QÜESTIIÓ, vol. 26, 1-2, p. 213-243, 2002

ESTIMADORES COMPUESTOS EN ESTADÍSTICA REGIONAL: UNA APLICACIÓN A LA ESTIMACIÓN DE LA TASA DE VARIACIÓN DE LA OCUPACIÓN EN LA INDUSTRIA*

À. COSTA[1]
A. SATORRA[2]
E. VENTURA[2]

Este trabajo es parte de un proyecto que estudia la aplicación de estimadores compuestos (combinación de estimadores directos e indirectos) para áreas pequeñas en estadística regional. Comparamos tres estimadores: uno directo basado en datos muestrales de cada Comunidad Autónoma (CA), otro sintético (indirecto) que combina los datos estatales con información específica de las CCAA, y un tercer estimador, el compuesto, basado en un modelo estadístico que se concreta en una combinación lineal de los dos anteriores. Ilustramos el método de análisis adoptado mediante la estimación de la tasa de variación de la ocupación industrial en las diferentes CCAA de España.

Composite estimators in regional statistics: an application to the estimation of the rate of change in industrial occupation

Palabras clave: Estadística regional, áreas pequeñas, estimadores compuestos

Clasificación AMS (MSC 2000): 62J07, 62J10, 62H12

---

* Este trabajo es fruto de un proyecto de investigación conjunto entre el Institut d'estadística de Catalunya (IDESCAT) y el Instituto Nacional de Estadística (INE). El INE y el IDESCAT no comparten necesariamente las opiniones expresadas en este trabajo, que son imputables únicamente a los firmantes del mismo. Los autores agradecen los comentarios de N.T. Longford a una versión preliminar de este trabajo.

[1] Institut d'Estadística de Catalunya. Via Laietana 58, 08003 Barcelona.

[2] Departament d'Economia i Empresa. Universitat Pompeu Fabra. Ramon Trias Fargas 25-27, 08005 Barcelona.

– Recibido en diciembre de 2001.
– Aceptado en marzo de 2002.

1. INTRODUCCIÓN

La mayoría de las grandes encuestas que llevan a cabo los organismos estadísticos nacionales han sido diseñadas en base a un tamaño muestral que garantiza la mayor fiabilidad de los estimadores a escala estatal. El tamaño de las muestras y su diseño suelen ser adecuados para proporcionar una precisión aceptable cuando se desea obtener estimadores para algunas áreas más pequeñas, por ejemplo a escala regional o provincial. Sin embargo, en ciertos casos el tamaño de las muestras es insuficiente. Éste es el caso si se quiere estimar no ya los niveles de una variable, sino su tasa de variación interanual, ya que este estadístico tiene una varianza mucho mayor. Esta es una situación normal y muy relevante en el ámbito de la estadística económica. Conocer la evolución de las variables es muchas veces tan o más importante que conocer su nivel.

Si se desea ofrecer información estadística adecuada para esas áreas más pequeñas sólo hay dos alternativas posibles: o se repite la encuesta adecuándola al tamaño del área de interés, o se recurre a la teoría estadística de estimación en pequeñas áreas para mejorar la calidad de los estimadores obtenidos a partir de la encuesta original.

En nuestro país existen encuestas diseñadas y aplicadas a algunas CA, que ofrecen un estimador ideal. Esta situación se da sobre todo en el País Vasco. Sin embargo debemos reconocer que los recursos empleados en la ejecución de estas encuestas son importantes, y la duplicación de cuestionarios en un mismo territorio representa un problema muy significativo. Por esta razón, algunas CCAA se han decantado hacia la utilización de algún tipo de estimador indirecto. Como ejemplo podemos citar el caso de Cataluña, donde el Institut d'Estadística de Catalunya (Idescat) ha elaborado un estimador sintético del Índice de Producción Industrial que ha resultado una alternativa muy adecuada en ese territorio (véase Costa, 1996 o Costa y Galter, 1994). Este estimador es un índice compuesto que utiliza 153 índices primarios estimados a escala nacional, y los pondera en función de su importancia relativa en la economía catalana. El Instituto Nacional de Estadística publicó durante un cierto tiempo unos estimadores regionales del IPI basados en esta metodología; sin embargo, estudios como los de Clar, Ramos y Suriñach (2000) han cuestionado la idoneidad de extender la metodología catalana al resto de territorios del Estado. El problema principal es que, para evitar que los sesgos de los estimadores sintéticos sean relevantes, es necesario que la estructura industrial de la CA sea similar a la del conjunto nacional, y eso no se da en todos los territorios considerados. Por lo tanto el estimador sintético puede ser excesivamente sesgado en algunos casos. En este trabajo pretendemos exponer y aplicar algunas de las soluciones ofrecidas por la teoría estadística para la elaboración de estimadores mejorados.

Existe una variada metodología que versa sobre el desarrollo de estimadores en áreas pequeñas. El lector puede consultar los artículos de Platek, Rao, Särndal y Singh (1987), Isaki (1990), Ghosh y Rao (1994), o Singh, Gambino y Mantel (1994) y obtener una visión de conjunto de las mismas. Una primera clasificación divide los diversos métodos existentes en dos categorías: métodos tradicionales, y métodos basados en modelos. A su vez, los métodos tradicionales pueden incluir estimadores directos, indirectos, o una combinación de ambos. Los estimadores directos tradicionales utilizan únicamente datos provenientes del área pequeña de interés. Son por lo general no sesgados pero un insuficiente tamaño muestral hace que cuenten con una elevada varianza. Los estimadores indirectos tradicionales y los estimadores basados en modelos obtienen más precisión al utilizar también observaciones muestrales que provienen de áreas relacionadas, o áreas vecinas. Los estimadores sintéticos son estimadores tradicionales indirectos que se obtienen a partir de estimadores no sesgados de un área grande. A partir de éstos últimos es posible derivar estimadores para áreas más pequeñas bajo el supuesto de que éstas poseen la misma estructura (a efectos del fenómeno estudiado) que la gran área inicial. El incumplimiento de esta condición puede resultar en estimadores sesgados. Los estimadores compuestos tradicionales son una combinación lineal de estimadores directos y estimadores sintéticos. Los estimadores basados en modelos pueden interpretarse como estimadores compuestos, pero a diferencia de los tradicionales en los que los pesos se determinan de antemano o simplemente se hacen dependientes del tamaño de la muestra, aquí esas ponderaciones dependen de la estructura de covarianzas del estimador. Para más información sobre este tema se pueden consultar los artículos recientes de Cressie (1995), Datta *et al.* (1999), Farrell, MacGibbon y Tomberlin (1997), Ghosh y Rao (1994), Pfefferman y Barnard (1991), Raghunathan (1993), Singh, Mantel y Thomas (1994), Singh, Stukel y Pfeffermann (1998), o Thomas, Longford y Rolph (1994).

El estimador compuesto propuesto en este artículo minimiza el error cuadrático medio, una vez especificado un modelo de comportamiento de la varianza. En este sentido, representa una mejora respecto de estimadores alternativos, directos o sintéticos (indirectos), en áreas pequeñas. Ilustraremos la técnica con ayuda de los datos de ocupación provenientes de la Encuesta de Población Activa, pero la metodología propuesta puede ser aplicada a la estimación de otros indicadores con muy pocas modificaciones. Nos centraremos en concreto en la variación trimestral del empleo total en el sector industrial. La varianza de los datos tiene múltiples componentes: dependerá del territorio considerado y del período considerado.

En la sección 2 exponemos los fundamentos teóricos de la estimación de áreas pequeñas que proponemos. Los datos son descritos en la sección 3 y en la 4 adaptamos la teoría al caso que nos ocupa. La sección 5 expone los resultados principales de la aplicación del estimador a las 17 CCAA españolas. La sección 6 concluye el presente trabajo, resume las conclusiones y sugiere líneas futuras de desarrollo de estimadores regionales. Los apéndices A, B y C muestran los datos originales, algunos cálculos estadísticos, y desarrollos teóricos que por su complejidad se ha preferido mantener al margen del texto principal.

2. FUNDAMENTOS TEÓRICOS DE LA ESTIMACIÓN EN ÁREAS PEQUEÑAS

Supongamos que disponemos de dos estimadores $\widehat{\theta}_1 \sim \left(\theta_1, \sigma_1^2\right)$ y $\widehat{\theta}_2 \sim \left(\theta_2, \sigma_2^2\right)$ que son (aproximadamente) incorrelacionados y deseamos estimar el parámetro θ_2 (en el presente trabajo, $\widehat{\theta} \sim \left(\theta, \sigma^2\right)$ denota que la media y la varianza del estadístico $\widehat{\theta}$ son respectivamente θ y σ_2). En el caso general en que $\theta_1 \neq \theta_2$, es obvio que el criterio de primar a un estimador centrado conduce a elegir $\widehat{\theta}_2$ y, por tanto, a ignorar por completo el valor de $\widehat{\theta}_1$. Sin embargo, cuando la varianza σ_2^2 es grande en comparación con σ_1^2 y con el sesgo al cuadrado $(\theta_1 - \theta_2)^2$, se puede demostrar que la información que proporciona $\widehat{\theta}_1$ no es despreciable en la estimación de θ_2. Un estimador alternativo al estimador centrado de θ_2, es el estimador compuesto

$$\widehat{\theta}_c = \pi\widehat{\theta}_1 + (1-\pi)\,\widehat{\theta}_2, \tag{1}$$

con un peso π adecuado para combinar de forma «óptima» la información sobre θ_2 contenida en los estimadores $\widehat{\theta}_1$ y $\widehat{\theta}_2$.

Un posible y comúnmente aceptado criterio de optimalidad consiste en minimizar la media de los errores de estimación al cuadrado, es decir, minimizar el MSE («mean square error») del estimador. Dado que $MSE\left(\widehat{\theta}_c\right) = \left(E(\widehat{\theta}_c) - \theta_2\right)^2 + Var(\widehat{\theta}_c)$, con

$$E(\widehat{\theta}_c) - \theta_2 = E(\pi\widehat{\theta}_1 + (1-\pi)\,\widehat{\theta}_2) - \theta_2 = \pi\theta_1 + (1-\pi)\,\theta_2 - \theta_2 = \pi\,(\theta_1 - \theta_2)$$

y $Var(\widehat{\theta}_c) = \pi^2\sigma_1^2 + (1-\pi)^2\,\sigma_2^2$ (en el caso de $\widehat{\theta}_1$ y $\widehat{\theta}_2$ incorrelacionados), obtenemos

$$MSE(\widehat{\theta}_c) = \pi^2\,(\theta_1 - \theta_2)^2 + \pi^2\sigma_1^2 + (1-\pi)^2\,\sigma_2^2$$

De manera que minimizar el MSE conduce a resolver la ecuación

$$\pi\,(\theta_1 - \theta_2)^2 + \pi\sigma_1^2 + (1-\pi)\,\sigma_2^2 = 0,$$

es decir,

$$\pi((\theta_1 - \theta_2)^2 + \sigma_1^2 + \sigma_2^2) = \sigma_2^2,$$

de manera que ell valor óptimo de π es $\pi = \sigma_2^2/v$, con $v = (\theta_1 - \theta_2)^2 + \sigma_1^2 + \sigma_2^2$.

Nótese que cuando la varianza σ_1^2 es pequeña con relación a los demás términos que componen v, obtenemos la siguiente expresión

$$\pi = \frac{\sigma_2^2}{(\theta_1 - \theta_2)^2 + \sigma_2^2} \tag{2}$$

Dado que en general el sesgo al cuadrado $(\theta_1 - \theta_2)^2$ es desconocido, aproximaremos dicho valor mediante $\delta^2 = E\,(\theta_1 - \theta_2)^2$, que a menudo se corresponderá con la denominada varianza *entre* grupos. En general σ_2^2 es una varianza *dentro* de grupos.

Cabe remarcar los casos extremos:

1. cuando la dispersión dentro de los grupos es pequeña en relación al sesgo, $\pi = 0$ y el estimador compuesto coincide con $\widehat{\theta}_2$;
2. cuando la dispersión entre grupos δ^2 es pequeña en relación a la varianza entre grupos, entonces el estimador compuesto coincide con $\widehat{\theta}_1$.

Un caso más general vendría descrito por dos estimadores $\widehat{\theta}_1 \sim \left(\theta_1, \sigma_1^2\right)$ y $\widehat{\theta}_2 \sim \left(\theta_2, \sigma_2^2\right)$ en donde la varianza σ_1^2 no es ignorable y la covarianza $\gamma = cov(\theta_1, \theta_2)$ es no nula. En este caso se demuestra que el valor óptimo π viene dado por (Longford, 2001)

$$\pi = \frac{\sigma_2^2 - \gamma}{(\theta_1 - \theta_2)^2 + \sigma_1^2 + \sigma_2^2 - 2\gamma} \tag{3}$$

Nótese que en dicho caso será necesaria también información sobre el valor de la covarianza γ. En el caso de muestreo aleatorio simple dicha covarianza será igual a $q \times \sigma_2^2$ con $0 \leq q \leq 1$, donde q es la fracción de muestra aportada por la k-ésima CA, de manera que la expresión anterior se transforma en

$$\pi = \frac{\sigma_2^2\,(1-q)}{(\theta_1 - \theta_2)^2 + \sigma_1^2 + \sigma_2^2(1-2q)} \tag{4}$$

Resultados parecidos se obtienen en un contexto multivariante. Si lo que se desea es estimar $w'\theta_2$, para un vector de pesos w arbitrario, a partir de los estimadores vectoriales $\widehat{\theta}_1 \sim (\theta_1, \Sigma_1)$ y $\widehat{\theta}_2 \sim (\theta_2, \Sigma_2)$, el estimador compuesto se obtendrá a partir de la combinación lineal $w'\widehat{\theta}_c$ con

$$\widehat{\theta}_c = \Pi\widehat{\theta}_1 + (1 - \Pi)\widehat{\theta}_2 \tag{5}$$

La expresión de la matriz de pesos Π que minimiza el MSE en el caso simplificado en que $\widehat{\theta}_1$ y $\widehat{\theta}_2$ no están correlacionados, es $\Pi = \Sigma_2 V^{-1}$ con $V = bb' + \Sigma_1 + \Sigma_2$ y $b = (\theta_1 - \theta_2)$ (Longford, 1999). La expresión de b, en general desconocida, se sustituirá por la siguiente matriz de varianzas y covarianzas *entre* grupos

$$\Delta = E\,(\theta_1 - \theta_2)\,(\theta_1 - \theta_2)'$$

Dichos estimadores vectoriales surgen al tratar la estimación de varios parámetros simultáneamente, por ejemplo las tasas de variación de paro de varios sectores económicos.

3. ANÁLISIS ESTADÍSTICO DE LOS DATOS

Los datos utilizados en este artículo provienen de la Encuesta de Población Activa (EPA) y de la Contabilidad Regional de España (CRE). Disponemos de datos trimestrales de la EPA de Ocupación industrial en España y en cada CCAA por ramas de actividad, desde 1987. También contamos con datos anuales de Empleo industrial y por ramas provenientes de la CRE, para España y las CCAA, desde 1995 a 1997.

En este trabajo nos centramos en el sector industrial. El Apéndice B muestra las ramas de actividad del sector industrial consideradas por la CRE, y como hemos construido su equivalencia con las ramas de actividad consideradas por la EPA, algo más desagregadas.

Los datos originales de la EPA representan miles de personas ocupadas en cada Comunidad. A partir de ellos, calculamos las tasas de variación trimestral de la ocupación en cada área geográfica en la forma habitual.

La notación básica utilizada es la siguiente:

- $y_{ij}(k,t)$ es la tasa de variación de la ocupación en la rama j del sector productivo i en el territorio k, y en el trimestre t.
- $y_i(k,t)$ es la tasa de variación de la ocupación en el sector productivo i y la Comunidad Autónoma[1] k, en el trimestre t.
- $y_{ij}(\cdot,t)$ es la tasa de variación de la ocupación en la rama j del sector productivo i en el conjunto del Estado Español, en el trimestre t.
- $y_i(\cdot,t)$ es la tasa de variación de la ocupación en el sector productivo i en el conjunto del Estado Español, en el trimestre t.

Las variables $y_{ij}(\cdot,t)$, $y_i(\cdot,t)$ y $y_i(k,t)$ pueden obtenerse como estimadores directos a partir de los datos disponibles en la EPA aunque el diseño de esta encuesta no permite obtener estimadores suficientemente precisos en el último de los casos.

También podemos crear un estimador sintético para cada territorio inspirado en la metodología del Idescat (véase Costa y Galter, 1994), combinando el estimador directo al nivel nacional y la información suplementaria extraída de la CRE, gracias a la cual podemos adaptar la información nacional a la estructura de la industria de cada una de las diferentes CCAA.

[1]No disponemos de datos detallados trimestrales por ramas de actividad dentro de cada sector productivo y autonomía, sólo anuales.

Calculamos

(6) $$z_i(k,t) = \sum_{j=1}^{J_i} w_{ij}(k) y_{ij}(\cdot,t)$$

donde J_i es el número de ramas en el sector i, y $w_{ij}(k)$ es el peso de la rama j en el sector i de la Comunidad k, en un periodo determinado y fijo (en concreto, el año 1997).

El Apéndice numérico A muestra los estimadores directos y sintéticos industriales de cada Comunidad para cada trimestre, así como las ponderaciones y valores nacionales utilizados en la elaboración del estimador sintético.

En este punto contamos con dos estimadores: el directo que es insesgado pero muy impreciso, y el sintético, que es sesgado aunque con una varianza más reducida.

4. ESTIMACIÓN DE LA TASA DE VARIACIÓN DE LA OCUPACIÓN EN LA INDUSTRIA

Proponemos aquí un estimador compuesto de la tasa de variación de la ocupación de cada CA, de periodicidad trimestral. El estimador compuesto es una combinación lineal de un estimador sintético y uno directo donde los pesos otorgados a cada componente dependen de la composición de la varianza de los indicadores de cada Comunidad.

Los dos estimadores $z_i(k,t)$ y $y_i(k,t)$ representan, respectivamente, los estimadores sintético y directo de la variación de la ocupación para el sector industrial de la Comunidad k en el momento t. Es decir suponemos que $\widehat{\theta}_2(k,t) = y_i(k,t) \sim (\theta_2(k,t)\ ,\ \sigma_2^2(k,t))$ y $\widehat{\theta}_1(k,t) = z_i(k,t) \sim (\theta_1(k,t)\ ,0)$. Si definimos $\delta(k,t) = \theta_2(k,t) - \theta_1(k,t)$, de acuerdo con las expresiones (1) y (2) de la Sección 2 obtenemos el estimador compuesto

(7) $$\widehat{\theta}_c(k,t) = \pi(k,t)\widehat{\theta}_1(k,t) + (1-\pi(k,t))\,\widehat{\theta}_2(k,t)$$

donde

(8) $$\pi(k,t) = \frac{\sigma_2^2(k,t)}{\delta^2(k,t) + \sigma_2^2(k,t)}$$

Dadas las características de muestreo de la EPA, y suponiendo constancia a lo largo del tiempo de los tamaños muestrales, consideramos que las varianzas muestrales $\sigma_2^2(k,t)$ son constantes en el tiempo, es decir $\sigma_2^2(k,t) = \sigma_2^2(k)$, y que los sesgos al cuadrado $\delta(k,t)^2$ se comportan (para cada Comunidad), como un proceso estocástico estacionario en media. Por tanto, aproximaremos $\delta(k,t)^2$ por su valor promedio $\varpi(k) =$

$\frac{1}{T}\sum_{t=1}^{T}\delta(k,t)^2$. Se observa que un estimador de la suma $\varpi(k)+\sigma_2^2(k)$ es precisamente la media muestral (para cada Comunidad) de la serie de diferencias al cuadrado

$$s_d^2(k)=\frac{1}{T}\sum_{t=1}^{T}d(k,t)^2=\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\theta}_2(k,t)-\widehat{\theta}_1(k,t)\right)^2 \tag{9}$$

y dado que $\widehat{\theta}_2(k,t)-\widehat{\theta}_1(k,t)=\theta_2(k,t)-\theta_1(k,t)+\varepsilon(k,t)$ donde $\varepsilon(k,t)$ es el error de muestreo, obtenemos (en el caso en que $\widehat{\theta}_1$ y $\widehat{\theta}_2$ no estén correlacionados)

$$s_d^2(k)=\frac{1}{T}\left(\sum_{t=1}^{T}\delta(k,t)^2+\sum_{t=1}^{T}\varepsilon(k,t)^2+2\sum_{t=1}^{T}\delta(k,t)\varepsilon(k,t)\right) \tag{10}$$

que obviamente converge hacia $\varpi(k)+\sigma_2^2(k)$, la expresión del denominador de $\pi(k,t)$. De este modo el peso óptimo vendrá dado por

$$\pi(k)=\frac{s_2^2(k)}{s_d^2(k)} \tag{11}$$

donde $s_2^2(k)$ es un estimador de la varianza muestral $\sigma_2^2(k)$. El Apéndice C muestra que el valor de esta varianza puede estimarse mediante la siguiente expresión

$$s_2^2(k)=2(CV(k))^2, \tag{12}$$

donde $CV(k)$ es el error relativo de muestreo medio calculado por el INE para la variable Ocupación en la k-ésima CA. Dicha expresión incorpora ya de forma automática la corrección necesaria para atender el efecto de diseño complejo de la muestra de la EPA.

En el caso en que la varianza del estimador sintético no sea ignorable, y exista una covarianza no nula entre los dos estimadores $\widehat{\theta}_1$ y $\widehat{\theta}_2$, puede verse que $s_d^2(k)$ continúa siendo un estimador consistente ahora del denominador de (4), de manera que en dicho caso la expresión del peso óptimo será

$$\pi(k)=\frac{s_2^2(k)-\widehat{\gamma}(k)}{s_d^2(k)} \tag{13}$$

donde $\widehat{\gamma}(k)$ es un estimador de la covarianza entre estimadores de la k-ésima CA[2].

Cuando la información histórica de cada comunidad no permita una estimación fiable de su sesgo específico (o en el caso extremo en que a priori se asigne una magnitud de sesgo común a todas las comunidades), es fácil ver que el valor común del sesgo será

[2] En general $\widehat{\gamma}(k)=q(k)s_2^2(k)$ donde $q(k)$ corresponde a la fracción de muestra equivalente aportada por la k-ésima CA.

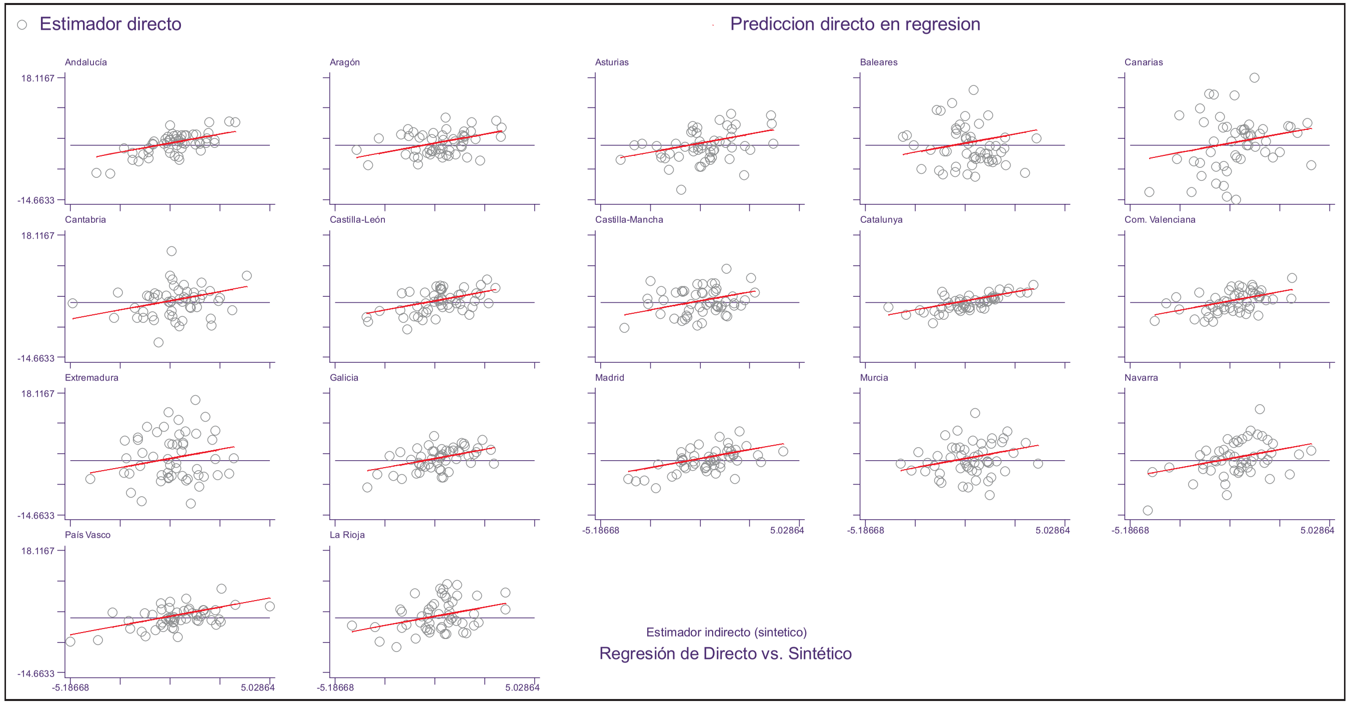

Figura 1. Regresión de Directo sobre Sintético

estimado por $[\frac{1}{K}\sum_k s_d^2(k) - \frac{1}{K}\sum_k s_2^2(k)]$, y la expresión del peso del estimador compuesto será

$$\pi(k) = \frac{s_2^2(k) - \widehat{\gamma}(k)}{[\frac{1}{K}\sum_k s_d^2(k) - \frac{1}{K}\sum_k s_2^2(k)] + s_2^2(k)}, \tag{14}$$

cuyo valor vemos que varía en cada CA a través solamente de la varianza muestral $s_2^2(k)$ y de la covarianza $\widehat{\gamma}(k)$.

5. ILUSTRACIÓN EN EL SECTOR INDUSTRIAL

En este apartado incluimos los cálculos efectuados para obtener el estimador compuesto y los resultados para las diferentes CCAA.

Primeramente hemos obtenido un estimador directo de la tasa de variación de la ocupación en la industria para cada CA. El estimador se obtiene directamente a partir de los datos de la EPA, restringidos a un área geográfica concreta. El cuadro 1 del Apéndice A muestra las tasas de variación calculadas mediante este procedimiento. Puede apreciarse que existen CCAA que presentan una gran variabilidad temporal en sus tasas de ocupación (Canarias, Baleares o La Rioja por ejemplo), mientras que otras como Madrid o Cataluña presentan una evolución más suave de la ocupación.

También hemos calculado el estimador sintético presentado en la ecuación (6). El estimador sintético resulta de ponderar las tasas de variación de las ramas industriales a nivel estatal provenientes de la EPA (véase el cuadro 3 del Apéndice A), con los porcentajes de ocupación industrial por ramas de cada Comunidad[3] que se obtienen de la CRE (véase el cuadro 4 del Apéndice A). Dado que la EPA cuenta con más de las 14 ramas industriales con las que cuenta la CRE, aplicamos las equivalencias que describimos en el Apéndice B con objeto de homogeneizar las series utilizadas en el cálculo del estimador. Los estimadores indirectos obtenidos para cada CA están recogidos en el cuadro 2 del Apéndice A.

La figura 4 permite comprobar la variabilidad del estimador directo en relación al indirecto. La serie de estimadores sintéticos de alguna manera «suaviza» la serie de estimadores directos, en cada CA. Se aprecia no obstante que la diferencia entre ambos estimadores es diferente según consideremos unas u otras CCAA. Por ejemplo cabe destacar el caso de Cataluña, donde ambos estimadores se encuentran bastante próximos; o los de Canarias o Extremadura en que la diferencia se acrecienta. Estas particularidades pueden observarse mejor con ayuda de la figura 6, que muestra mediante diagramas de

[3]Aunque en el presente artículo hemos optado por mantener unas ponderaciones fijas correspondientes al año 1997, el estimador sintético podria mejorarse utilizando ponderaciones variables en el tiempo.

caja la distribución de la diferencia entre los estimadores directo e indirecto trimestrales, para cada una de las 17 CCAA.

Es evidente que el estimador sintético suaviza en exceso la evolución de la tasa de variación de ocupación y dada la relativamente alta fiabilidad de los datos directos estimados en la EPA no resulta una buena alternativa en la mayoría de las CCAA.

Para calcular el estimador compuesto, necesitamos obtener los pesos otorgados a los estimadores directo e indirecto, específicos de cada CA, que hemos mostrado en la ecuación (2). Hemos optado por esta expresión simplificada dada la dificultad, en este momento, de contar con estimadores adecuados de la covarianza entre estimadores directos y sintéticos. El denominador del peso $\pi(k)$ se obtiene calculando el promedio de la diferencia entre los estimadores directo e indirecto, al cuadrado. Esta expresión es un estimador del sesgo al cuadrado más la varianza del error muestral del estimador directo. El numerador, que coincide con la varianza del error muestral del estimador directo, se obtiene multiplicando por dos el cuadrado del error de muestreo relativo que el INE otorga a la variable Ocupación en la EPA. La tabla 1 muestra los resultados intermedios y el valor final de los pesos.

Tabla 1. Cálculo de los pesos del estimador compuesto

	$CV^c(O)$	$s_2^2(k)$	$s_d^2(k)$	$\pi(k)$
Andalucía	1,202	2,89	4,748	0,609
Aragón	1,728	5,97	8,616	0,693
Asturias	2,512	12,62	14,498	0,875
Baleares	2,264	10,25	32,553	0,315
Canarias	2,342	10,97	49,223	0,223
Cantabria	2,738	14,99	17,084	0,878
Castilla-León	1,466	4,30	6,791	0,633
Castilla-La Mancha	1,526	4,66	12,005	0,388
Cataluña	1,382	3,82	2,086	1,00[4]
C. Valenciana	1,454	4,23	5,880	0,719
Extremadura	2,574	13,25	39,589	0,335
Galicia	1,708	5,835	6,591	0,885
Madrid	1,602	5,13	6,293	0,816
Murcia	2,442	11,93	18,664	0,639
Navarra	2,524	12,74	17,326	0,735
País Vasco	1,688	5,70	6,287	0,906
La Rioja	3,002	18,02	18,427	0,978

Nota: C.V. está calculado a partir de los valores medios de 1992.

[4] Este peso ha sido truncado a 1. En el caso de Cataluña se espera una correlación alta entre el estimador directo y el sintético, correlación que induce sesgo en la fórmula utilizada de cálculo del coeficiente para dicha CA. El cálculo del peso óptimo para Cataluña teniendo en cuenta dicha correlación involucraría características del diseño de muestra que creemos exceden el ámbito del presente trabajo.

En general, el estimador compuesto otorga un peso mayor al estimador indirecto en casi todas las CCAA. Se observa que el estimador compuesto otorga menor peso al estimador indirecto en aquellas Comunidades donde la varianza del error de muestreo es menor en relación a su sesgo promedio. Así ocurre en Castilla-La Mancha, Extremadura, Baleares o Canarias, como casos más extremos. En cambio, el estimador sintético recibe mayor peso en el caso de Cataluña, País Vasco o La Rioja entre otras, dado que en estas CCAA el sesgo promedio del estimador sintético es pequeño en relación a la varianza del error de muestreo. Se advierte como algunas de las regiones más industrializadas cuentan con un valor muy alto, mientras que los dos archipiélagos y algunas regiones menos industrializadas muestran valores más bajos. Estos resultados son llamativos ya que el estimador directo pierde peso precisamente en las CCAA mayores, mientras que lo pierde en las CCAA que pueden ser más propiamente consideradas pequeñas áreas.

A continuación examinamos el comportamiento de este estimador en el caso de dos Comunidades concretas que destacan por tener poco y mucho peso respectivamente otorgado al estimador indirecto : Canarias y el País Vasco[5].

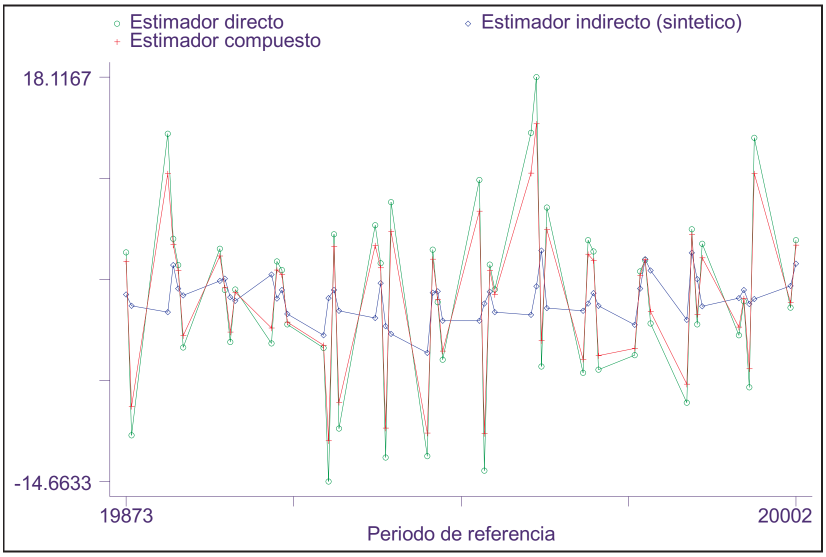

Figura 2. Comunidad Autónoma de Canarias

[5]El estimador compuesto en el caso de Cataluña o La Rioja, otorga un peso mayor todavía al estimador indirecto. En el caso de Cataluña ambos coinciden completamente, y en el caso de La Rioja, las discrepancias són mínimas, por lo que hemos preferido escoger otra Comunidad para mostrar un gráfico más claro de las diferencias entre los tres distintos estimadores.

En el caso de Canarias (figura 2), el estimador compuesto otorga casi el 78% del peso al estimador directo. Dado que el estimador directo de la tasa de variación de la ocupación en esta Comunidad a lo largo del período fluctúa entre los valores del 18,1% y −14,7%, las diferencias en las estimaciones puntuales de unos y otros estimadores pueden ser importantes. Esta última circunstancia se observa mejor en las figuras 6 y 7, que muestra mediante diagramas de caja la diferencia entre los valores del estimador directo e indirecto, y compuesto e indirecto de cada CA.

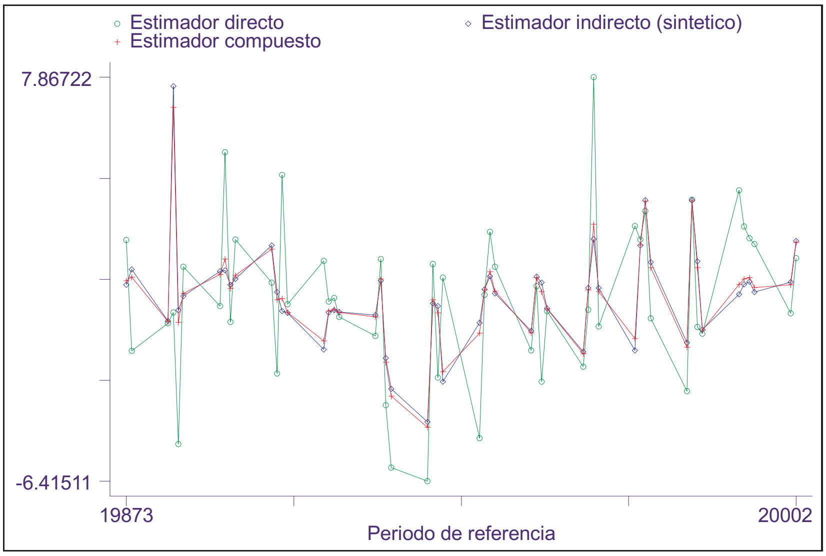

Figura 3. Comunidad Autónoma del País Vasco

En el caso del País Vasco (ver figura 3), donde la diferencia entre la estimación puntual del estimador directo y del sintético es también apreciable, con estimaciones directas de la tasa de variación de la ocupación que oscilan entre el 7,8% y el −6,4%, el estimador compuesto da más peso al estimador indirecto (la ponderación es 0,91). Esto es así porque el error de muestreo en esta Comunidad es bastante acusado en relación al sesgo incorporado en el estimador sintético.

Estos son dos casos extremos. La figura 5 muestra los tres estimadores: directo, indirecto y compuesto en las 17 CCAA.

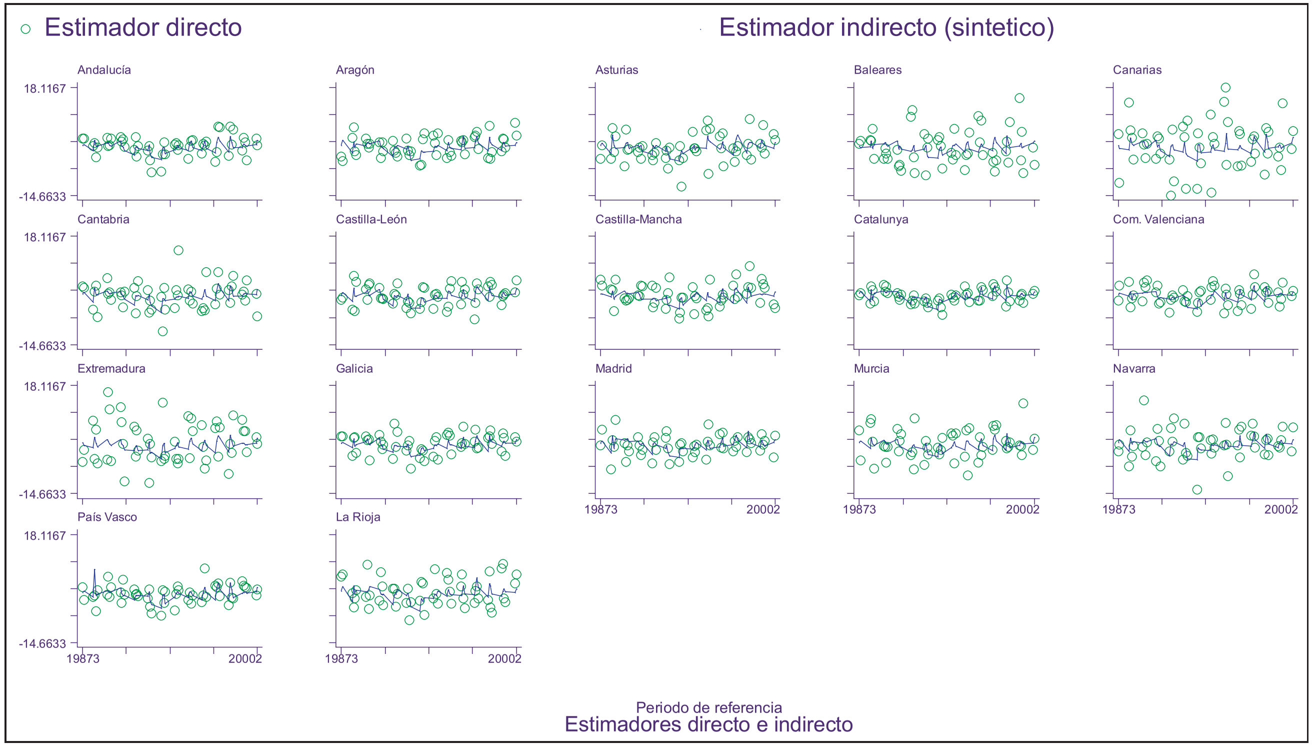

Figura 4. Estimadores directo e indirecto, por Comunidad

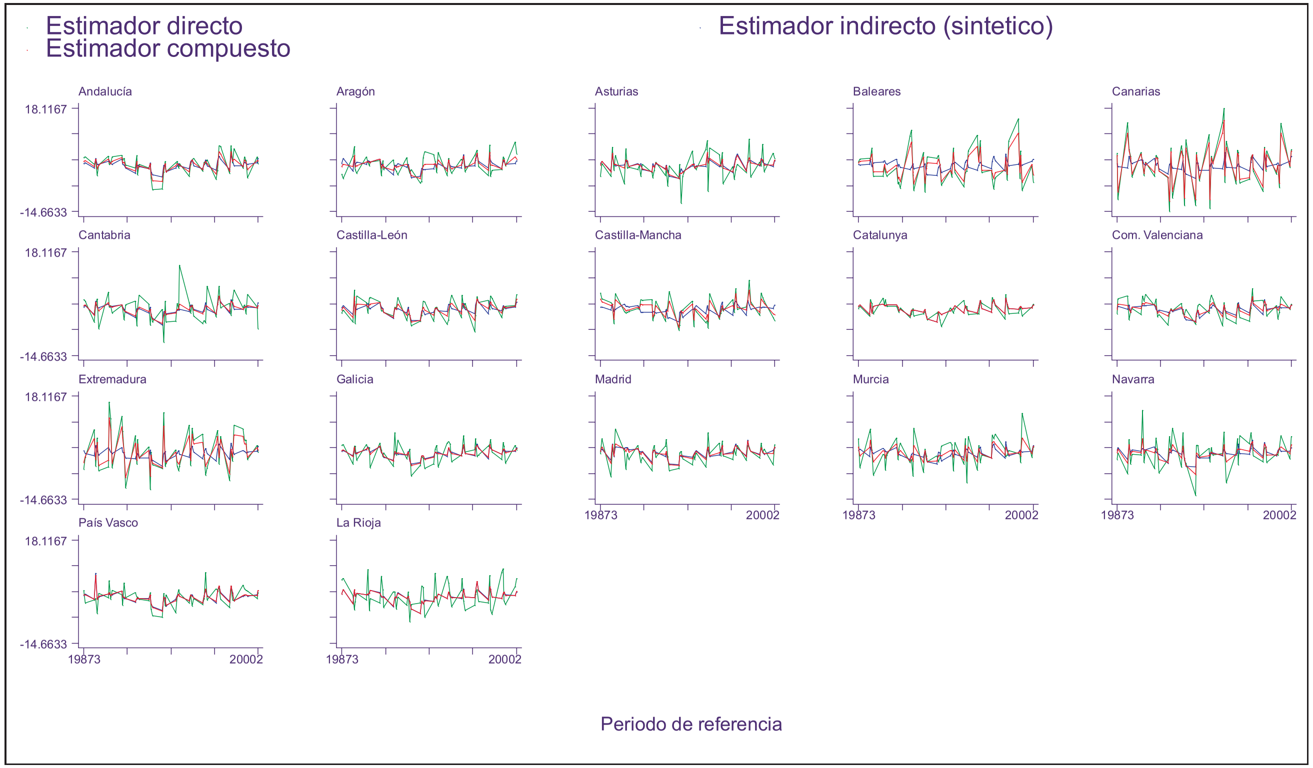

Figura 5. Comparación de estimadores directo, indirecto y compuesto, por Comunidad

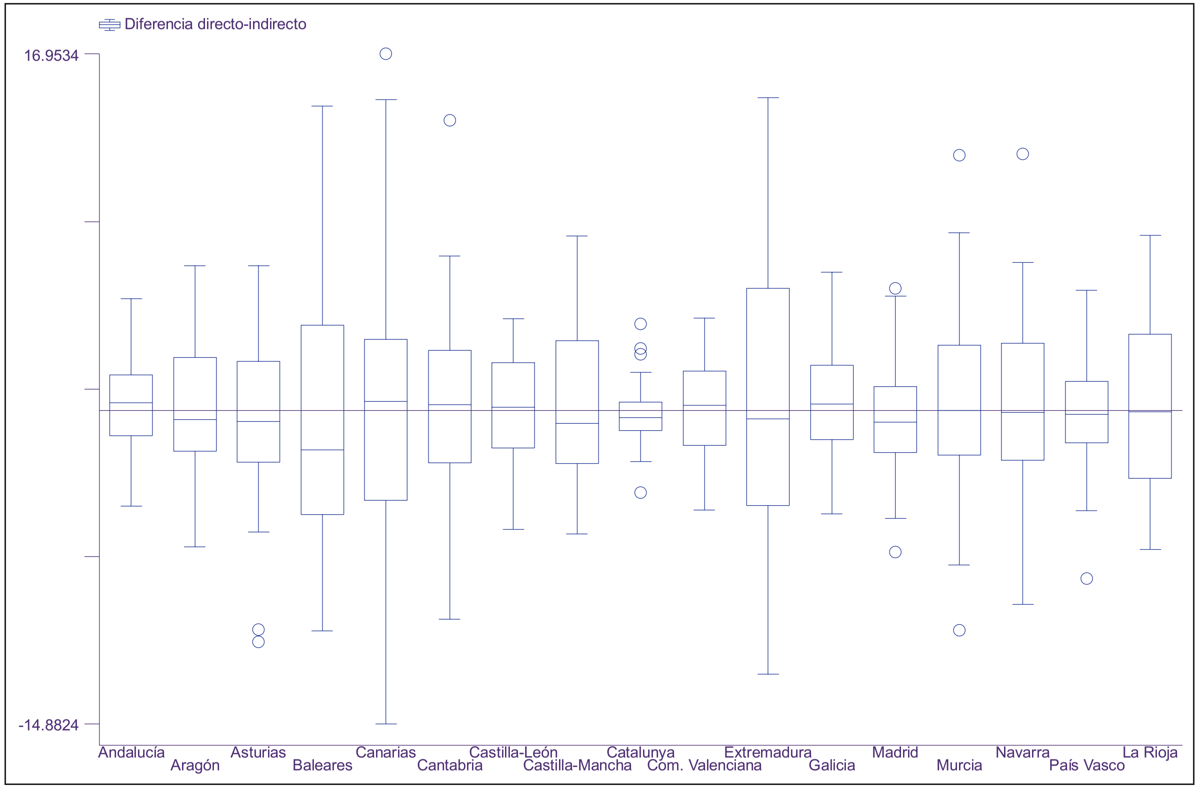

Figura 6. Distribución por Comunidad de la diferencia entre los estimadores directo e indirecto

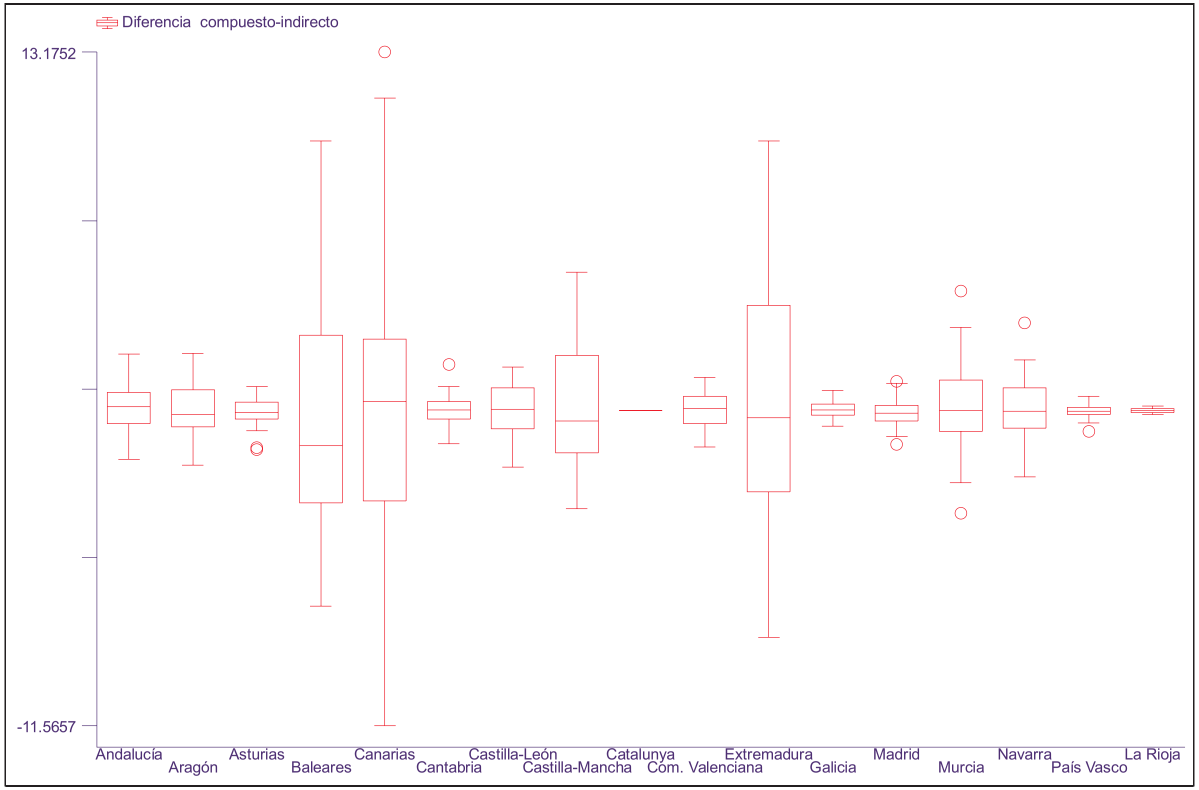

Figura 7. Diferencia por comunidades entre los estimadores compuesto e indirecto

6. CONCLUSIONES

En este trabajo hemos creado un estimador compuesto aplicado a la estimación de la variación de la tasa de ocupación de 17 CCAA españolas. El estimador compuesto es el resultado de combinar de forma lineal otros dos estimadores: el directo y el sintético. El primero es un estimador insesgado pero poco preciso, mientras que el segundo posee una varianza más reducida pero suele ser sesgado. Si asumimos un modelo de comportamiento para el estimador directo en función del indirecto, las ponderaciones utilizadas para obtener el estimador compuesto son función del sesgo del estimador sintético, y de la varianza del error muestral del estimador directo. De esta forma, el estimador compuesto resultante es el que minimiza el error cuadrático medio, dado el modelo adoptado.

El estimador resultante otorga *más peso al estimador directo en aquellas Comunidades donde la varianza del error de muestreo es menor en relación al sesgo promedio del estimador indirecto*. Así ocurre en las Comunidades de Canarias, Baleares, Castilla-La Mancha, o Extremadura, como casos más extremos. En cambio, el estimador sintético gana más protagonismo en el caso de Cataluña, La Rioja, País Vasco, o Madrid, dado que en estas Comunidades el sesgo promedio del estimador sintético es pequeño en relación a la varianza del error de muestreo.

Cabe señalar que la versión del estimador compuesto aquí presentada puede incorporar múltiples refinamientos, desde una modificación adecuada de los pesos en determinados períodos con características especiales, a un tratamiento diferenciado de unas determinadas CCAA. Asimismo la técnica utilizada puede aplicarse a otros conjuntos de datos.

Podemos señalar tres conclusiones principales:

1. Puede verse para Cataluña que el estimador compuesto es equivalente al sintético, lo que avalaría la utilización por parte del Idescat de indicadores sintéticos para aproximar tasas de variación interanual (tal como se está haciendo en los casos del IPI y del IPRI).
2. Por lo que respecta a la adopción por parte del INE (durante la segunda parte de los años noventa) de los estimadores sintéticos para hacer un IPI regional y de algunas CCAA (que lo están utilizando en la actualidad, tanto para el IPI como para el IPRI), habría que coincidir con Clar, Ramos y Suriñach (2000), en que estas estimaciones pueden ser adecuadas en algunos casos, pero no lo son en todos los casos. En algunas CCAA como Cataluña, claramente son aceptables, pero en otras, como en las islas, no parece ser un camino válido.
3. El anterior resultado no deja de ser curioso: en el estimador compuesto gana peso el estimador sintético cuando las CCAA son menos «pequeñas áreas», perdiéndolo en aquellas CCAA que de forma más natural podrían ser tomadas como auténticas pequeñas áreas.

Nos encontramos con un comportamiento paradójico de los estimadores compuestos. En ellos tiene mucho peso el estimador sintético en «pequeñas áreas» muy pequeñas (para las cuales el estimador directo tiene una varianza prácticamente infinita), y también en el caso de las «pequeñas áreas» muy grandes, como en el caso catalán (por tener el estimador sintético un sesgo irrelevante). Esto implicaría un comportamiento no lineal del peso del estimador directo ante incrementos en la dimensión de la pequeña área. También supone un prometedor papel de la estimación indirecta en el terreno de la estadística regional.

El proyecto más inmediato en nuestra agenda de trabajo consiste en estudiar las propiedades muestrales de los estimadores compuestos. Sesgo, varianzas poblacional y muestral del estadístico, y dato a estimar (nivel, proporción o tasa de variación), son aspectos determinantes de los pesos con los que los estimadores compuestos basados en modelos combinan la información disponible. Es por lo tanto fundamental explorar, a través de diversos escenarios, la manera en que las características poblacionales regionales o locales (que pueden estar muy centradas o ser muy diferentes del área general), el tamaño muestral, el tipo de dato a estimar, etc., influyen en su comportamiento. Asimismo será necesario aplicar técnicas específicas de estimación de sesgo y varianza dado que no todos los escenarios permiten derivar expresiones analíticas útiles. Este estudio puede llegar a tener implicaciones en los diseños muestrales y, en concreto, en las distribuciones de la muestra entre los distintos subdominios de las encuestas (afijaciones), en el caso de tener previamente presente que éstos serán estimados no por estimadores directos, sino por medio de estimadores compuestos.

7. REFERENCIAS

Clar, M., Ramos, R. y Suriñach, J. (2000). «Avantatges i inconvenients de la metodologia de l'INE per a elaborar indicadors de la producció industrial per a les regions espanyoles», *Qüestiió*, 24, 1, 151-186.

Costa, A. (1996). «El IPPI, un indicador de la coyuntura industrial catalana», *Fuentes estadísticas*, 17. INE.

Costa, A. y Galter, J. (1994). «L'IPPI, un indicador molt valuós per mesurar l'activitat industrial catalana». *Revista d'indústria*, 3, Generalitat de Catalunya, Departament d'Indústria i Energia, 6-15.

Cressie, N. (1995). «Bayesian Smoothing of Rates in Small Geographic Areas». *Journal of Regional Science*, 35 (4), 659-73.

Datta, G. S., *et al.* (1999). «Hierarchical Bayes Estimation of Unemployment Rates for the States of the U.S». *Journal of the American Statistical Association*, 94 (448), 1074-82.

Farrell, P. J, Macgibbon, B. y Tomberlin, T. J. (1997). «Empirical Bayes Small-Area Estimation Using Logistic Regression Models and Summary Statistics». *Journal of Business & Economic Statistics*, 15 (1), 101-8.

Ghosh, M. y Rao, J. N. K. (1994). «Small Area Estimation: an Appraisal». *Statistical Science*, 9, 1, Canada Statistics, 55-93.

INSTITUTO NACIONAL DE ESTADÍSTICA (1994). *Evaluación de la calidad de los datos de la Encuesta de Población Activa*. Año 1992.

Isaki, C. T. (1990). «Title Small-Area Estimation of Economic Statistics». *Journal of Business & Economic Statistics*, 8 (4), 435-41.

Longford, N. T. (1999). «Multivariate shrinkage estimation of small area means amd proportions», *Journal od the Royal Statistical Society*, A, 162, 227-246.

Longford, N. T. (2001). «Synthetic estimators with moderating influence: the carry-over in cross-over trials revisited». *Statistis in Medicine*, 20, 3189-3203.

Pfeffermann, D. y Barnard, C. H. (1991). «Some New Estimators for Small-Area Means with Application to the Assessment of Farmland Values». *Journal of Business & Economic Statistics*, 9 (1), 73-84.

Platek, R., Rao, J. N. K., Särndal, C. E. y Singh, M. P. (Eds.) (1987). *Small Area Statistics: An International Symposium*. New York; John Wiley and Sons.

Prasad, N. G. N. y Rao, J. N. K. (1990). «The estimation of Mean Squared Error of Small-Area Estimators», *Journal of the American Statistical Association*, 85, 163-171.

Raghunathan, T. E. (1993). «A Quasi-empirical Bayes Method for Small Area Estimation», *Journal of the American Statistical Association*, 88 (424), 1444-48.

Rao, C. R. (1973). *Linear statistical inferece and its applications*, 2nd edn., New York: John Wiley.

Singh, M. P., Gambino, J. y Mantel, H. J. (1994). «Issues and Strategies for Small Area Data», *Survey Methodology*, 20, 1, Statistics Canada, 3-22.

Singh, A. C., Mantel, H. J. y Thomas, B. W. (1994). «Time Series EBLUPs for Small Areas Using Survey Data», *Survey Methodology*, 20, 1, Canada Statistics, 33-43.

Singh, A. C., Stukel, D. M. y Pfeffermann, D. (1998). «Bayesian versus frequentist measures of error in small area estimation», *Journal of the Royal Statistical Society*, B, 60, 377-396.

Thomas, N., Longford, N. T. y Rolph, J. E. (1994). «Empirical Bayes Methods for Estimating Hospital-Specific Mortality Rates», *Statistics in Medicine*.

APÉNDICE

A. TABLAS ESTADÍSTICAS DE LOS ESTIMADORES

Este apéndice muestra las tablas de los estimadores directos e indirectos a partir de los cuales hemos calculado nuestro estimador compuesto.

Tasas de variación. Estimadores DIRECTOS por Comunidad. Sector Industrial

AÑO	TRIM	AND	ARA	AST	BAL	CAN	CANT	CAS-L	CAS-M	CAT	VAL	EXT	GAL	MAD	MUR	NAV	PVAS	RIO	VARIANZA
1987	3	2,496	-2,837	-3,456	0,972	3,928	2,997	-1,161	5,087	0,186	-1,386	-5,048	2,548	-0,179	4,485	-1,966	2,109	5,386	9,901
1987	4	2,628	-4,259	0,524	1,948	-10,920	2,613	-0,346	3,172	0,836	3,267	-1,410	2,674	4,663	-4,591	0,173	-1,804	5,975	16,633
1988	1	0,305	2,787	-3,550	2,136	13,548	-4,043	-4,183	1,982	-2,252	4,327	7,435	-3,341	-7,453	6,696	-6,508	-0,837	0,390	29,808
1988	2	1,284	5,984	5,703	1,981	5,026	1,313	6,197	1,734	2,985	-3,380	-4,068	2,157	1,407	7,774	4,394	-0,442	-4,133	12,935
1988	3	-3,153	-1,265	-1,491	5,569	2,892	3,268	-4,542	-4,454	1,835	-0,663	4,756	2,345	-0,487	0,949	-3,449	-5,109	-1,363	12,641
1988	4	-0,044	-2,506	3,207	-3,496	-3,802	-6,276	4,300	4,984	-1,532	0,165	-5,562	-2,336	7,742	-7,017	-0,366	1,164	1,121	20,000
1989	1	2,815	2,551	-5,788	-3,517	4,204	5,441	2,120	-0,760	3,483	-1,668	-4,327	2,646	0,697	-0,277	-0,644	-0,217	-0,850	45,489
1989	2	0,283	0,886	5,226	-0,395	0,867	0,688	-1,158	-1,299	1,284	4,143	16,122	1,526	-4,998	-0,625	13,657	5,216	8,867	30,969
1989	3	0,498	-0,243	-0,773	-2,028	-3,360	1,051	3,431	-1,029	2,945	2,647	10,927	1,663	-0,102	-5,407	-3,093	-0,777	-4,381	21,925
1989	4	2,632	0,993	-0,623	-3,465	0,911	-1,118	3,763	-0,766	1,012	4,756	-4,746	-4,642	-1,916	-0,591	0,924	2,126	-0,429	24,803
1990	1	3,130	1,530	-1,399	-0,932	-3,466	1,078	2,428	0,151	0,987	1,150	11,604	3,120	2,444	4,829	-4,944	0,609	-1,977	25,287
1990	2	1,634	-3,297	-0,136	-5,812	3,165	0,312	-0,594	0,000	-0,242	3,228	6,972	1,043	-0,197	-1,970	4,886	-2,603	6,637	12,590
1990	3	2,304	1,520	-3,444	-5,421	2,471	-3,320	-0,586	3,093	-1,079	-2,291	0,429	-1,190	-2,991	1,287	1,186	4,419	2,270	10,689
1990	4	0,010	-1,377	0,318	-7,264	-1,929	1,288	-0,913	2,871	0,356	-1,778	-10,959	0,173	-1,348	3,226	-6,386	-0,166	-3,329	28,962
1991	1	-0,841	1,801	-2,382	9,228	-3,852	2,596	-2,159	3,084	-2,551	-5,100	5,563	-4,093	-5,467	-2,973	-1,974	1,375	1,635	177,515
1991	2	3,091	-1,075	-3,907	11,265	-14,663	-5,241	0,638	-3,486	-1,441	-1,144	-3,422	6,442	0,869	-2,052	5,197	-0,066	1,574	35,464
1991	3	-2,934	-1,817	-1,701	-3,094	5,395	0,082	0,296	-1,957	0,131	0,167	4,359	-0,645	2,449	-7,261	1,026	0,062	-3,875	25,636
1991	4	0,267	-2,919	1,107	-7,956	-10,379	4,574	3,758	2,613	-2,195	-0,394	-1,593	4,080	-1,414	7,987	7,986	-0,613	0,105	51,263
1992	1	-1,000	2,496	-3,210	4,046	6,134	1,718	-3,622	-3,824	-1,511	1,848	2,046	-1,696	2,239	-6,088	-4,340	-1,287	-2,661	140,095
1992	2	0,397	-0,987	-0,975	-8,485	3,039	-4,991	-1,512	-1,215	-0,726	-1,488	-11,520	-0,142	4,289	1,284	6,470	1,433	1,619	37,932
1992	3	-4,043	-2,508	-6,580	-0,138	-12,718	-1,401	-1,932	0,984	-2,111	-1,788	0,440	1,253	-1,027	-0,779	-2,338	-3,735	-7,858	113,201
1992	4	-7,664	-1,281	-0,155	2,680	8,006	-4,016	-5,157	5,683	-3,358	-0,107	-3,670	-7,199	-5,700	-2,787	-3,245	-5,936	-2,343	43,085
1993	1	-7,496	-5,523	-3,914	2,234	-12,619	-0,114	-3,872	-6,736	-1,310	-4,842	-4,928	-3,829	-4,969	-0,824	-13,514	-6,415	-1,928	16,628
1993	2	-2,141	-5,270	0,352	3,027	4,139	-10,573	-1,915	-4,944	-0,726	-2,146	12,904	-1,175	0,181	-2,827	2,428	1,258	3,650	112,039
1993	3	1,351	2,417	-12,018	0,230	-0,113	2,518	2,650	-0,455	-3,324	1,182	-4,103	-1,576	0,626	-1,446	2,500	-2,754	3,135	66,695
1993	4	-2,176	4,269	-3,386	-6,398	-4,797	-4,134	2,418	-0,336	-5,591	1,527	-3,328	-4,965	-1,800	2,201	-1,076	0,777	-6,304	10,499
1994	1	0,852	3,421	4,587	-1,716	9,810	-3,729	-3,900	2,631	0,408	-1,181	1,054	-3,569	-3,766	3,508	1,450	-4,898	-1,241	20,683
1994	2	1,259	1,303	-0,345	-3,796	-13,792	1,819	0,393	-5,510	-0,353	2,674	-4,206	1,234	-0,333	2,774	-2,680	0,170	7,583	23,629
1994	3	-0,874	-2,331	-0,412	5,616	2,943	0,132	0,777	0,077	-1,280	4,560	-5,515	2,467	-2,122	-5,537	1,652	2,398	-0,038	63,332
1994	4	0,496	4,197	-1,374	-2,318	0,912	13,842	-0,218	-0,772	0,123	3,217	-4,071	-4,506	0,191	3,475	-0,722	1,174	-2,903	54,463
1995	1	-3,601	-3,208	4,967	-6,840	13,623	2,147	-4,772	-4,248	2,887	-5,207	8,887	3,171	-4,449	-3,546	-0,109	-1,797	6,524	33,269
1995	2	0,195	2,203	7,930	-2,397	18,117	-1,278	-3,245	-3,544	0,529	2,188	-3,934	3,741	1,163	4,490	6,829	0,482	4,448	31,017
1995	3	1,794	-1,427	-8,206	-1,934	-5,345	1,468	2,315	-6,074	0,887	0,410	7,998	2,912	6,381	-9,264	0,170	-2,901	0,663	49,270
1995	4	2,166	-2,652	5,567	4,508	7,568	-2,240	4,569	4,897	-1,768	-2,787	4,146	-2,725	-0,017	5,216	-9,309	-0,405	-2,982	18,715
1996	1	-1,864	1,955	3,317	9,467	-5,867	-4,642	-2,573	-3,130	-1,449	-4,527	6,022	-0,621	3,297	-5,365	-0,131	-2,374	2,252	16,997
1996	2	0,983	1,561	-0,473	1,478	4,897	-3,651	4,406	-1,268	-1,408	1,845	-0,642	0,508	-0,837	2,464	4,585	-0,358	5,942	43,953
1996	3	0,618	2,159	4,192	7,929	3,998	-4,294	0,491	0,547	2,771	0,756	-7,074	5,507	3,875	-2,977	-3,755	7,867	-1,254	16,774
1996	4	1,608	-2,143	-5,778	-2,374	-5,606	7,258	1,940	-0,555	1,816	-2,683	-4,519	0,308	-0,886	-0,624	5,600	-0,950	-4,210	180,002
1997	1	-4,383	3,104	-0,379	-1,638	-4,437	-2,399	-7,134	3,106	-0,008	-3,879	5,242	0,990	-1,928	-1,552	2,899	2,615	-2,616	12,324
1997	2	-1,610	3,566	-4,514	-4,372	2,357	2,175	-2,014	-0,878	3,661	0,541	7,160	4,632	1,382	5,811	6,597	2,123	1,003	50,921
1997	3	6,286	4,704	-1,289	-7,483	3,303	7,187	3,942	6,570	2,673	6,683	-3,551	-0,989	-0,496	4,605	2,659	3,132	2,270	29,284
1997	4	6,000	-0,521	0,371	-3,941	-1,846	0,518	1,977	-0,113	1,266	2,656	5,422	-0,653	2,470	5,987	-0,738	-0,650	-1,387	36,088
1998	1	-2,622	-3,342	-2,378	-1,317	-8,282	1,918	2,863	0,698	-3,219	2,732	-8,667	2,260	-2,168	1,193	-2,720	-3,240	-1,864	17,915
1998	2	6,166	6,495	0,889	1,769	5,778	-2,051	-0,114	2,723	4,805	1,083	0,487	2,750	2,438	-1,000	1,496	3,541	6,843	44,605
1998	3	0,794	-3,466	8,434	-4,604	-1,915	3,011	4,237	9,113	1,252	1,684	-2,421	4,377	1,655	-0,528	1,890	-0,966	-3,924	29,051
1998	4	5,226	-1,536	-3,791	7,670	4,605	6,208	3,440	2,447	-1,508	-2,141	8,898	0,377	0,654	0,727	-3,003	-1,206	-5,550	44,726
1999	1	-0,937	-1,311	0,907	14,812	-2,789	0,472	-2,672	-2,016	-1,283	1,217	7,715	-1,934	-1,590	-0,045	-0,324	3,868	7,797	26,161
1999	2	2,768	2,610	-0,217	-4,085	-0,047	1,383	-0,102	5,166	0,147	-3,093	4,231	3,445	3,297	-0,271	-0,667	2,590	9,050	10,543
1999	3	1,559	2,288	6,647	4,609	-7,045	-1,390	0,544	3,697	0,660	-0,973	4,233	2,165	0,170	-3,034	1,506	2,180	-1,226	51,920
1999	4	-4,019	-0,125	-1,980	-7,885	13,218	4,697	0,975	2,601	-2,064	4,070	-2,114	-3,208	1,256	12,593	5,596	1,962	-2,482	32,538
2000	1	2,644	7,353	3,922	-0,224	-0,563	0,773	1,155	-2,527	0,649	-0,039	2,415	2,469	-3,631	-1,341	-1,863	-0,485	3,264	47,821
2000	2	0,522	3,541	2,087	-5,355	4,918	-6,184	4,797	-3,673	1,618	1,703	0,388	1,144	2,788	1,892	5,384	1,476	5,942	45,204
VARIANZA		**8,809**	**9,592**	**16,959**	**28,563**	**52,765**	**17,222**	**9,315**	**11,976**	**4,419**	**8,309**	**40,388**	**8,812**	**9,794**	**19,322**	**21,205**	**7,560**	**18,091**	**41,988**
PESOS		0,210	0,228	0,404	0,680	1,257	0,410	0,222	0,285	0,105	0,198	0,962	0,210	0,233	0,460	0,505	0,180	0,431	

Cuadro 1. Estimadores directos

Tasas de variación. Estimadores SINTÉTICOS por Comunidad. Sector Industrial

AÑO	TRIM	AND	ARA	AST	BAL	CAN	CANT	CAS-L	CAS-M	CAT	VAL	EXT	GAL	MAD	MUR	NAV	PVAS	RIO
1987	3	0,504	0,525	0,521	0,272	0,516	0,788	0,331	0,682	0,879	0,358	0,827	0,503	0,661	0,734	0,802	0,877	0,528
1987	4	0,770	2,228	1,251	-0,034	-0,418	1,143	1,202	0,745	1,458	0,959	-0,084	0,876	1,370	1,244	1,840	2,284	1,067
1988	1	-0,979	-1,814	-1,876	0,186	-0,917	-1,751	-1,686	-0,307	-1,647	-0,008	-0,738	-0,860	-2,346	-1,079	-1,972	-2,191	-0,758
1988	2	2,195	1,790	3,614	1,987	2,896	2,500	2,588	1,577	2,073	1,356	2,358	2,045	2,876	2,100	2,115	2,453	1,548
1988	3	0,356	-0,032	0,616	-0,364	0,965	0,380	-0,052	-0,858	-1,121	-1,104	0,612	-0,209	-1,069	0,408	0,352	0,343	-0,361
1988	4	0,302	0,973	-0,265	0,724	0,441	0,409	-0,012	0,043	0,893	0,764	-0,527	0,172	1,941	-0,192	1,002	1,280	0,121
1989	1	1,274	0,326	-0,031	1,062	1,615	1,552	1,121	1,661	1,682	0,887	1,679	1,126	1,490	1,552	0,493	0,752	1,015
1989	2	1,472	1,380	1,319	1,503	1,772	1,195	1,853	0,998	1,243	1,079	1,235	1,920	1,727	1,139	1,459	0,775	1,045
1989	3	0,567	1,244	1,557	0,710	0,269	0,760	0,254	0,821	1,126	1,099	0,356	0,321	1,845	0,464	0,980	1,368	0,523
1989	4	0,276	0,957	-0,427	0,361	-0,028	0,961	0,204	0,754	1,321	1,069	-0,149	0,262	0,971	0,407	1,116	2,240	0,736
1990	1	1,888	2,005	0,641	1,949	2,105	1,459	2,046	1,234	1,367	1,414	1,750	2,105	1,327	1,832	1,967	1,321	1,919
1990	2	0,236	0,313	0,076	-0,530	0,168	0,580	0,569	0,436	0,710	-0,143	0,413	0,163	0,460	0,071	0,328	0,065	0,266
1990	3	0,132	-0,233	0,424	0,262	0,851	-0,097	0,216	-0,884	-0,479	-0,728	-0,170	-0,389	0,258	-0,285	-0,220	-0,076	-0,399
1990	4	-0,946	-0,365	-1,315	-0,601	-1,071	-0,901	-0,973	-0,951	0,001	-0,384	-1,519	-1,282	0,660	-0,559	-0,342	0,156	-0,471
1991	1	-2,442	-2,954	-2,091	-1,347	-2,789	-2,729	-2,899	-1,016	-2,205	-0,959	-1,709	-2,205	-2,920	-2,125	-3,206	-3,029	-1,754
1991	2	0,038	0,154	0,463	-0,724	0,219	0,198	0,320	-0,750	-0,191	-0,959	-0,201	-0,136	0,392	-0,312	0,267	-0,142	-0,443
1991	3	0,152	-0,229	-0,417	0,436	0,879	-0,054	-0,125	-0,653	-0,342	-0,246	-0,121	0,024	0,725	0,063	0,159	-0,030	-0,368
1991	4	-0,371	0,632	0,633	-1,212	-0,809	0,148	0,242	-1,302	-0,325	-0,632	-1,604	-0,395	0,241	-0,481	0,951	1,322	-0,431
1992	1	-1,154	-0,706	-1,172	-0,489	-1,417	-1,143	-1,149	-0,714	-0,904	0,152	-1,467	-0,581	-1,051	-1,277	-0,414	-0,457	-0,537
1992	2	0,683	0,319	-0,012	0,445	1,428	0,568	0,648	-0,094	-0,129	-0,079	1,007	0,482	-0,122	0,606	0,829	0,293	0,665
1992	3	-1,999	-1,562	-1,655	-2,709	-2,065	-1,875	-1,865	-2,764	-2,373	-2,688	-2,323	-2,390	-2,211	-2,015	-1,485	-1,531	-2,057
1992	4	-3,117	-4,079	-3,448	-3,075	-2,684	-2,942	-3,510	-2,629	-3,108	-3,355	-2,163	-3,541	-3,340	-2,842	-4,063	-3,757	-3,157
1993	1	-3,847	-3,489	-4,127	-3,250	-4,233	-5,059	-3,574	-3,959	-3,986	-3,928	-4,135	-3,048	-3,745	-3,390	-4,295	-5,187	-4,322
1993	2	-0,053	-0,962	-3,058	-0,539	0,661	-0,657	-0,434	-0,994	-0,804	-1,324	-0,159	-0,885	-0,963	0,265	-0,655	-1,381	-0,134
1993	3	-0,093	-1,436	-1,030	0,080	0,770	-0,833	-0,586	-0,530	-1,188	-0,856	0,583	-0,768	-1,019	0,024	-1,301	-2,124	-0,230
1993	4	-1,979	-1,455	-2,047	-2,304	-1,623	-1,440	-1,762	-2,740	-1,708	-2,838	-2,411	-1,778	-0,886	-2,465	-1,460	-0,988	-2,895
1994	1	-0,891	-1,171	-0,224	-1,026	-1,617	-0,953	-0,940	0,028	-0,586	-0,391	-0,925	-1,039	-1,286	-0,738	-1,614	-1,308	-0,809
1994	2	-0,056	0,347	0,491	0,463	-0,251	-0,265	0,120	0,024	-0,145	0,395	0,000	0,221	-0,424	-0,131	0,131	0,122	0,365
1994	3	0,508	0,150	-0,861	0,623	0,712	-0,160	0,126	0,430	0,146	0,390	0,766	-0,128	0,189	1,115	-0,072	-0,479	0,839
1994	4	-0,210	0,511	-0,158	0,965	-0,923	0,031	-0,567	1,030	1,058	1,502	-0,480	0,349	0,738	-0,351	0,103	1,061	0,216
1995	1	-1,153	-1,137	-0,434	-1,400	-1,168	-0,696	-0,720	-1,406	-1,240	-1,384	-1,257	-0,973	-1,493	-1,389	-0,881	-0,539	-1,104
1995	2	1,235	1,466	3,551	1,071	1,163	1,303	1,698	1,252	1,089	1,187	1,368	1,446	1,410	1,242	1,311	1,497	0,817
1995	3	1,812	-0,194	2,201	1,761	4,068	1,164	1,265	-0,198	-0,331	-0,375	2,249	0,262	0,925	1,181	0,609	0,125	0,612
1995	4	-0,190	-0,322	1,736	-0,539	-0,599	0,025	0,186	-0,219	-0,404	-0,739	0,292	0,238	-0,975	0,325	-0,236	-0,078	-0,307
1996	1	-1,150	-0,867	-0,523	-1,512	-0,804	-0,895	-0,671	-2,097	-1,291	-1,906	-1,719	-0,903	-0,565	-1,395	-0,526	-0,400	-1,849
1996	2	0,382	0,862	-0,168	-0,028	-0,243	0,046	0,346	0,832	0,762	0,572	0,263	0,183	0,975	0,810	0,117	-0,250	0,423
1996	3	1,800	3,314	1,573	1,106	0,628	2,036	1,967	2,236	2,773	2,109	1,446	2,340	2,294	1,978	2,963	2,564	2,144
1996	4	-0,288	-0,473	-0,089	-0,006	-0,413	-0,079	-0,137	0,245	0,118	0,418	-0,127	-0,228	-0,344	-0,056	-0,192	0,117	0,417
1997	1	-1,692	-0,146	-2,308	-1,150	-1,943	-1,444	-1,479	-2,047	-0,929	-1,177	-2,382	-0,789	0,233	-1,638	-0,395	-0,117	-1,791
1997	2	1,522	1,751	0,186	1,151	0,961	1,572	1,339	2,005	2,157	1,773	1,288	1,302	1,624	1,904	1,575	1,636	1,915
1997	3	2,972	3,351	2,504	2,993	3,369	3,878	3,026	2,534	3,113	3,075	2,973	2,934	3,060	3,018	4,077	5,029	3,517
1997	4	1,947	1,285	3,675	1,545	2,465	2,415	2,227	1,674	1,437	1,385	2,088	1,759	1,505	1,289	1,659	2,040	1,324
1998	1	-1,429	-0,960	-1,631	-1,845	-1,560	-1,257	-1,395	-1,950	-1,413	-1,655	-2,088	-1,845	-0,807	-1,490	-1,071	-0,806	-1,527
1998	2	3,258	3,062	0,933	3,593	3,863	3,132	2,677	2,722	3,415	3,061	3,191	2,800	4,192	3,647	3,436	3,291	3,520
1998	3	1,409	1,561	1,533	1,355	1,736	1,974	1,396	1,274	1,605	1,442	1,330	1,409	1,612	1,051	2,020	2,533	1,356
1998	4	-0,085	0,683	1,431	-0,070	-0,443	0,045	0,228	0,075	0,129	0,064	-0,386	1,021	0,597	-0,812	0,362	0,078	-1,060
1999	1	-0,127	-1,058	-0,106	0,346	0,188	0,010	-0,523	0,710	-0,038	0,435	0,704	-0,288	-0,469	0,002	-0,814	-0,542	0,192
1999	2	1,100	1,338	2,224	1,305	0,863	0,647	1,032	0,543	0,689	0,733	0,602	0,876	1,408	1,082	0,742	0,908	0,537
1999	3	0,471	1,405	0,915	-0,079	-0,263	0,989	0,960	0,774	0,948	0,929	0,053	1,181	0,447	0,508	1,524	1,665	0,642
1999	4	0,411	0,238	0,452	0,049	0,142	0,634	0,280	0,593	0,154	0,545	0,258	0,607	-0,365	0,440	0,546	0,631	0,275
2000	1	0,677	0,461	-0,452	1,174	1,191	0,404	0,691	0,078	0,429	0,326	0,538	0,661	0,735	0,628	0,485	0,482	0,627
2000	2	2,200	1,428	1,533	1,990	3,009	2,053	2,295	1,335	1,348	1,414	2,248	1,658	1,755	2,386	1,824	1,748	2,082

Cuadro 2. Estimadores indirectos

Tasas variación España, por RAMA (en porcentage)

	RAMA 1	RAMA 2	RAMA 3	RAMA 4	RAMA 5	RAMA 6	RAMA 7	RAMA 8	RAMA 9	RAMA 10	RAMA 11	RAMA 12	RAMA 13	RAMA 14
87-3	-1,10	-3,85	1,61	0,49	0,84	3,54	5,29	-5,04	-1,20	1,16	6,34	-2,86	-0,08	-2,30
87-4	3,77	-5,81	-0,60	-0,12	-3,35	-3,09	2,27	4,99	-1,66	0,43	5,75	5,43	4,54	6,81
88-1	-5,23	7,41	-2,89	3,14	-0,22	-4,47	-4,17	-4,86	2,98	-0,45	-3,78	-8,81	-2,03	1,33
88-2	8,11	1,40	2,66	-2,47	7,93	7,95	5,06	2,29	1,07	1,81	2,79	2,45	-0,37	1,49
88-3	-0,52	4,03	4,32	-5,41	-3,32	-7,43	-3,28	-9,57	0,87	2,06	7,02	-0,73	-0,65	0,51
88-4	-4,82	3,39	-3,53	-0,22	-3,23	7,03	-0,83	-1,58	4,75	1,48	2,01	4,08	1,68	0,62
89-1	-3,96	-1,05	3,13	-0,26	7,63	2,32	12,68	2,63	-0,35	-0,48	-2,42	1,86	-2,50	1,17
89-2	2,41	3,43	1,03	0,15	5,99	4,86	1,05	2,79	1,24	-0,73	-3,13	-0,32	4,76	0,94
89-3	3,02	-3,43	-1,43	1,36	-2,83	5,01	-0,49	-6,29	1,23	3,04	2,90	6,84	-0,51	1,80
89-4	-5,97	0,83	-3,05	0,49	0,19	0,06	4,67	6,37	4,73	0,94	9,10	1,13	-1,19	0,21
90-1	-1,39	8,46	3,30	1,55	1,26	-1,24	-3,35	5,98	0,22	-0,83	0,90	0,59	5,35	1,66
90-2	1,41	-3,68	2,09	0,84	-2,01	-0,36	4,38	4,52	1,98	-2,10	0,06	4,83	-0,15	-6,42
90-3	3,00	11,14	0,22	-2,80	-4,49	0,54	-2,85	3,93	1,40	-1,14	-0,95	1,00	-0,54	-0,67
90-4	-3,03	-4,96	-3,03	-1,09	-3,47	8,63	0,00	2,84	-1,59	-2,15	3,90	-1,49	-1,46	3,45
91-1	-1,97	-5,43	-3,89	3,50	2,12	-1,20	-3,24	-6,25	-3,89	0,13	-5,83	-6,42	-5,89	-0,27
91-2	2,59	-1,46	2,04	-2,53	-4,56	0,55	-0,25	-0,69	-1,46	0,70	-3,80	4,30	2,45	-3,07
91-3	-4,60	-0,11	0,08	-4,05	4,02	6,45	-1,74	-5,53	0,11	1,14	-2,42	-1,37	0,40	4,42
91-4	1,69	-5,95	-2,37	-4,86	-5,53	-0,26	-2,65	4,91	0,97	2,92	0,20	1,16	4,51	3,69
92-1	-5,09	-5,11	-4,81	1,34	1,77	2,70	-6,43	-0,70	2,20	2,24	-1,30	-6,08	1,31	0,16
92-2	-3,12	1,67	3,99	-1,55	-0,33	-0,10	-5,07	0,10	2,15	0,05	3,11	-1,59	0,57	-3,03
92-3	-0,39	-2,65	-0,26	-4,30	-9,80	-5,31	-5,70	-1,40	-1,94	-1,39	0,96	0,56	-0,34	-2,48
92-4	-4,66	-5,05	0,03	-2,97	0,24	-3,96	1,24	-5,79	-5,51	-2,76	-5,52	0,19	-9,42	-4,80
93-1	2,31	3,27	-5,76	-3,29	5,18	-1,22	-1,68	-6,36	-8,63	-13,76	0,20	-11,70	0,84	4,81
93-2	-9,81	4,62	4,94	-2,99	-7,45	-2,14	1,64	3,00	1,47	-6,18	1,81	-4,81	-0,21	0,17
93-3	-2,06	1,14	4,79	-0,65	-2,72	0,86	-3,91	-4,70	-1,38	-1,85	-5,40	-2,49	-3,13	-0,39
93-4	-3,15	5,87	-3,33	-6,36	1,53	-2,68	2,07	-1,41	-2,36	-2,33	-0,14	5,03	-0,60	-4,05
94-1	4,34	-2,24	-2,17	1,83	-3,76	-4,63	6,02	0,12	0,78	-1,64	-2,79	-1,79	-2,47	0,98
94-2	2,97	5,19	-1,20	2,82	-1,56	-1,91	-6,63	3,45	-1,49	0,50	1,00	-1,68	1,46	-1,03
94-3	-2,16	-1,15	3,51	0,50	-1,72	0,75	-2,76	0,46	0,59	-4,70	3,27	0,89	-2,49	5,09
94-4	-2,80	5,68	-6,55	5,40	1,48	1,62	3,41	-0,69	-0,20	3,13	1,99	-1,10	0,22	0,44
95-1	0,46	0,44	-1,29	-1,91	-1,99	-4,35	-1,73	4,84	-2,82	2,59	-4,92	-1,34	0,31	-3,78
95-2	12,39	-1,75	-0,25	-0,75	7,98	2,62	0,96	-0,88	3,17	-0,59	5,82	2,03	-0,40	3,02
95-3	3,76	13,24	6,78	-5,51	-6,39	6,49	-4,35	-5,76	8,17	0,23	3,22	-3,39	-3,94	-3,43
95-4	8,03	-4,22	1,70	-0,80	3,35	-1,99	3,97	-4,71	-11,06	3,30	-0,39	-7,56	1,68	1,26
96-1	0,72	0,82	-2,59	-6,39	3,11	1,85	0,72	-0,12	-0,34	-1,85	1,56	-2,81	1,42	-0,23
96-2	2,62	-6,41	1,21	1,07	0,50	-0,94	1,90	-0,12	0,95	-5,46	0,71	9,18	-0,63	5,84
96-3	1,62	-4,57	1,40	4,94	-2,88	0,12	4,65	-1,11	-0,81	0,86	8,22	3,89	7,69	-0,43
96-4	-1,03	-5,47	-0,89	1,76	0,64	2,72	-0,15	5,01	-0,82	1,94	-1,65	-4,42	-1,52	-0,59
97-1	-6,81	-1,69	-5,38	-3,57	8,82	2,65	-5,64	-1,79	-7,34	-0,67	-1,44	5,88	2,83	5,25
97-2	-2,97	-1,96	2,01	3,12	-3,17	-0,34	7,15	4,00	2,07	-0,67	3,95	2,07	1,75	2,67
97-3	-5,61	2,00	3,20	0,51	4,00	1,52	-2,27	8,98	2,18	9,29	5,90	5,29	1,89	1,20
97-4	7,84	4,53	1,56	-0,74	2,22	-0,11	3,60	2,89	7,39	4,65	-0,50	2,75	-0,56	-3,56
98-1	-2,51	-2,58	-1,64	-3,95	-6,62	-3,00	-3,55	1,46	-0,33	-0,68	-3,16	5,90	-0,77	2,00
98-2	-9,13	-0,12	5,67	1,49	2,08	10,70	1,50	0,72	-1,26	4,15	1,55	3,86	2,73	6,59
98-3	-1,84	3,37	0,05	0,15	1,32	2,02	1,85	2,24	5,78	3,93	6,03	1,34	0,60	-3,30
98-4	6,05	2,80	-4,46	-0,18	8,75	0,46	2,91	-8,37	1,59	0,17	-0,43	2,16	4,46	-1,85
99-1	-2,31	-4,08	0,71	2,02	3,70	1,97	1,84	-2,83	0,44	2,50	-2,88	-1,94	-5,44	-2,19
99-2	8,48	8,16	-0,18	-0,14	-3,04	-0,10	-0,55	-2,57	-2,99	1,02	-0,72	2,70	2,76	7,52
99-3	0,64	-5,79	-1,85	0,17	4,54	-2,38	2,58	2,87	2,89	1,22	4,71	-0,06	4,00	0,85
99-4	-2,17	-3,60	-0,29	-0,75	1,96	-3,25	3,81	-3,79	3,37	3,10	0,79	-3,80	1,46	2,01
00-1	-3,26	9,87	0,91	-0,35	1,52	1,05	-1,90	6,03	-1,93	-0,71	-1,52	-0,49	1,44	1,75
00-2	-0,27	1,53	5,05	-2,56	3,90	2,02	-1,54	7,11	3,38	0,64	-0,43	2,57	-0,47	5,42

Cuadro 3. Estimador directo por ramas de actividad industrial. Ámbito estatal

Ponderaciones Industria 97

	AND	ARA	AST	BAL	CAN	CANT	CAS-L	CAS-M	CAT	VAL	EXT	GAL	MAD	MUR	NAV	PVAS	RIO
Rama 1	2,83%	2,51%	21,15%	0,93%	1,54%	2,35%	2,31%	6,36%	0,50%	0,66%	3,60%	2,80%	1,25%	1,84%	0,77%	0,31%	0,89%
Rama 2	4,74%	1,93%	2,42%	9,01%	8,79%	2,65%	1,45%	2,85%	1,98%	1,79%	4,68%	3,65%	4,69%	2,40%	0,92%	1,26%	2,21%
Rama 3	24,59%	11,90%	13,75%	18,63%	33,19%	20,29%	17,35%	22,58%	12,41%	10,50%	33,09%	18,02%	10,08%	27,86%	15,82%	23,58%	6,48%
Rama 4	10,00%	11,22%	3,32%	18,94%	2,20%	5,29%	26,41%	5,70%	16,89%	25,83%	15,83%	13,09%	6,64%	12,73%	4,76%	19,81%	1,88%
Rama 5	3,55%	2,80%	4,38%	5,59%	5,27%	4,41%	5,38%	5,36%	2,79%	4,83%	7,19%	9,13%	1,64%	3,39%	3,69%	2,52%	2,87%
Rama 6	5,26%	4,55%	4,68%	9,01%	9,89%	4,12%	2,39%	4,77%	8,30%	5,01%	3,96%	3,10%	16,82%	4,10%	6,14%	4,72%	5,35%
Rama 7	3,79%	1,84%	1,96%	0,31%	1,54%	6,47%	6,75%	3,31%	8,59%	1,95%	2,52%	3,16%	6,55%	4,24%	1,38%	0,94%	4,56%
Rama 8	2,23%	3,19%	1,21%	1,55%	1,98%	4,71%	1,79%	6,23%	5,00%	3,11%	0,72%	2,01%	2,89%	3,39%	4,61%	6,29%	8,13%
Rama 9	8,81%	4,55%	7,85%	7,14%	10,99%	7,35%	8,89%	7,75%	4,04%	9,95%	6,83%	7,06%	3,95%	4,95%	6,61%	6,60%	3,62%
Rama 10	11,24%	10,54%	23,11%	12,11%	12,09%	18,82%	10,68%	10,33%	10,81%	13,51%	11,51%	10,59%	9,56%	9,76%	15,82%	12,58%	25,74%
Rama 11	3,39%	9,86%	5,29%	2,48%	1,76%	6,18%	3,25%	3,44%	6,78%	4,27%	3,60%	3,47%	5,17%	6,08%	10,29%	5,66%	14,61%
Rama 12	3,19%	9,19%	1,96%	1,55%	1,98%	5,59%	3,16%	2,98%	7,58%	3,01%	1,80%	2,68%	12,84%	1,70%	5,99%	1,57%	7,61%
Rama 13	8,85%	19,15%	5,14%	2,48%	3,96%	8,24%	2,56%	13,71%	8,90%	4,88%	0,72%	16,80%	10,78%	5,37%	19,20%	5,66%	9,86%
Rama 14	7,53%	6,77%	3,78%	10,25%	4,84%	3,53%	7,61%	4,64%	5,44%	10,69%	3,96%	4,44%	7,16%	12,16%	3,99%	8,49%	6,20%
	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Cuadro 4. Ponderaciones de las ramas de actividad industrial por Comunidad

B. EQUIVALENCIAS EN LA CLASIFICACIÓN DE LAS RAMAS INDUSTRIALES

Clasificación CRE	Clasificación CNAE 93
Extracción de productos energéticos, otros minerales y refino de petróleo	B31+B35+B5D
Energía eléctrica, gas y agua	B70
Alimentación, bebidas y tabaco	B41
Textil, confección, cuero y calzado	B44+B47
Madera y corcho	B49
Papel; edición y artes gráficas	B4B
Industria química	B5F
Caucho y plástico	B5H
Otros productos minerales no metálicos	B5J
Metalurgia y productos metálicos	B5L
Maquinaria y equipo mecánico	B5O
Equipo eléctrico, electrónico y óptico	B5Q
Fabricación de material de transporte	B60
Industrias manufactureras diversas	B63

donde, en la clasificación CNAE:

B31: Extracción de productos energéticos
B35: Extracción de otros minerales excepto productos energéticos
B41: Alimentación, Bebidas y Tabaco
B44: Textiles, Confeccion
B47: Cuero y calzado
B49: Madera, Corcho
B4B: Papel, Artes Gráficas y Edicion
B5D: Coquerias, Refino de petroleo
B5F: Industrias Quimicas
B5H: Caucho y Materias Plasticas
B5J: Otros productos minerales no metálicos
B5L: Metalurgia y fabricación de productos metálicos
B5O: Maquinaria y equipo mecánico
B5Q: Material y equipo electrico, electrónico
B60: Material de transporte
B63: Otras Industrias Manufactureras
B70: Energia electrica, gas y agua

C. DERIVACIÓN DE LA FÓRMULA DEL ERROR MUESTRAL

En este apéndice relacionaremos el error de muestreo relativo proporcionado por el INE para la encuesta de la EPA y la varianza muestral $\sigma_2^2(k)$.

De lo reseñado anteriormente en la Sección 4, obtenemos

$$\widehat{\theta}_2(k,t) = 100 \times \frac{o(k,t) - o(k,t-1)}{o(k,t-1)},$$

donde $o(k,t)$ denota el número de ocupados en la muestra de la comunidad k en el instante t (para el sector industrial concreto). La información que obtenemos de los documentos del INE sobre la EPA es la relativa al coeficiente de variación de la ocupación, es decir una estimación del valor C.V. $= \sigma_o/\mu_0$, donde μ_0 y σ_o son respectivamente la media y la desviación estándar de la variable ocupación para la comunidad k en el momento t. Introduciendo la proporción muestral $\widehat{p}(k,t) = o(k,t)/N$, con un tamaño muestral N que suponemos constante, obtenemos

$$\widehat{\theta}_2(k,t) = 100 \times \frac{\widehat{p}(k,t) - \widehat{p}(k,t-1)}{\widehat{p}(k,t-1)}$$

cuya varianza $\sigma_2^2(k,t)$, inducida por la varianza muestral de $\widehat{p}(k,t)$, es necesario determinar.

Dado que $\widehat{p}(k,t)$ es una proporción muestral, obtenemos *bajo muestreo aleatorio simple* que $var(\widehat{p}(k,t)) = p(k,t)(1- p(k,t))/N$, donde $p(k,t)$ indica la proporción poblacional. Dado que $\widehat{p}(k,t)$ converge en probabilidad a $p(k,t)$ cuando el tamaño muestral N tiende a infinito, utilizando el método-δ (Rao, 1973) obtenemos

$$\begin{aligned}
avar(\widehat{\theta}_2(k,t)) &= \left(\frac{1}{p(k,t)}, -\frac{p(k,t)}{p(k,t-1)^2}\right)\left[var\begin{pmatrix}\widehat{p}(k,t)\\ \widehat{p}(k,t-1)\end{pmatrix}\right]\left(\frac{1}{p(k,t)}, -\frac{p(k,t)}{p(k,t-1)^2}\right)' \\
&= \frac{1}{N}\left(\frac{1}{p(k,t)}, -\frac{p(k,t)}{p(k,t-1)^2}\right)\begin{pmatrix} p(k,t)(1-p(k,t)) & 0 \\ 0 & p(k,t-1)(1-p(k,t-1))\end{pmatrix} \\
&\times \left(\frac{1}{p(k,t)}, -\frac{p(k,t)}{p(k,t-1)^2}\right)' = \frac{2}{N}\frac{1-p(k)}{p(k)}
\end{aligned}$$

para todo t, donde *avar* denota varianza asintótica. En el cálculo de dicha expresión, hemos sustituido $p(k,t)$ por su valor promedio en cada Comunidad, pongamos $p(k)$ (en este punto utilizamos el supuesto de que $p(k,t)$ es estacionario en media) y se ha considerado que las muestras en dos periodos consecutivos son independientes. En el caso que este último supuesto no sea correcto, entonces la expresión anterior se modifica para obtener:

$$avar(\widehat{\theta}_2(k,t)) = E.D. \times \frac{2}{N} \times \frac{1-p(k)}{p(k)} \times (1-\rho(k))$$

donde $\rho(k)$ es el coeficiente de la correlación muestral en dos periodos sucesivos. La EPA utiliza un diseño de muestra complejo, por lo tanto la expresión derivada bajo el supuesto de muestreo aleatorio simple debe ser modificada multiplicando por el efecto de diseño E.D.

Por otra parte, en el caso de muestreo complejo con efecto de diseño E.D., obtenemos que

$$\begin{aligned}
\text{C.V.}\,(o(k,t)) \;&=\; \widehat{\sigma}_o/\widehat{\mu}_0 \;=\sqrt{\frac{var(\widehat{p}(k,t))}{\widehat{p}(k,t)^2}} = \sqrt{\text{E.D.}} \times \sqrt{\frac{\widehat{p}(k,t)(1-\widehat{p}(k,t))}{N\widehat{p}(k,t)^2}} \\
&=\; \sqrt{\text{E.D.}} \times \sqrt{\frac{(1-\widehat{p}(k,t))}{N\widehat{p}(k,t)}}
\end{aligned}$$

de modo que al comparar las expresiones de la varianza muestral de $\widehat{\theta}_2(k,t)$ y el coeficiente de variación anterior, obtenemos la fórmula simple:

$$avar\,(\widehat{\theta}_2(k,t)) = 2 \times \text{C.V.}^2(k,t) \times (1-\rho(k))$$

ENGLISH SUMMARY

COMPOSITE ESTIMATORS IN REGIONAL STATISTICS: AN APPLICATION TO THE ESTIMATION OF THE RATE OF CHANGE IN INDUSTRIAL OCCUPATION*

À. COSTA[1]
A. SATORRA[2]
E. VENTURA[2]

This work is part of a project that studies the application of small area composite estimators (a combination of direct and indirect estimators) in regional statistics. We compare three estimators: a direct one based on sample data for each of the Spanish Autonomous Communities, a synthetic (indirect) one that combines Estate wide information with the Community specific information, and a third estimator, the composite one, which is based on a model and results in a linear combination of the two previous estimators. We show the adopted method of analysis by estimating the industrial occupation rate of change in the different Spanish Autonomous Communities.

Keywords: Regional statistics, small areas, composite estimators

AMS Classification (MSC 2000): 62J07, 62J10, 62H12

* This article is a result from a joint research project between the Institut d'Estadística de Catalunya (IDESCAT) and the Instituto Nacional de Estadística (INE). The INE and the IDESCAT do not necessarily share the opinions expressed in this study, which belong exclussively to the authors. We thank the comments of N. T. Longford to a preliminary version of this article.

[1] Institut d'Estadística de Catalunya. Via Laietana 58, 08003 Barcelona.
[2] Departament d'Economia i Empresa. Universitat Pompeu Fabra. Ramon Trias Fargas 25-27, 08005 Barcelona.

– Received December 2001.
– Accepted March 2002.

Most of the surveys that are carried out by the national statistics offices have been designed to yield estimators with small variance for national wide figures. The size and design of the sample may still allow us to get precise estimators in cases in which the areas under consideration are smaller, such as regions or provinces. But in some cases the size of the sample is just too small for that purpose. It is even worst when one wants to estimate the rate of variation instead of the level of a variable, since rates have larger variances. This is quite a frequent situation in economics statistics.

There are two alternatives which allow us to deal with the issue of providing valid statistical information for the smallest areas. The first one requires to repeat the survey with a new and more suitable sample size. The second one is to use statistical theory on small areas estimation to improve the quality of the estimators obtained from the original survey.

In this article we take the second alternative. If part of the national sample has been extracted from a particular small area, we can obtain a direct estimation of the variable of interest for that area. Such estimator is unbiased but often lacks precision. On the other hand the survey allows us to obtain an estimator which is unbiased and precise at the national level. We can use it to create an indirect estimator for the small area which will exhibit a smaller variance than the direct estimator but will usually be biased.

Here we propose to use a composite estimator. First we specify a model of variance behavior of the direct and synthetic estimators, and then we consider a linear combination of both estimators as our composite estimator. The weights of the linear combination are chosen so as to minimize the mean square error of the composite estimator. In this sense, the composite estimator represents an improvement relative to alternative estimators, either direct or synthetic (indirect) for small areas.

The method proposed is applied to data of the Encuesta de Población Activa, a labor survey conducted at the national Spanish level. The quantity to be estimated is the quarterly variation of the occupation rate in the industrial sector in each of the 17 Autonomous Communities of Spain. This rate of variation has multiple components as it depends on the geographical area and the time period being considered.

Our composite estimator assigns automatically larger weight to the direct estimator in those Communities where the variance of the sampling error is smaller than the bias of the indirect estimator, and larger weight to the synthetic estimator when bias is small relative to the variance of the sampling error. We observe that for most of the Communities, the composite estimator gives more weight to the synthetic estimator, given that the size of the sample in the small area is quite insufficient and therefore the sampling error has a large variance. But there are also some Communities in which the composite estimator assigns higher weight to the synthetic estimator. This is the case of Catalonia, whose industrial web is very similar to the national wide one, and the difference between the synthetic and the direct estimators is very small in this Community.

From this study it is realized that the weights determining the composite estimator depend on the bias, the population and sample variances, the type of variable being estimated (levels or rates), and other elements which are determinant of the way in which the estimator combines the available information. It is therefore important to explore, through different scenarios, the way in which all those elements affect the behavior of the composite estimator. In future work we plan to investigate the sample properties of this composite estimators. The study may have useful implications for the sample design and, particularly, for the sample distribution among the different surveys' subdomains if it is decided that composite instead of direct estimators will be used for estimation in the small areas.