

# **HOMOGENEIZACIÓN DE LAS SERIES DE DATOS OBSERVADOS PARA EL ANÁLISIS DE TENDENCIAS**

**Carlos M. Fdez-Jardón**

Escuela de Negocios Caixavigo  
VIGO

**Anabel Caneda**

Facultad de Economía  
Universidad de Navarra

## **Resumen**

La mayoría de los datos económicos proceden de estimaciones, pues es prácticamente imposible tener datos totalmente exactos de la realidad económica, más aún cuando ésta no es sólo cuantitativa, sino que existen muchos elementos que junto al valor numérico, condicionan su interpretación. Si además se tienen en cuenta los errores humanos al tomar las mediciones, los de los instrumentos de medida y el hecho de considerar cifras agregadas, encontramos que son muchos los elementos que condicionan los resultados de las observaciones. En un reciente trabajo Trívez (1994) analiza el efecto que los distintos tipos de valores anómalos tiene sobre el comportamiento de una serie temporal, tanto para su predicción como para su estimación. Detectar éstos es importante pero en muchos casos que son debidos a errores ajenos al comportamiento real de la economía interesa eliminarlos para poder detectar el verdadero comportamiento de la tendencias.

En este trabajo se sugiere una metodología de homogeneización de datos cara a estudiar comportamientos tendenciales de variables macroeconómicas, incluyendo un ejemplo sobre el desempleo en Navarra.

## HOMOGENEIZACIÓN DE LAS SERIES DE DATOS OBSERVADOS PARA EL ANÁLISIS DE TENDENCIAS

### INTRODUCCIÓN

Es habitual que muchos datos económicos procedan de estimaciones pues, en bastantes casos, es prácticamente imposible tener datos totalmente exactos de la realidad económica, más aún cuando ésta no es sólo cuantitativa, sino que existen muchos elementos que junto al valor numérico, condicionan su interpretación. Si además se tienen en cuenta los errores humanos al tomar las mediciones, los de los instrumentos de medida y el hecho de considerar cifras agregadas, encontramos que son muchos los elementos que afectan a los resultados de las observaciones.

Este hecho limita la actuación práctica, cuando se debe realizar un análisis de la situación económica a partir de una fuente de datos estimados. Por ello, el primer paso que suele tenerse en cuenta es evaluar la calidad de los datos [Stafford (1991)]. En algunos casos también es frecuente hacer uso de técnicas econométricas que tengan en cuenta los errores en la medición de estos.

Generalmente siempre que se valora la bondad de los datos y se observan algunos con baja fiabilidad, bien sea por la fuente, por el transcriptor o por otras diferentes causas, se reestiman estos datos y la nueva serie se utiliza para el análisis, corrigiéndolos de acuerdo a unos parámetros de coherencia con la situación económica observada.

Este hecho, normal en la mayor parte de las economías modernas, hace que cada vez se vaya mejorando la fiabilidad de las observaciones, y como consecuencia, al analizar las de periodos anteriores se es consciente de que puede ocurrir que éstas no tengan la calidad deseable. Baste recordar el considerable esfuerzo que diferentes centros de investigación están realizando por obtener unos datos fiables y de gran calidad, desde el propio I.N.E, prácticamente desde su fundación, los Institutos de estadística regionales<sup>1</sup>, Instituto de estudios fiscales<sup>2</sup>, el propio Banco de España, Servicios de estudios de Bancos y Cajas de Ahorros, como FIES y el servicio de estudios del BBV<sup>3</sup>, y otros centros de investigación privados<sup>4</sup>, que intentan mejorar la fiabilidad de los datos.

A título de ejemplo de los problemas que plantea la contaminación en los datos, Trívez (1994) analiza el efecto que los distintos tipos de valores anómalos tiene sobre el comportamiento de una serie temporal, tanto para su predicción como para su estimación. Detectar estos es importante, y

---

<sup>1</sup>En particular, en Galicia, desde donde se escribe esta comunicación existe una publicación 'Galicia en cifras' del Instituto Galego de Estadística, que además de recopilar datos de diferentes fuentes, en algunos casos propias, homogeneiza estos para su presentación. Por lo que conozco en bastantes otras comunidades autónomas se realizan funciones o publicaciones similares.

<sup>2</sup>Por ejemplo se puede ver la monografía N75 del IEF, de Corrales y Fraguas sobre Series Macroeconómicas para el periodo 1954-88: un intento de homogeneización.

<sup>3</sup>Desde hace varios años publican una estimación bianual de la Renta Nacional de España y un Informe económico.

<sup>4</sup>A la sombra de muchas universidades se han ido creando diferentes institutos o centros de investigación económica, como por ejemplo el Instituto Klein y CEPREDEN, Hispalink, etc...

una vez detectados, tiene interés averiguar las causas de esos efectos. En el caso de que sean debidos a errores ajenos al comportamiento real de la economía interesa eliminarlos para poder estudiar la verdadera evolución de la serie.

Cuando las series son multivariantes y hay que hacer uso de sistemas de ecuaciones los problemas generados por los valores extraños se multiplican de forma exponencial.

Aunque existen muchos métodos para el análisis y tratamiento de la fiabilidad de un conjunto de datos económicos no hay un acuerdo claro sobre cual es el método ideal y por ello son muchas las opciones propuestas para "depurar" los datos, bien sea porque el interés se centra en determinar las tendencias o eliminar estacionalidades, para lo que se hace uso generalmente del filtraje de las series económicas, ARIMA X-11, etc...<sup>5</sup>, o simplemente para eliminar datos que claramente tienen un comportamiento que tergiversa el modelo. (Fair, 1984).

Simplificando al máximo la exposición vamos a considerar dos actitudes extremas, aunque ninguna se da en sentido puro, por lo menos en el ámbito de la economía aplicada, y cada uno puede encontrar su situación según esté mas cerca de una u otra.

El exponer los problemas que presenta cada uno de los métodos que se deducen de cada actitud nos ira sugiriendo criterios en los que se puede basar una suavización de datos. A continuación, como aplicación de esos criterios expondremos la homogeneización en series económicas para el estudio de relaciones a largo plazo, para terminar con un ejemplo concreto referido al desempleo en Navarra.

## **MÉTODOS DE ANÁLISIS DE DATOS ECONÓMICOS**

Supongamos un conjunto de observaciones de una variable aleatoria  $X$  a lo largo del tiempo. Ante este conjunto de datos caben dos opciones extremas:

- No hacer ninguna suposición sobre la variable subyacente.
- Suponer que procede de un proceso estocástico  $X_t$  conocido.

En el segundo caso se puede suponer que sólo se conoce la ley de distribución pero los parámetros que la caracterizan son desconocidos.

Analizaremos los pros y contras de cada de los casos expuestos y a partir de ellos trataremos de diseñar un conjunto de principios que nos ayuden a buscar criterios para la homogeneización de los datos económicos.

### **Datos sin teoría**

El primer caso considera que no hay suposiciones subyacentes. Los datos son los que se observan y así deben tomarse. Por tanto no debe hacerse ningún tratamiento de homogeneización. Quizás el mejor enfoque de esta visión sea dado por unas palabras de Sir Arthur Conan Doyle, que pone en boca del famoso detective Sherlock Holmes *"Recuerda que nosotros nos aproximábamos al caso con la mente totalmente en blanco, lo cual siempre es una ventaja. Sin ninguna teoría previa.*

---

<sup>5</sup>Aunque existen muchos trabajos analizando los peligros de trabajar con series desestacionalizadas por diferentes métodos y los efectos que esto produce, ver por ejemplo Raveh(1984)

*Fuimos allí y observamos y diseñamos nuestras inferencias a partir de nuestras observaciones."*

6

El fuerte crecimiento que al inicio de siglo tuvo la estadística inferencial, ha hecho que se tienda a minusvalorar el análisis descriptivo de datos. Por reacción algunos autores como Mallows y Walley (1980)<sup>7</sup> proponían en la reunión anual de estadísticos celebrada en Houston la posibilidad de una nueva disciplina: "A Theory of Data Analysis?", aunque lo hacían recalando que ponían gran énfasis en la interrogación, pues se trata de delimitar los problemas que esa nueva teoría - de darse - podía llevar consigo.<sup>8</sup>

Este planteamiento parece se presenta bastante lógico cuando los datos observados reflejan la realidad tal como es, pero la mayor parte de las veces los datos son "aproximaciones" de la realidad, bien porque es difícil observar exactamente todos los aspectos que corresponden a un fenómeno real, debido a factores subjetivos o ambientales, o dificultades técnicas, o bien porque al cuantificar la información se pierden matices cualitativos adicionales que son básicos para la interpretación. Por ello en la practica se suele hablar de *errores de medida*.

El efecto que estos errores pueden tener sobre la metodología econométrica ha sido estudiado en algunos casos particulares, por ejemplo en la regresión, (Judge et al. (1988), produciendo heterocedasticidad, unas veces, falsa identificación, otras, etc. según las suposiciones que se hagan sobre esos errores de medición. En otros casos hace falta hacer transformaciones de datos que eliminen esos comportamientos (Andrews, 1971; Breusch and Godfrey, 1987) o ajusten mejor los datos a la teoría subyacente (Box and Cox, 1964)

Por otro lado existen muchas series, que debido a comportamientos ajenos a la situación económica, legislación política, etc..., están afectadas de comportamientos extraños que tergiversan la evolución lógica de las series. Estos hechos no se puede negar que existen, sin embargo si sólo afectan a una variable y no a las demás producen efectos sobre las relaciones globales que son muy difíciles de evaluar y muchas veces el uso de variables ficticias, no es conveniente, pues produce muchas dificultades para la diagnosis del modelo (Mc Aleer, 1995).

También se presentan bastantes casos en que hay varias observaciones ausentes, y estas deben ser *estimadas* de alguna forma, para completar dichas series, si se quiere hacer uso de los cálculos. Harvey y Pierce (1984) sugieren procedimientos para series temporales, Orchard y Woodbury (1972) y Beale y Little (1975) para máxima verosimilitud, y una discusión mas amplia puede encontrarse en Frane (1978).

En resumidas cuentas necesitaremos hacer algún tipo de suposición sobre el comportamiento de los datos , con lo que nos encontraríamos en que datos sin teoría propiamente no se pueden dar.

---

<sup>6</sup>Ver McAleer, (1995)

<sup>7</sup>Ver Fernández-Jardón(1986)

<sup>8</sup>Como un compendio de este enfoque se puede considerar Hoaglin et al. (1983)

## Teoría sin datos

El extremo contrario se presenta cuando se enfocan los problemas desde el ámbito teórico, sin tener datos previamente. Los datos son utilizados como un auxiliar de la teoría para demostrar la validez o no de ésta, pero al observarlos, dada la complejidad de errores que estos acumulan no se aceptan los datos como son y se corrigen para que coincidan lo mas posible con el proceso que los genera y que lógicamente debe adaptarse a la teoría, por lo menos en parte. También para este caso tiene una frase nuestra querido Holmes: *"Yo ya tengo datos. Es un error capital teorizar sin tener ningún dato. Insensiblemente uno comienza a coger los hechos para vestir las teorías en lugar de buscar teorías que vistan a los hechos"*

En muchos libros de texto las suposiciones que se hacen son "heroicas", pero a menudo poco realistas. Por ejemplo Benzecri (1973) comentaba *"estadística no es probabilidad. Algunos autores .... han edificado una pomposa disciplina rica en hipótesis que no son jamás satisfechas."*<sup>9</sup> En este caso interesa mas buscar suposiciones útiles que verídicas. Sin embargo la teoría busca ser contrastada con los datos para que sea legítimo hacer uso de ella a la hora de explicar la realidad. Los datos son los que tienen la última palabra para decir si una teoría es válida o no. En algunos casos parece que existe una enorme brecha entre la teoría económica y la Econometría a la hora de contrastar modelos concretos. Por ejemplo: ¿Cual es la relación apropiada entre la teoría y los métodos utilizados para contrastarla? ¿Podemos obtener todos los datos que necesitamos, y son todos de toda confianza?

Cuando la actitud es de prioridad de la teoría, en el fondo de este planteamiento los datos se buscan para justificar una teoría y esto bien puede hacerse "homogeneizando" los datos, o bien haciendo uso de un número insuficiente de ellos que desde el punto de vista econométrico son capaces de adaptarse a cualquier modelo. Esta es una de las causas del desprecio que en algunos ámbitos se le dan a las técnicas econométricas. Quizás el extremo de esta concepción venga dado por el famoso adagio, que cuentan de un sabio matemático: "Si la realidad no se adapta a la teoría, peor para la realidad".

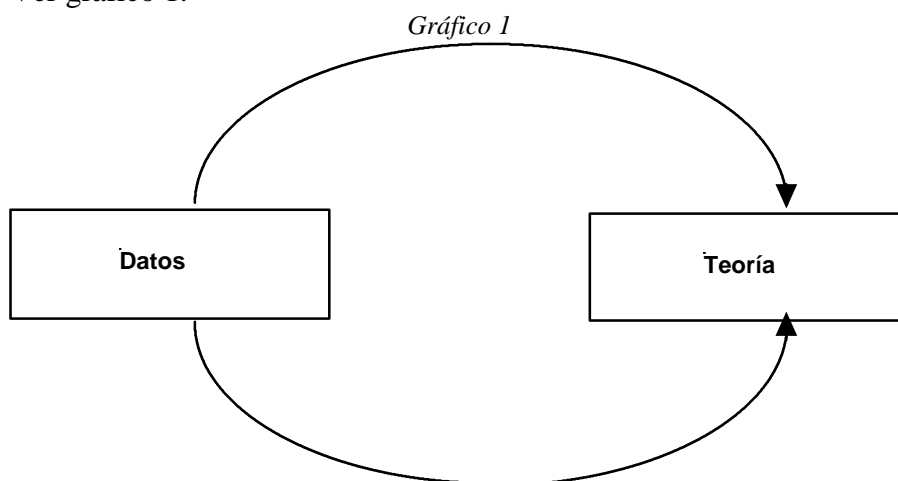
## Teoría y datos

La metodología propuesta en este trabajo está basada en Mc Aleer (1995) de forma que se tengan en cuenta tanto la teoría como los datos conjuntamente. Es un proceso cíclico, que en muchos casos, es casi indiferente el punto de partida, aunque casi siempre se exige en un mínimo de teoría: los datos contrastan esta y a la vez le añaden mejoras, y la teoría al analizar los datos nos sugiere como deberían comportarse y donde presentan formas anómalas que no tienen interés para el estudio que se está realizando y por lo tanto conviene eliminarlos o corregirlos de acuerdo al comportamiento esperado. La justificación de fondo se basa en que la realidad es de suyo muy compleja, y cada modelo es una simplificación de esta realidad que presenta sólo algunos aspectos de la verdad. Por ello nuevas observaciones ayudan a mejorar el modelo o descubrir aspectos complementarios de la verdad presente en los datos que mejora la simplificación inicial. Si sin

---

<sup>9</sup>Ver Fernández-Jardón(1986)

embargo, no interesa, por el tipo de estudio entrar en complicaciones que apenas mejoran la realidad estudiada, nos quedaríamos con el modelo inicial y reajustaríamos las nuevas observaciones al valor que sería esperado, por tratarse de un fenómeno adicional que lo hace variar. Los diferentes criterios de selección de modelos ver, por ejemplo, (Aznar et al., 1988) Granger, C. W. J. (ed. lit.) (1991), et....) y de diagnóstico de modelos (Achen, 1986; Atkinson, 1987; Fox, 1991, etc...) nos pueden ayudar a comprender mejor todo este ámbito. Cabe decir que se presupone que la teoría y los datos están muy relacionados, y bien se puede plantear el modelo de la teoría a los datos o viceversa aunque lo habitual es una retroalimentación entre ambos. Ver gráfico 1.



En el texto se plantea en primer lugar los principios generales para la homogeneización de datos. La mayor parte se deducen de los criterios generales de validez de una teoría, que van a ser necesarios para justificar la teoría elegida. A continuación se expone la aplicación de esos principios al estudio de relaciones a largo plazo entre variables económicas.

### **PRINCIPIOS GENERALES PARA LA HOMOGENEIZACIÓN DE DATOS**

Consideremos un conjunto de observaciones de una variable aleatoria  $X$  multidimensional. ¿Como comprobar si los datos son todos fiables? ¿Que pasaría si cambiáramos alguno de ellos? ¿Que efectos se producen al homogeneizarlos?

La idea básica de la homogeneización consiste en presuponer que el conjunto de datos proviene de un proceso generador de datos, con una ley de distribución multidimensional aunque desconocida. Homogeneizar consiste en eliminar o transformar los datos que no se ajusten a dicho proceso de forma que se pueda suponer que todos han sido generados por el mismo proceso.

La justificación fundamental de la necesidad de la homogeneización de datos se basa en la calidad de éstos y en localizar aquellos que son fundamentales para el estudio: *Es importante reconocer que datos son incidentales y cuales son fundamentales.*"

Es conveniente usar una mezcla de ambos: la información que suministran los datos y el ingenio del investigador.

Se trata de buscar una guía que nos ayude a determinar cuales son las relaciones causa-efecto en los datos: detectar factores que puedan ser usados como ayuda a la interpretación de los hechos

reales, y para ello hay que distinguir entre los factores que son importantes de los que son meramente coincidentes.

Para ver la validez de los datos y los límites de la homogeneización pensamos deben tenerse en cuenta una serie de principios básicos que rijan dicho comportamiento:

***1. Tener un problema concreto a estudiar:***

Es claro que el hecho de utilizar datos es debido a que se quiere resolver algún problema concreto y más en el ámbito de la economía aplicada. La homogeneización va a depender del tipo de problema que se aborde. No es lo mismo, por ejemplo, hacer las predicciones de ventas de una empresa, que ver como afectó la crisis del petróleo a esas ventas.

***2. Elaborar una teoría económica coherente que pueda ser contrastable con los datos***

La teoría nos ayuda siempre a enmarcar el problema concreto y por lo tanto el comportamiento de las series económicas, para ver en que ámbito nos movemos y que elementos interesa analizar. De igual forma las series suelen indicar comportamientos característicos de algunas variables que sugieren fundamentos de la teoría económica que debe usarse.

***3. Dar prioridad a los datos sobre la teoría***

Las tendencias y comportamientos de los datos no deben someterse a una rígida base teórica son que deben servir para contrastar la validez de ésta o ver si necesita variables o elementos nuevos, pero los datos están ahí y la teoría es algo elaborado por la mente humana, y por lo tanto susceptible de error. Es lógico que cuanto mayor sea el acuerdo de la comunidad científica sobre un punto, mas fácilmente se puede aceptar esa teoría pero si los datos la contradicen, en algo falla.

***4. Debe existir un proceso generador de datos***

Quizás su ley o su comportamiento concreto no sea conocido, pero si éste no existe es imposible elaborar hipótesis estadísticas que puedan luego ayudarnos a realizar contrastes econométricos. De esa forma comportamientos extraños en alguno de los datos permiten la corrección de ese dato por uno mas uniforme con el proceso genérico.

Definidos los principios generales y dado que el problema a estudiar limita el proceso de homogeneización, vamos a modo de ejemplo a analizar la homogeneización de series económicas para el estudio de relaciones a largo plazo. Este es un proceso ya tradicional en el análisis económico, donde un elemento de homogeneización típico es la desestacionalización de las series, pues no afectan a los comportamientos de las tendencias, hasta el punto de que bastantes instituciones como por ejemplo, el F.M.I. publican las series económicas estacionalizadas y desestacionalizadas.

**Homogeneización de series económicas para estudiar comportamientos a largo plazo**

Nos interesa analizar fundamentalmente los efectos debidos a las relaciones a largo plazo de las variables que afectan al desempleo<sup>10</sup>,. Por lo tanto no debemos tener en cuenta otros efectos de la

---

<sup>10</sup>En realidad para que el estudio fuera completo deberíamos definir el modelo en el que nos movemos y las variables que tienen interés, pero por simplicidad utilizaremos únicamente ésta

economía que podrían afectar a esta variable. Por ello estableceremos un método de homogeneización de las series de datos, de manera que se eliminen los efectos estacionales o puramente coyunturales, ajenos al comportamiento a largo plazo del desempleo.

Dicho método se basa generalmente, en una sucesión de pasos, partiendo del estudio de la coherencia de los datos con la evolución temporal de la economía y de la posibilidad de eliminar aquellos efectos coyunturales no específicos del modelo, esto es, los que afectan al corto plazo y no afectan a la relación a largo plazo entre las variables de interés.

En primer lugar debemos tener en cuenta el comportamiento de las series, pues si estas no son estacionarias no se observará ninguno de los efectos coyunturales ocultos por la fuerte tendencia de la serie. Por lo tanto, estableceremos las suposiciones básicas del proceso para luego elaborar los pasos a tener en cuenta para la homogeneización:

Sea la serie  $X_t$  un proceso estocástico cuyas observaciones vienen dadas por  $(x_1, \dots, x_T)$ . Para ver los efectos coyunturales debemos trabajar con series estacionarias que reflejan los comportamientos instantáneos en el tiempo. Inicialmente supondremos que se puede descomponer la serie en dos procesos:

$$X_t = T_t + Z_t$$

donde  $T$  nos mide el efecto estructural de la serie y  $Z$  el efecto coyuntural. Si la serie es  $I(d)$  el efecto coyuntural no es observable en los datos, por tanto para realizar el análisis posterior conviene presentar el modelo de la siguiente forma:

$$\begin{aligned} X_t &= f(t) + T_t \\ \Delta^d T_t &= u_t + Z_t \end{aligned}$$

donde  $u_t$  es estacionario y aglutina los efectos aleatorios,  $f(t)$  nos mide la tendencia determinística de la serie y  $Z_t$  mide los efectos coyunturales determinísticos o estocásticos especiales que denominamos *variables de contaminación*.

En general, partiendo del modelo anterior, una vez obtenida la serie estacionaria el esquema general podría establecerse constando de los siguientes pasos:

- 1. Selección de datos atípicos que contaminan al comportamiento estructural de la serie
- 2. Análisis de efectos estacionales determinísticos o estocásticos no estacionarios
- 3. Selección de variables que miden los efectos que conjuntamente cambian coyunturalmente el comportamiento tendencial de la serie, y eliminación de esos efectos
- 4. Diagnóstico de las series revisadas

Desarrollamos brevemente cada uno de los pasos anteriores

### ***1. Selección de posibles datos atípicos***

Para analizar los valores atípicos el primer paso será convertir la serie en estacionaria para lo cual debemos diferenciarla  $d$  veces. Una vez diferenciada se realiza una regresión de cada componente de la serie respecto al tiempo, pues es posible que aún queden efectos determinísticos de la tendencia de la serie sin diferenciar, y a partir de ella se estudia que valores son atípicos e influyentes, y se define una variable ficticia para cada uno de esos puntos, sea como valores de

impulso, innovacional, de cambio de nivel o cambio temporal<sup>11</sup>. Después indicaremos si esos puntos se consideran como extraños o no en todo el proceso.

## ***2. Selección de efectos estacionales determinísticos***

Para seleccionar los efectos estacionales determinísticos de la serie observada, en primer lugar se define una variable ficticia para cada uno de los periodos estacionales considerados, que pueden producir estacionalidad y a continuación se realiza una regresión paso a paso de cada componente de la serie diferenciada sobre estas variables ficticias, las definidas en el apartado anterior y la tendencia temporal esperada. A partir de ella se estudia que variables de las definidas tienen un efecto significativo sobre el comportamiento de la serie. El hecho de hacer uso de una regresión paso a paso tiene su justificación en indicar cual o cuales de los periodos son importantes de forma que nunca se incurra en la denominada "trampa de las variables ficticias" (Gujarati, 1981). De igual forma se ve si existe una tendencia temporal determinística además de la estocástica ya definida por la diferenciación de la serie.

## ***3. Eliminación de efectos coyunturales ajenos a la relación de la ecuación***

Para eliminar los efectos estacionales determinísticos de la serie observada y los efectos de puntos anómalos, no propios del comportamiento de la ecuación, se deben considerar aquellos que no sean comunes a la variable dependiente y a alguna de las independientes, por la cual pueden haber sido causados. Para ello se analizan las variables ficticias definidas que no verifican ambas condiciones y se define una nueva variable a partir de la serie diferenciada, restándole el efecto de esas variables. Sin embargo las tendencias no las eliminamos pues hacen relación a comportamientos a largo plazo.

## ***4. Diagnósis***

A partir de esa serie diferenciada ya homogeneizada, se realiza un diagnóstico de su comportamiento. Para ello, se comprueba si los residuos de la regresión del paso 2 tienen un comportamiento de series estacionarias, y si no quedan efectos adicionales sobre las variables así homogeneizadas.

A partir del cuarto punto se repiten desde el primero al tercero, si la diagnosis nos dice que aún quedan efectos en la serie de datos observados. Una vez comprobada que la diagnosis es buena, se integra  $d$  veces la serie diferenciada para volver al orden  $I(d)$  original, y esas serán las series con las que trabajaremos.

A lo largo de lo que resta en este apartado, se analizan los cuatro puntos indicados para cada una de las series de datos de desempleo en la economía navarra.

## **Homogeneización de la tasa de desempleo de la economía navarra**

Definimos un proceso  $X_t$  que recoge el comportamiento del desempleo a lo largo del tiempo. Expondremos previamente un comentario más desarrollado de esta variable, para analizar luego su homogeneización a partir del modelo global. Estudiamos primero su orden de integrabilidad,

---

<sup>11</sup>Balke(1993) sugiere la dificultad que tiene el hecho de detectar cambios de nivel en series temporales. En su artículo pueden considerarse otras metodologías mas complejas. En este caso nos reduciremos a la mas elemental de todas, para facilitar la exposición del trabajo.

para una vez diferenciada  $d$  veces para convertirla en estacionaria analizar los valores extraños. Para comprobar si las series son  $I(1)$  o no realizamos los test de Dickey-Fuller (1981). La tabla 1 recoge los resultados del desempleo

Tabla 1

NO CONSTANT, NO TREND			
A(1)=0	T-TEST	-.66717	-1.62
-----			
CONSTANT, NO TREND			
A(1)=0	T-TEST	-.81059	-2.57
A(0)=A(1)=0		.46171	3.78
-----			
CONSTANT, TREND			
A(1)=0	T-TEST	-2.8728	-3.13
A(0)=A(1)=A(2)=0		2.9722	4.03
A(1)=A(2)=0		4.2929	5.34
-----			

Por lo tanto debemos diferenciar las series para analizar los comportamientos extraños.

1. Realizamos la regresión del diferencial de la tasa de desempleo diferenciada respecto al tiempo, medido por una variables de evolución lineal. Obtenemos los resultados expresados en la tabla A.11.1.

La impresión es que la tendencia no es significativa, y que el único valor que es atípico e influyente corresponde a la observación realizada en el tercer trimestre del 85.

2. Una vez definidas las variables ficticias estacionales que denominaremos por  $s_1$ ,  $s_2$ ,  $s_3$  y  $s_4$ , indicando el trimestre correspondiente realizamos una regresión paso a paso del diferencial de la tasa de desempleo diferenciada respecto al tiempo, medido por una variable de evolución lineal, a la variable ficticia, que denominamos  $d_{10}$ , y a las variables estacionales. Obtenemos los resultados expresados en la tabla A.11.2.

Como se puede observar la tendencia no es significativa, y que el único efecto estacional se da en los segundos trimestres en que la diferencial de la tasa de desempleo baja respecto a su valor normal. También tiene un efecto significativo el valor atípico e influyente que corresponde a la observación realizada en el tercer trimestre del 85.

3. Definimos la serie diferenciada corregida, eliminando ambos efectos de la serie original diferenciada, y comprobamos que los residuos de la regresión anterior son estacionarios

Integramos por lo tanto la serie para obtener la diferencial de la tasa de desempleo homogeneizada.

## CONCLUSIONES

Hemos definido unos principios generales de homogeneización de series económicas que consideramos tienen una cierta validez universal. A continuación estudiamos su aplicación al estudio de tendencias a largo plazo y lo aplicamos a la homogeneización de la tasa de desempleo en el caso de la comunidad autónoma navarra.

## BIBLIOGRAFÍA

- ACHEN, C. H. (1986) Interpreting and using regression. Sage, Beverly Hills
- ANDREWS, D.F. , (1971) A note on the selection of data transformations *Biometrika*, 249 254.
- ATKINSON, A.C. (1987) Plots, transformations and regression : an introduction to graphical methods of diagnostic regression analysis. Clarendon Press, Oxford

- AZNAR , A., APARICIO, M. T., TRIVEZ , F. J. (1988) Selección y fiabilidad de modelos econométricos para la política económica. Instituto de Estudios Fiscales, Madrid
- BALKE N.S. (1993) Detecting level shifts in time series. J. of Bus. & Econ. Stast. V11N1p81-92.
- BEALE E.M.L. AND LITTLE R.J.A. (1975) Missing values in multivariate analysis. J. Roy Statist. Soc. 37 B 129-145.
- BOX, G. E. P. AND COX, D. R. (1964) An analysis of transformations, *Journal of the Royal Statistical Society B*, 26, 211-243.
- BREUSCH, T. S. AND GODFREY, L. G. (1987) Data transformation tests, *Economic Journal*, 96. 47-58.
- CHATTERJEE, SAMPRIT / HADI, ALI S. (1988) Sensitivity analysis in linear regression. Wiley & Sons, New York
- DICKEY, D. A Y FULLER, W. A. (1981): " The likelihood ratio statist for autorregresive tiem series with a unit roots", *Econometría*, Vol. 49, pp. 1057-1072.
- FAIR, RAY C. (1984) Specification, estimation, and analysis of macroeconomic models. Harvard University Press, Cambridge [Mass.]
- FERNANDEZ-JARDON C.M. (1986) Proyecto Docente de Estadística. Universidad de Santiago de Compostela.
- FOX, J. (1991) Regression diagnostics. Sage, Newbury Park
- FRANE J.W. (1976) Some simple procedures for handling missing data in multivariate analysis. *Psychometrika* 41, 409-415.
- GRANGER, C. W. J. (ed.lit.) (1991) Modelling economic series : readings in econometric methodology. Clarendon Press, Oxford
- GUJARATI R. (1991) Econometría . McGrawHill
- HARVEY, A. C. AND PIERCE, R. G. (1984) Estimating missing observations in economic time series, *Journal of the American Statistical Association*, 79, 125-131.
- HOAGLIN D.C., MOSTELLER, F. & TUKEY J.W. (eds) (1983) Understanding Robust and exploratory Data Analysis. Wiley. NY.
- MCALLER M. (1995): Sherlock Holmes and the search for thuth, in: Les Oxley, Donald A. R. George, Colin J. Roberts and Stuart sayer, eds., *Surveys in econometrics*, Blackwell.
- ORCHARD AND WOODBURY (1972) A missing information principles: theory and applications. Proc 6th Berkeley Symp. 1, 697-715
- RAVEH A. (1984) Comments on some properties of X-11. J. of Economic Studies V66p343-348
- STAFFORD F. (1991): La función de los microdatos en la investigación de economía laboral, o como prevenir el fallecimiento de la economía empírica, en: Orley, C. A. y Layard, R. (comps.), *Manual de economía del trabajo*. I, Ministerio de trabajo y Seguridad social.
- TRIVEZ F.J. (1994) Efectos de los distintos tipos de outliers en las prdiicones de os modelos ARIMA. *Estadística Española* V35N135p21-58

# APÉNDICES

Tabla A.11.1

VARIABLE	ESTIMATED	STANDARD	T-RATIO	PARTIAL STANDARDIZED ELASTICITY
NAME	COEFFICIENT	ERROR	37 DF	P-VALUE CORR. COEFFICIENT AT MEANS
T	-0.99016E-02	0.1444E-01	-0.6857	.249 -.112 -.1120 -.4832
CONSTANT	.63830	.3440	1.855	.964 .292 .0000 1.4832
RESIDUAL	RSTUDENT	HT	COVRAT	DFRITS DFFIT
2	.76390	.7886	.0987	1.1325 .2610 0.83670E-01
3	-.53253	-.5452	.0912	1.1434 -.1727 -0.53459E-01
4	-.31066	-.3159	.0841	1.1470 -.0958 -0.28542E-01
5	-.43597	-.4423	.0775	1.1327 -.1282 -0.36608E-01
6	1.1044	1.1334	.0712	1.0603 .3138 0.84646E-01
7	.48361	.4878	.0653	1.1154 .1289 0.33795E-01
8	.39040	.3921	.0599	1.1140 .0989 0.24853E-01
9	.40286	.4036	.0548	1.1075 .0972 0.23353E-01
10	2.9604	3.3909	.0501	6.388 .7790 .15625
11	-1.2994	-1.3240	.0459	1.0067 -.2904 -0.62491E-01
12	0.61584E-01	.0612	.0420	1.1025 .0128 0.27025E-02
13	-.26877	-.2667	.0386	1.0944 -.0534 -0.10790E-01
14	-1.1091	-1.1165	.0356	1.0232 -.2144 -0.40895E-01
15	-0.33448E-01	-.0331	.0329	1.0922 -.0061 -0.11389E-02
16	-1.0464	-1.0487	.0307	1.0261 -.1866 -0.33145E-01
17	.59281	.5874	.0289	1.0672 .1013 0.17630E-01
18	-1.2214	-1.2287	.0275	1.0005 -.2065 -0.34490E-01
19	-1.0780	-1.0789	.0265	1.0181 -.1778 -0.29289E-01
20	.93707	.9338	.0258	1.0337 .1521 0.24860E-01
21	-0.69745E-01	-.0687	.0256	1.0838 -.0111 -0.18354E-02
22	-1.3780	-1.3930	.0258	.9763 -.2269 -0.36557E-01
23	-1.5753	-1.6063	.0265	.9448 -.2648 -0.42799E-01
24	1.6328	1.6703	.0275	.9355 .2807 0.46108E-01
25	-0.59749E-01	-.0589	.0289	1.0875 -.0102 -0.17769E-02
26	-1.5186	-1.5482	.0307	.9580 -.2755 -0.48099E-01
27	1.2611	1.2742	.0329	1.0001 .2351 0.42941E-01
28	-1.4081	-1.4327	.0356	.9803 -.2751 -0.51918E-01
29	1.5063	1.5415	.0386	.9669 .3089 0.60470E-01
30	-.73488	-.7352	.0420	1.0703 -.1540 -0.32248E-01
31	-0.41177E-01	-.0410	.0459	1.1070 -.0090 -0.19802E-02
32	.72987	.7333	.0501	1.0796 .1685 0.38529E-01
33	-.31772	-.3181	.0548	1.1113 -.0766 -0.18417E-01
34	.23487	.2356	.0599	1.1201 .0594 0.14952E-01
35	-.29463	-.2965	.0653	1.1246 -.0784 -0.20589E-01
36	-.22329	-.2253	.0712	1.1341 -.0624 -0.17114E-01
37	.54900	.5579	.0775	1.1255 .1617 0.46098E-01
38	.79941	.8194	.0841	1.1115 .2484 0.73445E-01
39	-.35625	-.3639	.0912	1.1539 -.1153 -0.35762E-01
40	.90276	.9353	.0987	1.1171 .3096 0.98880E-01

\*\*\*\* STEPWISE REGRESSION \*\*\*\*

\*\*\*STEPPING PARAMETERS ARE: FE= 4.0000 AND FX= 3.9000.

**** STEPPING SEQUENCE ****									
STEP	VARIABLE					D.F.	D.F.		
NUMBER	LABEL	STATUS	F-VALUE	NUM.	DEN.	IF-PROBABILITY			
1	S2	[STEPPED IN]	29.9006	1	37	.000003			
2	D10	[STEPPED IN]	13.3957	1	36	.000802			
*SUMMARY FOR POTENTIAL VARIABLES NOT ENTERED INTO THE REG. EQUATION*									
T	[IF ENTERED]		.0568	1	35	.813065			
S1	[IF ENTERED]		.7381	1	35	.396112			
S3	[IF ENTERED]		2.4405	1	35	.127235			
S4	[IF ENTERED]		4.727	1	35	.496301			
**** END OF STEPPING SEQUENCE ****									
R-SQUARE = .5969 R-SQUARE ADJUSTED = .5745									
VARIANCE OF THE ESTIMATE-SIGMA**2 = .78138									
STANDARD ERROR OF THE ESTIMATE-SIGMA = .88396									
SUM OF SQUARED ERRORS-SSE = 28.130									
MEAN OF DEPENDENT VARIABLE = -0.94785E-01									
LOG OF THE LIKELIHOOD FUNCTION = -48.9674									

VARIABLE	ESTIMATED	STANDARD	T-RATIO	PARTIAL STANDARDIZED ELASTICITY
NAME	COEFFICIENT	ERROR	36 DF	P-VALUE CORR. COEFFICIENT AT MEANS
S2	-2.3891	.3373	-7.083	0.000 -.763 -.7799 6.4630
D10	3.4103	.9318	3.660	1.000 .521 .4030 -.9226
CONSTANT	.43037	.1641	2.622	.994 .400 .0000 -4.5405