



Estudios de Economía Aplicada

ISSN: 1133-3197

secretaria.tecnica@revista-eea.net

Asociación Internacional de Economía
Aplicada
España

Rojo García, J.L.; Domínguez Martín, Rafael
Estimaciones para pequeñas áreas: un enfoque bayesiano al problema de la distribución de una
magnitud económica
Estudios de Economía Aplicada, vol. 20, núm. 1, abril, 2002, pp. 217-240
Asociación Internacional de Economía Aplicada
Valladolid, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=30120104>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Estimaciones para pequeñas áreas: un enfoque bayesiano al problema de la distribución de una magnitud económica

*ROJO GARCÍA, J.L. y **SANZ GÓMEZ, J.A.

Facultad de Ciencias Económicas. Departamento de Economía Aplicada (Estadística y Econometría). Universidad de Valladolid.

Telf.: 983 42 33 21-Fax: 983 42 33 20; *e-mail: rojo_s@eco.uva.es; **e-mail: tono@eco.uva.es

RESUMEN

En este trabajo se desarrolla un método para distribuir un agregado económico en áreas geográficas menores.

Las estimaciones de los valores desagregados se realizan mediante procedimientos bayesianos, con distribuciones a priori aproximadamente no informativas para ciertos parámetros o de Laplace para dichos valores. La implementación de dichas estimaciones se realiza mediante muestreo de Gibbs.

El procedimiento se utiliza para resolver problemas de congruencia en estimaciones de macromagnitudes derivadas de modelos econométricos regionales.

Se sugiere la posibilidad de aplicar el procedimiento para problemas temporales, relacionados con trimestralización de magnitudes económicas

Palabras clave: Métodos bayesianos, Muestro de Gibbs, Congruencia, Análisis regional.

ABSTRACT

This paper provides a method to distribute an economic aggregate among smaller areas, using indicators. It makes use of Bayesian tools, implemented by Gibbs sampling in order to obtain the estimates. The method, with minor changes, is applied to adjust predictions provided by individual (single country) econometric models, with the overall prediction arising from an aggregate model. Authors also explore the adequacy of their method in solving problems of time series interpolation or temporal disaggregation.

Código UNESCO: 62C10, 62C12.

Artículo recibido el 20 de enero de 2001. Aceptado el 16 de mayo de 2001.

1. INTRODUCCIÓN. PLANTEAMIENTO DEL PROBLEMA DE CONGRUENCIA. SOLUCIONES CONOCIDAS

En este trabajo nos planteamos la distribución de una magnitud económica en áreas geográficas menores. Por ejemplo, por referirnos al caso español, puede repartirse el PIB

nacional de un determinado año entre las distintas Comunidades Autónomas, o bien, distribuir entre comarcas el valor añadido industrial a precios de mercado de una provincia. Asimismo, puede utilizarse para desagregar el valor añadido en Productos Manufacturados de una provincia entre los correspondientes a Bienes de equipo, Bienes intermedios o Bienes de consumo.

Con pequeñas modificaciones, el procedimiento puede aplicarse cuando la magnitud se obtiene como media, en lugar de como suma.

Dos son las situaciones más habituales. En la primera de ellas, se conoce una aproximación a los valores desagregados, por lo que el problema planteado se resume en cuadrar dichas aproximaciones, de manera que sean congruentes con la magnitud agregada. Este problema se denomina por ello de congruencia. Ejemplos de esta situación surgen en modelizaciones multinacionales y multirregionales, en los que se estiman separadamente modelos que posteriormente se ajustan a estimaciones globales.

En la segunda, no se dispone de aproximaciones, sino que conocemos la evolución temporal de indicadores cuyos ritmos de crecimiento informan sobre las variaciones de la magnitud en las áreas geográficas menores. Ejemplos de dicha situación se encuentran en los intentos de realizar estimaciones provinciales de ciertas magnitudes a partir de indicadores relacionados y de la correspondiente magnitud agregada.

Aparentemente, esta técnica podría utilizarse para realizar desagregaciones temporales (por ejemplo, la elaboración de Contabilidades trimestrales desde un enfoque no contable). Aunque la idea general del trabajo es aplicable, las modificaciones son en este caso de cierta importancia, por el papel relevante que juega en este caso el diseño del perfil temporal de la magnitud trimestral.

En este artículo resolvemos únicamente la primera de las dos situaciones. Se conoce, para un año base, la magnitud desagregada en las distintas regiones, y por tanto, la magnitud en el ámbito nacional. Para otro año que llamaremos corriente, conocemos el valor nacional de la magnitud y aproximaciones que denominaremos iniciales (probablemente no congruentes) a las magnitudes regionales. El procedimiento permitirá obtener estimaciones aproximadamente congruentes de estas últimas.

Esta técnica puede encadenarse secuencialmente de manera simple aunque no óptima si se desea hacer congruentes estimaciones de dos o más años. No abordaremos en este trabajo la obtención de un procedimiento óptimo para varios años.

Callealta (1991, 1995) plantea y resuelve el problema mediante un procedimiento de triangularización matricial ordenada, que permite detectar la presencia de restricciones lineales dependientes y de medidas no observables a través de los balances.

Obtiene, así, el valor buscado de la congruencia sujeto a restricciones lineales, el cual se obtiene optimizando cuadráticamente.

En el siguiente apartado formulamos el problema y obtenemos su solución. En el tercero estableceremos el procedimiento de simulación de valores regionales; el cuarto describe el diagnóstico de convergencia así como su implementación. El quinto describe una aplicación al reparto entre 18 regiones del valor añadido nacional. Por último se detallan brevemente las conclusiones y se plantean posibilidades de ampliación a otros escenarios.

2. PLANTEAMIENTO DE NUESTRO PROBLEMA. OBJETIVO

Sea V_n la magnitud nacional¹ para la que disponemos de su valor en el periodo base, V_0 , y en el periodo corriente, V_1 . Asimismo, se dispone de los valores añadidos regionales en el periodo base, v_{r0} , $r = 1, \dots, P$. Obviamente, para el periodo base se cumple que la magnitud nacional es la suma de los P valores regionales,

$$V_0 = \sum_{r=1}^P v_{r0}.$$

Se conoce, además, una estimación inicial de las magnitudes en el periodo corriente, $v_{r,ini}$, $r = 1, \dots, P$, y se desea hacer una estimación congruente de las mismas, v_{r1} , $r = 1, \dots, P$.

La distribución a priori de estas magnitudes debe recoger la información proveniente de los valores iniciales. Se nos plantean dos posibilidades:

1. Tomar unas distribuciones a priori que sean función de $|v_{r1} - v_{r,ini}|$ para cada región, r .
2. Que las distribuciones a priori dependan de los errores iniciales cometidos en la estimación de las tasas de variación de la magnitud regional.

Entre estas dos posibilidades, y tras desarrollar ambas, los autores optamos por la segunda. La primera, más simple desde el punto de vista analítico, produce importantes errores porcentuales de estimación en regiones para las que la magnitud toma valores menores.

En definitiva, si para $r = 1, \dots, P$, llamamos

$$Tv_{r,ini} = \frac{v_{r,ini} - v_{r0}}{v_{r0}}$$

y

$$Tv_r = \frac{v_{r1} - v_{r0}}{v_{r0}}$$

1. Aunque, como hemos dicho, puede aplicarse a otros tipos de desagregación, a partir de ahora llamaremos regiones a las áreas menores y nación a su conjunto.

se adopta como distribución a priori de Tv_r una distribución de Laplace. Esta situación, habitual por otro lado, puede verse en Akaike (1980) y Young (1996). La distribución de Laplace² (también llamada exponencial doble o distribución del error de primer tipo) es una distribución de probabilidades que mantiene cierta similitud con la distribución normal. Como ella, es unimodal, simétrica, si bien las colas son algo más gruesas (más probables) que las de la distribución normal de la misma media y varianza. En definitiva, no concentra tanto las probabilidades como la distribución normal en torno a su media, esto es, justifica la posibilidad de mayores desviaciones entre la tasa de variación regional y la estimación inicial de la misma. Por tanto, escribimos

$$p(v_{r1} | I_r, D) \propto I_r e^{-I_r |Tv_r - Tv_{r,ini}|}$$

donde I_r es un parámetro positivo y D son los datos

$$D = \{V_0, V_1, v_{10}, K, v_{P0}, v_{1,ini}, K, v_{P,ini}\}$$

Supondremos asimismo que v_{r1} , $r = 1, \dots, P$, son independientes a priori³.

En definitiva, la distribución a priori conjunta de los valores regionales resulta

$$p(v_{11}, v_{21}, K, v_{P1} | I_1, K, I_P, D) \propto \prod_{r=1}^P I_r e^{-I_r |Tv_r - Tv_{r,ini}|}$$

Para los parámetros, I_r , $r = 1, \dots, P$, implicados en esta distribución a priori suponemos independencia⁴, y distribuciones a priori aproximadamente uniformes en \mathbf{R}^+ . Puesto que la

2. Una variable, X , sigue una distribución de Laplace (doble exponencial) de parámetro λ ($\lambda > 0$) si su función de densidad es

$$f(x) = \frac{I}{2} e^{-I|x|}, -\infty < x < \infty$$

Su esperanza es igual a cero, y su varianza vale $Var(X) = \frac{2}{I^2}$.

3. Obviamente, la suposición de independencia entre los valores regionales es meramente simplificadora, ya que existen relaciones entre los crecimientos regionales basadas en la proximidad geográfica o económica (especializaciones sectoriales, economías de escala, similar climatología, etc.). No obstante, la consideración de las relaciones de dependencia a priori exige especificar el tipo, forma y grado de dependencia, lo que complicaría mucho el modelo. En esta versión se elude, por tanto, la asunción de dependencias a priori entre los valores regionales. Ello no elimina, con todo, la dependencia entre dichos valores, a través de la función de verosimilitud.

4. Es habitual suponer independencia a priori entre estos parámetros, así como entre los implicados en la función de verosimilitud, ya que su papel se limita a añadir elasticidad a las distribuciones correspondientes, sin que la dependencia a priori entre los mismos tenga un sentido económico en el problema.

distribución uniforme en \mathbf{R}^+ es impropia, tomaremos distribuciones gamma aproximadamente planas (este recurso es habitual. Por ejemplo, puede consultarse Spiegelhalter et al. (1995), Young (1996) o Zellner et al. (1991), entre otros).

$$p(\mathbf{I}_r) \propto e^{-I_r a_{I_r}}, r=1, K, P$$

con a_{I_r} pequeño por ejemplo, $a_{I_r}=0.01$ (nótese que, entonces, $Var(\mathbf{I}_r)=10^4$).

En cuanto a la verosimilitud⁵, supondremos que la relación en el periodo corriente entre la magnitud nacional y las regionales denota el carácter aproximado de estas últimas, pero expresado en términos de tasas,

$$TV = TV_{sum} + \mathbf{e}$$

donde $TV = \frac{V_1 - V_0}{V_0}$ es la tasa de variación nacional de la magnitud, y

$$TV_{sum} = \frac{\sum v_{r1} - V_0}{V_0}$$

es la tasa de variación nacional que se deriva de tomar, en el periodo corriente, el valor nacional de la magnitud como suma de los valores regionales. Se supone asimismo \mathbf{e} que sigue una distribución de Laplace de parámetro \mathbf{b} , esto es,

$$l(v_{11}, v_{21}, K, v_{P1} | \mathbf{b}, D) \propto \mathbf{b} e^{-\mathbf{b}|TV - TV_{sum}|}.$$

5. El modelo que propone la función de verosimilitud puede aparecer como contradictorio con la distribución de los valores regionales. En efecto, si se considera que TV es un dato, la distribución de \mathbf{e} se deduciría de la relación $\mathbf{e} = TV - TV_{sum}$. Pero TV es una variable aleatoria antes de extraer la muestra (de tamaño 1) que proporciona su valor. Más en detalle, la generación de los datos se desarrolla de la siguiente forma: En primer lugar, los valores regionales, v_{r1} , $r=1, K, P$, intervienen con su distribución *a priori*,

$$p(v_{11}, v_{21}, K, v_{P1} | \mathbf{I}_1, K, \mathbf{I}_P, D)$$

El modelo de comportamiento establece que

$$\frac{V_1^* - V_0}{V_0} = \frac{\sum v_{r1} - V_0}{V_0} + \mathbf{e}$$

donde \mathbf{e} tiene la distribución de Laplace anterior y V_1^* es la variable que representa el valor añadido nacional (aleatorio) en el periodo corriente. Este modelo determina la distribución de V_1^* .

Se obtiene una realización muestral, V_1 , de V_1^* , que de hecho, es el valor nacional proporcionado por las estadísticas agregadas (Contabilidad Nacional en nuestro caso). Esta realización nos permite escribir la función de verosimilitud, que sólo será función de los valores regionales, y de los datos.

Obsérvese que el modelo planteado es un modelo básico de errores, en el que se supone que la magnitud a medir (la tasa de variación nacional) es estimada aproximadamente mediante la tasa que se calcula a partir de los valores regionales. Es admisible un cierto margen de error, para permitir que los valores regionales se adecuen, al menos parcialmente, a las propuestas iniciales (lo que el modelo recoge a través de las distribuciones a priori). El desajuste entre ambas tasas no puede ser elevado, por lo que la distribución de \mathbf{e} propuesta (doble exponencial) concede poca probabilidad a errores amplios.

Se acepta, finalmente, que \mathbf{b} tiene una distribución a priori independiente del resto de los parámetros y aproximadamente plana, para lo que haremos

$$p(\mathbf{b}) \propto e^{-b \cdot a_b}$$

con a_b suficientemente pequeño, por ejemplo, $a_b = 0.01$.

En definitiva, la distribución a posteriori para las magnitudes regionales y el resto de los parámetros resulta

$$\begin{aligned} p(v_{11}, v_{21}, K, v_{p1}, \mathbf{I}_1, K, \mathbf{I}_p, \mathbf{b} | D) &\propto \\ &\propto p(\mathbf{b}) \cdot p(\mathbf{I}_1, K, \mathbf{I}_p | \mathbf{b}) \cdot p(v_{11}, K, v_{p1} | \mathbf{I}_1, K, \mathbf{I}_p, \mathbf{b}, D) \cdot l(v_{11}, K, v_{p1}) \propto \\ &\propto e^{-ba_b} \cdot \prod_{r=1}^P e^{-I_r a_{I_r}} \cdot \prod_{r=1}^P I_r e^{-I_r |TV_r - TV_{r,ini}|} \cdot \mathbf{b} e^{-\mathbf{b} |TV - TV_{sum}|} \end{aligned}$$

Como es conocido (véase Ferguson (1967)), el estimador óptimo es el estimador de Bayes que, suponiendo una función de pérdida cuadrática en los parámetros, es la media de la distribución a posteriori. Su complejidad aconseja utilizar procedimientos de simulación. En nuestro caso, las distribuciones condicionadas son manejables, por lo que proponemos el muestreo de Gibbs, uno de los procedimientos englobados en las denominadas *Markov Chain Monte Carlo* (MCMC). Una referencia actualizada de dicho procedimiento, y que puede ser consultada por el lector, es Gamerman (1997).

No obstante, el comportamiento, en la simulación, de los parámetros $\mathbf{I}_1, \dots, \mathbf{I}_p, \mathbf{b}$ es muy volátil, por lo que conviene extraer la distribución a posteriori de los valores regionales, para realizar a partir de ahí la simulación. Así, integrando en dichos parámetros se obtiene sin dificultad

$$\begin{aligned} \int_{\mathbf{b}} \int_{\mathbf{I}_1} \int_{\mathbf{I}_p} p(v_{11}, K, v_{p1}, \mathbf{I}_1, K, \mathbf{I}_p, \mathbf{b} | D) d\mathbf{b} d\mathbf{I}_1 d\mathbf{I}_p &\propto \\ &\propto \frac{1}{(a_b + |TV - TV_{sum}|)^2} \cdot \frac{1}{\prod_{r=1}^P (a_{I_r} + |TV_r - TV_{r,ini}|)^2} \end{aligned}$$

El muestreo de Gibbs exige que las distribuciones condicionadas sean conocidas (y, obviamente, suficientemente manejables como para que sus valores puedan simularse sin dificultad). Obsérvese que la distribución condicionada de un valor regional v_{h1} , $h = 1, \dots, P$, resulta

$$p(v_{h1} | v_1^{(h)}, D) \propto \frac{1}{(a_b + |TV - TV_{sum}|)^2} \cdot \frac{1}{(a_{I_h} + |Tv_h - Tv_{h,ini}|)^2}$$

donde $v_1^{(h)}$ denota el vector (v_{11}, \dots, v_{P1}) , excluido v_{h1} .

Podemos escribir esta distribución condicionada explícitamente como función de v_{h1} . Así,

$$|Tv_h - Tv_{h,ini}| = \frac{1}{v_{h0}} |v_{h1} - a|$$

siendo $a = v_{h,ini}$. Análogamente

$$|TV - TV_{sum}| = \frac{1}{V_0} |v_{h1} - b|$$

donde $b = V_1 - \sum_{r \neq h} v_{r1}$. Obsérvese que mientras que a es la aproximación inicial a v_{h1} , b es

la aproximación resultante de hacer congruente el valor nacional de la magnitud con las aproximaciones de las restantes regiones. Sustituyendo estas expresiones en la distribución condicionada y operando con los factores de su divisor, obtenemos

$$p(v_{h1} | v_1^{(h)}, D) \propto \frac{1}{(c + |v_{h1} - a|)^2} \cdot \frac{1}{(d + |v_{h1} - b|)^2}$$

donde

$$\begin{aligned} a &= v_{h,ini} \\ b &= V_1 - \sum_{r \neq h} v_{r1} \\ c &= a_{I_h} \cdot v_{h0} \\ d &= a_b \cdot V_0 \\ a_{I_h} &= a_b = 0.01 \end{aligned}$$

3. SIMULACIÓN DE UN VALOR REGIONAL

En esta sección describiremos el método empleado para la simulación de un valor de v_{h1} a partir de la distribución condicionada del final del apartado anterior. Para aliviar en lo posible la notación, llamaremos x al valor v_{h1} , y escribiremos la densidad condicionada como

$$p(x|v_1^{(h)}, D) \propto \frac{1}{(c+|x-a|)^2} \cdot \frac{1}{(d+|x-b|)^2} \equiv g(x)$$

De cara a facilitar la lectura del apartado, hemos trasladado la demostración de los resultados al Apéndice I.

En primer lugar, obtenemos la expresión completa de la función de densidad condicionada, necesaria para realizar la simulación. Esto es, obtenemos la constante k tal que

$$p(x|v_1^{(h)}, D) = \frac{1}{k} \cdot g(x)$$

El lema 1 del Apéndice I establece el valor de dicha constante.

Como comentaremos más adelante, el procedimiento empleado exige obtener una función mayorante de la densidad condicionada. Ello exige que nos informemos sobre la forma de dicha densidad. Describimos a continuación el contenido esencial de los resultados obtenidos, cuya demostración y enunciado riguroso se presenta en el Apéndice I:

Resultado 1. La función de densidad, $p(x|v_1^{(h)}, D)$ alcanza su máximo (absoluto) en $x=a$.

Ocasionalmente, puede tener un máximo relativo en $x=b$. (Este resultado se demuestra en el lema 2 del Apéndice I).

Resultado 2. Dicha densidad es creciente a la izquierda del intervalo determinado por a y b , y decreciente a su derecha, lo que es obvio a la vista de su expresión. Además, entre a y b puede tener un mínimo, x_{min} , cuya expresión, si existe, se expresa en el Corolario 3 del Apéndice I. En dicho caso, se calcula, asimismo, el valor $p(x_{min}|v_1^{(h)}, D)$ de la función de densidad en el mínimo (lema 4 del Apéndice I).

Los dos resultados anteriores pueden ilustrarse a través de la figura 1, que muestra la función de densidad condicionada. El aspecto de la misma es, bien el que presenta la figura, o bien con monotonía decreciente entre el punto **A** y el **B**.

Como indicamos anteriormente, el procedimiento de simulación es el denominado de aceptación-rechazo (*acceptance-rejection*). Este método viene sugerido por la expresión analítica de la función de densidad, f , que hace difícil utilizar otros métodos de simulación

(transformación inversa, composición o convolución, entre los más importantes). Haremos sólo una breve descripción de dicho método, remitiendo al lector interesado, bien a la referencia original (von Neumann (1951)) o bien a versiones más depuradas y didácticas (por ejemplo, Law y Kelton (1991), págs. 478 a 484). Sin ánimo de citar referencias exhaustivas, pueden verse aplicaciones de este método a la generación de distribuciones de uso frecuente en Ahrens y Dieter (1974) o Schmeiser y Lal (1980) para distribuciones gamma, o Schmeiser y Babu (1980) para distribuciones beta. Esta última referencia utiliza en la acotación una función lineal a trozos, como en nuestra aplicación.

Para una variable continua, X , con función de densidad f , el método de aceptación y rechazo exige proporcionar una función, t , que acote superiormente a la función de densidad f , esto es, $f(x) \leq t(x)$, para todo x . Se construye a partir de t una densidad, s , $s \propto t$. El procedimiento se desarrolla mediante los siguientes pasos:

1. Se genera Y con densidad s .
2. Se genera $U \propto U(0,1)$, independiente de Y .
3. Si $U \leq \frac{f(Y)}{t(Y)}$, se toma $X=Y$. Si no, se vuelve al paso primero.

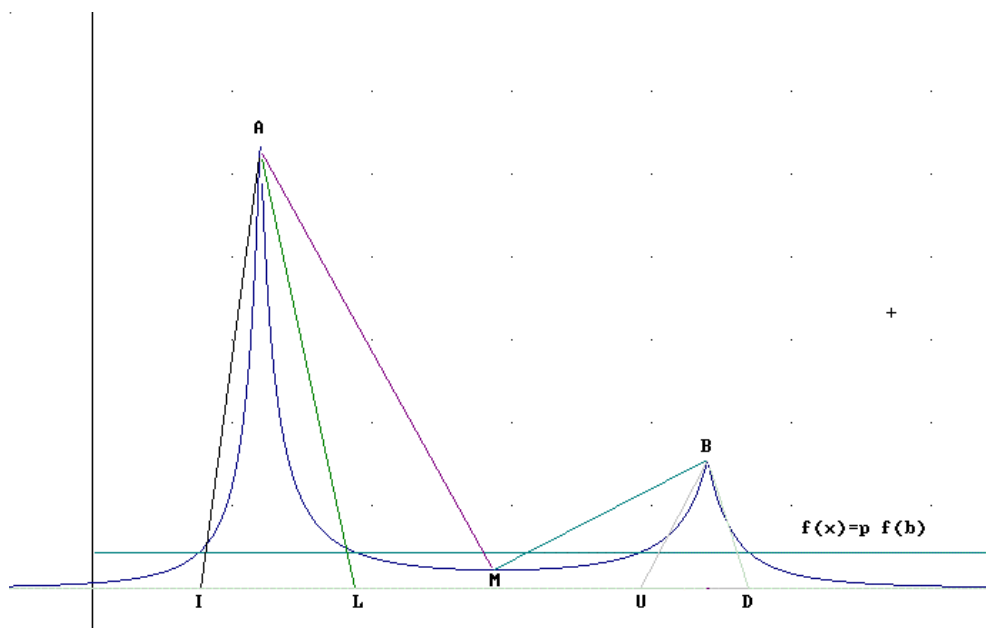
Como puede observarse, la rapidez del método depende del número de simulaciones “en falso”. Se comprueba que la probabilidad de aceptación en el paso tercero es

$$\frac{1}{\int_{-\infty}^{\infty} t(x) dx}$$

por lo que conviene elegir una función mayorante tan pequeña como sea posible (y, obviamente, sencilla de simular). En nuestro caso, las características de la función de densidad, como muestra la figura 1, aconsejan realizar una aproximación al método. El objeto de ésta es, por un lado, que permite una mayorante que puede simularse mediante el método de la transformación inversa, ya que está formada por distribuciones triangulares. Por otro lado, esta aproximación va a permitir la aceleración de la simulación que exige la alta dimensionalidad del problema.

Así, fuera del intervalo determinado por a y b se simplifica la mayoración si suponemos que la función de densidad se anula en ambas colas por debajo de un cierto umbral suficientemente pequeño, valor que nosotros tomaremos como $p \cdot p(b|v_1^{(h)}, D)$, con $p=0.01$. Consecuentemente, anularemos también en ambas colas de la distribución la función mayorante. Adicionalmente, si existe un mínimo x_{min} entre a y b , anularemos la densidad (y

Figura 1: Función de densidad condicionada y su mayorante



la mayorante) en el intervalo de valores, x , para los que $p(x|v_1^{(h)}, D) \leq p \cdot p(b|v_1^{(h)}, D)$. La figura 1 ilustra también esta modificación. Así, la mayorante será la poligonal **[IALUBD]** o bien **[IAMBD]**, según que entre a y b la función de densidad tome valores inferiores a $p \cdot p(b|v_1^{(h)}, D)$ o no los tome.

El lema 5 del Apéndice I proporciona las abscisas de los puntos L y U de la figura, y en los lemas 6 y 7 se obtienen las de los puntos I y D .

4. IMPLEMENTACIÓN. DIAGNÓSTICO DE CONVERGENCIA

Como es conocido, existe una amplia discusión en la literatura en relación con la inicialización de la cadena, la selección de valores muestrales que garanticen cuasi-independencia y las reglas para contrastar la convergencia. En el citado trabajo de Dani Gamerman (Gamerman (1997)) puede encontrarse un resumen de las referencias más notables, así como una breve discusión entre las técnicas.

En general, existe un compromiso entre eficiencia y rapidez de simulación, dependiente a su vez del tamaño del problema. En nuestro caso, el alto número de variables exige procedimientos sencillos y automáticos de diagnóstico. En concreto:

- **Elección de valores iniciales y del número de cadenas:**

Las ventajas de emplear múltiples cadenas para evitar que la solución se concentre en una moda local y para ayudar en el diagnóstico de convergencia, han sido discutidas en Gelman y Rubin (1992). En nuestro caso, hemos optado por una única cadena de 3000 iteraciones, tanto debido al elevado número de variables como a la disponibilidad de valores iniciales adecuados que, previsiblemente, nos conducirán a modas globales. En concreto, para una región, r , el valor inicial ${}_{(0)}v_{r1}$ resulta de promediar dos aproximaciones, la primera,

$$v_{r1}^{(1)} = v_{r0} \cdot \frac{V_1}{V_0}$$

resulta de aplicar a las regiones la tasa de crecimiento nacional, y la segunda, $v_{r1}^{(2)}$ es el valor inicial $v_{r,ini}$. En definitiva,

$${}_{(0)}v_{r1} = \frac{v_{r1}^{(1)} + v_{r1}^{(2)}}{2}$$

- **Diagnóstico de convergencia:**

También aquí, la literatura es abundante, por lo que remitiremos al lector interesado al libro de Gamerman (1997), págs. 133 a 144. Nuevamente nuestra elección viene condicionada por el alto número de variables y por el necesario automatismo del procedimiento, y sigue la propuesta de Geweke (1992). Cada 500 simulaciones separamos los 125 valores iniciales y los 125 finales, haciendo en cada paquete 25 grupos de 5 observaciones, para los que calculamos su media.

Entonces, se calculan esperanzas y varianzas de los dos paquetes de 25 números obtenidos y se realiza un test de diferencia de medias asintóticamente normal al nivel de significación del 10%. Si dicho test rechaza la igualdad de medias las 500 observaciones son rechazadas (*burn-in*) y la simulación continúa.

- **Estimación de los valores regionales:**

Una vez alcanzada la convergencia, las técnicas habituales proponen estimar las medias de la distribución marginal a posteriori como

$$\hat{v}_{r1} = \frac{1}{n} \sum_{i=1}^n {}_{(i)}v_{r1}$$

donde ${}_{(i)}v_{r1}$ es la iteración i -ésima del valor añadido para la región r -ésima en el periodo corriente, esto es, se estiman los valores como las medias ergódicas de la cadena.

En cuanto a la precisión de las estimaciones, la dependencia entre iteraciones sucesivas modifica la estimación de la varianza de las medias ergódicas con respecto a la práctica habitual para el muestreo aleatorio simple.

En concreto, si con una notación general denominamos \mathbf{t}^2 a la varianza de la distribución límite de $\sqrt{n}\bar{X}$, donde \bar{X} es la media ergódica, estimaremos dicha dispersión mediante un procedimiento denominado *estimación por lotes* (véase Gamerman (1997) o Schmeiser (1982)) que proporciona

$$\frac{\mathbf{t}^2}{n} = \frac{1}{k(k-1)} \sum_{i=1}^k (\bar{X}_i - \bar{X})^2$$

donde \mathbf{t}^2 es la estimación de \mathbf{t}^2 , a es la longitud de la cadena de la que procede la media ergódica, \bar{X}_i , $i=1, \dots, k$ son medias de k lotes de valores consecutivos de la cadena y \bar{X} su media. Esta dispersión estimada de las medias ergódicas se usa para construir intervalos de confianza para la esperanza, $E_p(X)$, de la distribución a posteriori, teniendo en cuenta que

$$\sqrt{n} \frac{\bar{X}_n - E_p(X)}{\mathbf{t}}$$

se distribuye aproximadamente como una $N(0,1)$.

- **Precisión de las estimaciones e intervalos de confianza:**

Aunque pueden obtenerse a partir de los datos simulados (basados en los cuantiles empíricos), resulta más eficiente el uso de la información contenida en la distribución condicionada, a través del estimador de la función de densidad marginal a posteriori Rao-Blackwelizado (véase, por ejemplo, Gelfand y Smith (1990)), en concreto,

$$f_{v_{h1}}(x|D) = \frac{1}{n} \sum_{j=1}^n f(x|_{(j)} v_1^{(h)}, D)$$

donde j indica la iteración, y se utilizan las últimas n iteraciones ($n=200$ en nuestra aplicación).

Esta densidad se utiliza para estimar la varianza y los cuantiles al 5% y al 95%.

- **Implementación:**

El procedimiento ha sido implementado en TSP, versión 4.2b.

5. CONGRUENCIA DE LAS ESTIMACIONES DEL VAB DE LAS REGIONES ESPAÑOLAS

Nos planteamos hacer congruentes con el VAB nacional a precios de mercado del año 1999 (proporcionado por Contabilidad Nacional⁶) los 18 valores añadidos regionales obtenidos por el Grupo HISPALINK para las 17 Comunidades Autónomas españolas, mas Ceuta y Melilla conjuntamente. Estos valores añadidos regionales son el resultado de estimaciones (predicciones) individuales de los grupos que forman parte del Proyecto HISPALINK.

La congruencia se establece para los totales y para cada una de las 6 ramas en que desagregamos el total nacional, Productos agrarios, Productos energéticos, Productos industriales, Construcción, Servicios de mercado y Servicios de no mercado.

Aunque el problema planteado puede tildarse de concreto, es sólo una muestra de cómo el procedimiento implementado puede emplearse en cualesquiera situaciones en que se combinan modelos del tipo bottom-up (predicciones para las áreas menores que posteriormente se integran en una predicción conjunta) con modelos top-down (para los que una predicción agregada se incrusta en modelos para las áreas menores). En este sentido, el ejemplo propuesto no es sino un reflejo de una necesidad de una solución general para este tipo de situaciones.

Además, como indicamos en la introducción y más adelante comentaremos, los problemas de desagregación para pequeñas áreas pueden plantearse en el contexto del aquí propuesto. Los valores que aquí denominamos iniciales serían el resultado de aplicar al periodo base los crecimientos previstos de ciertos indicadores relevantes o provenientes de opiniones de expertos.

Volviendo sobre nuestra aplicación concreta, digamos que las estimaciones iniciales de cada uno de los equipos suelen utilizar como input (en su fase top-down) la evolución nacional esperada, por lo que los desajustes con los totales nacionales no son excesivos (la tabla 1 posterior muestra estos desajustes, que van desde el 0.01% para los Servicios de mercado hasta el 2.95% de la subrama energética).

Adicionalmente se dispone de valores para 1998 desagregados para cada una de las 6 ramas y de las 18 regiones, congruentes con Contabilidad Nacional y que provienen del banco de datos HISPADAT (banco de datos del Grupo HISPALINK, véase HISPALINK (1999)).

6. Véase Contabilidad Nacional de España (2000). En el momento de la publicación, el INE ha publicado ya sus estimaciones para los años 1999 y 2000.

Como puede verse en la tabla 2 que, por su amplitud, se presenta al final de este trabajo, el procedimiento suministra valores estimados para cada una de las regiones y ramas, que son las medias ergódicas en cada una de las cadenas del muestreo de Gibbs. Asimismo proporciona el error estándar de las estimaciones, y un intervalo de confianza asintóticamente normal para las mismas, utilizando los resultados de la sección 3.

Dicha tabla muestra la riqueza estadística del método empleado. A diferencia de los métodos denominados “aritméticos”, proporciona una evaluación del error estándar de la estimación, más un intervalo de valores razonables para el valor estimado (un intervalo de confianza o, como diríamos en ambiente bayesiano, un intervalo de credibilidad al 95%).

Pueden estimarse los valores añadidos a partir de la función de densidad marginal a posteriori estimada, en lugar de hacerlo a través de las medias ergódicas. De hecho Gamerman (1997) sugiere esta idea (véase la página 128 de la referencia anterior). No obstante, las aproximaciones empleadas en la estimación de la función de densidad marginal predictiva ocasionan que las medias obtenidas a partir de ella cuadren peor con el total nacional. La tabla 1 muestra este hecho, y en ella puede observarse que los errores de esta estimación son muy similares a los iniciales, por lo que esta opción es de poca utilidad práctica. Esta aproximación se realiza en el rango de valores simulados. Así, si

$$\min v_{h1} = \min_{j=1,K,n} ({}_{(j)}v_{h1}) \quad y \quad \max v_{h1} = \max_{j=1,K,n} ({}_{(j)}v_{h1})$$

la aproximación obtiene una malla de valores (200 en la aplicación) y para cada uno de ellos estima la función de densidad marginal predictiva promediando las distribuciones condicionadas de v_{h1} en un rango selecto de iteraciones del resto de las variables. Puesto que, habitualmente, dichas condicionadas son fuertemente leptocúrticas, con frecuencia se ignorarán los valores en los que la función de densidad condicionada alcanza su moda. Este es el motivo de que la calidad de la estimación de medidas de posición a partir de la función de densidad estimada pueda verse disminuida.

Tabla 1: Desajustes entre las estimaciones regionales y el total nacional

	<i>Valores iniciales</i>	<i>Estimación media ergódica</i>	<i>Estimaciones f.d. predictiva</i>
Total	0.54	0.13	0.54
Agricultura	0.80	0.22	0.72
Energía	2.95	1.23	2.74
Productos manufacturados	1.01	0.30	0.90
Construcción	1.53	0.52	1.37
Servicios de mercado	0.01	-0.04	0.01
Servicios de no mercado	0.38	0.07	0.36

Como puede verse en dicha tabla, los menores errores corresponden a las estimaciones mediante las medias ergódicas; salvo para los Servicios de mercado, son sensiblemente inferiores, tanto a los que se deducen de los valores iniciales (lo cual resulta natural) como a los que se deducen de las medias que proporciona la función de densidad marginal predictiva estimada.

La tabla 3 muestra las diferencias entre las tasas de variación propuestas por los equipos y los resultados de nuestras estimaciones. En general, las correcciones van en la misma dirección para todas las regiones, como consecuencia de las desviaciones en el total nacional.

6. CONCLUSIONES Y AMPLIACIONES

En este trabajo se propone un método para estimar magnitudes económicas en pequeñas áreas cuando se conoce su valor en un área agregada y aproximaciones iniciales de las mismas. La solución propuesta consiste en la obtención del estimador de Bayes para distribuciones a priori aproximadamente uniformes, mediante la realización de muestreo de Gibbs a partir de las distribuciones condicionadas.

Se estiman adicionalmente los errores estándar de los estimadores, y la función de densidad marginal predictiva para cada una de las variables.

El procedimiento se aplica para cuadrar las cifras del VAB a precios de mercado para las regiones españolas, a partir de estimaciones iniciales proporcionadas por el grupo HISPALINK, mejorando sustancialmente la congruencia con la cifra nacional.

Aunque el trabajo puede aplicarse secuencialmente para cuadrar varios años, este procedimiento secuencial no es óptimo. Los autores trabajan en la búsqueda de un procedimiento más eficiente que no sea excesivamente costoso en términos de tiempo de simulación.

Frente a los métodos tradicionales, basados en procedimientos aritméticos (habitualmente son métodos numéricos de maximización o minimización de una distancia en un convexo) que o bien no son métodos propiamente estadísticos o sus propiedades estadísticas no se ponen de manifiesto, el método aquí propuesto explicita la estructura probabilística de las variables implicadas a través de un modelo de errores.

Estas estimaciones de valores añadidos no resultan comparables con las que el INE obtiene en sus contabilidades regionales, y ello por varios motivos:

1. El INE utiliza un enfoque *contable*, con una información proveniente de indicadores mucho más amplia, si bien con un mayor retraso.
2. Nuestro método es deudor de la calidad de las estimaciones (predicciones) iniciales, siendo un procedimiento de ajuste más que de estimación.

3. El INE no proporciona medidas estadísticas sobre sus estimaciones, por lo que no son posible las comparaciones de precisión entre ambas.

De cara a un futuro próximo, los autores orientan las aplicaciones del método a la estimación de datos provinciales o municipales, donde las carencias estadísticas son, en nuestro país, muy elevadas.

Por otro lado, en muchas ocasiones se conocen o imponen cotas superiores o inferiores a los valores estimados, de manera que las variaciones interanuales no sean muy acusadas. No resulta complicada la inclusión de estas restricciones en el problema, aunque los autores no hemos realizado aún una propuesta en este sentido.

Finalmente, desde un enfoque más próximo a los de tipo contable, los autores trabajan en procedimientos de estimación sin la presencia de valores iniciales, sino mediante la inclusión en el modelo del comportamiento de indicadores representativos.

Agradecimientos

Los autores agradecen al profesor Javier Callealta, de la Universidad de Alcalá de Henares, las sugerencias a una versión preliminar de este trabajo. Asimismo los comentarios de un evaluador anónimo han enriquecido notablemente el texto. En cualquier caso, las deficiencias que pudieran encontrarse son de la exclusiva responsabilidad de los autores.

7. BIBLIOGRAFÍA

- AHRENS, J.H. y DIETER, U. (1974): Computer Methods for Sampling from Gamma, Beta, Poisson, and Binomial Distributions, *Computing*, 12: 223-246.
- AKAIKE, H.: Seasonal adjustment by a Bayesian modelling, *Journal of time series analysis*, 1, (1980), 1-13.
- CALLEALTA, F.J. (1991): Un método alternativo para reconciliación de datos en tablas input-output, V reunión ASEPELT-España.
- CALLEALTA, F.J.: Una visión integrada regionalmente de la economía Española: El modelo de congruencia en el modelo HISPALINK, *La integración económica regional en España. La Comunidad Valenciana*, Mundi-Prensa, 29-46 (18) (1995).
- FERGUSON, T.S. (1967): *Mathematical Statistics: a Decision Theoretic Approach*, New York: Academic Press.
- GAMERMAN, D. (1997): Markov Chain Monte Carlo (Stochastic Simulation for Bayesian Inference), Capítulo 5. New York: Chapman & Hall.
- GELFAND, A.E. y SMITH, A.F.M. (1990): *Sampling-Based Approaches to Calculating Marginal Densities*, *Journal of the American Statistical Association*, 85 (140), 398-409.

- GELMAN, A. y RUBIN, D.B. (1992): *Inference from Iterative Simulation Using Multiple Sequences*, *Statistical Science*, 7 (4), 457-511.
- GEWEKE, J. (1992): *Evaluating the accuracy of Sampling-based approaches to calculating posterior moments*, *Bayesian Statistics 4* (Ed. J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith), Clarendon Press, Oxford, UK.
- HISPALINK (1999): Banco de datos HISPADAT. Edición de diciembre de 1999. Coordinadora Ana López. Madrid.
- INE (2000): Boletín Trimestral de Coyuntura, nº 75-marzo 2000. Madrid.
- RAO, J.N.K. (2000): Statistical methodology for indirect estimations in small areas. Eustat.
- SCHMEISER, B.W., y LAL, R.: Squeeze Methods for Generating Gamma Variates, *J. Am. Statist. Assoc.*, 75: 679-682 (1980)
- SCHMEISER, B.W., y BABU, A.J.G.: Beta Variate Generation via Exponential Majorizing Functions, *Operations Res.*, 28: 917-926 (1980)
- SPIEGELHALTER, D., THOMAS, A. y BEST, N. (1995): BUGS 0.5, Bayesian Inference using Gibbs Sampling Manual, *MRC Biostatistics Unit*, 37-38.
- LAW, A.M. y KELTON, W.D.: *Simulation Modeling & Analysis*, Industrial Engineering Series, McGraw-Hill International Editions, (1991).
- VON NEUMANN, J.: Various Techniques Used in Connection with Random Digits, *Natl. Bur. Std. Math. Ser.*, 12: 36-38 (1951)
- YOUNG, M.R. (1996): Robust Seasonal Adjustment by Bayesian Modelling, *Journal of Forecasting*, 15, 355-367.
- ZELLNER, A., HONG, C. y MIN, C. (1991): Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques, *Journal of Econometrics*, 49, 275-304.

APÉNDICE I

En este apéndice se demuestran resultados auxiliares que ya han sido presentados en la sección tercera. Las notaciones y la motivación del problema son los descritos en dicha sección.

Lema 1: La función de densidad v_{h1} de condicionada por $(v_1^{(h)}, D)$ resulta

$$f(x) = \frac{1}{k} \cdot g(x), \quad -\infty < x < \infty$$

$$k = \frac{2 \ln \frac{c}{M+d}}{(M+d-c)^3} + \frac{M+N}{c(M+d)(M+d-c)^2} - \frac{2 \ln \frac{cd}{M^2+MN+cd}}{(M+N)^3} +$$

con

$$+ \frac{M^2N+MN^2-2Mcd}{cd(M^2+MN+cd)(M+N)^2} + \frac{M+N}{d(M+c)(M+c-d)^2} - \frac{2 \ln \frac{M+c}{d}}{(M+c-d)^3}$$

siendo $M = |a-b|$ y $N = c+d$.

Demostración: Es inmediata sin más que calcular

$$k = \int_{-\infty}^{\infty} g(x) dx.$$

El siguiente lema establece que el máximo de la función de densidad condicionada se encuentra en el punto a .

Lema 2: Si a y b son los valores definidos anteriormente, se cumple que $f(a) > f(b)$. Además, f alcanza su máximo en a y, ocasionalmente, un máximo relativo en b .

Demostración: En efecto,

$$g(a) = \frac{1}{c^2(d+|a-b|)^2}$$

y

$$g(b) = \frac{1}{d^2(c+|a-b|)^2}$$

Si planteamos $\frac{1}{g(a)} - \frac{1}{g(b)}$ resulta

$$c^2(d+|a-b|)^2 - d^2(c+|a-b|)^2 = (a-b)^2(c^2 - d^2) + 2cd(c-d)|a-b|$$

y, como siempre se cumple $0 < c < d$, la expresión anterior es la suma de dos cantidades negativas, con lo que $\frac{1}{g(a)} - \frac{1}{g(b)} < 0$; por tanto, $g(a) > g(b)$ y, en consecuencia, $f(a) > f(b)$.

Resulta obvio que la función de densidad, f , es creciente a la izquierda del intervalo determinado por a y b y decreciente a su derecha. Dentro de dicho intervalo si, por ejemplo $a < b$, la función g resulta

$$g(x) = \frac{1}{(c+x-a)^2(d+b-x)^2}$$

y, por tanto,

$$\ln(g(x)) = -2[\ln(c+x-a) + \ln(d+b-x)]$$

Derivando,

$$\frac{d \ln(g(x))}{dx} = -2 \left[\frac{a+b-c+d-2x}{(c+x-a)(d+b-x)} \right]$$

y esta derivada es negativa si $x < \frac{a+b}{2} + \frac{d-c}{2}$ (recordemos que es un valor del intervalo (a, b)).

Teniendo en cuenta que $c < d$ y, por tanto, que $\frac{d-c}{2} > 0$, la función es decreciente a la derecha de a , pudiendo tener un mínimo entre a y b si $\frac{d-c}{2} < \frac{a+b}{2}$.

Puede realizarse un razonamiento similar si $a > b$. Los resultados se resumen en el siguiente

Corolario 3: El mínimo de f en el intervalo determinado por a y b , si existe, se encuentra en

$$x_{\min} = \frac{a+b-(c-d)}{2} \quad \text{si } a < b \text{ y } a < \frac{a+b-(c-d)}{2} < b$$

y en

$$x_{\min} = \frac{a+b+(c-d)}{2} \quad \text{si } a > b \text{ y } a > \frac{a+b+(c-d)}{2} > b$$

Demostración: El comentario anterior al enunciado de este corolario demuestra el resultado si $a < b$. De la misma forma se comprueba si $a > b$.

Lema 4: La función de densidad en el mínimo, x_{min} , obtenido en el Corolario 3, es, si existe,

$$f(x_{min}) = \begin{cases} \frac{1}{k} \frac{1}{(c + x_{min} - a)^2 (d + b - x_{min})^2} & a < x_{min} < b \\ \frac{1}{k} \frac{1}{(c + a - x_{min})^2 (d + x_{min} - b)^2} & b < x_{min} < a \end{cases}$$

donde k es la constante obtenida en el Lema 1.

Demostración: Sustituyendo, obtenemos el valor de la función de densidad en el mínimo, si éste existe.

Lema 5: Sea p un valor entre 0 y 1. Si x_{min} se encuentra en el intervalo delimitado por a y b y si $f(x_{min}) < p f(b)$, entonces las soluciones de $f(x) = p f(b)$ en dicho intervalo son:

- Si $a < x < b$,

$$x_1 = \frac{a + b - c + d}{2} - \sqrt{\left(\frac{a - b - c - d}{2}\right)^2 + \frac{1}{\sqrt{kpf(b)}}}$$

$$x_2 = \frac{a + b - c + d}{2} + \sqrt{\left(\frac{a - b - c - d}{2}\right)^2 + \frac{1}{\sqrt{kpf(b)}}}$$

- Si $b < x < a$,

$$x_1 = \frac{a + b + c - d}{2} - \sqrt{\left(\frac{a - b + c + d}{2}\right)^2 + \frac{1}{\sqrt{kpf(b)}}}$$

$$x_2 = \frac{a + b + c - d}{2} + \sqrt{\left(\frac{a - b + c + d}{2}\right)^2 + \frac{1}{\sqrt{kpf(b)}}}$$

Además, en ambas situaciones, $x_1 < x_2$.

Demostración: Para $a < b$, las soluciones de la ecuación

$$f(x) = \frac{1}{k(c+x-a)^2(d+b-x)^2} = p \cdot f(b)$$

coinciden con las de

$$(c+x-a)^2(d+b-x)^2 - \frac{1}{kp f(b)} = 0$$

Entonces, si $f(x_{\min}) < p f(b)$, se comprueba inmediatamente que la ecuación admite cuatro soluciones reales, de las que dos, x_1 y x_2 , las definidas en el enunciado, se encuentran en el intervalo (a, b) . La relación $x_1 < x_2$ es obvia sin más que observar sus expresiones.

La solución para $b < a$ es análoga.

Los dos siguientes lemas recogen los puntos de anulación a la izquierda del menor valor entre a y b o a la derecha del mayor. A dichos puntos les llamaremos x_{izq} y x_{der} .

Lema 6: Sea p un valor entre 0 y 1. Una solución de $f(x) = p f(b)$ está a la izquierda del $\min(a, b)$ y resulta

$$x_{izq} = \frac{a+b+c+d}{2} - \sqrt{\left(\frac{a-b+c-d}{2}\right)^2 + \frac{1}{\sqrt{kp f(b)}}}$$

Demostración: Para $x < \min(a, b)$, la ecuación

$$f(x) = \frac{1}{k(c-x+a)^2(d-x+b)^2} = p \cdot f(b)$$

conduce a una ecuación similar a la del lema anterior

$$(x-(a+c))^2(x-(b+d))^2 - \frac{1}{kp f(b)} = 0$$

obteniéndose el resultado.

Lema 7: Sea p un valor entre 0 y 1. Una solución de $f(x) = p f(b)$ está a la derecha del $\max(a, b)$ y resulta

$$x_{der} = \frac{a+b-c-d}{2} + \sqrt{\left(\frac{a-b-c+d}{2}\right)^2 + \frac{1}{\sqrt{kp f(b)}}}$$

Demostración: Análoga a la anterior.

Tabla 2. Estimación del VABPM y medidas de dispersión

		Total	Agricultura	Energía	Productos Industriales	Construcción	Servicios	
							Destinados	No destinados
Andalucía	v.a.	10385181	990740	380500	1233911	959364	5111725	1707144
	err.est.	1697,2	202,0	109,8	341,8	193,2	990,9	372,9
	ex.inf.	10381854	990344	380285	1233241	958986	5109782	1706413
	ex.sup.	10388507	991136	380716	1234581	959743	5113667	1707875
Aragón	v.a.	2464903	120235	66921	577066	175256	1137020	390292
	err.est.	793,8	48,8	67,4	114,4	75,5	413,3	98,7
	ex.inf.	2463347	120139	66789	576841	175108	1136210	390099
	ex.sup.	2466459	120330	67053	577290	175404	1137830	390486
Asturias	v.a.	1795381	68935	94507	329987	177566	831670	295832
	err.est.	655,8	34,8	64,8	147,5	87,3	398,6	105,9
	ex.inf.	1794096	68867	94380	329698	177394	830888	295624
	ex.sup.	1796666	69003	94634	330277	177737	832451	296039
Balears	v.a.	1869469	34958	38883	107950	128316	1358624	206495
	err.est.	685,6	16,5	37,0	99,6	60,0	310,3	64,1
	ex.inf.	1868125	34926	38810	107755	128198	1358016	206369
	ex.sup.	1870813	34991	38955	108145	128433	1359233	206620
Canarias	v.a.	3015741	121867	112991	170601	284257	1830271	501196
	err.est.	790,3	56,1	90,3	86,8	90,3	668,5	145,6
	ex.inf.	3014192	121757	112814	170431	284080	1828961	500910
	ex.sup.	3017290	121977	113168	170771	284434	1831581	501481
Cantabria	v.a.	990297	64366	12092	233516	66864	465946	144541
	err.est.	419,4	19,1	25,2	119,0	42,5	280,5	81,2
	ex.inf.	989475	64328	12042	233282	66781	465396	144382
	ex.sup.	991119	64403	12141	233749	66948	466496	144700
Castilla-León	v.a.	4466380	498545	209066	880627	364683	1712608	805283
	err.est.	1136,5	72,3	83,4	269,9	101,9	514,7	194,8
	ex.inf.	4464152	498404	208903	880098	364483	1711599	804902
	ex.sup.	4468607	498687	209230	881156	364882	1713616	805665
Castilla-La Mancha	v.a.	2758089	307210	263170	441478	278835	1020979	446975
	err.est.	698,0	69,4	126,1	145,7	92,5	498,0	137,1
	ex.inf.	2756721	307074	262923	441192	278653	1020003	446706
	ex.sup.	2759457	307346	263417	441763	279016	1021955	447244
Cataluña	v.a.	14530534	265345	519742	3815287	1001909	7507015	1404102
	err.est.	1747,8	64,7	128,1	588,8	151,4	1351,4	222,5
	ex.inf.	14527108	265218	519491	3814133	1001612	7504366	1403666
	ex.sup.	14533960	265471	519993	3816441	1002206	7509664	1404538

v.a.: valor añadido (media ergódica)

err.est.: error estándar

ex.inf.: extremo inferior

ex.sup.: extremo superior

Tabla 2b. Estimación del VABPM y medidas de dispersión

		Total	Agricultura	Energía	Productos Industriales	Construcción	Servicios	
							Destinados	No destinados
C. Valenciana	v.a.	7280066	278055	252891	1716767	542468	3596921	894252
	err.est.	1719,9	81,9	129,6	411,7	115,4	772,7	175,9
	ex.inf.	7276695	277894	252637	1715960	542242	3595406	893908
	ex.sup.	7283437	278215	253145	1717574	542695	3598435	894597
Extremadura	v.a.	1375983	135030	97520	93567	148311	579989	319816
	err.est.	650,9	44,9	48,3	93,3	82,8	368,4	107,4
	ex.inf.	1374708	134941	97426	93384	148149	579267	319605
	ex.sup.	1377259	135118	97615	93750	148473	580711	320026
Galicia	v.a.	4332200	419867	294804	790512	382020	1757471	691578
	err.est.	1365,0	93,4	83,3	156,0	122,8	500,5	131,3
	ex.inf.	4329524	419684	294641	790206	381779	1756490	691320
	ex.sup.	4334875	420050	294967	790818	382260	1758452	691835
C. de Madrid	v.a.	12044511	28794	206282	1877744	1060065	6819284	2078804
	err.est.	1925,4	18,1	94,8	370,7	175,1	1209,8	304,1
	ex.inf.	12040737	28759	206097	1877018	1059722	6816913	2078208
	ex.sup.	12048284	28830	206468	1878471	1060409	6821655	2079400
R. de Murcia	v.a.	1863779	176180	132131	276553	161346	797413	316201
	err.est.	707,5	47,9	79,9	83,9	81,4	451,9	107,4
	ex.inf.	1862392	176086	131975	276388	161186	796528	315990
	ex.sup.	1865166	176274	132288	276717	161505	798299	316411
C. F. Navarra	v.a.	1221807	63731	13407	434261	85321	482118	144260
	err.est.	608,8	46,7	19,1	105,8	41,5	328,2	101,1
	ex.inf.	1220614	63640	13369	434054	85240	481475	144062
	ex.sup.	1223000	63823	13444	434469	85402	482761	144459
P. Vasco	v.a.	4816115	110658	380442	1500048	290840	1998533	526744
	err.est.	904,2	44,0	152,3	360,1	131,4	530,4	149,3
	ex.inf.	4814343	110572	380144	1499342	290583	1997493	526451
	ex.sup.	4817887	110745	380741	1500753	291097	1999572	527036
La Rioja	v.a.	718734	67184	8150	300354	33115	228450	79292
	err.est.	458,5	36,9	19,2	155,7	33,0	161,9	51,1
	ex.inf.	717836	67111	8113	300049	33051	228132	79192
	ex.sup.	719633	67256	8188	300659	33180	228767	79392
Ceuta y Melilla	v.a.	221017	1422	2704	3196	12116	100480	100294
	err.est.	212,7	2,9	5,1	20,2	13,1	177,9	63,6
	ex.inf.	220600	1416	2694	3156	12091	100131	100170
	ex.sup.	221434	1428	2714	3236	12142	100829	100419

v.a.: valor añadido (media ergódica)

err.est.: error estándar

ex.inf.: extremo inferior

ex.sup.: extremos superior

**Tabla 3. Tasa de variación real del VAB a precios de mercado regionales
(Valores estimados y valores iniciales propuestos por los equipos)**

