

# **“ESTUDIO DE LA IMPLICACIÓN DE LOS DIFERENTES NIVELES DE INFORMACIÓN SOBRE LOS RESULTADOS DE AJUSTE EN EL CASO DE LA COMUNIDAD CASTELLANO-LEONESA.”**

Esteban García, Jesús.

López Rodríguez, M<sup>a</sup> Isabel.

Ruiz Ponce, Félix.

Departamento de Economía Aplicada.

Facultad de Ciencias Económicas y Empresariales.

Universidad de Valencia.

## **Resumen**

Considerando los datos desagregados, correspondientes a los Ingresos y Gastos familiares de cada una de las provincias que componen la comunidad de Castilla - León, aportados por las dos últimas Encuestas de Presupuestos Familiares el objetivo del presente estudio es doble: por una parte validar el modelo Gamma biparamétrico para las densidades de las dos variables anteriormente citadas y por otra contrastar la validez de estudios realizados hasta la década de los 80 en los que se utilizaba únicamente la información agrupada en decilas.

Para llevar a cabo el análisis propuesto se utilizará, al trabajar con toda la muestra, el test de Kolmogorov - Smirnov, incluyéndose un test de ajuste adicional, al trabajar con datos agregados, consistente en la “verificación de las cotas de Gastwirth”, así como la propuesta de Rojo Olivas para el representante de la decila.

## 1. INTRODUCCIÓN.

Múltiples son los trabajos desarrollados acerca de la “distribución de la renta”, escapando al objetivo del presente estudio hacer una descripción exhaustiva de los distintos enfoques bajo los cuales han sido realizados así como de las distribuciones de probabilidad propuestas. En cuanto a los trabajos realizados cuya meta principal era obtener un modelo de probabilidad que sirviera para describir adecuadamente la renta, destacar la gran cantidad de los mismos y tener en cuenta que dado que uno de los objetivos del presente análisis es la validación del modelo Gamma, se procederá a realizar un recorrido histórico que permita tener una visión global, sin menoscabo de los que no se citen, de los resultados obtenidos por algunos de los autores más relevantes al plantear el mismo. En este sentido hay que destacar las aportaciones hechas por Salem y Mount, Dagum y Rojo Olivas. Previamente hay que tener en cuenta algunas cuestiones en los estudios realizados por los autores nombrados:

- Los métodos de estimación empleados para la obtención de los parámetros de las distribuciones propuestas, no coinciden en todos los casos. Así Salem y Mount y Rojo Olivas utilizan para la obtención de los estimadores de los parámetros de la distribución Gamma el método de máxima-verosimilitud, Dagum respeta los estimadores obtenidos por Salem y Mount para los casos de la distribución Log-Normal y Gamma (mediante máxima-verosimilitud), así mismo trabaja con los que dedujeron Singh y Maddala (utilizando el método no lineal de mínimos cuadrados y aplicando el algoritmo de David-son-Fletcher-Powell) y utiliza para estimar los parámetros del modelo por él propuesto el método no lineal de mínimos cuadrados y el algoritmo de Birta<sup>1</sup>.
- Los test de adherencia utilizados para la validación de los modelos propuestos en los estudios referidos son básicamente dos: el test de Kolmogorov-Smirnov y la verificación de las cotas de Gastwirth en el caso de que los datos muestrales que se posean sean agregados.

---

<sup>1</sup> Birta, L.G. (1978): “OPTPAK, A Program Package for Unconstrained Function Minimization”. Technical Report TR76-02. University of Ottawa.

La utilización del segundo queda ampliamente justificada en el trabajo desarrollado por Gastwirth y Smith según los cuales “de cualquier distribución aplicada cuyo índice teórico de Gini cae fuera de los límites por una cantidad significativa se debe declarar que se ajusta inadecuadamente a los datos”<sup>2</sup>. La idea subyacente está basada en que si los datos con los que se trabaja son agregados el cálculo del índice de Gini de la muestra completa no es posible; sin embargo deberá encontrarse en el intervalo definido por las cotas propuestas por Gastwirth, obtenidas teniendo en cuenta los datos de los que se dispone.

Así Salem y Mount<sup>3</sup> al decantarse por el modelo Gamma, tropiezan con el inconveniente que supone el desconocimiento de la media geométrica (necesaria para la estimación de los parámetros del modelo) en el caso de que los datos de los que se dispongan sean agregados. Esto se refleja en el test de bondad de ajuste según el representante del intervalo considerado sea uno u otro. Dos son las propuestas realizadas; por un lado considerar como representante del intervalo la marca de clase y por otro la media aritmética de dicho intervalo. En el primer caso Salem y Mount llegan a la conclusión de que la distribución estimada verifica las cotas de Gastwirth, no cumpliéndose esto bajo el segundo supuesto cuando la muestra está agrupada en diez intervalos.

Por otra parte Dagum<sup>4</sup>, que aunque en un principio incluye en su análisis el modelo Pareto desestimándolo más tarde por considerarlo eficaz únicamente cuando se estudian grupos de rentas altas, realiza el estudio proponiendo las distribuciones Log-Normal, Gamma, Singh-Maddala, Dagum-tipo I, Dagum-tipo II, a la renta familiar de los EE.UU. en 1969 (con datos agregados en 10 intervalos), llegando a la conclusión de que aunque todas ellas verifican el test que se deduce del cumplimiento de las cotas de Gastwirth, tan solo las de Dagum-tipo I y Dagum-tipo II soportan el test de Kolmogorov-Smirnov, mientras que tanto el modelo Log-Normal como el Gamma y el Singh-Maddala se ajustan inadecuadamente.

---

<sup>2</sup> Gastwirth, J.L.; Smith, J.T. (1972): “A New Goodness of Fit Test”. *Proceeding of American Statistical Association, Business and Economic Statistics*. pág. 320.

<sup>3</sup> Salem, A.B.; Mount, T.D. (1974): “A Convenient Descriptive Model of Income Distribution: The Gamma Density”. *Econometrica*. Vol 42. N° 6.

<sup>4</sup> Dagum, C. (1991): “Renta y distribución de la riqueza, desigualdad y pobreza: teoría, modelos y aplicaciones”. Cuaderno 22. Seminario Internacional de Estadística en Euskadi.

En este contexto Rojo Olivas propone la distribución Gamma y demuestra, en su tesis doctoral, que tomando un representante adecuado del intervalo para obtener la estimación de la media geométrica el modelo propuesto por Salem y Mount no solo verifica el test que se deduce del cumplimiento de las cotas de Gastwirth sino que soporta además el de Kolmogorov-Smirnov. Los datos de los que parte corresponden a la tabla formada por los valores medios y los extremos de las decilas (excepto la última para el que, mediante una función que relaciona los valores medios y los extremos superiores de las decilas novena y décima, obtiene una estimación) de los Ingresos de los hogares aportados por la Encuesta de Presupuestos Familiares (E.P.F.) correspondientes al período 1980-81. Recogiendo el problema suscitado por el desconocimiento de la media geométrica, ya denunciado por Salem y Mount, considera como representante de cada intervalo decil una combinación lineal convexa del extremo inferior de la decila y el valor medio de la misma:

$$\text{representante del intervalo} = x_{ri} = x_{li} + k(x_{mi} - x_{li})^5$$

donde la determinación del valor del coeficiente K se realiza con el objetivo de obtener una estimación de la media geométrica que permita el cumplimiento de las cotas de Gastwirth así como del test de Kolmogorov-Smirnov.

Los resultados a los que llega son francamente buenos, ya que exceptuando Segovia y la Comunidad Valenciana, todas las provincias y comunidades soportan ambos test de ajuste.

Una vez expuestos los antecedentes, el presente estudio pretende, al estar en disposición de los datos desagregados de Ingresos y Gastos correspondientes a las E.P.F. de 1980-81 y 1990-91, por un lado, como ya se ha indicado anteriormente, validar el modelo Gamma y por otro, teniendo en cuenta que un núcleo común a los trabajos presentados radica en la utilización de datos agregados, normalmente en decilas, contrastar la validez de los estudios realizados hasta la década de los 80 y principio de los 90 en los que se utilizaba únicamente la información agrupada en decilas.

---

<sup>5</sup> Rojo Olivas, C. (1993): “Análisis de la distribución de la renta en España”. Tesis Doctoral. Universidad de Valencia. pág. 89.

## 2. METODOLOGÍA.

El modelo teórico que se pretende validar y que, por tanto, se usará a lo largo del análisis es el Gamma biparamétrico, con función de densidad:

$$f(x) = \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)} \quad \forall x \in ]0, \infty[, \alpha > 0, \lambda > 0$$

Los datos (proporcionados por el I.N.E.) con los que se trabajará corresponden a los Ingresos y Gastos de los hogares recogidos en las E.P.F., en dos períodos de tiempo: 1980-81 y 1990-91, presentando la siguiente estructura

DISEÑO DEL FICHERO DE LA ENCUESTA DE PRESUPUESTOS FAMILIARES 1980/81 PROPORCIONADO POR EL I.N.E.	
POSICIONES	CONTENIDO
1-7	NÚMERO DE ORDEN
8-9	PROVINCIA
10-17	GASTO TOTAL
18-25	INGRESO TOTAL
26-46	FACTOR DE ELEVACIÓN

DISEÑO DEL FICHERO DE LA ENCUESTA DE PRESUPUESTOS FAMILIARES 1990/91 PROPORCIONADO POR EL I.N.E.	
POSICIONES	CONTENIDO
1-5	NÚMERO DE IDENTIFICACIÓN DEL HOGAR
6-7	PROVINCIA
8-15	GASTO TOTAL DEL HOGAR (se incluyen: autoconsumo, autosuministro, salario en especie y alquileres imputados. para los bienes procedentes del salario en especie se considera como gasto total el valor en el mercado del bien consumido. Para la obtención del gasto total no se han tenido en cuenta los códigos de recogida en campo 126510 y 902536, por no considerarse los conceptos que en ellos se registran, gastos de consumo)
16-23	TOTAL DE INGRESOS DEL HOGAR (MONETARIOS Y NO MONETARIOS)
24-34	FACTOR DE ELEVACIÓN POBLACIONAL

En cuanto a los métodos de estimación y los test de adherencia seleccionados, se opta por los mismos que fueron utilizados por los autores citados en el epígrafe anterior, es decir:

**1.-** El método que se utilizará para obtener los estimadores de los parámetros  $\alpha$  y  $\lambda$  de una distribución Gamma será el de máxima-verosimilitud, que conduce al sistema:

$$\hat{\lambda} = \frac{n\hat{\alpha}}{\sum_{i=1}^n X_i} = \frac{n\hat{\alpha}}{n\bar{X}} = \frac{\hat{\alpha}}{\bar{X}}$$

$$\ln(\hat{\alpha}) - \Psi(\hat{\alpha}) = \ln\left(\frac{\bar{X}}{\tilde{X}}\right)$$

siendo  $\Psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$ , conocida como función digamma, y  $\bar{X}$  y  $\tilde{X}$  las media aritmética y geométrica respectivamente.

Por tanto, las expresiones de los estimadores máximo-verosímiles de  $\alpha$  y  $\lambda$  no son deducibles de forma analítica, sino que es necesaria la utilización de técnicas de cálculo numérico. En este caso concreto se utilizarán de forma combinada los métodos de Newton-Raphson y Gauss-Laguerre.

**2.-** El test de bondad de ajuste que se utilizará será el de Kolmogorov - Smirnov que, basado en el teorema de Glivenko - Cantelli o teorema fundamental de la Estadística Matemática según el cual la función de distribución empírica converge a la función de distribución teórica con probabilidad 1 o casi seguramente, utiliza como variable para realizar el contraste:

$$D = \max_{i \in \{1, 2, \dots, n\}} \left\{ \left| F_{0i} - F_{Ti} \right| \right\}$$

siendo  $n$  el tamaño muestral,  $F_{0i}$  y  $F_{Ti}$  las funciones de distribución empírica y teórica, respectivamente, correspondiente al valor  $i$ -ésimo de la muestra. Los valores críticos,  $D_\alpha$ , para esta variable están tabulados, según el tamaño muestral y el nivel de significación  $\alpha$ . En general, si  $n > 100$  estos pueden obtenerse atendiendo a la expresión:

$$D_{\alpha} = \sqrt{\frac{-\ln(\alpha / 2)}{2n}}$$

A este test se incorporará el que se deduce de la “verificación de las cotas de Gastwirth” cuando se trabaje con los datos agregados en decilas.

### **3. AJUSTE DEL MODELO $\Gamma(\alpha, \lambda)$ , A NIVEL PROVINCIAL, CON DATOS DESAGREGADOS.**

En este epígrafe se va a proceder a analizar la bondad de ajuste del modelo propuesto considerando los datos desagregados de cada una de las provincias que componen la comunidad de Castilla-León.

En cualquier caso hay que indicar que dado que la información disponible es toda la muestra, es posible un conjunto de opciones de las que se carecía en anteriores trabajos. Así, una de las implicaciones importantes del trabajo con datos desagregados, cuando se propone el modelo Gamma como explicativo de la distribución de renta, es poder obtener la media geométrica muestral necesaria para la estimación de los parámetros  $\alpha$  y  $\lambda$ .

Para ello se hará uso de los factores de elevación, teniendo en cuenta que a partir de ellos se obtienen los factores de ponderación para cada dato muestral, de forma que si  $D_i$  es un dato muestral (ingreso o gasto) con factor de elevación  $F_i$ , esto se traduce en que dicho dato representa a  $F_i$  hogares, por lo que:

$$\frac{F_i}{F_T} \quad \text{con} \quad F_T = \sum_i F_i$$

será el factor de ponderación de  $D_i$ .

Una vez fijadas las bases del análisis que pretende realizarse con la información que se posee, se obtienen las estimaciones de los parámetros  $\alpha$  y  $\lambda$  de la distribución gamma para cada una de las provincias, tanto para la variable Ingreso como para la variable Gasto

durante los dos períodos de tiempo considerados. Dada la ingente cantidad de datos con los que se está trabajando se han utilizado un programa ad-hoc (programa KOLMOP02) que facilita la entrada de ficheros y permite al mismo tiempo obtener otros resultados que pudieran ser interesantes. Las tablas que aparecen a continuación reflejan el resultado del test de bondad de ajuste, siguiendo la siguiente estructura:

1ª Columna: nombre de la provincia.

2ª Columna: nivel de significación, para el que no se rechaza, el contraste de Kolmogorov-Smirnov.

3ª Columna: valor que toma el estadístico del test utilizado.

4ª Columna: valor del parámetro  $\alpha$  estimado de la Gamma correspondiente.

5ª Columna: valor del parámetro  $\lambda$  estimado de la Gamma correspondiente.

Tabla 3.1: Test Kolmogorov-Smirnov. Provincias. INGRESOS 80-81.

PROVINCIAS	NIVEL DE SIGNIFICACIÓN	VALOR DEL ESTADÍSTICO D	$\alpha$	$\lambda$
ÁVILA	11%	0.06755	2.16683	0.0000038442172
BURGOS	>20%	0.03154	2.60552	0.0000032018520
LEÓN	>20%	0.04268	2.50140	0.0000039906658
PALENCIA	>20%	0.02788	2.55693	0.0000032153666
SALAMANCA	13%	0.05740	2.45324	0.0000043386899
SEGOVIA	2%	0.07990	2.67384	0.0000039192582
SORIA	>20%	0.04990	2.35306	0.0000038406306
VALLADOLID	>20%	0.05153	3.43150	0.0000039702786
ZAMORA	0.5%	0.09246	2.44914	0.0000047138128

Fuente: Elaboración propia.

Tabla 3.2: Test Kolmogorov-Smirnov. Provincias. INGRESOS 90-91.

PROVINCIAS	NIVEL DE SIGNIFICACIÓN	VALOR DEL ESTADÍSTICO D	$\alpha$	$\lambda$
ÁVILA	17%	0.06115	2.97575	0.0000019673976
BURGOS	5%	0.07356	2.73593	0.0000011686711
LEÓN	10%	0.05625	2.72073	0.0000012194253
PALENCIA	>20%	0.04285	3.38087	0.0000016087017
SALAMANCA	14%	0.06438	2.58074	0.0000016570800
SEGOVIA	>20%	0.04636	3.12081	0.0000015574083
SORIA	>20%	0.05024	3.04037	0.0000014057098
VALLADOLID	13%	0.06060	2.87461	0.0000012756476
ZAMORA	7%	0.07063	2.86898	0.0000016344807

Fuente: Elaboración propia.

Tabla 3.3: Test Kolmogorov-Smirnov. Provincias. GASTOS 80-81.

PROVINCIAS	NIVEL DE SIGNIFICACIÓN	VALOR DEL ESTADÍSTICO D	$\alpha$	$\lambda$
ÁVILA	17%	0.06171	1.96506	0.0000031210795
BURGOS	>20%	0.03474	2.03759	0.0000024738365



LEÓN	>20%	0.03919	2.06073	0.0000026319171
PALENCIA	>20%	0.03281	2.05717	0.0000021486585
SALAMANCA	8%	0.06217	1.80090	0.0000026990585
SEGOVIA	>20%	0.05353	2.27606	0.0000026735361
SORIA	>20%	0.06163	2.15271	0.0000029383907
VALLADOLID	>20%	0.03725	3.14288	0.0000034379414
ZAMORA	1%	0.08514	2.11801	0.0000036171116

Fuente: Elaboración propia.

Tabla 3.4: Test Kolmogorov-Smirnov. Provincias. GASTOS 90-91.

PROVINCIAS	NIVEL DE SIGNIFICACIÓN	VALOR DEL ESTADÍSTICO D	$\alpha$	$\lambda$
ÁVILA	>20%	0.05707	2.03310	0.0000010335480
BURGOS	>20%	0.03404	2.29387	0.0000009187651
LEÓN	>20%	0.03648	2.65031	0.0000010983185
PALENCIA	>20%	0.03150	2.79.849	0.0000012949133
SALAMANCA	>20%	0.05626	2.05034	0.0000012264686
SEGOVIA	>20%	0.03156	2.54954	0.0000012731001
SORIA	>20%	0.04309	2.65697	0.0000011017920
VALLADOLID	>20%	0.05180	2.63882	0.0000011119479
ZAMORA	>20%	0.05001	2.24984	0.0000011765936

Fuente: Elaboración propia.

De su observación se deduce que:

1º Una provincia (concretamente Zamora) no supera el test de adherencia con respecto a la variable Ingreso en el 80-81.

2º Todas las provincias superan dicho test con respecto a la variable Ingresos en el período 90-91 y para Gastos, tanto en el 80-81 como en el 90-91.

En términos porcentuales, y considerando globalmente los 36 análisis realizados, esto supone un 2.7% de no ajuste a nivel provincial, aunque la situación es diferente si se analiza el porcentaje teniendo en cuenta el año en que fue realizada la encuesta y la variable considerada, así el 11% de las provincias no verifican el test para los Ingresos en el 80-81, mientras que el 100% lo superan para la misma variable en el 90-91 y para la característica Gastos en ambos períodos. De lo que se deduce que el modelo Gamma puede aceptarse para las densidades de las dos variables con las que se ha trabajado.

#### 4. AJUSTE DEL MODELO $\Gamma(\alpha, \lambda)$ , A NIVEL PROVINCIAL, CON DATOS AGREGADOS.

En lo que respecta al segundo de los objetivos propuestos en el presente estudio, recuérdese que la idea que suscita el mismo es poder conocer si con los datos agregados en decilas y teniendo conocimiento únicamente de los valores extremos de las mismas y valores medios de la variable considerada (en este caso Ingresos o Gastos) en cada uno de ellos, se sigue aceptando el ajuste a la distribución Gamma, dado que han sido estos los datos con los que más frecuentemente se ha estado trabajando en la década de los 80 y principio de los 90. Lógicamente, los casos más interesantes serán aquellos en los que no se acepta el modelo Gamma al utilizar la muestra completa.

La forma de proceder para extraer alguna conclusión será la siguiente: considerar aquellas provincias en las que se ha aceptado el ajuste para Ingresos y Gastos en ambas encuestas, y repetir el análisis con los datos agregados en decilas, teniendo como conocimiento únicamente el extremo de las diez decilas así como sus valores medios y, estimando el representante de los diez intervalos según la propuesta de Rojo Olivas, obtener la media geométrica necesaria para la estimación de los parámetros  $\alpha$  y  $\lambda$  de una Gamma.

A partir de estos datos se utiliza como test de bondad de ajuste el de Kolmogorov - Smirnov, incluyéndose en este caso un test de ajuste adicional, al trabajar con datos agregados, consistente en la verificación de las cotas de Gastwirth.

En las tablas siguientes se exponen los resultados obtenidos, en Ingresos y Gastos (80-81 y 90-91) al agregar los datos en decilas (para las provincias en las que se aceptaba el ajuste a una Gamma con datos originales) con la siguiente estructura:

1ª Columna: nombre de la provincia.

2ª Columna: intervalo formado por los valores de  $k$  para los que se acepta el ajuste por el test de Kolmogorov - Smirnov, verificándose al mismo tiempo las cotas de Gastwirth.

3ª Columna: valor de  $k$ , perteneciente al intervalo anterior, para el que se produce el mejor ajuste.

4ª Columna: nivel de significación del contraste.

5ª Columna: cota inferior de Gastwirth.

6ª Columna: valor del índice de Gini.

7ª Columna: cota superior de Gastwirth.

8ª Columna: valor de  $\alpha$  de la Gamma ajustada.

9ª Columna: valor de  $\lambda$  de la Gamma ajustada.

Tabla 4.1: Test Kolmogorov-Smirnov. Provincias. INGRESOS 80-81.

PROVINCIAS	INTERVALO	K	N.S.	C.I.G.	I. GINI	C.S.G.	$\alpha$	$\lambda$
ÁVILA	[0.76, 0.86]	0.86	8%	0.36724	0.36802	0.38225	2.06873	0.000003670194
BURGOS	[0.91, 0.96]	0.96	>20%	0.33035	0.33129	0.33987	2.61583	0.000003214532
LEÓN	[0.85, 0.92]	0.92	>20%	0.33872	0.33910	0.34968	2.48470	0.000003964044
PALENCIA	[0.94, 0.99]	0.99	>20%	0.32893	0.32931	0.33767	2.65050	0.000003333041
SALAMANCA	[0.85, 0.91]	0.85	>20%	0.34494	0.35485	0.35618	2.24616	0.000003972481
SEGOVIA	[0.80, 0.87]	0.80	5%	0.33872	0.35078	0.35126	2.30481	0.000003378356
SORIA	[0.83, 0.89]	0.88	>20%	0.35184	0.35425	0.36232	2.25466	0.000003680032
VALLADOLID	[0.96, 0.99]	0.99	>20%	0.28172	0.28484	0.29016	3.63038	0.000004200394

Fuente: Elaboración propia.

Tabla 4.2: Test Kolmogorov-Smirnov. Provincias. INGRESOS 90-91.

PROVINCIAS	INTERVALO	K	N.S.	C.I.G.	I. GINI	C.S.G.	$\alpha$	$\lambda$
ÁVILA	[0.86,0.91]	0.86	>20%	0.31954	0.32814	0.32862	2.67130	0.000001766123
BURGOS	[0.86,0.92]	0.91	10%	0.32967	0.33188	0.33985	2.60554	0.000001112980
LEÓN	[0.90,0.95]	0.95	>20%	0.30186	0.30219	0.31095	3.19634	0.000001432594
PALENCIA	[0.90,0.94]	0.90	>20%	0.29821	0.30625	0.30648	3.10540	0.000001477628
SALAMANCA	[0.81,0.87]	0.86	>20%	0.34508	0.34735	0.35545	2.35584	0.000001512675
SEGOVIA	[0.88,0.92]	0.88	>20%	0.31202	0.32001	0.32015	2.82210	0.000001408343
SORIA	[0.87,0.91]	0.87	>20%	0.31720	0.32469	0.32623	2.73381	0.000001263975
VALLADOLID	[0.89,0.94]	0.89	>20%	0.32085	0.33009	0.33014	2.63671	0.000001170082
ZAMORA	[0.84,0.90]	0.84	>20%	0.32699	0.33693	0.33698	2.52025	0.000001435806

Fuente: Elaboración propia.

Tabla 4.3: Test Kolmogorov-Smirnov. Provincias. GASTOS 80-81.

PROVINCIAS	INTERVALO	K	N.S.	C.I.G.	I. GINI	C.S.G.	$\alpha$	$\lambda$
ÁVILA	[0.81,0.91]	0.91	>20%	0.37757	0.39032	0.39122	1.95045	0.000003097888
BURGOS	[0.86,0.94]	0.86	>20%	0.37085	0.38127	0.38154	1.90734	0.000002315713
LEÓN	[0.81,0.90]	0.89	>20%	0.36910	0.38121	0.38131	2.03085	0.000002593759
PALENCIA	[0.85,0.93]	0.85	>20%	0.36934	0.38006	0.38083	1.92137	0.000002006815
SALAMANCA	[0.70,0.80]	0.72	>20%	0.40731	0.41915	0.41996	1.54038	0.000002308614
SEGOVIA	[0.82,0.90]	0.82	>20%	0.35682	0.36831	0.36888	2.06509	0.000002425729
SORIA	[0.80,0.88]	0.80	>20%	0.37004	0.38075	0.38188	1.91331	0.000002611613
VALLADOLID	[0.96,0.99]	0.99	>20%	0.29465	0.30303	0.30315	3.29269	0.000003601818
ZAMORA	[0.76,0.83]	0.78	>20%	0.37871	0.38833	0.38949	1.85413	0.000003166474

Fuente: Elaboración propia.

Tabla 4.4: Test Kolmogorov-Smirnov. Provincias. GASTOS 90-91.

PROVINCIAS	INTERVALO	K	N.S.	C.I.G.	I. GINI	C.S.G.	$\alpha$	$\lambda$
ÁVILA	[0.80,0.87]	0.80	>20%	0.38135	0.39119	0.39137	1.79594	0.000000912986
BURGOS	[0.85,0.92]	0.85	>20%	0.35541	0.36493	0.36529	2.10878	0.000000844631

LEÓN	[0.89,0.94]	0.89	>20%	0.33284	0.34122	0.34160	2.45078	0.000001015633
PALENCIA	[0.91,0.96]	0.96	>20%	0.31851	0.32736	0.32774	2.82944	0.000001309232
SALAMANCA	[0.79,0.86]	0.79	>20%	0.37977	0.38915	0.38924	1.81814	0.000001087571
SEGOVIA	[0.92,0.97]	0.97	>20%	0.33308	0.34059	0.34189	2.58343	0.000001290022
SORIA	[0.91,0.95]	0.91	>20%	0.33159	0.33842	0.33911	2.49575	0.000001034941
VALLADOLID	[0.92,0.97]	0.92	>20%	0.32580	0.33536	0.33581	2.54642	0.000001073016
ZAMORA	[0.82,0.89]	0.82	>20%	0.36391	0.37345	0.37401	2.00073	0.000001046319

Fuente: Elaboración propia.

De su observación se deduce que se ha aceptado el ajuste a la distribución Gamma en todas aquellas provincias donde se obtenía este resultado con todos los datos desagregados.

Por otra parte cabe preguntarse, hasta qué punto la información a nivel de decilas aportada por el I.N.E. es lo suficientemente completa como para que la duplicación de resultados sea total, es decir ¿se seguirá rechazando el ajuste al modelo Gamma en aquellos casos donde este era la conclusión a la que se llegaba, con la muestra exhaustiva de la que ahora se dispone?. Si la respuesta fuera afirmativa, se produciría una validación de los resultados de estudios realizados por otros autores que dispusieran únicamente de los datos agregados en decilas. Si por el contrario aquellas provincias que no ajustaban, considerando todos los datos muestrales, sí lo hacen al tener como información solo las decilas muestrales esto abriría una vía de reflexión a tener en cuenta y que en cualquier caso no va a abordarse en este análisis, por escapar a los objetivos del mismo.

Así, realizando el estudio para dichos casos, teniendo en cuenta que en el presente trabajo tan sólo dejaba de superar el test la provincia de Zamora para Ingresos en 1980-81, el resultado ha sido el siguiente:

Tabla 4.5: Test Kolmogorov-Smirnov. Provincia que no ajustaba. INGRESOS 80-81.

PROVINCIA	INTERVALO	K	N.S.	C.I.G.	I. GINI	C.S.G.	$\alpha$	$\lambda$
ZAMORA	[0.78,0.87]	0.78	>1%	0.34947	0.36252	0.36268	2.14055	0.000004119886

Fuente: Elaboración propia.

De su observación se concluye que, en el caso del estudio de la provincia que no ajustaba con datos desagregados esta supera el test de Kolmogorov - Smirnov y se verifica las cotas de Gastwirth.

En forma resumida, dado que las diferentes situaciones que podrían darse eran:

Cuadro 1: Ajuste según muestra completa o decilas.

	TRABAJANDO CON DECILAS LA PROVINCIA AJUSTA	TRABAJANDO CON DECILAS LA PROVINCIA NO AJUSTA
TRABAJANDO CON LA MUESTRA COMPLETA LA PROVINCIA AJUSTA	SITUACIÓN 1	SITUACIÓN 2
TRABAJANDO CON LA MUESTRA COMPLETA LA PROVINCIA NO AJUSTA	SITUACIÓN 3	SITUACIÓN 4

Fuente: Elaboración propia.

Donde las únicas que implicaban algún tipo de “alerta” eran la segunda y la tercera situación, se ha observado que dentro de esta última se encuentra la provincias de Zamora (Ingresos 80-81).

Por tanto la replicación no se produce, ya que en el caso del trabajo con decilas “todas” las provincias que componen la comunidad de Castilla-León ajustan al modelo Gamma, cosa que no sucedía al trabajar con la muestra completa.

## 5. CONCLUSIONES.

Teniendo en cuenta los dos objetivos marcados en el presente trabajo (validación del modelo Gamma biparamétrico, para las densidades de los Ingresos y Gastos familiares, y contrastar la validez de estudios realizados hasta la década de los 80 y principio de los 90, en los que se hacía uso de la información agrupada en decilas) se concluye que:

- I) El primero de los objetivos se cubre, es decir, dado que de los 36 análisis realizados tan solo en uno de ellos se rechaza la distribución propuesta, lo que se traduce en un 97.3% de ajuste, al trabajar con datos desagregados, el modelo  $\Gamma(\alpha, \lambda)$  puede aceptarse para las densidades de las dos variables consideradas.
- II) Del estudio comparativo datos agregados - datos desagregados, en cuanto a ajuste, se deduce que los resultados obtenidos no coinciden para ambos niveles informativos. Efectivamente, cuando se utiliza toda la muestra (datos desagregados) la provincia de Zamora no soporta el test de Kolmogorov-Smirnov en el caso de Ingresos 1980-81, mientras que no solamente verifica este test cuando se utiliza el otro tipo

de información muestral (datos agregados en decilas), sino que además supera el test que se deduce de la verificación de las cotas de Gastwirth, contraste aconsejable, como se explicitó en epígrafes anteriores, cuando se trabaja con datos agregados. Este hecho abre una vía de reflexión, en tanto que, los resultados de los estudios en los que se utiliza únicamente la información agrupada en decilas deberán tomarse, como mínimo, con cierta cautela.

## **6. BIBLIOGRAFÍA.**

Birta, L.G. (1978): "OPTPAK, A Program Package for Unconstrained Function Minimization". Technical Report TR76-02. University of Ottawa.

Dagum, C. (1991): "Renta y distribución de la riqueza, desigualdad y pobreza: teoría, modelos y aplicaciones". Cuaderno 22. Seminario Internacional de Estadística en Euskadi.

Gastwirth, J.L.; Smith, J.T. (1972): "A New Goodness of Fit Test". Proceeding of American Statistical Association, Business and Economic Statistics. pág. 320.

Rojo Olivas, C. (1993): "Análisis de la Distribución de la Renta en España". Tesis Doctoral. Universidad de Valencia.

Salem, A.B.; Mount, T.D. (1974): "A Convenient Descriptive Model of Income Distribution: The Gamma Density". *Econometrica*. Vol. 42. Nº 6.