

Series enlazadas de algunos agregados económicos regionales, 1995-2007

Versión 1.1

Angel de la Fuente*

Instituto de Análisis Económico (CSIC)

Noviembre de 2008

Resumen

En este trabajo se construyen series “homogéneas” de VAB a precios corrientes y constantes, empleo y población para las regiones españolas y Ceuta y Melilla que cubren el periodo 1955-2007 a intervalos bienales. Estas series se obtienen enlazando la Contabilidad Regional del INE con las series elaborada por Julio Alcalde y colaboradores para la Fundación BBVA. El “punto de corte” en el que se abandona esta última fuente como referencia para la construcción de la serie enlazada se determina utilizando un procedimiento que permite estimar cuál de las dos series disponibles generará un estimador con menor error cuadrático medio cuando se utiliza como variable dependiente en una regresión sobre una variable independiente arbitraria.

Abstract

I construct “homogeneous” series of GVA at current and constant prices, employment and population for the Spanish regions plus Ceuta and Melilla covering the period 1955-2007 at biennial intervals. The series are obtained by linking the Regional Accounts of the National Statistical Institute with the series constructed by Julio Alcaide and his team for the BBVA Foundation. The “switching point” at which this last source stops being used as a reference to construct the linked series is determined using a procedure that allows me to estimate which of the two competing series would produce an estimator with the lowest MSE when it is used as dependent variable in a regression on an arbitrary independent variable.

* Este trabajo forma parte de un proyecto de investigación cofinanciado por el Fondo Europeo de Desarrollo Regional y la Fundación Caixa Galicia. Agradezco también la financiación del Ministerio de Ciencia e Innovación a través del proyecto SEJ 2005-06357 y los comentarios, sugerencias y ayuda con los datos de Antonio Díaz, Andrés de Bustos y Javier Escibá.

1. Introducción

Una de las dificultades con las que nos enfrentamos los analistas de la economía regional española es la inexistencia de series homogéneas comúnmente aceptadas de los principales agregados económicos regionales para períodos largos. Las únicas series existentes que se remontan a 1955 son las elaboradas por Julio Alcaide y diversos colaboradores, primero para el Servicio de Estudios del Banco de Bilbao y posteriormente para la Fundación BBVA y la Fundación FIES. Por otro lado, disponemos también de la Contabilidad Regional de España (CRE) que elabora el Instituto Nacional de Estadística (INE) desde 1980. La CRE ha sufrido importantes cambios metodológicos a lo largo del tiempo, sin que esté disponible hasta el momento una serie histórica homogénea oficial para el conjunto del período. Ambas fuentes, además, presentan diferencias muy apreciables durante su período de solapamiento.

El presente trabajo es un informe provisional sobre un proyecto en curso en el que se comparan las distintas series disponibles y se intenta elaborar series históricas "homogéneas" mediante el enlace entre las últimas series disponibles de Contabilidad Regional (base 2000) y otras fuentes entre las que destacan las series de la Fundación BBVA. El autor agradecería cualquier comentario que pueda ayudar a mejorar la calidad de las series enlazadas, incluyendo sugerencias sobre fuentes estadísticas adicionales que puedan ser de interés, los resultados de intentos previos de homogenización o cualquier otra información relevante sobre la fiabilidad de las distintas series consideradas.

El objetivo del ejercicio que aquí se presenta ha sido el de construir series largas de algunos agregados regionales (VAB a precios corrientes y constantes, empleo y población) que enlacen con los datos más recientes de Contabilidad Regional en base 2000 (CRE00) a partir de 1995. El enlace se ha hecho con el objetivo de maximizar el contenido informativo de los datos con vistas a su utilización en el tipo de ejercicios estadísticos más habituales en la literatura, como puede ser la estimación de funciones regionales de producción.

Esto quiere decir dos cosas. Primero, que para aquellos períodos anteriores a 1995 en los que existen varias fuentes alternativas, la selección de la fuente de referencia que utilizaré para extender hacia atrás las series de la CRE00 se ha basado en un indicador estadístico de la calidad de los datos. Este indicador se construye a partir del siguiente experimento conceptual. Supongamos que queremos analizar los determinantes de una de las variables con las que estamos trabajando (renta o empleo) regresándola sobre alguna otra variable, x . ¿Cuál de las series disponibles generará un estimador con mejores propiedades y, en particular, con un menor error cuadrático medio?

La segunda implicación de mi preocupación por la fiabilidad de los resultados de determinados ejercicios estadísticos es que me centraré en las tasas de crecimiento de las distintas series en vez de en sus niveles. Esto es, el indicador de calidad se estimará con las series en primeras diferencias y las series enlazadas se construirán "retropolando" las series de CRE00, utilizando en cada período la tasa de crecimiento de la serie que corresponda, sin intento alguno (al menos por el momento) de ajustar los niveles resultantes de forma que se respeten sus valores

originales en determinados años base. He elegido esta opción fundamentalmente porque en las especificaciones econométricas más habituales en la literatura de referencia (con los datos en diferencias o introduciendo efectos fijos) los parámetros del modelo se identifican en base a la variación temporal de las series y no a sus niveles, lo que no hace aconsejable introducir ruido adicional en la primera para intentar corregir los segundos (salvo que existan muy buenas razones para ello).

Por otra parte, la insistencia en mantener inalterables los valores en anteriores años base a la hora de realizar un enlace entre dos series podría ser bastante discutible, especialmente si el período transcurrido desde el año base no es muy largo. La razón es que respetar el valor de la vieja serie en su año base implica distribuir toda la discrepancia que observamos entre la nueva y la vieja serie en su primer año común sobre el período comprendido entre esta fecha y el antiguo año base. Implícitamente, por tanto, estamos suponiendo que el “error” en la vieja serie se ha generado únicamente a partir de su año base, lo que no parece muy plausible, especialmente si este período es corto. Un supuesto alternativo que podría ser más razonable en muchos casos es que el error en la vieja serie tiene un peso constante a lo largo del tiempo. Si éste es el caso, la vieja serie contendrá un error de nivel, pero su tasa de crecimiento reflejará correctamente la de la variable subyacente. El procedimiento correcto de enlace bajo estos supuestos sería el que aquí se utiliza: extender la nueva serie hacia atrás utilizando las tasas de crecimiento de la vieja serie, sin corregir los niveles resultantes.

El trabajo está organizado como sigue. En la sección 2 se describe la información disponible y el procedimiento adoptado para fijar la “fecha de corte” en la que abandonaré las series del BBVA y adoptaré las del INE como referencia para construir la serie enlazada. La sección 3 describe algunas tareas preliminares, incluyendo la construcción de la serie enlazada de población y la normalización de los datos de VAB nominal y empleo. En la sección 3 se realiza el enlace de estas dos últimas series y en la sección 4 el de los índices de precios y el VAB a precios constantes. La sección 5 concluye con un listado de tareas pendientes para futuras extensiones del trabajo. El Anexo 1 contiene los detalles de la construcción del indicador de calidad relativa de las series que se utiliza para fijar la fecha de corte y el Anexo 2 una serie de gráficos en los que se comparan las distintas series existentes y la nueva serie enlazada.

2. Información disponible y planteamiento general

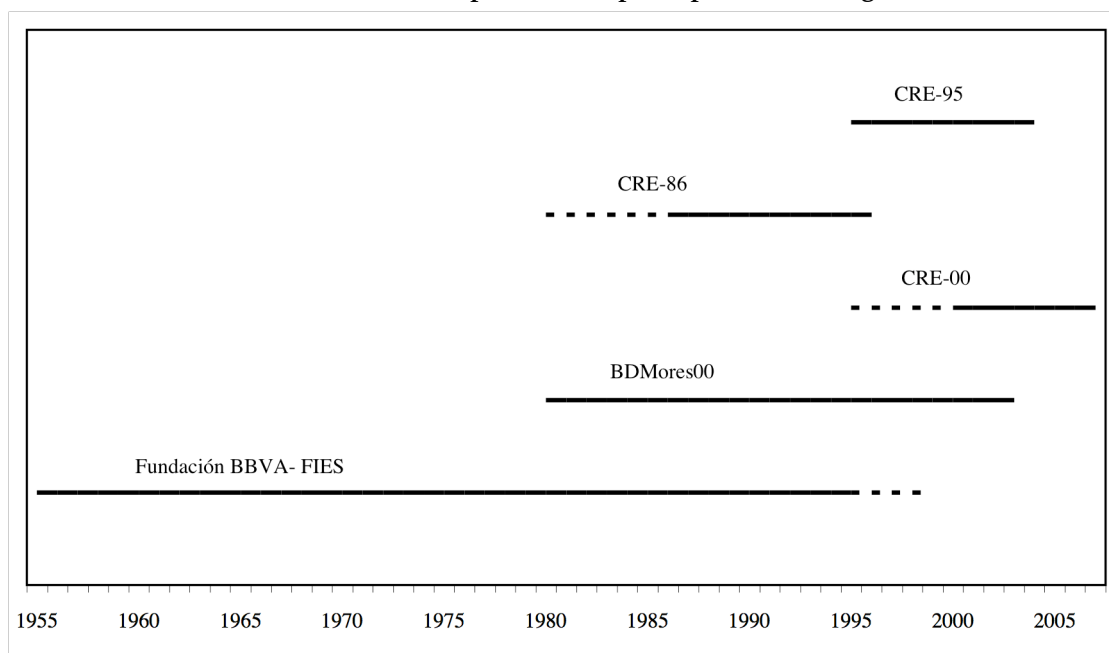
El Gráfico 1 resume la cobertura temporal de las fuentes disponibles para las variables de interés. La serie homogénea elaborada en Fundación BBV (1999) cubre el período 1955-93 a intervalos bienales. Esta serie se extiende hasta 1995 en Fundación BBVA (2000), donde también se ofrecen avances de las magnitudes relevantes hasta 1999.

El INE ha elaborado cuatro series diferentes de Contabilidad Regional con bases 1980, 1986, 1995 y 2000. Además del cambio en el año base, las nuevas series han incorporado cambios metodológicos para adaptarlas a las sucesivas directrices internacionales en materia de contabilidad nacional y regional así como información adicional proveniente de nuevas encuestas y otras fuentes. Las dos primera y las dos últimas series han sido enlazadas por el propio INE para construir series “homogéneas” para 1980-96 en base 1986 y 1995-2007 en base

2000.¹ En el primer caso, sin embargo, el enlace parece haber sido sólo parcial, pues la mayor parte de las magnitudes que se ofrecen en el sitio web del INE comienzan en 1986 (la excepción es el PIB).

Finalmente, en el gráfico se incluyen también las series recogidas en la base de datos BDMores en base 2000 (de Bustos et al, 2008), donde se construye un cuidadoso enlace entre las diversas series de Contabilidad Regional. Aunque casi no utilizaré estos datos en lo que sigue, la serie enlazada que aquí se construye se compara también con la BD Mores en el Anexo 2 con el fin de que los potenciales usuarios de las distintas bases de datos puedan hacerse una idea de las diferencias entre ellas.

Gráfico 1: Cobertura temporal de las principales series regionales

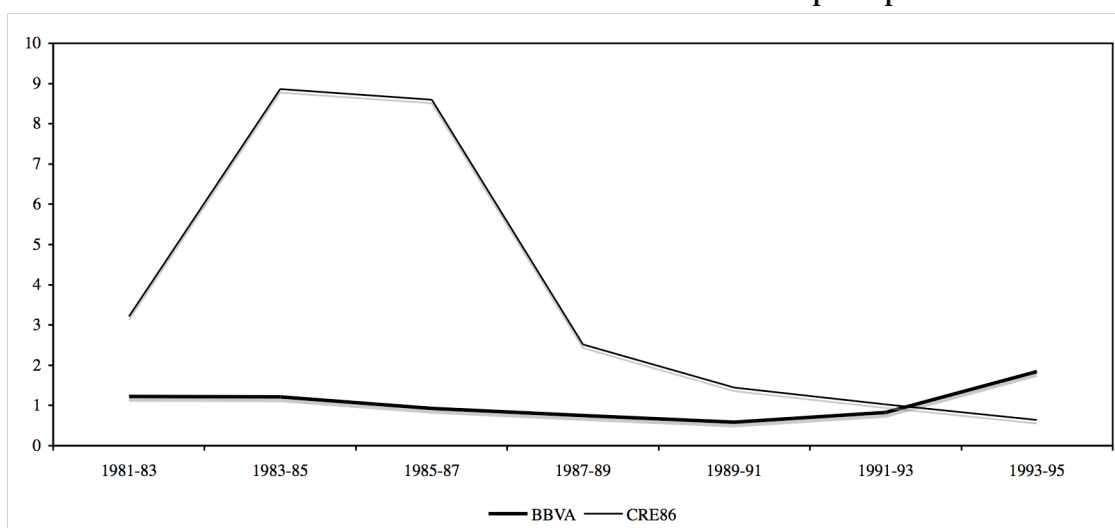


Tomando como dadas las series enlazadas de Contabilidad Regional para el período 1995-2007, el problema al que nos enfrentamos es el de cómo extenderlas hacia atrás de manera óptima y en particular, cuál de las distintas fuentes disponibles utilizar como referencia para ello en cada período. Dado que el INE dispone en principio de mucho mejores medios para estimar la renta y el empleo regionales que un servicio de estudios privado parece que lo más lógico sería utilizar las distintas series de CRE disponibles desde 1980 y recurrir a las de la Fundación BBVA únicamente para el período anterior, cuando éstas son la única fuente disponible.

Un examen preliminar de los datos, sin embargo, sugiere que ésta podría no ser la mejor forma de proceder. La razón es que, tal como se muestra en el Gráfico 2, las series del INE presentan durante sus primeros años una elevada volatilidad que las hace un tanto sospechosas. Este hecho sugiere que podría ser buena idea elegir una “fecha de corte” algo más tardía para abandonar la serie del BBVA y pasar a utilizar los datos del INE como referencia para construir la serie enlazada.

¹ En la versión de esta serie que aquí se utiliza, los datos de 2004 y 2005 son provisionales, el de 2006 es un avance y el de 2007 una primera estimación.

Gráfico 2: Varianza de las diferencias anualizadas del PIB per cápita relativo



- Nota: El PIB per cápita regional es un índice con España = 100. Las diferencias anualizadas se calculan como $(y_{it+2} - y_{it})/2$ donde y_{it} es el PIB per cápita de la región i en el año t , normalizada por el promedio nacional. Se utilizan datos bienales porque la serie del BBVA sólo está disponible para años impares.

Una posible explicación del patrón que observamos en el Gráfico 2 es que los primeros años de la serie del INE podrían contener una elevada cantidad de ruido debido a las lógicas dificultades que comporta la puesta en marcha de una nueva serie estadística (a lo que hay que añadir en el caso del PIB la complicación que supuso la introducción del IVA en 1986). Existe, sin embargo, una explicación alternativa: las series de la CRE podrían ser más volátiles que las del BBVA no porque contengan más ruido que ellas, sino porque capturan mejor la volatilidad real de la economía. El perfil temporal de las volatilidades que observamos en el gráfico sugiere que no es probable que esta última sea la única explicación porque en caso contrario esperaríamos que (otras cosas iguales) la volatilidad de la serie del BBV fuese siempre menor y no es así. Con todo, conviene no perder de vista esta posibilidad aunque sólo sea como una forma de modelizar el plus de credibilidad que en principio parece lógico atribuir a las series del INE.

En este contexto, el problema que se nos plantea se puede reformular de la manera siguiente. Sea y_i la (tasa de crecimiento de) la renta relativa por habitante en el territorio i en un período determinado. Supongamos que queremos estimar una ecuación de la forma

$$(1) y_i = \beta x_i + u_i$$

con datos de corte transversal de un año dado para analizar el efecto sobre el crecimiento de la renta de alguna otra variable, x . Dado que disponemos de dos estimaciones alternativas de y (las series de CRE86 y BBVA) ¿cuál de ellas deberíamos utilizar?

Para responder esta pregunta resulta útil tratar a las dos series de referencia como "proxies ruidosas" de la variable subyacente, y , en el marco de un modelo de error de medición. Para recoger las dos hipótesis indicadas más arriba, supondré que la serie de CRE sólo contiene un error de medición clásico, mientras que la serie del BBVA podría sufrir además de un sesgo

sistemático hacia la media muestral (que es cero tal como he normalizado las variables). Tenemos, por tanto,

$$(2) y_i^{CRE} = y_i + e_i^{CRE}$$

$$(3) y_i^{BBV} = (1 - \phi)y_i + e_i^{BBV}$$

donde $\phi \in [0,1]$ y e_i^{CRE} y e_i^{BBVA} son errores de medición "puros" distribuidos idéntica e independientemente con media cero y varianzas σ_{CRE}^2 y σ_{BBV}^2 respectivamente y no correlacionados con x e y o entre sí.

El caso más sencillo es aquel en el que las dos series contienen sólo un error clásico de medición ($\phi = 0$). Bajo esta hipótesis, la serie con menor volatilidad es siempre preferible porque una mayor varianza indica necesariamente un mayor contenido de ruido y cuánto mayor sea éste, menor será la precisión con la que podemos estimar el parámetro de interés.

Cuando $\phi > 0$ la situación se complica porque una menor varianza ya no equivale a menos ruido y porque el sesgo en la medición de y genera además un sesgo en la estimación de β . El criterio natural en este caso es la minimización del error cuadrático medio (*mean squared error*, MSE) del estimador de β . Como veremos enseguida, bajo ciertos supuestos es posible obtener una estimación consistente del MSE del estimador de β que se obtendría con cada una de las series, pero únicamente si conocemos el valor de ϕ , lo que no es el caso. Sin embargo, también es posible estimar (i) el sesgo mínimo que tendrían que tener los datos del BBVA para que la serie de CRE pudiese ser preferible para alguna variable x , y (ii) una cota superior para ϕ . La combinación de ambas estimaciones nos da una condición suficiente pero no necesaria para establecer la superioridad de la serie del BBVA ("en valor esperado"). Como veremos más adelante, esto no elimina por completo la necesidad de algún juicio subjetivo de valor pero ciertamente permite acotar el problema.

Sean MSE_{BBV} y MSE_{CRE} los errores cuadráticos medios de los estimadores de β que se obtienen estimando por MCO la ecuación (1) con cada una de las series de referencia. Como se muestra en el Anexo 1, la función $F()$ dada en

$$(4) \quad \begin{aligned} \frac{n-1}{\beta^2} (\widehat{MSE}_{BBV} - \widehat{MSE}_{CRE}) &= F(\phi, R^2(x)) \\ &= (n-2)\phi^2 + \left(2 + \frac{1}{R^2(x)} \left(\frac{\hat{r}_{BBV} - \hat{r}_{CRE}}{\hat{r}_{BBV}\hat{r}_{CRE}} \right) \right) \phi - \frac{1}{R^2(x)} \left(\frac{\hat{r}_{BBV} - \hat{r}_{CRE}}{\hat{r}_{BBV}\hat{r}_{CRE}} \right) \\ &\equiv (n-2)\phi^2 + \left(2 + \frac{A}{R^2(x)} \right) \phi - \frac{A}{R^2(x)} \end{aligned}$$

nos da un estimador consistente de la diferencia entre estas magnitudes. En esta expresión, $n = 18$ es el número de observaciones en el corte transversal (las 17 comunidades autónomas más Ceuta y Melilla), $R^2(x)$ el R^2 de la regresión (1) si ésta se pudiese estimar directamente con la variable y "correctamente medida" y \hat{r}_{BBV} y \hat{r}_{CRE} los valores estimados de las *ratios de fiabilidad* de

las dos series, que se obtienen regresando cada una de ellas sobre la otra con datos de corte transversal para el período de interés (véase el Anexo 1).

Inspeccionando la ecuación (4) vemos que si $\hat{r}_{CRE} > \hat{r}_{BBV}$ entonces $F() > 0$ para todo ϕ y x , lo que implica que la serie del INE es siempre la mejor. Bajo el supuesto contrario (que es el que nos interesa aquí dado que es el más habitual en los datos durante el período muestral), es fácil comprobar que $F()$ es una función creciente en ϕ y en R^2 en todo su dominio $[0,1] \times [0,1]$ con

$$F(0, R^2(x)) = -\frac{A}{R^2(x)} < 0$$

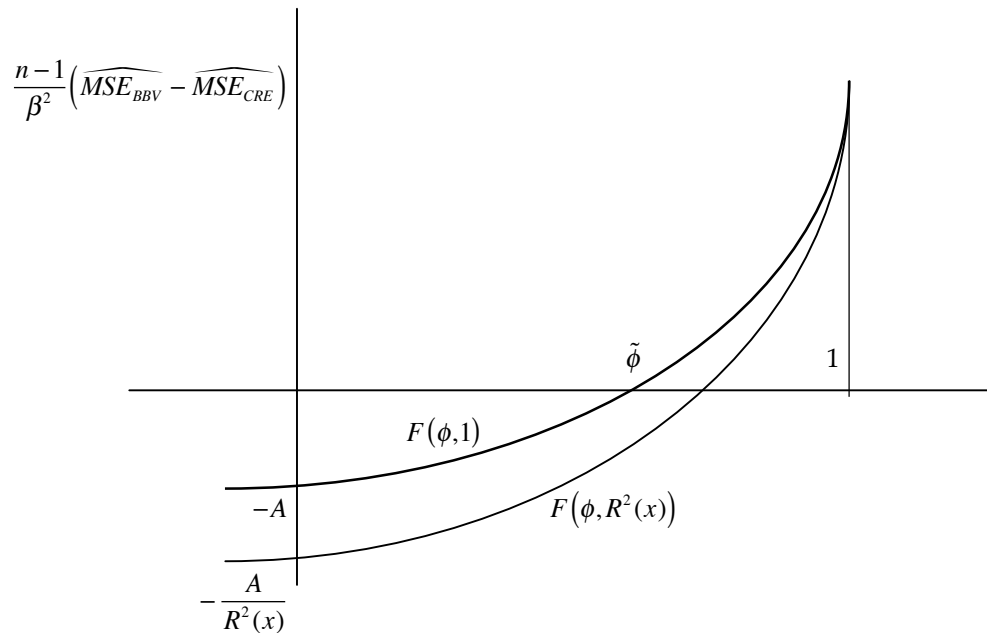
$$(5) F(1, R^2(x)) = n > 0$$

$$F(\phi, R^2(x)) \leq F(\phi, 1) = (n-2)\phi^2 + (2+A)\phi - A$$

tal como se ilustra en el Gráfico 3. Por tanto, si la serie del BBVA tiene una mayor ratio de fiabilidad entonces es preferible a la del INE siempre que no presente un sesgo sistemático ($\phi = 0$) y continúa siéndolo para valores suficientemente bajos de ϕ . En términos generales, el valor de ϕ que iguala los MSEs obtenidos con las dos series depende de x , pero existe un valor crítico de este parámetro, $\tilde{\phi}$ tal que la serie del BBVA es preferible para cualquier variable x siempre que $\phi < \tilde{\phi}$. Este valor es el mayor de los ceros de la función cuadrática $F(\phi, 1)$ y viene dado por

$$(6) \tilde{\phi} = \frac{-(2+A) + \sqrt{(2+A)^2 + 4(n-2)A}}{2(n-2)}$$

Gráfico 3: La función $F()$



Alternativamente, podemos también fijar un valor "razonable" de $R^2(x)$ para obtener una estimación de $\tilde{\phi}$ bajo un criterio algo menos exigente.

Por otra parte, también se muestra en el Anexo 1 que es posible estimar una cota superior para ϕ utilizando

$$\hat{\phi}_{\max} = 1 - \max_t \hat{r}_{CRE,t}$$

donde $\hat{r}_{CRE,t}$ es el ratio de fiabilidad estimado para la serie CRE86 en el año t . Puesto que $\phi \leq \hat{\phi}_{\max}$, una condición suficiente (en "valor esperado") pero no necesaria para que la serie del BBVA sea preferible a la del INE en un año dado es que $\hat{\phi}_{\max} \leq \tilde{\phi}$.

En lo que sigue, calcularé los valores de $\tilde{\phi}$ y $\hat{\phi}_{\max}$ en cada año para distintos agregados regionales (de renta y empleo) y los utilizaré de guía para intentar determinar una fecha razonable de corte. Como cabría esperar a la vista del Gráfico 2, el procedimiento sugiere que no es aconsejable utilizar los primeros años de la serie del INE. En consecuencia, el cambio en las series de referencia se pospondrá hasta finales de los años ochenta.

3. Algunas cuestiones preliminares

En esta sección se describen algunas tareas previas a la construcción del enlace entre las series de interés. Puesto que he decidido trabajar con las series en términos per cápita, el primer paso consiste en construir una serie enlazada de población que se utilizará para normalizar todas las demás. Una vez hecho esto, se procede a la construcción de una serie enlazada "completa" de CRE86 para el período 1980-95, extendiendo hacia atrás las series que el INE sólo suministra a partir de 1986 utilizando otra información elaborada por el propio INE. Esta serie enlazada CRE86 es la que se confronta seguidamente con la del BBVA para realizar el enlace final.

3.1. Construcción de la serie enlazada de población

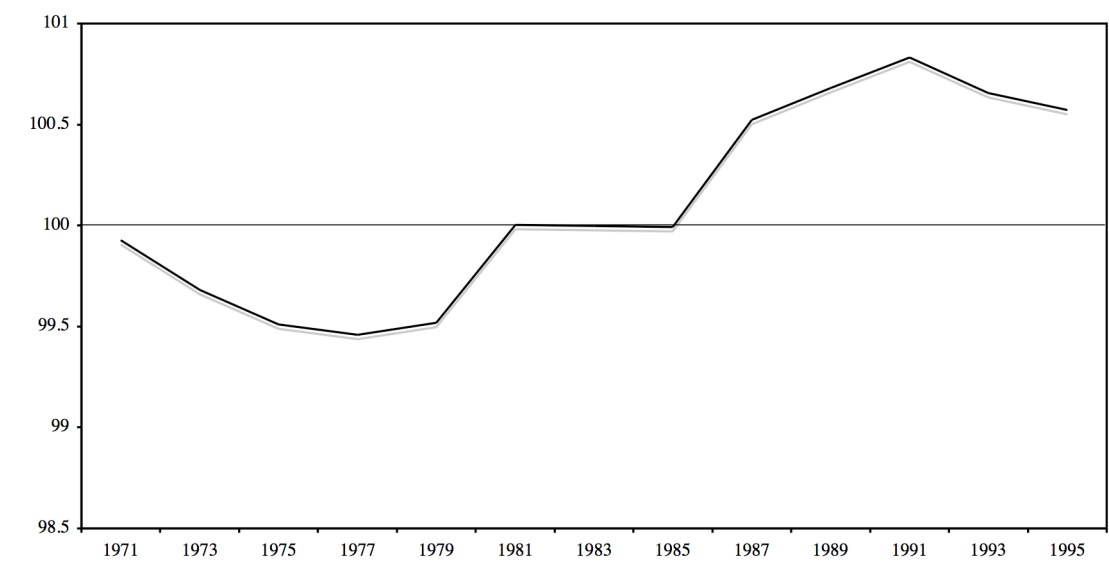
En principio, tanto la CRE como la Fundación BBVA utilizan el mismo agregado poblacional (población residente o de derecho a 1 de julio) y parten de datos del INE. El BBVA, sin embargo, introduce algunas correcciones en estas series (sobre las que no ofrece detalles), lo que hace las dos fuentes no coincidan exactamente (véase el Gráfico 4). En cualquier caso, las discrepancias entre ellas son menores.

Las series de referencia del INE en materia de población son las estimaciones intercensales de población, que cubren el período 1971-2001, y las estimaciones de la población actual a partir del Censo de 2001, que extienden la serie anterior hasta el presente. Para años anteriores a 1971 no se dispone (al menos en la página web del INE) de series anuales sino únicamente de datos referidos a años censales.

La serie enlazada de población se construye como se indica en el Cuadro 1. Desde 1971 en adelante, se utilizan directamente las series del INE. Para períodos anteriores, la serie del INE

para cada región se extiende desde 1971 hacia atrás utilizando las tasas de crecimiento de la correspondiente serie de población del BBVA.

Gráfico 4: Población nacional según el BBVA (dato del INE = 100)



Cuadro 1: Construcción de la serie enlazada de población residente (de derecho) a 1 de julio

<i>Período</i>	<i>Fuente:</i>
2002-2008	INE, estimaciones de la población actual a partir del Censo de 2001
1971-2001	INE, estimaciones intercensales de población Extendida hacia atrás utilizando las tasas de crecimiento de:
1955-1971	Fundación BBV (1999)

3.2. Normalización de las series

Buena parte de los cálculos que se presentan en lo que sigue se han realizado con series normalizadas. Las series de VAB a precios corrientes y empleo se expresan en primer lugar en términos per cápita dividiéndolas por la serie enlazada de población descrita en el apartado anterior y se transforman a continuación en índices, normalizando cada magnitud por su promedio nacional en el mismo año.

Los enlaces se realizan en primera instancia con las series así normalizadas, comenzando por el final del período y yendo hacia atrás. En cada punto de enlace, la serie "homogénea" se extiende hacia atrás utilizando las tasas de crecimiento de la serie elegida como referencia (también normalizada). El resultado de este cálculo preliminar se renormaliza de forma que el promedio nacional calculado a partir de las nuevas series regionales sea exactamente igual a 100 en cada período.

El cálculo de los MSEs de las series se realiza con las primeras diferencias temporales de las series estandarizadas. Puesto que los datos del BBVA sólo están disponibles en años alternos, he trabajado sólo con los años impares de todas las series. Las primeras diferencias de la serie

bienal se dividen por dos para obtener las "diferencias anualizadas" con las que se realizan los cálculos.

3.2. Construcción de la serie enlazada de CRE86 para 1980-95

Las series de empleo y de VAB a coste de los factores (cf) a precios corrientes de la CRE86 que ofrece el INE en su página web comienzan en 1986. Para construir la serie completa 1980-95 que se confronta con la del BBVA, he extendido estas series hacia atrás utilizando las tasas de crecimiento de las magnitudes más cercanas disponibles en la serie homogénea 1980-89 en base 1986 publicada en su día por el INE.² Puesto que el VAB a coste de los factores no está disponible en esta fuente, esta magnitud se extiende hacia atrás utilizando la tasa de crecimiento del VAB a precios (corrientes) de mercado. Como ya se ha indicado, el enlace se realiza con las series estandarizadas de la forma descrita en el apartado anterior.

4. Series de VABcf a precios corrientes y empleo

En esta sección se comparan las series de VAB cf a precios corrientes y de empleo de CRE86 y BBVA y se construye una serie "homogénea" para ambas magnitudes enlazando la "mejor" de ambas con la serie de CRE00. En ambos casos el empleo se mide en puestos de trabajo "brutos", sin ponderar por el número de horas trabajadas.

He decidido trabajar con el VAB a coste de los factores/precios básicos (y no con el PIB) porque el primero es el agregado de renta menos afectado por los cambios metodológicos que ha experimentado la contabilidad regional y, puesto que excluye el IVA, también por las modificaciones de nuestro sistema fiscal. Los Gráficos A1.1-A.1.18 del Anexo 2 comparan la evolución de las distintas series de empleo y de VAB regional a precios corrientes por habitante y muestran también la serie enlazada cuya construcción se describirá más adelante.

4.1. Comparación de las dos fuentes

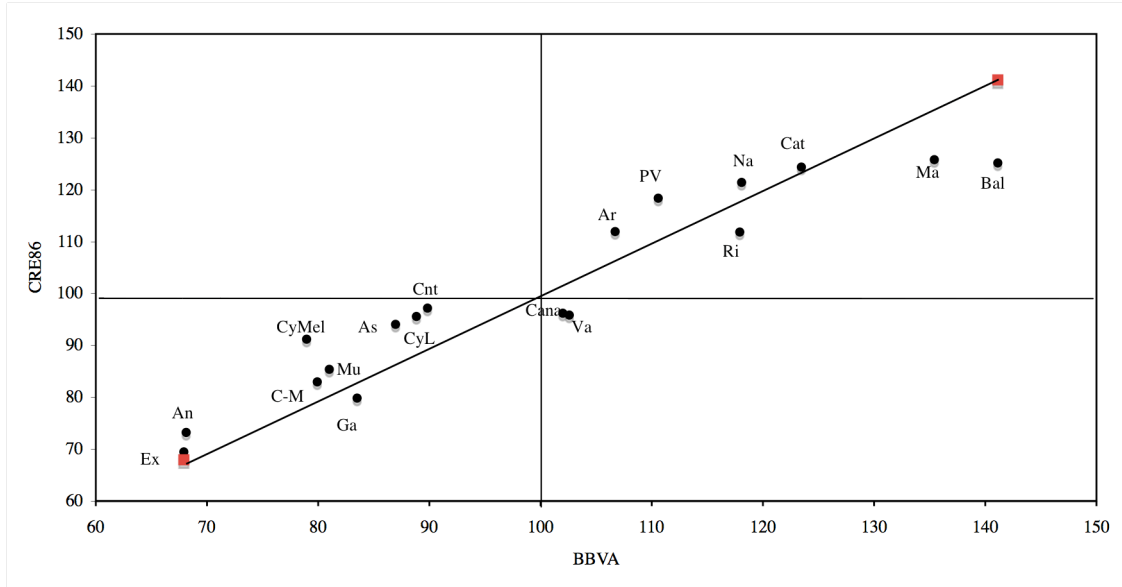
En el Gráfico 5 se comparan las estimaciones de VAB y empleo per cápita en 1995 que ofrecen el BBVA y el INE. Si ambas fuentes coincidiesen exactamente, todos los puntos deberían situarse sobre la línea recta que se muestra en el gráfico (la "diagonal"). Cuando no es así, la distancia vertical entre cada punto y la diagonal nos da la diferencia entre la estimación del INE y la del BBVA.

Como cabría esperar, la correlación entre las dos series es positiva y elevada (0.95 para el VAB y 0.85 para el empleo). Sin embargo, las diferencias entre ambas son muy apreciables en algunos casos. Los más llamativos son el País Vasco, Madrid y Baleares. El INE es mucho más "pesimista" que el BBVA relación con las dos últimas comunidades, tanto en relación con la renta como con el empleo, y bastante más optimista en lo que concierne a la primera de ellas. En el caso de Baleares, la diferencia entre ambas fuentes está en torno a los veinte puntos, tanto en renta como en empleo.

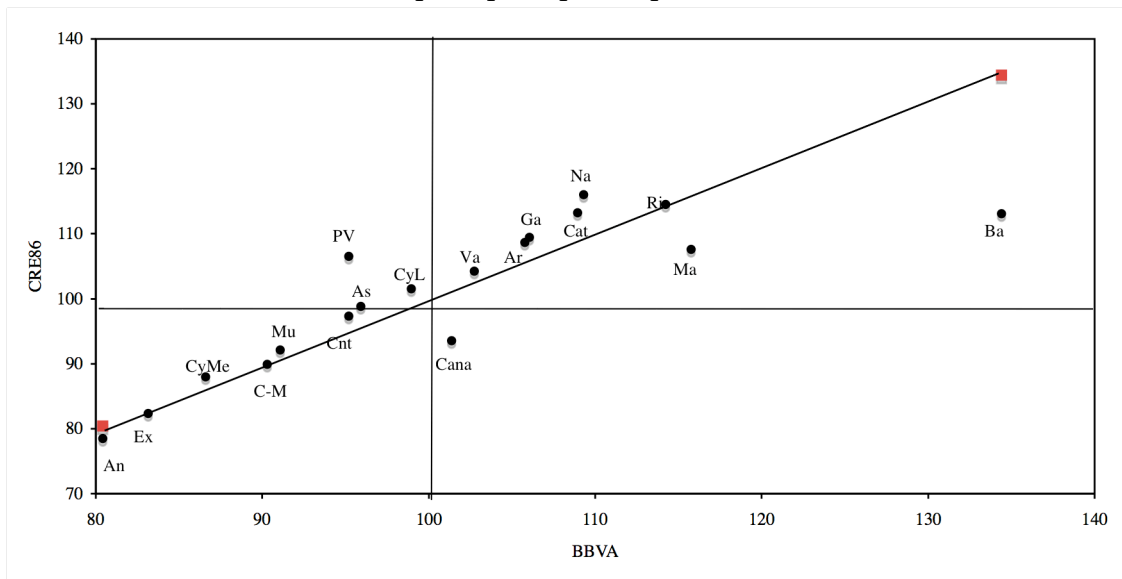
² Agradezco a Andrés de Bustos el haberme proporcionado estas series. (Pedir referencia exacta).

Gráfico 5: CRE86 vs. BBVA en 1995

a. VABcf per cápita (España = 100)



b. Empleos per cápita (España = 100)



Las discrepancias entre la CRE86 y el BBVA se han ido reduciendo a lo largo del período analizado. Entre 1981 y 1995, la distancia media entre las dos series (medida por el promedio de los valores absolutos de las diferencias entre las mismas en cada región) se redujo en un 23% en el caso del empleo por habitante y en un 18.8% en el del VAB per cápita (trabajando siempre con variables estandarizadas). (Véase el Gráfico 6). Al mismo tiempo, la correlación entre las series aumentó en un 4.8% y un 7.5% respectivamente (Gráfico 7).

Gráfico 6: Distancia media en niveles (estandarizados) entre BBVA y CRE86

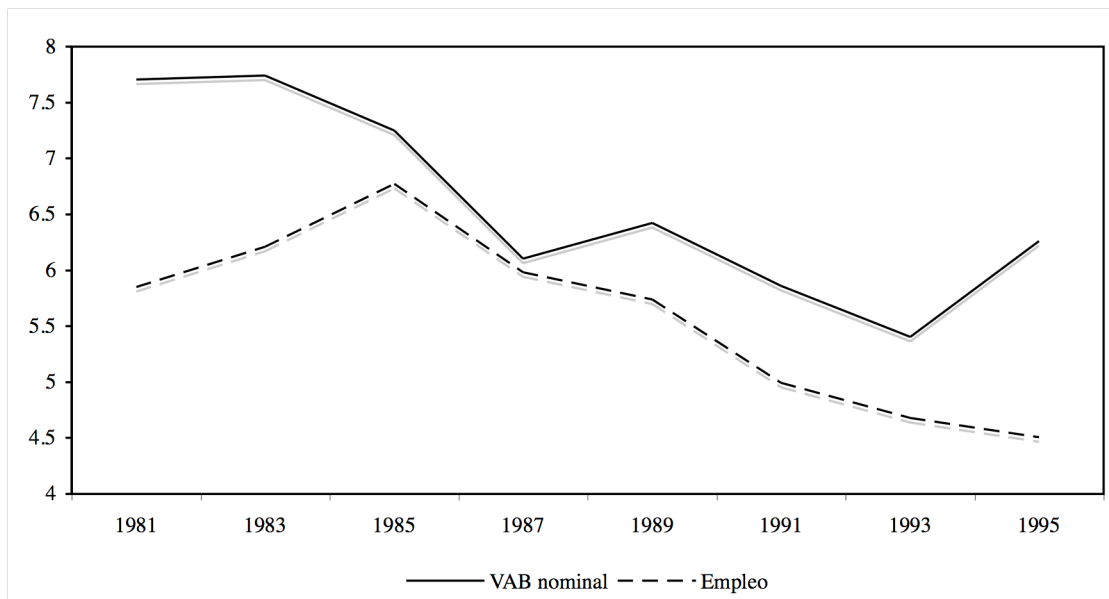
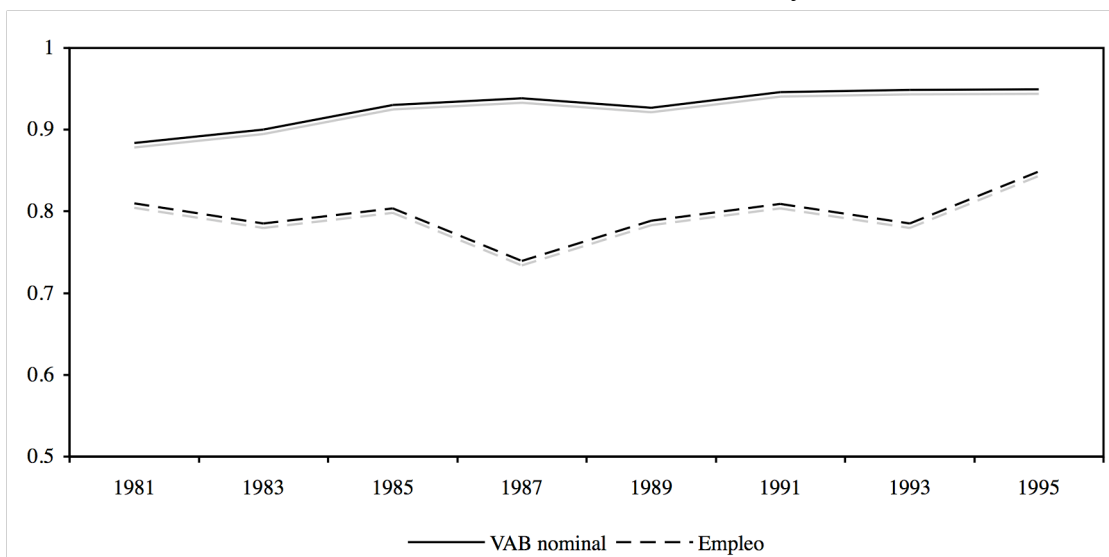
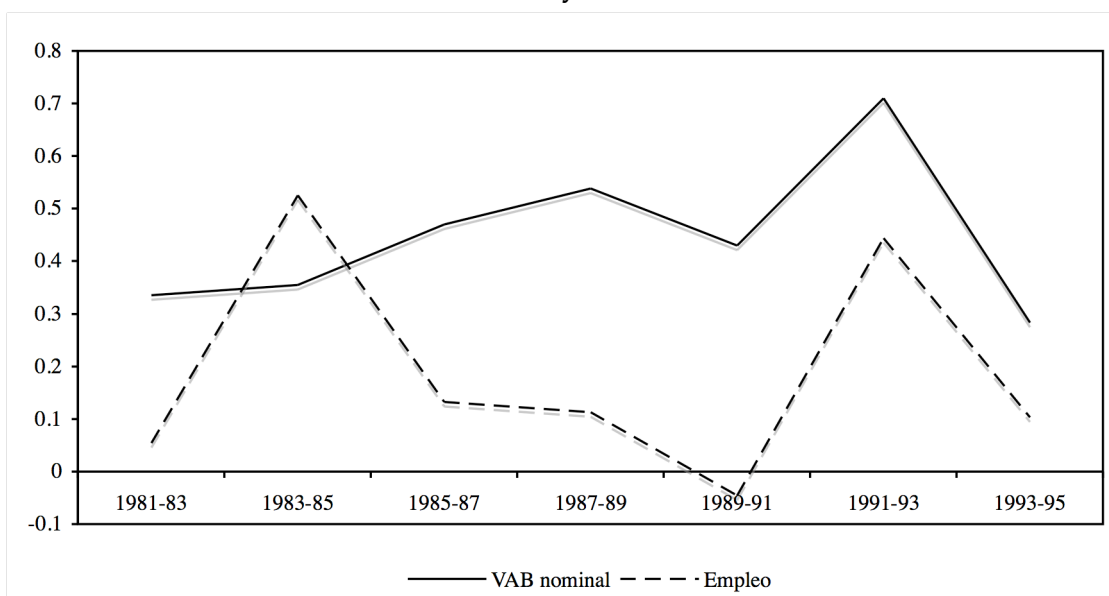


Gráfico 7: Correlación en niveles entre BBVA y CRE86



Si consideramos las primeras diferencias de las series en vez de sus niveles, el grado de coincidencia entre los datos del BBVA y los de Contabilidad Regional es significativamente menor. La correlación entre las diferencias del empleo es generalmente inferior a 0.5 y llega a ser negativa en el bienio 1989-91. En el caso del VAB, la correlación es más elevada pero aún así no excede el 0.7. (Véase el Gráfico 8).

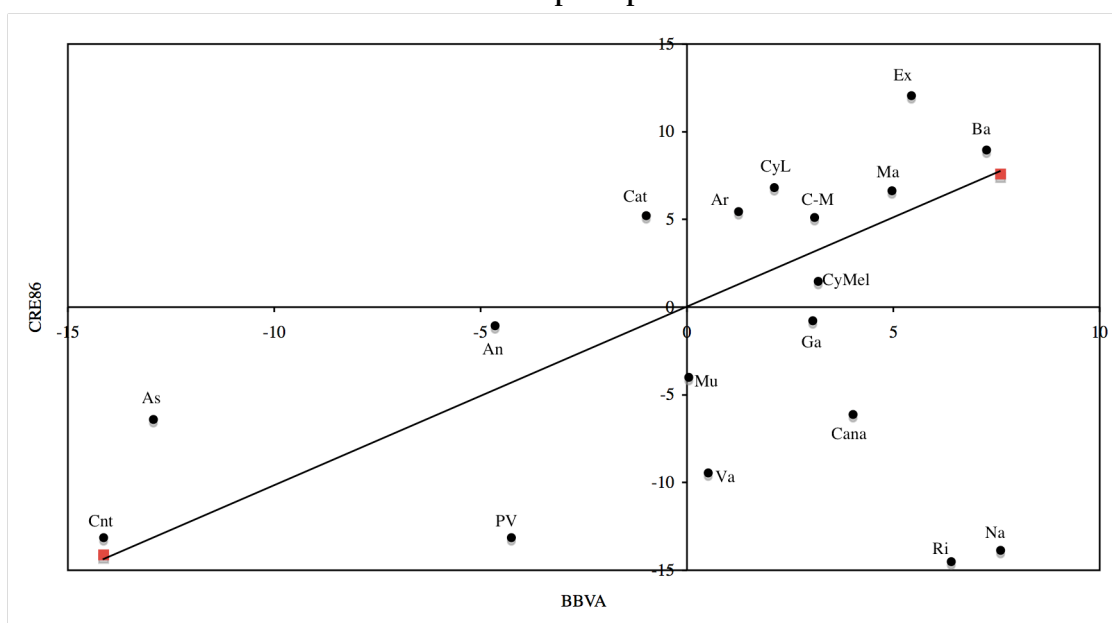
Gráfico 8: Correlación entre BBVA y CRE86 en diferencias anualizadas



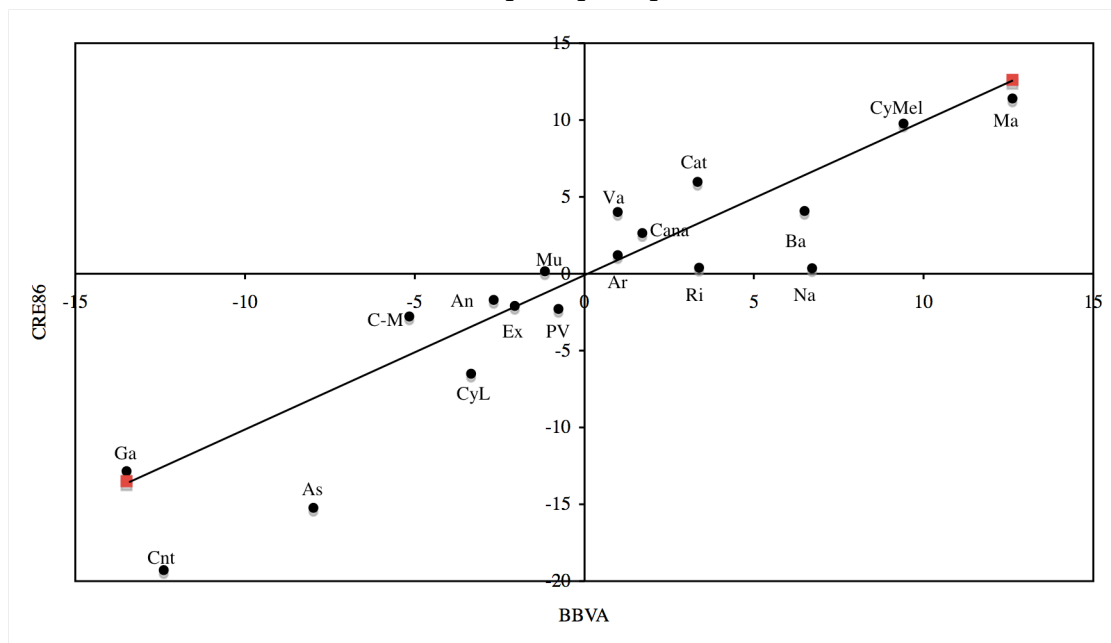
Sorprendentemente, cuando consideramos los cambios acumulados durante el conjunto del período, las tornas se invierten: la correlación entre ambas fuentes es de 0.92 para el empleo y sólo de 0.35 para el VAB per cápita. (Véase el Gráfico 9). En el caso de esta última variable, algunas diferencias son muy llamativas: más de veinte puntos en la Rioja y Navarra y más de diez en Valencia y Canarias. En siete casos, además, las dos fuentes ni siquiera coinciden sobre el signo del cambio registrado en el VAB per cápita relativo durante el período.

Gráfico 9: CRE86 vs. BBVA, incremento acumulado entre 1981 y 1995

a. VAB per cápita



b. Empleos per cápita



4.2. Selección del punto de corte

Retomando el hilo central del trabajo, para construir la serie enlazada necesitamos determinar la fecha óptima en la que abandonar las series del BBVA y adoptar como referencia la CRE86. En principio, esta fecha podría ser distinta para cada una de las series que estamos considerando. El método adoptado también permite la posibilidad de ir alternando las dos fuentes de acuerdo con los resultados que obtengamos en cada bienio. Sin embargo, he preferido excluir ex-ante estas dos posibilidades y elegir una única fecha de corte común para todas las series. El combinar series de VAB y empleo de fuentes diferentes parece arriesgado porque, aunque ninguna de las dos fuentes da demasiados detalles sobre sus procedimientos de estimación, uno tiende a pensar que las dos magnitudes no se estiman de forma independiente. Por otra parte, saltar de una serie a otra sería un tanto inconsistente con la intuición básica que guía el ejercicio: que hay problemas con los primeros años de la serie del INE pero que éstos seguramente se han ido corrigiendo con el paso del tiempo.

El Gráfico 10 muestra (en escala logarítmica) el ratio entre las varianzas de las diferencias anualizadas del VAB y el empleo per cápita en nuestras dos fuentes, con el INE en el numerador. El patrón temporal de las volatilidades relativas es similar al observado en el Gráfico 2 para el caso del PIB per cápita: con una única excepción puntual (el empleo en el bienio 1981-83), las series de la CRE86 son mucho más volátiles que las del BBVA durante los primeros años del período y se acercan después gradualmente a estas segundas, llegando incluso a presentar varianzas inferiores en la parte final de la muestra.

Como hemos visto en la sección 2, esto implica que las series del BBVA han de considerarse “mejores” que las de la CRE durante estos años siempre que no presenten un sesgo sistemático hacia la media demasiado elevado. El Gráfico 11 muestra los valores estimados del nivel crítico del parámetro que mide este sesgo, $\tilde{\phi}$, y de la cota superior para el mismo, $\hat{\phi}_{\max}$. En el caso del

VAB, los valores estimados de estos parámetros indican con claridad que deberían descartarse las tres primeras observaciones de la serie del INE (1981-87) y muy probablemente también las dos siguientes (1987-1991). En el caso del empleo, los resultados son menos concluyentes. Dado el valor estimado de la cota superior $\hat{\phi}_{\max}$, la única observación claramente descartable sería la correspondiente a 1985-87. Por otra parte, no parece probable que el sesgo sistemático de la serie del BBVA pueda ser mucho mayor en la estimación del empleo que en la del VAB. Si utilizamos para la serie de empleo la misma cota de ϕ que para la de VAB, la situación es similar en ambos casos y apunta hacia 1989 como la fecha más razonable de corte.

Gráfico 10: Varianza CRE/Varianza BBVA, series en diferencias anualizadas (escala logarítmica)

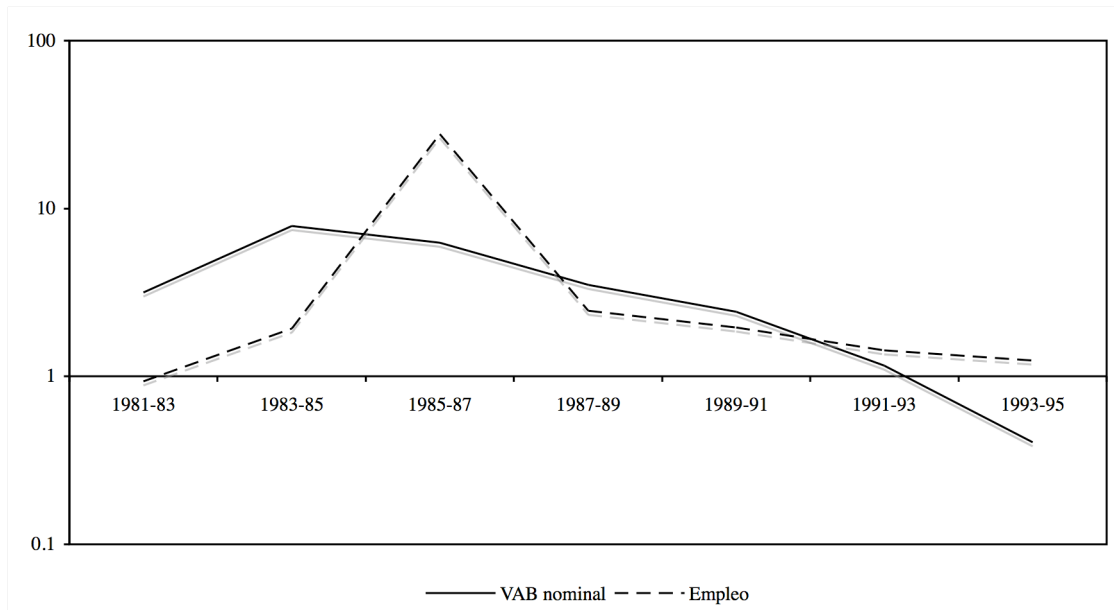
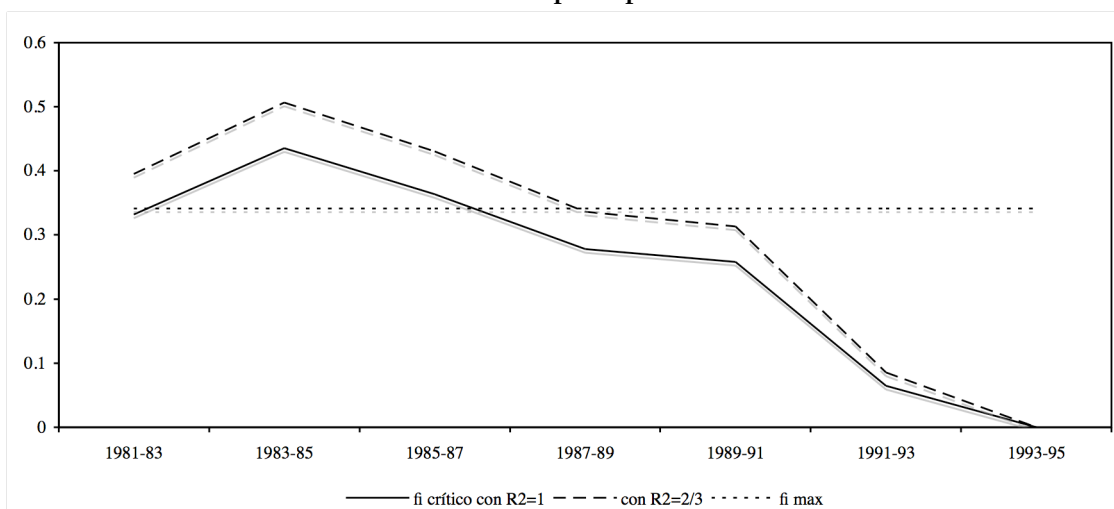
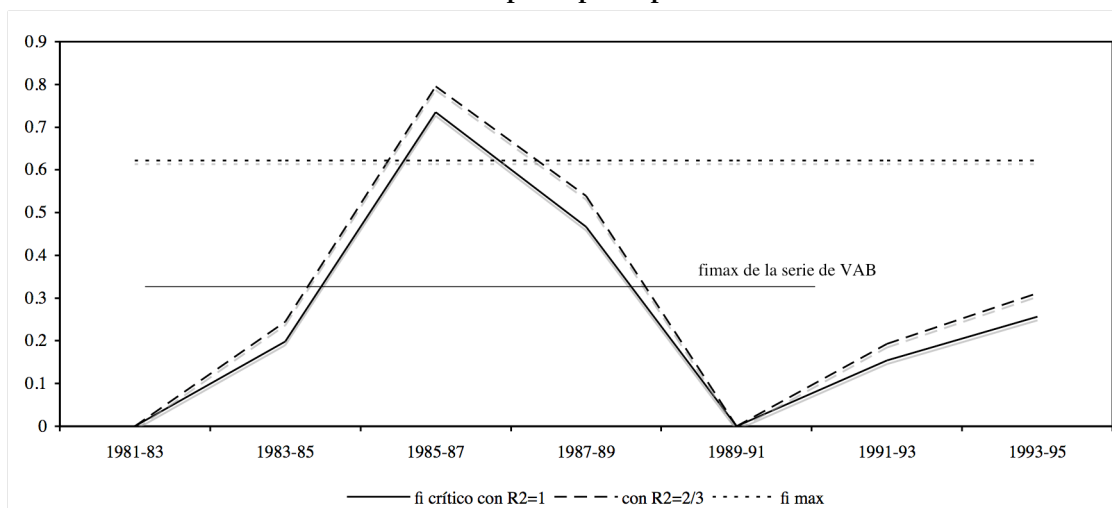


Gráfico 11: Valor crítico y cota superior de ϕ

a. VAB per cápita



b. Empleos per cápita



4.3. Construcción de la serie enlazada

Una vez fijado el punto de corte en 1989, el enlace se realiza por el procedimiento de "retropolación" descrito en el apartado 3.2. Los niveles (normalizados) del VAB y el empleo per cápita se construyen extendiendo hacia atrás la serie homogénea de CRE00 a partir de 1995 utilizando la tasa de crecimiento de la serie de referencia en cada período (CRE86 desde el año citado hasta 1989 y BBVA para años anteriores).

Una vez enlazados los valores normalizados por habitante, se procede a reconstruir los totales de cada magnitud de la forma siguiente. En primer lugar, se recuperan la renta por habitante (en euros corrientes) y el empleo (en puestos de trabajo por habitante) para el conjunto de España extendiendo hacia atrás la correspondiente serie nacional de CRE00 con las tasas de crecimiento de las series nacionales seleccionadas para períodos anteriores. Multiplicando la serie resultante por el índice de cada región se recuperan los niveles de VAB y empleo per cápita (en euros y puestos de trabajo respectivamente) y multiplicando estas cifras por la población nacional y regional se recuperan el VAB y el empleo totales de cada región y del conjunto del país (excluyendo la "extra-región").

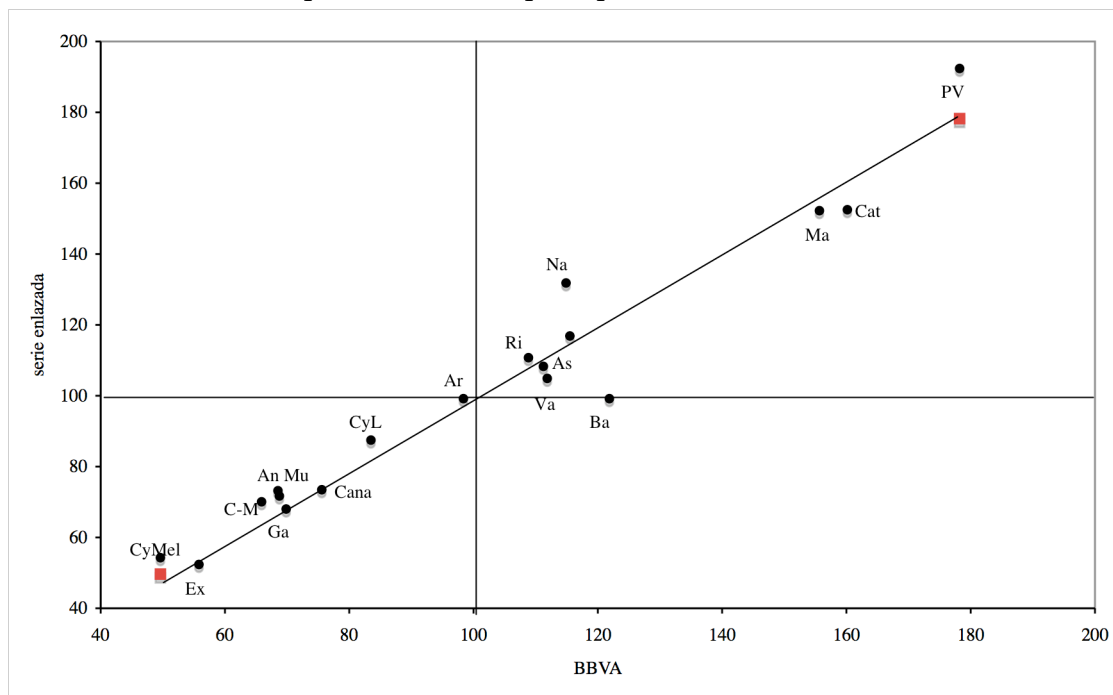
Cuadro 2: Construcción de las series enlazadas de VAB a precios corrientes y empleo por habitante

Período	Fuente:
1995-2007	CRE00 serie homogénea 1995-2007
	Extendida hacia atrás utilizando las tasas de crecimiento de:
1989-1995	CRE86
1955-1989	Fundación BBV (1999)

4.4. Algunos comentarios sobre la serie enlazada

Los Gráficos 12 y 13 y los Cuadros 3 y 4 comparan las series enlazadas de VAB a precios corrientes y empleo per cápita con sus homólogas del BBVA en el año 1955. La correlación entre

Gráfico 12: VAB a precios corrientes per cápita en 1955. Serie enlazada vs. BBVA



Cuadro 3: VAB per cápita en 1955. Serie enlazada vs BBVA

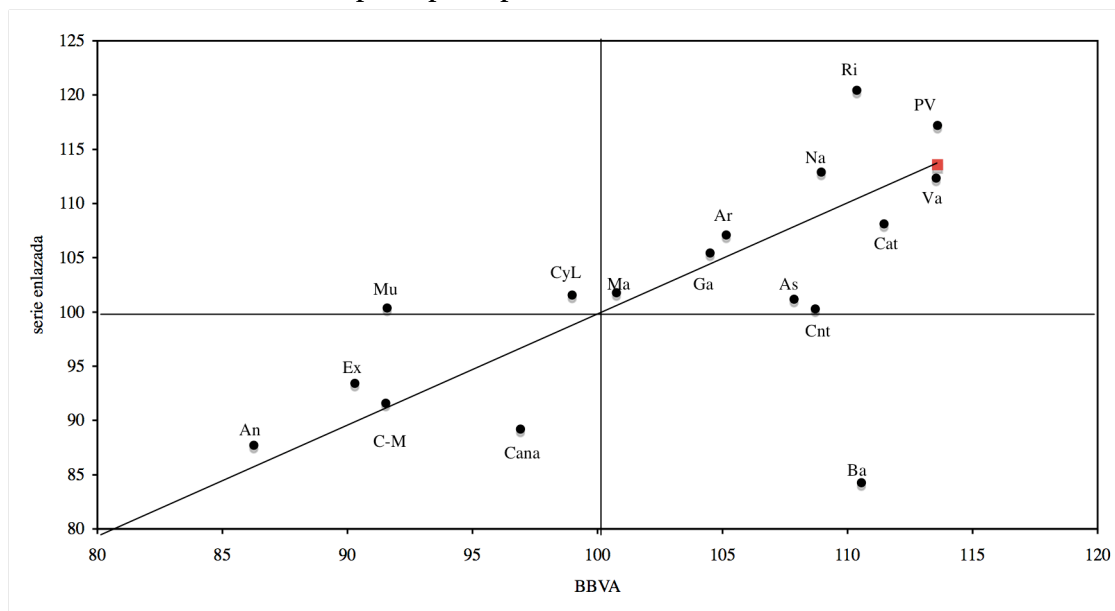
	BBV	rank BBV	SE	rank SE	SE- BBVA	%	rankBBV- rank SE
<i>País Vasco</i>	178.2	1	192.4	1	14.2	8.0%	0
<i>Cataluña</i>	160.2	2	152.6	2	-7.6	-4.8%	0
<i>Madrid</i>	155.7	3	152.3	3	-3.4	-2.2%	0
<i>Baleares</i>	121.8	4	99.2	10	-22.7	-18.6%	-6
<i>Cantabria</i>	115.5	5	116.9	5	1.4	1.2%	0
<i>Navarra</i>	114.9	6	131.8	4	16.9	14.7%	2
<i>Valencia</i>	111.9	7	104.9	8	-7.0	-6.3%	-1
<i>Asturias</i>	111.2	8	108.3	7	-2.9	-2.6%	1
<i>Rioja</i>	108.9	9	110.8	6	1.9	1.8%	3
<i>Aragón</i>	98.4	10	99.2	9	0.8	0.9%	1
<i>Cast. y León</i>	83.5	11	87.4	11	3.9	4.7%	0
<i>Canarias</i>	75.6	12	73.5	12	-2.0	-2.7%	0
<i>Galicia</i>	69.9	13	68.0	16	-1.9	-2.7%	-3
<i>Murcia</i>	68.7	14	71.7	14	2.9	4.3%	0
<i>Andalucía</i>	68.5	15	73.2	13	4.7	6.8%	2
<i>Cast. la Mancha</i>	65.9	16	70.0	15	4.2	6.4%	1
<i>Extremadura</i>	55.8	17	52.3	18	-3.5	-6.3%	-1
<i>Ceuta y Melilla</i>	49.6	18	54.3	17	4.7	9.4%	1

- Nota: SE = serie enlazada.

ambas fuentes es muy elevada: 0.974 en el caso del VAB y 0.872 en el del empleo. En términos estadísticos, por tanto, las discrepancias entre las fuentes en el momento del enlace son lo suficientemente reducidas como para que el procedimiento de enlace por retroprolación genere un serie cuyos niveles no son muy distintos de los de la serie original del BBVA. Esto implica que, en principio, podemos utilizar también los niveles de la nueva serie con un cierta

tranquilidad. Sin embargo, también se detectan diferencias significativas en algunas regiones, lo que supone un problema si queremos utilizar las series no sólo para estimar los parámetros de algún modelo estadístico sino también para analizar la evolución de regiones específicas.

Gráfico 13: Empleos per cápita en 1955. Serie enlazada vs. BBVA



- Nota: se omite Ceuta y Melilla.

Cuadro 4: Empleos per cápita en 1955. Serie enlazada vs BBVA

	BBV	rank BBV	SE	rank SE	SE- BBVA	%	rankBBV- rank SE
<i>País Vasco</i>	113.6	1	117.2	2	3.6	3.2%	-1
<i>Valencia</i>	113.5	2	112.3	4	-1.2	-1.1%	-2
<i>Cataluña</i>	111.5	3	108.1	5	-3.3	-3.0%	-2
<i>Baleares</i>	110.6	4	84.2	17	-26.3	-23.8%	-13
<i>Rioja</i>	110.4	5	120.4	1	10.1	9.1%	4
<i>Navarra</i>	109.0	6	112.9	3	4.0	3.6%	3
<i>Cantabria</i>	108.7	7	100.3	12	-8.4	-7.8%	-5
<i>Asturias</i>	107.9	8	101.2	10	-6.7	-6.2%	-2
<i>Aragón</i>	105.2	9	107.1	6	2.0	1.9%	3
<i>Galicia</i>	104.5	10	105.4	7	0.9	0.9%	3
<i>Madrid</i>	100.7	11	101.8	8	1.0	1.0%	3
<i>Cast. y León</i>	99.0	12	101.6	9	2.6	2.6%	3
<i>Canarias</i>	96.9	13	89.2	15	-7.7	-7.9%	-2
<i>Murcia</i>	91.6	14	100.4	11	8.8	9.6%	3
<i>Cast. la Mancha</i>	91.5	15	91.6	14	0.0	0.0%	1
<i>Extremadura</i>	90.3	16	93.4	13	3.1	3.5%	3
<i>Andalucía</i>	86.3	17	87.7	16	1.4	1.7%	1
<i>Ceuta y Melilla</i>	45.9	18	49.5	18	3.6	7.9%	0

- Nota: SE = serie enlazada.

El caso más llamativo es el de Baleares. Tomando como referencia la estimación del BBVA, la nueva serie reduce la renta per cápita del archipiélago en 22.7 puntos y su empleo por habitante en 26.3 puntos. En términos de la primera variable, la región pasaría de la cuarta a la

décima posición y en términos de ocupación, de la cuarta a la penúltima posición.³ Menos extremas pero también preocupantes son las discrepancias sobre Navarra (en términos de renta) y la Rioja (en términos de empleo).

5. Precios y VAB a precios constantes

Los índices de precios del VAB implícitos en las series del BBVA se obtienen dividiendo la serie de VAB a precios corrientes por la correspondiente serie a precios constantes (de 1986). En FBBV (1999, pp. 36-37) se dan algunas indicaciones sobre la construcción de las series de precios. En el caso de la industria y los servicios, el VAB a precios corrientes de cada rama productiva se deflacta en todas las regiones utilizando el deflactor nacional que proporciona la CNE (para 21 ramas). En estos sectores, por tanto, las diferencias de niveles de precios entre regiones de acuerdo con las series del BBVA recogen únicamente diferencias en su estructura productiva. En el caso de la agricultura, sin embargo, se parte de los datos que ofrece el Ministerio de Agricultura sobre los precios de diversos productos y la producción de los mismos en cada región para construir un índice específico para cada territorio. También se construye un índice de precios específico para el sector de la construcción, pero no se ofrece detalle alguno sobre la metodología utilizada, remitiendo al lector a un trabajo que no he podido encontrar (Alcaide y Pascualena, 1985).

En la CRE86 no se ofrecen estimaciones del VAB a precios constantes. Hay dos formas rápidas de recuperar un índice de precios para esta serie. La primera es trabajar con el deflactor del PIB regional que se puede recuperar a partir de las series de PIB a precios corrientes y constantes que ofrece la misma fuente. La segunda es trabajar con las series de VAB a precios corrientes y constantes que se ofrecen en la BD Mores, donde también se parte de los deflactores nacionales por ramas de actividad que proporciona el propio INE para construir índices de precios regionales. Puesto que el deflactor del PIB de CRE86 muestra un comportamiento un tanto errático (véase el Anexo 2), he optado por utilizar el deflactor del VAB de la BDMores para deflactar el VAB nominal de la CRE86 en los años en los que ésta es la serie de referencia para el enlace.⁴

Por último, en la CRE00 ya no se ofrecen series de VAB a precios constantes sino índices enlazados de volumen con base 2000 (esto es, con un valor normalizado a 100 en el año 2000 para todas las regiones). Estas series seguramente recogen las variaciones del output real regional a lo largo del tiempo mejor que sus antecesoras a precios constantes pero no contienen información alguna sobre sus niveles. Dado que esta última es necesaria en muchas ocasiones, he reconstruido una serie de VAB "a precios constantes de 2000" (lo que estrictamente hablando no es cierto) multiplicando los índices de volumen por el VAB a precios corrientes de cada región en el año 2000. Dividiendo la serie de VAB a precios corrientes por la serie de VAB a

³ Maluquer (2006) considera muy implausibles las estimaciones de la FBBVA para Baleares.

⁴ Pregunta: ¿cómo construye el INE los índices de precios regionales? No parece que estén deflactando cada rama con el índice de precios nacional. Un cálculo rápido comparando dos regiones en la serie 2000-07 muestra que incluso dentro de cada rama hay variaciones de precios muy significativas entre regiones.

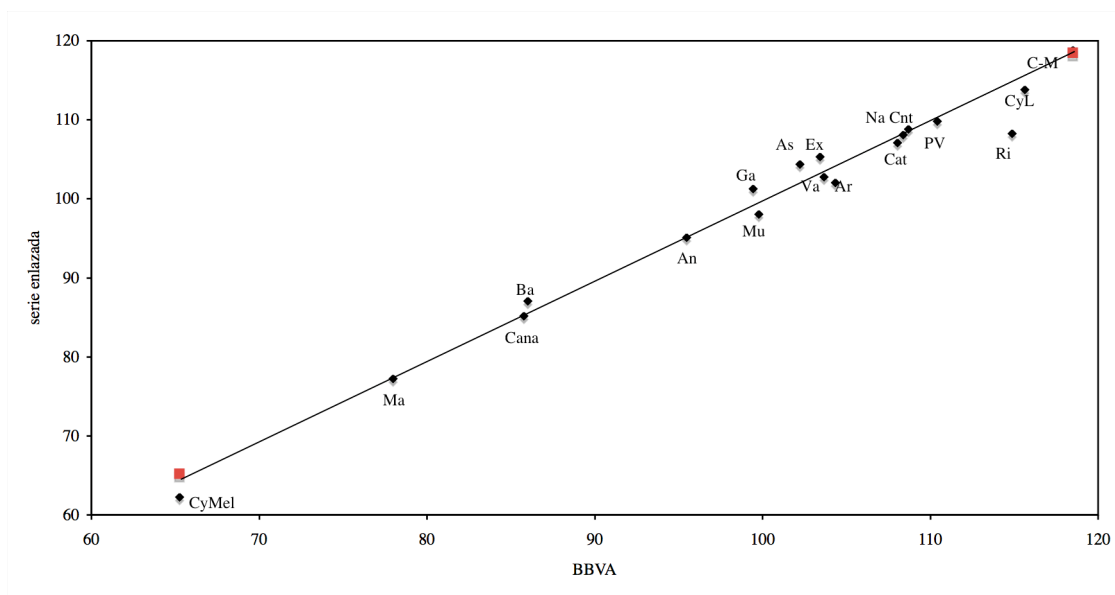
precios constantes así construida, obtengo un índice aproximado de precios del VAB con base 2000.

La serie enlazada de índices de precios del VAB en base 2000 se obtiene extendiendo hacia atrás la serie de precios del VAB de CRE00 utilizando las tasas de crecimiento de las series de precios del VAB de CRE86 (de hecho, de la BDMores) y del BBVA, manteniendo el año de corte entre ambas fuentes fijado en una sección anterior (1989).⁵ En esta ocasión el enlace no se realiza con las series normalizadas por el promedio nacional sino con índices "brutos" de precios para cada región. Los promedios nacionales se recuperan ex-post tal como se indica seguidamente.

Una vez construida la serie de precios, ésta se utiliza para deflactar la serie enlazada de VAB nominal, obteniendo una serie de VAB real valorada (aproximadamente) a precios constantes de 2000. Sumando las regiones, obtengo una serie de VAB real a nivel nacional. Finalmente, el índice de precios nacional se obtiene dividiendo el VAB nacional a precios corrientes por el VAB nacional a precios de 2000.

Los perfiles de las distintas series regionales de precios y VAB real por habitante se comparan en el Anexo 2.

Gráfico 14: Nivel relativo de precios del VAB en 1955. Serie enlazada vs BBVA



- Nota: Ambas series de precios se ponen en base 1995 antes de normalizarlas por el promedio nacional.

El Gráfico 14 compara la serie enlazada de precios con la del BBBVA en 1955 tras expresar ambas en una base común y normalizarlas por el promedio nacional. La correlación entre ambas series es de 0.989 y las diferencias entre ellas son generalmente poco significativas. La

⁵ El criterio estadístico utilizado más arriba para valorar el contenido informativo de las distintas series no parece aplicable en el caso de los índices de precios. Dado que en las fuentes que manejamos estos índices se construyen fundamentalmente a partir de los deflatores nacionales por ramas, hemos de esperar que su volatilidad subestime la de los precios "reales" dado que sólo se recogen las variaciones medias nacionales dentro de cada rama. En este contexto, una menor volatilidad no es necesariamente una buena señal. Dada la dificultad de encontrar un criterio razonable para intentar valorar estas series, he optado por respetar el punto de corte fijado arriba (y por no incluir a las series de precios en el análisis realizado para fijarlo).

mayor de ellas corresponde a la Rioja, con un valor de 6.6 puntos. En términos generales, por tanto, el enlace no genera diferencias preocupantes en niveles iniciales que aconsejen considerar la posibilidad de introducir correcciones adicionales.

6. Conclusión y algunos temas pendientes

En este trabajo se presenta un primer intento de construir una serie "homogénea" larga mediante el enlace de los datos de la Fundación BBVA con los de diversas versiones de la Contabilidad Regional. El enlace se ha realizado con el objetivo de optimizar las propiedades estadísticas de las series enlazadas con vistas a su utilización como variables dependientes en modelos de crecimiento o determinación del empleo y se construye utilizando un procedimiento de "retropolación pura" que conserva las tasas de crecimiento de las series de referencia pero no necesariamente sus niveles.

Puesto que las discrepancias existentes entre las distintas series en el momento de su enlace son generalmente reducidas, los niveles de la serie enlazada son razonablemente parecidos a los de las series originales de referencia. Este hecho sugiere que podemos utilizar los niveles de la serie enlazada con una cierta tranquilidad a la hora de estimar el modelo relevante. Para algunas regiones, sin embargo, se detectan diferencias muy significativas en niveles entre la serie enlazada y la serie original. Esto introduce un considerable grado de incertidumbre sobre los niveles de algunos agregados y aconseja una cierta precaución a la hora de utilizar los datos para analizar la evolución de algunas regiones (muy especialmente en el caso de Baleares). En esta situación, parece aconsejable intentar buscar fuentes alternativas que nos permitan reducir el grado de incertidumbre sobre las variables relevantes. Si la información disponible lo permite, debería investigarse la posibilidad de introducir alguna corrección en los niveles de las series en las regiones más problemáticas.

Otra anomalía que merecería una investigación cuidadosa tiene que ver con la evolución del empleo en Galicia entre 1995 y 1999. De acuerdo con la serie enlazada de CRE00 (que por lo demás he aceptado sin discusión hasta el momento) esta comunidad pierde 17 puntos de empleo relativo en cuatro años, diez más que la siguiente (Asturias). Esto es resultado de un descenso del 4.7% en el empleo gallego (con un descenso de la población del 1.06%) y de un incremento del 16.3% en el empleo nacional (con un crecimiento de la población del 1.37%). De acuerdo con esta fuente, Galicia sería la única comunidad con una pérdida neta de empleo durante el período (mientras que Asturias, con un descenso más acusado de la población, registraba un aumento del 4.3% en el número de puestos de trabajo).

Más generalmente, podría ser conveniente revisar el enlace realizado por el INE entre las series de CRE en bases 1995 y 2000 dado que éste se ha realizado con la restricción (no necesariamente razonable como hemos visto) de que la nueva serie ha de respetar, excepto por correcciones metodológicas, los valores de la serie original en el año base de 1995 (INE, 2008b).

Referencias

- Alcaide, J. (2007). *Evolución de la población española en el siglo XX por provincias y comunidades autónomas*. Fundación BBVA, Bilbao.
- Alcaide, C. y R. Pascualena (1985). "Deflatores provinciales y regionales de las macromagnitudes económicas." Caja de Ahorros de Alicante y Murcia, Alicante.
- de Bustos, A., A. Cutanda, A. Díaz, F. J. Escribá, M. J. Murgui y M. J. Sanz (2008). "La BD MORES en base 2000: nuevas estimaciones y variables." Documento de Trabajo D2008-02, Dirección General de Presupuestos, Ministerio de Hacienda. Madrid.
- de la Fuente, A. y R. Doménech (2006). "Human capital in growth regressions: how much difference does data quality make?" *Journal of the European Economic Association* 4(1), pp. 1-36.
- 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, 2008a). Contabilidad Regional de España. En Base de datos electrónica INEbase. Economía: Cuentas Económicas. Madrid.
<http://www.ine.es/inebase/cgi/um?M=%2Ft35%2Fp010&O=inebase&N=&L=>
- Instituto Nacional de Estadística (INE, 2008b). Cifras de población. En Base de datos electrónica INEbase. Demografía y población. Madrid.
http://www.ine.es/inebmenu/mnu_cifraspop.htm
- Instituto Nacional de Estadística (INE, 2008c). "Metodología de la serie homogénea 1995-2006." En Base de datos electrónica INEbase. Economía: Cuentas Económicas. Contabilidad Regional de España. Madrid.
<http://www.ine.es/inebase/cgi/um?M=%2Ft35%2Fp010&O=inebase&N=&L=>
- Krueger, A. y M. Lindahl (2001). "Education for growth: why and for whom?" *Journal of Economic Literature* XXXIX, pp. 1101-36.
- Maluquer, J. (2006?). "El desarrollo económico de las Islas Baleares en perspectiva comparada (1950-2000)." Mimeo, Universidad Autónoma de Barcelona.