

Series anuales de algunos agregados económicos y demográficos regionales, 1955-2009 (RegDat versión 2.3)

Angel de la Fuente*
Instituto de Análisis Económico (CSIC)

Septiembre de 2010

Resumen

En una serie de trabajos anteriores he construido series largas de diversos agregados económicos regionales enlazando las distintas bases de la Contabilidad Regional de España entre sí y con las series históricas elaboradas por la Fundación BBVA. Puesto que esta última fuente sólo ofrece datos para años impares, las series enlazadas heredan esta característica desde 1955 hasta 1989. En la presente nota se construyen series anuales completas de las variables de interés utilizando un sencillo procedimiento de interpolación que incorpora la información anual disponible a nivel nacional para estas variables.

1. Introducción

En una serie de trabajos anteriores (de la Fuente, 2009 y 2010 a y b) he construido series largas de diversos agregados económicos regionales enlazando las distintas bases de la Contabilidad Regional de España (CRE) entre sí y con las series históricas elaboradas por Julio Alcaide y colaboradores para la Fundación BBVA y sus antecesoras. Puesto que esta última fuente sólo ofrece datos para años impares, las series enlazadas heredan esta característica durante el período en el que las series del BBVA constituyen la principal referencia para su construcción. Esto hace que las series enlazadas tengan frecuencia bienal desde 1955 hasta 1989 en el caso de las magnitudes tomadas de la Contabilidad Regional y hasta 1971 en el de las series de población que complementan a las anteriores. Puesto que esta característica de los datos es una limitación importante de cara a algunas aplicaciones, en esta nota se construyen series anuales de las variables de interés. Para ello se utiliza un sencillo procedimiento de interpolación que incorpora la información anual disponible a nivel nacional para muchas de estas variables.

2. El procedimiento de interpolación

Sea X_{it} una de las series regionales de interés y X_t el correspondiente agregado nacional. Durante una parte del período muestral, la serie regional se conoce sólo para años impares y ha de estimarse para los años pares. En la mayor parte de los casos, sin embargo, se dispone de una serie nacional de frecuencia anual para el período completo.

* Este trabajo forma parte de un proyecto de investigación cofinanciado por el Fondo Europeo de Desarrollo Regional y el Servicio de Estudios del BBVA. Agradezco también la financiación del Ministerio de Ciencia e Innovación a través del proyecto ECO2008-04837/ECON.

Sean g_{it}^1 y g_{it}^2 las tasas de crecimiento de X_{it} durante los dos años incluidos en el bienio centrado en t . Utilizando letras minúsculas para indicar que estamos trabajando con logaritmos de X , g_{it}^1 y g_{it}^2 satisfacen la siguiente condición:

$$(1) \quad g_{it}^1 + g_{it}^2 \equiv (x_{it} - x_{it-1}) + (x_{it+1} - x_{it}) = x_{it+1} - x_{it-1} \equiv \Delta_{it}$$

donde la suma Δ_{it} es una variable conocida para todo t par. Si no disponemos de ninguna información adicional, la forma natural de extender la serie regional a los años pares es la *interpolación geométrica* pura de la serie bienal. Este procedimiento equivale a suponer que la tasa de crecimiento anual se mantiene constante dentro de cada bienio, de forma que

$$(2) \quad g_{it}^1 = g_{it}^2 = \frac{\Delta_{it}}{2} \equiv \bar{g}_{it}$$

Aplicando este incremento al valor inicial de la variable (en $t-1$), el valor “esperado” de su logaritmo en t vendría dado por

$$(3) \quad x_{it-1} + \bar{g}_{it} = x_{it-1} + \frac{x_{it+1} - x_{it-1}}{2} = \frac{x_{it+1} + x_{it-1}}{2} \equiv \bar{x}_{it}$$

esto es, por la simple interpolación lineal entre sus valores (medidos en logaritmos) en los años impares que rodean a t . En lo que sigue me referiré a \bar{x}_{it} como a la *tendencia* de x_i en t .

En el presente contexto, sin embargo, contamos con información adicional que debería ayudarnos a mejorar la estimación del valor de X en años pares. En particular, disponemos de datos nacionales de periodicidad anual que podemos utilizar para introducir algún tipo de corrección a la interpolación pura descrita arriba. Así, si observamos que el valor nacional de X está por encima de su tendencia en t (esto es, que $x_t > \bar{x}_t$), sabremos que lo mismo ha de ser cierto al menos para una parte importante de las regiones. Si estamos dispuestos a suponer que todas ellas se comportan de la misma forma en algún sentido, podemos utilizar esta información para rellenar los huecos en la variable regional de interés de una forma que, al menos en promedio, será mejor que la pura interpolación basada en la hipótesis dada en (2).

Antes de proceder, conviene introducir algunos términos. Si definimos

$$(4) \quad \phi_t = \frac{x_t - x_{t-1}}{x_{t+1} - x_{t-1}} = \frac{g_t^1}{\Delta_t}$$

como la fracción del crecimiento agregado de X que se produce durante el primer año del bienio centrado en t , podemos expresar g_{it}^1 y g_{it}^2 como

$$(5) \quad g_{it}^1 = \phi_t \Delta_t$$

y

$$(6) \quad g_{it}^2 = \Delta_t - g_{it}^1 = \Delta_t - \phi_t \Delta_t = (1 - \phi_t) \Delta_t$$

La hipótesis de que todas las regiones se comportan de la misma forma (y por lo tanto de la misma forma que el agregado) se puede concretar al menos dos formas diferentes. La primera consiste en suponer que el crecimiento total registrado durante el bienio de interés se reparte entre el primer y el segundo año del período en las mismas proporciones en todas las regiones. Esto es, estaremos suponiendo que el crecimiento durante el primer año del bienio en la región i viene dado por

$$(7) \quad \hat{g}_{it}^1 = \phi_t \Delta_{it}$$

lo que podemos escribir también en la forma

$$(8) \quad \hat{g}_{it}^1 = \phi_t \Delta_{it} = \frac{\Delta_{it}}{2} + \left(\phi_t \Delta_{it} - \frac{\Delta_{it}}{2} \right) = \bar{g}_{it} + \left(\phi_t - \frac{1}{2} \right) \Delta_{it}$$

Esto es, cuando el agregado nacional esté por encima de su tendencia en t ($\phi_t > 1/2$) supondremos que también lo están todas las regiones e introduciremos en cada una de ellas una corrección al alza a su tasa tendencial de crecimiento, \bar{g}_{it} , que será proporcional al crecimiento total observado en la región i durante el conjunto del período, Δ_{it} . Fijándome en esta propiedad del ajuste, me referiré a este procedimiento como el *método de corrección proporcional* (a la interpolación geométrica pura).

La segunda posibilidad es similar a la primera excepto en que la corrección que se introduce cuando el agregado se desvía de su tendencia es idéntica para todas las regiones. En este caso, estimaríamos el valor de x_{it} como

$$(9) \quad \tilde{x}_{it} = \bar{x}_{it} + (x_t - \bar{x}_t) = (x_{it-1} + \bar{g}_{it}) + x_t - (x_{t-1} + \bar{g}_t)$$

Con el fin de facilitar la comparación con (8), observamos que la estimación de la tasa de crecimiento del primer año vendría dada en este caso por

$$(10) \quad \begin{aligned} \hat{g}_{it}^1 &= \tilde{x}_{it} - x_{it-1} = \bar{g}_{it} + x_t - x_{t-1} - \bar{g}_t = \bar{g}_{it} + (g_t^1 - \bar{g}_t) = \\ &= \bar{g}_{it} + \left(\phi_t \Delta_t - \frac{\Delta_t}{2} \right) = \bar{g}_{it} + \left(\phi_t - \frac{1}{2} \right) \Delta_t \end{aligned}$$

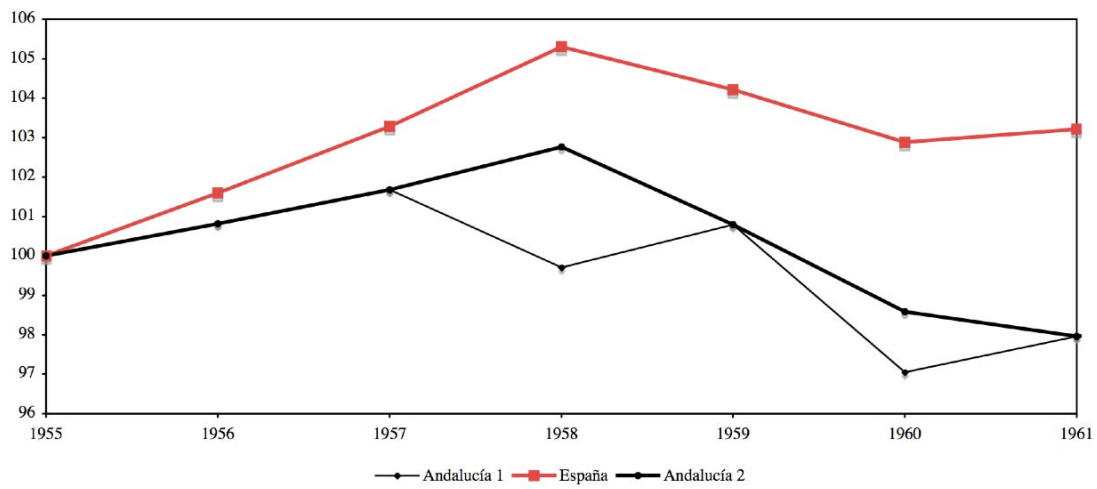
Esta expresión tiene la misma forma que la ecuación (8) y, como ella, exige una corrección al alza de la tasa de crecimiento de cada región (sobre su valor tendencial) cuando la serie nacional crece por encima de su tendencia en t . La única diferencia es que ahora la corrección es la misma para todas las regiones en vez de ser proporcional al crecimiento observado en cada una de ellas durante el conjunto del bienio. Hablaré por tanto del *método de corrección uniforme* (a la interpolación geométrica pura).

¿Cuál de los dos procedimientos es mejor?

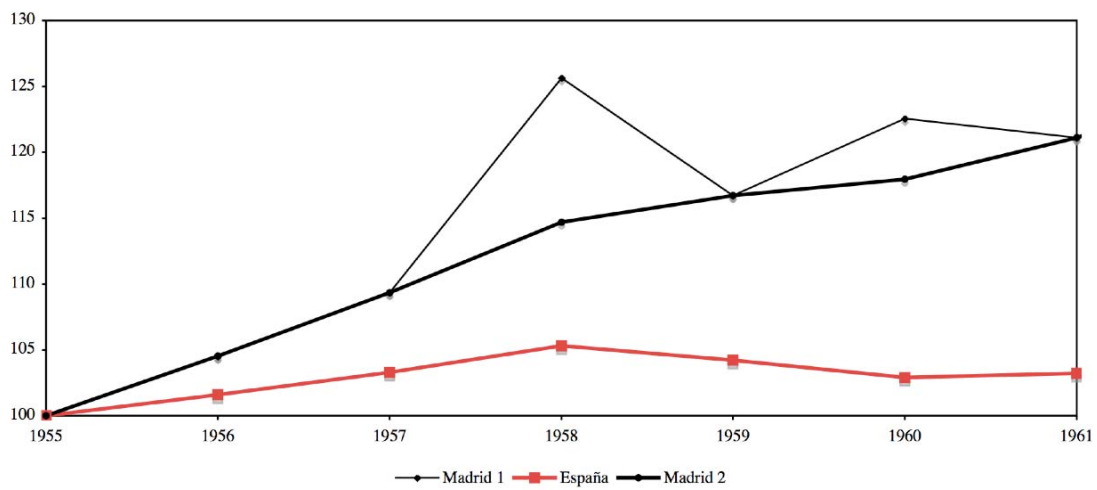
Aunque resulta difícil intentar responder esta pregunta en abstracto, la comparación entre los resultados de los dos procedimientos en ciertos casos concretos sugiere que el método de

**Gráfico 1: Evolución del empleo total (1955 = 100)
comparación entre métodos alternativos de interpolación**

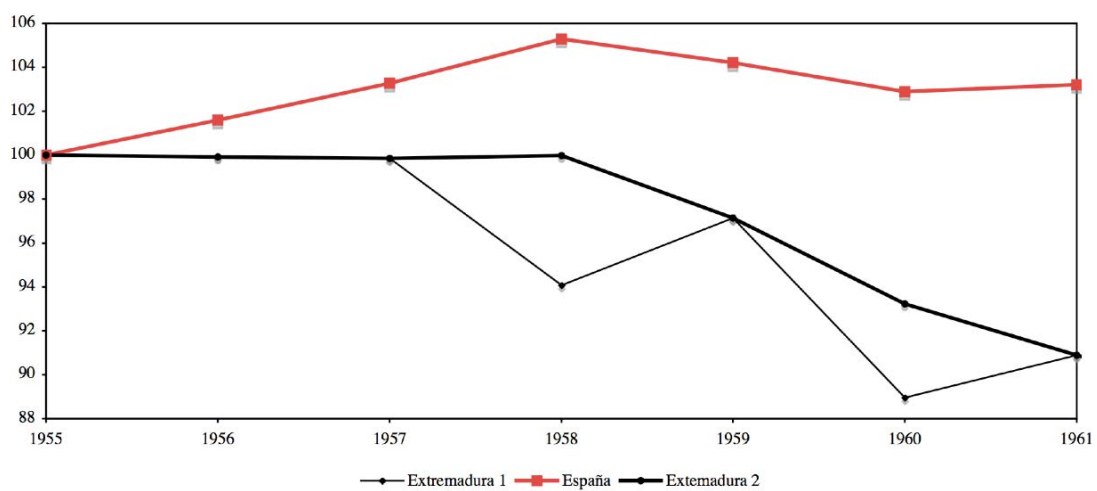
Andalucía



Madrid



Extremadura



corrección proporcional tiende a generar un patrón de dientes de sierra en los datos que resulta generalmente poco plausible.

El Gráfico 1 ilustra este hecho utilizando los primeros años de la serie de empleo total. En cada panel se compara la evolución del empleo en una región determinada con la senda del empleo nacional, con ambas series normalizadas a 100 en su año inicial de 1955. La serie original tiene frecuencia anual en el caso nacional y bienal en el regional. En este último caso se muestran los resultados de aplicar ambos procedimientos de interpolación para extender la serie a los años pares (el método 1 es el de corrección proporcional).

Como se observa en el Gráfico, la serie nacional presenta cambios de tendencia en los años pares de 1958 y 1960, lo que hace que el valor de ϕ_t correspondiente a ambos años sea mayor que 1. Cuando este coeficiente se aplica a valores de Δ_{it} que pueden ser muy distintos de la media nacional la corrección a la tasa tendencial dada en (8) resulta con frecuencia "excesiva," en una dirección en el primer año del bienio y en la contraria en el segundo.

Por el contrario, el método de corrección uniforme produce series con un perfil en principio más plausible. En consecuencia, he optado por este segundo método para extender las series regionales a los años pares. El Cuadro 1 resume el procedimiento utilizado para anualizar cada una de las series e indica también la parte del período muestral afectada en cada caso – esto es, la parte del período en la que es necesario interpolar la serie original para obtener una serie de periodicidad anual.

Cuadro 1: Método de anualización de las distintas series y parte del período muestral afectada por la interpolación

<i>- series construidas por procedimiento 2, con corrección uniforme</i>	
VAB nominal	1955-89
Deflactor del VAB	1955-89
Empleo total (puestos de trabajo)	1955-89
Empleo asalariado	1955-89
Remuneración de asalariados	1955-89
Rentas totales del trabajo	1955-89
Población de hecho	1955-71
<i>- series obtenidas a partir de otras series tras anualizar éstas</i>	
VAB a precios constantes de 2000	1955-89
Participación del trabajo en el VAB	1955-89
Sueldo medio	1955-89
Deflactor del VAB a precios medios	1955-89
VAB a precios medios de 1955-2009	1955-2009
<i>- series construidas por pura interpolación geométrica</i>	
Población de derecho	1955-71
Población residente corregida	1955-71

3. Otros detalles

Renormalización de las series

Las series preliminares obtenidas por el procedimiento indicado arriba se reescalan proporcionalmente de forma que sean exactamente consistentes con el correspondiente agregado nacional en los años pares. En la mayor parte de los casos, el agregado nacional es simplemente la suma de las series regionales, con lo que el necesario factor de corrección se obtiene como la razón entre tal agregado y la suma de los valores preliminares de las series regionales.

La excepción es el deflactor del VAB, donde la serie nacional es una media ponderada de las series regionales en vez de su suma. Con el fin de que los cálculos necesarios se puedan realizar utilizando como ponderaciones los pesos regionales en el VAB nacional medido a precios corrientes, resulta conveniente trabajar con el inverso del deflactor del VAB, $1/P$, que puede expresarse como la razón entre el VAB a precios constantes y el VAB a precios corrientes, Q/Y . De esta forma, el inverso del deflactor del VAB nacional puede escribirse como una media ponderada de los inversos de los deflactores de los VABs regionales,

$$(11) \quad \frac{1}{P_t} = \frac{Q_t}{Y_t} = \frac{\sum_i Q_{it}}{Y_t} = \sum_i \frac{Q_{it}}{Y_t} = \sum_i \frac{Y_{it} / P_{it}}{Y_t} = \sum_i \frac{Y_{it}}{Y_t} \frac{1}{P_{it}} = \sum_i sh_i Y_t \frac{1}{P_{it}}$$

en la que las ponderaciones son las participaciones regionales en el VAB nacional a precios corrientes, $sh_i Y_t$. Tras interpolar los niveles de precios regionales por el procedimiento de corrección uniforme, la ecuación (11) se utiliza para calcular el deflactor nacional que se derivaría de las series preliminares de precios regionales. El cociente entre esta variable y el dato directo de deflactor nacional nos da el factor de ajuste que hay que aplicar a las series de precios regionales para que sean plenamente consistentes con la serie nacional.

Series a precios medios de 1955-2009

Siguiendo el procedimiento expuesto en de la Fuente (2009), he construido series de VAB a precios medios del período 1955-2009. El deflactor relevante se ha recalculado a partir de la serie completa anualizada de deflactores del VAB en base 2000, a diferencia de los trabajos anteriores en los que se utilizaban sólo los años impares de la serie. Esto hace que se produzcan pequeños cambios en los años impares de las series de deflactores y de VAB a precios medios (a diferencia del resto de las series, cuyos valores en años impares no se ven afectados por la extensión de las mismas a años pares).

Series de población

La base de datos incluye tres series de población regional: la población de hecho, la población de derecho o residente y la población residente corregida (por la parte de los residentes ausentes que se estima puedan ser realmente emigrantes). En el caso de la primera de estas series se dispone de una serie nacional de periodicidad anual (construida por Maluquer, 2008),

lo que permite aplicar el procedimiento preferido de corrección uniforme a la interpolación. En los otros dos casos no se dispone de una serie de tales características, por lo que ambas series se anualizan por interpolación pura -- esto es, las series se extienden a años pares partiendo del supuesto de que sus tasas de crecimiento se mantienen constantes durante los dos años de cada bienio.

Archivo de datos

El trabajo va acompañado de un archivo de Excel (*RegDat_v23_5509.xls*) que contiene las nuevas series anualizadas.

Referencias

de la Fuente, A. (2009). "Series enlazadas de algunos agregados económicos nacionales y regionales, 1955-2007. Versión 2.1." Mimeo, Instituto de Análisis Económico, CSIC, Barcelona.

http://fedea.es/033_Publicaciones_VerMas.asp?id=861

de la Fuente, A. (2010a). "Un enlace alternativo de los agregados de VAB y empleo entre las bases 1995 y 2000 de la Contabilidad Nacional y Regional." *Estadística Española* 52 (173), pp. 67-90.

de la Fuente, A. (2010b). "Series enlazadas de empleo asalariado y rentas del trabajo regionales (RegDat versión 2.2)." Mimeo, Instituto de Análisis Económico, CSIC, Barcelona.

Fundación BBV (FBBV, 1999). *Renta nacional de España y su distribución provincial. Serie homogénea. Años 1955 a 1993 y avances 1994 a 1997*. Bilbao.

Fundación BBVA (FBBVA, 2000). *Renta nacional de España y su distribución provincial. Año 1995 y avances 1996-1999*. Bilbao.

Instituto Nacional de Estadística (INE, 2010). *Contabilidad Regional de España*. En Base de datos electrónica INEbase. Economía: Cuentas Económicas. Madrid.

Maluquer, J. (2008). "El crecimiento moderno de la población en España de 1850 a 2001: una serie homogénea anual." *Investigaciones de Historia Económica* 10, pp. 129-62.