0,106	9,94
1509	0,186	13,05	0,147	14,21	0,206	15,07
1510	0,141	18,15	0,128	21,43	0,294	11,11
1511	0,206	13,11	0,169	12,14	0,191	14,01
1512	0,159	11,64	0,159	13,58	0,125	13,13
1513	0,180	15,45	0,203	19,51	0,206	17,34
1514	0,093	7,86	0,115	7,36	0,113	7,60
1515	0,174	15,49	0,054	36,39	0,215	15,50
1516	0,231	14,96	0,209	14,74	0,216	15,85
1517	0,197	13,96	0,204	15,37	0,239	17,93
1518	0,111	16,78	0,148	18,80	0,198	18,64
2701	0,162	16,14	0,122	15,21	0,229	18,53
2702	0,153	13,61	0,165	15,50	0,100	19,88
2703	0,121	19,68	0,103	18,63	0,129	20,61
2704	0,134	6,73	0,111	7,38	0,107	8,25
2705	0,155	14,24	0,138	11,10	0,096	16,87
2706	0,119	14,91	0,137	20,45	0,146	14,47
2707	0,156	15,59	0,028	50,99	0,156	25,04
2708	-	-	-	-	-	-
2709	0,178	18,68	0,230	16,26	0,201	16,53
2710	0,126	14,67	0,078	16,41	0,181	11,32
2711	0,188	8,20	0,214	9,07	0,105	15,52
2712	0,212	9,84	0,176	10,08	0,187	10,57
2713	0,230	15,84	0,268	13,51	0,193	15,48
3201	0,167	14,49	0,210	11,15	0,201	13,43
3202	0,237	14,41	0,194	18,48	0,238	16,81
3203	0,178	10,41	0,172	10,03	0,194	10,99

(segue na páxina seguinte)

Cadro 4: Continuación da táboa

codcoma	Ano 2007		Ano 2008		Ano 2009	
	Estimador	CV	Estimador	CV	Estimador	CV
3204	0,221	12,86	0,238	10,72	0,267	11,10
3205	0,148	5,71	0,166	5,46	0,132	6,22
3206	0,184	12,30	0,284	8,62	0,247	12,19
3207	-	-	-	-	-	-
3208	0,216	10,99	0,296	10,65	0,299	9,85
3209	0,253	15,17	0,228	15,21	0,096	22,32
3210	0,137	14,25	0,191	12,24	0,180	11,12
3211	0,271	8,67	0,226	11,58	0,231	11,43
3212	0,352	11,84	0,376	10,72	0,212	15,48
3601	0,214	9,69	0,161	12,71	0,238	9,24
3602	0,130	13,66	0,124	13,14	0,152	14,68
3603	0,230	9,25	0,157	10,95	0,204	12,08
3604	0,144	15,55	0,213	10,81	0,135	15,54
3605	0,143	10,89	0,119	11,06	0,155	9,15
3606	0,254	13,96	0,180	14,71	0,278	12,98
3607	0,132	7,25	0,119	7,36	0,114	7,73
3608	0,160	7,92	0,120	8,88	0,164	8,42
3609	0,181	11,63	0,199	12,15	0,111	16,20
3610	0,152	4,27	0,121	4,89	0,136	5,19

(Fin da táboa)

## E. Resultados para o estimador EB

Cadro 5: Resultados co estimador EB

codcoma	Ano 2007		Ano 2008		Ano 2009	
	EB	CV(EB)	EB	CV(EB)	EB	CV(EB)
1501	0,221	7,70	0,236	6,07	0,195	9,01
1502	0,170	6,09	0,190	6,10	0,188	5,49
1503	0,180	14,66	0,258	8,34	0,237	12,07
1504	0,193	5,61	0,219	5,11	0,207	4,91
1505	0,155	9,01	0,162	8,42	0,196	8,71
1506	0,083	5,90	0,076	6,59	0,094	5,91
1507	0,144	12,78	0,097	26,37	0,153	13,49
1508	0,146	4,69	0,141	4,80	0,116	6,65
1509	0,281	5,84	0,187	7,61	0,265	8,48
1510	0,197	11,21	0,181	14,37	0,253	8,04
1511	0,210	6,41	0,193	7,29	0,209	7,59
1512	0,226	5,35	0,167	8,62	0,192	6,72
1513	0,262	8,52	0,269	11,23	0,266	9,54
1514	0,084	8,37	0,099	6,77	0,105	6,64
1515	0,220	8,92	0,163	17,71	0,236	8,19
1516	0,263	8,70	0,250	9,38	0,301	8,90
1517	0,296	6,82	0,256	7,52	0,386	9,36
1518	0,184	9,85	0,210	8,46	0,220	12,77
2701	0,243	7,94	0,224	7,67	0,188	14,30
2702	0,201	8,98	0,185	10,49	0,222	8,77
2703	0,190	12,02	0,205	9,87	0,250	10,60
2704	0,129	5,58	0,103	6,67	0,119	6,33
2705	0,158	8,00	0,163	7,60	0,175	8,27
2706	0,145	9,49	0,156	11,24	0,144	10,02
2707	0,167	11,08	0,082	26,73	0,147	16,47
2708	-	-	0,172	21,49	-	-
2709	0,166	16,88	0,178	14,44	0,214	13,38
2710	0,178	6,93	0,153	9,77	0,178	7,00
2711	0,220	4,67	0,227	5,25	0,196	7,33
2712	0,194	7,04	0,245	5,04	0,208	6,50
2713	0,275	8,96	0,294	8,64	0,218	9,56
3201	0,224	8,33	0,222	6,53	0,220	7,73
3202	0,308	7,75	0,215	13,48	0,259	12,06
3203	0,255	5,65	0,226	6,02	0,227	6,16

(segue na páxina seguinte)

Cadro 5: Continuación da táboa

codcoma	Ano 2007		Ano 2008		Ano 2009	
	EB	CV(EB)	EB	CV(EB)	EB	CV(EB)
3204	0,237	8,33	0,288	5,06	0,217	7,97
3205	0,136	4,66	0,151	3,97	0,150	4,51
3206	0,231	5,91	0,313	4,73	0,302	6,26
3207	0,273	13,21	0,350	14,10	-	-
3208	0,240	7,80	0,374	5,06	0,386	4,89
3209	0,326	9,11	0,264	10,12	0,189	14,87
3210	0,178	7,15	0,203	7,58	0,195	6,66
3211	0,297	5,00	0,354	4,35	0,251	6,86
3212	0,334	8,88	0,391	7,14	0,247	9,50
3601	0,254	5,18	0,199	6,43	0,245	4,59
3602	0,198	7,37	0,202	6,38	0,210	7,78
3603	0,280	4,82	0,236	5,46	0,264	6,23
3604	0,196	7,75	0,225	5,48	0,200	6,41
3605	0,175	5,52	0,177	5,28	0,171	5,51
3606	0,366	7,36	0,263	7,62	0,344	7,39
3607	0,128	5,36	0,144	5,58	0,135	5,86
3608	0,181	4,83	0,180	4,67	0,191	5,38
3609	0,224	5,90	0,178	7,67	0,163	9,64
3610	0,151	3,35	0,136	3,82	0,144	3,88

(Fin da táboa)



## F. Software

Toda a análise estatística deste traballo foi realizado co software R, que está dispoñible de modo gratuito en: <http://www.r-project.org/>.

Os métodos implementáronse empregando esta linguaxe, para o cal se empregaron as seguintes librerías *forecast* e *nlme*. Ademais, para a representación gráfica dos mapas comarcais úsáronse as librerías *maptools* e *spdep*.



# Bibliografía

- Enquisa de condicións de vida das familias. metodoloxía.* Instituto Galego de Estatística (IGE).
- Renda dos fogares municipal. contas de distribución da renda dos fogares por concellos.* Instituto Galego de Estatística (IGE).
- Esteban, M.D., Morales, D., Pérez, A., & Santamaría, L. 2011a. Small area estimation of poverty proportions under area-level time models. *Computational statistics and data analysis.* 26
- Esteban, M.D., Morales, D., Pérez, A., & Santamaría, L. 2011b. Two area-level models for estimating small area poverty indicators. *Journal of the indian society of agricultural statistics.* 26, 27, 28
- 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.*
- Foster, J., Greer, J., & Thorbecke, E. 1984. A class of decomposable poverty measures. *Econometrica.* 6
- González Manteiga, W., Lombardía, M.J., Molina, I., Morales, D., & Santamaría, L. 2008. Analytic and bootstrap approximations of prediction errors under a multivariate fay-herriot model. *Computational statistics and data analysis,* 52, 5242–5252. 21
- Lombardía, M.J., & Morales, D. 2008. *Estimación en áreas pequenas.* Curso de estimación en áreas pequenas.
- Molina, I., & Morales, D. 2009. Small area estimation of poverty indicators. *Boletín de estadística e investigación operativa,* 25, 218–225. 18
- Molina, I., & Rao, J.N.K. 2010. Small area estimation of poverty indicators. *The canadian journal of statistics,* 38, 369–385. 22, 24, 25, 35
- Molina, I., & Rao, J.N.K. Marzo 2009. Small area estimation of poverty indicators. *Statistics and econometrics series 05.*
- Morales, D., & Esteban Lefler, M.D. 2012. Bootstrap procedures in time-correlated area-level linear mixed models. *In: Actas del xxxiii congreso nacional de estadística e investigación operativa. seio 2012.* 45

- Prasad, N.G.N., & Rao, J.N.K. 1990. The estimation of the mean squared error of small-area estimators. *Journal of the american statistical association*. 21
- Rao, J.N.K. 2003. *Small area stimation*. Wiley-Interscience. 20
-