

ESTIMACIONES PARAMETRICAS Y NO PARAMETRICAS DE LA DISTRIBUCION DEL INGRESO DE LOS OCUPADOS DEL GRAN BUENOS AIRES, 1992-1997

Patricia Botargues¹ y Diego Petrecolla^{2°}

¹ Ministerio de Economía y Obras y Servicios Públicos

² Instituto y Universidad Torcuato Di Tella.

INTRODUCCION

La investigación en el área de la distribución del ingreso ha seguido dos corrientes. La primera estudia la participación de las remuneraciones a los factores productivos como porcentaje del ingreso nacional. La segunda estudia la distribución del ingreso de una población de unidades económicas. Esta población generalmente se define en base a características socioeconómicas o regionales (población urbana o rural, perceptores de ingresos, ocupados, etc.). A este enfoque se lo denomina distribución personal del ingreso y uno de los principales objetivos de investigación del mismo es el de proporcionar una descripción matemática de la distribución que se aproxime a la distribución observada.

Debe encontrarse entonces una función que describa la distribución personal del ingreso. Esta función puede describir toda la distribución o un rasgo particular de la misma. Tradicionalmente el rasgo más importante de la distribución personal ha sido la desigualdad de ingresos. Sin embargo, la literatura también ha prestado atención a la “cola inferior” para analizar problemas relacionados con la pobreza (Metcalf, 1969).

Según Dagum (1990) un “buen modelo” de distribución del ingreso es el resultado de una inferencia precisa de procesos económicos observados. Es un instrumento que sintetiza la información contenida en un número muy grande de observaciones. El mismo puede ser utilizado para análisis posteriores, investigación e interpretación socioeconómica y para analizar políticas que afecten la distribución del ingreso y la riqueza. Los parámetros del modelo se pueden utilizar en la especificación de modelos macroeconómicos de ciclos económicos, crecimiento económico y planificación para el desarrollo.

[°] Los autores agradecen especialmente al Dr. Camilo Dagum por su invaluable apoyo y por sus enriquecedores comentarios, como así también las sugerencias de los árbitros designados por Económica. Las opiniones expresadas son de los autores y no necesariamente de las instituciones a las que pertenecen.

Los modelos que se han utilizado con más frecuencia en la literatura son el de Pareto, el lognormal y el gamma. Usualmente se acepta al de Pareto como el modelo que mejor ajusta la distribución de los ingresos altos. Los modelos lognormal y gamma generalmente ajustan todo el rango de ingresos pero suelen no ajustar correctamente ambas colas de la distribución. Estudios empíricos realizados para EE.UU y Holanda obtuvieron mejores ajustes para el modelo gamma que para el modelo lognormal (Salem and Mount, 1974; Dagum, 1977b), pero los resultados obtenidos para la distribución gamma distan de ser completamente satisfactorios, ya que exageran la asimetría (Kakwani, 1980). En 1976 Singh y Maddala propusieron un modelo¹ que ajustó mejor que las distribuciones lognormal y gamma datos de ingresos de los EE.UU en 1960 y 1969. Los modelos propuestos por Dagum en 1977 superaron la bondad de ajuste de los modelos lognormal, gamma y de Singh-Maddala para la distribución de los ingresos familiares de los EE.UU en 1960, 1969 (Dagum 1980b) y 1978 (Dagum 1983).

Una forma alternativa de obtener estimadores de la densidad de los ingresos es mediante métodos no paramétricos. De estos métodos uno de los más antiguos y utilizados es el conocido histograma, el cual tiene ciertas deficiencias relacionadas con la continuidad que llevaron a considerar otras técnicas. El estimador de núcleos (kernels) fue propuesto por Rosenblatt en 1956 y Parzen en 1962 (DiNardo, Fortin y Lemieux, 1996).

Los métodos paramétricos y no paramétricos son complementarios en el análisis de distribuciones. Härdle y Linton (1994) citan entre los principales usos de los procedimientos no paramétricos los siguientes:

- las técnicas de suavizado no paramétricas se pueden utilizar como un medio conveniente y sucinto para mostrar las características de una base de datos y por lo tanto para ayudar en la construcción de modelos paramétricos,
- se pueden usar para hacer diagnósticos de un modelo paramétrico estimado.

Este trabajo se divide en cuatro secciones. La primera reseña los modelos de distribución del ingreso que serán estimados en la segunda sección y presenta una breve discusión acerca de la relevancia de las estimaciones no paramétricas y de la metodología que se utiliza en la estimación que se presenta en la segunda sección. En ella se ajustan los modelos de distribución del ingreso lognormal, de Singh y Maddala y de Dagum tipos I y II² a los datos de ingresos de los ocupados³ de la EPH del GBA entre octubre de 1992

¹ La misma función de distribución fue deducida inicialmente por Burr en 1942, pero con distinto propósito y razonamiento (Singh y Maddala, 1976) y aplicada a datos de ingresos por Sargan en 1958 (Kakwani, 1980). Singh y Maddala la proponen en 1976 utilizando un enfoque alternativo de tiempos de falla.

² La distribución gamma fue ajustada estimando sus parámetros por el método de los momentos, por máxima verosimilitud utilizando dos aproximaciones: una sugerida por Thom en 1958 (Garibotti, 1996) y otra propuesta por Greenwood y Durand en

y octubre de 1997 y se estiman las funciones de densidad de los ingresos de los ocupados del GBA en el mismo período con el método no paramétrico de núcleos. La tercera sección contiene el análisis de la bondad de ajuste de los modelos estimados, una comparación entre los mismos y una comparación entre los resultados de las estimaciones paramétricas y las no paramétricas. La cuarta y última sección presenta las conclusiones del trabajo.

1. MODELOS DE DISTRIBUCIÓN DEL INGRESO Y METODO DE NUCLEOS

La distribución lognormal

En 1931 fue propuesta por Gibrat como modelo para describir una distribución unimodal de ingresos (Dagum, 1977b).

La manera más usual de introducir la distribución lognormal es: si Y es una variable aleatoria con distribución normal de parámetros μ y σ^2 , entonces la variable aleatoria $X = \exp(Y)$ tiene distribución lognormal dependiente de los mismos parámetros⁴. Por lo tanto, su fdp resulta:

$$f(x) = \frac{1}{\sqrt{2\pi}sx} e^{-0.5\left(\frac{\ln(x)-m}{s}\right)^2}, \quad x > 0$$

Sus momentos de orden r centrados en el origen son

$$E(X^r) = e^{rm+r^2s^2/2}$$

y el coeficiente de Gini se puede obtener como (Dagum, 1980)

$$G = 2 N(s\sqrt{2}; 0, 1) - 1$$

donde $N(y; 0, 1)$ representa la función de distribución acumulativa de una normal con media cero y varianza uno.

1960 (Kotz y Johnson, 1983), y también por mínimos cuadrados no lineales. El ajuste resultó en todos los casos no satisfactorio. Los resultados de la estimación (por el método de los momentos) para los ocupados de EPH para el año 1995 se muestran en Botargues y Petrecolla (1997).

³ Se define a los ocupados como el conjunto de personas que tiene, por lo menos, una ocupación (operacionalmente se considera que un individuo es ocupado cuando responde que ha trabajado por lo menos una hora en forma remunerada o 15 hs. de manera no remunerada durante la semana de referencia de la EPH). De este modo se reconocen cuatro categorías ocupacionales: patrón o empleador, trabajador por cuenta propia, obrero o empleado y trabajador sin salario (Encuesta Permanente de Hogares, 1984).

⁴ μ y σ^2 no representan parámetros de localización o escala, pero son de fácil interpretación y éste es uno de los motivos por los cuales se ha popularizado esta distribución (Cowell, 1995).

Los percentiles se deducen de

$$x_p = e^{\ln^{-1}(p;0,1)+m}$$

La distribución de Singh y Maddala

Es un modelo de tres parámetros que fue deducido especialmente para describir la distribución de los ingresos y otras distribuciones de formas similares. Es una generalización de la distribución de Pareto y la distribución de Weibull utilizada en análisis de falla de equipos (Singh y Maddala, 1976).

Su fdp es:

$$f(x) = abcx^{b-1}(1+ax^b)^{-c-1}, x>0$$

Los momentos centrados de orden r se pueden escribir como (Dagum, 1980)

$$E(X^r) = \frac{r}{ba^{r/b}} B(r/b, c-r/b)$$

donde $B(p,q)$ es la función beta de Euler definida para todo $p>0, q>0$ como

$$B(p,q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx$$

por lo tanto existen para $r<bc$.

El coeficiente de Gini se puede obtener mediante la expresión (Dagum, 1980)

$$G = 1 - B(c, 2c-1/b) / B(2c, c-1/b)$$

Los percentiles se deducen como

$$x_p = \left[\frac{\left(\frac{1}{1-p} \right)^{1/c} - 1}{a} \right]^{1/b}$$

El modelo de Dagum tipo I

El modelo de Dagum tipo I es un modelo de tres parámetros β, λ y $\delta > 0^5$ cuya fdp es:

$$f(x) = \mathbf{b} \mathbf{l} \mathbf{d} x^{-\mathbf{d}-1} (1 + \mathbf{l} x^{-\mathbf{d}})^{-\mathbf{b}-1}, \quad x > 0$$

Los momentos de orden r centrados en el origen se obtienen como

$$E(X^r) = \mathbf{b} \mathbf{l}^{r/\mathbf{d}} B(1 - r/\mathbf{d}, \mathbf{b} + r/\mathbf{d}), \quad r < \mathbf{d}$$

La fórmula para obtener el coeficiente de Gini es

$$G = -1 + B(\mathbf{b}, \mathbf{b}) / B(\mathbf{b}, \mathbf{b} + 1/\mathbf{d}), \quad \mathbf{d} > 1$$

Los percentiles se deducen mediante la expresión

$$x_p = \mathbf{l}^{1/\mathbf{d}} (p^{-1/\mathbf{b}} - 1)^{-1/\mathbf{d}}$$

El modelo de Dagum tipo II

Es un modelo de cuatro parámetros cuya fdp es

$$0 < \alpha < 1; \quad \beta, \lambda, \delta > 0$$

$$f(x) = \begin{cases} (1 - \mathbf{a}) \mathbf{b} \mathbf{l} \mathbf{d} x^{-\mathbf{d}-1} (1 + \mathbf{l} x^{-\mathbf{d}})^{-\mathbf{b}-1}, & x > 0 \\ \mathbf{a}, & x = 0 \\ 0, & x < 0 \end{cases}$$

Los momentos centrados de orden r se obtienen mediante la expresión

$$E(X^r) = (1 - \mathbf{a}) \mathbf{b} \mathbf{l}^{r/\mathbf{d}} B(1 - r/\mathbf{d}, \mathbf{b} + r/\mathbf{d}), \quad r < \mathbf{d}$$

El coeficiente de Gini se calcula mediante

$$G = [\mathbf{a} + (1 - \mathbf{a})] [-1 + B(\mathbf{b}, \mathbf{b}) / B(\mathbf{b}, \mathbf{b} + 1/\mathbf{d})], \quad \mathbf{d} > 1$$

⁵ En los dos modelos de Dagum los parámetros tienen interpretaciones económicas bien definidas (Dagum, 1977b).

Los percentiles se deducen como

$$x_p = \begin{cases} \mathbf{I}^{1/d} \left\{ \left[\frac{(1-\mathbf{a})}{(p-\mathbf{a})} \right]^{1/b} - 1 \right\}^{-1/d}, & p > \mathbf{a} \\ 0, & p \leq \mathbf{a} < 1 \end{cases}$$

Estimación no paramétrica: Método de núcleos

La idea en la que se basa este método es la siguiente. Si X es una variable aleatoria con función de distribución F y densidad f , entonces en cada punto de continuidad x de f se tiene que

$$f(x) = \lim_{h \rightarrow 0} \frac{1}{2h} (F(x+h) - F(x-h)).$$

Dada una muestra X_1, \dots, X_n proveniente de la distribución F , para cada h fijo, $F(x+h) - F(x-h)$ se puede estimar por la proporción de observaciones que yacen en el intervalo $(x-h, x+h)$. Por lo tanto, tomando h pequeño, un estimador natural de la densidad es

$$\hat{f}_{n,h}(x) = \frac{1}{2hn} \#\{X_i : X_i \in (x-h, x+h)\},$$

donde $\#A$ indica el número de elementos del conjunto A .

Otra manera de expresar este estimador es considerando la función de peso w definida como

$$w(x) = \begin{cases} 1/2 & \text{si } |x| < 1 \\ 0 & \text{cc} \end{cases}$$

de manera que el estimador de la densidad f en el punto x se puede expresar como

$$\hat{f}_{n,h}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} w\left(\frac{x - X_i}{h}\right)$$

Pero este estimador no es una función continua, tiene saltos en los puntos $X_i \pm h$ y su derivada es 0 en todos los otros puntos. Por ello se ha sugerido reemplazar a la función w por funciones más suaves K , llamadas núcleos, lo que da origen a los estimadores de núcleos.

El estimador de núcleos de una función de densidad f calculado a partir de una muestra aleatoria X_1, \dots, X_n de dicha densidad se define como

$$\hat{f}_{n,h}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

donde la función K se elige generalmente entre las funciones de densidad conocidas y h es un parámetro llamado ventana, parámetro de suavizado o ancho de banda, el cual determina las propiedades estadísticas del estimador: el sesgo crece y la varianza decrece con h (Härdle y Linton, 1994). Es decir que si h es grande, los estimadores están sobresuavizados y son sesgados, y si h es pequeño, los estimadores resultantes están subsuavizados, lo que equivale a decir que su varianza es grande.

A pesar que la elección del núcleo K determina la forma de la densidad estimada, la literatura sugiere que esta elección no es crítica, al menos entre las alternativas usuales (Deaton, 1997). Más importante es la elección del tamaño de ventana. Deaton (1997) dice que en muchas aplicaciones un procedimiento adecuado es considerar diferentes tamaños de ventana, graficar las densidades asociadas y observar si los estimadores están subsuavizados o sobresuavizados. Sugiere además que debe preferirse subsuavizar al utilizar estos métodos gráficos. Para determinar un ancho de banda con el cual comenzar, una alternativa es calcular el ancho de banda óptimo si se supone que la densidad tiene una forma específica.

La ventana óptima en el sentido de minimizar el error medio cuadrático integrado (definido como la esperanza de la integral del error cuadrático sobre toda la densidad) fue calculada por Silverman (1986) y depende de la verdadera densidad f y del núcleo K . Al suponer que ambos la densidad y el núcleo son normales, la ventana óptima resulta

$$h^* = 1.06 \sigma n^{-1/5}$$

donde σ es la desviación standard de la densidad⁶. El mismo autor sugiere utilizar medidas robustas de dispersión en lugar de σ , con lo cual el ancho de banda óptimo se obtiene como

$$h^* = 1.06 \min(\sigma, 0.75 IQR) n^{-1/5}$$

donde IQR es el rango intercuartílico, es decir la diferencia entre los percentiles 75vo. y 25vo (Deaton, 1997).

⁶ Un método alternativo que elige el tamaño de la ventana minimizando aproximadamente estimadores de buena calidad del error medio cuadrático integrado puede verse en Sheather y Jones (1991).

2. ESTIMACIONES DE LA DENSIDAD DE LOS INGRESOS DE LOS OCUPADOS DE LA EPH PARA EL GBA. 1992 – 1997.

Estimación de los modelos

Factores importantes a tener en cuenta al hacer comparaciones de modelos son el número de intervalos de ingreso, la composición de los mismos, y el método de estimación de los parámetros, ya que esto puede tener impactos de consideración en las medidas de bondad de ajuste (McDonald y Mantrala, 1995). En este trabajo se utilizaron 16 tramos de ingreso, los mismos para todos los modelos, y los parámetros fueron estimados por un método de mínimos cuadrados no lineales, el cual minimiza la suma de las desviaciones cuadráticas entre las distribuciones observadas y las teóricas.

En los cuadros 1, 2, 3 y 4 se presentan, respectivamente, los parámetros estimados⁷ correspondientes a los modelos lognormal, Singh-Maddala y Dagum tipos I y II, las medias y las medianas de los ingresos de las distribuciones observadas y ajustadas, los coeficientes de Gini observados y los estimados, los estadísticos de Kolmogorov-Smirnov (K-S)⁸ y sus valores críticos asintóticos (VCA) al 5% y al 1%, y las sumas de los cuadrados de las desviaciones entre las distribuciones observadas y las ajustadas (SSE) tanto para las funciones de distribución acumulativas (fda) como para las funciones de densidad de probabilidad (fdp), para cada año entre 1992 y 1997 para los ingresos de la población ocupada del GBA.

Cuadro 1. Lognormal

Año	Parámetros Estimados		Media (\$)		Mediana (\$)		Coef. de Gini		K-S (%)	V.C.A. (5%/1%)	S.S.E.	
	μ	σ	Obs.	Est.	Obs.	Est.	Obs.	Est.			fda	fdp
92	-0,711	0,675	688,8	617,2	472,6	491,3	0,402	0,367	3,25	2,33/2,79	0,00443	0,00822
93	-0,609	0,739	777,6	714,5	526,6	543,7	0,409	0,398	2,92	2,27/2,72	0,00240	0,00425
94	-0,570	0,729	801,6	737,5	549,8	565,2	0,410	0,394	2,76	2,28/2,74	0,00254	0,00367
95	-0,656	0,763	779,6	694,0	509,8	518,9	0,440	0,410	3,03	2,28/2,73	0,00380	0,00400
96	-0,670	0,777	766,5	691,8	512,8	511,6	0,439	0,417	2,20	2,34/2,81	0,00383	0,00506
97	-0,666	0,771	756,7	691,4	519,2	513,8	0,425	0,414	2,27	2,22/2,67	0,00288	0,00347

⁷ Utilizando el método de mínimos cuadrados no lineales por medio de un proceso iterativo.

⁸ Es sabido que los tests de bondad de ajuste no tienen potencia suficiente contra alternativas que pueden ser de importancia en la práctica (Bickel y Doksum, 1977). Además, como se utiliza la distribución ajustada, el test de bondad de ajuste de Kolmogorov-Smirnov debe tomarse a título informativo solamente (Dagum, 1985).

Cuadro 2. Singh-Maddala

Año	Parámetros Estimados			Media (\$)		Mediana (\$)		Coef. de Gini		K-S (%)	V.C.A. (5%/1%)	S.S.E.	
	a	b	c	Obs.	Est.	Obs.	Est.	Obs.	Est.			fda	fdp
92	26,869	3,269	0,570	688,8	716,1	472,6	475,7	0,402	0,449	1,25	2,33/2,79	0,00049	0,00297
93	8,469	2,675	0,726	777,6	803,8	526,6	536,2	0,409	0,460	1,74	2,27/2,72	0,00086	0,00205
94	8,147	2,737	0,712	801,6	830,6	549,8	557,6	0,410	0,456	0,83	2,28/2,74	0,00038	0,00125
95	11,938	2,749	0,655	779,6	810,1	509,8	510,6	0,440	0,485	1,14	2,28/2,73	0,00030	0,00080
96	7,427	2,452	0,787	766,5	777,9	512,8	508,3	0,439	0,475	1,89	2,34/2,81	0,00112	0,00226
97	5,707	2,351	0,883	756,7	754,6	519,2	513,7	0,425	0,459	1,33	2,22/2,67	0,00088	0,00145

Cuadro 3. Dagum tipo I.

	Parámetros Estimados			Media (\$)		Mediana (\$)		Gini		K-S (%)	V.C.A. (5%/1%)	S.S.E.	
	β	λ^*	δ	Obs.	Est.	Obs.	Est.	Obs.	Est.			fda	fdp
92	2,882	0,059	2,047	688,8	698,6	472,6	474,1	0,402	0,434	0,91	2,33/2,79	0,00037	0,00113
93	1,767	0,133	2,051	777,6	795,1	526,6	533,9	0,409	0,452	1,55	2,27/2,72	0,00068	0,00158
94	1,730	0,145	2,083	801,6	817,9	549,8	555,9	0,410	0,445	0,95	2,28/2,74	0,00035	0,00091
95	1,821	0,122	1,983	779,6	782,4	509,8	510,4	0,440	0,467	1,60	2,28/2,73	0,00059	0,00138
96	1,230	0,184	2,093	766,5	756,8	512,8	509,2	0,439	0,463	1,87	2,34/2,81	0,00134	0,00285
97	1,073	0,212	2,189	756,7	742,5	519,2	514,7	0,425	0,452	1,49	2,22/2,67	0,00096	0,00151

* El parámetro de escala λ corresponde a un ingreso de \$ 1000.

Cuadro 4. Dagum tipo II.

	Parámetros Estimados			Media (\$)		Mediana (\$)		Gini		K-S (%)	V.C.A. (5%/1%)	SSE		
	α	β	λ^*	δ	Obs.	Est.	Obs.	Est.	Obs.			Est.	fda	fdp
92	0,006	2,883	0,059	2,047	688,82	698,76	472,60	474,63	0,402	0,437	1,00	2,33/2,79	0,00038	0,00113
93	0,009	2,104	0,111	2,019	777,58	799,04	526,60	532,99	0,409	0,456	1,60	2,27/2,72	0,00064	0,00127
94	0,013	2,192	0,114	2,041	801,63	822,61	549,80	554,87	0,410	0,452	0,84	2,28/2,74	0,00025	0,00054
95	0,030	4,089	0,054	1,877	779,56	799,29	509,80	507,40	0,440	0,484	0,71	2,28/2,73	0,00017	0,00040
96	0,039	2,405	0,094	1,971	766,47	771,37	512,80	507,71	0,439	0,480	1,50	2,34/2,81	0,00071	0,00182
97	0,029	1,577	0,144	2,100	756,66	750,70	519,24	514,00	0,425	0,462	1,33	2,22/2,67	0,00072	0,00133

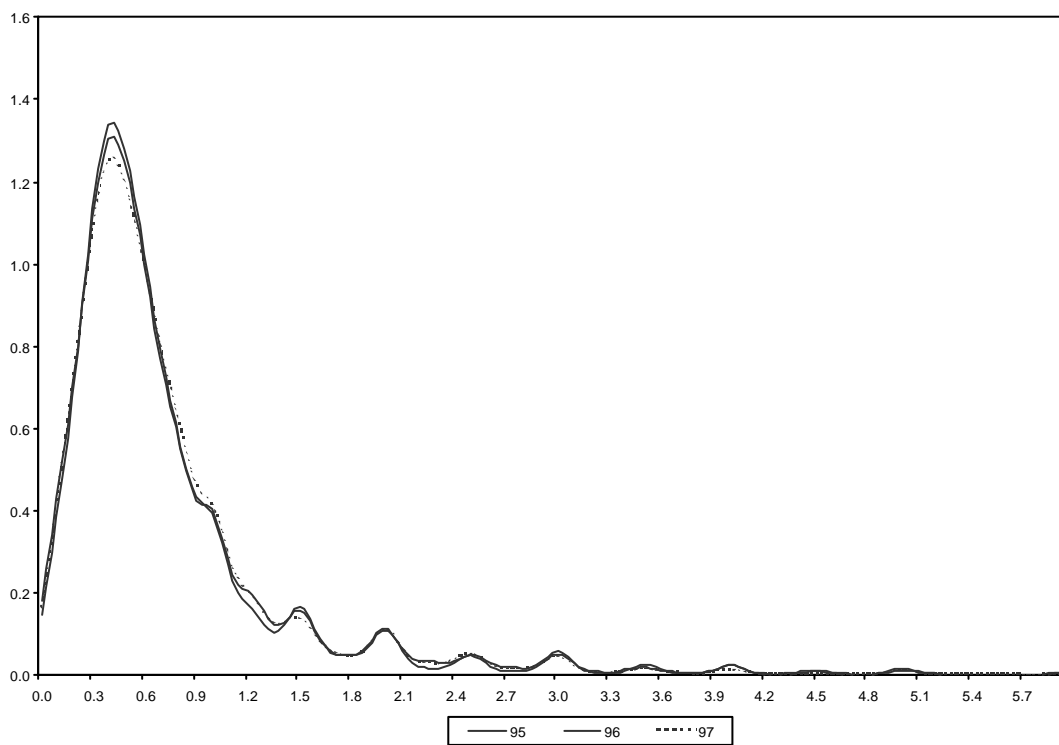
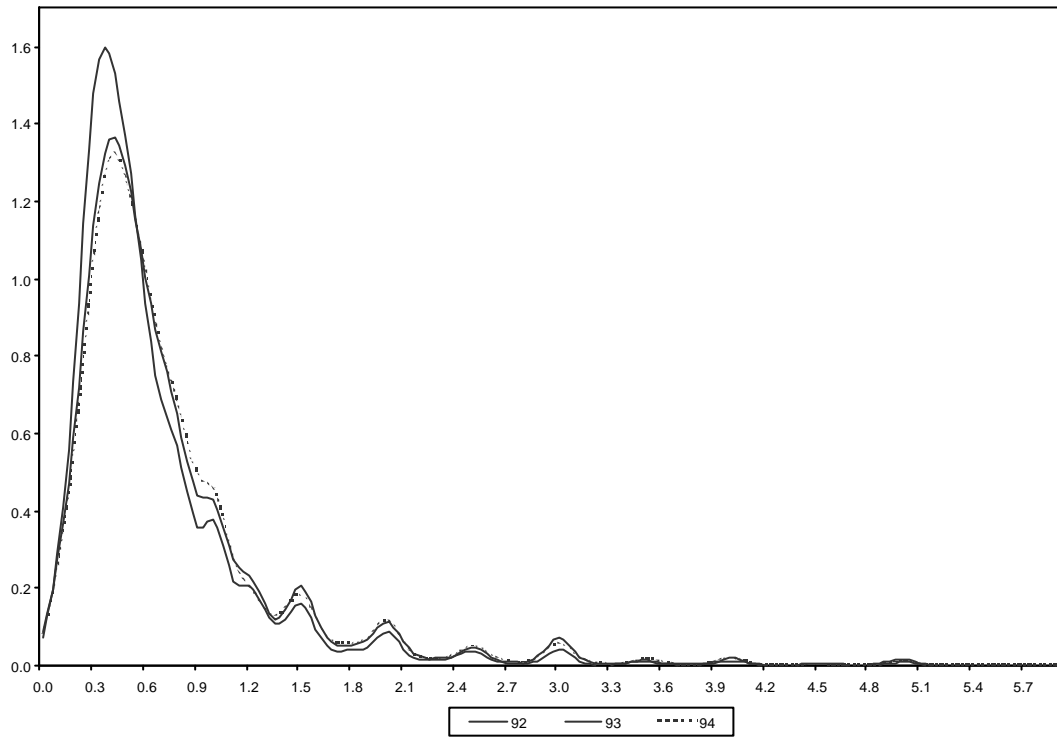
* El parámetro de escala λ corresponde a un ingreso de \$ 1000.

Estimación no paramétrica

En el Gráfico 1 se muestran las estimaciones de las funciones de densidad de los ingresos de los ocupados del GBA por el método de núcleos para las ondas de octubre de la EPH entre 1992 y 1997. Estas se realizaron considerando un núcleo K normal. Con la finalidad de elegir los tamaños de ventana se probaron anchos de banda mayores y menores a los óptimos, no encontrándose evidencia que justifique no utilizar los mismos. Los tamaños óptimos de ventana se calcularon suponiendo que los logaritmos de los ingresos se distribuyen normalmente. Los mismos son: para 1992, $h = 0.0719$; para 1993, $h = 0.0773$; para 1994, $h = 0.0806$; para 1995, $h = 0.0775$; para 1996, $h = 0.0783$; para 1997, $h = 0.0813$.

Gráfico 1: Estimaciones no paramétricas 1992-1997*

* En el eje de las abscisas se mide el ingreso de los ocupados del GBA en miles de pesos.



3. COMPARACION DE LA BONDAD DE AJUSTE DE LOS MODELOS ESTIMADOS

Para analizar la bondad del ajuste de los modelos en el Cuadro 5 se presentan las sumas de los cuadrados de las desviaciones entre las distribuciones observadas y las ajustadas (SSE) de las fda y de las fdp, las diferencias entre las medias y las medianas estimadas y observadas, los estadísticos de K-S y las diferencias entre los coeficientes de Gini observados y estimados⁹.

Cuadro 5: Bondad de ajuste

	Modelo	SSE		Dif. Medias (%)	Dif. Medianas (%)	K-S	V.C.A. (5% / 1%)	Dif. Coef. Gini (%)
		fda	fdp					
1992	Singh-Maddala	0,00049	0,00297	4,0	0,7	1,25	2,33/2,79	11,7
	lognormal	0,00443	0,00822	-10,4	4,0	3,25		-8,7
	Dagum I	0,00037	0,00113	1,4	0,3	0,91		8,0
	Dagum II	0,00038	0,00113	1,4	0,4	1,00		8,7
1993	Singh-Maddala	0,00086	0,00205	16,7	13,5	1,74	2,27/2,72	14,4
	lognormal	0,00240	0,00425	3,7	15,0	2,92		-0,7
	Dagum I	0,00068	0,00158	15,4	13,0	1,55		12,4
	Dagum II	0,00064	0,00127	16,0	12,8	1,60		13,4
1994	Singh-Maddala	0,00038	0,00125	20,6	18,0	0,83	2,28/2,74	13,4
	lognormal	0,00254	0,00367	7,1	19,6	2,76		-2,0
	Dagum I	0,00035	0,00091	18,7	17,6	0,95		10,7
	Dagum II	0,00025	0,00054	19,4	17,4	0,84		12,4
1995	Singh-Maddala	0,00034	0,00081	17,6	8,0	1,14	2,28/2,73	20,9
	lognormal	0,00383	0,00404	0,7	9,8	3,03		2,0
	Dagum I	0,00059	0,00138	13,6	8,0	1,60		16,2
	Dagum II	0,00017	0,00039	16,0	7,4	0,71		20,4
1996	Singh-Maddala	0,00112	0,00226	12,9	7,6	1,89	2,34/2,81	18,4
	lognormal	0,00383	0,00506	0,4	8,3	2,20		3,7
	Dagum I	0,00134	0,00285	9,9	7,7	1,87		15,2
	Dagum II	0,00071	0,00181	12,0	7,4	1,50		19,4
1997	Singh-Maddala	0,00088	0,00145	-0,3	-1,1	1,33	2,22/2,67	8,0
	lognormal	0,00287	0,00347	-8,6	-1,0	2,27		-2,6
	Dagum I	0,00096	0,00151	-1,9	-0,9	1,49		6,1
	Dagum II	0,00072	0,00133	-0,8	-1,0	1,33		8,7

De acuerdo a los estadísticos de K – S, los modelos que ajustan satisfactoriamente las distribuciones observadas de ingresos de los ocupados del GBA son el de Singh-Maddala y los modelos de Dagum tipos I y II. El modelo lognormal sólo ajusta en los años 1996 y 1997.

⁹ Botargues y Petrecolla (1997) compararon los deciles de los ingresos de los ocupados, obtenidos de la estimación de los modelos lognormal, Singh-Maddala, gamma, y Dagum tipos I y III para 1995. Una comparación de los deciles de los ingresos de los ocupados, estimados por los modelos de Dagum tipos I y II entre los años 1992 y 1996 puede verse en Botargues y Petrecolla (1998).

Para decidir cuál modelo elegir, Dagum (1980a) propone seguir el enfoque de Cramer para testear si la reducción en la suma de las desviaciones al cuadrado entre la fdp observada y ajustada es significativa. En el Cuadro 6 se presentan los resultados del test aplicado.

El análisis del Cuadro 6 revela que el modelo de Singh-Maddala ajusta significativamente mejor que el modelo lognormal entre 1992 y 1997. Cuando se compara el modelo de Singh-Maddala con el modelo de Dagum tipo I se observa que las SSE del primero es mayor que la del segundo entre los años 1992 - 1994, y es menor entre 1995 – 1997. Sin embargo, el resultado de la aplicación del test de Cramer indica que las diferencias entre los modelos no son significativas. Las SSE del modelo de Dagum tipo II son siempre menores que las correspondientes al modelo de Singh-Maddala. El test de Cramer revela que estas diferencias son significativas al 5 % solamente para los años 92, 94 y 95. Las SSE del modelo de Dagum tipo II son siempre menores que las correspondientes al modelo tipo I excepto para el año 1992¹⁰, para el cual debe preferirse el modelo tipo I por ser más parsimonioso. El test F indica que las diferencias son significativas al 5 % para todos los años salvo 1993, y que para el año 1995 también lo son al 1 %.

En los Gráficos 2.1 a 2.6 se presentan las estimaciones paramétricas y no paramétricas de la distribución de los ingresos de los ocupados del GBA para cada uno de los años entre octubre de 1992 y octubre de 1997. Las estimaciones no paramétricas revelan por un lado una menor concentración en los tramos más bajos de los ingresos que la que muestran los modelos estimados. Esto puede deberse a que el método de núcleos tiende a sesgar los resultados en las colas. En particular, en la cola inferior, al utilizar información desbalanceada (más observaciones a la derecha que a la izquierda) para estimar la densidad para un ingreso muy bajo, el estimador suele sesgarse hacia arriba. Por otro lado, en la parte central de la distribución se observa que las estimaciones no paramétricas que las estimaciones no paramétricas muestran una mayor concentración del ingreso que la que surge de las distribuciones paramétricas.

¹⁰ Una comparación más detallada entre las estimaciones de los modelos de Dagum tipos I y II para los años 1992-1996 puede consultarse en Botargues y Petrecolla (1998).

Cuadro 6. Ocupados GBA. Modelos estimados de distribución del ingreso. Análisis de la varianza de la SSE. 1992 – 1997.

	SSE	df	MSE	F
92 Singh-Maddala	0,002974	12	0,000248	
Lognormal	0,008217	13	0,000632	
Dagum tipo I	0,001126	12	0,000094	
Dagum tipo II	0,001131	11	0,000103	
Diferencia LN-(S-M)	0,005243	1	0,005243	21,16
Diferencia (S-M)-Dtl	0,001848			2,64
Diferencia (S-M)-DtII	0,001843	1	0,001843	7,43
93 Singh-Maddala	0,002055	12	0,000171	
Lognormal	0,004254	13	0,000327	
Dagum tipo I	0,001577	12	0,000131	
Dagum tipo II	0,001275	11	0,000116	
Diferencia LN-(S-M)	0,002199	1	0,002199	12,84
Diferencia (S-M)-Dtl	0,000478			1,30
Diferencia (S-M)-DtII	0,000780	1	0,000780	4,55
Diferencia Dtl - DtII	0,000302	1	0,000302	2,60
94 Singh-Maddala	0,001255	12	0,000105	
Lognormal	0,003674	13	0,000283	
Dagum tipo I	0,000905	12	0,000075	
Dagum tipo II	0,000540	11	0,000049	
Diferencia LN-(S-M)	0,002419	1	0,002419	23,14
Diferencia (S-M)-Dtl				1,39
Diferencia (S-M)-DtII	0,000715	1	0,000715	6,84
Diferencia Dtl - DtII	0,000365	1	0,000365	7,44
95 Singh-Maddala	0,000809	12	0,000067	
Lognormal	0,004043	13	0,000311	
Dagum tipo I	0,001384	12	0,000115	
Dagum tipo II	0,000399	11	0,000036	
Diferencia LN-(S-M)	0,003234	1	0,003234	48,01
Diferencia Dtl -(S-M)	0,000575			1,71
Diferencia (S-M)-DtII	0,000410	1	0,000410	6,08
Diferencia Dtl - DtII	0,000985	1	0,000985	27,17
96 Singh-Maddala	0,002264	12	0,000189	
Lognormal	0,005060	13	0,000389	
Dagum tipo I	0,002848	12	0,000237	
Dagum tipo II	0,001819	11	0,000165	
Diferencia LN-(S-M)	0,002797	1	0,002797	14,83
Diferencia Dtl -(S-M)	0,000584			1,26
Diferencia (S-M)-DtII	0,000444	1	0,000444	2,36
Diferencia Dtl - DtII	0,001029	1	0,001029	6,22
97 Singh-Maddala	0,001450	12	0,000121	
Lognormal	0,003470	13	0,000267	
Dagum tipo I	0,001510	12	0,000126	
Dagum tipo II	0,001330	11	0,000121	
Diferencia LN-(S-M)	0,002020	1	0,002020	16,72
Diferencia Dtl -(S-M)	0,000060			1,04
Diferencia (S-M)-DtII	0,000120	1	0,000120	0,99
Diferencia Dtl - DtII	0,001029	1	0,001029	8,51

Nota: df = grados de libertad = número de intervalos de clase - número de parámetros del modelo-1. Los valores críticos del estadístico F para niveles de significación del 5 y 1 % son $F_{0,95}(1,11)=4,84$, $F_{0,99}(1,11)=9,65$, $F_{0,95}(1,12)=4,75$, $F_{0,99}(1,12)=9,33$, $F_{0,95}(12,12)=2,69$ y $F_{0,99}(12,12)=4,16$.

Gráficos 2.1 a 2.6. Estimaciones paramétricas y no paramétricas de la distribución del ingreso. Octubre 1992 a octubre 1997.

Gráfico 2.1: Año 1992

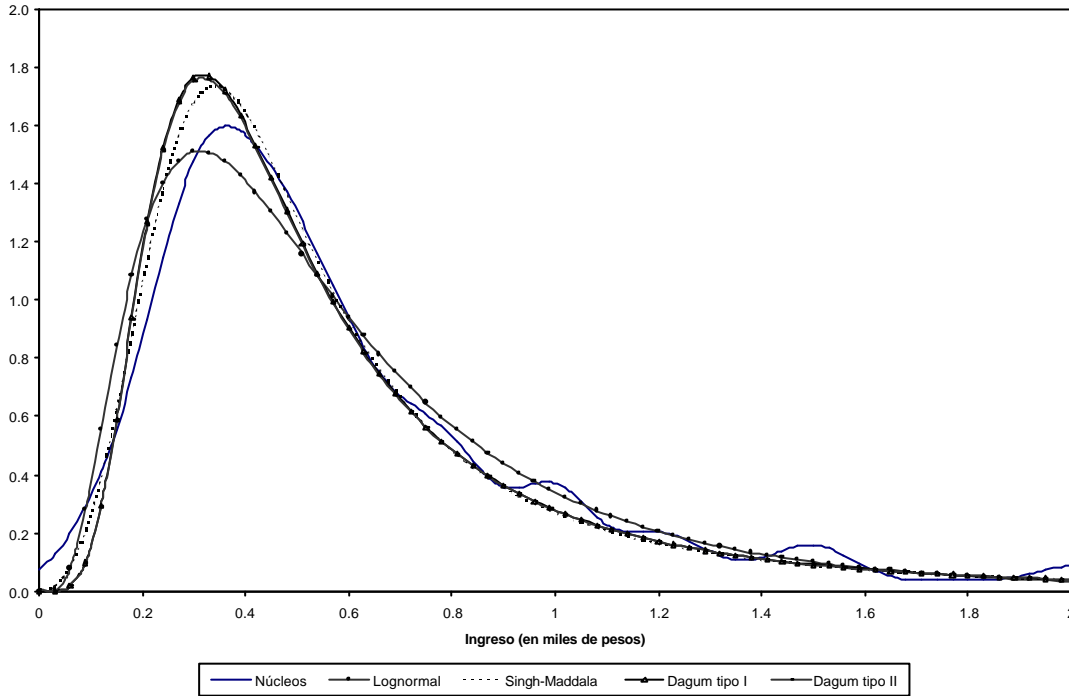


Gráfico 2.2: Año 1993

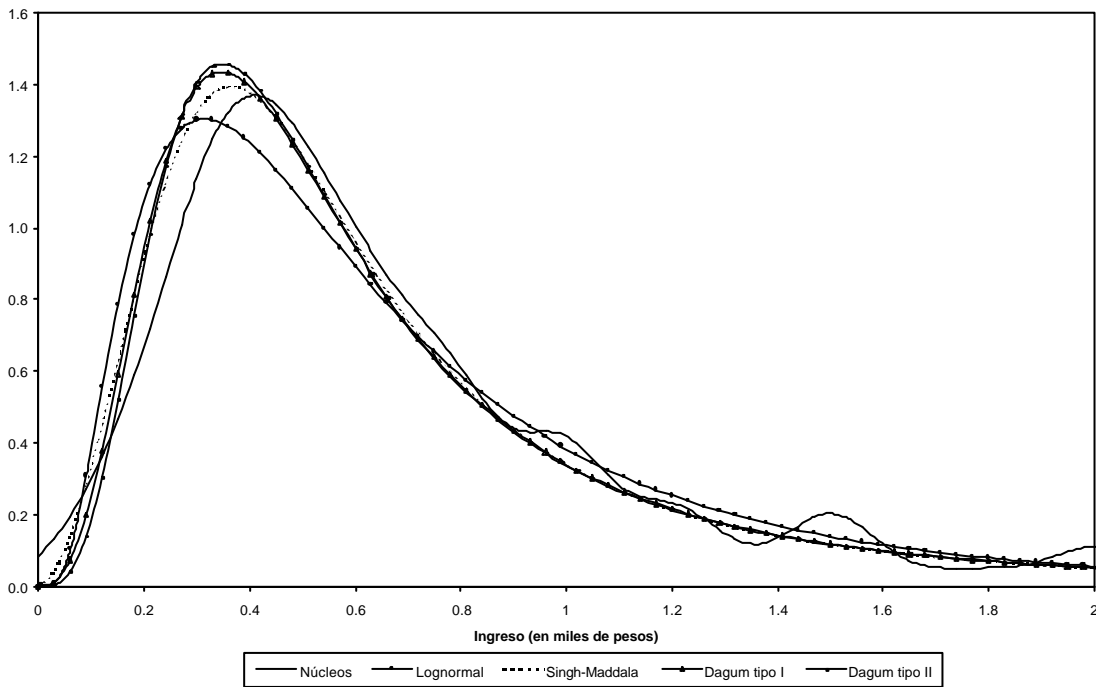


Gráfico 2.3: Año 1994

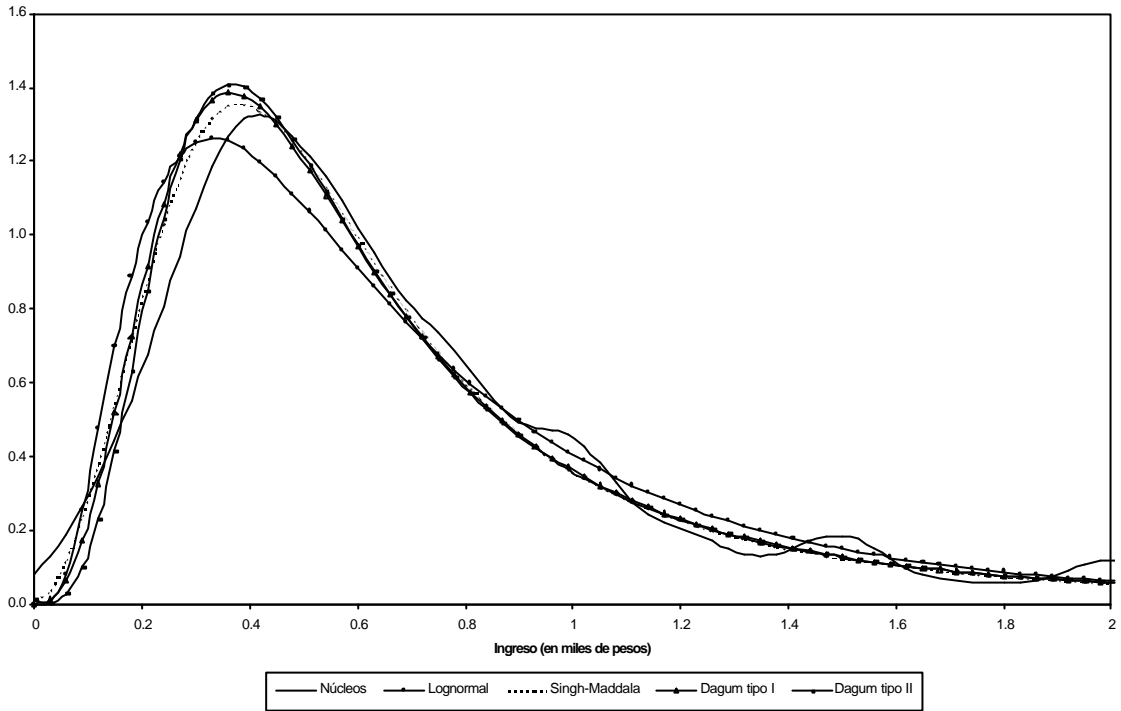


Gráfico 2.4: Año 1995

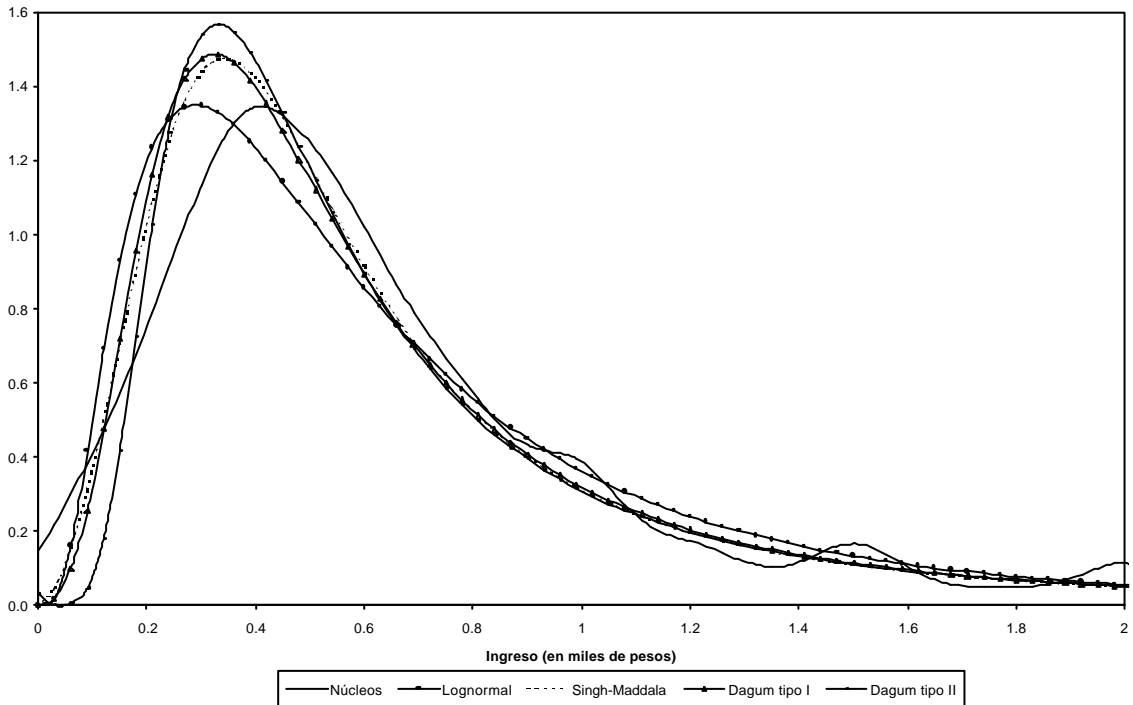


Gráfico 2.5: Año 1996

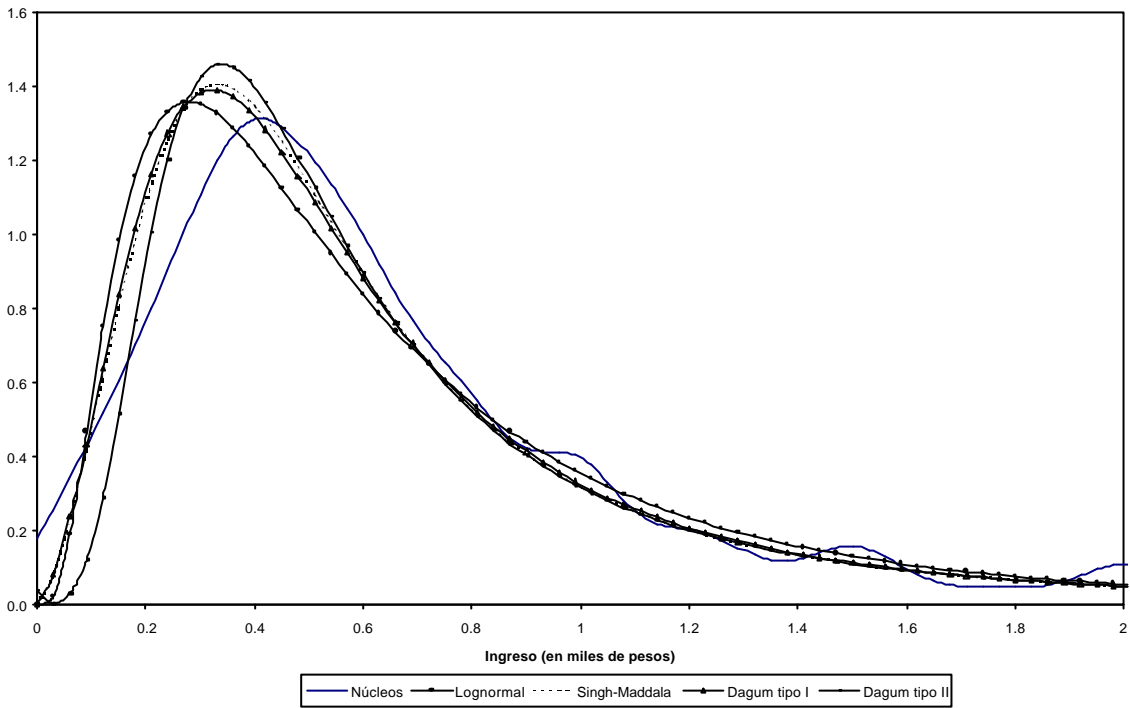
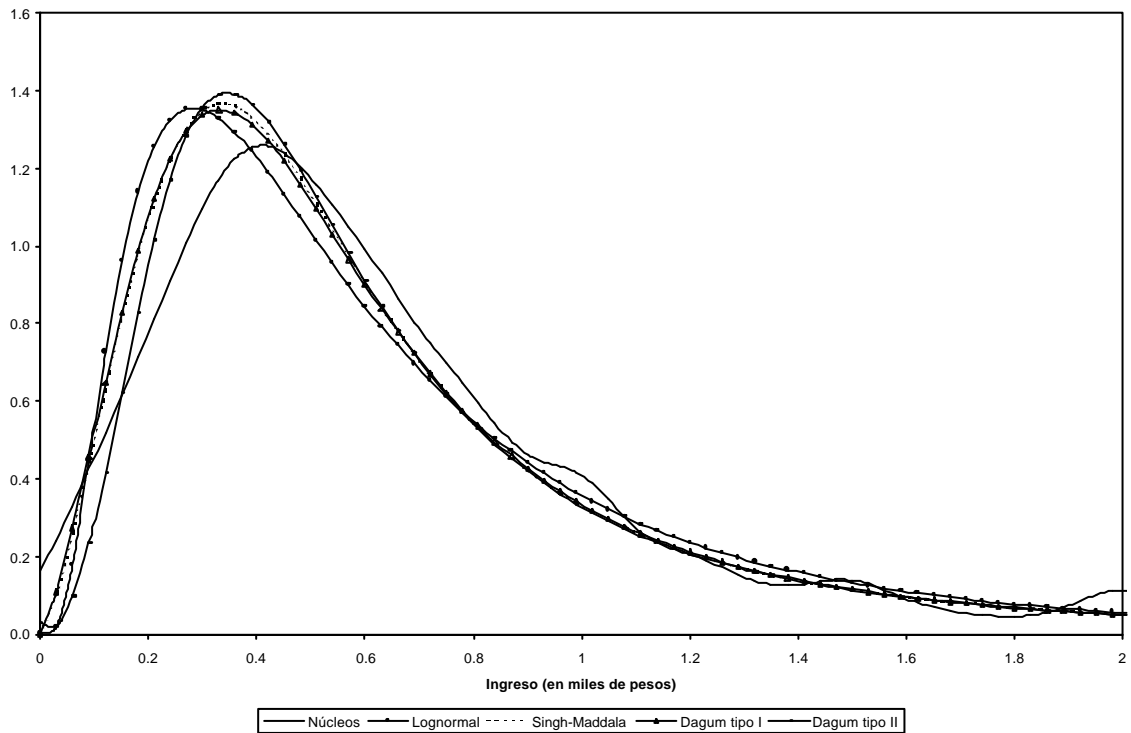


Gráfico 2.6: Año 1997



4. CONCLUSIONES

Con el objetivo de analizar la distribución del ingreso de la población ocupada en el Gran Buenos Aires se ajustaron los modelos de Dagum tipos I y II, lognormal y de Singh y Maddala y se estimaron, con el método de núcleos, las funciones de densidad de los ingresos de los ocupados con datos de la EPH correspondientes a las ondas de octubre entre 1992 y 1997.

Para comparar la bondad de ajuste de los modelos estimados se utilizaron la suma de los cuadrados de las desviaciones entre las distribuciones observadas y las ajustadas (SSE) de las fda y de las fdp, las diferencias entre las medias y las medianas estimadas y observadas, las diferencias entre los coeficientes de Gini observados y estimados y los estadísticos de Kolmogorov-Smirnov, y un test F aplicado por Dagum (1980a). Los modelos de Dagum tipos I y II y de Singh y Maddala ajustan satisfactoriamente las distribuciones observadas de ingresos entre los años 1992-1997. El modelo lognormal ajusta solamente las distribuciones analizadas en los años 1996 y 1997. Las estimaciones no paramétricas revelan por un lado una menor concentración en los tramos más bajos de los ingresos que la que muestran los modelos estimados. Por otro lado, en la parte central de la distribución se observa que las estimaciones no paramétricas muestran una mayor concentración del ingreso que la que surge de las distribuciones paramétricas.

Bibliografía:

Bickel, P.J. and Doksum, K.A. (1977): *Mathematical Statistics*. San Francisco: Holden Day.

Botargues, P. y Petrecollo, D. (1997): Funciones de Distribución del Ingreso en Gran Buenos Aires, Argentina, 1990 – 1996. Presentado en el XXV Coloquio Argentino de la SAE, Mar del Plata, noviembre de 1997.

Botargues, P. y Petrecollo, D. (1998): Funciones de distribución del ingreso y afluencia económica relativa para ocupados según nivel de educación en GBA, Argentina, 1992-1996. Documento de Trabajo Instituto Torcuato Di Tella (en prensa).

Cowell, F. A. (1995): *Measuring Inequality*. Second Edition. Prentice – Hall.

Cramer, J.S. (1978): A Function for Size Distribution of Incomes: Comment, *Econometrica*, March, Vol. 46, No. 2.

Dagum, C. (1977a): El modelo log-logístico y la distribución del ingreso en la Argentina, *El Trimestre Económico*, Vol. XLIV (4), No. 176.

Dagum, C. (1977b): A New Model of Personal Income Distribution: Specification and Estimation, *Economie Appliquée*, Vol. XXX, No. 3.

Dagum, C. (1980 a): The Generation and Distribution of Income, the Lorenz Curve and the Gini Ratio, *Economie Appliquée*, Vol. XXXIII, No.2.

Dagum, C. (1980 b): Inequality Measures Between Income Distributions with Applications, *Econometrica*, vol. 48, No. 7, November.

Dagum, C. (1983): Income Distribution Models and Income Inequality Measures, in *Encyclopedia of Statistical Sciences*, vol. 4, John Wiley & Sons.

Dagum, C. (1985): Analyses of Income Distribution and Inequality by Education and Sex in Canada, *Advances in Econometrics*, vol. 4, pp. 167-227.

Dagum, C. (1987): Measuring Economic Affluence Between Populations of Income Receivers, *Journal of Business & Economic Statistics*, January, Vol.5, No. 1.

Dagum, C. (1990): Generation and Properties of Income Distribution Functions, In *Studies in Contemporary Economics*, C. Dagum and M. Zenga (Eds.), Springer-Verlag Berlin Heidelberg.

Dagum, C. and Chiu, K. (1991): User's Manual for the program EPID (Econometric Package for Income Distribution) for Personal Computers. Revised Version, Ottawa: Statistics Canada.

Deaton, A. (1997): *The Analysis of Household Surveys. A Microeconomic Approach to Development Policy*. The World Bank. The Johns Hopkins University Press.

DiNardo, J., Fortin, N. and Lemieux, T. (1996): Labor Market Institutions and the Distribution of Wages, 1973-1992: a Semiparametric Approach. *Econometrica*, Vol. 64, No.5. September.

Encuesta Permanente de Hogares. Resultados Preliminares. INDEC, abril 1984.

Garibotti, G. (1996): Métodos Estadísticos para Estimar los Parámetros de una Distribución Pluviométrica. Tesis de Magister Scientiae - Biometría, Facultad de Agronomía, UBA.

Härdle, W. and Linton, O. (1994): Applied Nonparametric Methods. Handbook of Econometrics, Volume IV, Chapter 38. Elsevier Science.

Kakwani, N. C. (1980): Income Inequality and Poverty, Methods of Estimation and Policy Applications. Oxford University Press.

Kotz, S. And Johnson, N.L. (1983): "Gamma Distribution". Encyclopedia of Statistical Sciences. John Wiley & Sons, pp. 292-297.

McDonald, J.B. and Mantrala, A. (1995): The Distribution of Personal Income: Revisited, *Journal of Applied Econometrics*, Vol. 10, 201-204.

Metcalf, C. E. (1969): The Size Distribution of Personal Income During the Business Cycle, *American Economic Review*, 59, pp. 657-669.

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

Sheather, S. J. and Jones, M. C. (1991): A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation, *Journal of the Royal Statistical Society*, B, 53, No.3.

Silverman, B. W. (1986): Density Estimation for Statistics and Data Analysis, London and New York, Chapman and Hall.

Singh, S.K. and Maddala, G.S. (1976): A Function for Size Distribution of Incomes, *Econometrica*, September, Vol.44, No.5.

ESTIMACIONES PARAMETRICAS Y NO PARAMETRICAS DE LA DISTRIBUCION DEL INGRESO DE LOS OCUPADOS DEL GRAN BUENOS AIRES, 1992-1997.

RESUMEN

Con el objetivo de analizar la distribución del ingreso de la población ocupada en el Gran Buenos Aires se ajustaron los modelos de Dagum tipos I y II, lognormal y de Singh y Maddala para los años 1992 a 1997 con los datos de ingresos de los ocupados de la Encuesta Permanente de Hogares. Además se estimó la función de densidad de los ingresos de los ocupados del GBA entre 1992 y 1997 con el método de núcleos, que a diferencia de los modelos presentados es no paramétrico.

Para comparar la bondad de ajuste de los modelos estimados se utilizó un conjunto de indicadores propuesto por Dagum (1980a). Los modelos de Dagum tipos I y II y de Singh y Maddala ajustan satisfactoriamente la distribución de ingresos observados entre los años 1992-1997. El modelo lognormal ajusta la distribución de ingresos analizada sólo en 1996 y 1997. Las estimaciones no paramétricas revelan: a) una menor concentración en los tramos más bajos de los ingresos que la que muestran los modelos estimados, b) una mayor concentración del ingreso en la parte central de la distribución que la que surge de las distribuciones paramétricas.

PARAMETRIC AND NON-PARAMETRIC ESTIMATIONS OF INCOME DISTRIBUTION IN GREAT BUENOS AIRES, 1992-1997

SUMMARY

In order to analyze the income distribution of the employed population in Great Buenos Aires the Dagum types I and II, Lognormal and Singh and Maddala Models were adjusted with 1992-1997 income data from the Permanent Household Survey. The probability density function was also estimated with the Kernel Method, which is a non parametric technique.

To compare the goodness of fit of the estimated models a set of indicators proposed by Dagum (1980a) was used. Dagum types I and II and Singh and Maddala Models fitted the data between 1992-1997. The Lognormal distribution fitted the data in 1996 and 1997. The non-parametric estimations showed less concentration for the lower income groups and a higher degree of concentration for middle income groups than the estimated models.