

WORKING PAPER

Series enlazadas de algunos agregados económicos regionales, 1955-2014

Parte I: Metodología, VAB, PIB y puestos de trabajo

Ángel de la Fuente

Series enlazadas de algunos agregados económicos regionales, 1955-2014.

Parte I: Metodología, VAB, PIB y puestos de trabajo

Ángel de la Fuente

Resumen

En este trabajo se construyen series “homogéneas” anuales de VAB y PIB a precios corrientes y constantes y de puestos de trabajo para las regiones españolas durante el periodo 1955-2014. Estas series se obtienen enlazando la Contabilidad Regional del INE con las series elaboradas por Julio Alcaide y colaboradores para la Fundación BBVA. El “punto de corte” en el que se abandona esta última fuente en favor de la anterior como referencia para la construcción de la serie larga 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. La discrepancia entre ambas series que aflora en el momento del enlace se reparte entre los niveles iniciales y las tasas de crecimiento de la serie más antigua utilizando, en la medida de lo posible, estimaciones externas del valor de las variables de interés a comienzos del período muestral.

Abstract

Distributed ledgers are a technology that can support a digitized version of cash while potentially withholding its four major features: universality, anonymity, peer-to-peer exchangeability (P2P) and a constant nominal value. Based on that, several central banks are assessing the issuance of central bank digital currencies (CBDCs) under potentially very disruptive assumptions that would broaden their policy scope (including unbounded negative rates); but there is a range of possible schemes to be implemented depending exactly on which features of physical cash to withhold. This paper identifies and analyzes the opportunities and challenges of four key schemes, which taken as a whole offer advice on a strategy of gradual testing and implementation. Such a strategy would minimize the potentially large disruptive costs to the current financial architecture and credit markets, while ensuring efficiency gains linked to lower intermediation costs (especially in reference to the payment system), an eventual broadening of the policy scope and a reduction of informality, illicit activities and tax evasion.

Palabras clave: Contabilidad Regional de España, series homogéneas

Keywords: Regional Accounts for Spain, homogeneous series

Clasificación JEL/JEL Classification: E01, R1

Series enlazadas de algunos agregados económicos regionales, 1955-2014.

Parte I: Metodología, VAB, PIB y puestos de trabajo

RegData_55-14, Versión 5.0-parte I

Angel de la Fuente*

FEDEA e Instituto de Análisis Económico (CSIC)

Abril de 2017

Resumen

En este trabajo se construyen series “homogéneas” anuales de VAB y PIB a precios corrientes y constantes y de puestos de trabajo para las regiones españolas durante el periodo 1955-2014. Estas series se obtienen enlazando la Contabilidad Regional del INE con las series elaboradas por Julio Alcaide y colaboradores para la Fundación BBVA. El “punto de corte” en el que se abandona esta última fuente en favor de la anterior como referencia para la construcción de la serie larga 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. La discrepancia entre ambas series que aflora en el momento del enlace se reparte entre los niveles iniciales y las tasas de crecimiento de la serie más antigua utilizando, en la medida de lo posible, estimaciones externas del valor de las variables de interés a comienzos del período muestral.

Abstract

I construct “homogeneous” series of employment and GVA and GDP at current and constant prices for Spain and its regions covering the period 1955-2014. 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. To the extent that it is possible, the discrepancy between the two series at the point of linkage is distributed between the initial levels of the older series and its subsequent growth using external estimates of the relevant variables at the beginning of the sample period.

Palabras clave: Contabilidad Regional de España, series homogéneas

Clasificación JEL/JEL Classification: E01, R1

Key words: Regional Accounts for Spain, homogeneous series

* Este trabajo forma parte de un proyecto de investigación cofinanciado BBVA Research. Agradezco también la financiación del Ministerio de Economía y Competitividad a través del proyecto ECO2014-59196-P y la asistencia de Lucía Gorjón, Analía Viola, Esther Gómez y Carmen Arias.

1. Introducción

Una de las dificultades a 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 de tiempo. 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 sus sucesores posteriormente para la Fundación BBVA y la Fundación FIES. Una versión revisada y homogénea de estas series que cubre el período 1955-93 a intervalos bienales está disponible en Fundación BBV (1999). Por otro lado, disponemos también de la Contabilidad Regional de España (CRE) que elabora el Instituto Nacional de Estadística (INE) desde 1980. Las distintas series de la CRE han sido enlazadas entre sí en de la Fuente (2016b), obteniéndose así una serie homogénea para 1980-2014 que, como se verá más adelante, presenta diferencias muy apreciables con la del BBVA durante su período de solapamiento.

En el presente estudio se construyen series históricas "homogéneas" de algunos agregados económicos regionales para el período 1955-2014 mediante el enlace de las dos fuentes citadas, intentando resolver las discrepancias entre ellas con ayuda de otras fuentes externas.¹ El trabajo es fundamentalmente una actualización de otro anterior (de la Fuente, 2008 y 2009) excepto en que ahora se incorpora la serie completa enlazada de la CRE construida en de la Fuente (2016b) y se construyen algunas series adicionales.² El estudio se divide en dos partes. En esta primera parte se desarrolla la metodología de enlace y se construyen series anuales de VAB y PIB a precios corrientes y constantes y de puestos de trabajo. En una segunda parte, la base de datos se completará con series adicionales de empleo, salarios y rentas del trabajo, que en algunos casos cubrirán sólo parte del período de interés.

El trabajo está organizado como sigue. En la sección 2 se describe la información disponible sobre la evolución de los agregados de interés en nuestras dos fuentes básicas. Tomando como dadas estas series, la tarea que aquí se aborda es la de combinarlas de la mejor forma posible para cubrir el conjunto del período desde 1955. Esto plantea dos problemas fundamentales: primero, ¿qué serie deberíamos utilizar como referencia cuando ambas están disponibles? y

¹ La idea de enlazar las series de la CRE con las del BBVA no es novedosa. Doménech et al (1999) y de Bustos et al (2008) realizan sendos ejercicios en esta línea. En ambos trabajos se toman como dadas las series de la CRE desde su inicio en 1980 y se extienden hacia atrás tomando como referencia las series del BBVA. Doménech et al (1999) ofrecen dos enlaces alternativos entre ambas fuentes: el primero de ellos se construye por retropolación (esto es, las tasas de crecimiento de las series del BBVA se utilizan para prolongar hacia atrás las series de la CRE), mientras que en el segundo se respetan los valores iniciales de la serie más antigua, repartiéndose a lo largo del período las diferencias entre ambas fuentes que afloran en 1980. De Bustos et al (2008) optan por el primero de los métodos citados. En ambos casos, las series así obtenidas se ajustan de forma que su agregación coincida con una serie nacional de referencia basada en la CNE. En ambos trabajos, el enlace se realiza con datos desagregados por sectores (4 grandes ramas en Doménech et al y 17 en de Bustos et al). Las principales diferencias entre este trabajo y los dos citados tienen que ver con el nivel de agregación, con el método de enlace y con la determinación del "punto de corte" entre las series del BBVA y de la CRE. Los dos últimos aspectos se discutirán en detalle más adelante.

² El principal cambio que se introduce en este trabajo en relación con su versión anterior (de la Fuente, 2009) es que las series terminales son distintas. Mientras que en de la Fuente (2008) se trabajaba con una extensión hacia atrás hasta 1995 de la CRE00, en el presente trabajo se parte directamente de la serie enlazada de CRE 1980-2014 construida en de la Fuente (2016b).

segundo, ¿cómo exactamente debería construirse el enlace entre ellas? El primer problema, que se aborda en la sección 3, puede reformularse en la práctica en términos de la fijación de la "fecha de corte" en la que se abandonan las series del BBVA para adoptar como referencia los datos de CRE elaborados por el INE. Para abordar el problema, se desarrolla una metodología que permite escoger la serie con mayor contenido informativo. El procedimiento de selección se basa en el siguiente experimento conceptual: supongamos que queremos analizar los determinantes de las variables con las que estamos trabajando (renta y empleo) regresándolas 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?

Una vez fijado el punto de corte, en la sección 4 se procede a la construcción de las series enlazadas. El ejercicio se realiza en primera instancia en términos relativos, recuperándose después los niveles absolutos de las series a partir de los correspondientes valores nacionales. Esto es, para cada una de las dos fuentes disponibles, se construyen índices de *output* y puestos de trabajo por habitante, normalizados por el correspondiente promedio nacional, que se enlazan entre sí. El enlace se realiza utilizando el llamado procedimiento mixto (de la Fuente, 2014), utilizando información tomada de fuentes externas para intentar resolver las discrepancias que existen entre las series originales en cuanto a la posición relativa de algunas regiones. Una vez enlazados los índices, los niveles per cápita de las series se recuperan multiplicando los índices enlazados por los niveles medios nacionales de la variable correspondiente medidos en términos absolutos, que se toman de de la Fuente (2016a). Finalmente, los valores totales de las series se recuperan multiplicando sus valores per cápita por la población de cada región.

Puesto que las series del BBVA tienen periodicidad bienal, las series enlazadas construidas en la sección 4 se refieren sólo a los años impares durante la primera parte del período muestral. En la sección 5 los huecos que presentan estas series se rellenan para obtener series anuales utilizando un procedimiento de interpolación que incorpora la información anual existente a nivel agregado. En la sección 6 se construyen series de VAB y PIB real valorados a los precios medios del conjunto del período muestral. En la sección 7 se describen los archivos de datos que acompañan al trabajo y la sección 8 concluye. En el Anexo 1 se recogen los detalles técnicos del procedimiento utilizado para fijar la fecha de corte entre los datos del BBVA y los de la CRE y en el Anexo 2 se incluye una serie de gráficos que muestran los perfiles de las series originales y enlazadas de las principales variables de interés.

2. Información disponible

Como ya se ha indicado, en este trabajo se parte fundamentalmente de dos fuentes. La primera son las series homogéneas elaboradas en Fundación BBV (1999), que cubren el período 1955-93 a intervalos bienales. Esta serie se extiende hasta 1995 en Fundación BBVA (FBBVA, 2000), donde también se ofrecen avances de las magnitudes relevantes hasta 1999. La segunda fuente son las series enlazadas de la CRE-e que se construyen en de la Fuente (2016b) para el período 1980-2014 a partir de las distintas series de Contabilidad Regional que ha ido publicando el INE.

El Cuadro 1 resume la información disponible sobre las principales variables de interés. Ambas fuentes ofrecen información sobre la mayor parte de las mismas, lo que permite enlazarlas directamente o con ajustes menores en algún caso. La principal excepción es que la FBBVA no ofrece información sobre el número de personas ocupadas o asalariadas, ni sobre horas de trabajo, lo que nos obligará a recurrir a otras fuentes auxiliares para estimar algunas variables con el fin de completar la base de datos en la medida de lo posible.

Cuadro 1: Series disponibles en las fuentes básicas

		<i>FBBVA</i> <i>1955-1995</i> <i>años impares</i>	<i>RegDat</i> <i>v4.2</i> <i>1980-2014</i>
<i>VABcf/pb nom</i>	VAB al coste de los factores/precios básicos, a precios corrientes	X	X
<i>VABcf/pb real</i>	VAB al coste de los factores/precios básicos, a precios constantes	X	X
<i>PIBpm nom</i>	PIB a precios de mercado, a precios corrientes	X	X
<i>PIBpm real</i>	PIB a precios de mercado, a precios constantes	X	X
<i>PISB/SIFMI</i>	Producción imputada de servicios bancarios o servicios de intermediación financiera	X	X
<i>IMPPRODN</i>	Impuestos netos sobre los productos	X	X
<i>RAS</i>	Remuneración de los asalariados	X	X
<i>RTL</i>	Rentas totales del trabajo estimadas	X	X
<i>PT</i>	Puestos de trabajo	X	X
<i>OCU</i>	Personas ocupadas	no	X
<i>H</i>	Horas trabajadas	no	X
<i>PTEJC</i>	Puestos de trabajo equivalentes a jornada completa	no	X
<i>PTAS</i>	Puestos de trabajo asalariados	X	X
<i>AS</i>	Personas asalariadas	no	X
<i>HAS</i>	Horas trabajadas por los asalariados	no	X
<i>PTASEJC</i>	Puestos de trabajo asalariados equivalentes a jornada completa	no	X

Las dos fuentes ofrecen series de PIB y VAB a precios corrientes y constantes. En ambos casos, el PIB se mide a precios de mercado y el VAB a coste de los factores o precios básicos, por lo que las series son en principio aproximadamente consistentes entre sí.³ Sin embargo, hay algunas

³ El Valor Añadido Bruto (VAB) se define genéricamente como la diferencia entre el valor de la producción y el coste de adquisición de los consumos intermedios. Dependiendo de cómo se valore la producción, se obtienen distintos conceptos de valor añadido que difieren entre sí en el tratamiento dado a ciertos impuestos indirectos y a las subvenciones a las empresas. Antes de la introducción de la base 1995, los dos principales agregados de valor añadido ofrecidos en la Contabilidad Nacional y Regional eran el VAB a precios de mercado (VABpm) y el VAB a coste de los factores (VABcf). El primero de estos agregados incluía los impuestos sobre la producción (aunque no el IVA ni los impuestos sobre las importaciones) y excluía las subvenciones a las empresas, mientras que el segundo incluía las subvenciones (fundamentalmente las ayudas de la política agrícola comunitaria y las subvenciones de explotación a empresas públicas) y excluía los impuestos sobre la producción (fundamentalmente los impuestos especiales y sobre transmisiones patrimoniales y actos jurídicos documentados así como otros impuestos menores). Comenzando con la base 1995, el principal agregado de valor añadido que se ofrece en la CRE y la CNE pasa a ser el VAB a precios básicos (VABpb). Esta magnitud es similar al VAB a coste

diferencias entre ambas fuentes que hay que corregir antes de proceder a su enlace. En particular, la serie de VABcf agregado regional que ofrece la FBBVA se construye como la suma de los VABs sectoriales y no excluye el coste imputado de los servicios financieros que utilizan los sectores productivos (PISB o producción imputada de servicios bancarios), mientras que la CRE-e excluye esta partida del VAB agregado. Antes de enlazar ambas series, por tanto, en el caso de la FBBVA resulta necesario construir una nueva serie revisada de VABcf neta de PISB. Por otra parte, puesto que no se dispone de un deflactor específico para la PISB, el deflactor del VABcf con el que trabajaremos en la primera parte del período muestral corresponde realmente a la serie de VABcf antes de sustraer la PISB.

Durante la primera parte del período muestral, los deflactores regionales disponibles se construyen fundamentalmente a partir de índices de precios sectoriales elaborados a nivel nacional. En FBBV (1999, pp. 36-37) se dan algunas indicaciones sobre la construcción de las series de precios que se utilizan en esta fuente. 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 índice de precios nacional que proporciona la CNE (para 21 ramas productivas). 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 conseguido encontrar (Alcaide y Pascualena, 1985). Para los primeros años de la serie de la CRE-e la situación es similar: el VAB regional se deflacta con un índice de precios construido a partir de los deflactores sectoriales nacionales (de la Fuente, 2016b, p. 8).

Finalmente, las series de remuneración de asalariados de la CRE y de costes del factor trabajo en FBBVA también son consistentes entre sí, pues incluyen las cotizaciones sociales a cargo del empleador además de los salarios brutos (de retenciones por IRPF y cotizaciones sociales a cargo del trabajador). Ambas fuentes también permiten estimar las rentas totales del trabajo (RTL) por el procedimiento que se describe en la segunda parte de este estudio.

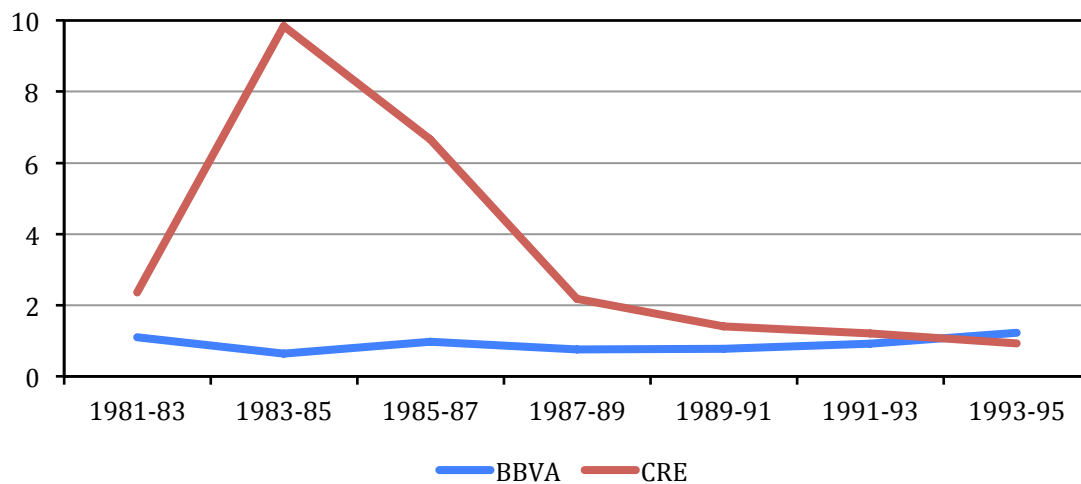
3. ¿Qué fuente es mejor en cada período? Determinación del punto de corte

Como ya se ha visto, las series del BBVA se solapan con las de Contabilidad Regional entre 1980 y 1995, lo que nos obliga a elegir entre ellas como referencia para la construcción de la serie larga enlazada. Dado que el INE dispone de mucho mejores medios para estimar la renta y el empleo regionales que un servicio de estudios privado, en principio lo más lógico sería utilizar

de los factores pero no idéntica. En particular, el VABpb excluye los principales impuestos sobre los productos pero incluye algunos impuestos menores ligados a la producción (como los impuestos sobre actividades económicas y sobre bienes inmuebles). De la misma forma, se incluyen las principales subvenciones a las empresas, pero se excluyen algunas partidas menores (como las ayudas al fomento del empleo y las bonificaciones de intereses). Dado el reducido importe de estas partidas, el VABpb y el VABcf son los indicadores de valor añadido más comparables entre sí que encontramos en las series de CNE y CRE anteriores y posteriores a la introducción de la base 95 (INE, 1999).

las series enlazadas de la CRE disponibles desde 1980 y recurrir a las de la Fundación BBVA únicamente para el período anterior, cuando éstas son la única fuente disponible, tal como hacen Doménech et al (1999) y de Bustos et al (2008). 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 1 y en el Anexo 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.

Gráfico 1: Varianza de las tasas de crecimiento anualizadas del PIB per cápita relativo



- *Nota:* El PIB per cápita regional es un índice con media España = 100. Las tasas de crecimiento anualizadas se calculan como $(pib_{i,t+2} - pib_{i,t})/2$ donde $pib_{i,t}$ es el logaritmo del PIB per cápita relativo 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 1 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 estas últimas, 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 entonces cabría esperar que (otras cosas iguales) la volatilidad de la serie del BBVA 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.

3.1. Un modelo formal para elegir entre dos series alternativas

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 o el empleo relativo 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 período dado para analizar el efecto sobre el crecimiento de la renta o el empleo de alguna otra variable, x . Dado que disponemos de dos estimaciones alternativas de y (las series de CRE-e y BBVA) ¿cuál de ellas deberíamos utilizar?

Para responder esta pregunta resulta útil tratar a las dos series de referencia como aproximaciones "ruidosas" a la misma 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 la 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 aproximadamente 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 y_i es el valor real de la variable de interés, el parámetro $\phi \in [0,1]$ mide el tamaño del sesgo hacia cero de las series del BBVA y e_i^{CRE} y e_i^{BBV} 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.

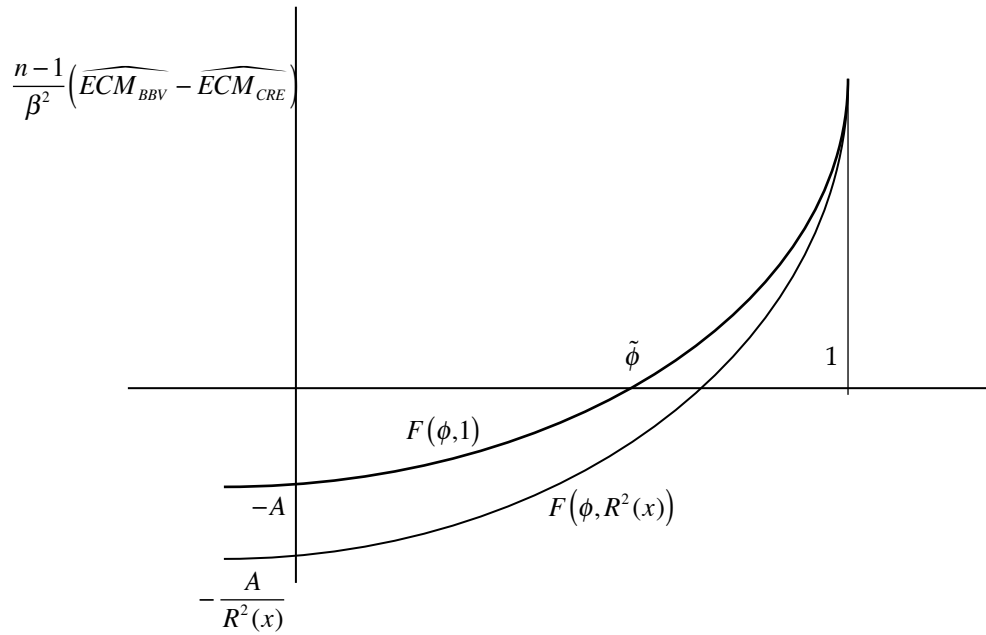
Cuando $\phi > 0$ la situación se complica porque una menor varianza ya no equivale necesariamente a menos ruido y porque el error sistemático en la medición de y genera un sesgo en la estimación de β . El criterio natural en este caso es la minimización del error cuadrático medio (ECM) del estimador de β . Como veremos enseguida, bajo ciertos supuestos es posible obtener una estimación consistente del ECM del estimador de β que se obtendría con cada una de las series disponibles, 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 la 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"). Aunque este resultado no elimina necesariamente toda la incertidumbre sobre qué serie es mejor, ciertamente permite acotar el problema.

Sean ECM_{BBV} y ECM_{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

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

nos da un estimador consistente de la diferencia entre estas magnitudes, de forma que la serie del BBVA será preferible a la de la CRE si y sólo si $F() < 0$. En esta expresión, $n = 18$ es el número de observaciones en el corte transversal (las 17 comunidades autónomas más el agregado de 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 los *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 la sección 1 del Anexo 1).

Gráfico 2: La función $F()$ con $A > 0$



Inspeccionando la ecuación (4) vemos que si $\hat{r}_{CRE} > \hat{r}_{BBV}$ entonces $A < 0$ y $F() > 0$ 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), tenemos $A > 0$ y 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

$$\begin{aligned}
(5) \quad F(0, R^2(x)) &= -\frac{A}{R^2(x)} < 0 \\
F(1, R^2(x)) &= n > 0
\end{aligned}$$

tal como se ilustra en el Gráfico 2. Por tanto, si la serie del BBVA tiene un 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 ECMs 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 $\phi < \tilde{\phi}$. Para calcular este valor, observamos que

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

de forma que si $F(\phi, 1)$ es negativo también lo será $F(\phi, R^2(x))$ para cualquier x , lo que asegura la superioridad de la serie del BBVA para cualquier regresor. El valor crítico de ϕ es, por tanto, el mayor de los ceros de la función cuadrática $F(\phi, 1)$ y viene dado por

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

Alternativamente, podemos fijar un valor "razonable" de $R^2(x)$ para obtener una estimación de $\tilde{\phi}$ bajo un criterio algo menos exigente. Por otra parte, en el Anexo 1 se muestra también que es posible estimar una cota superior para ϕ utilizando

$$(7) \quad \hat{\phi}_{\max} = 1 - \max_t \hat{r}_{CRE,t}$$

donde $\hat{r}_{CRE,t}$ es el ratio de fiabilidad estimado para la serie CRE-e 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 $\hat{\phi}_{\max} \leq \tilde{\phi}$, esto es, que la cota superior de ϕ esté por debajo de su valor crítico.

En el apartado siguiente, calcularé los valores de $\hat{\phi}_{\max}$ y de $\tilde{\phi}$ 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.

3.2. Determinación del punto de corte

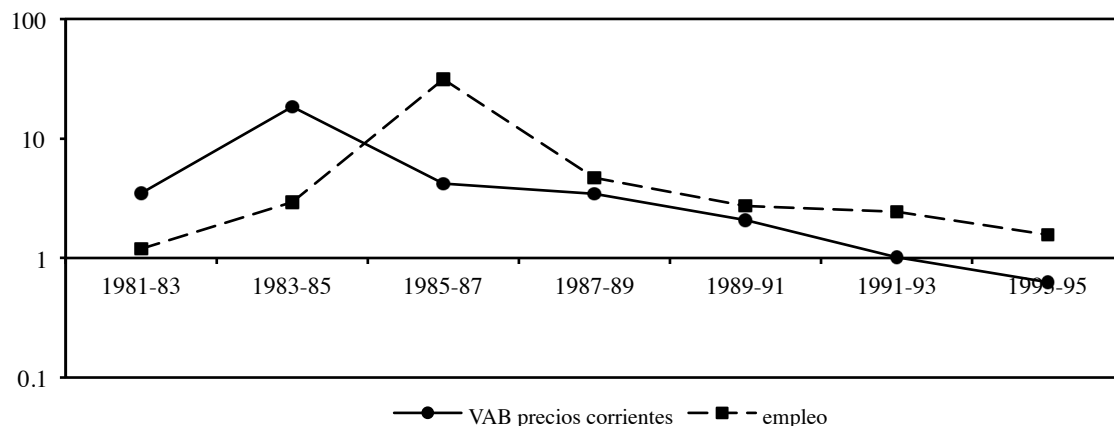
Aplicando el procedimiento desarrollado en el apartado anterior, es posible determinar la fecha óptima en la que abandonar las series del BBVA y adoptar como referencia la CRE-e. En principio, esta fecha podría ser distinta para cada una de las series que estamos considerando. El método adoptado también 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 inconsistente con la intuición básica que guía el ejercicio: que hay problemas con los primeros años de la serie del INE pero que seguramente éstos se han ido corrigiendo con el paso del tiempo.

El cálculo de los ECMs de los estimadores obtenidos con las series del BBVA y de la CRE-e se realiza trabajando con las tasas de crecimiento anual (calculadas como diferencias logarítmicas) de las series estandarizadas de VAB y puestos de trabajo por habitante. Puesto que los datos del BBVA sólo están disponibles en años alternos, he utilizado sólo los años impares de cada una de las series. Las primeras diferencias logarítmicas de la serie bienal se dividen por dos para obtener las tasas de crecimiento anualizadas con las que se realizan los cálculos.

El Gráfico 3 muestra (en escala logarítmica) el ratio entre las varianzas de las tasas de crecimiento anualizadas del VAB y el empleo per cápita relativos en nuestras dos fuentes, con el INE en el numerador. El patrón temporal de las volatilidades relativas es similar al observado en el Gráfico 1 para el caso del PIB per cápita: las series de la CRE-e son generalmente mucho más volátiles que las del BBVA durante los primeros años del período y se acercan después gradualmente a estas últimas, llegando incluso a presentar varianzas inferiores en la parte final de la muestra.

Gráfico 3: Varianza CRE/Varianza BBVA, series en tasas de crecimiento anualizadas (escala logarítmica)

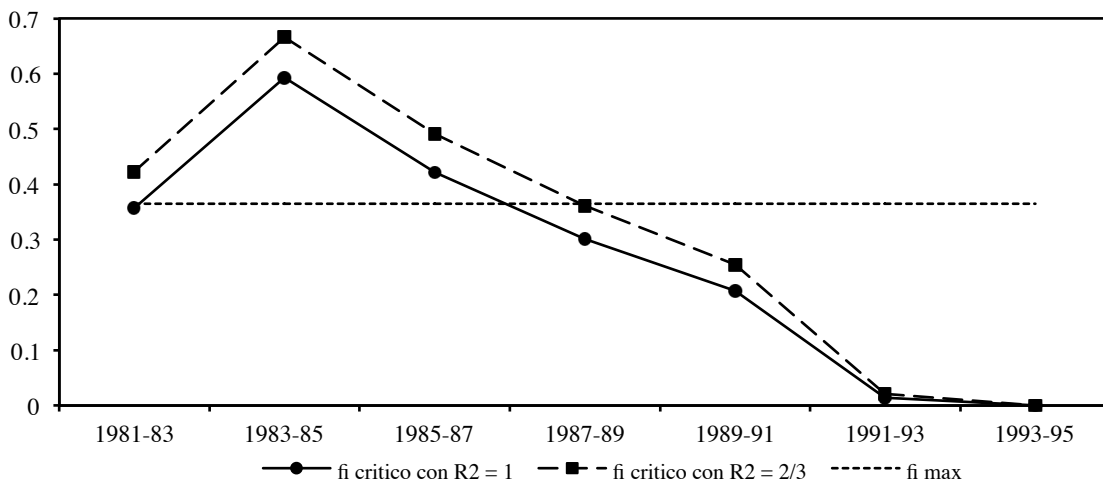


Como hemos visto en el apartado anterior, esto implica que las series del BBVA han de considerarse “mejores” que las de la CRE durante estos años siempre que no presenten un sesgo sistemático hacia la media demasiado elevado. El Gráfico 4 muestra los valores estimados del nivel crítico del parámetro 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 seguramente también la siguiente (1987-89). En el caso del empleo, las estimaciones de los parámetros relevantes sugieren un rechazo claro en 1985-87 y 1987-89 pero no así en los dos bienios anteriores. Sin embargo, no parece probable que el sesgo sistemático de la serie del BBVA pueda ser mucho

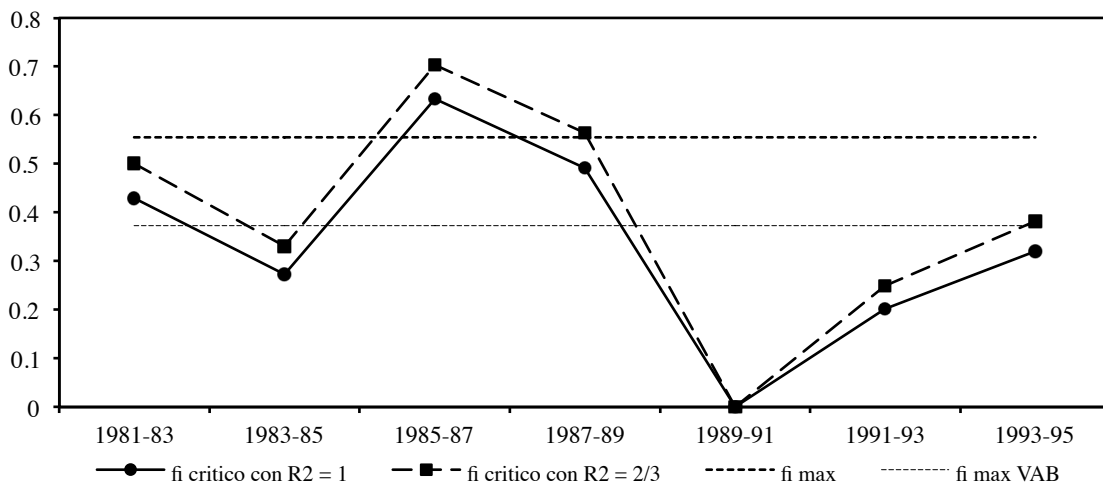
mayor en la estimación del empleo que en la del VAB. Si utilizamos para la serie de empleo la misma cota de ϕ que para la serie de VAB, las dos primeras observaciones también se sitúan por encima o cerca del umbral de rechazo. En conjunto, por tanto, parece razonable fijar la fecha de corte en 1989. Esto es, la serie final que queremos construir se identifica con la de la CRE-e de 1989 en adelante y se extiende hacia atrás desde esta fecha enlazándola con la del BBVA por el procedimiento que se describirá más adelante.

Gráfico 4: Valor crítico y cota superior de ϕ

a. VAB per cápita



b. Puestos de trabajo per cápita



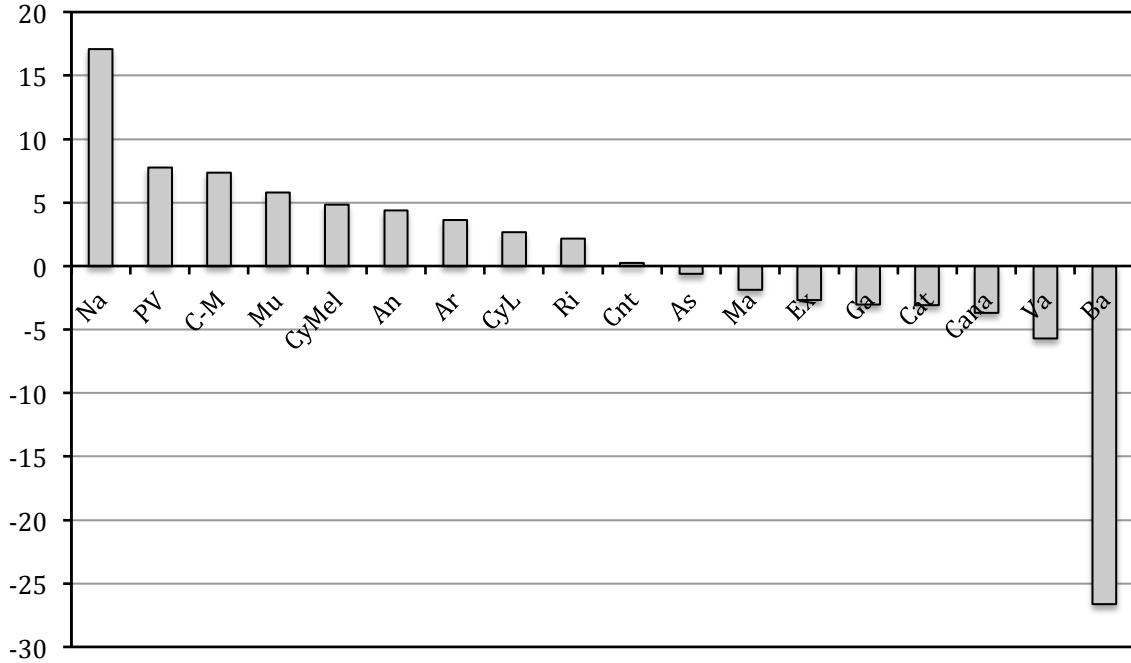
3.3. Discrepancias entre ambas fuentes en el punto de corte

El Gráfico 5 muestra la discrepancia existente en 1989 entre las series normalizadas de VAB y empleo per cápita de las dos fuentes con las que estamos trabajando. En el caso del VAB, las diferencias son llamativamente grandes en dos regiones: Baleares, que “pierde” más de veinticinco puntos de output per cápita relativo al pasar de la serie del BBVA a la de la CRE-e y Navarra, que “gana” más de quince puntos con el cambio. En el caso del empleo, Baleares se mantiene destacada en la parte inferior de la distribución, mientras que la Rioja sustituye a

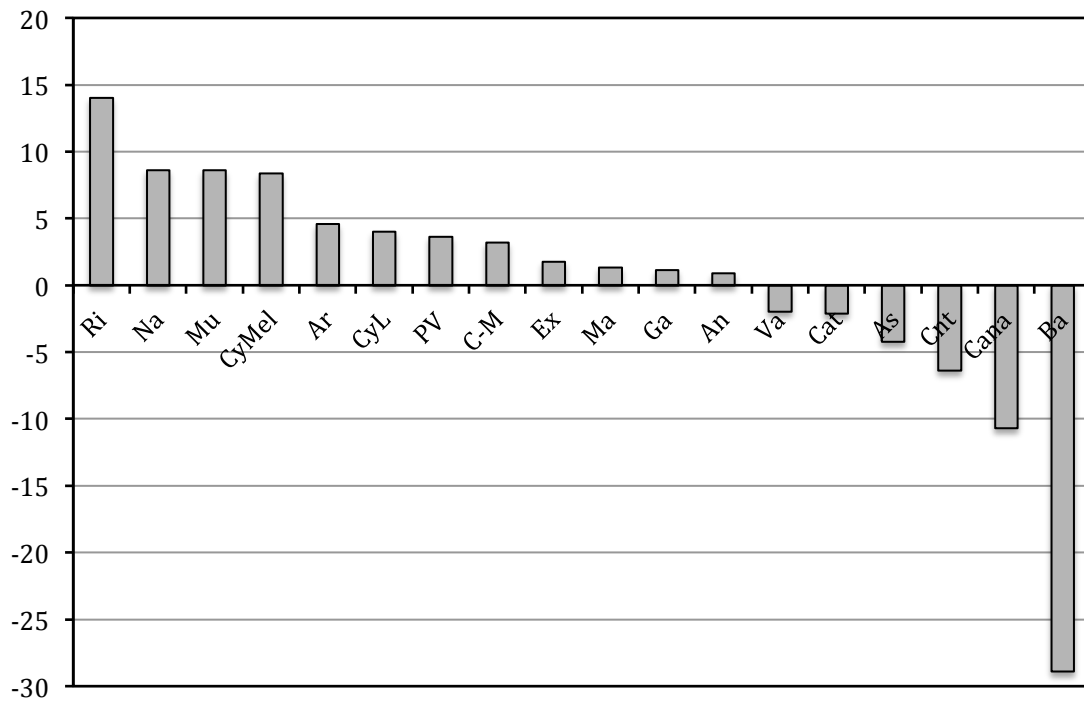
Navarra en el otro extremo. A estas comunidades habría que añadir Canarias, que pierde más de 10 puntos de empleo y Navarra, Murcia y Ceuta y Melilla, que ganan claramente más de 5.

Gráfico 4: Diferencia en 1989 entre las series normalizadas del BBVA y de la CRE-e

a. VAB per cápita relativo a precios corrientes



b. Empleo per cápita relativo



4. Construcción de las series enlazadas de VAB, PIB y puestos de trabajo

En esta sección se construyen series homogéneas de empleo y de output para el período 1955-2014 mediante el enlace de las series del BBVA con las de la CRE-e fijando como punto de corte el año 1989. Las series a enlazar son el número de puestos de trabajo, el VAB (a coste de los factores/precios básicos) y el PIB a precios de mercado. Una vez enlazados el VAB y el PIB a precios corrientes, ambas variables se deflactan utilizando series de índices de precios que se obtiene por retropolación a partir de los deflatores de la CRE-e y del BBVA.

Buena parte de los cálculos necesarios se realiza con series normalizadas, expresadas en términos relativos al promedio nacional por habitante. De esta forma, se neutralizan las posibles diferencias de nivel entre las distintas series y se utiliza únicamente la información que éstas contienen sobre la posición relativa de cada territorio. Los enlaces del empleo y el VAB a precios corrientes se realizan en primera instancia con las series así normalizadas, recuperándose después los niveles por habitante y los valores totales de cada magnitud a partir de la población de derecho y de los promedios nacionales de las variables de interés, que se toman de las series agregadas construidas en de la Fuente (2016a).

4.1. Procedimiento de enlace

La mayor parte de los enlaces que se realizan en este trabajo se construyen por el llamado *procedimiento mixto*, desarrollado en de la Fuente (2014). Esta metodología proporciona una vía intermedia entre los dos procedimientos más habituales en la literatura: la *interpolación* y la *retropolación*. El nombre del método hace referencia al hecho de que éste, a diferencia de los otros dos citados, reparte a lo largo del tiempo la discrepancia entre las dos series que aflora en el momento del enlace de una forma que afecta tanto a las tasas de crecimiento de la serie más antigua como a su valor en el año base -- mientras que la retropolación conserva la primera característica y la interpolación la segunda. El método mixto parte del supuesto de que el "error de medición" que presenta la serie más antigua es proporcional a un componente "mal medido" de la variable de interés cuya tasa bruta de crecimiento es un múltiplo constante de la tasa bruta de crecimiento del resto del agregado correspondiente.

Para concretar, consideremos los detalles de la construcción del enlace entre dos series, X_t e Y_t . La serie de construcción más reciente y en principio más fiable, Y_t , comienza en T , mientras que la serie más antigua, X_t , comienza en 0 y se extiende al menos hasta T . En el punto de enlace, T , los valores de ambas series generalmente no coincidirán debido a la introducción de cambios metodológicos o a la incorporación de nueva información de base en la construcción de la nueva serie. Típicamente, esto se traduce en el "afloramiento" de nueva actividad, de forma que $Y_T > X_T$. Llamaré D_T a la discrepancia existente entre la serie nueva y la serie antigua en el momento del enlace, y

$$(8) d_T = \ln Y_T - \ln X_T = y_T - x_T$$

a la diferencia proporcional o logarítmica entre ambas en el mismo momento.

Los dos procedimientos más habituales de enlace siguen un esquema común. Tanto en la interpolación como en la retrapolación, la serie enlazada se construye añadiendo a la serie más antigua una estimación del error de medición contenido en la misma que se basa en la única observación directa de esta magnitud de la que se dispone, la correspondiente al momento del enlace. Así pues, la serie enlazada por el método j vendrá dada por

$$(9) \hat{y}_t^j = x_t + \hat{d}_t^j \quad \text{para } 0 \leq t \leq T \quad \text{con } \hat{d}_t^j = \begin{cases} \hat{d}_t^r = d_T \\ \hat{d}_t^i = \frac{t}{T} d_T \end{cases}$$

En el caso de la retrapolación ($j = r$) se supone que el error de medición que contiene la serie más antigua se mantiene constante en el tiempo, mientras que en el de la interpolación ($j = i$) la hipótesis es que el error se ha generado enteramente entre 0 y T , y que lo ha hecho a un ritmo uniforme.

La diferencia entre los dos procedimientos estándar también puede expresarse en términos de las hipótesis implícitas en cada uno de ellos sobre el valor del error de medición contenido en la serie más antigua en el momento 0. En la interpolación se supone que tal error es nulo ($d_o = 0$), mientras que en la retrapolación la hipótesis es que este error es igual al observado en el momento del enlace ($d_o = d_T$). En el procedimiento mixto que utilizaremos en este trabajo, se elige una opción más flexible consistente en parameterizar el error de medición inicial de una forma que permita acomodar cualquier hipótesis intermedia. Tendremos ahora

$$(10) \hat{d}_o^m = \rho d_T$$

donde ρ es un parámetro, cuyo valor podemos fijar libremente, que mide la magnitud del error inicial. En ocasiones, resulta útil observar que este procedimiento equivale a fijar el valor inicial de la serie enlazada, \hat{y}_o^m , como una media ponderada de los valores iniciales de la serie más antigua y de la serie enlazada que se obtendría por retrapolación,

$$(11) \hat{y}_o^m = x_o + \hat{d}_o^m = x_o + \rho d_T = (1 + \rho - \rho)x_o + \rho d_T = (1 - \rho)x_o + \rho(x_o + d_T) = (1 - \rho)x_o + \rho \hat{y}_o^r$$

La segunda modificación que se introduce en el método mixto tiene que ver con el perfil temporal del error de medición de la serie más antigua. En la retrapolación se supone que este error se mantiene constante y en la interpolación que aumenta linealmente con el tiempo. Sin embargo, ninguna de estas opciones resulta demasiado plausible, al menos si la fuente del error es, como parece probable, la deficiente medición de ciertas actividades cuyo peso relativo aumenta con el paso del tiempo. En este caso, el crecimiento del error dependerá del ritmo relativo de crecimiento de tales actividades en relación con el resto de la economía. Si suponemos que la razón entre ambas tasas de crecimiento es aproximadamente constante, podemos modelizar la evolución del error de una forma sencilla, lo que permite derivar la siguiente expresión para la senda del error proporcional de medición:⁴

⁴ Véase de la Fuente (2014) para más detalles.

$$(12) d_t \equiv d_o \left(\frac{1}{\rho} \right)^{\frac{t}{T}} = \rho d_T \left(\frac{1}{\rho} \right)^{\frac{t}{T}} = d_T \rho^{\frac{T-t}{T}}$$

Por tanto, la serie enlazada por el procedimiento mixto vendrá dada por

$$(13) \hat{y}_t^m = x_t + \hat{d}_t^m \text{ para } t \leq T \text{ con } \hat{d}_t^m = d_T \rho^{\frac{T-t}{T}}$$

El problema obvio que se plantea con el método mixto es el de cómo fijar el valor de ρ . Siempre que resulte posible, en lo que sigue se intentará basar esta decisión en algún tipo de información externa que nos pueda dar alguna pista sobre el valor de la variable relevante hacia el comienzo del período muestral.

Enlaces de series de precios y magnitudes "reales"

Cuando X e Y son magnitudes monetarias, se plantea también la cuestión de cómo enlazar sus valores "reales" (en volumen o a precios constantes del año base). En principio, esto puede hacerse de dos formas alternativas: enlazando los correspondientes índices de precios para deflactar después la serie enlazada a precios corrientes, o enlazando directamente las magnitudes reales para recuperar seguidamente un índice de precios para la serie enlazada.

En el caso de las series de precios, el procedimiento natural de enlace a nuestros efectos es la retroprolación. La razón es que la única información real que contienen las series de precios regionales que existen para España es la que recogen sus variaciones a lo largo del tiempo. Su nivel, sin embargo, es totalmente arbitrario y generalmente se fija igualando a la unidad para todas las regiones el valor del índice en el año que se toma como base. Las discrepancias de nivel entre series de precios en el momento del enlace, por tanto, tampoco contienen información alguna porque reflejan únicamente el hecho de que las series a enlazar tienen distintos años base. Si eliminamos esta diferencia, tomando como referencia común para ambas series el año del enlace, la diferencia entre ellas en ese año será siempre cero por definición. No tiene mucho sentido, por tanto, pensar en cómo debemos distribuir el "error" aflorado en el momento del enlace a lo largo del tiempo.⁵

En consecuencia, la forma más sencilla de construir series enlazadas de magnitudes reales consiste en enlazar en primer lugar las magnitudes medidas a precios corrientes por el procedimiento que se considere más oportuno para, seguidamente, enlazar por retroprolación los correspondientes índices de precios y, finalmente, deflactar la serie nominal construida en la primera etapa utilizando esta serie de precios.

4.2. Construcción de las series largas de empleo y VAB regional para años impares

En esta sección se construyen series largas de empleo, VAB a coste de los factores/precios básicos y PIB a precios de mercado para el período 1955-2007 mediante el enlace de las

⁵ La situación sería diferente si dispusiéramos de índices que capturasen las diferencias existentes entre regiones en niveles de precios y no sólo en sus tasas de crecimiento. En tal caso, los niveles de las series regionales en relación al deflactor agregado sí contendrían información útil y podría ser conveniente utilizar algún otro procedimiento de enlace.

correspondientes series de la FBBVA y la CRE-e. Las dos últimas variables se miden tanto a precios corrientes como a precios constantes de 2010.

El ejercicio se realiza en varias etapas. En primer lugar, se enlazan las series de VAB a precios corrientes y de puestos de trabajo por habitante, trabajando con series normalizadas por el promedio nacional y referidas únicamente a años impares, que son los únicos para los que el BBVA ofrece información. Las series de VAB a precios corrientes y de puestos de trabajo de las dos fuentes utilizadas se expresan en primer lugar en términos per cápita, dividiendo cada una de ellas por la serie enlazada "homogénea" de población de derecho (no corregida) tomada de de la Fuente (2016c) y se transforman a continuación en índices, normalizando cada magnitud por su promedio nacional en el mismo año (excluyendo a la extra-regio⁶). Los enlaces del empleo y el VAB a precios corrientes se realizan en primera instancia con las series así normalizadas, recuperándose después los niveles por habitante a partir del correspondiente promedio nacional, calculado con las series agregadas enlazadas construidas en de la Fuente (2016a) tras ajustar estas últimas para excluir la extra-regio. Finalmente, estas magnitudes se multiplican por la población de derecho de cada región para obtener los totales nacionales de VAB y empleo.

a. Discrepancias entre la serie del BBVA y la serie de la CRE-e retropolada

Para construir las series enlazadas de VAB y empleo relativos, necesitamos asignar un valor al parámetro ρ que aparece en las ecuaciones (10)-(13). Como ya se ha visto, este parámetro puede interpretarse como el peso que se le asigna a la serie de la CRE-e retropolada hasta 1955 cuando fijamos el valor inicial de la serie enlazada como una media ponderada de esta variable y del valor original de la serie del BBVA, esto es

$$(11') \hat{y}_o^m = (1 - \rho)x_o + \rho\hat{y}_o^r$$

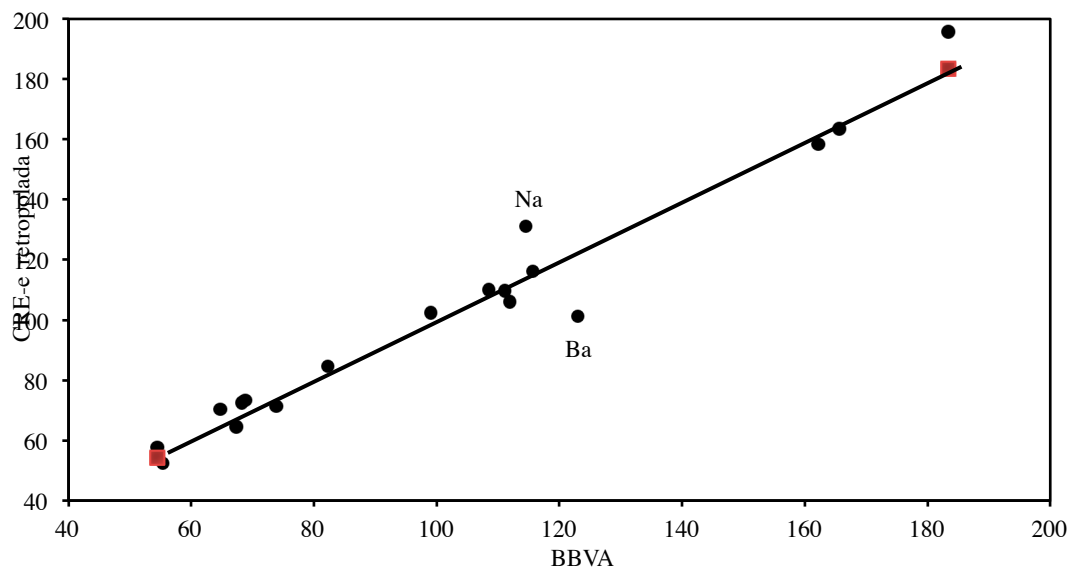
Para comenzar, resulta interesante comprobar las diferencias existentes entre estas dos series. Recordando que

$$(9') \hat{y}_o^r = x_o + d_{89}$$

resulta sencillo calcular el valor correspondiente a 1955 de la serie de la CRE-e retropolada y compararlo con el valor correspondiente al mismo año de la serie original de la Fundación BBVA para 1955. Los resultados de la comparación se resumen en los Gráficos 5 y 6 y los Cuadros 2 y 3.

⁶ La "extra-regio" o territorio extrarregional es un territorio ficticio al que se imputan aquellas actividades económicas que no se pueden asignar a una región determinada. En ella se incluyen, por ejemplo, las embajadas españolas y las bases militares y científicas en países extranjeros.

Gráfico 5: VAB a precios corrientes per cápita en 1955.
CRE-e retropolada vs. BBVA



Cuadro 2: VAB per cápita en 1955 a precios corrientes
CRE-e retropolada vs BBVA

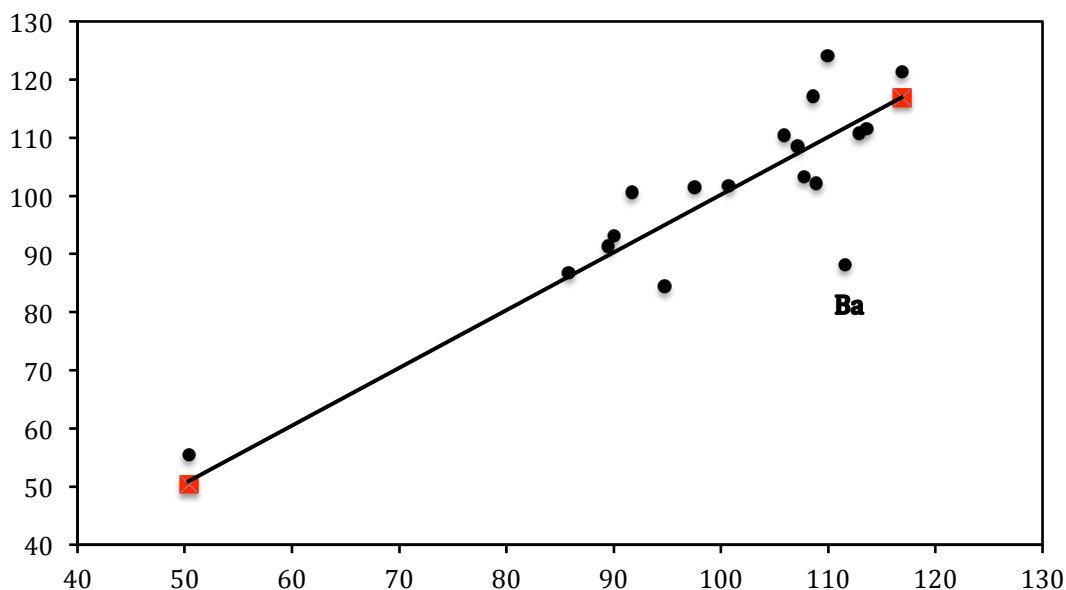
	BBVA	rank BBVA	CRE-e	rank CRE-e	CRE- BBVA	%	rankBBVA- rank CRE
<i>País Vasco</i>	183,3	1	196,0	1	12,7	6,7%	0
<i>Madrid</i>	165,3	2	162,9	2	-2,4	-1,5%	0
<i>Cataluña</i>	162,3	3	158,3	3	-4,0	-2,5%	0
<i>Baleares</i>	122,9	4	100,5	10	-22,4	-20,1%	-6
<i>Cantabria</i>	115,7	5	116,0	5	0,3	0,2%	0
<i>Navarra</i>	114,5	6	131,4	4	16,9	13,8%	+2
<i>Valencia</i>	111,9	7	105,8	8	-6,1	-5,6%	-1
<i>Asturias</i>	111,2	8	110,4	7	-0,8	-0,7%	+1
<i>Rioja</i>	108,4	9	110,5	6	2,1	1,9%	+3
<i>Aragón</i>	99,1	10	102,4	9	3,3	3,3%	+1
<i>Castilla y León</i>	82,3	11	84,8	11	2,5	3,0%	0
<i>Canarias</i>	73,7	12	71,0	14	-2,7	-3,7%	-2
<i>Murcia</i>	68,8	13	73,5	12	4,8	6,7%	+1
<i>Andalucía</i>	68,2	14	72,5	13	4,3	6,1%	+1
<i>Galicia</i>	67,3	15	64,8	16	-2,5	-3,8%	-1
<i>Cast. la Mancha</i>	64,9	16	70,9	15	6,0	8,9%	+1
<i>Extremadura</i>	55,4	17	53,2	18	-2,2	-4,1%	-1
<i>Ceuta y Melilla</i>	54,4	18	58,0	17	3,5	6,3%	1

- Nota: CRE-BBVA es la diferencia en puntos de renta relativa entre las dos fuentes, % es la diferencia logarítmica entre las mismas variables.

Un resultado alentador del ejercicio es que los valores iniciales de las dos series que estamos comparando no son demasiado distintos en términos estadísticos. La correlación entre la serie de la CRE-e retropolada y la serie del BBVA en 1955 es de 0,979 en el caso del VAB y de 0,864 en el del empleo. Para la mayoría de las regiones, por tanto, el enlace entre las dos series (por cualquier procedimiento) no exige una revisión radical de la historia. Como ya hemos visto, sin embargo, hay algunas excepciones. El caso más llamativo es el de Baleares. Tomando como

referencia la estimación del BBVA, la retropolación de la serie de la CRE reduce la renta per cápita relativa del archipiélago en 22,4 puntos y su empleo por habitante relativo en 23,3 puntos. En términos de la primera variable, la región pasaría de la cuarta a la décima posición y en términos de la segunda, de la cuarta a la decimoquinta. También son preocupantes las discrepancias sobre Navarra (16,9 puntos en términos de renta y 8,5 en términos de empleo), la Rioja (14,1 puntos en términos de empleo), País Vasco (12,7 puntos de VAB) y Canarias y Murcia (10,3 y 8,9 puntos de empleo).

Gráfico 6: Puestos de trabajo per cápita en 1955. Serie CRE-e retropolada vs. BBVA



Cuadro 3: Puestos de trabajo per cápita en 1955. Serie CRE-e retropolada vs BBVA

	BBVA	rank BBVA	CRE retr.	rank CRE	CRE-BBVA	%	rankBBVA-rank CRE
<i>País Vasco</i>	116.9	1	121.3	2	4.4	3.7%	-1
<i>Valencia</i>	113.6	2	111.5	4	-2.2	-1.9%	-2
<i>Cataluña</i>	112.9	3	110.7	5	-2.2	-2.0%	-2
<i>Baleares</i>	111.6	4	88.2	15	-23.3	-23.5%	-11
<i>Rioja</i>	109.9	5	124.1	1	14.1	12.1%	+4
<i>Cantabria</i>	108.9	6	102.1	9	-6.8	-6.4%	-3
<i>Navarra</i>	108.6	7	117.1	3	8.5	7.6%	+4
<i>Asturias</i>	107.8	8	103.2	8	-4.5	-4.3%	0
<i>Madrid</i>	107.2	9	108.5	7	1.3	1.2%	+2
<i>Aragón</i>	105.9	10	110.4	6	4.5	4.2%	+4
<i>Galicia</i>	100.7	11	101.7	10	1.0	1.0%	+1
<i>Cast. y León</i>	97.5	12	101.5	11	3.9	3.9%	+1
<i>Canarias</i>	94.7	13	84.4	17	-10.3	-11.5%	-4
<i>Murcia</i>	91.7	14	100.6	12	8.9	9.2%	+2
<i>C. la Mancha</i>	90.0	15	93.1	13	3.1	3.4%	+2
<i>Extremadura</i>	89.5	16	91.4	14	1.9	2.1%	+2
<i>Andalucía</i>	85.8	17	86.8	16	1.0	1.1%	+1
<i>Ceuta y Melilla</i>	50.4	18	55.5	18	5.1	9.6%	0

- Nota: CRE-BBVA es la diferencia en puntos de empleo relativo entre las dos fuentes, % es la diferencia logarítmica entre las mismas variables.

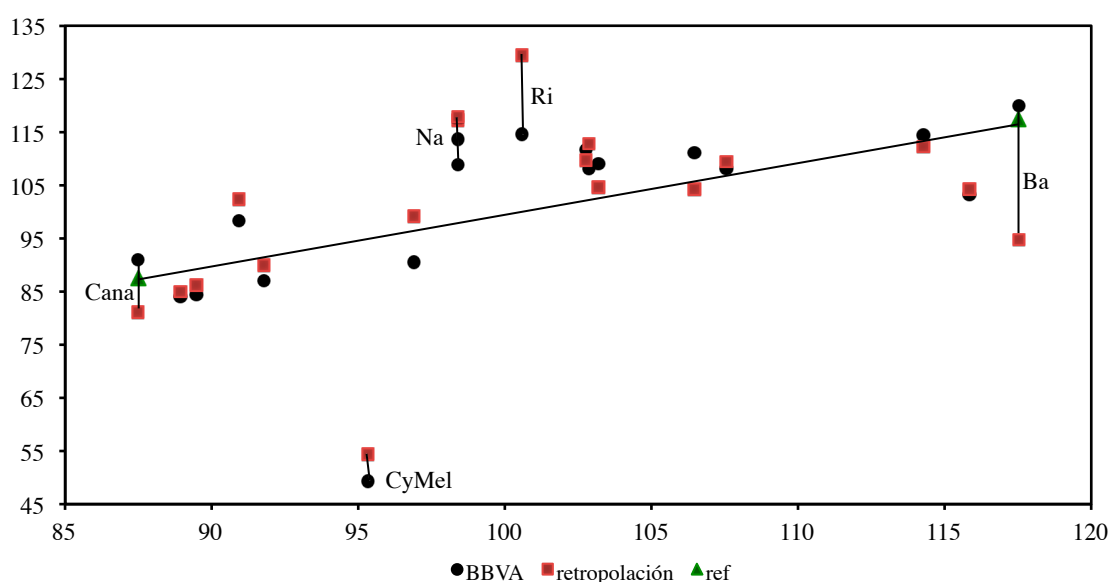
b. Comparación con fuentes externas y selección del valor de ρ

Para intentar basar el valor del parámetro ρ en algo más firme que la mera conjetura, he intentado buscar fuentes “externas” de información que nos puedan dar alguna idea de la fiabilidad relativa de las dos estimaciones alternativas de las que disponemos para la renta y el empleo iniciales: la estimación de la Fundación BBVA para el primer año del período muestral y el valor correspondiente al mismo ejercicio de la serie de la CRE retropolada.

Puestos de trabajo relativos por habitante

En el caso del empleo, el Censo de 1960 nos proporciona una estimación externa en principio fiable de la tasa de ocupación relativa de las regiones españolas al comienzo del período muestral. Un problema con esta fuente es que en la misma no se identifica a los parados, que parecen incluirse en una categoría residual dentro de la población activa. Por lo tanto, el número de parados en 1960 (calculado como el promedio de los registrados a finales de los meses de junio y de diciembre) se toma del *Anuario Estadístico de España* de 1961. Substrayendo esta cifra de la población activa censal, se obtiene la población ocupada, que se divide por la población de derecho de la misma fuente y se normaliza por el promedio nacional para obtener el dato más cercano posible a las series normalizadas de empleo con las que estamos trabajando. El dato censal así obtenido se compara con una estimación del empleo per cápita relativo de cada región en 1960 de acuerdo con la serie del BBVA y con la serie de la CRE retropolada, que se obtiene como el promedio de los valores correspondientes a 1959 y 1961 de cada una de estas series.

**Gráfico 7: Comparación con el censo de 1960.
Serie de empleo del BBVA y serie CRE-e retropolada**



El Gráfico 7 muestra los resultados. La serie de la CRE-e retropolada y la serie del BBVA se miden en la escala vertical, mientras que el dato censal se mide en el eje horizontal. Puesto que la recta que aparece en el gráfico corresponde a la diagonal o línea de 45° , la distancia vertical entre cada punto y esta recta mide la diferencia entre cada una de las series de interés y el dato

censal correspondiente. En términos generales, la serie del BBVA se parece más al dato censal que la serie retropolada. En aquellos casos en los que hay diferencias significativas entre ambas fuentes, los puntos negros que representan a la primera tienden a estar más cerca de la recta de 45° que los cuadrados rojos que corresponden a la serie retropolada. De la misma forma, la correlación con el dato censal (excluyendo a Ceuta y Melilla) es bastante más alta en el caso del BBVA (0,784) que en el de la serie enlazada (0,458).

En principio, por tanto, parece razonable asignar un peso superior a la estimación del BBVA que a la serie retropolada, esto es, fijar un valor reducido para ρ a la hora de realizar el enlace por el método mixto. En consecuencia, he asignado un valor de 0,10 a este parámetro.

Ceuta y Melilla

Ceuta y Melilla constituyen un caso muy atípico. Como se observa en el Gráfico 7, tanto la retropolación de la serie terminal de empleo como la estimación del BBVA toman un valor muy inferior al dato censal de ocupación (en torno a 40 puntos), lo que aconseja examinar en mayor detalle la información disponible sobre la evolución del empleo en este territorio.

Cuadro 4: Ceuta y Melilla, empleos por habitante relativos, diversas fuentes (España = 100)

	1960	1970	1981	1991
<i>Censo</i>	95,3	79,8		78,2
<i>BBVA</i>	49,4	58,4	77,2	86,6
<i>CRE-e</i>			79,8	83,3

- *Nota:* los datos del BBVA para años censales pares se obtienen promediando los valores correspondientes al año anterior y posterior a cada censo. Los censos de 1960 y 1970 contienen información sobre la población activa pero no sobre el número de parados. Este último dato se toma del *Anuario Estadístico de España*. Los datos censales corresponden al número de personas ocupadas por habitante de derecho, mientras que los del BBVA y la CRE-e se basan en el número de puestos de trabajo.

El Cuadro 4 muestra la información que he podido reunir sobre la evolución del empleo relativo de las dos ciudades autónomas en años seleccionados (medido por el número de ocupados o por el número de puestos de trabajo por habitante de derecho en relación al promedio nacional). Los datos censales sugieren un fuerte descenso del empleo relativo entre 1960 y 1970, seguido de una estabilización a un nivel inferior en unos veinte puntos al promedio nacional. La serie del BBVA, por contra, indica un incremento fuerte y sostenido de esta variable durante todo el período muestral a partir de un nivel inicial muy inferior al promedio nacional. Puesto que la estimación del BBVA coincide aproximadamente con la de la CRE-e a partir de 1980 y no está muy lejos del dato censal de 1991, parece razonable conservarla como serie de referencia desde el primer año citado hasta el punto de corte. En la primera parte del período muestral, sin embargo, la fuerte discrepancia con el dato censal aconseja algún tipo de corrección al alza.

En consecuencia, he optado por revisar la serie de empleo relativo del BBVA para Ceuta y Melilla antes de proceder a su enlace con la serie terminal de la CRE-e. La nueva serie de referencia se construye como sigue. De 1981 en adelante, se mantiene la serie original del BBVA.

Para 1960 y 1970 la nueva serie se obtiene como una media ponderada de las estimaciones del BBVA (con peso 2/3) y del correspondiente dato censal (con peso 1/3). El dato de 1960 se extiende hacia atrás hasta 1955 por retropolación, utilizando las tasas de crecimiento de la serie del BBVA. Entre 1960 y 1970 la serie se completa respetando en lo posible el perfil temporal de la serie original del BBVA, pero reescalando sus incrementos de forma que la serie pase por los puntos deseados en los años citados. Esto es, para cada t entre 1960 y 1970 construimos

$$(14) \Delta_t = \frac{EMBBVA_t - EMPBBVA_{1960}}{EMPBBVA_{1970} - EMPBBVA_{1960}}$$

de forma que Δ_t mide la fracción del incremento total del empleo durante la década de interés que se ha producido ya en el año t . El valor en t de la nueva serie de referencia de empleo se estima como

$$(15) EMREF_t = EMREF_{1960} + \Delta_t (EMREF_{1970} - EMREF_{1960})$$

El mismo procedimiento se utiliza también para completar la serie ente 1970 y 1981, año éste en el que se vuelve a la serie original del BBVA. La serie así construida de índices se utilizar para recuperar una serie corregida en niveles que reemplaza a la serie original del BBVA como serie de referencia entre 1955 y 1981 para Ceuta y Melilla. Esta serie se enlaza por el procedimiento habitual con la serie terminal de la CRE-e utilizando un valor de 0.10 para ρ .⁷

VAB relativo por habitante

En el caso del VAB resulta más complicado encontrar una estimación verdaderamente externa del valor de esta magnitud cerca del comienzo del período muestral. Alvarez Llano (1986) y Martín Rodríguez (1992 y 1993) ofrecen estimaciones del nivel de renta de las regiones españolas en 1950 o 1960 como parte de ambiciosos intentos por aproximar la evolución económica de estos territorios durante un período de varios siglos. Para los años que nos interesan, sin embargo, ambos autores se apoyan en datos del BBVA o en trabajos de Alcaide relacionados con ellos. La única estimación aparentemente independiente de esta última fuente que he podido encontrar es la que construye Plaza Prieto (1953) para el año 1949 utilizando procedimientos indirectos para provincializar por ramas una estimación del producto agregado español. Dadas las enormes limitaciones de la información disponible, existen serias dudas sobre la fiabilidad de la estimación de Plaza. Sin embargo, cabe esperar que estas limitaciones sean algo menos importantes a la hora de aproximar la situación relativa de cada uno de los territorios y, en cualquier caso, parece tratarse de la única fuente externa existente.

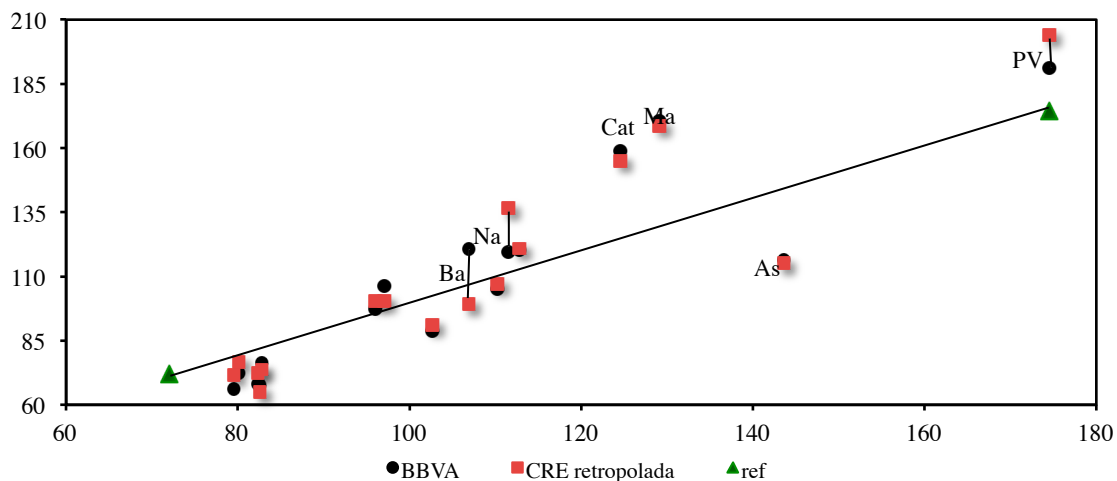
Las estimaciones del BBVA y la retropolación de la serie terminal de la CRE se comparan con los datos de Plaza Prieto en el Gráfico 8, que se construye de la misma forma que el Gráfico 7, con la serie de Plaza medida en el eje horizontal y las otras dos en el eje vertical.⁸

⁷ No hace falta insistir en que la serie de empleo para Ceuta y Melilla que aquí se ofrece es de dudosa fiabilidad en sus primeros años. En aplicaciones, podría ser prudente excluir a este territorio de la muestra, especialmente si constituye una observación atípica en materia de empleo.

⁸ La serie de VAB del BBVA se extiende hacia atrás desde 1955 hasta 1950 utilizando la tasa de crecimiento de la misma variable durante este período de acuerdo con Alcaide (2007), donde se ofrecen estimaciones

En primer lugar, hay que resaltar que la correlación entre las estimaciones de este autor y las otras dos series es muy elevada y adopta un valor similar en ambos casos: 0,908 con la serie del BBVA y 0,915 con la serie terminal retropolada. En algunos territorios, estas dos últimas series se alejan significativamente de la estimación de Plaza pero, en términos generales, las dos fuentes coinciden entre sí cuando esto sucede. Los casos más problemáticos son los del País Vasco, Navarra, y Baleares (véase el Cuadro 4). En los dos primeros territorios, la estimación del BBVA está más cercana a la de Plaza, mientras que en el último sucede lo contrario.

**Gráfico 8: Comparación con la estimación de Plaza Prieto.
Serie de VAB del BBVA y serie terminal retropolada, 1950**



La información disponible, por tanto, sugiere que la serie del BBVA sobrevalora significativamente la renta de Baleares a mediados del siglo pasado,⁹ pero no necesariamente infravalora la de las otras dos regiones citadas. A la vista de estos resultados, el valor de ρ se ha fijado como sigue. Como criterio general, he elegido un valor de 0,5, lo que supone dar el mismo peso a la serie del BBVA y a la retropolación de la serie terminal a la hora de estimar el nivel inicial de renta. Las excepciones son el País Vasco y Navarra, para las que he fijado $\rho = 1/3$, otorgando así un peso mayor a la estimación del BBVA que a la retropolación de la serie terminal, y Baleares, para la que utilizaré $\rho = 2/3$, invirtiendo las ponderaciones de ambas fuentes.

c. Enlace de las series de empleo y de VAB a precios corrientes

El enlace entre las series del BBVA y de la CRE-e de puestos de trabajo y de VAB ccf/pb a precios corrientes se realiza por el procedimiento mixto, utilizando los valores de ρ indicados en la sección anterior y resumidos en el Cuadro 5. El resultado obtenido por este procedimiento

de renta regional a intervalos quinquenales para todo el siglo XX. El valor de la serie terminal retropolada en el mismo año se obtiene añadiendo al dato estimado del BBVA en 1950 la diferencia logarítmica entre la serie del BBVA y la serie terminal en 1989. El dato de Alcaide (2007) para 1950 no se utiliza directamente porque existen pequeñas discrepancias entre esta fuente y las series del BBVA en 1955.

⁹ Maluquer (2006) también considera muy poco plausibles las estimaciones de renta la FBBVA para Baleares.

se renormaliza de forma que el promedio nacional (ponderado por población) calculado a partir de las nuevas series regionales enlazadas sea exactamente igual a 100 en cada período.

Cuadro 5: Valores de ρ utilizados para enlazar las distintas series

	VAB	empleo
<i>caso general</i>	0.50	0.10
excepciones:		
<i>País Vasco</i>	1/3	
<i>Navarra</i>	1/3	
<i>Baleares</i>	2/3	

Una vez obtenidas las series definitivas de VAB y empleo por habitante en términos relativos, los niveles absolutos de VAB y empleo por habitante se recuperan aplicando los índices enlazados al correspondiente promedio nacional (excluyendo la extra regio).¹⁰ La cifra resultante se multiplica a su vez por la población de derecho regional para recuperar el empleo total de cada territorio (medido en puestos de trabajo) y su VAB a precios básicos (medido en euros corrientes).

4.3. Construcción de la serie de PIB nominal

La serie de PIB nominal a precios de mercado se obtiene sumando a la serie enlazada de VABcf/pb una serie de impuestos netos sobre los productos. Esta serie también se obtiene mediante el enlace entre las series de la CRE-e y las del BBVA con punto de corte en 1989. El enlace de estas series se realiza por interpolación, respetando el valor original de la serie del BBVA en 1955 y el de la serie de la CRE-e en 1989. La diferencia entre las dos fuentes en el punto de corte es modesta en el agregado (en torno a un 5%) pero sorprendentemente grande en algunas regiones, excediendo el 40% en Baleares y Ceuta y Melilla. Esto plantea algunas dudas sobre la territorialización de esta variable, lo que me lleva a considerar la serie de VAB como más fiable que la serie de PIB.

4.4. Deflatores y series de VAB y PIB a precios constantes

Las series de deflatores del VAB de la CRE-e y de la Fundación BBVA se enlazan por retroprolación, extendiendo desde 1989 hacia atrás la serie de precios del VABpb de la CRE-e utilizando las tasas de crecimiento de las series de precios del VAB del BBVA, manteniendo el punto de corte entre ambas fuentes en 1989, el año fijado en una sección anterior.¹¹ En esta

¹⁰ Estos promedios se construyen utilizando los totales nacionales de VAB, PIB y puestos de trabajo estimados en de la Fuente (2016a), de los que se subtrae una estimación de la parte que corresponde a la extra regio. Desde 1980 en adelante, esta última estimación se toma de de la Fuente (2016b). Para años anteriores la producción y empleo de la extra regio se aproximan a partir del supuesto de que su peso en el total nacional es el mismo que el observado en 1980 (el 0,083% en el caso del VAB y el 0,036% en el del empleo medido en puestos de trabajo). El mismo procedimiento se aplica al resto de variables de interés.

¹¹ 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

ocasión, el enlace no se realiza con las series normalizadas por el promedio nacional sino con índices "brutos" de precios para cada región.

Una vez construida una serie preliminar de precios por este procedimiento, esta serie se utiliza para deflactar la serie enlazada de VABpb nominal, obteniendo una serie preliminar de VAB real valorada (aproximadamente) a precios constantes de 2010. Esta serie preliminar se ajusta proporcionalmente de forma que su suma sobre regiones coincida con la serie nacional de VABpb a precios constantes construida en de la Fuente (2016a), de la que se sustrae una estimación de la parte que corresponde a la extra regio.¹² Dividiendo la serie enlazada de VAB a precios corrientes por esta serie final de VAB a precios constantes, se recupera la serie final de precios en base 2010.

Las series de deflactores del PIB y de PIB a precios de 2010 se construyen por el mismo procedimiento. En ambos casos, se recupera también un deflactor para el agregado de España sin la extra-regio como el cociente entre las correspondientes series de VAB y PIB a precios corrientes y constantes de 2010.

5. Construcción de series anuales

Puesto que la Fundación BBVA sólo ofrece datos para años impares, las series enlazadas construidas en la sección anterior heredan esta característica durante el período en el que las series del BBVA constituyen la principal referencia para su construcción, esto es, desde 1955 hasta 1989. Dado que esta característica de los datos es una limitación importante de cara a algunas aplicaciones, en esta sección 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 sobre las variables relevantes.

5.1. Procedimiento de interpolación

Sea X_{it} una de las series regionales de interés y X_t el correspondiente agregado nacional. Supongamos que la serie regional se conoce sólo para años impares y ha de estimarse para los años pares mientras que se dispone de una serie nacional de frecuencia anual para el período completo.

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:

$$(16) \quad g_{it}^1 + g_{it}^2 \equiv (x_{it} - x_{it-1}) + (x_{it+1} - x_{it}) = x_{it+1} - x_{it-1} \equiv \Delta_{it}$$

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).

¹² Como en el caso del VAB y el PIB nominal, el dato de la extra regio se toma de la CRE-e de 1980 en adelante. Para años anteriores, el valor estimado del VAB y el PIB nominal de la extra-regio se deflacta con el correspondiente deflactor nacional.

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

$$(17) \quad g_{it}^1 = g_{it}^2 = \frac{\Delta_{it}}{2} \equiv \bar{g}_{it}$$

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

$$(18) \quad x_{it-1} + \bar{g}_{it} = x_{it-1} + \frac{x_{it+1} - x_{it-1}}{2} = \frac{x_{it+1} + x_{it-1}}{2} \equiv \bar{x}_{it}$$

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 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

$$(19) \quad \phi_t = \frac{x_t - x_{t-1}}{x_{t+1} - x_{t-1}} = \frac{g_t^1}{\Delta_t}$$

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

$$(20) \quad g_{it}^1 = \phi_t \Delta_t$$

y

$$(21) \quad g_{it}^2 = \Delta_t - g_{it}^1 = \Delta_t - \phi_t \Delta_t = (1 - \phi_t) \Delta_t$$

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 de 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

$$(22) \bar{g}_{it}^1 = \phi_t \Delta_{it}$$

lo que podemos escribir también en la forma

$$(23) \bar{g}_{it}^1 = \phi_t \Delta_{it} = \frac{\Delta_{it}}{2} + \left(\phi_t \Delta_{it} - \frac{\Delta_{it}}{2} \right) = \bar{g}_{it} + \left(\phi_t - \frac{1}{2} \right) \Delta_{it}$$

Esto es, cuando el agregado nacional esté por encima de su tendencia en t ($\phi > 1/2$) 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 . 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

$$(24) \tilde{x}_{it} = \bar{x}_{it} + (x_t - \bar{x}_t) = (x_{it-1} + \bar{g}_{it}) + x_t - (x_{t-1} + \bar{g}_t)$$

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

$$(25) \begin{aligned} \tilde{g}_{it}^1 &= \tilde{x}_{it} - x_{it-1} = \bar{g}_{it} + x_t - x_{t-1} - \bar{g}_t = \bar{g}_{it} + (g_t^1 - \bar{g}_t) = \\ &= \bar{g}_{it} + \left(\phi_t \Delta_t - \frac{\Delta_t}{2} \right) = \bar{g}_{it} + \left(\phi_t - \frac{1}{2} \right) \Delta_t \end{aligned}$$

Esta expresión tiene la misma forma que la ecuación (23) 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 a esta pregunta en abstracto, la comparación entre los resultados de los dos procedimientos en ciertos casos concretos sugiere que el método de corrección proporcional tiende a generar un patrón de dientes de sierra en los datos que resulta generalmente poco plausible.¹³ 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.

¹³ Véase de la Fuente (2010). Los problemas surgen cuando la serie nacional presenta cambios de tendencia en años pares, lo que hace que el valor de ϕ_t pueda ser 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.

5.2. 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,

$$(26) \frac{1}{P_t} = \frac{Q_t}{Y_t} = \frac{\sum_i Q_{it}}{Y_t} = \sum_i \frac{Q_{it}}{Y_t} = \sum_i \frac{Y_{it} / P_{it}}{Y_t} = \sum_i \frac{Y_{it}}{Y_t} \frac{1}{P_{it}} = \sum_i sh_t Y_{it} \frac{1}{P_{it}}$$

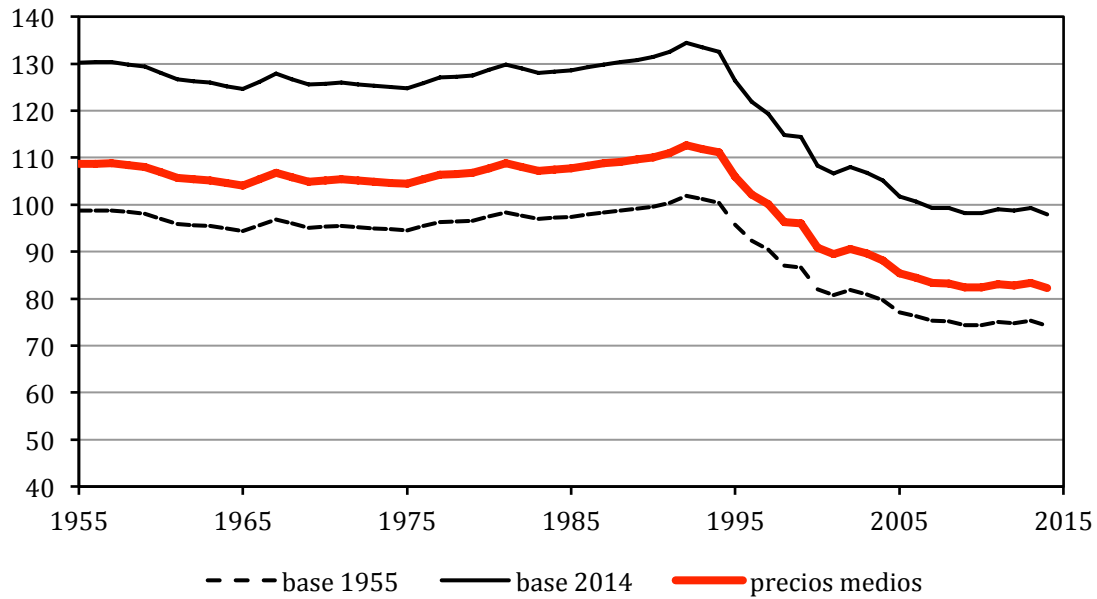
en la que las ponderaciones son las participaciones regionales en el VAB nacional a precios corrientes, $sh_t Y_{it}$. Tras interpolar los niveles de precios regionales por el procedimiento de corrección uniforme, la ecuación (26) se utiliza para calcular el deflactor nacional que se derivaría de las series preliminares de precios regionales. El cociente entre esta variable y el dato directo del 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. Para obtener la serie anual del deflactor del PIB se proceder de la misma forma.

Finalmente, las series anuales de VAB y PIB a precios constantes de 2010 se obtienen aplicando a las series anualizadas de VAB y PIB a precios corrientes la correspondiente serie anual de deflactores.

6. Output a precios medios del período muestral

En ocasiones nos interesa comparar el *nivel* de productividad de distintas regiones además de su tasa de crecimiento. El ejercicio, sin embargo, puede ser problemático cuando se realiza con datos a precios constantes porque los niveles relativos de productividad regional pueden ser muy sensibles al año que se tome como base en la serie de precios. El Gráfico 19 ilustra el problema con datos de Baleares. Puesto que la economía de esta comunidad autónoma es muy intensiva en servicios y los precios de los servicios han crecido bastante más que los de los bienes industriales o agrícolas, Baleares aparenta ser mucho más productiva si su VAB por puesto de trabajo se valora a precios constantes del final del período muestral que cuando se trabaja con los precios del comienzo del mismo.

**Gráfico 19: Producto por puesto de trabajo, Baleares
a precios constantes de distintos años y a precios medios del período muestral
España = 100**



Para intentar mitigar este problema, he construido series de VAB y PIB valoradas a precios medios del período muestral. Esto es, en vez de dividir la serie de *output* valorada a precios corrientes por un índice de precios con base en un año determinado, la serie se deflacta utilizando la media geométrica de los distintos índices de precios que se obtendrían tomando como base cada uno de los años incluidos en el período muestral.

La construcción del deflactor apropiado se realiza como sigue. Sea P_{it}^{BT} el deflactor de la región i correspondiente al año t en base T ($= 2010$ en nuestro caso). Para cambiar la referencia de la serie de precios, obteniendo así un índice "en base S " tenemos que dividir el índice original de precios en base T por el valor correspondiente (también en base T) al nuevo año de referencia,

$$(27) P_{it}^{BS} = \frac{P_{it}^{BT}}{P_{is}^{BT}}$$

o, utilizando minúsculas para indicar que estamos trabajando con logaritmos,

$$(28) p_{it}^{BS} = p_{it}^{BT} - p_{is}^{BT}$$

El índice de precios que queremos calcular es la media geométrica de los que se obtendrían tomando como base cada uno de los años del período muestral, esto es,

$$(29) \bar{P}_{it} = \left(\prod_{n=1}^N P_{it}^{Bn} \right)^{1/N}$$

donde N es el número de observaciones por región. Tomando logaritmos de esta expresión y utilizando (28), tenemos

$$(30) \bar{p}_{it} = \frac{1}{N} \sum_n P_{it}^{Bn} = \frac{1}{N} \sum_n (P_{it}^{BT} - P_{in}^{BT}) = P_{it}^{BT} - \frac{1}{N} \sum_n P_{in}^{BT}$$

Obsérvese que esta ecuación nos permite calcular \bar{P}_{it} directamente a partir de los índices de precios en base 2010 (=T), sin necesidad de construir explícitamente todas las series de precios con base en cada uno de los años del período muestral. Obsérvese también que el resultado no depende del año que se tome como base para la serie de precios. Si partimos de una serie de precios con referencia S, tendremos

$$(31) \bar{p}_{it}^* = P_{it}^{BS} - \frac{1}{N} \sum_n P_{in}^{BS} = (P_{it}^{BT} - P_{iS}^{BT}) - \frac{1}{N} \sum_n (P_{in}^{BT} - P_{iS}^{BT}) = P_{it}^{BT} - \frac{1}{N} \sum_n P_{in}^{BT} = \bar{p}_{it}$$

donde he utilizado (27) para pasar los índices de precios de base S a base T.

Como en el caso anterior, las series regionales de VAB y PIB a precios medios se agregan para recuperar el total nacional y el deflactor nacional se obtiene dividiendo la serie nacional a precios corrientes por la serie a precios medios.

7. Archivos de datos

El trabajo va acompañado de dos archivos de Excel. El primero (*REGDAT_v50_5514pI.xlsx*) contiene las nuevas series enlazadas y el segundo (*datos_originales.xlsx*) todos los datos de otras fuentes que se han utilizado para construir tales series. Los Cuadros 6 y 7 resumen el contenido de ambos archivos. Se incluye también un archivo adicional (*series_bienales_enlazadas_55-95.xlsx*) con las series bienales enlazadas de las principales magnitudes.

Cuadro 6: Contenido del archivo de datos *REGDAT_v50_5514pI.xlsx*

-
- VAB a coste de los factores/precios básicos, medido a precios corrientes, constantes de 2010 y precios medios del período 1955-2014
 - PIB, medido a precios corrientes, constantes de 2010 y precios medios del período 1955-2014
 - Empleo (puestos de trabajo)
 - Población residente a 1 de julio
-

Cuadro 7: Contenido del archivo de datos *datos_originales_v50_I.xlsx*

hoja 1: Series regionales de la Fundación BBVA, 1955-95, años impares

-
- Valor Añadido Bruto a coste de los factores a precios corrientes y constantes de 1986, suma de sectores, antes de substraer la PISB.
 - Producto Interior Bruto a precios corrientes y constantes de 1986
 - Empleo total (puestos de trabajo)
 - Empleo asalariado (puestos de trabajo asalariados)
 - Población residente a 1 de julio
 - Remuneración de asalariados (incluyendo cotizaciones sociales pagadas por el trabajador y por el empleador)
 - Servicios de intermediación financiera imputados a los sectores productivos (producción imputada de servicios bancarios o PISB)
 - Impuestos netos sobre la producción y los productos
 - Serie corregida de puestos de trabajo para Ceuta y Melilla (sección 4.2.b)

Fuentes: Fundación BBV (1999) y Fundación BBVA (2000).

Cuadro 7: continuación

hoja 2: CRE-e, 1980-2014

- Empleo total: personas ocupadas y puestos de trabajo
- Valor Añadido Bruto a precios básicos, medido a precios corrientes y constantes de 2010
- Producto Interior Bruto, medido a precios corrientes y constantes de 2010
- Empleo Asalariado: personas y puestos de trabajo asalariados
- Remuneración de asalariados, a precios corrientes
- Población residente a 1 de julio (desde 1955)

Fuentes: de la Fuente (2016b y c), RegData v4.3 y RegData-Dem v5.0."

hoja 3: Datos de empleo en 1960

- Activos, parados y población de derecho por regiones (y provincias para las dos primeras variables).

Fuentes: Censo de 1960 y Anuario Estadístico de 1961 (para parados)

hoja 4: Estimaciones de renta para 1949

- Renta per cápita en 1949 a precios corrientes por provincias y regiones.

Fuente: Plaza Prieto (1953).

hoja 5: Series nacionales y estimaciones para la extra-regio

- VAB y PIB a precios corrientes y constantes y puestos de trabajo.

Fuente: de la Fuente (2016a y b)

8. Conclusión

En este trabajo se construyen series "homogéneas" largas de algunos agregados económicos regionales mediante el enlace de los datos de la Fundación BBVA con series previamente enlazadas de la Contabilidad Regional. El "punto de corte" en el que la primera de estas fuentes se abandona como referencia para la construcción de la serie larga 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.

En la mayor parte de los casos, el enlace no es problemático en el sentido de que las discrepancias existentes entre las distintas series en el momento de su enlace son generalmente reducidas, al menos cuando se trabaja con series normalizadas por el promedio nacional por habitante. En los casos en los que esto no es cierto, he intentado apoyarme en evidencia externa a la hora de distribuir la diferencia que aflora en el momento del enlace entre los niveles iniciales de las series de la Fundación BBVA y sus tasas de crecimiento durante el período muestral. Aún así, ha de reconocerse que el procedimiento utilizado deja un cierto margen para decisiones subjetivas que son seguramente discutibles. Como contrapartida, tiene la ventaja de hacer explícita la necesidad de realizar algún tipo de hipótesis sobre la senda temporal del error aflorado, afrontando así de una forma transparente una dificultad que los métodos más habituales de enlace ocultan pero no resuelven.

ANEXO 1: ¿Qué serie de referencia? Detalles técnicos

En este anexo se presentan los detalles técnicos del procedimiento utilizado en el texto para fijar la fecha de corte en la que se abandonan las series del BBVA para pasar a tomar como referencia las series de la CRE-e. Los apartados 1 a 4 recogen algunos resultados y desarrollos preliminares. Este material se utiliza en los apartados 5 y 6 para calcular el error cuadrático medio (ECM) de las distintas series, el nivel crítico del parámetro ϕ que mide el posible sesgo hacia cero de las series del BBVA y la cota superior para el mismo parámetro.

1. Estimación por MCO cuando no hay errores de medición

Como referencia, conviene repasar algunos resultados básicos sobre el modelo estándar de regresión que utilizaré como punto de partida en lo que sigue. Supongamos que y viene dada por

$$(A1) \quad y_i = \beta x_i + u_i$$

donde u_i es una perturbación aleatoria iid con media cero y varianza σ_u^2 . Supongamos también que se cumplen las hipótesis del modelo clásico de regresión y, en particular, que el término de error no está correlacionado con el regresor ($Ex_i u_i = 0$) de forma que la estimación del modelo por mínimos cuadrados ordinarios (MCO) producirá estimadores consistentes del parámetro β . Para simplificar los cálculos he supuesto también (sin pérdida de generalidad) que todas las variables tienen media cero, lo que entre otras cosas implica que la constante de la regresión desaparece.

La ecuación (A1) implica que la varianza de y_i dada x_i es igual a

$$\text{var}(y_i | x_i) = \sigma_u^2 \quad \text{para todo } i$$

donde conviene observar que la varianza “condicional” $\text{var}(y_i | x_i)$ no es lo mismo que la varianza “total” o incondicional de y , $\text{var } y$.

El estimador por MCO de β viene dado por

$$(A2) \quad b = \frac{\sum x_i y_i}{\sum x_i^2} = \sum \left(\frac{x_i}{\sum x_i^2} \right) y_i$$

y, por tanto, su varianza (condicional en x) será igual a

$$(A3) \quad \text{var } b = \sum \left(\frac{x_i}{\sum x_i^2} \right)^2 \text{var}(y_i | x_i) = \sigma_u^2 * \sum \left(\frac{x_i}{\sum x_i^2} \right)^2 = \sigma_u^2 * \frac{\sum x_i^2}{(\sum x_i^2)^2} = \frac{\sigma_u^2}{\sum x_i^2}$$

Por último, el R^2 de la regresión es la fracción de la variación total de y que viene explicada por el modelo. Utilizando (A2), tenemos

$$(A4) R^2(y|x) = \frac{\sum (bx_i)^2}{\sum y_i^2} = \frac{b^2 \sum x_i^2}{\sum y_i^2} = \frac{\frac{(\sum x_i y_i)^2}{(\sum x_i^2)^2} \sum x_i^2}{\sum y_i^2} = \frac{(\sum x_i y_i)^2}{(\sum x_i^2)(\sum y_i^2)}$$

2. Implicaciones del error de medición en la variable dependiente

Consideremos ahora la posibilidad de que la variable dependiente, y , esté medida con error y pueda presentar un sesgo sistemático hacia cero. Supondremos que lo que podemos observar no es y , sino una o varias *proxies* ruidosas de la forma

$$(A5) y_i^k = (1 - \phi_k)y_i + e_i^k$$

donde $\phi_k \in [0,1]$ captura el posible sesgo sistemático en la medición de y y e_i^k es un error "clásico" de medición iid y con

$$(A6) Ee_i^k = 0, E(e_i^k)^2 = \sigma_k^2 \text{ para todo } i \text{ y } Ee_i^k x_i = Ee_i^k y_i = 0$$

Observemos que

$$(A7) \text{var}(y_i^k | x_i) = (1 - \phi_k)^2 \text{var}(y_i | x_i) + \text{var } e_i^k = (1 - \phi_k)^2 \sigma_u^2 + \sigma_k^2$$

Supongamos ahora que estimamos la regresión dada en (A1) utilizando y^k en vez de y como variable dependiente. El estimador de MCO vendrá dado por

$$(A8) b_k = \frac{\sum x_i y_i^k}{\sum x_i^2} = \sum \left(\frac{x_i}{\sum x_i^2} \right) y_i^k$$

y su varianza será

$$(A9) \text{var } b_k = \frac{\text{var}(y_i^k | x_i)}{\sum x_i^2} = \frac{(1 - \phi_k)^2 \sigma_u^2 + \sigma_k^2}{\sum x_i^2}$$

Es fácil comprobar que el sesgo sistemático en la medición de y se traslada a b . Usando (A8) y (A5) tenemos

$$(A10) b_k = \frac{\sum x_i y_i^k}{\sum x_i^2} = \frac{\sum x_i [(1 - \phi_k)y_i + e_i^k]}{\sum x_i^2} = \frac{(1 - \phi_k) \sum x_i y_i}{\sum x_i^2} + \frac{\sum x_i e_i^k}{\sum x_i^2} = (1 - \phi_k)b + \frac{\sum x_i e_i^k}{\sum x_i^2}$$

donde b es el estimador de MCO que se obtendría si y fuese directamente observable. Puesto que b es consistente y por hipótesis $Ee_i^k x_i = 0$, tenemos

$$(A11) Eb_k = (1 - \phi_k)\beta$$

Examinando (A9) y (A11), vemos que el componente clásico del error de medición en y no sesga el estimador pero aumenta su varianza. El componente sistemático, por el contrario, genera un sesgo a la baja en el estimador pero también reduce su varianza.

Seguidamente, queremos calcular el error cuadrático medio (ECM) del estimador MCO de β .

Definiendo el sesgo asintótico de b_k como

$$(A12) \quad s_k = Eb_k - \beta = (1 - \phi_k)\beta - \beta = -\phi_k\beta$$

tenemos

$$(A13) \quad \begin{aligned} E(b_k - \beta)^2 &= E((b_k - Eb_k) + (Eb_k - \beta))^2 = E((b_k - Eb_k) + s_k)^2 \\ &= E((b_k - Eb_k)^2 + 2(b_k - Eb_k)s_k + s_k^2) = \\ &= E(b_k - Eb_k)^2 + 2s_k E(b_k - Eb_k) + s_k^2 = E(b_k - Eb_k)^2 + 0 + s_k^2 \\ &= \text{var } b_k + \phi_k^2 \beta^2 \end{aligned}$$

Substituyendo (A9) en esta expresión y dividiendo por β^2 tenemos

$$(A14) \quad E_k \equiv \frac{E(b_k - \beta)^2}{\beta^2} = \phi_k^2 + \frac{\text{var } b_k}{\beta^2} = \phi_k^2 + \frac{(1 - \phi_k)^2 \sigma_u^2 + \sigma_k^2}{\beta^2 \sum x_i^2}$$

Nuestro objetivo es estimar el valor de E_k para las dos series alternativas con las que estamos trabajando (la serie del BBVA y la de la CRE-e) con el fin de determinar cuál de ellas es “mejor” en el sentido de generar estimadores con menor ECM en cada uno de los años en los que ambas se solapan. Más adelante volveré a esta expresión e intentaré escribirla en términos de magnitudes observables o estimables (que a poder ser no dependan de las x 's, dado que x puede ser en principio cualquier variable). Antes, sin embargo, necesitamos introducir algunos conceptos y realizar algunos cálculos preliminares.

3. Ratios de fiabilidad

Supongamos que disponemos de dos *proxies* alternativas para y , y^j e y^k . Regresando cada una de estas variables sobre la otra podemos obtener un estimador del ratio entre la varianza de la variable subyacente, y , y la varianzas de cada una de las *proxies*. Este cociente se suele denominar *ratio de fiabilidad* y, bajo ciertas condiciones, constituye un buen indicador del contenido informativo de las distintas series (esto es, un indicador inverso de la importancia del “ruido” debido a la existencia de errores de medición).¹⁴

Para concretar, identificaremos y^k con la serie del BBVA e y^j con la serie enlazada de la CRE. Esto es, supondremos que la segunda serie no presenta un sesgo sistemático ($\phi_j = 0$) pero a cambio podría contener más ruido puro ($\sigma_j^2 > \sigma_k^2$). Tenemos, por tanto,

$$(A15) \quad y_i^k = (1 - \phi_k)y_i + e_i^k$$

$$(A16) \quad y_i^j = y_i + e_i^j$$

¹⁴ Véase por ejemplo Krueger y Lindahl (2001) y de la Fuente y Doménech (2006).

donde supondremos que los errores e_i^k y e_i^j , además de no estar correlacionados con x e y , tampoco lo están entre sí ($Ee_i^j e_i^k = 0$). Utilizando estas hipótesis, tenemos

$$(A17) E y_i^k y_i^j = E[(1 - \phi_k) y_i + e_i^k][y_i + e_i^j] = (1 - \phi_k) E y_i^2$$

Consideremos ahora una regresión de la forma $y_i^j = r_k y_i^k + \varepsilon_{ki}$. El valor esperado del estimador del coeficiente de pendiente (el ratio de fiabilidad de y_i^k) viene dado por

$$\hat{r}_k = \frac{\sum y_i^k y_i^j}{\sum (y_i^k)^2} = (1 - \phi_k) \frac{\sum y_i^2}{\sum (y_i^k)^2} + (1 - \phi_k) \frac{\sum y_i e_i^j}{\sum (y_i^k)^2} + (1 - \phi_k) \frac{\sum y_i e_i^k}{\sum (y_i^k)^2} + \frac{\sum e_i^k e_i^j}{\sum (y_i^k)^2}$$

Esta expresión es difícil de simplificar porque contiene cocientes de variables aleatorias (obsérvese que a diferencia de x , no podemos tratar a y_i^k como una constante). Sin embargo, sí podemos calcular el límite en probabilidad del estimador, que viene dado por

$$(A18) \tilde{E} \hat{r}_k \equiv \text{plim } \hat{r}_k = \frac{E y_i^k y_i^j}{E (y_i^k)^2} = \frac{(1 - \phi_k) E y_i^2}{E (y_i^k)^2}$$

Utilizando esta expresión, tenemos

$$(A19) \frac{E y_i^2}{E (y_i^k)^2} = \frac{\tilde{E} \hat{r}_k}{(1 - \phi_k)}$$

Intercambiando los papeles de las dos series, tenemos también

$$(A20) \tilde{E} \hat{r}_j = \frac{E y_i^k y_i^j}{E (y_i^j)^2} = \frac{(1 - \phi_k) E y_i^2}{E (y_i^j)^2} \Rightarrow \frac{E y_i^2}{E (y_i^j)^2} = \frac{\tilde{E} \hat{r}_j}{(1 - \phi_k)}$$

Más adelante, utilizaremos estas expresiones para escribir el ECM de cada serie en términos de magnitudes “observables.”

4. Otros preliminares

La expresión que hemos derivado para el ECM del estimador de β en presencia de errores de medición (ecuación (A14)) incluye las varianzas de la perturbación de la regresión original y del componente clásico del error de medición (σ_u^2 y σ_k^2). Queremos reemplazar estos dos términos por otras magnitudes más fáciles de estimar o aproximar sin usar las x 's.

Utilizando

$$(A1) y_i = \beta x_i + u_i$$

podemos calcular la varianza “total” (no condicional) de y :

$$(A21) \text{var } y_i = \beta^2 \text{var } x_i + \sigma_u^2$$

y escribir σ_u^2 en la forma

$$(A22) \quad \sigma_u^2 = \text{var } y_i - \beta^2 \text{var } x_i$$

De la misma forma, dada

$$(A15) \quad y_i^k = (1 - \phi_k)y_i + e_i^k$$

tenemos

$$(A23) \quad \text{var } y_i^k = (1 - \phi_k)^2 \text{var } y_i + \sigma_k^2 \Rightarrow \sigma_k^2 = \text{var } y_i^k - (1 - \phi_k)^2 \text{var } y_i$$

Substituyendo (A22) y (A23) en (A14), tenemos:

$$(A24) \quad \begin{aligned} E_k &= \phi_k^2 + \frac{(1 - \phi_k)^2 \sigma_u^2 + \sigma_k^2}{\beta^2 \sum x_i^2} = \phi_k^2 + \frac{(1 - \phi_k)^2 [\text{var } y_i - \beta^2 \text{var } x_i] + \text{var } y_i^k - (1 - \phi_k)^2 \text{var } y_i}{\beta^2 \sum x_i^2} \\ &= \phi_k^2 + \frac{-(1 - \phi_k)^2 \beta^2 \text{var } x_i + \text{var } y_i^k}{\beta^2 \sum x_i^2} = \phi_k^2 - \frac{(1 - \phi_k)^2 \text{var } x_i}{\sum x_i^2} + \frac{\text{var } y_i^k}{\beta^2 \sum x_i^2} \\ &= \phi_k^2 - \frac{(1 - \phi_k)^2 \text{var } x_i}{\sum x_i^2} + \frac{\text{var } y_i^k}{\text{var } y_i} \frac{\text{var } y_i}{\beta^2 \sum x_i^2} \end{aligned}$$

5. Estimación del ECM

Seguidamente, construimos un estimador de E_k reemplazando los momentos poblacionales que aparecen en (A24) por sus homólogos muestrales y los parámetros (excepto ϕ_k que quedará como un parámetro "libre") por estimadores insesgados o al menos consistentes. En particular, reemplazamos

$$\text{var } x_i \quad \text{por} \quad \frac{1}{n-1} \sum x_i^2 \quad \text{donde } n \text{ es el número de observaciones}$$

$$\frac{\text{var } y_i^k}{\text{var } y_i} \quad \text{por} \quad \frac{1 - \phi_k}{\hat{r}_k} \quad \text{véase la ecuación (A19)}$$

$$\beta^2 \quad \text{por} \quad b^2 = \frac{(\sum x_i y_i)^2}{(\sum x_i^2)^2} \quad \text{véase la ecuación (A2) y}$$

$$\text{var } y_i \quad \text{por} \quad \frac{1}{n-1} \sum y_i^2 \quad \text{y}$$

$$\tilde{E}\hat{r}_k \quad \text{por} \quad \hat{r}_k$$

para obtener un estimador consistente de E_k

$$\begin{aligned}
\hat{E}_k &= \phi_k^2 - \frac{(1-\phi_k)^2}{n-1} \frac{\sum x_i^2}{\sum x_i^2} + \frac{1-\phi_k}{\hat{r}_k} \frac{\frac{1}{n-1} \sum y_i^2}{\frac{(\sum x_i y_i)^2}{(\sum x_i^2)^2} \sum x_i^2} \\
(A25) \quad &= \phi_k^2 - \frac{(1-\phi_k)^2}{n-1} + \frac{1-\phi_k}{(n-1)\hat{r}_k} \frac{(\sum y_i^2)(\sum x_i^2)}{(\sum x_i y_i)^2} = \phi_k^2 - \frac{(1-\phi_k)^2}{n-1} + \frac{1-\phi_k}{(n-1)\hat{r}_k R^2(x)} \\
&= \phi_k^2 - \frac{1}{n-1} \left((1-\phi_k)^2 - \frac{1-\phi_k}{\hat{r}_k R^2(x)} \right)
\end{aligned}$$

donde hemos hecho uso de la ecuación (A4) para escribir el último término de la expresión en función del R^2 de una regresión de y sobre x , $R^2(x)$. Obsérvese que el último término tiende a cero cuando $n \rightarrow \infty$. Esto es, el ECM se reduce al cuadrado del sesgo cuando el tamaño de la muestra tiende a infinito.

Procediendo de la misma forma en el caso de y^j , recordando que por hipótesis $\phi_j = 0$ y utilizando la ecuación (A20), tenemos:

$$(A26) \quad \hat{E}_j = -\frac{\frac{1}{n-1} \sum x_i^2}{\sum x_i^2} + \frac{1-\phi_k}{\hat{r}_j} \frac{\frac{1}{n-1} \sum y_i^2}{\frac{(\sum x_i y_i)^2}{(\sum x_i^2)^2} \sum x_i^2} = -\frac{1}{n-1} + \frac{1-\phi_k}{(n-1)\hat{r}_j R^2(x)} = \frac{1}{n-1} \left(\frac{1-\phi_k}{\hat{r}_j R^2(x)} - 1 \right)$$

Definamos la función

$$(A27) \quad F(\phi_k, R^2(x)) = (n-1)(\hat{E}_k - \hat{E}_j)$$

Diremos que la serie y^k es "mejor" que la serie y^j si tiene un ECM menor, esto es, si $F(\phi_k, R^2(x)) < 0$. Puesto que $F()$ es una función de x , por el momento la afirmación de que una serie es mejor que otra dependerá de la variable x que estemos considerando y, por supuesto, del valor del parámetro ϕ_k , sobre el que no sabemos nada.

Combinando (A25) y (A26) tenemos

$$\begin{aligned}
(A28) \quad F(\phi_k, R^2(x)) &= (n-1)(\hat{E}_k - \hat{E}_j) = (n-1)\phi_k^2 - (1-\phi_k)^2 + \frac{1-\phi_k}{\hat{r}_k R^2(x)} - \frac{1-\phi_k}{\hat{r}_j R^2(x)} + 1 \\
&= (n-1)\phi_k^2 - (1-2\phi_k + \phi_k^2) + \frac{1-\phi_k}{R^2(x)} \left(\frac{1}{\hat{r}_k} - \frac{1}{\hat{r}_j} \right) + 1 \\
&= (n-1)\phi_k^2 + (2\phi_k - \phi_k^2) + \frac{1-\phi_k}{R^2(x)} \left(\frac{\hat{r}_j - \hat{r}_k}{\hat{r}_k \hat{r}_j} \right) \\
&= (n-2)\phi_k^2 + 2\phi_k - \frac{1-\phi_k}{R^2(x)} \left(\frac{\hat{r}_k - \hat{r}_j}{\hat{r}_k \hat{r}_j} \right) \\
&= (n-2)\phi_k^2 + \left(2 + \frac{1}{R^2(x)} \left(\frac{\hat{r}_k - \hat{r}_j}{\hat{r}_k \hat{r}_j} \right) \right) \phi_k - \frac{1}{R^2(x)} \left(\frac{\hat{r}_k - \hat{r}_j}{\hat{r}_k \hat{r}_j} \right) \\
&\equiv (n-2)\phi_k^2 + \left(2 + \frac{A}{R^2(x)} \right) \phi_k - \frac{A}{R^2(x)}
\end{aligned}$$

que es la ecuación (4) en el texto.

6. Una cota para ϕ

El ratio de fiabilidad de y_i^j puede utilizarse para estimar una cota superior para ϕ_k . Dado que

$$(A16) \quad y_i^j = y_i + e_i^j$$

tenemos

$$E(y_i^j)^2 = Ey_i^2 + E(e_i^j)^2 = Ey_i^2 + \sigma_j^2$$

Dividiendo por Ey_i^2

$$\frac{E(y_i^j)^2}{Ey_i^2} = 1 + \frac{\sigma_j^2}{Ey_i^2}$$

Utilizando

$$(A20) \quad \tilde{E}\hat{r}_j = \frac{Ey_i^k y_i^j}{E(y_i^j)^2} = \frac{(1 - \phi_k)Ey_i^2}{E(y_i^j)^2} \Rightarrow \frac{Ey_i^2}{E(y_i^j)^2} = \frac{\tilde{E}\hat{r}_j}{(1 - \phi_k)}$$

en la expresión anterior

$$\frac{(1 - \phi_k)}{\tilde{E}\hat{r}_j} = 1 + \frac{\sigma_j^2}{Ey_i^2} \Rightarrow (1 - \phi_k) = \tilde{E}\hat{r}_j \left(1 + \frac{\sigma_j^2}{Ey_i^2} \right) \geq \tilde{E}\hat{r}_j$$

de donde

$$\phi_k \leq 1 - \tilde{E}\hat{r}_j$$

Puesto que \hat{r}_j es (por definición) un estimador consistente de $\tilde{E}\hat{r}_j$, podemos utilizarlo para construir una cota tentativa para ϕ_k . Si además suponemos que este último parámetro es constante durante el período que estamos analizando (lo que parece razonable, dado que la metodología del BBVA ya estaba bien asentada en 1981 y más aún en el momento en que se revisan las series para construir la serie homogénea 1955-95), tendremos

$$\phi_k \leq 1 - \tilde{E}\hat{r}_{jt} \quad \text{para todo } t$$

y por lo tanto

$$\phi_k \leq 1 - \max_t \tilde{E}\hat{r}_{jt} \equiv \phi_{\max}$$

lo que nos permite estimar ϕ_{\max} como

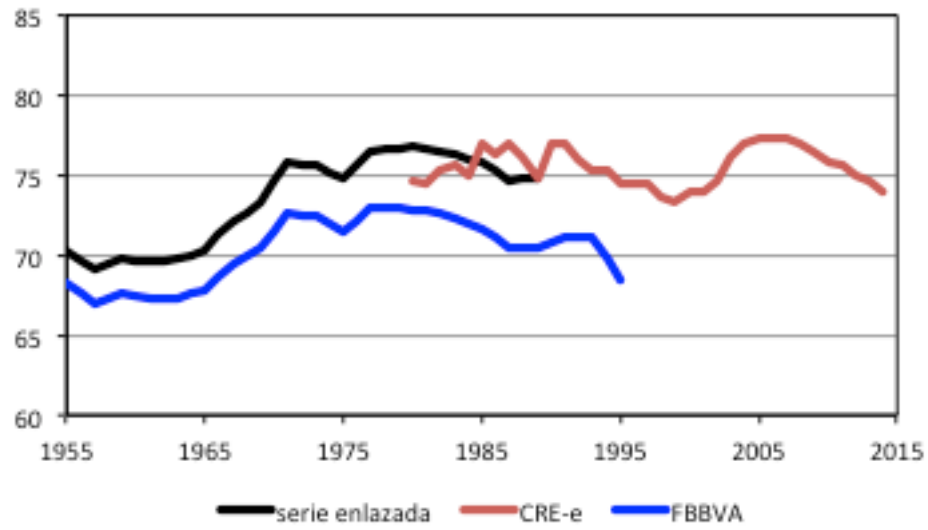
$$(A29) \quad \hat{\phi}_{\max} = 1 - \max_t \hat{r}_{jt}$$

ANEXO 2: Gráficos regionales

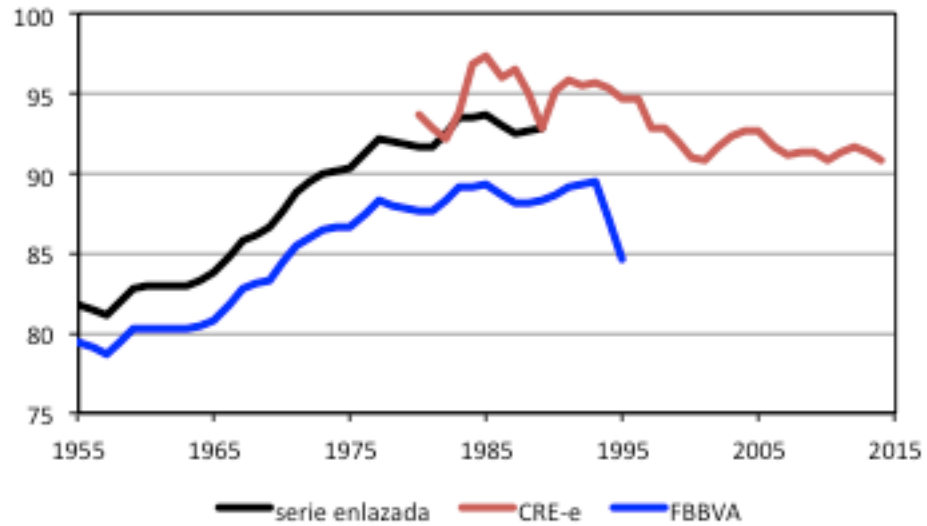
En este Anexo se recogen una serie de gráficos que muestran los perfiles de las series originales y enlazadas de las principales variables de interés. Para cada comunidad autónoma se muestra la evolución del VAB a precios corrientes por habitante y la de sus dos grandes componentes: el VAB por puesto de trabajo (*productividad*) y el número de puestos de trabajo por habitante (*empleo*). Las tres variables se normalizan por el promedio nacional (excluyendo a la *extra regio*) para convertirlas en índices con media = 100.

Los gráficos ilustran que, como ya se ha comentado en el texto, las series de la CRE muestran una volatilidad sospechosamente elevada durante los años ochenta en algunas regiones. Véanse por ejemplo los casos de Aragón, Murcia y la Rioja.

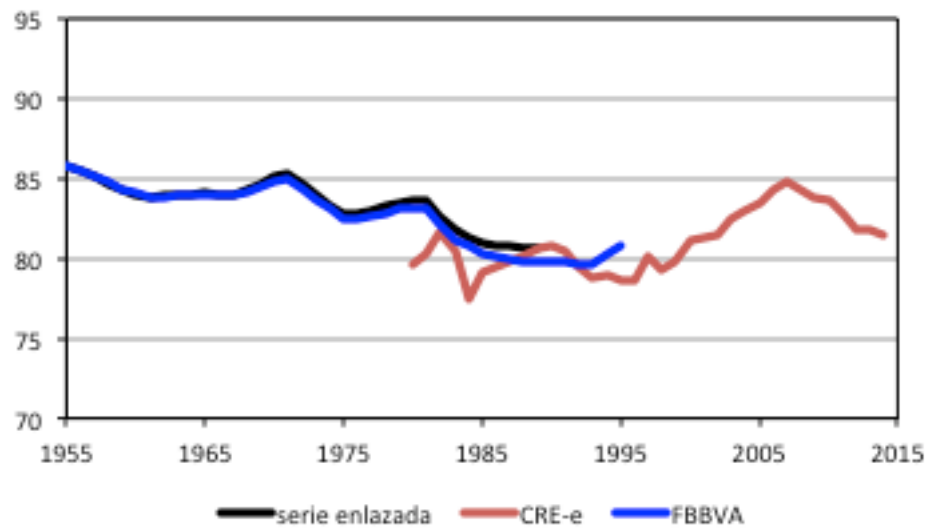
A2.1. Andalucía
 VAB por habitante relativo (a precios corrientes)



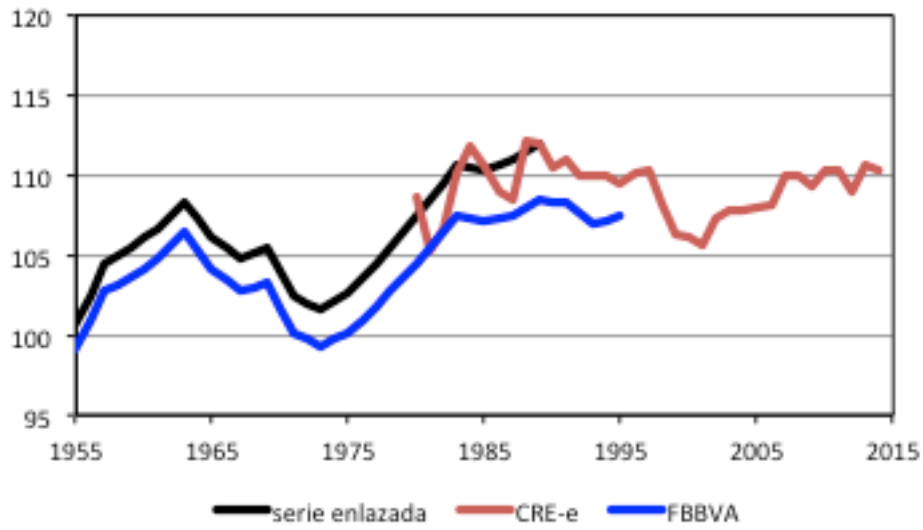
Productividad relativa (VAB por PT)



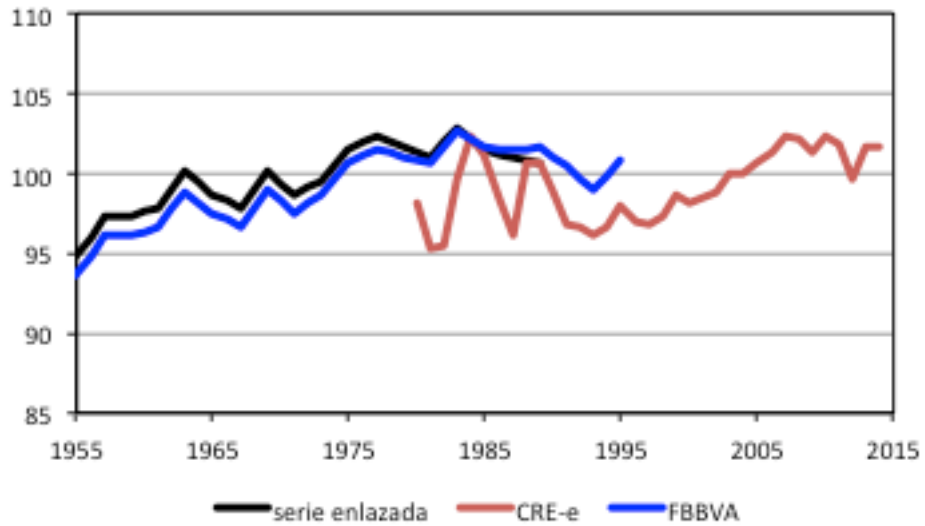
Empleo relativo



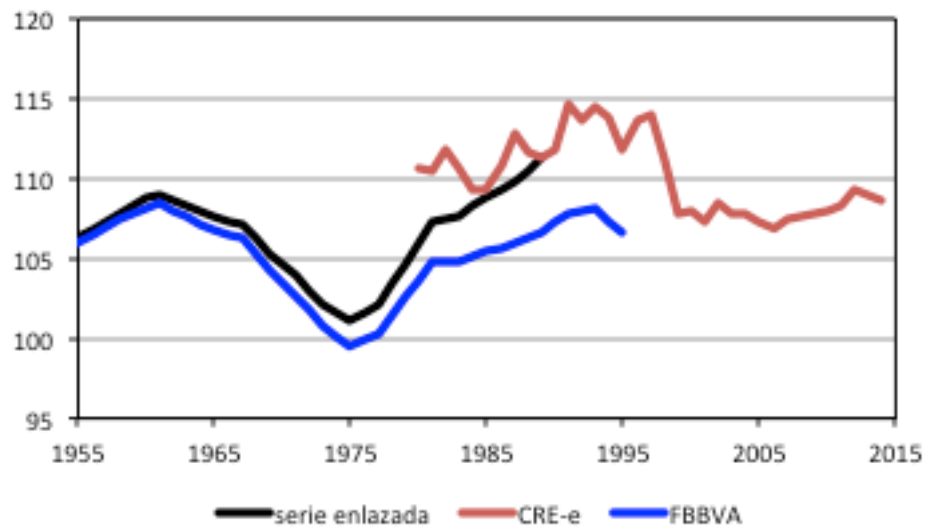
A2.2. Aragón
VAB per capita relativo



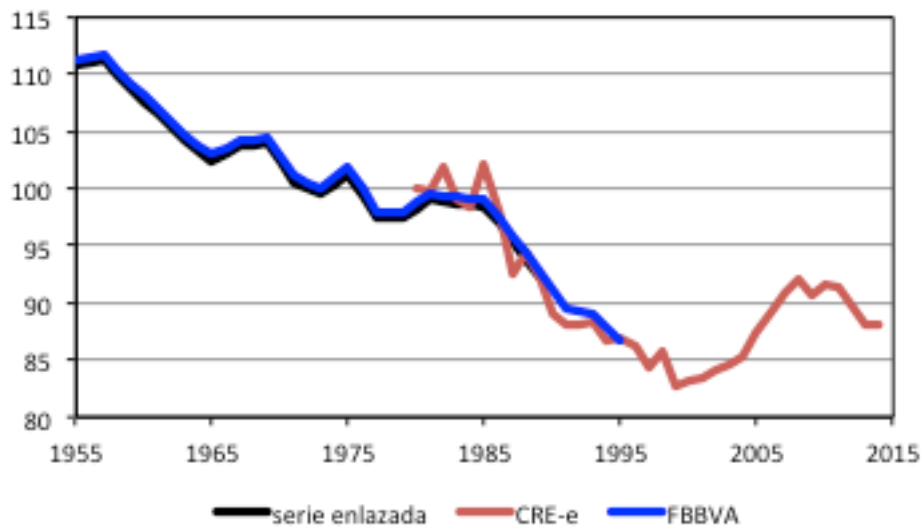
Productividad relativa



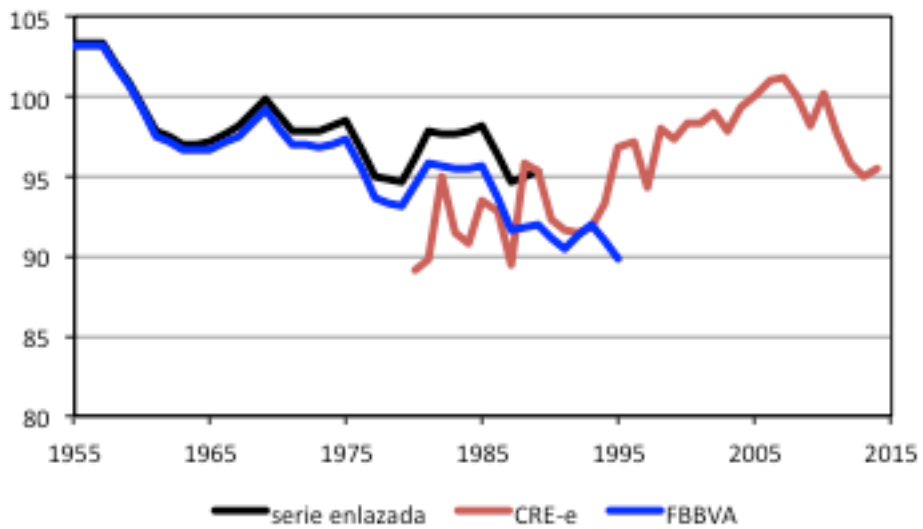
Empleo relativo



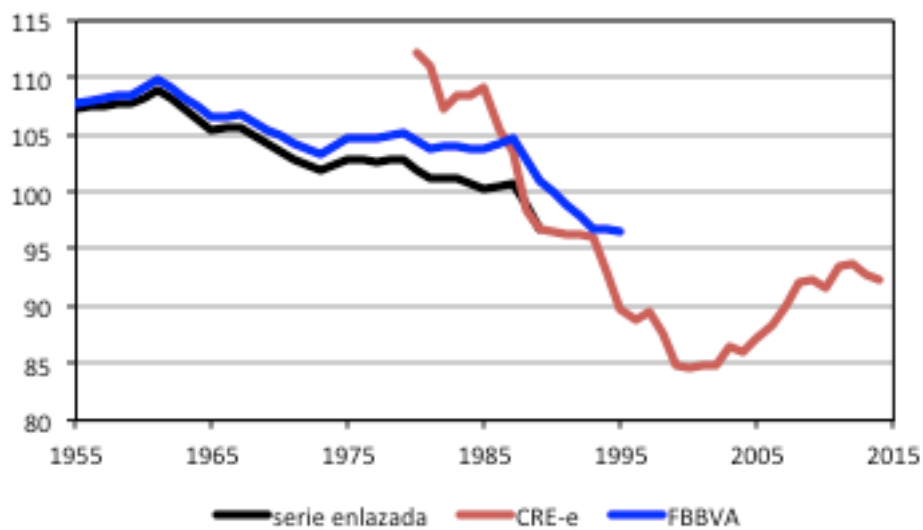
A2.3. Asturias
VAB per capita relativo



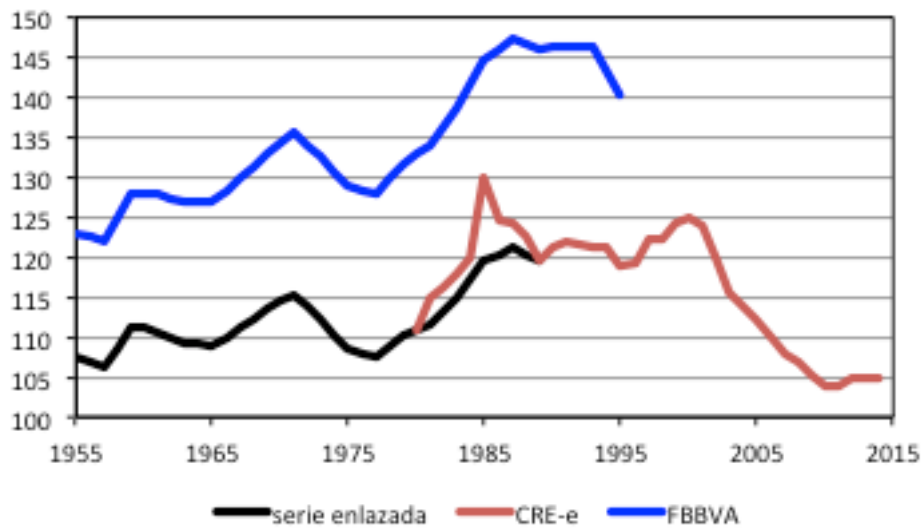
Productividad relativa



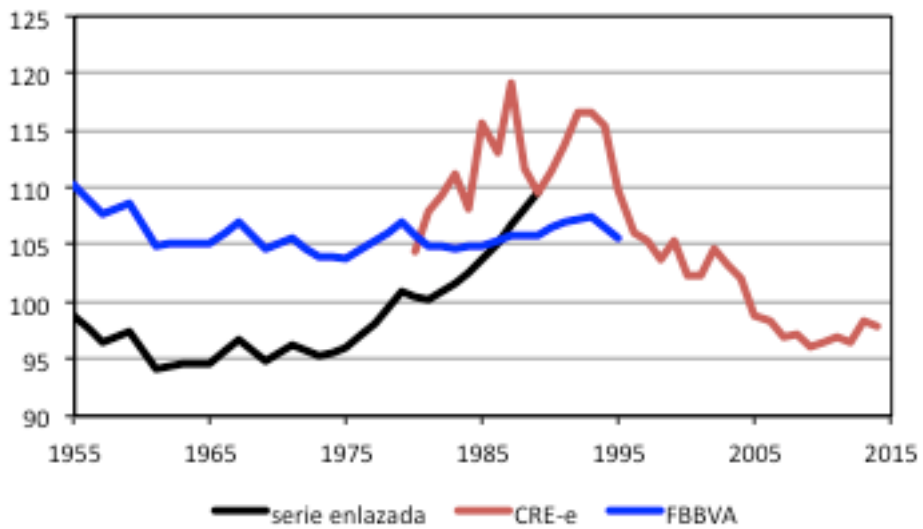
Empleo relativo



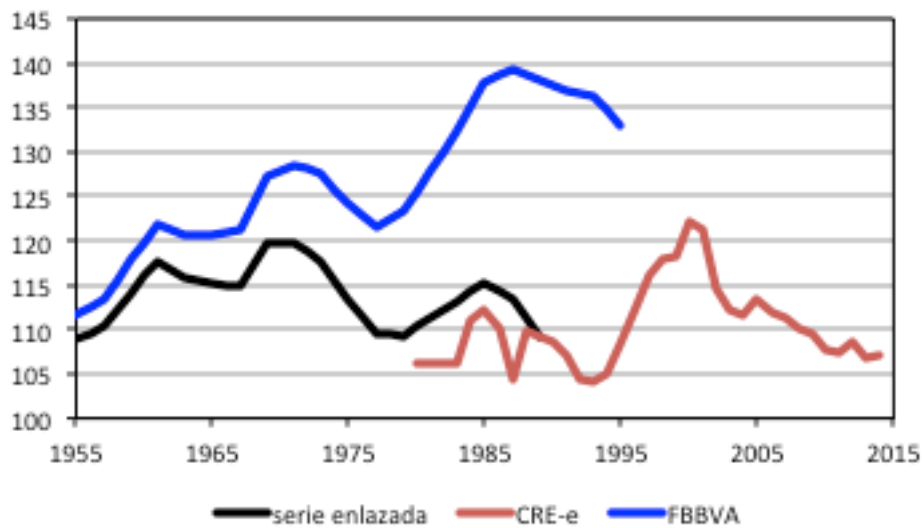
A2.4. Baleares
VAB per capita relativo



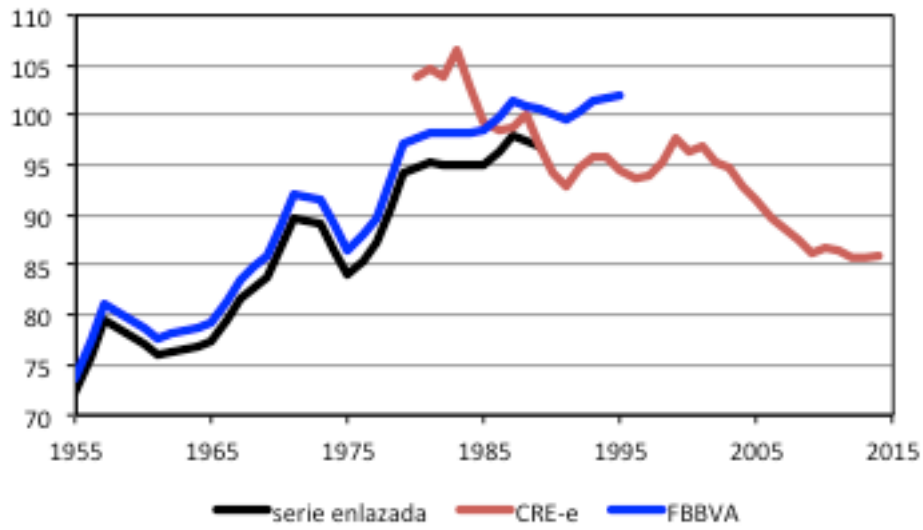
Productividad relativa



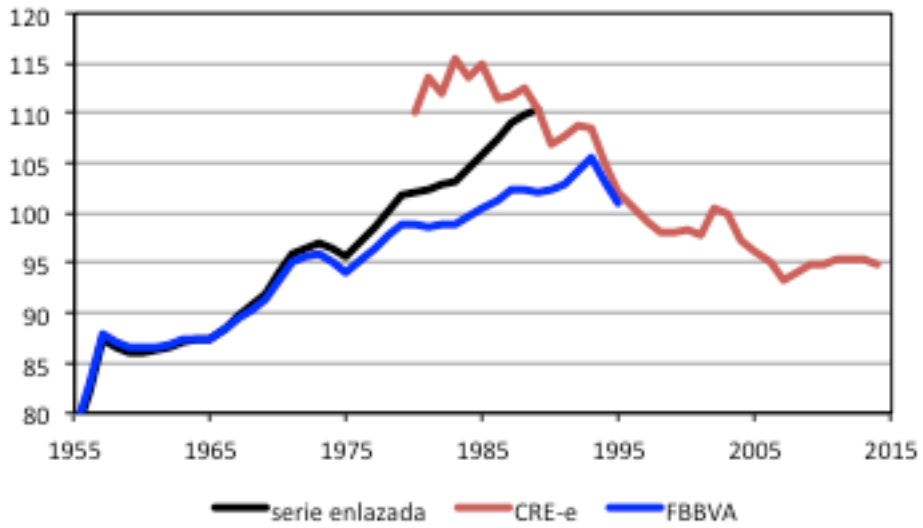
Empleo relativo



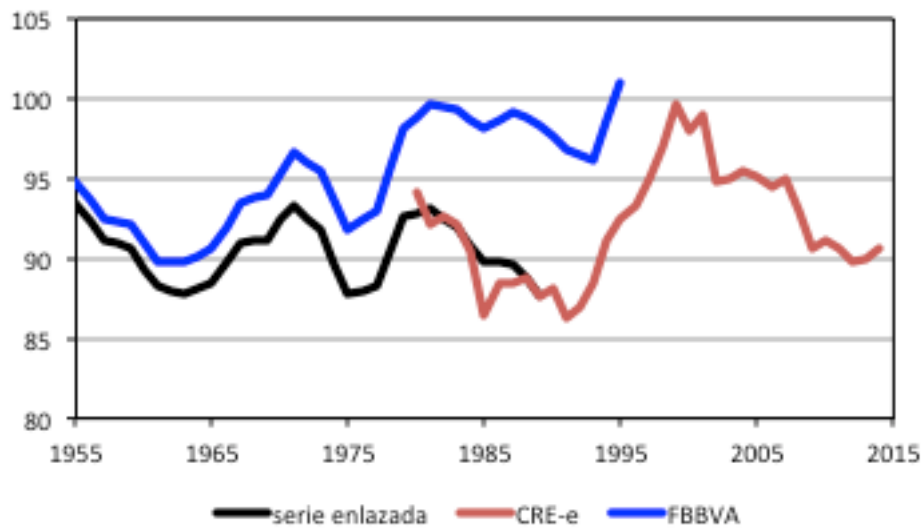
A2.5. Canarias
VAB per capita relativo



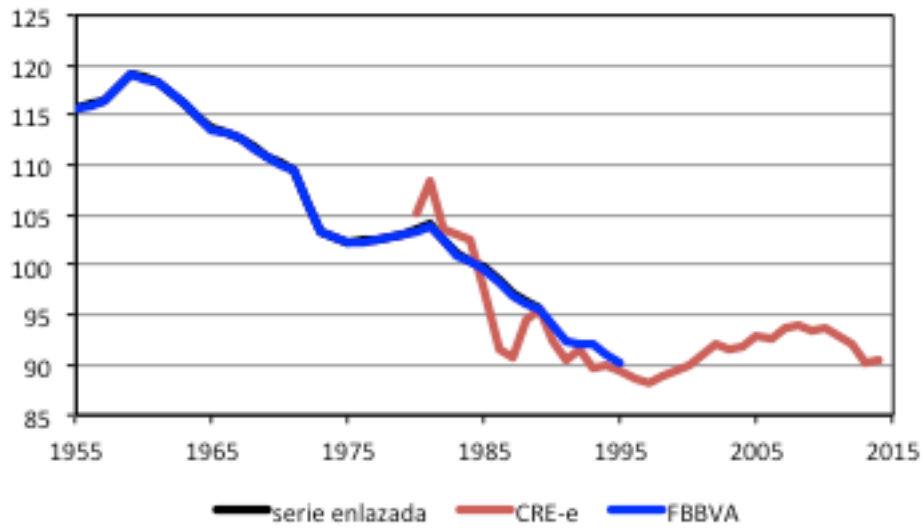
Productividad relativa



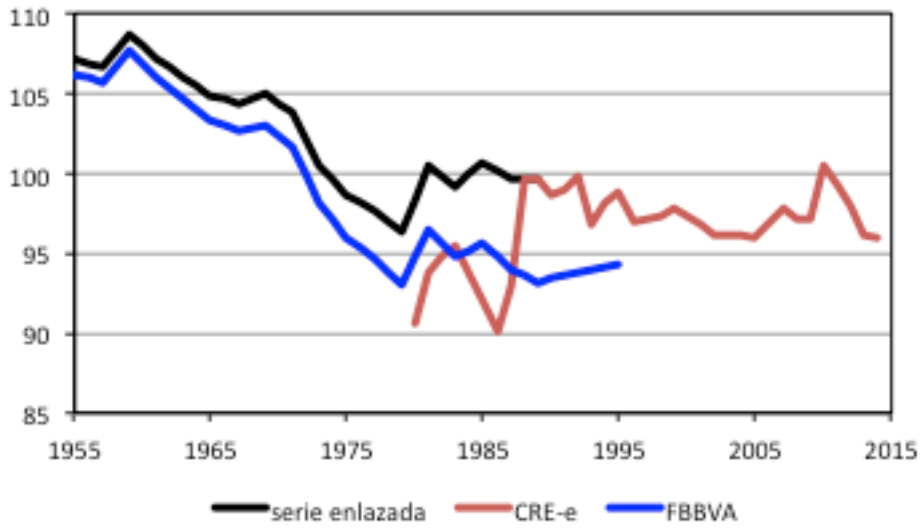
Empleo relativo



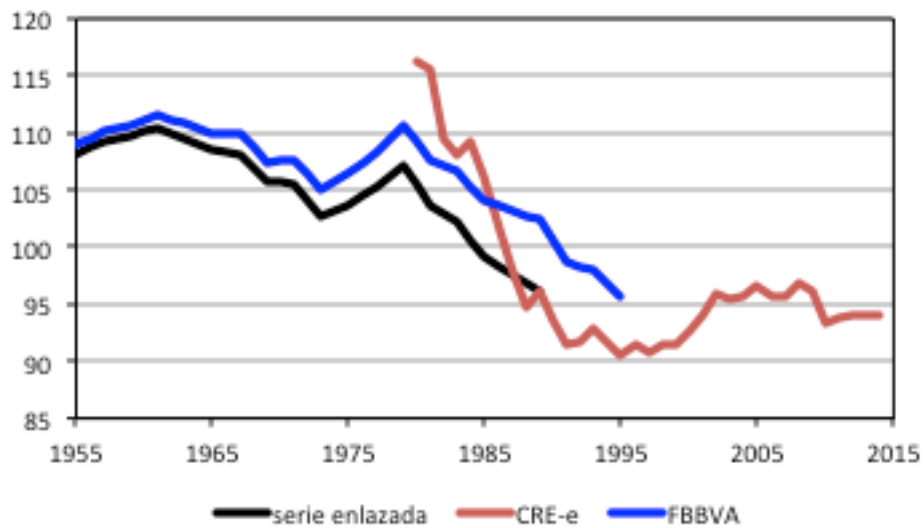
A2.6. Cantabria
VAB per capita relativo



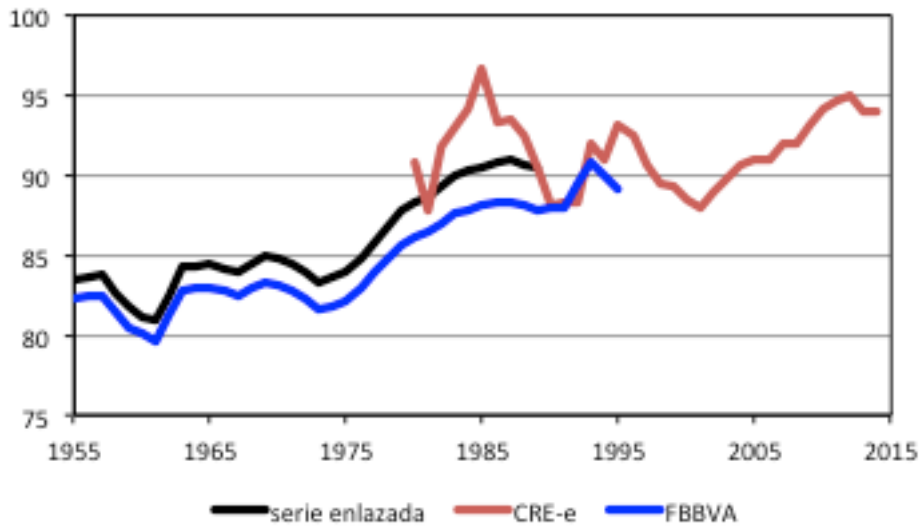
Productividad relativa



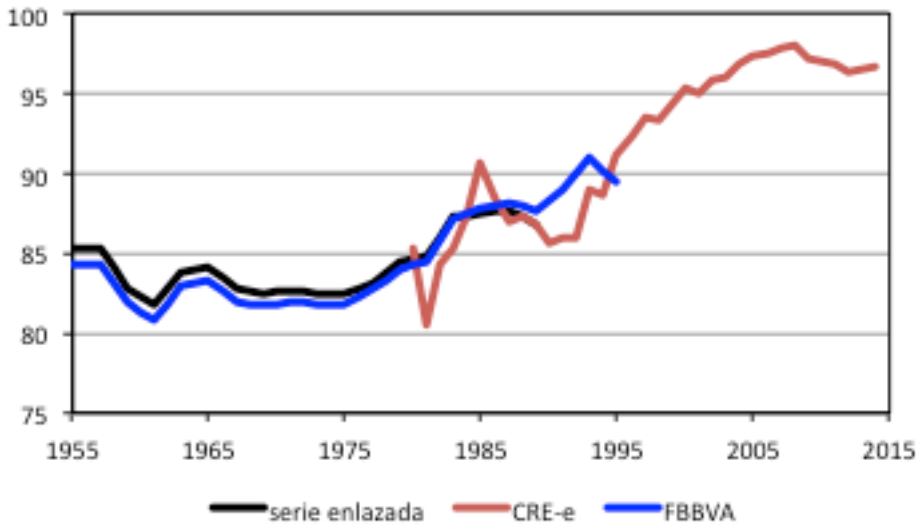
Empleo relativo



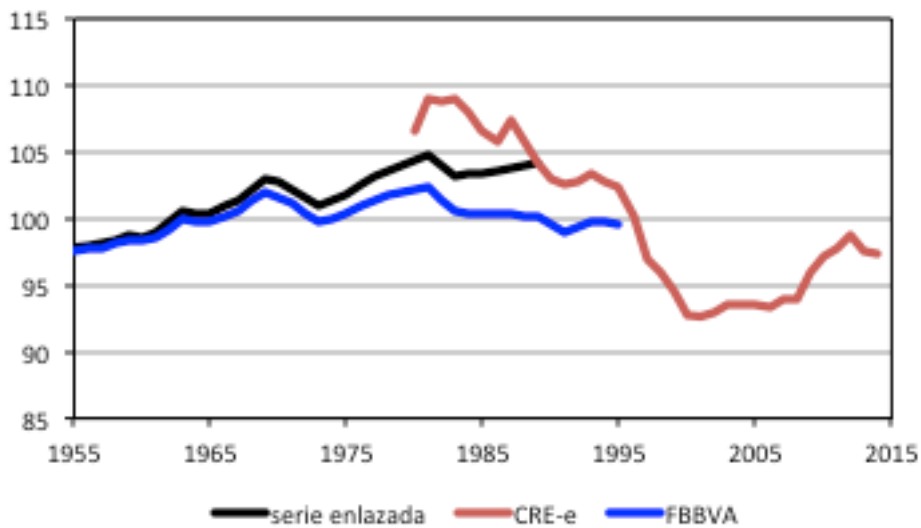
A2.7. Castilla y León
VAB per capita relativo



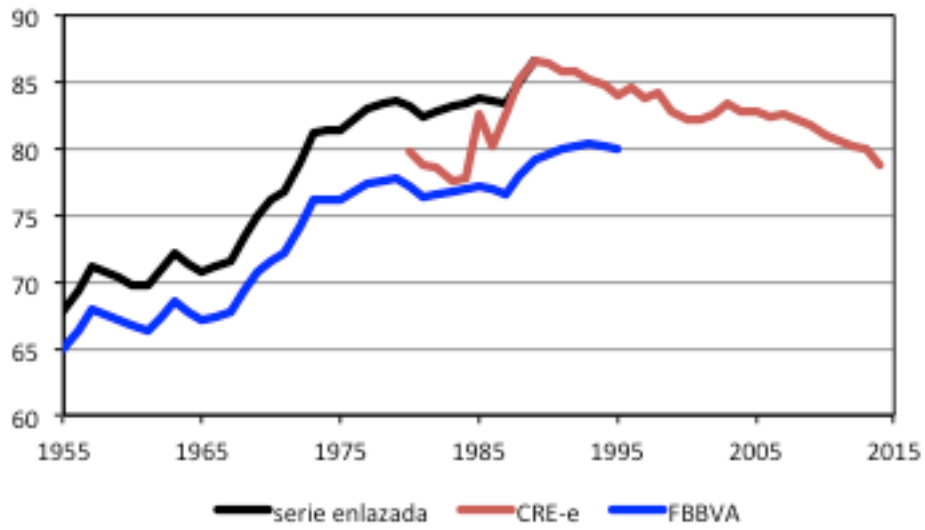
Productividad relativa



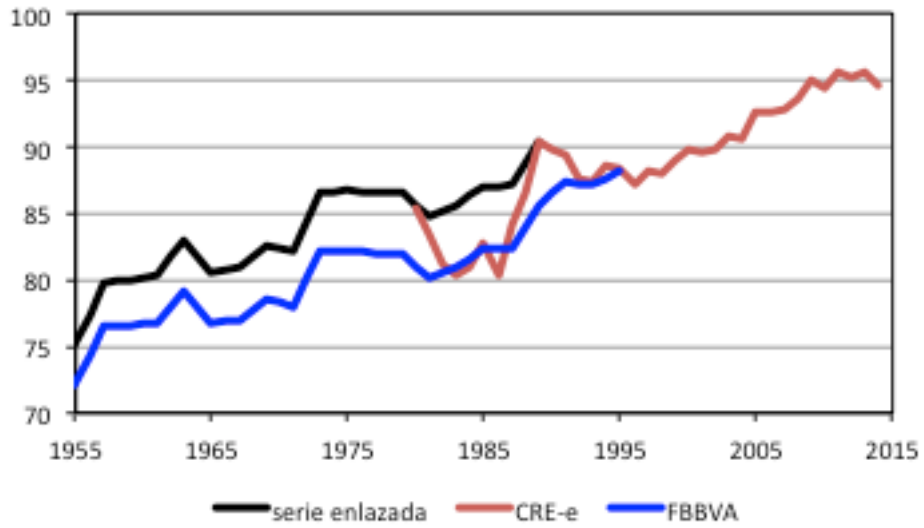
Empleo relativo



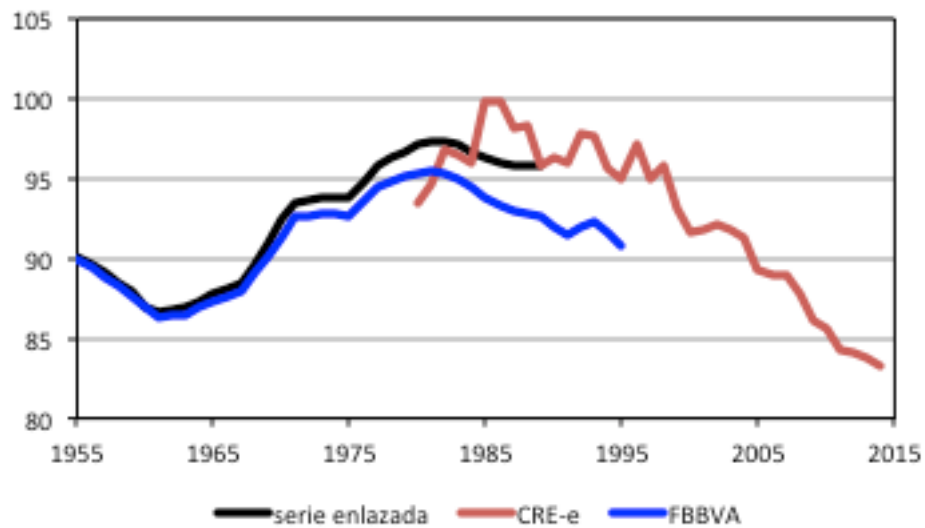
A2.8. Castilla la Mancha
VAB per capita relativo



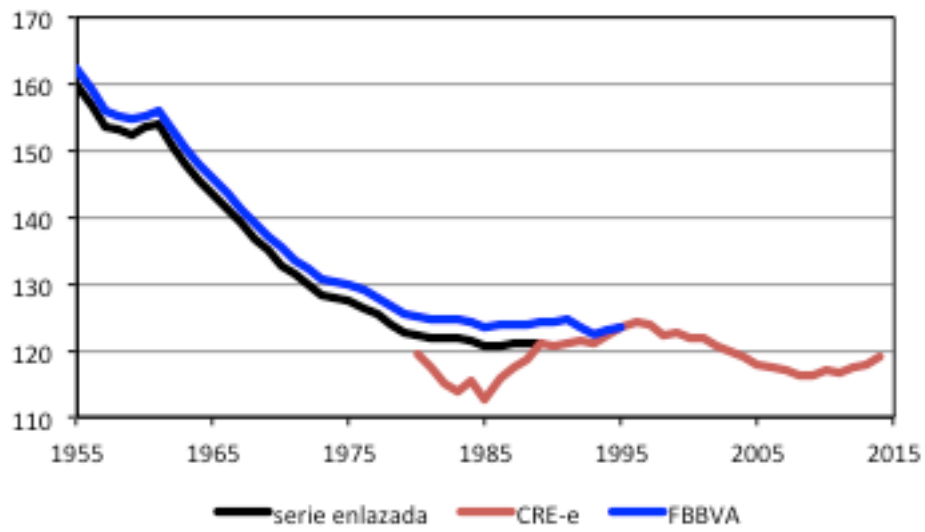
Productividad relativa



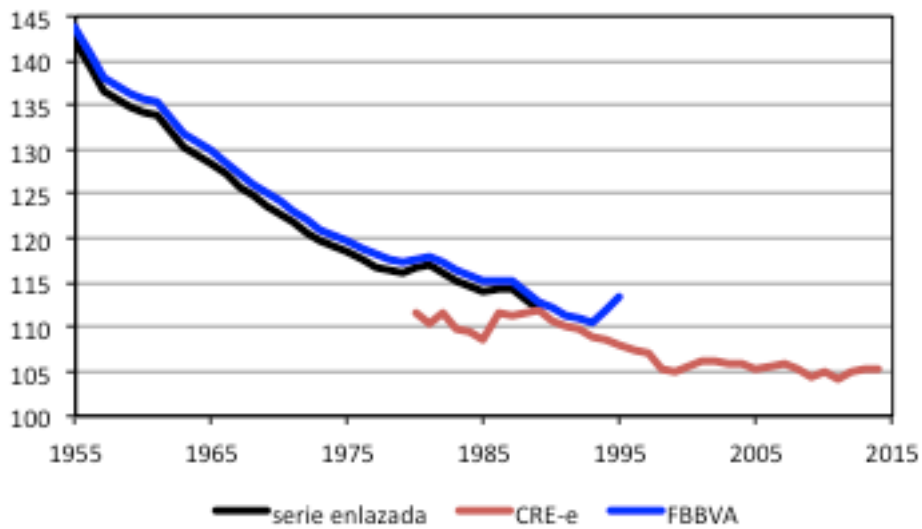
Empleo relativo



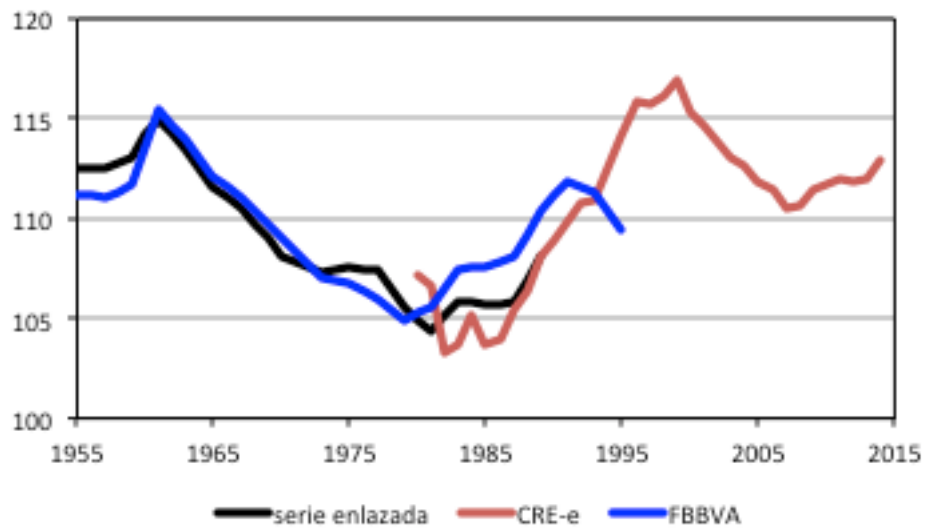
A2.9. Cataluña
VAB per capita relativo



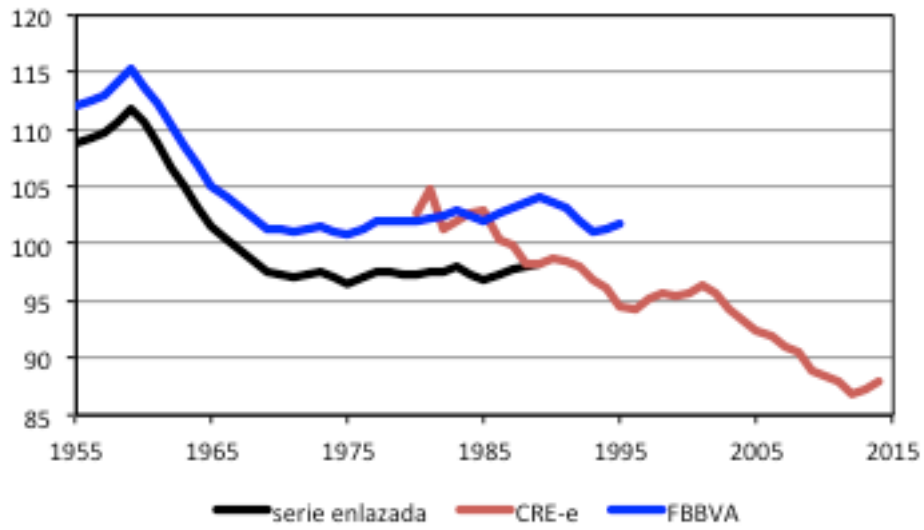
Productividad relativa



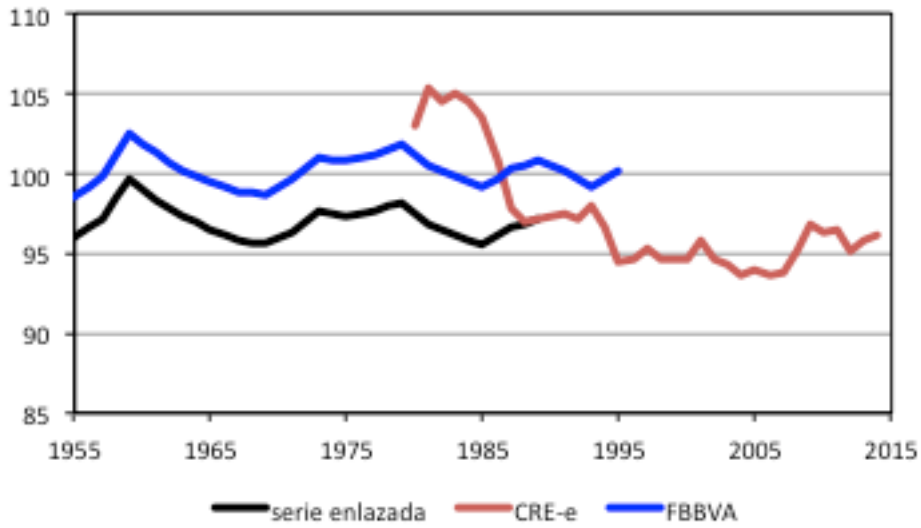
Empleo relativo



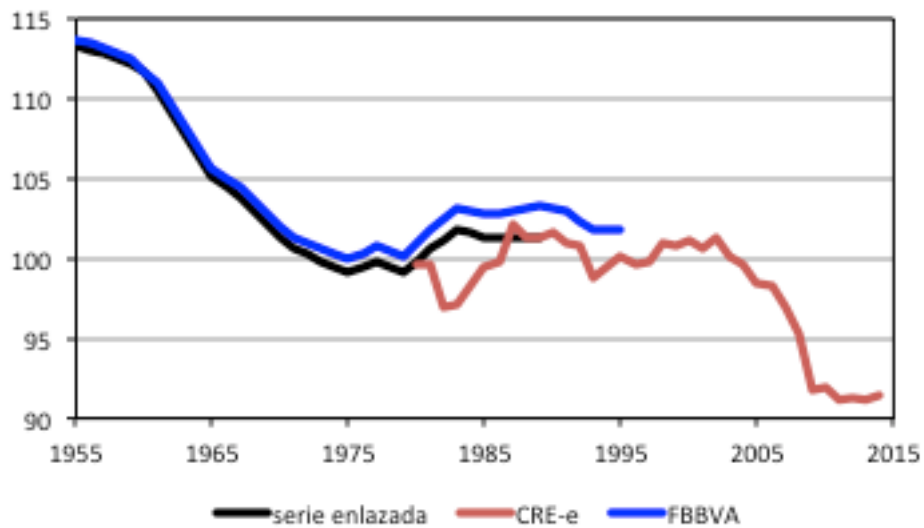
A2.10. Valencia
VAB per capita relativo



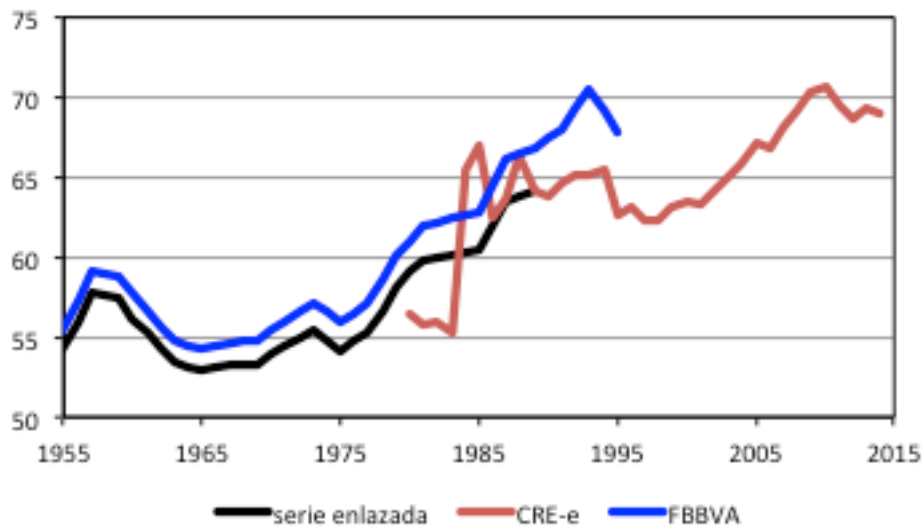
Productividad relativa



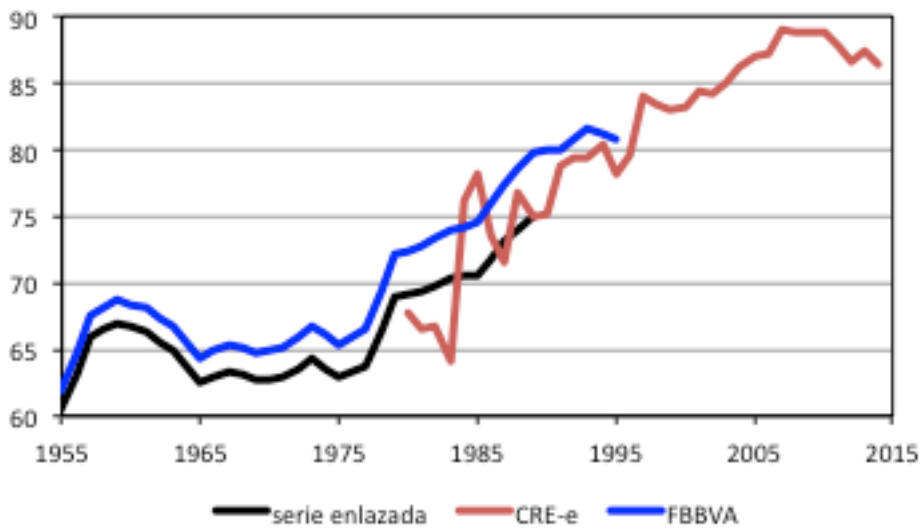
Empleo relativo



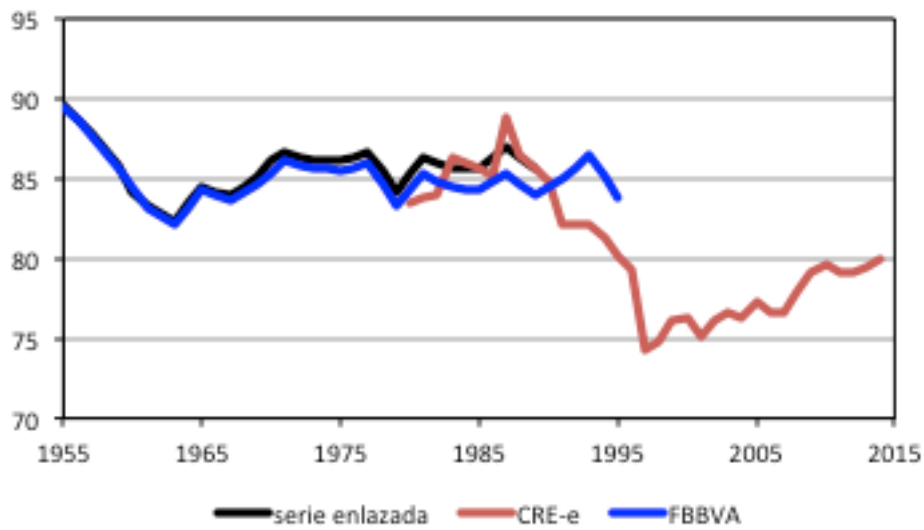
A2.11. Extremadura
VAB per capita relativo



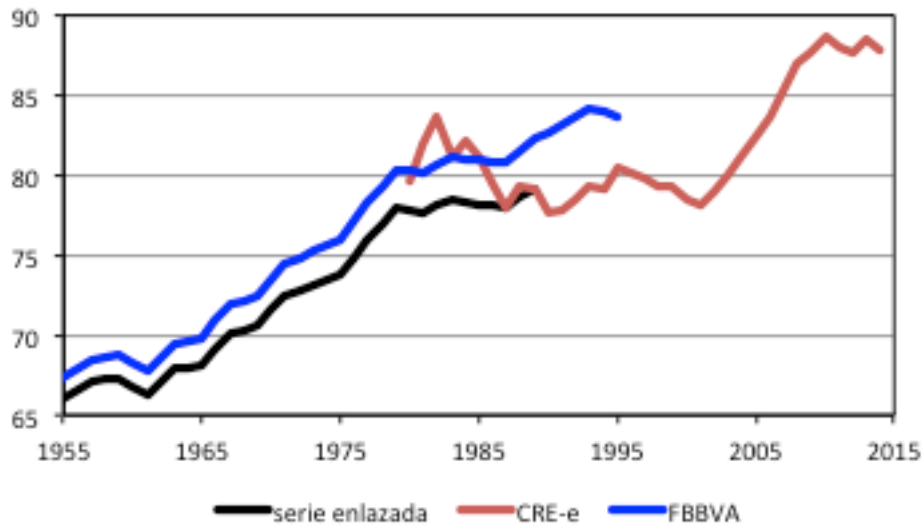
Productividad relativa



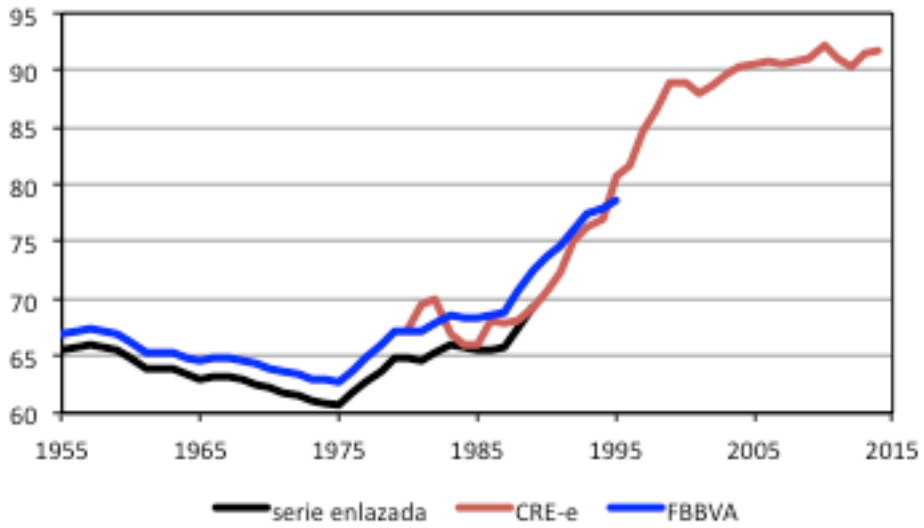
Empleo relativo



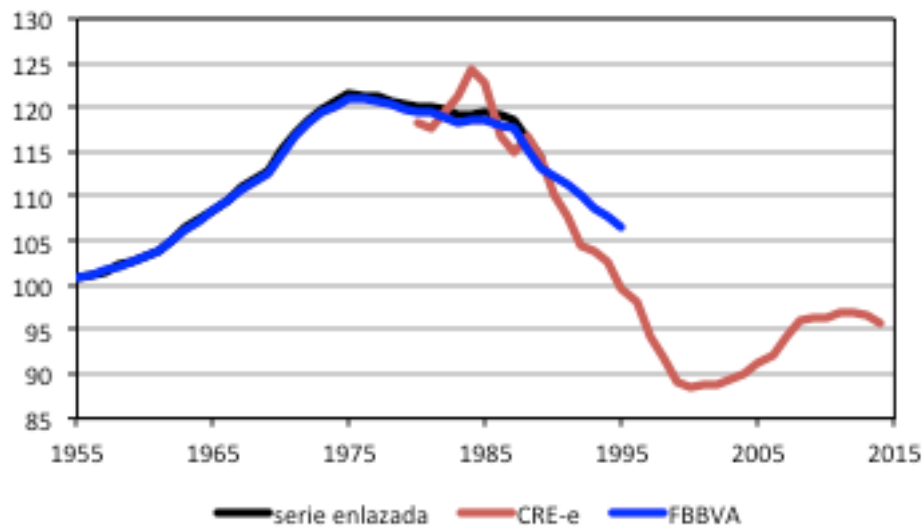
A2.12. Galicia
VAB per capita relativo



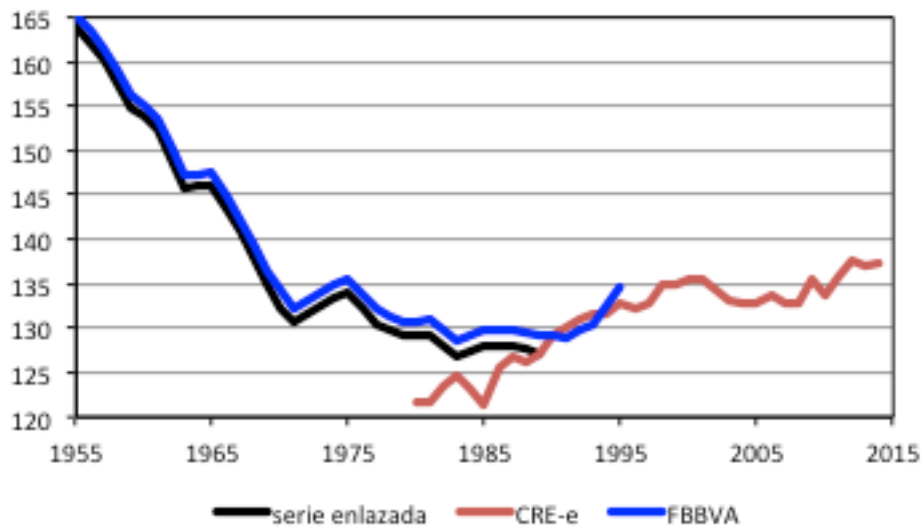
Productividad relativa



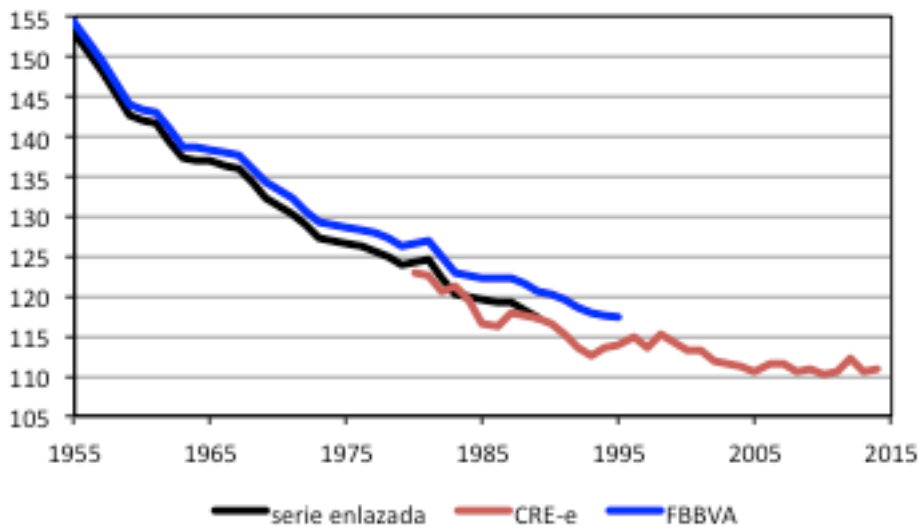
Empleo relativo



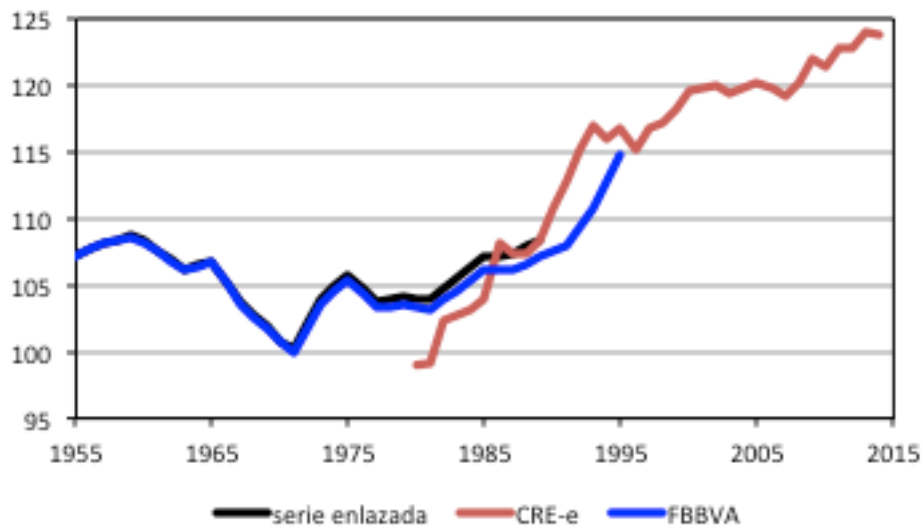
A2.13. Madrid
VAB per capita relativo



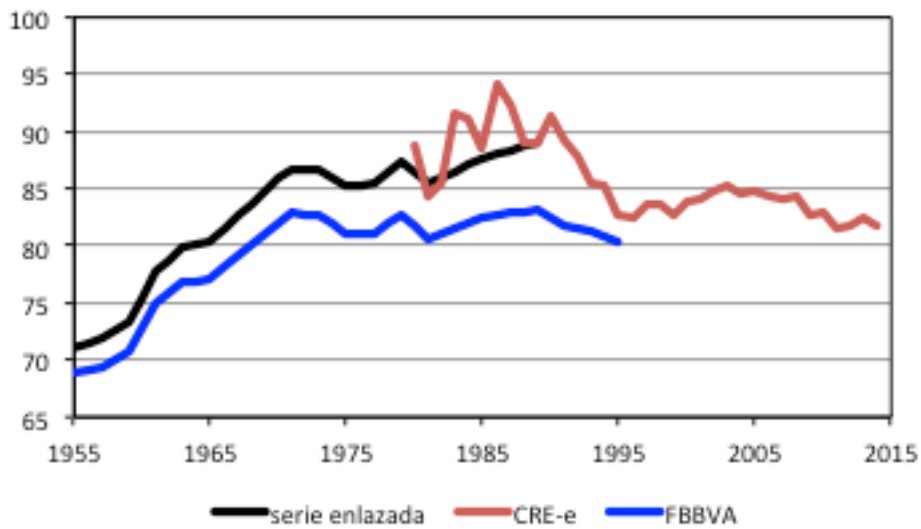
Productividad relativa



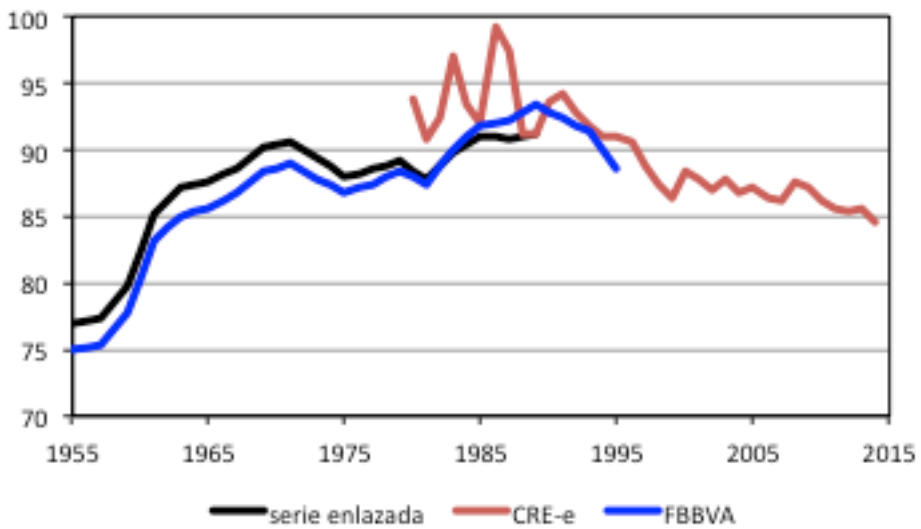
Empleo relativo



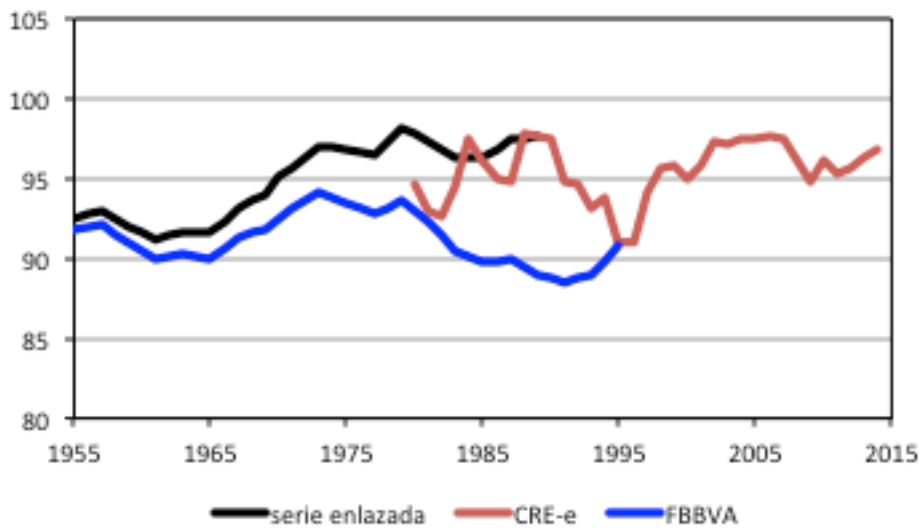
A2.14. Murcia
VAB per capita relativo



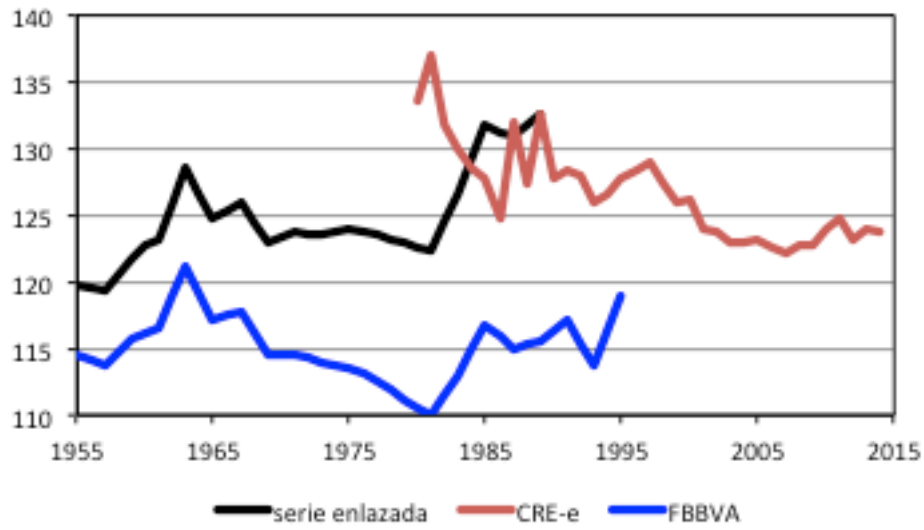
Productividad relativa



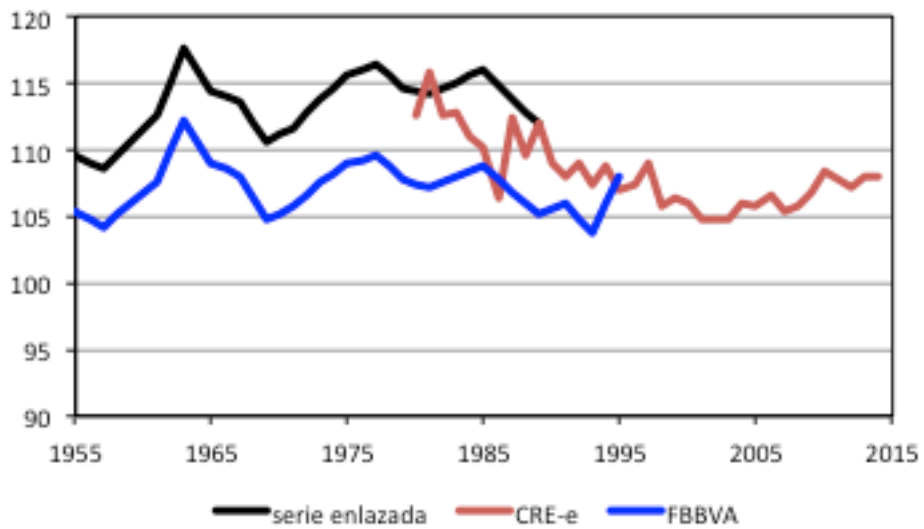
Empleo relativo



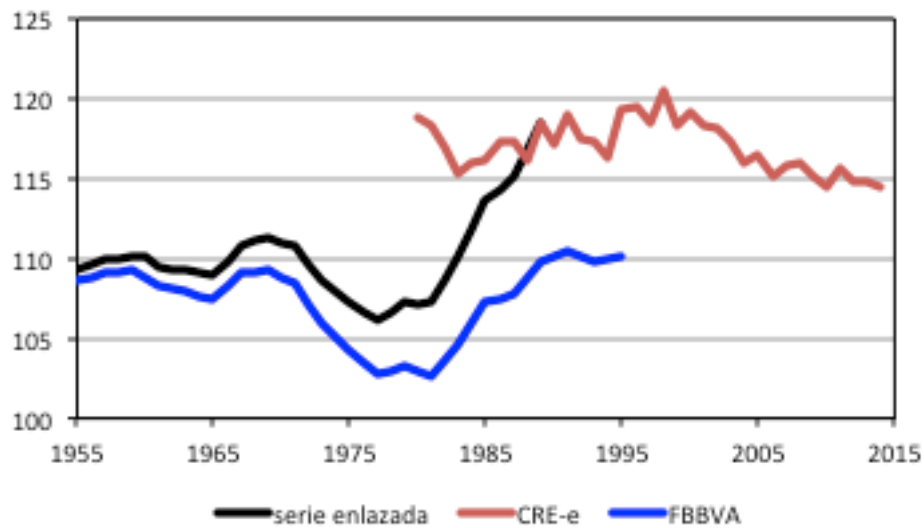
A2.15. Navarra
VAB per capita relativo



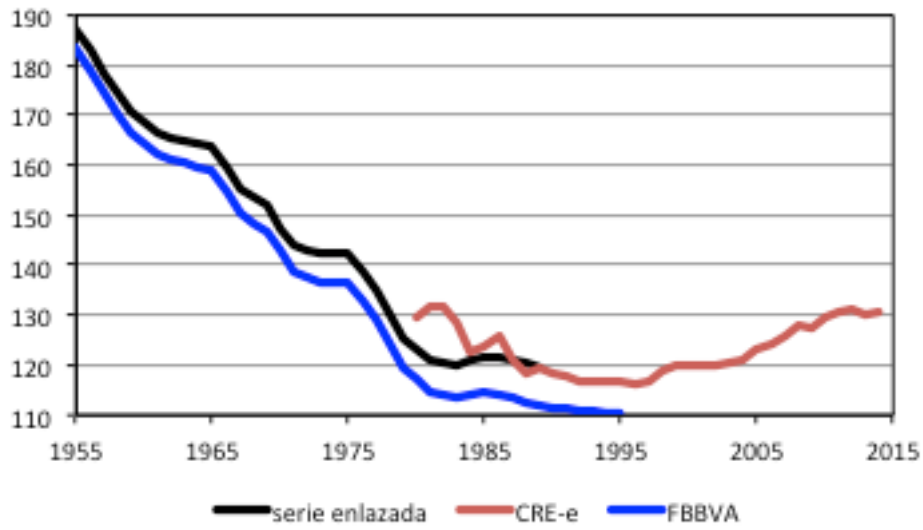
Productividad relativa



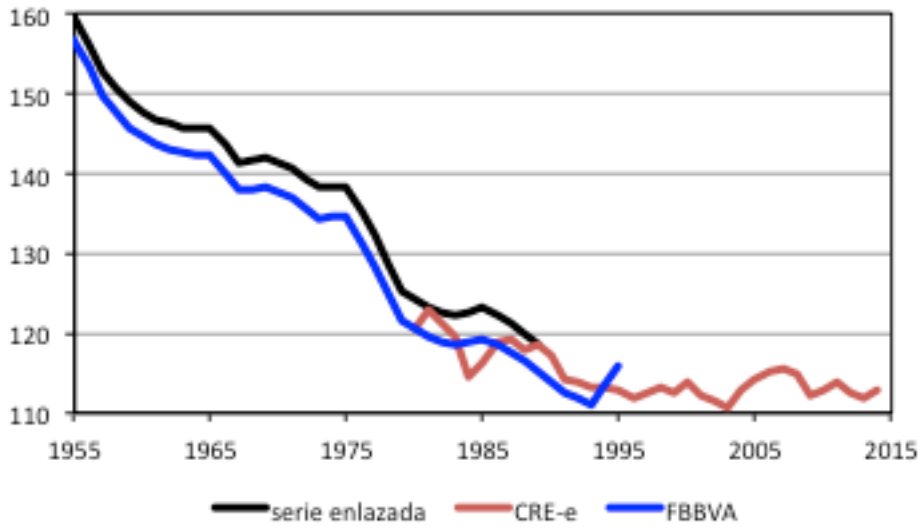
Empleo relativo



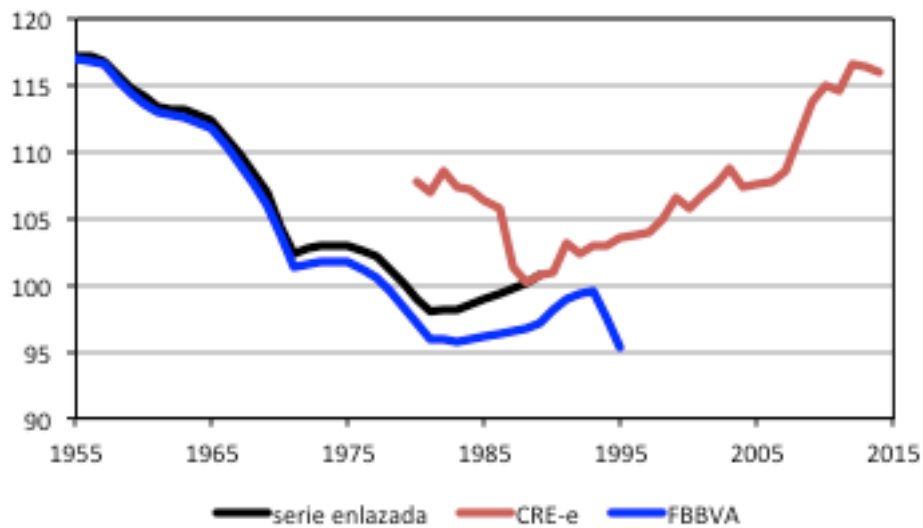
A2.16. País Vasco
VAB per capita relativo



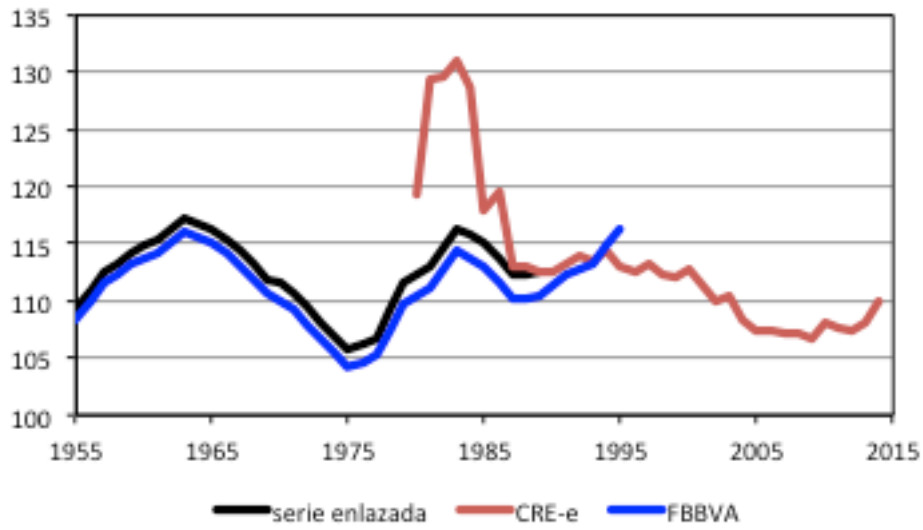
Productividad relativa



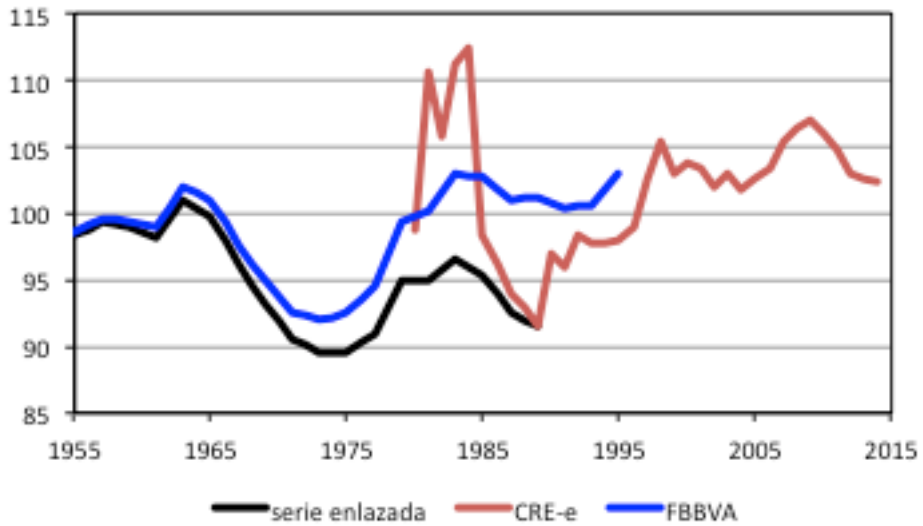
Empleo relativo



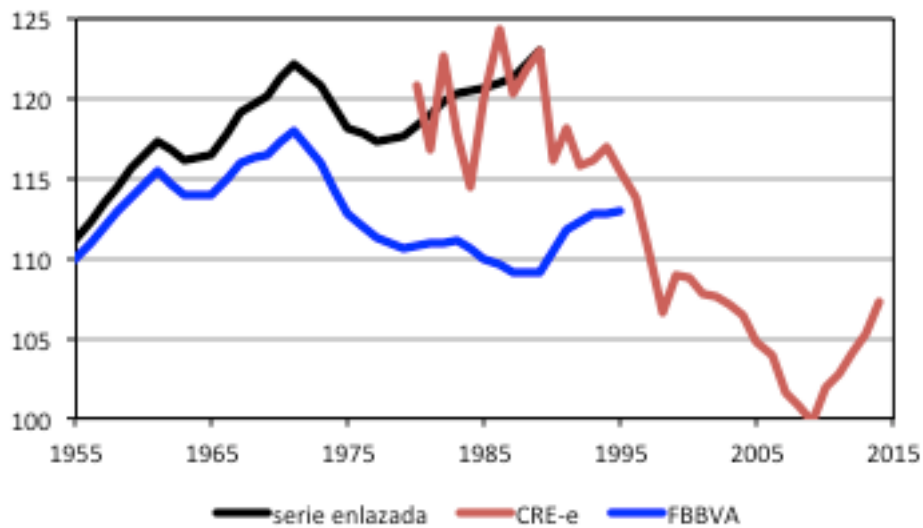
A2.17. Rioja
VAB per capita relativo



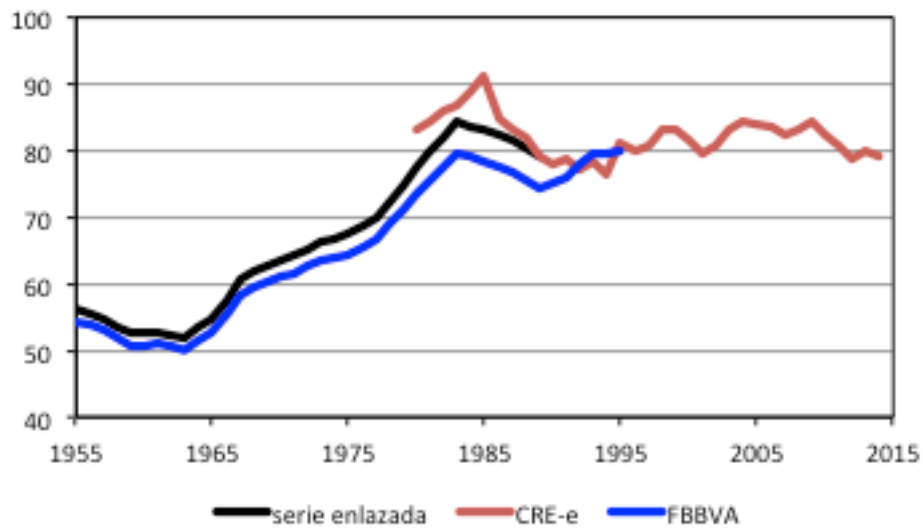
Productividad relativa



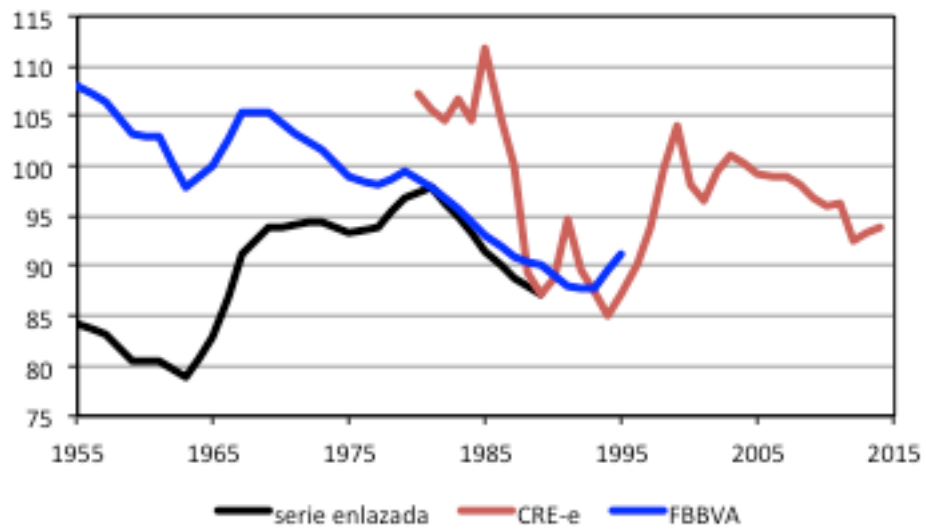
Empleo relativo



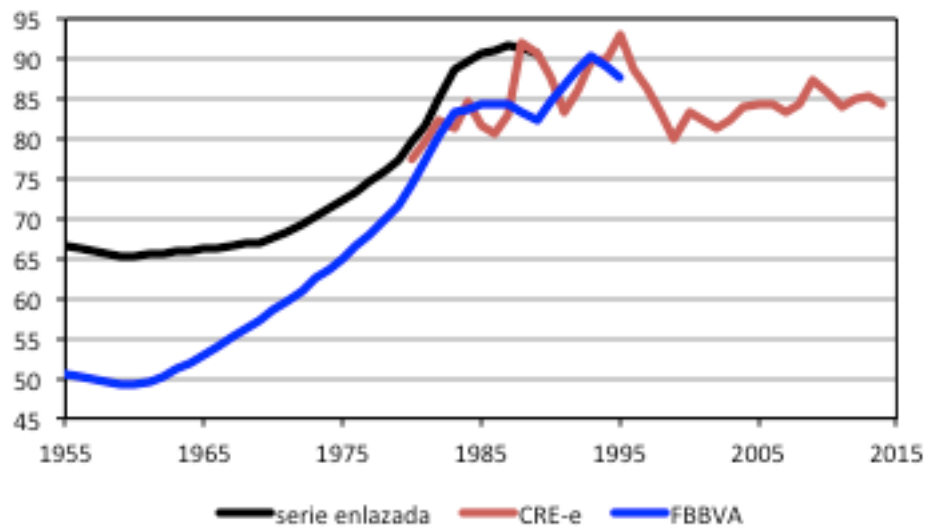
A2.18. Ceuta y Melilla
VAB per capita relativo



Productividad relativa



Empleo relativo



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.
- Alvarez Llano, R. (1986). "Evolución de la estructura regional de España en la historia. Una aproximación." *Situación 1*, pp. 5-61.
- Cañada, A. (1995). "Algunos aspectos metodológicos de la estimación del empleo en la Contabilidad Nacional de España." *Estadística Española* 37(138), pp. 45-73.
- 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.
http://www.igae.meh.es/SGPG/Cln_Principal/Presupuestos/Documentacion/Basesdatos_estudiosregionales.htm
- de la Fuente, A. (2008). "Series enlazadas de algunos agregados económicos regionales, 1995-2007. Versión 1.1." Documento de Trabajo D2008-03. Dirección General de Presupuestos, Ministerio de Economía y Hacienda, Madrid.
<http://ideas.repec.org/p/fda/fdacee/02-09.html>
- de la Fuente, A. (2009). "Series enlazadas de algunos agregados económicos nacionales y regionales, 1955-2007. Versión 2.1." Documento de Trabajo D2009-06, Dirección General de Presupuestos, Ministerio de Economía y Hacienda. Madrid, septiembre de 2009.
- de la Fuente, A. (2010). "Series anuales de algunos agregados económicos y demográficos regionales, 1955-2009 (RegDat versión 2.3)." Mimeo, Instituto de Análisis Económico (CSIC).
- de la Fuente, A. (2014). "A mixed splicing procedure for economic time series." *Estadística Española* 56, 183, 2014, pp. 107-21.
- de la Fuente, A. (2016a). "Series enlazadas de PIB y otros agregados de Contabilidad Nacional para España, 1955-2014. (RegDat_Nac versión 4.2)." FEDEA, Estudios sobre Economía Española no. 2016-02, Madrid.
<http://documentos.fedea.net/pubs/eee/eee2016-02.pdf>
- de la Fuente, A. (2016b). "Series enlazadas de Contabilidad Regional para España, 1980-2014 (RegData_8014 versión 4.3)." FEDEA, Estudios sobre la Economía Española no. 2016-35. Madrid.
<http://documentos.fedea.net/pubs/eee/eee2016-35.pdf>
- de la Fuente, A. (2016c). "Series largas de algunos agregados demográficos regionales, 1950-2015. RegDat-Dem versión 5.0." FEDEA, Estudios de Economía Española no. 2016-14, Madrid.
<http://documentos.fedea.net/pubs/eee/eee2016-14.pdf>
- 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.
- Doménech, R., J. Escribá y M. J. Murgui (1999). "Cambios en precios relativos y crecimiento económico en las regiones españolas." *Revista de Economía Aplicada* VII, pp. 5-29.
- 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, 1961). *Anuario Estadístico de España, Año 1961*. En base de datos INEbase Historia. Madrid. <http://www.ine.es/inebaseweb/25687.do>
- Instituto Nacional de Estadística (INE, 1950). *Censo de la Población de España y territorios de su soberanía y protectorado según el empadronamiento realizado el 31 de diciembre de 1950*. In INEbase, electronic database: Demografía y población. Cifras de población y censos demográficos. Censos de población desde 1900. <http://www.ine.es/inebaseweb/71807.do?language=0>
- Instituto Nacional de Estadística (INE, 1962). *Censo de la Población y de las Viviendas de España según la inscripción realizada el 31 de diciembre de 1960*. En base de datos INEbase: Demografía y población. Cifras de población y censos demográficos. Censos de población desde 1900. <http://www.ine.es/inebaseweb/71807.do?language=0>
- Instituto Nacional de Estadística (INE, 2005a). "Cambio de base de la Contabilidad Nacional de España." 19 de mayo de 2005.
- Instituto Nacional de Estadística (INE, 2005b). "Contabilidad Nacional de España (CNE). Base 2000. Nota informativa." 19 de mayo de 2005.
- Instituto Nacional de Estadística (INE, 2007). "Contabilidad regional de España, base 2000. Serie homogénea 1995-2006. Nota metodológica." En Base de datos electrónica INEbase. Economía: Cuentas Económicas: Contabilidad Regional de España. Madrid. <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft35%2Fp010&file=inebase&L=0>
- Instituto Nacional de Estadística (INE, 2009a). Contabilidad Nacional de España. Datos históricos, Base 1986. En Base de datos electrónica INEbase. Economía. Cuentas Económicas. Contabilidad Nacional de España. Madrid. <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t35/p008/cne/ba86&file=pcaxis>
- Instituto Nacional de Estadística (INE, 2009b). Contabilidad Regional de España, Base 1986. En Base de datos electrónica INEbase. Economía: Cuentas Económicas. Contabilidad Regional de España. Madrid. <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t35/p010/a1996&file=pcaxis>
- Instituto Nacional de Estadística (INE, 2009c). Cifras de población. En Base de datos electrónica INEbase. Demografía y población. Madrid. http://www.ine.es/inebmenu/mnu_cifraspob.htm
- 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=>
- 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.
- 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.
- Maluquer, J. y M. Llonch (2005). "Trabajo y relaciones laborales." En A. Carreras y X. Tafunell, coordinadores. *Estadísticas históricas de España, siglos XIX-XX*, segunda edición. Fundación BBVA, Bilbao, pp. 1155-1245.
- Martin Rodriguez, M. (1992). "Pautas y tendencias de desarrollo económico regional en España: una visión retrospectiva." En J. L. García Delgado y A. Pedreño, directores, *Ejes territoriales de desarrollo: España en la Europa de los noventa*. Madrid, Economistas, pp. 133-55.

- Martín Rodríguez, M. (1993). "Evolución de las disparidades económicas regionales: una perspectiva histórica." En J. L. García Delgado, editor, España, Economía. Espasa Calpe, Madrid, pp. 891-927.
- Plaza Prieto, J. (1953). "El producto nacional de España y su distribución espacial: un análisis hipotético." *De Economía* VI(22), pp. 9-31. Reproducido en J. Velarde, editor (1969). *Lecturas de economía española*. Madrid, Gredos, pp. 198-218.
- Uriel, E., M. L. Moltó y V. Cucarella (2000). "*Contabilidad Nacional de España. Series enlazadas 1954- 1997. (CNEe-86)*". Fundación BBV, Bilbao.

Working Papers

2017

17/10 **Ángel de la Fuente**: Series enlazadas de algunos agregados económicos regionales, 1955-2014. Parte I: Metodología, VAB, PIB y puestos de trabajo

17/09 **José Félix Izquierdo**: Modelos para los flujos de nuevo crédito en España

17/08 **Ángel de la Fuente**: Series enlazadas de algunos agregados económicos regionales, 1955-2014. Parte I: Metodología, VAB, PIB y puestos de trabajo

17/07 **Víctor Adame García, Javier Alonso Meseguer, Luisa Pérez Ortiz, David Tuesta**: Infrastructure and economic growth from a meta-analysis approach: do all roads lead to Rome?.

17/06 **Víctor Adame García, Javier Alonso Meseguer, Luisa Pérez Ortiz, David Tuesta**: Infraestructuras y crecimiento: un ejercicio de meta-análisis.

17/05 **Olga Cerqueira Gouveia, Enestor Dos Santos, Santiago Fernández de Lis, Alejandro Neut y Javier Sebastián**: Monedas digitales emitidas por los bancos centrales: adopción y repercusiones.

17/04 **Olga Cerqueira Gouveia, Enestor Dos Santos, Santiago Fernández de Lis, Alejandro Neut and Javier Sebastián**: Central Bank Digital Currencies: assessing implementation possibilities and impacts.

17/03 **Juan Antolín Díaz and Juan F. Rubio-Ramírez**: Narrative Sign Restrictions for SVARs.

17/02 **Luis Fernández Lafuerza and Gonzalo de Cadenas**: The Network View: applications to international trade and bank exposures.

17/01 **José Félix Izquierdo, Santiago Muñoz, Ana Rubio and Camilo Ulloa**: Impact of capital regulation on SMEs credit.

2016

16/21 **Javier Sebastián Cermeño**: Blockchain in financial services: Regulatory landscape and future challenges for its commercial application

16/20 **Máximo Camacho and Matías Pacce**: Forecasting travelers in Spain with Google queries.

16/19 **Javier Alonso, Alfonso Arellano, David Tuesta**: Factors that impact on pension fund investments in infrastructure under the current global financial regulation.

16/18 **Ángel de la Fuente**: La financiación regional en Alemania y en España: una perspectiva comparada.

16/17 **R. Doménech, J.R. García and C. Ulloa**: The Effects of Wage Flexibility on Activity and Employment in the Spanish Economy.

16/16 **Ángel de la Fuente**: La evolución de la financiación de las comunidades autónomas de régimen común, 2002-2014.

16/15 **Ángel de la Fuente**: La liquidación de 2014 del sistema de financiación de las comunidades autónomas de régimen común: Adenda.

16/14 **Alicia García-Herrero, Eric Girardin and Hermann González**: Analyzing the impact of monetary policy on financial markets in Chile.

16/13 **Ángel de la Fuente**: La liquidación de 2014 del sistema de financiación de las comunidades autónomas de régimen común.

16/12 **Kan Chen, Mario Crucini**: Trends and Cycles in Small Open Economies: Making The Case For A General Equilibrium Approach.

16/11 **José Félix Izquierdo de la Cruz**: Determinantes de los tipos de interés de las carteras de crédito en la Eurozona.

- 16/10 **Alfonso Ugarte Ruiz**: Long run and short run components in explanatory variables and differences in Panel Data estimators.
- 16/09 **Carlos Casanova, Alicia García-Herrero**: Africa's rising commodity export dependency on China.
- 16/08 **Ángel de la Fuente**: Las finanzas autonómicas en 2015 y entre 2003 y 2015.
- 16/07 **Ángel de la Fuente**: Series largas de algunos agregados demográficos regionales, 1950-2015.
- 16/06 **Ángel de la Fuente**: Series enlazadas de Contabilidad Regional para España, 1980-2014.
- 16/05 **Rafael Doménech, Juan Ramón García, Camilo Ulloa**: Los efectos de la flexibilidad salarial sobre el crecimiento y el empleo.
- 16/04 **Ángel de la Fuente, Michael Thöne, Christian Kastrop**: Regional Financing in Germany and Spain: Comparative Reform Perspectives.
- 16/03 **Antonio Cortina, Santiago Fernández de Lis**: El modelo de negocio de los bancos españoles en América Latina.
- 16/02 **Javier Andrés, Ángel de la Fuente, Rafael Doménech**: Notas para una política fiscal en la salida de la crisis.
- 16/01 **Ángel de la Fuente**: Series enlazadas de PIB y otros agregados de Contabilidad Nacional para España, 1955-2014.

Haga clic para acceder a los documentos de trabajo publicados
español e inglés

El análisis, las opiniones y las conclusiones incluidos en el presente documento son propiedad exclusiva del autor del informe y no son necesariamente propiedad de Grupo BBVA.

Las publicaciones de BBVA Research se pueden consultar en la siguiente página web: <http://www.bbvarresearch.com>

Datos de contacto:

BBVA Research

Calle Azul, 4

Edificio de la Vela - 4º y 5º plantas

28050 Madrid (España)

Tel.: +34 91 374 60 00 y +34 91 537 70 00

Fax: +34 91 374 30 25

bbvaresearch@bbva.com

www.bbvarresearch.com