

**La modelización paramétrica
de las distribuciones salariales.
Un estudio aplicado al caso español**

Carmelo García

Universidad de Alcalá

Mercedes Prieto

Universidad de Valladolid

Hipólito Simón

Universidad de Alicante-IEI-IEB

Resumen

En este artículo se modeliza paramétricamente la distribución salarial en España a partir de los microdatos de la *Encuesta de Estructura Salarial*. La novedosa evidencia obtenida corrobora que, de forma similar a otros ámbitos de análisis en la literatura económica, estas técnicas de modelización son también apropiadas y útiles en el caso de las distribuciones salariales. Entre los principales hallazgos destaca asimismo el hecho de que la distribución beta generalizada de segunda especie es el modelo teórico que mejor se ajusta por lo general a la distribución salarial empírica de la economía española.

Palabras clave: Distribución salarial, modelización paramétrica, beta generalizada de segunda especie, España.

Clasificación JEL: J31, D31, C46.

Direcciones:

Carmelo García (carmelo.garcia@uah.es)
Facultad de Ciencias Económicas y Empresariales
Plaza de la Victoria, 8
28802 Alcalá de Henares

Mercedes Prieto (prietoal@eae.uva.es)
Facultad de Ciencias Económicas y Empresariales
Universidad de Valladolid
Avda. Valle Esgueva, 6
47011 Valladolid

Hipólito Simón (hsimon@ua.es)
Aulario I
Universidad de Alicante
03690 San Vicente del Raspeig

1. Introducción

La modelización paramétrica de las distribuciones empíricas de variables económicas presenta numerosas ventajas analíticas en el ámbito del estudio económico y causal de dichas distribuciones. Por este motivo, estas técnicas han sido ampliamente utilizadas en el análisis empírico de las distribuciones de variables como la riqueza o, muy especialmente, la renta (véase, por ejemplo, Chotikapanich, 2008 y Cowell, 2002).

La principal ventaja de este enfoque es que toda la información contenida en la distribución empírica se puede condensar en un número limitado de parámetros (véase Maasoumi, 1994). De esta manera, cualquier medida descriptiva de interés, como pueden ser la media, los cuantiles, indicadores de desigualdad o de polarización, se pueden expresar en función de este reducido número de parámetros. Además, se da la circunstancia de que los parámetros pueden servir de indicadores de distintos aspectos de la distribución no captados por las medidas e índices habituales (por ejemplo, sobre el comportamiento de las colas), de modo que su análisis arroja significado estadístico y económico. Así, por ejemplo, mediante el estudio de la variación de los parámetros en el tiempo y en el espacio, pueden explicarse los cambios que experimenta el proceso distributivo (véase García y Prieto, 2011, para la renta en el caso español). Alternativamente, la introducción de los parámetros en modelos econométricos permite construir sistemas de ecuaciones a partir de los cuales se pueden simular los efectos de determinadas variables de tipo macroeconómico o demográfico sobre la distribución, actuando los parámetros como variables dependientes de las mencionadas ecuaciones. Estos modelos econométricos no presentan los problemas de estimación de los modelos habituales basados en regresiones en las que la variable independiente suele ser el índice de Gini u otros índices de desigualdad (Jäntti y Jenkins, 2010).

La modelización paramétrica de la distribución permite también la posibilidad de reconstruir la distribución de la variable de interés en su totalidad, utilizando únicamente los valores de determinados indicadores obtenidos de los datos. Así, si se da la circunstancia de que la distribución teórica modeliza aceptablemente la distribución empírica, se pueden generar distribuciones completas (o una parte de las mismas: colas, muestras truncadas por límites legales, etc.), sin disponer de los microdatos. Como muestra en este sentido, Graf y Nedyalkova (2010) son capaces de reconstruir, utilizando la distribución gamma generalizada de segunda especie, las distribuciones completas de rentas para diversos países europeos exclusivamente a partir de un número reducido de indicadores, entre los que se incluye el porcentaje de pobres o el índice de Gini.

Esta técnica ofrece finalmente también ciertas ventajas frente al uso alternativo de métodos no paramétricos como la estimación *kernel*. Entre las mismas destaca la posibilidad de aportar justificaciones teóricas de la generación de la distribución y la posibilidad de corregir los problemas habituales de falta de información en las encuestas sobre las colas de la distribución -motivada por la ocultación de salarios y rentas altas y las dificultades de acceso a los receptores de rentas y salarios muy bajos - (Pinkovsky y Sala-i-Martin, 2009).

El objetivo de este artículo es modelizar mediante técnicas paramétricas la distribución de los salarios en España. Esta aproximación es en buena medida novedosa, en la medida en que las estimaciones no paramétricas de las funciones de densidad constituyen en la actualidad el método claramente predominante en el análisis empírico de las distribuciones salariales (véase, por ejemplo, Fortin et al., 2011). Además, en acusado contraste con los estudios de renta, en el ámbito de la modelización paramétrica de distribuciones salariales las aportaciones existentes en

la literatura económica son mucho más limitadas. Entre las mismas pueden destacarse los sistemas de estructuras jerárquicas propuestos por Roy (1950) y Lydall (1959, 1968) y los trabajos de Creedy, Lye y Martin (1996), Sattinger (1996) y Parker (1996, 1999) sobre modelos de mercados de trabajo que generan situaciones de equilibrio en forma de distribuciones paramétricas. A su vez, con un carácter más reciente, cabe mencionar también ciertos análisis sobre aspectos concretos como el uso de mixturas de distribuciones para el análisis de distribuciones salariales o la vigencia de la distribución exponencial para la modelización de las colas (véase Schneider, 2008).

En el artículo se traslada al estudio de los salarios un esquema metodológico de reciente aplicación en el terreno de las distribuciones personales de renta (véanse Dastrup et al. 2007 y Prieto y García, 2009). Dicha metodología, que se utiliza por primera vez aplicada a las distribuciones salariales, pone en valor la distribución beta generalizada de segunda especie (GB2) y los modelos anidados en la misma, entre los que se encuentran los mejores y más utilizados para representar distribuciones, tales como los modelos Dagum o Singh-Maddala. Cabe destacar, además, que la distribución beta generalizada de segunda especie produce ajustes muy adecuados a las distribuciones empíricas de rentas tanto en el caso español (Prieto y García, 2009) como en otros trabajos aplicados en el contexto internacional (Dastrup et al., 2007). Esta adecuación ha generado, además, una línea de investigaciones teóricas sobre este modelo que permiten disponer de las expresiones de la mayor parte de los indicadores de desigualdad y de pobreza que se derivan del mismo (McDonald, 1984; Buttler y McDonald, 1989; Wilfling, 1996; Kleiber, 1999; Sarabia et al., 2002 y Jenkins, 2007), así como de algoritmos de computación de los mismos (Jenkins, 2007).

El análisis empírico se desarrolla a partir de los microdatos de la *Encuesta de Estructura Salarial*, elaborada por el INE. Se trata de la fuente de información más completa sobre salarios a nivel macroeconómico que existe para la economía española (lo que se refleja, entre otras circunstancias, en el amplio número de investigaciones desarrolladas a partir de la misma (véanse, por ejemplo, Alcalá y Hernández, 2010 ó Card y De la Rica, 2006). La encuesta se elabora conforme a una metodología común para todos los países de la Unión Europea, por lo que sus resultados son, además, homologables a nivel europeo (pueden encontrarse análisis comparativos internacionales basados en los microdatos de esta encuesta en Simón, 2009 y 2010). El análisis se desarrolla sobre la última ola disponible de la encuesta ya que, a diferencia de las anteriores, incluye a las empresas de tamaño muy reducido, por lo que la misma presenta una cobertura prácticamente exhaustiva del sector privado. Asimismo, es importante destacar también que el análisis empírico se ha desarrollado sobre la muestra completa de la encuesta (lo que supone una excepción reseñable, ya que el acceso de los investigadores a los microdatos de la misma se produce habitualmente sólo para muestras restringidas de la encuesta) y que, dada la riqueza de la información que incluye sobre los salarios y sus distintos componentes, la aproximación a la distribución salarial teórica se ha realizado para diversos conceptos salariales (salario por hora, mensual y anual)¹.

¹ Además de los aspectos señalados con anterioridad, la correcta modelización de la distribución de los salarios en la economía española presenta un importante aspecto práctico de gran utilidad, en la medida en que una de las principales fuentes de información sobre el ámbito del mercado laboral puesta recientemente a disposición de los investigadores, la *Muestra Continua de Vidas Laborales*, elaborada por la Seguridad Social, presenta información sobre salarios (en concreto, las bases de cotización de los afiliados al sistema), pero con un carácter truncado debido a los mínimos y máximos legales. En este sentido, puede encontrarse un intento de tratar con esta limitación a través de distintas aproximaciones, incluyendo el uso de ciertos supuestos distribucionales en Bonhomme y Hospido (2011).

La estructura del artículo es la siguiente. En el segundo apartado se expone la metodología de modelización propuesta. En el tercer apartado se presentan los datos utilizados procedentes de la *Encuesta de Estructura Salarial*. El cuarto apartado se dedica a la exposición de los resultados de los ajustes realizados a los distintos conceptos salariales. El artículo finaliza con un resumen de las principales conclusiones obtenidas en el estudio.

2. Metodología: selección y estimación de modelos

Desde un enfoque paramétrico, la distribución de los salarios puede representarse por un miembro del conjunto de funciones de distribución continuas, perfectamente especificadas, excepto por un vector $(p \times 1)$ de parámetros desconocidos, θ . Es decir, por un miembro de

$$\Psi := \{F_\theta, \theta \in \Theta\}$$

donde $\Theta \subset \mathcal{R}^p$ es el espacio paramétrico y $f(x; \theta)$ es la función de densidad asociada a la función de distribución F_θ .

Antes de comenzar el análisis empírico, en un estudio de modelización paramétrica es preciso acometer la elección, entre todas las funciones paramétricas, de aquellas que cumplen una serie de propiedades deseables según la naturaleza de los datos de rentas o salarios: asimetría a la derecha, fundamento económico del modelo y flexibilidad de su forma funcional para anidar diferentes modelos de interés, entre otras².

En la literatura se encuentran numerosos modelos propuestos para la modelización de variables sujetas a procesos de percepción individual. Entre dichos modelos pueden destacarse las distribuciones de Pareto (Pareto, 1897), la distribución lognormal, la distribución beta biparamétrica (Thurow, 1970), la distribución gamma biparamétrica (Salem y Mount, 1974), la distribución Singh-Maddala (Singh y Maddala, 1976), las distribuciones de Dagum (Dagum, 1977) y la distribución beta generalizada de segunda especie (desde ahora, GB2), introducida por McDonald (1984)³.

Las propuestas de modelos más recientes se refieren a transformaciones y mixturas de las distribuciones existentes previamente. Es el caso de la distribución doble Pareto-lognormal (Reed y Jorgensen, 2005), la distribución log-Dagum (Domma y Perri, 2009) y los modelos de mixturas de distribuciones logarítmico-normales o de mixturas de gammas (Chotikapanich y Griffiths, 2008). Las mismas, no obstante, pierden la simplicidad y las ventajas de aplicación que aporta la utilización de un único modelo, capaz de cubrir el rango completo de la variable.

Para evitar esta pérdida de simplicidad, principal ventaja del enfoque paramétrico, en este artículo el punto de partida será una única distribución dotada de gran flexibilidad: la distribución GB2. Se trata de una distribución tetraparamétrica que, por una parte, produce muy buenos ajustes para las distribuciones de ingresos de algunos países (véanse Brachmann, Stich y Trede, 1996; Dastrup et al., 2007 y Jenkins, 2007) y que, por otra parte, anida diferentes

² Puede verse una relación de propiedades deseables de las funciones paramétricas en Callealta y otros (1996) y Dagum (1997).

³ Las principales distribuciones utilizadas en la modelización paramétrica de la distribución de la renta, entre ellas las contenidas en la GB2, pueden encontrarse estudiadas con detalle en el trabajo de Kleiber y Kotz (2003).

modelos de distribuciones de la renta que también se han mostrado especialmente adecuadas para las rentas en numerosos países (y, en concreto, en el caso español en periodos precedentes al que se analiza en este artículo: Callealta et al., 1996, Prieto y García, 2011). En este sentido, dado que es obvio que las rentas individuales resultan fuertemente influidas por los ingresos salariales (puede encontrarse evidencia al respecto en OCDE, 2011), objeto de análisis en este artículo, no parece inapropiado plantear de partida la comprobación de la adecuación a las distribuciones salariales de los modelos paramétricos utilizados tradicionalmente para distribuciones de renta.

La expresión de la función de densidad de la distribución GB2 es la siguiente:

$$f(x) = \frac{ax^{ap-1}}{b^{ap} B(p, q) (1 + (x/b)^a)^{p+q}}, \quad x \geq 0 \quad (1)$$

donde $a, b, p, q > 0$ y $B(p, q)$ es la función beta.

Más concretamente, b es un parámetro de escala y a, p y q son parámetros de forma. Así, a es un parámetro que controla el grosor de ambas colas y la forma general de la distribución, mientras que p y q determinan, respectivamente, la forma de la cola izquierda y derecha. Estos últimos parámetros son también responsables del grado de asimetría de la distribución del logaritmo de la variable modelizada. Con fines ilustrativos, la influencia de las variaciones de los parámetros sobre la forma de la distribución queda reflejada en el gráfico 1 en el que se representan diferentes formas de la función de densidad ante variaciones de cada uno de los tres parámetros de forma (manteniendo fijos, en cada caso, los valores del resto de los parámetros). En el anexo se presentan también simulaciones de curvas de Lorenz para diferentes valores de los parámetros (gráfico A.1), las cuales permiten apreciar claramente las partes de la distribución sobre las que cada uno de los parámetros tiene mayor influencia (para más detalles, véase Kleiber y Kotz, 2003).

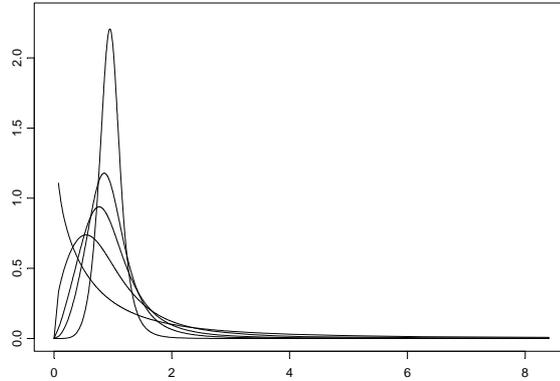
Las distribuciones triparamétricas beta de segunda especie Dagum⁴ y Singh-Maddala corresponden, por su parte, a casos particulares de la distribución GB2 cuando $a=1$, $q=1$ y $p=1$, respectivamente. En el caso de que simultáneamente $q=1$ y $p=1$ se obtiene la distribución Fiskla cual, dentro de las distribuciones biparamétricas, suele presentar excelentes ajustes de las distribuciones de renta (Prieto y García, 2009). La distribución GB2 también anida los modelos Lomax y Lomax inversa. Éstos, no obstante, suelen presentar ajustes deficientes en distribuciones de salarios y de rentas (además de otros inconvenientes adicionales: véase Kleiber y Kotz, 2003). En función de lo anterior, en el análisis empírico se ha optado por realizar los ajustes correspondientes a los modelos GB2, Dagum, Singh-Maddala, beta de segunda especie y Fisk⁵.

⁴ El modelo Dagum anidado por la GB2 es en concreto la distribución Dagum triparamétrica o de tipo I. De aquí en adelante, lo denominaremos simplemente modelo Dagum.

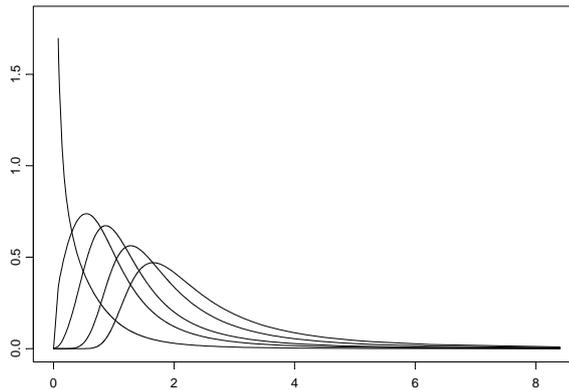
⁵ Con fines comparativos, también se han realizado ajustes para algunos modelos habituales en los estudios de rentas y salarios, como la distribución gamma y logarítmico normal, entre otros. Los resultados han sido, en todos los casos, significativamente peores que los que se presentan para las distribuciones seleccionados en este artículo (los mismos están disponibles por parte de los autores ante su requerimiento).

Gráfico 1. Simulaciones de la función de densidad de la GB2 ante variaciones de los parámetros.

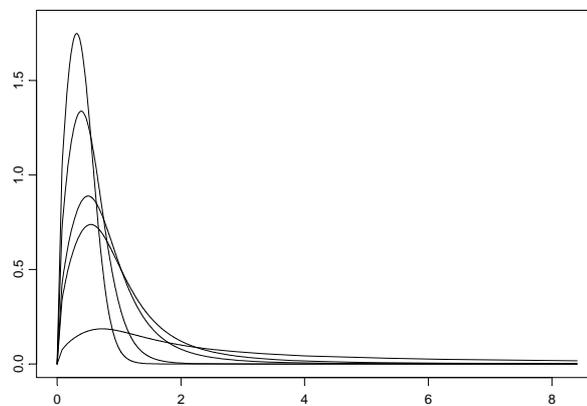
- a) Variaciones del parámetro α sobre la función de densidad de la GB2 ($p=0,5$; $q=0,7$; $\alpha = 1,5, 3, 4,5, 6$ y 12 , de izquierda a derecha).



- b) Variaciones del parámetro p sobre la función de densidad de la GB2 ($\alpha=3$; $q=0,7$; $p=0,1, 0,5, 1, 2,5$ y 5 , de izquierda a derecha).



- c) Variaciones del parámetro q sobre la función de densidad de la GB2 ($\alpha=3$; $p=0,5$; $q=0,1, 0,7, 1, 2,5$ y 5 , de derecha a izquierda).



Una vez seleccionados los modelos teóricos más adecuados, se estiman, en primer lugar, los vectores de parámetros desconocidos mediante el método de máxima verosimilitud, el cual proporciona estimadores que son asintóticamente eficientes (Rao, 1973). Las estimaciones máximo verosímiles se obtienen como solución de un sistema de ecuaciones no lineales (resuelto con el algoritmo de Newton-Raphson implementado en lenguaje C mediante programas originales propios, diseñados específicamente para este fin). La utilización de programas específicos nos permite controlar todo el procedimiento de estimación y seguir exactamente los pasos descritos en la metodología. En los programas de estimación, que se ponen a disposición del lector, los valores iniciales de los parámetros se eligen a partir de las estimaciones por mínimos cuadrados no lineales y a través de una búsqueda reticular con una precisión de una décima.

En segundo lugar, se analiza gráficamente en qué medida los modelos paramétricos ajustados reproducen adecuadamente los datos originales. En este sentido, inicialmente, como primera aproximación, se comprueba si la función de densidad estimada se adecua al histograma. La medida gráfica más potente que permite juzgar la bondad del ajuste es el gráfico de probabilidad (*PP-plot*) que muestra las discrepancias entre todos y cada uno de los valores de la función de distribución empírica (en el eje de abscisas) frente a los valores correspondientes de la función de distribución del modelo teórico (en el eje de ordenadas). Así pues, los gráficos *PP-plots* constituyen una potente medida de bondad del ajuste, ya que reflejan las discrepancias entre la función de distribución teórica y empírica observación a observación, las cuales van dibujando una línea que, cuanto más próxima es a la diagonal del cuadrante, indicará una mejor calidad del ajuste.

Adicionalmente, se usarán diferentes estadísticos que permiten medir la bondad del ajuste de los diferentes modelos teóricos. Se trata, en concreto, del logaritmo de la función de verosimilitud, el coeficiente de correlación entre los valores de la función de distribución empírica y los valores de la función de distribución ajustada y los estadísticos de Kolmogorov-Smirnov y de Cramer-Von Mises (estos últimos muy adecuados en particular para el caso de datos no agrupados, como son los utilizados en este trabajo).

Finalmente, una vez comprobada la bondad de los ajustes, los modelos más adecuados serán seleccionados comparando los valores obtenidos de los estadísticos mencionados. Para comparar entre modelos anidados en la distribución GB2 se realizarán, además, contrastes de hipótesis para seleccionar el mejor modelo entre los anidados. Para ello se comparan las funciones de verosimilitud de la muestra de los diferentes modelos. Así pues, mediante ratios de logaritmos de la función de verosimilitud se detecta si son significativas las mejoras que introducen los modelos no restringidos (con más parámetros) sobre los anidados en ellos. En concreto, para los contrastes de modelos anidados se utiliza el siguiente estadístico de razón de verosimilitudes:

$$2[\ln L - \ln L^*] \xrightarrow{n \rightarrow \infty} \chi_r^2$$

Donde $\ln L$ y $\ln L^*$ son, respectivamente, los logaritmos neperianos de las funciones de verosimilitud de la muestra para el modelo sin restricciones y para el modelo anidado. Bajo la hipótesis nula (es decir, si el modelo restringido es el verdadero), dicho estadístico se distribuye asintóticamente como una distribución χ_r^2 , donde los grados de libertad (r) son la diferencia entre el número de parámetros del modelo sin restricciones y el número de parámetros del modelo anidado (Cox y Hinkley, 1974).

3. Datos

En este apartado del artículo se procede a describir la fuente de información empleada en el análisis empírico. La *Encuesta de Estructura Salarial* (desde ahora, EES) es una encuesta elaborada para el ámbito español por el INE conforme a una metodología armonizada para todos los países de la Unión Europea para la realización de encuestas nacionales que contengan información estrictamente comparable entre países sobre la distribución de los salarios⁶. Se trata de una encuesta que se elabora con periodicidad cuatrienal, existiendo en la actualidad tres olas disponibles, correspondientes concretamente a los años 1995, 2002 y 2006⁷.

El diseño de la EES corresponde a un muestreo en dos etapas de trabajadores a partir de las cuentas de cotización de las empresas en la Seguridad Social. Su cobertura ha ido creciendo con el transcurso del tiempo, siendo éste el motivo básico por el que el análisis empírico se circunscribe a los microdatos de la ola más reciente, la de 2006. Así, a diferencia de las anteriores, esta ola abarca a todos los empleados por cuenta ajena que trabajan en establecimientos de cualquier tamaño (las de 2002 y 1995 no incluían a los establecimientos con menos de 10 trabajadores) y cubre de forma prácticamente exhaustiva el sector privado de la economía española (en concreto, incluye todas las actividades comprendidas en las secciones C a O de la clasificación NACE-93, con la excepción de la sección L - administración pública-, mientras que la ola de 1995 no incluía las secciones M, N y O -educación, sanidad y otras actividades sociales-).

La EES proporciona una amplia información sobre la remuneración y las características de los trabajadores y sus puestos de trabajo (sexo, edad, educación, antigüedad, ocupación, tipo de contrato, tipo de jornada, nacionalidad y si el trabajador realiza tareas de supervisión) y sobre sus establecimientos (sector, tamaño, tipo de convenio colectivo y región)⁸. Además, en la medida en que incluye información detallada sobre los distintos componentes que conforman el salario y para distintas referencias temporales, la misma permite el cálculo de distintas medidas salariales (en concreto el salario por hora, mensual y anual). A efectos de este trabajo, esto permite que la modelización de las distribuciones salariales tenga un carácter más amplio (y, en consecuencia, de mayor interés).

La información salarial contenida en la EES se obtiene para un mes de referencia, octubre, cuya elección por parte del INE se realiza sobre la base de que se trata de un mes representativo del conjunto del año en términos de percepciones salariales. No obstante, la misma abarca componentes salariales de carácter tanto mensual (salario base y complementos salariales percibidos en octubre) como anual (complementos salariales de tipo fijo y variable). Los salarios están expresados en términos brutos y en su cálculo se incluye cualquier tipo de pago por parte de las empresas (incluyendo comisiones, dietas, pluses por trabajo nocturnos y en fines de semana), con la única excepción del pago de horas extraordinarias. Así, el salario mensual incluye entre sus componentes (además del salario base y los componentes salariales percibidos durante el mes de octubre) la parte prorrateada de los pagos extraordinarios de periodicidad no mensual (como vacaciones pagadas, pagas extraordinarias o pluses anuales por incentivos). El salario por hora ha sido

⁶ Puede encontrarse información detallada sobre la metodología común de elaboración de la encuesta en Eurostat (2006) y sobre la *Encuesta de Estructura Salarial* en la página web del INE (www.ine.es).

⁷ El INE elabora también con periodicidad anual la *Encuesta Anual de Estructura Salarial*, la cual proporciona únicamente estimaciones de la ganancia bruta anual por trabajador, clasificada por tipo de jornada, sexo, actividad económica y ocupación. Pueden encontrarse más detalles sobre la misma en Moral-Arce y Martín (2009).

⁸ Pueden encontrarse estadísticos descriptivos de la muestra de la ola de 2006 de la EES en la tabla A.1 del anexo 1.

calculado a partir del salario mensual (en concreto, dividiéndolo por la jornada semanal normal en la empresa en una semana de octubre multiplicada por 4,42)⁹.

Un aspecto adicional a destacar es que el análisis empírico se ha desarrollado a partir de la muestra completa de la *Encuesta de Estructura Salarial*. Esta circunstancia es inusual, en la medida en que, por motivos de confidencialidad, el INE únicamente facilita a los investigadores los microdatos correspondientes a una parte de la muestra de esta encuesta (esto ocurre cuando entre las variables sobre las que se requiere información se encuentran atributos relativos a los establecimientos -lo que suele ser habitual-, ya que se considera que los mismos podrían ser susceptibles de ser utilizados para identificar a establecimientos concretos). Esta muestra de carácter restringido, conformada por menos observaciones, se selecciona por parte del INE mediante criterios desconocidos para los investigadores y, en consecuencia, con un alcance que no se puede valorar. Con el fin de salvar los posibles sesgos asociados a esta circunstancia, se ha realizado una petición ex profeso de microdatos al INE en la que se ha renunciado a obtener cualquier tipo de información sobre los establecimientos, lo que ha permitido el acceso al conjunto de la muestra de la encuesta.

En la tabla A.1 del anexo se desarrolla una comparación de esta muestra (a la cual denominamos muestra *completa*) con una muestra representativa de aquellas a las que suelen tener acceso los investigadores (a la que denominamos muestra *restringida*). Dicha comparación revela que, de forma general, existen en la práctica diferencias muy significativas entre ambas muestras. Entre las mismas destaca el tamaño mucho más elevado de la muestra *completa* (la restringida está formada por 140.321 observaciones, las cuales suponen únicamente el 59,7% de las 234.896 observaciones de la muestra *completa*), así como su distinta composición en función de las características sociodemográficas de los asalariados (con una mayor presencia en términos comparativos de hombres, nativos, individuos con estudios secundarios, con jornada a tiempo completo y con contrato indefinido). Asimismo, en aquella se aprecian valores más elevados tanto en las medidas centrales de la distribución (media y mediana) como en los distintos cuartiles (lo cual sugiere, en consecuencia, que la eliminación de observaciones por parte del INE en la *muestra restringida* no tiene un carácter aleatorio, sino que afecta en mayor medida a establecimientos que pagan salarios comparativamente elevados). Finalmente, también se observan diferencias en los niveles de desigualdad, los cuales tienden a ser sistemáticamente mayores para todas las medidas salariales cuando se calculan a partir de la muestra *completa*. En consecuencia, a la luz de esta evidencia, la posibilidad de desarrollar el análisis empírico a partir de la muestra completa de la *Encuesta de Estructura Salarial* es una circunstancia especialmente relevante.

4. Resultados

4.1 Modelización de las distribuciones del salario por hora y mensual

Tal y como se puede observar en los histogramas¹⁰ de los gráficos 2, 3 y 4, las formas de las distribuciones empíricas del salario por hora y mensual presentan características similares entre sí, pero claramente diferenciadas de la de la distribución de los salarios anuales. Entre las diferencias más notables destaca la

⁹ Hay que señalar que esta forma de calcular el salario por hora no constituye un simple cambio de escala sobre el salario mensual, ya que el factor utilizado no es el mismo para todos los trabajadores.

¹⁰ Cabe destacar que si bien el conjunto del análisis empírico se desarrolla sobre la muestra completa de la encuesta, en los histogramas se ha eliminado un 1% de las observaciones extremas en la cola derecha, con el fin de poder mostrar en detalle las características más relevantes de las distribuciones.

bimodalidad patente de esta última distribución. Además, la comparación de las formas de los tres histogramas muestra, entre otras circunstancias, una mayor dispersión de la distribución de los salarios anuales frente a los salarios mensuales y por hora. Todas estas diferencias se constatan con la comparación de los resultados obtenidos para las medidas descriptivas de cada distribución (tabla A.1 del anexo). Por esta razón y por el diferente tratamiento analítico que requiere una distribución bimodal, a continuación se analizarán de forma independiente las distribuciones del salario por hora y mensual, dejando para el siguiente epígrafe el estudio específico de la distribución de salarios anuales.

Siguiendo la metodología presentada en el segundo apartado, en primer lugar se han realizado las estimaciones por máxima verosimilitud de los cinco modelos considerados para el salario por hora y el salario mensual. Como resultado de este proceso se han obtenido las estimaciones de los parámetros de cada una de las distribuciones teóricas, así como sus errores estándar (tablas 1 y 2). Las estimaciones de los parámetros de forma de la distribución GB2, contenidas en las tablas, presentan valores por encima de 4 para el caso del parámetro a y valores pequeños, inferiores o cercanos a 1, para los parámetros p y q . Se trata, en consecuencia, de valores similares a los obtenidos en estimaciones para las distribuciones de rentas de España y otros países desarrollados (véase, por ejemplo, Jenkins, 2007), lo que revela que en general se da una notable similitud de la forma de la distribución de los salarios con las distribuciones de rentas en los tramos controlados por estos parámetros. En el caso español esta similitud es evidente, por ejemplo, en el caso del salario mensual, donde se llega a obtener, para el modelo de Dagum y Fisk, estimaciones de los parámetros muy próximas a las de las distribuciones de renta (Prieto y García, 2009). Esta evidencia es, pues, indicativa de la similitud en la forma de las distribuciones de rentas y salarios.

Tabla 1. Estimaciones de los parámetros de los modelos teóricos y errores estándar (entre paréntesis). Distribución del salario por hora.

Parámetros	Distribuciones				
	GB2	Dagum	Singh-Maddala	Beta II	Fisk
a	4,043 (0,062)	2,683 (0,006)	6,006 (0,027)	-	3,474 (0,006)
b	5,655 (0,036)	4,635 (0,033)	6,305 (0,012)	0,404 (0,022)	8,995 (0,009)
p	1,938 (0,054)	4,163 (0,059)	-	93,936 (4,879)	-
q	0,570 (0,011)	-	0,353 (0,002)	4,525 (0,017)	-

Tabla 2. Estimaciones de los parámetros los modelos teóricos y errores estándar (entre paréntesis). Distribución del salario mensual.

Parámetros	Distribuciones				
	GB2	Dagum	Singh-Maddala	Beta II	Fisk
a	5,239 (0,100)	3,062 (0,009)	3,104 (0,010)	-	3,045 (0,005)
b	1404,177 (3,793)	1467,209 (5,121)	1418,703 (5,258)	1694,092 (26,217)	1455,389 (1,700)
p	0,510 (0,011)	0,983 (0,007)	-	5,677 (0,041)	-
q	0,473 (0,012)	-	0,949 (0,007)	6,523 (0,053)	-

Los gráficos de probabilidad (*PP-plots*) permiten comprobar la adecuación de los modelos a la distribución empírica de una forma rigurosa y a la vez intuitiva (gráficos A.2 y A.3). Además, arrojan luz sobre el porqué de los resultados globales en las diferentes medidas de bondad del ajuste que analizaremos posteriormente. Con fines interpretativos, hay que tener en cuenta que un ajuste perfecto se correspondería con un gráfico en el que todos los puntos están situados sobre la línea diagonal del cuadrante, mientras que las desviaciones de los puntos por encima y por debajo de la diagonal recogen las discrepancias entre la distribución empírica y el modelo teórico. Así pues, en los diferentes gráficos puede apreciarse, entre otros hechos, la gran proximidad a la diagonal de los puntos del gráfico de la distribución GB2, así como la similitud entre funciones empíricas y teóricas para los modelos Singh-Maddala y Dagum. Los ajustes son mejores, además, para los salarios por hora (excepto en el caso del modelo de Fisk, que se comporta mejor para los salarios mensuales, sobre todo en la parte superior de la distribución). El conjunto de esta evidencia revela, en suma, la superioridad de la distribución GBII en el ajuste tanto de la distribución de los salarios por hora como de la de los salarios mensuales. Estas conclusiones se desprenden también de los histogramas que, aunque menos precisos que los *PP-plots*, permiten valorar visualmente la bondad de los ajustes en una primera aproximación. En ambos gráficos se observan los buenos resultados de los ajustes de la distribución GB2, la cual, sobre todo, modeliza de forma excelente la cola derecha de la distribución. En el caso del salario por hora puede apreciarse que los ajustes obtenidos son muy adecuados en general, pero en particular en el caso de los modelos GB2 y Singh-Maddala.

Gráfico 2. Histograma de la distribución de los salarios por hora y funciones de densidad de los modelos teóricos ajustados.

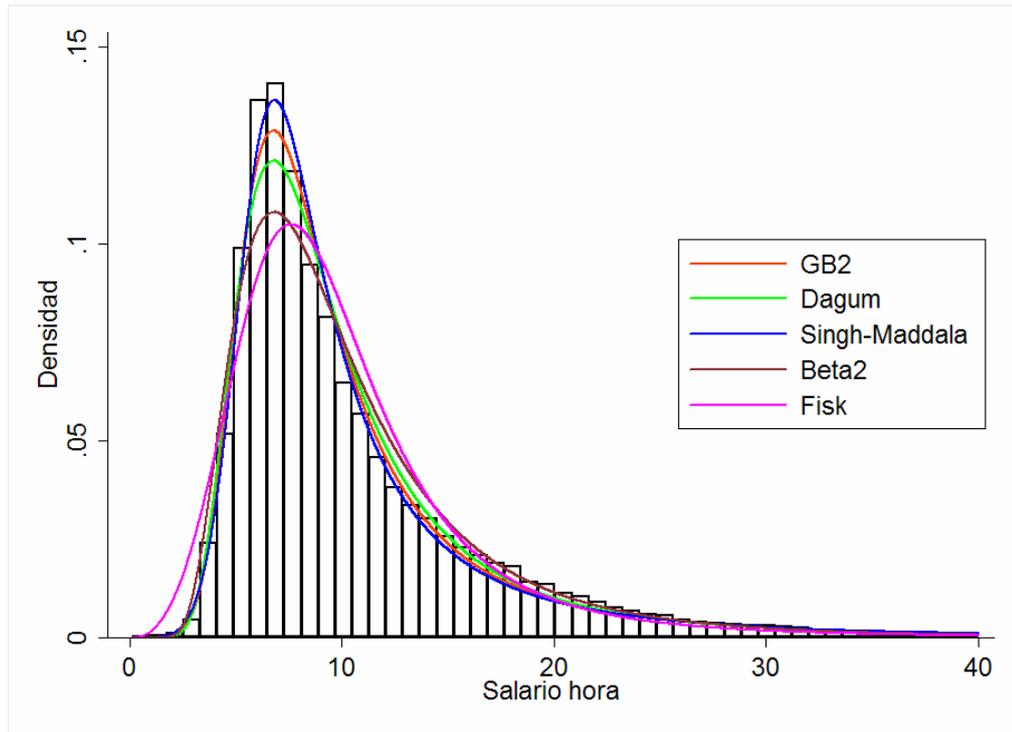
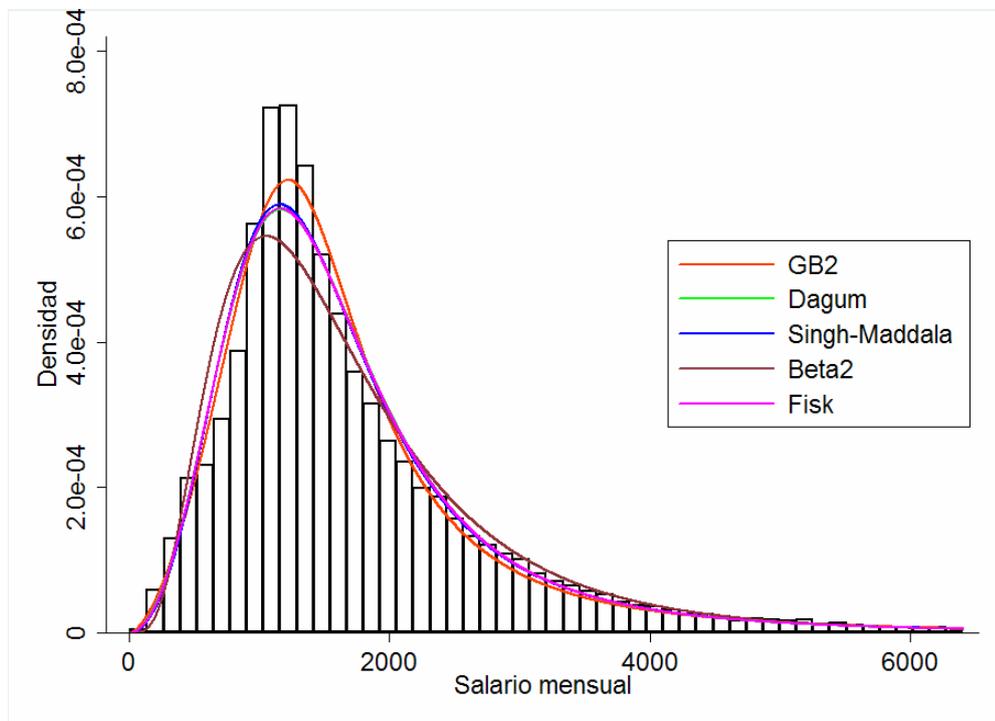


Gráfico 3. Histograma de la distribución de los salarios mensuales y funciones de densidad de los modelos teóricos ajustados.



Para completar el siguiente paso metodológico, consistente en el análisis de la adecuación de los modelos y la selección de los mejores, se han obtenido las medidas previamente indicadas para estudiar la bondad del ajuste (logaritmo de la función de verosimilitud, coeficiente de correlación lineal entre los valores de las funciones de distribución empírica y teórica y los estadísticos de Kolmogorov-Smirnov y de Cramer-Von Mises) y las ordenaciones de modelos (del mejor al peor) que se generan de acuerdo a los valores de cada medida (tablas 3 y 4). Para la correcta interpretación de las tablas, es necesario tener en cuenta que valores elevados del logaritmo de la función de verosimilitud y del coeficiente de correlación lineal indican buenos ajustes, ya que el logaritmo de la función de verosimilitud es la función objetivo que se maximiza en el proceso de estimación de los parámetros y el coeficiente de correlación entre los valores de las funciones de distribución empírica y teórica refleja la proximidad de ambas. Por su parte, los otros dos estadísticos se construyen sobre el concepto de distancia entre las funciones de distribución empírica y teórica, por lo que valores más pequeños indican mayor proximidad y, por tanto, mejores ajustes¹¹.

Partiendo de estas premisas, todas las medidas de las tablas 3 y 4 muestran la clara superioridad del modelo GB2 sobre los demás. No obstante, también presentan valores reseñables las medidas de bondad de las distribuciones de Singh-Maddala y Dagum, las cuales producen buenos ajustes tal como sugerían los *PP-plots*.

En el caso de los salarios por hora, los resultados de todas las medidas de bondad de ajuste generan ordenaciones idénticas, lo que da idea de la robustez de los resultados obtenidos ante los diferentes criterios de bondad de ajuste utilizados. En el caso de los salarios mensuales, la coincidencia también es casi total, con una única variación en la ordenación generada por el logaritmo de la función de verosimilitud, consistente en el intercambio de dos posiciones.

De la comparativa de las medidas de bondad del ajuste de los modelos se deduce también la capacidad de la distribución Fisk, con sólo dos parámetros, para modelizar la distribución, sobre todo de los salarios mensuales. En este caso, esta distribución biparamétrica presenta un mejor ajuste que otras distribuciones dotadas de un mayor número de parámetros como es el caso de la distribución beta de segunda especie que, dentro de las triparamétricas, produce los peores ajustes¹².

¹¹ Cabe señalar que aunque estos dos últimos estadísticos son los empleados para realizar sendos contrastes de bondad del ajuste, su distribución bajo una hipótesis nula compuesta es desconocida (Stephens, 1986) y, consecuentemente, no se puede evaluar si las diferencias entre la función de distribución empírica y la teórica son estadísticamente significativas. Por lo tanto, estos dos estadísticos se utilizan esencialmente como medidas descriptivas de magnitud de tales diferencias. Así, cuanto mayores son los valores de estos dos estadísticos mayor distancia habrá entre la función de distribución empírica y la teórica.

¹² La distribución Fisk presenta ciertas ventajas para el análisis de la distribución al depender de un único parámetro a (éste, además, es el inverso del índice de Gini, hecho que permite una clara interpretación económica del parámetro como medida de igualdad). Así, la aplicación de los resultados sobre dominancia de Lorenz obtenidos por Kleiber (1999) permite establecer ordenaciones sobre desigualdad muy sencillas, dependientes únicamente del parámetro a . En nuestro caso, se puede concluir, a partir de las estimaciones del parámetro a de los modelos de Fisk, que la distribución de los salarios por hora domina en sentido de Lorenz (esto es, se caracteriza por una menor desigualdad) a la de salarios mensuales y ésta a la de los salarios anuales.

Tabla 3. Medidas de bondad del ajuste de los modelos teóricos y ordenaciones que generan.
Distribución de los salarios por hora.

Modelos	Logaritmo de la función de verosimilitud	Orden	Coficiente de correlación	Orden	Estadístico de Cramer Von Mises	Orden	Estadístico Kolmogorov-Smirnov	Orden
GB2	-691855,77	1	0,9994	1	30,4121	1	0,0209	1
Dagum	-692244,81	3	0,9989	3	45,7409	3	0,0259	3
Singh-Maddala	-692168,22	2	0,9993	2	39,8110	2	0,0246	2
Beta II	-695378,00	4	0,9969	4	138,3597	4	0,0427	4
Fisk	-701524,25	5	0,9947	5	209,2573	5	0,0462	5

Nota: La ordenación va del mejor al peor modelo en función de la correspondiente medida de bondad del ajuste.

Tabla 4. Medidas de bondad del ajuste de los modelos teóricos y ordenaciones que generan.
Distribución de los salarios mensuales

Modelos	Logaritmo de la función de verosimilitud	Orden	Coficiente de correlación	Orden	Estadístico de Cramer Von Mises	Orden	Estadístico Kolmogorov-Smirnov	Orden
GB2	-1920041,7	1	0,9993	1	26,9800	1	0,0174	1
Dagum	-1920531,8	3	0,9988	4	47,7950	4	0,0266	4
Singh-Maddala	-1920509,5	2	0,9989	2	41,6354	2	0,0257	2
Beta II	-1923520,9	5	0,9973	5	128,9696	5	0,0448	5
Fisk	-1920534,8	4	0,9988	3	46,6158	3	0,0265	3

Nota: La ordenación va del mejor al peor modelo en función de la correspondiente medida de bondad del ajuste.

Finalmente, para la comparación entre modelos anidados, realizaremos contrastes para descartar aquellas distribuciones en las que la introducción de un parámetro adicional no supone una mejora significativa en la bondad del ajuste. Conforme se puede comprobar en las dos primeras columnas de la tabla 5, en todos los casos los valores de los estadísticos de prueba para los salarios por hora y mensuales son sistemáticamente elevados, y los correspondientes p -valores se sitúan por debajo del 1%. Estos resultados indican que en todos los contrastes realizados se rechazan las hipótesis de restricciones de los parámetros con un nivel de significación por debajo del 1%. Así pues, en todas las comparaciones la introducción de un nuevo parámetro mejora significativamente la bondad del ajuste, por lo que siempre son mejores los modelos más flexibles (con mayor número de parámetros) que los que tienen menos parámetros. En consecuencia, los resultados de los contrastes anidados para el caso de los salarios por hora y para los mensuales revelan la superioridad del modelo GB2 sobre los demás modelos anidados en él.

Tabla 5. Resultados de los contrastes anidados de comparación entre modelos teóricos.

Hipótesis Nulas	Salarios por hora	Salarios mensuales	Salarios anuales		
			Total	Contratos indefinidos	Contratos temporales
$\alpha=1$ (Beta II mejor que GB2)	$\chi_1^2=2416,90$ (0,000)	$\chi_1^2=1791,34$ (0,000)	$\chi_1^2=1394,65$ (0,000)	$\chi_1^2=6373,15$ (0,000)	$\chi_1^2=36,63$ (0,000)
$p=1$ (Singh-Maddala mejor que GB2)	$\chi_1^2=307,23$ (0,000)	$\chi_1^2=1827,04$ (0,000)	$\chi_1^2=1275,52$ (0,001)	$\chi_1^2=2957,92$ (0,000)	$\chi_1^2=50,03$ (0,000)
$q=1$ (Dagum mejor que GB2)	$\chi_1^2=1452,13$ (0,000)	$\chi_1^2=2089,95$ (0,000)	$\chi_1^2=11,90$ (0,001)	$\chi_1^2=584,59$ (0,000)	$\chi_1^2=71,73$ (0,000)
$p=q=1$ (Fisk mejor que GB2)	$\chi_2^2=20513,96$ (0,000)	$\chi_2^2=2123,37$ (0,000)	$\chi_2^2=13738,23$ (0,000)	$\chi_2^2=9204,62$ (0,000)	$\chi_2^2=75,96$ (0,000)

Nota: En cada casilla se presentan los valores del estadístico de prueba y los p-valores del contraste (entre paréntesis). En todos los casos se rechaza la hipótesis planteada con una significación del 1%.

4.2 Modelización de la distribución del salario anual

El histograma que aparece en el gráfico 4, las diferentes medidas descriptivas obtenidas para la distribución de salarios anuales y el análisis de inferencia sobre las mismas revelan, como hemos señalado con anterioridad, características diferentes en relación con las distribuciones del salario por hora y mensual. Entre ellas destaca la presencia de dos modas en la distribución, un hecho que dificulta la modelización cuando se utilizan modelos teóricos unimodales, los más habitualmente considerados para ajustar distribuciones de ingresos. Esta circunstancia nos ha llevado a realizar una aproximación diferente en el caso de la distribución de salarios anuales. Así, tras analizar las causas de la bimodalidad, y una vez detectada la potencial razón de la misma, se ha procedido al análisis diferenciando dos subpoblaciones de asalariados.

En cualquier caso, dada la flexibilidad de los modelos seleccionados en este trabajo, capaz de producir aceptables ajustes incluso en distribuciones multimodales, también se han realizado los ajustes de la distribución general de salarios anuales, obteniéndose las distintas estimaciones de los parámetros que se presentan en la tabla 6.

Los gráficos de probabilidad (gráfico A.5) indican, en términos generales, la adecuación del modelo GB2 y Dagum para modelizar los salarios anuales. Sin embargo, el gráfico 4, aunque más impreciso por la agrupación por intervalos en el histograma, muestra la dificultad de los modelos teóricos para reproducir la zona intermodal de la distribución, razón que nos lleva a plantear el análisis posterior por subpoblaciones.

Respecto a la selección de modelos, de las medidas de bondad del ajuste (tabla 7) y de los resultados de los contrastes anidados (tabla 5) puede deducirse que los modelos con mayor número de parámetros presentan siempre los mejores ajustes en este caso. En cierto modo, la presencia de una distribución bimodal hacía esperable la necesidad de mayor flexibilidad para ajustar la distribución empírica y, en consecuencia, la ordenación de los mejores modelos está determinada por el número de parámetros. Así pues, el mejor modelo según todos los criterios es la GB2, situándose en un segundo escalón los modelos Dagum y Singh-Maddala.

Gráfico 4. Histograma de la distribución de los salarios anuales y funciones de densidad de los modelos teóricos ajustados.

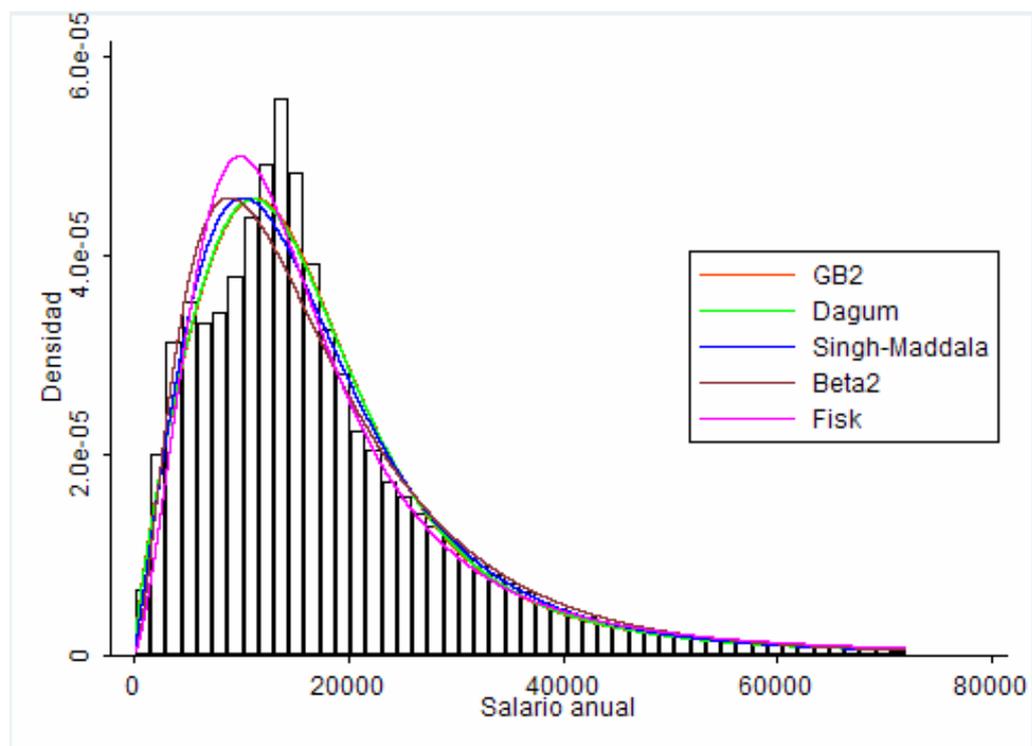


Tabla 6. Estimaciones de los parámetros de los modelos teóricos y errores estándar (entre paréntesis). Distribución del salario anual.

Parámetros	Distribuciones				
	GB2	Dagum	Singh-Maddala	Beta II	Fisk
a	3,184 (0,058)	3,002 (0,010)	2,030 (0,006)	-	2,391 (0,004)
b	18867,32 (106,735)	19165,42 (66,017)	21517,56 (151,235)	47221,84 (850,499)	14324 (21,338)
p	0,554 (0,012)	0,595 (0,004)	-	2,729 (0,013)	-
q	0,914 (0,025)	-	1,844 (0,018)	8,174 (0,105)	-

Tabla 7. Medidas de bondad del ajuste de los modelos teóricos y ordenaciones que generan.
Distribución de los salarios anuales

Modelos	Logaritmo de la función de verosimilitud	Orden	Coefficiente de correlación	Orden	Estadístico de Cramer Von Mises	Orden	Estadístico Kolmogorov Smirnov	Orden
GB2	-2502414,9	1	0,9994	1	23,0424	1	0,0193	1
Dagum	-2502420,3	2	0,9994	2	24,8370	2	0,0204	2
Singh-Maddala	-2502741,2	3	0,9990	3	42,1038	3	0,0286	3
Beta II	-2504106,5	4	0,9981	5	90,9479	5	0,0411	5
Fisk	-2505371,4	5	0,9983	4	70,2289	4	0,0360	4

Notas: La ordenación va del mejor al peor modelo en función de la correspondiente medida de bondad del ajuste.

Una vez analizados los resultados del ajuste a la distribución global, profundizaremos en el estudio de la modelización del salario anual, partiendo del análisis de las causas que podrían provocar la bimodalidad de la distribución. Con este fin, se han comparado las subpoblaciones que se generan a partir de las modalidades de cada una de las variables de clasificación de la encuesta (sexo del trabajador, tipo de empresa, tipo de contrato, etc.), contrastando las diferencias entre distribuciones utilizando, entre otras técnicas, contrastes de Kolmogorov-Smirnov y Kruskal-Wallis de varias poblaciones¹³. Los resultados indican que la principal causa de bimodalidad es la convivencia en la misma distribución de salarios anuales de dos subpoblaciones: la de los trabajadores con contratos temporales y la de los trabajadores con contratos indefinidos. Esta conclusión puede intuirse con claridad mediante la comparación de los histogramas de la distribución global (gráfico 4) con la de ambos tipos de salarios (gráficos 5 y 6). En relación con esta cuestión, es importante destacar que Motellón et al. (2010) documentan la presencia de importantes diferencias en las distribuciones salariales cuando se distingue por tipo de contrato, así como que el diferente comportamiento de estas distribuciones explica buena parte del cambio en la distribución salarial global en España durante el período 1995-2002.

Los salarios de los trabajadores con contratos indefinidos presentan una distribución más habitual en muestras de rentas o ingresos, mientras que los salarios de los trabajadores con contratos temporales presentan una mayor dispersión y heterogeneidad, además de contar con una moda inferior a la que presentan los salarios anuales. Esta moda se corresponde con la primera de las dos modas de la distribución global de salarios anuales, como puede deducirse de la comparación de los intervalos modales de ambas distribuciones. Obviamente, se trata de una moda relativa dado, entre otros hechos, el menor número de individuos con contratos temporales (62.542) frente al número de individuos con contratos indefinidos (172.354) presentes en la muestra.

¹³ Los resultados detallados están disponibles por parte de los autores ante su requerimiento.

Gráfico 5. Histograma de la distribución de salarios anuales de contratos indefinidos y funciones de densidad de los modelos teóricos ajustados.

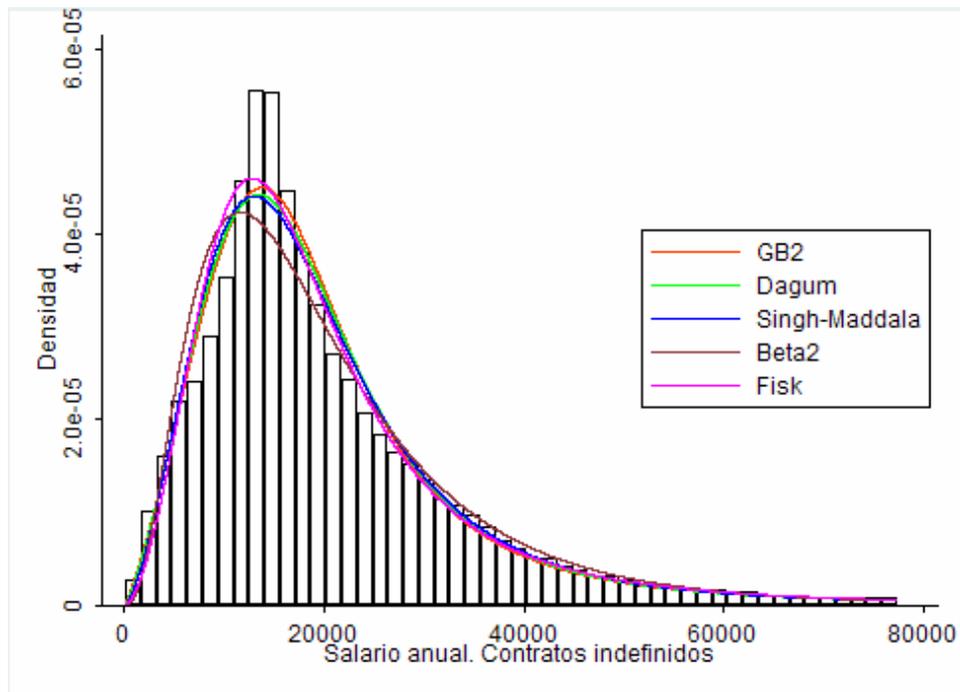
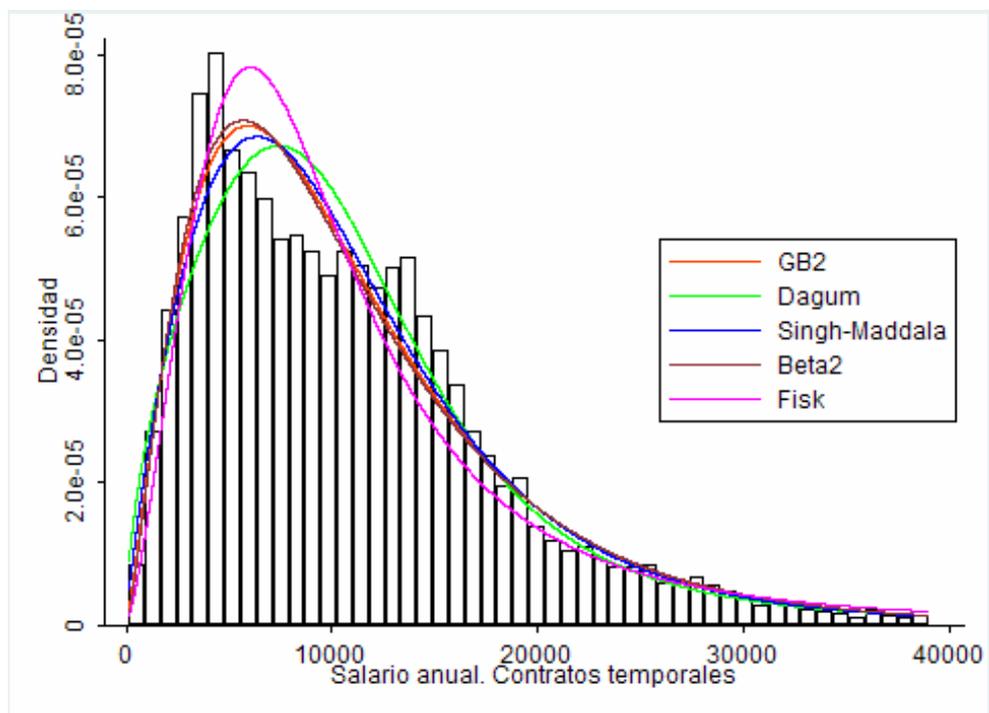


Gráfico 6. Histograma de la distribución de salarios anuales de contratos temporales y funciones de densidad de los modelos teóricos ajustados.



Una vez establecida esta diferenciación, se ha procedido a ajustar los modelos teóricos a estas dos subpoblaciones, mejorándose los ajustes en el caso de los salarios anuales de contratos indefinidos, sobre todo, para las distribuciones Singh-Maddala y Fisk (gráfico A.5). En el caso de los salarios anuales de contratos temporales, la adecuación empeora en general, detectándose por tanto que son estos datos los que provocan un peor funcionamiento de los ajustes cuando se incorporan a la distribución global (gráfico A.6). A su vez, todos los resultados obtenidos son muy diferentes entre sí, tanto en los parámetros estimados como en las medidas de bondad de ajuste (tablas 9 y 10).

En el caso de los contratos indefinidos, el modelo mejor en dos de los criterios de bondad es la distribución GB2. Sin embargo, de acuerdo a los otros dos criterios, el modelo que mejor se ajusta es el de Fisk, que cuenta con la ventaja de tener únicamente dos parámetros. La observación de los gráficos de probabilidad, presentados en el anexo, deja claro esta gran capacidad de la distribución Fisk para modelizar adecuadamente toda la distribución con excepción del cuartil inferior (algo que también se observa en trabajos sobre la modelización de la distribución personal de la renta, en especial para el caso español: Prieto y García, 2009).

Tabla 9. Estimaciones de los parámetros de los modelos teóricos y errores estándar (entre paréntesis). Distribuciones del salario anual para contratos indefinidos y temporales.

Parámetros	Contratos indefinidos				
	GB2	Dagum	Singh-Maddala	Beta II	Fisk
a	3,622 (0,033)	3,013 (0,012)	2,499 (0,009)	-	2,719 (0,006)
b	18742,6 (45,914)	19452,43 (78,401)	19712,01 (118,038)	35854,07 (357,389)	16946,4 (425,878)
p	0,614 (0,007)	0,767 (0,006)	-	3,746 (0,012)	-
q	0,757 (0,010)	-	1,304 (0,013)	7,499 (0,048)	-
Parámetros	Contratos temporales				
	GB2	Dagum	Singh-Maddala	Beta II	Fisk
a	1,299 (0,049)	3,297 (0,028)	1,836 (0,010)	-	2,317 (0,008)
b	28852,22 (2110,448)	13691,43 (100,477)	19255,58 (376,807)	55131,95 (1400,165)	8968,646 (27,073)
p	1,662 (0,094)	0,475 (0,007)	-	2,451 (0,011)	-
q	6,235 (0,618)	-	2,950 (0,076)	13,113 (0,269)	-

Los resultados que se obtienen para los contratos temporales son distintos a los anteriores, aunque también se constata la superioridad de la GB2 seguida muy de cerca por la distribución Singh-Maddala (ésta produce los mejores ajustes según dos criterios de bondad de ajuste, el coeficiente de correlación y el estadístico de Kolmogorov-Smirnov). La distribución Fisk resulta inadecuada en este caso, ya que con un solo parámetro no es capaz de captar toda la variabilidad de la forma de una distribución con muchas irregularidades, alejada de la forma acampanada habitual.

Los valores de las estimaciones de los parámetros son también marcadamente diferentes, indicando así las peculiaridades en desigualdad y forma de esta distribución. Véase, como ejemplo, en la distribución GB2 el reducido valor del parámetro α , responsable de la forma global de la distribución, o el elevado valor del parámetro q , que indica un cambio importante en el comportamiento de la cola derecha, el cual es muy diferente al de las distribuciones de renta habituales o al de las demás distribuciones salariales analizadas en este trabajo.

Así pues, en suma, la separación en subpoblaciones (contratos temporales e indefinidos) para el estudio de los salarios anuales permite diferenciar dos regularidades empíricas distintas que, a su vez, responden a realidades económicas diferentes. Una modelización por separado de ambas subpoblaciones, como herramienta de análisis, facilitará sin duda el posterior estudio económico de las distribuciones.

Tabla 10. Medidas de bondad del ajuste de los modelos teóricos y ordenaciones que generan. Distribuciones de los salarios anuales de contratos indefinidos y temporales.

Contratos indefinidos								
Modelos	Logaritmo de la función de verosimilitud	Orden	Coeficiente de correlación	Orden	Estadístico de Cramer Von Mises	Orden	Estadístico Kolmogorov Smirnov	Orden
GB2	-1848259.7	1	0,9992	2	21,5866	2	0,0195	1
Dagum	-1848314.3	2	0,9991	3	25,5167	3	0,0225	2
Singh-Maddala	-1848484	3	0,9990	4	28,2156	4	0,0256	4
Beta II	-1850160.8	5	0,9980	5	72,3396	5	0,0416	5
Fisk	-1848895.3	4	0,9993	1	19,6129	1	0,0249	3
Contratos temporales								
Modelos	Logaritmo de la función de verosimilitud	Orden	Coeficiente de correlación	Orden	Estadístico de Cramer Von Mises	Orden	Estadístico Kolmogorov Smirnov	Orden
GB2	-637096.71	1	0,9988	2	13,1137	1	0,0282	2
Dagum	-637530.88	4	0,9985	4	19,3218	4	0,0370	4
Singh-Maddala	-637148.87	3	0,9989	1	12,6159	2	0,0263	1
Beta II	-637115.92	2	0,9987	3	13,7525	3	0,0299	3
Fisk	-638600.3	6	0,9961	5	42,9033	5	0,0459	5

Nota: La ordenación va del mejor al peor modelo en función de la correspondiente medida de bondad del ajuste.

5. Conclusiones

Aun a pesar de que se trata de una técnica que presenta numerosas ventajas analíticas en el ámbito del estudio económico y causal de las distribuciones empíricas de variables económicas como la riqueza o la renta, la modelización paramétrica de las distribuciones salariales es una técnica apenas utilizada en los estudios sobre el mercado de trabajo. En este artículo se procede a modelizar paramétricamente la distribución salarial en España y, por ello, se contribuye a una línea de investigación novedosa, consistente en la aplicación en el campo de las distribuciones salariales de un tipo de técnica limitado hasta ahora en la literatura económica a otros ámbitos de análisis.

En el análisis se ha utilizado una metodología rigurosa para seleccionar y estimar los mejores modelos paramétricos para la distribución salarial de España, para lo que se ha examinado la adecuación del modelo beta generalizada de segunda especie y de cuatro modelos anidados en él. Se ha empleado para ello la fuente de datos más completa de la que puede disponerse sobre la distribución salarial en España, la muestra completa de la *Encuesta de Estructura Salarial*, diferenciando tres conceptos salariales de interés habitual para los trabajos especializados: salario por hora, mensual y anual.

Los resultados muestran que el funcionamiento de la modelización paramétrica es apropiada y similar al que presenta en distribuciones de renta, con medidas de bondad de ajuste aceptables. Independientemente del concepto de salario y la técnica de análisis empleados, se comprueba que la distribución beta generalizada de segunda especie es en general la más adecuada para la distribución salarial de la economía española. También hay que destacar el comportamiento del modelo Singh-Maddala, en el caso del salario por hora, y del modelo Fisk para ajustar, sobre todo, la parte superior de todas las distribuciones.

Así pues, en suma, en el artículo se comprueba que las distribuciones de salarios se pueden representar adecuadamente por modelos paramétricos, resultado novedoso para en el caso español, pero también en los estudios desarrollados en el ámbito internacional, donde tampoco son frecuentes este tipo de análisis. El estudio de las distribuciones salariales introduce, por otra parte, elementos enriquecedores sobre los procedimientos de modelización en distribuciones de rentas, un terreno en el que se ha venido aplicando más habitualmente la modelización paramétrica. Entre estos elementos destaca la consideración de varios conceptos salariales, así como la necesidad de la identificación de subpoblaciones para explicar la bimodalidad de algunas distribuciones y proceder a una posterior mejora en su modelización por separado. Este hecho facilitará sin duda el posterior estudio económico de las distribuciones.

Por último, cabe destacar que a partir de la evidencia presentada en este trabajo, en el futuro sería posible explorar nuevas líneas de investigación sobre la utilización del modelo beta generalizada de segunda especie para el análisis de distribuciones salariales. Como muestra, partiendo de modelos paramétricos podrían desarrollarse estudios causales sobre los factores determinantes de las distribuciones salariales, mediante modelos econométricos que permitan reproducir el efecto de indicadores como la tasa de paro, la inflación o el crecimiento del PIB sobre los parámetros de las distribuciones salariales.

Referencias bibliográficas

- Alcalá, F.; Hernández, P.J. (2010): "Firms' Main Market, Human Capital, and Wages", *SERIEs-Journal of the Spanish Economic Association*, vol. 1(4), 433-458.
- Bloomquist, K. M. and M. Tsvetovat (2007): Pareto Distribution of U.S. Wage and Salary Income. ESHIA / WEHIA 2007.
- Bonhomme, S. y Hospido, L. (2011): "Using Social Security Data to Estimate Earnings Inequality", mimeo.
- Callealta, J., J.M. Casas y J. Núñez (1996): " Distribución de la renta per capita disponible en España: descripción, desigualdad y modelización", en Pena B., J. Callealta, J.M. Casas, A. Merediz y J. Núñez (eds.) Cap5.
- Card, D. y S. De la Rica (2006): "Firm-level Contracting and the Structure of Wages", *Industrial and Labor Relations Review*, vol. 59 (4), 573-593.
- Chotikapanich, D. (2008): *Modeling Income Distributions and Lorenz Curves*, Springer, Nueva York
- Chotikapanich, D. y W. E Griffiths (2008): "Estimating Income Distributions Using a Mixture of Gamma Densities", Department of Economics - Working Papers Series 1034, Universidad de Melbourne.
- Cowell, F. (2002): "Measurement of Inequality", en Handbook of Income distribution, A. B. Atkinson y F. Bourguignon (eds.), North Holland, Amsterdam.
- Cox, D.R., y D. V. Hinkley (1974): *Theoretical Statistics*. Chapman & Hall, Londres.
- Creedy, J., J. Lye y V.L. Martin (1996): "A Labour Market Equilibrium Model of The Personal Distribution of Earnings", *Journal of Income Distribution*, 6, 127-144.
- Dagum, C. (1977): "A new model of personal expenditure distribution: Secification and estimation", *Economiee Appliquée*, 30, 413-437.
- Dagum, C. (1980): "The generation and distribution of income, the Lorenz curve and the Gini ratio", *Economiee Appliquée* 33, 327-367.
- Dastrup, S.; R. Hartshorn y J. McDonald (2007): "The impact of taxes and transfer payments on the distribution of income: A parametric comparison," *Journal of Economic Inequality*, vol. 5(3), 353-369.
- Domma, F. y Perri, P. (2009): "Some developments on the log-Dagum distribution", *Statistical Methods and Applications*, 18(2), 205-220.
- Eurostat (2006): Structure of Earnings Survey 2002 Quality Report.
- Fortin, N.; Lemieux, T.; Firpo, S. (2011): "Decomposition Methods in Economics", *Handbook of Labor Economics*, ed. Elsevier.
- García, C. y Prieto, M. (2011): "Using the Dagum model to explain changes in personal income distribution", *Applied Economics*, 43 (28), 4377-4386.
- Graf, M. y D. Nedyalkova (2010): "Fitting the Generalized Beta Distribution of the Second Kind to the Empirical Income Distribution from the Aggregate Laeken Indicators". International Conference on Indicators and Survey Methodology, Viena, Austria.
- Jäntti, M. y S.P. Jenkins (2010): "Examining the Impact of Macro-Economic Conditions on Income Inequality", *Journal of Economic Inequality*, 8 (2), 221-240.

- Jenkins, S. P. (2007): Inequality and the GB2 Income Distribution, in: ECINEQ Working Paper Series, Society for the Study of Economic Inequality.
- Kleiber, C. (1999): "On the Lorenz order within parametric families of income distributions", *Sankhyā: The Indian Journal of Statistics*, 61, 514-517.
- Kleiber, C. y S. Kotz (2003): "Statistical Size Distributions in Economics and Actuarial Sciences". Wiley Series in Probability and Statistics.
- Lydall, H. A. (1959): The Distribution of Employment Incomes. *Econometrica* 27 (1), 110-115.
- Lydall, H.F. (1968): The Structure of Earnings. Oxford Clarendon Press, Oxford.
- Maasoumi, E. (1994): "Empirical analysis of inequality and welfare", en P. Schmidt y H. Pesaran (eds.): Handbook of Applied Microeconomics. Oxford: Blacwell Publisher.
- McDonald, J.B. (1984): "Some Generalized Functions for the Size Distribution of Income", *Econometrica*, 52, 647-663.
- Moral-Arce, I.; Martín, E. (2009): "Integración de información administrativa y muestral en estadísticas económicas estructurales. La Encuesta Anual de Estructura Salarial", *Estadística Española*, Vol. 51, núm. 172, 487-504.
- Motellón, E.; López-Bazo, E.; El-Attar, M. (2010): "Cambios en la distribución salarial en España, 1995-2002. Efectos a través del tipo de contrato", *Revista de Economía Aplicada*, Número 53 (vol. XVIII), 5-38.
- OCDE (2011): *Divided We Stand: Why Inequality Keeps Rising*, ed. OCDE.
- Pareto, V. (1897): *Cours de Economie Politique*, traducido al inglés en A.S. Scheiwer y A.N. Page (eds.), (1972), McMillan.
- Parker, S.C. (1996): "A Labor Market Theory of the Earnings Distribution". *Journal of Income Distribution*. 6(1), 33-52.
- Parker, S.C. (1999): "The Generalised Beta as a Model of the Earnings Distribution", *Economics Letters*, 62(2), 197-200.
- Pinkovsky, M. y Xavier Sala-i-Martin (2009): "Parametric Estimations of the World Distribution of Income", NBER Working Paper 15433.
- Prieto, M. (1998): Modelización paramétrica de la distribución personal de la renta para España mediante métodos robustos, Tesis doctoral, Universidad de Valladolid.
- Prieto, M. y C. García (2009): "La distribución beta generalizada de segunda especie como modelo de la distribución personal de la renta en España", *Estadística Española*, 51 (170), 33-62.
- Reed, W. y M. Jorgensen, (2005): "The double Pareto-lognormal distribution. A new parametric model for size distribution", *Communications in Statistics*, 34, 1733-1753.
- Roy, A.D. (1950). "The Distribution of Earnings and of Individual Output". *Economic Journal*, 60 (239), 489-505.
- Salem, S.K. y G.S. Mount (1974): "A convenient descriptive model of income distribution: the Gamma distribution", *Econometrica*, 42, 115-127.
- Sarabia, J.M., E. Castillo, y D.J. Slottje (2002): "Lorenz Ordering Between McDonald's Generalized Functions of the Income Size Distributions". *Economics Letters*, 75, 265-270.
- Sattinger, M. (1996); "Choice, Order Statistics and the Distribution of Earnings". *Journal of Income Distribution*, 6(1), 9-32.

- Schneider, P.A. (2008): "A Comparative Entropy Analysis of the Distribution of Wages and Salaries" en *The Annual UMass Amherst / NSSR Workshop*. 2008.
- Simón, H. (2009): "La desigualdad salarial en España: Una perspectiva internacional y temporal", *Investigaciones Económicas*, vol. XXXIII(3), 439-471.
- Simón, H. (2010): "International Differences in Wage Inequality: A New Glance with European Matched Employer-Employee Data", *British Journal of Industrial Relations*, 48(2), 310-346.
- Singh, S.K. y G.S. Maddala (1976): "A Function for the Size Distribution of Incomes". *Econometrica*, 44, 963-970.
- Stephens, M.A. (1986): "Tests based on EDF statistics". En D'Agostino, R.B. y Stephens, M.A.,(eds.): *Goodness-of-Fit Techniques*. Marcel Dekker, Nueva York.
- Thurow, L.C. (1970): "Analyzing the American Income Distribution". *American Economic Review*, 60, 261-269.
- Wifling, B. (1996): "Lorenz ordering of generalized beta-II income distributions", *Journal of Econometrics*, 71(1-2), 381-388.

ANEXO

Tabla A.1. Estadísticos descriptivos.
Comparación de las muestras *completa* y *restringida* de la EES 2006.

	Muestra <i>completa</i>			Muestra <i>restringida</i>		
	Salario Hora	Salario Mensual	Salario Anual	Salario Hora	Salario Mensual	Salario Anual
Media muestral	10,913	1751,199	18022,18	10,690	1712,194	1744,47
Primer cuartil	6,501	1054,210	9322,59	6,452	1045,21	9064,7
Mediana muestral	8,555	1413,732	14629,87	8,345	1384,045	14313,71
Tercer Cuartil	12,688	2084,528	22289,23	12,225	2010,148	21212,94
Desviación típica	7,915	1270,828	14764,31	7,685	1233,196	14320,77
Índice de Gini	0,310	0,328	0,379	0,307	0,324	0,378
P90/P10	3,590	4,262	6,858	3,544	4,176	6,805
P90/P50	2,228	2,182	2,315	2,249	2,176	2,283
P50/P10	1,669	1,915	2,939	1,576	1,919	2,980
Hombre		0,611			0,590	
Inmigrante		0,63			0,074	
Edad		38,61			38,44	
Estudios primarios		0,250			0,265	
Estudios secundarios		0,533			0,515	
Estudios terciarios		0,218			0,221	
Jornada parcial		0,146			0,156	
Contrato temporal		0,266			0,288	
Observaciones		234.896			140.321	

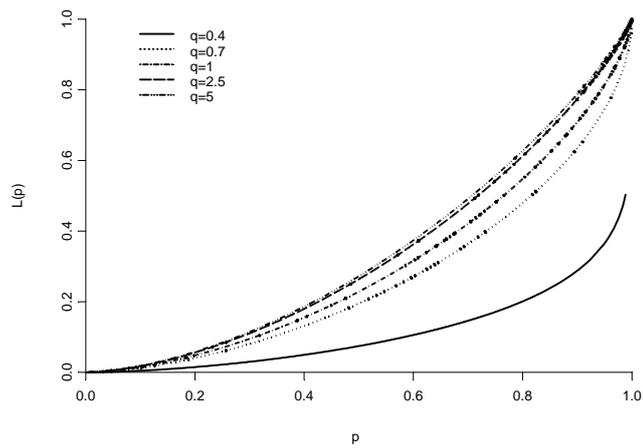
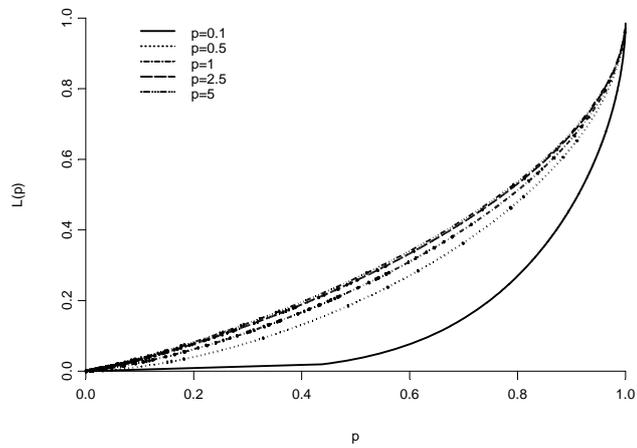
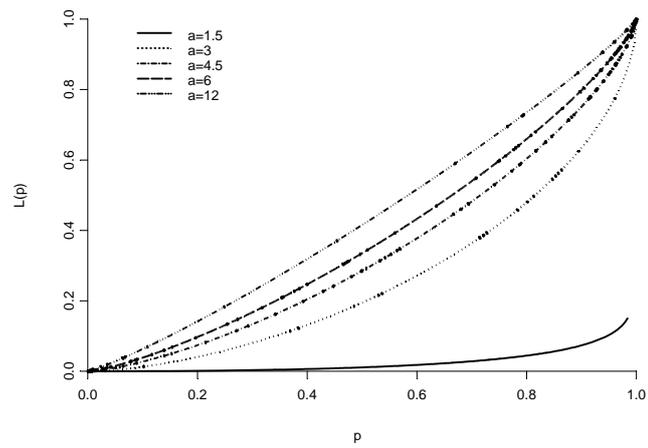


Gráfico A.1. Sensibilidad de las curvas de Lorenz a las variaciones de los parámetros de forma de la distribución GB2.

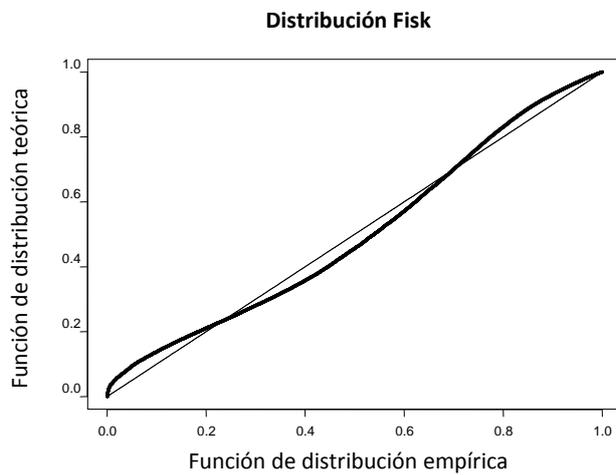
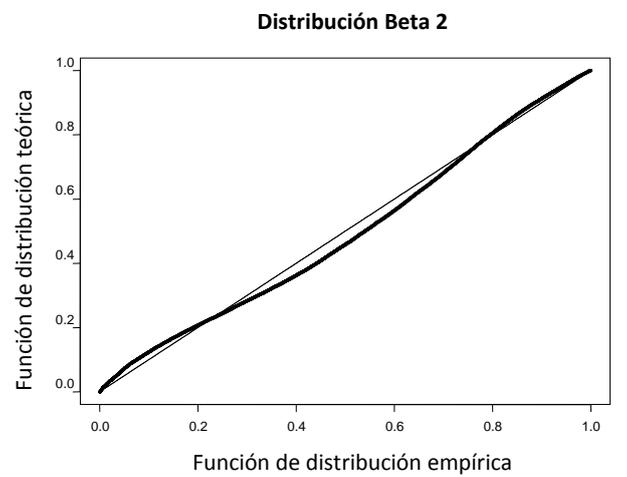
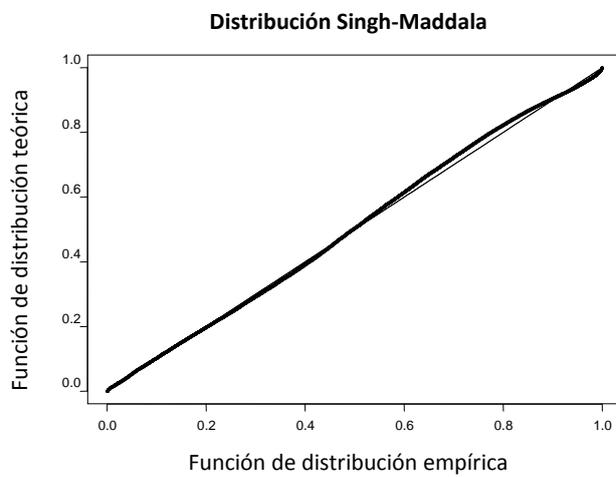
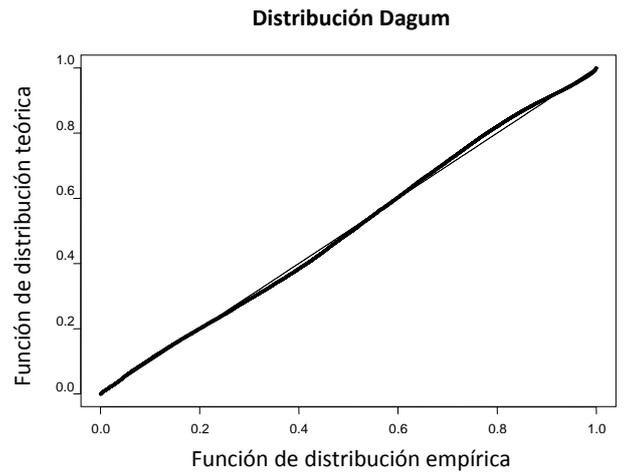
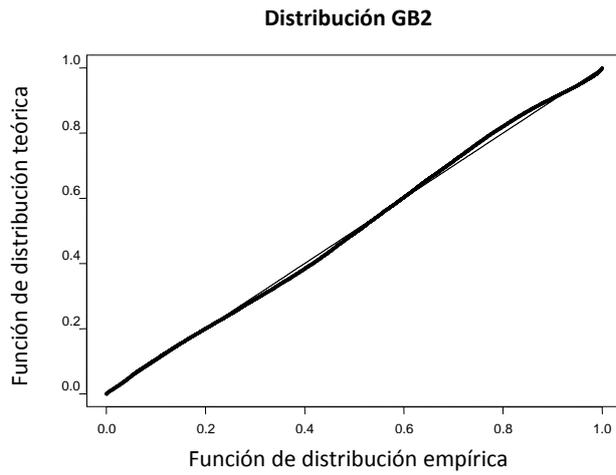


Gráfico A.2. Gráficos de probabilidad de los ajustes de la distribución del salario por hora.

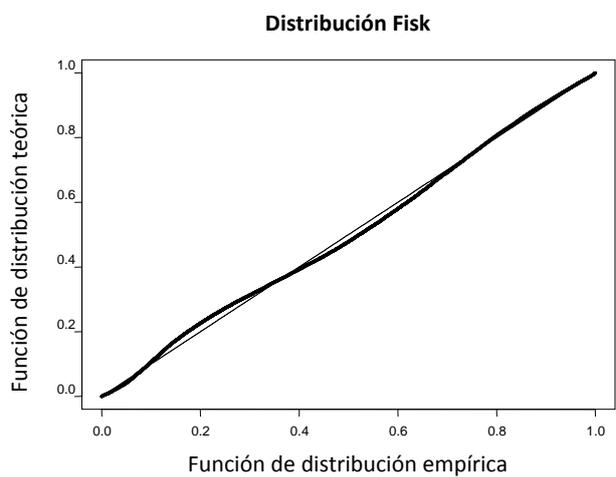
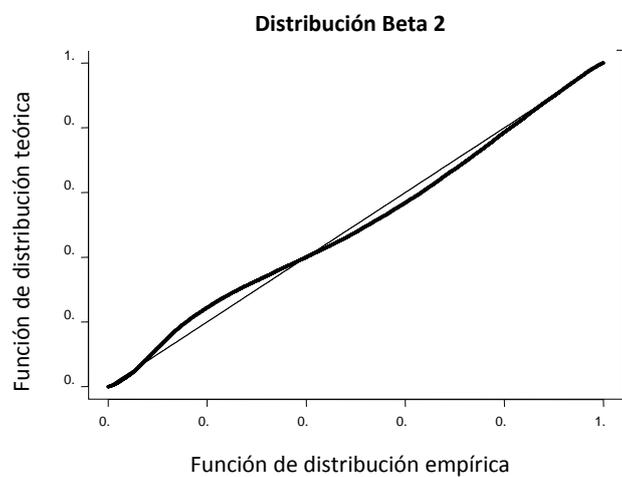
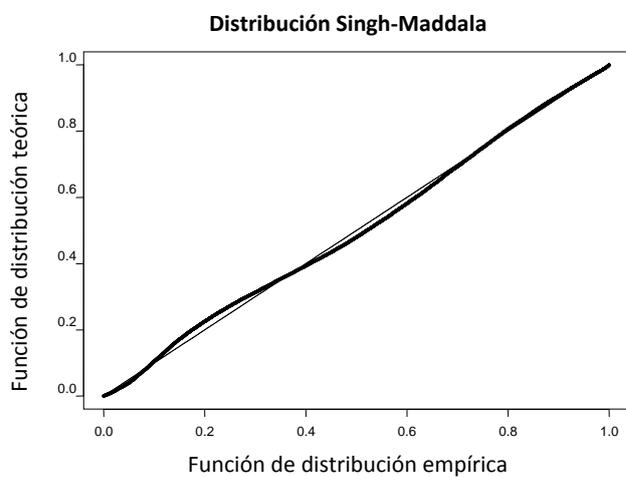
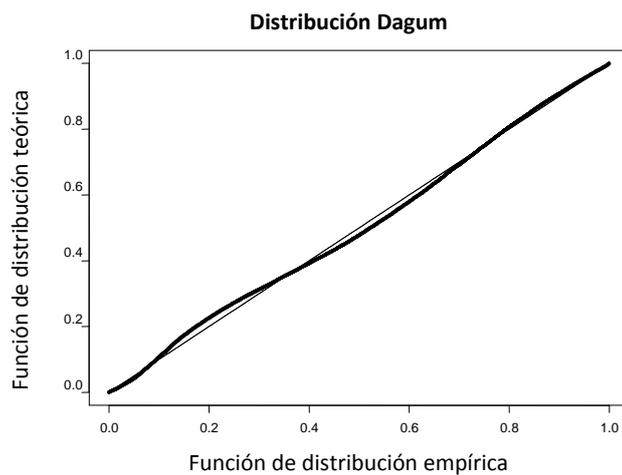
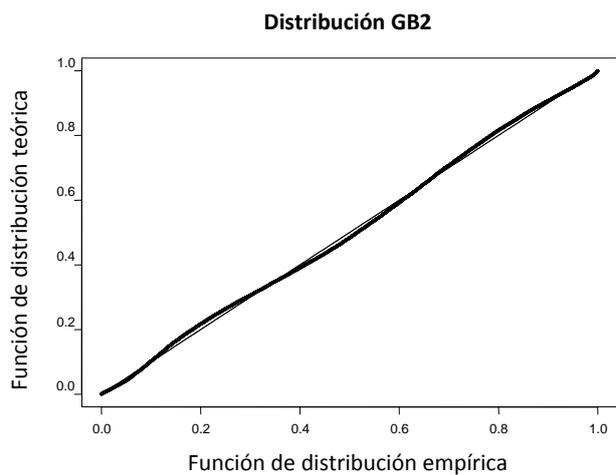


Gráfico A.3. Gráficos de probabilidad de los ajustes de la distribución del salario mensual.

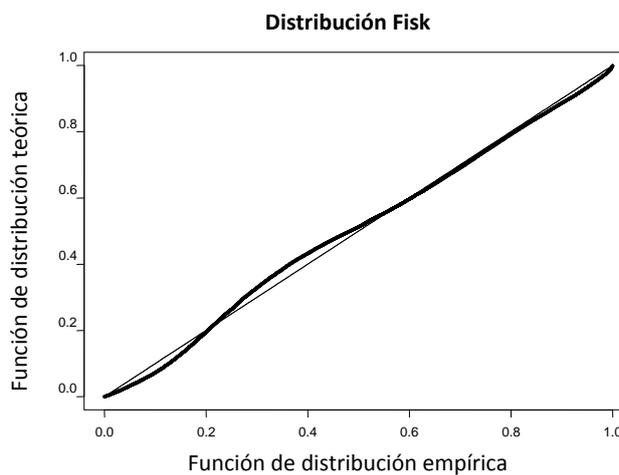
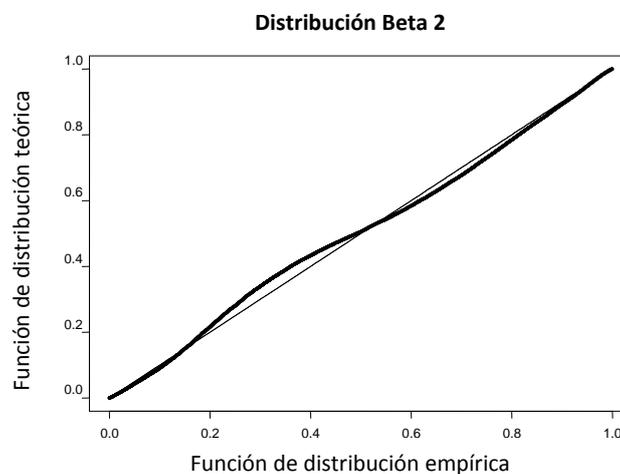
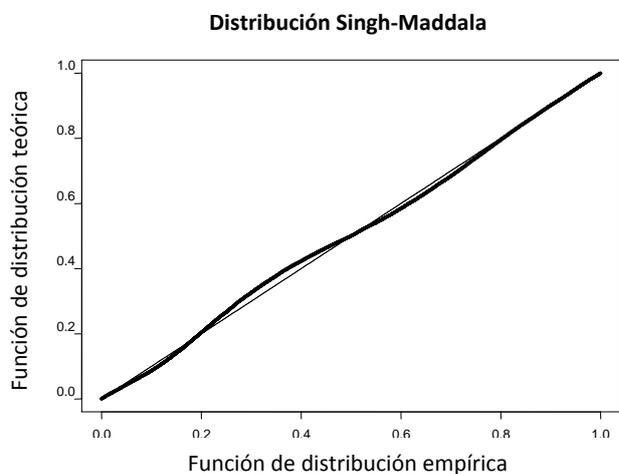
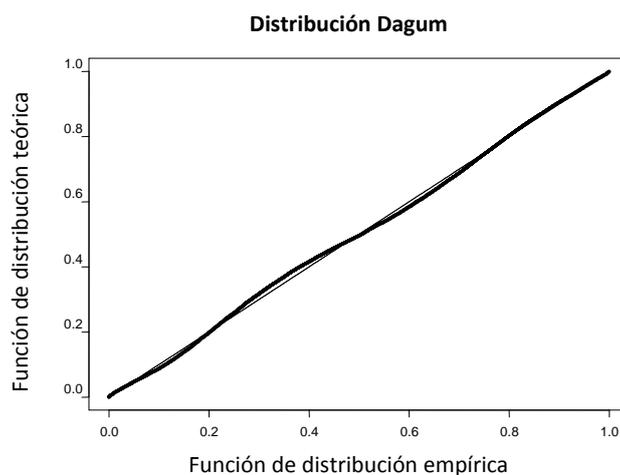
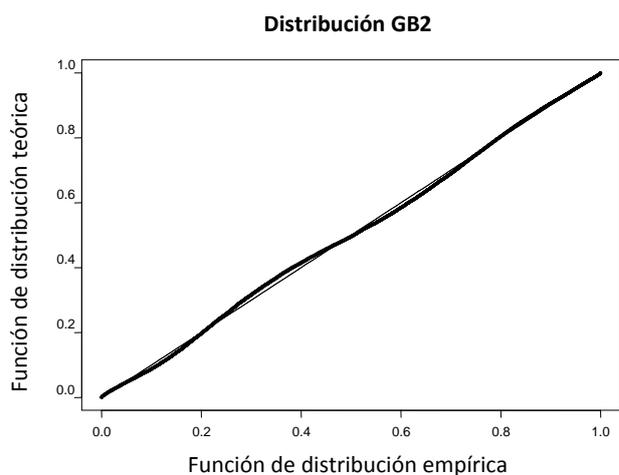


Gráfico A.4. Gráficos de probabilidad de los ajustes de la distribución del salario anual.

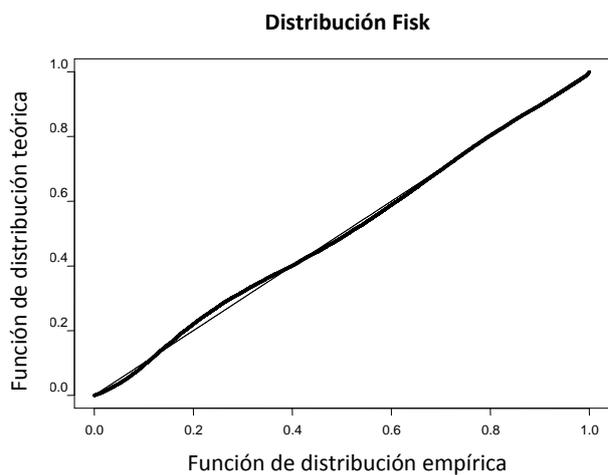
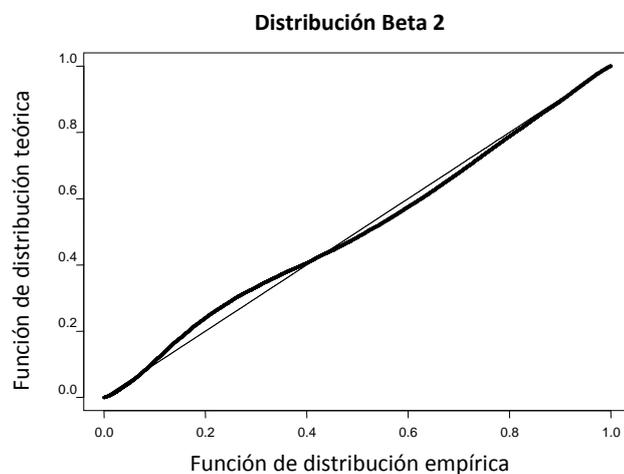
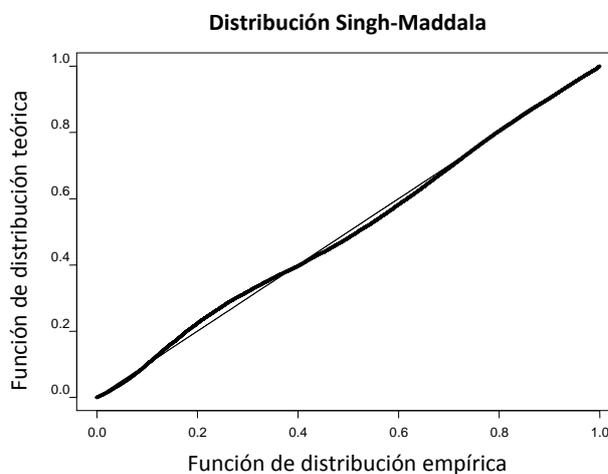
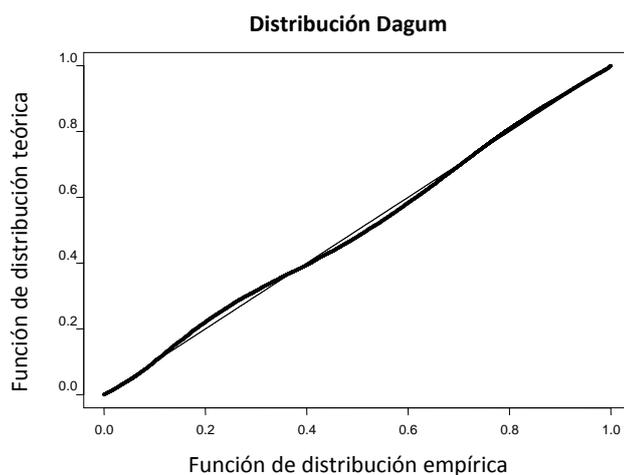
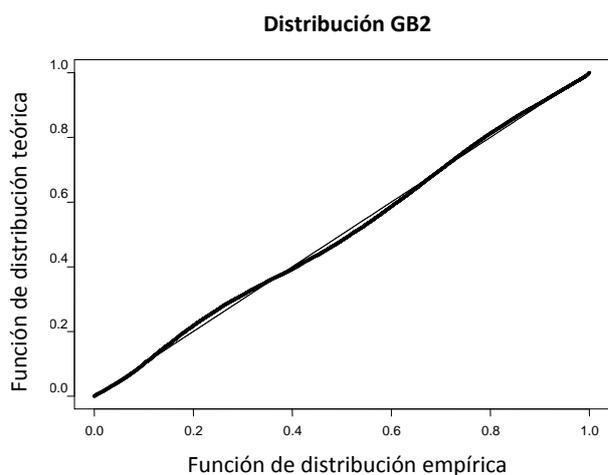


Gráfico A.5. Gráficos de probabilidad de los ajustes de la distribución del salario anual en contratos indefinidos.

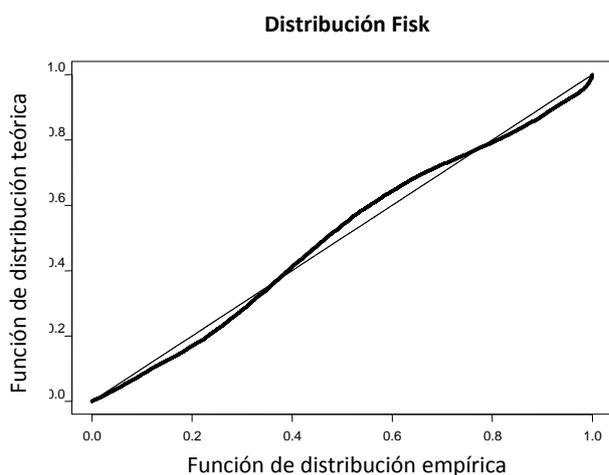
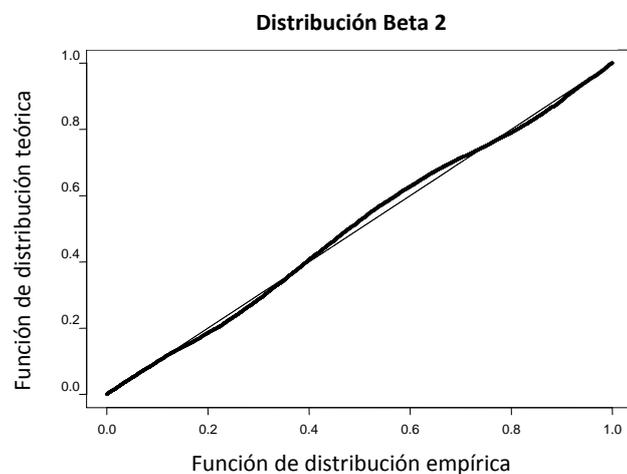
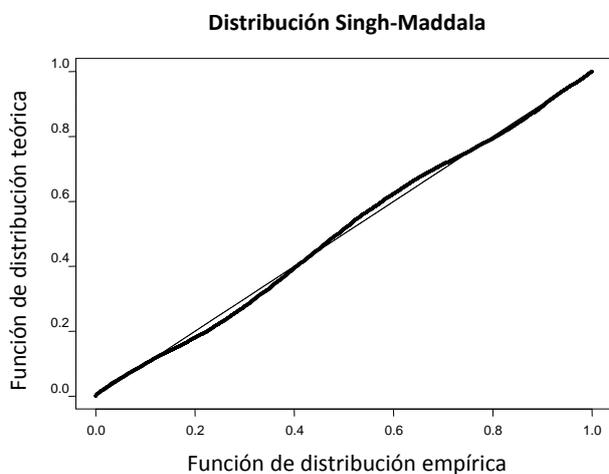
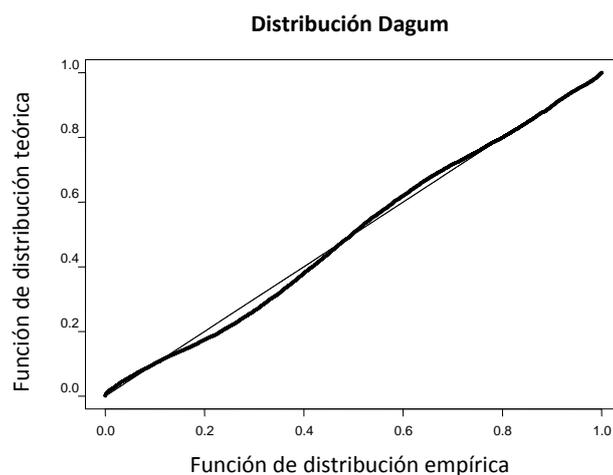
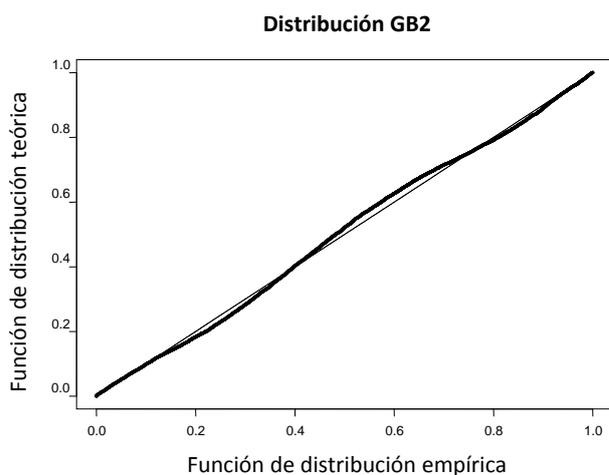


Gráfico A.6. Gráficos de probabilidad de los ajustes de la distribución del salario anual en contratos temporales.