

EVOLUCIÓN DEL COEFICIENTE DE ZIPF PARA COLOMBIA EN EL SIGLO XX*

**Gloria Lucía Bernal Nisperuza
Carlos Enrique Nieto Trujillo**

Julio de 2005

Resumen

El cumplimiento de la ley de Zipf para la distribución del tamaño de las ciudades es uno de los temas de investigación en la literatura sobre economía urbana y regional. La ley de Zipf constituye un caso particular de la distribución de rango- tamaño (rank-size distribution), en la que el rango de una ciudad o región es inversamente proporcional a su tamaño. El objetivo de este estudio consiste en estimar el coeficiente de la distribución de rango-tamaño (o coeficiente de Zipf) para Colombia en diferentes periodos del siglo veinte, describir su evolución, y ofrecer posibles explicaciones para los cambios en el valor del coeficiente que tienen en cuenta hechos históricos ocurridos a lo largo del período que se analiza. Se encuentra que la ley de Zipf no se cumple para ninguno de los censos de población entre 1918 y 1993, y que el coeficiente de la distribución de rango tamaño decrece a lo largo del siglo veinte excepto el período entre 1973 y 1985. La caída del coeficiente antes del 1973 se explica principalmente por mejoras en la red nacional de transportes y el aumento de la participación de la industria en la economía. El leve aumento del coeficiente entre 1973 y 1985 se explica por la bonanza cafetera de la década de los 70 y por la disminución de la participación de la industria en el producto interno bruto. La caída del coeficiente entre 1985 y 1993 se explica por factores como el desplazamiento forzado de la población y el final de la bonanza cafetera en la década de los 80. La evidencia empírica es consistente con la migración de la población hacia los grandes centros urbanos del país.

Clasificación JEL: R10, R12

Palabras clave: Economía regional, Zipf, distribución de rango-tamaño, Colombia.

* Artículo presentado como requisito para obtener el título de Economista de la Pontificia Universidad Javeriana. Agradecemos la colaboración de Andrés Rosas como director del trabajo de grado, puesto que sin su ayuda no hubiera sido posible esta investigación.

Introducción

George Kingsley Zipf (1902-1950), profesor de Harvard, observó que la frecuencia de uso de la n -ésima palabra más frecuente en cualquier lenguaje natural es inversamente proporcional a n^1 . Posteriormente esta observación recibió el nombre de ley de Zipf, y tuvo aplicaciones en áreas del conocimiento como la biología, la medicina y las finanzas. En años recientes, algunos sectores de la literatura sobre economía urbana y regional han investigado el cumplimiento de la ley de Zipf y sus posibles consecuencias.

En el caso de la distribución del tamaño de la población de ciudades o regiones, el cumplimiento de la ley de Zipf implica que la probabilidad de encontrar una ciudad o región con un tamaño (s) superior a un valor S dado es proporcional a S^{-1} (ecuación 1).

$$(1) P(s>S) = \beta S^{-1}$$

Donde β es una constante positiva. La ley de Zipf constituye a su vez un caso particular de la denominada distribución de rango-tamaño (*rank-size distribution*), la cual está expresada en la ecuación (2). El rango R_j de la ciudad o región j corresponde al ordenamiento conforme al tamaño: se asigna rango 1 a la ciudad o región más poblada, rango 2 a la segunda ciudad o región con mayor población y así sucesivamente hasta la asignación de

¹ Lingüista norteamericano, profesor de Harvard desde finales de los años veinte. Se destacó por estudiar las restricciones en la expresión oral cotidiana y el uso del vocabulario en la expresión escrita. Su obra más conocida, *Human Behavior and the Principle of Least Effort*, fue publicada en 1949.

rango n para la n -ésima ciudad o región del ordenamiento. En este caso general, dicho rango es inversamente proporcional al tamaño, pero el exponente que determina la proporción puede ser una constante positiva diferente de 1.

$$(2) R_j = \alpha S_j^{-\zeta} e^{\mu}$$

$$(3) \ln R = \ln \alpha - \zeta \ln S + \mu$$

Donde α es una constante positiva y μ es el término de error de la regresión lineal. La estimación de la ecuación 3, que es una transformación logarítmica de la ecuación 2, mediante mínimos cuadrados ordinarios, arroja como uno de sus resultados el valor del parámetro ζ , o coeficiente de Zipf². Si la ley de Zipf se cumple (el coeficiente de Zipf es estadísticamente igual a 1) quiere decir que la ciudad con mayor población será exactamente k veces tan grande como la k -ésima ciudad más poblada; pero si dicho parámetro resulta ser menor que 1, entonces encontramos una amplia gama de tamaños de ciudad en una amplitud reducida del rango, lo cual implica que las ciudades más grandes son aún más grandes de lo que predice la ley de Zipf. Si el parámetro es superior a 1, ocurre lo contrario.

Establecer si la ley de Zipf es una regularidad empírica que se presenta en la economía urbana y regional cobra importancia, porque el hecho de que el coeficiente de Zipf sea estadísticamente igual a 1 implica restricciones para la construcción de modelos de

² Hill (1975) propone un estimador alternativo para calcular el coeficiente de Zipf porque demuestra que el estimador de MCO es sesgado para muestras pequeñas.

crecimiento regional³. Por otro lado, las variaciones del valor de ζ pueden tener como significado la existencia de factores que producen efectos de aglomeración en los grandes centros urbanos, o problemas de congestión que tienen consecuencias para la distribución de los tamaños de población de ciudades o regiones.

Cuando no se cumple la ley de Zipf, la distribución de probabilidad de las tasas de crecimiento de la población depende del tamaño inicial de ciudades o regiones (Gabaix, 1999). Este hecho se debería tener en cuenta a la hora de diseñar las políticas de desarrollo urbano y regional. Si el gobierno observa, por ejemplo, que la distribución de las tasas de crecimiento de la población de ciudades y regiones depende del tamaño inicial de las mismas, y que las ciudades más grandes crecen a tasas superiores a las de ciudades más pequeñas, puede diseñar políticas que eviten un posible problema de congestión en el largo plazo.

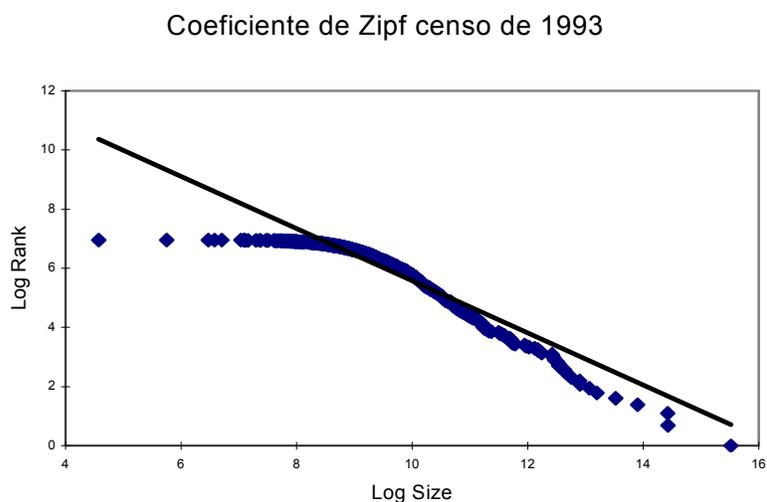
El objetivo de este estudio consiste en estimar el coeficiente de Zipf para la distribución de los tamaños de población de los municipios colombianos entre 1918 y 1993. Más adelante se ofrecen posibles explicaciones para el comportamiento del parámetro ζ relacionadas con hechos históricos que se presentaron a lo largo del siglo veinte⁴.

³ Más adelante en la revisión de literatura, se trata más a fondo este tema.

⁴ Utilizamos los censos poblacionales correspondientes a 1918, 1928, 1938, 1951, 1964, 1973, 1985 y 1993.

Aplicación de la ley de Zipf a municipios colombianos

A continuación se presenta una gráfica que ilustra los resultados de la estimación de la ecuación 3 con datos del censo de 1993 para 1053 municipios colombianos:



$$\ln R_j = 14.0190 - 0.8385 \ln S_j$$

(0.07546) (0.00780)

$$R^2 = 0.9159$$

El R cuadrado refleja una alta bondad de ajuste, y los coeficientes de la regresión son estadísticamente significativos al 1%⁵. La ley de Zipf no se cumple, en vista de que el coeficiente de Zipf es estadísticamente diferente de 1⁶. No obstante, el valor del coeficiente varía a lo largo de los censos de población del siglo veinte, e inclusive no es consistente con

⁵ Errores estándar entre paréntesis.

⁶ Se obtienen resultados similares cuando se corren regresiones en las que se suprimen el 5% y el 10% de los municipios más pequeños. En los anexos 1 y 2 se incluyen los resultados de estas regresiones y las pruebas para la hipótesis nula $H_0: \zeta = 1$.

la regularidad empírica que se ha presentado en algunos períodos para otros países como los Estados Unidos (Gabaix, 1999) y Dinamarca (Knudsen, 2001)⁷.

Dado que existen variaciones en el coeficiente de Zipf para los datos de población de Colombia en diferentes períodos de tiempo, es importante determinar qué fuerzas hacen que el parámetro ζ cambie. Para ello se revisan algunas explicaciones que se encuentran en la literatura relacionada.

Literatura Relacionada

En la literatura relacionada con la ley de Zipf, los estudios se dividen en teóricos y empíricos. Una de las investigaciones teóricas más influyentes en el área es la de Gabaix (1999). Un resultado de dicha investigación consiste en que una condición suficiente para el cumplimiento de la ley de Zipf es la ley de Gibralt, la cual se cumple cuando la distribución de probabilidad de las tasas de crecimiento para las ciudades es independiente del tamaño inicial de las mismas. En el rango de datos en el que la distribución se ajusta a la ley de Gibralt, automáticamente se ajusta a la ley de Zipf.

Un resultado adicional del modelo de Gabaix es que por lo menos en un cierto rango la distribución del tamaño de la población se aproxima a la ley de Zipf, lo cual implica una condición mínima para cualquier modelo de crecimiento local. En este modelo, el

⁷ Gabaix encuentra que el coeficiente de Zipf es estadísticamente igual a 1 para las 135 áreas metropolitanas estadounidenses incluidas en la muestra en 1991. Knudsen encuentra que la ley de Zipf se cumple para las 61 ciudades danesas de mayor población en el año 2000.

crecimiento de las ciudades se descompone en dos partes: el valor esperado de la tasa de crecimiento y choques estocásticos de varianza que, dadas las preferencias de la población, afectan las decisiones de los habitantes del país en la escogencia de una ciudad para vivir. En este caso, las decisiones dependen del comportamiento de choques aleatorios que afectan a las ciudades y que pueden ser de política, regionales, climáticos o relacionados con una industria específica.

Los choques de política (regional, nacional o federal) y climáticos son componentes de la varianza que no dependen del tamaño de las ciudades, porque afectan indistintamente a ciudades grandes y pequeñas. Por otro lado, los choques que afectan a una industria determinada pueden tener mayores efectos en las ciudades pequeñas, debido a que la variedad de industrias es menor y la dependencia de alguna industria en particular es mayor. Sin embargo, si el tamaño de las ciudades es lo suficientemente grande, los efectos de choques específicos de una industria son similares de una ciudad a otra y por lo tanto el crecimiento no depende significativamente del tamaño. Esto implica que se cumple la ley de Gibralt y se converge a la ley de Zipf.

En el modelo de Gabaix las ecuaciones que describen el comportamiento de los agentes se derivan de un proceso de maximización tanto de la utilidad de los consumidores, como de los beneficios de las firmas que producen con rendimientos constantes a escala. Esto permite que los resultados del modelo sean consecuencia de supuestos consistentes con la teoría clásica del consumidor y de la firma. Por otra parte, el modelo de Steindl (1965) parte del hecho de que las ciudades existentes crecen a una tasa G y que el nacimiento de

nuevas ciudades ocurre a una tasa V . El resultado del modelo es una distribución de rango-tamaño con exponente $K=V/G$. Para que la ley de Zipf se cumpla es necesario que V sea igual a G , y no existen razones aparentes para pensar que esta condición es cierta. Otro modelo, propuesto por Simon (1955), asume la existencia de una probabilidad P de que los pobladores que migran en una región formen una nueva ciudad, y de una probabilidad $(1-P)$ de que lleguen a una ciudad existente. El resultado del modelo es una distribución de rango-tamaño con exponente $K=1/(1-P)$. Este modelo tiene dos problemas: primero, para que la ley de Zipf se cumpla se requiere que P sea cercano a 0, y en este punto la convergencia del modelo es infinitamente lenta (Gabaix, 1999). Segundo, algunos valores de la distribución de rango-tamaño son consistentes con que la tasa de crecimiento del número de ciudades sea superior a la tasa de crecimiento de la población de las ciudades existentes, lo cual es poco plausible.

Además del componente teórico, los estudios de Gabaix, Steindl y Simon estiman el coeficiente de Zipf y corroboran empíricamente sus hipótesis. De igual forma, encontramos en la literatura estudios primordialmente empíricos del análisis de la ley de Zipf. Knudsen (2001) encuentra que para el caso de las 61 ciudades danesas de mayor tamaño la ley de Zipf se cumple. Gabaix y Ioannides (2002) proponen la estimación del parámetro ζ mediante el estimador de Hill, puesto que encuentran que mínimos cuadrados ordinarios arroja un estimador sesgado para muestras pequeñas⁸. Tong Soo (2004) estima el

⁸ El estimador de Hill es el estimador de máxima verosimilitud para la distribución de la forma $R_j = \alpha S_j^{-\zeta}$.

coeficiente de Zipf para 73 países y obtiene resultados diferentes⁹. La ley de Zipf se cumple para países como Egipto en 1996, Argentina en 1999 o Ucrania en 1998, y no se cumple en países como Francia en 1999 y Grecia en 1991¹⁰. El autor explica las diferencias observadas en el coeficiente de Zipf con variables políticas y geográficas, y encuentra que las primeras son significativas y tienen los signos esperados de acuerdo con la teoría, mientras que las segundas tienen signos contrarios a lo esperado y también son significativas¹¹.

Brakman et Al.(1999), de ahora en adelante BGMB, hace un aporte importante a la literatura al proponer explicaciones para el comportamiento histórico del coeficiente de Zipf en Holanda. En este estudio se examinan los exponentes de la distribución de rango-tamaño para tres períodos: 1600, 1900 y 1990. Los resultados obtenidos para el parámetro ζ fueron 1.81, 0.97 y 1.38, respectivamente y en todos los casos el R cuadrado fue superior a 0.95¹². BGMB busca en eventos históricos explicaciones para el cambio en los

⁹ Este estudio utiliza tanto el estimador de Hill, como el de Mínimos Cuadrados Ordinarios. El estimador de Hill rechaza la hipótesis de la ley de Zipf en 30 países, mientras que el de MCO la rechaza en 53 casos.

¹⁰ Los coeficientes de Zipf obtenidos mediante MCO son: Egipto (0.995), Argentina (1.043), Ucrania (1.024), Francia (1.4505), Grecia (1.413), y en los dos últimos casos se rechaza la ley de Zipf al 5% de significancia. En los cinco casos citados los resultados sobre la hipótesis son iguales para los dos estimadores.

¹¹ Tong Soo utiliza como variables políticas el grado de dictadura en el país, y el gasto total del gobierno. Se espera que la primera variable afecte negativamente el coeficiente de Zipf, porque las dictaduras tienden a concentrar la actividad económica en las ciudades grandes donde es más fácil obtener las rentas, y que la segunda lo afecte positivamente, ya que es una proxy de la mayor capacidad para garantizar la seguridad fuera de las ciudades grandes. Las variables de geografía son los costos de transporte y el grado de las economías de escala. Se espera que la primera variable afecte positivamente el coeficiente de Zipf, porque existen incentivos para que la población se ubique cerca de los centros agrícolas de producción y se formen más ciudades con menor tamaño. La segunda variable debería afectar negativamente el coeficiente, puesto que a mayor grado en las economías de escala, más atractivos resultan para la población los grandes centros urbanos.

¹² En BGMB los valores estimados del coeficiente de Zipf son 0.55, 1.03 y 0.72. Estos valores se obtienen de la regresión: $\ln S = \ln \alpha - \zeta \ln R + e$. Aquí se presentan los valores inversos porque las estimaciones se realizan con base en la ecuación 3.

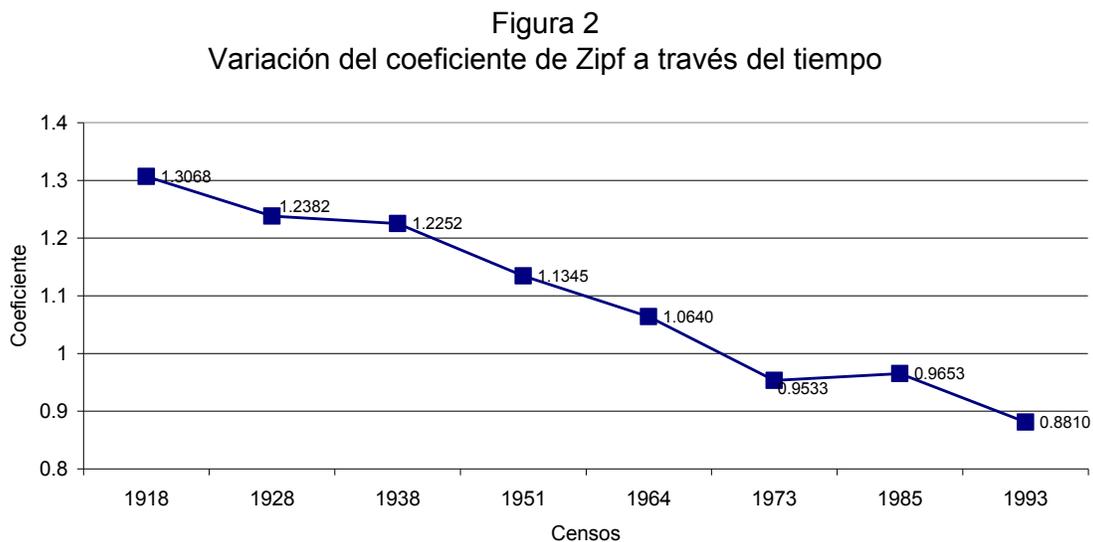
parámetros estimados: el alto valor para el coeficiente en 1600 se explica porque los altos costos de transporte y la poca actividad industrial de las ciudades dificultaban la integración nacional y la movilidad de la población en ese período. Estas condiciones hacen que la distribución del tamaño de las ciudades sea relativamente equitativa. Por otro lado, la creciente proliferación de las actividades industriales con altos retornos y la sofisticación de la red nacional de transporte en 1900 tienen como consecuencia un auge de las ciudades más grandes, y esto a su vez hace que la distribución sea más concentrada¹³. Finalmente, el aumento del exponente de la distribución de rango-tamaño en 1990 se explica por factores que aumentan la congestión en las grandes ciudades (polución o tráfico pesado). Estos cambios generan incentivos para que más personas decidan vivir en otras ciudades más pequeñas y de esta forma la distribución del tamaño de la población pase a ser más equitativa.

Los autores proponen un modelo de equilibrio general y hacen simulaciones a partir de la escogencia de parámetros de acuerdo con argumentos históricos. Esto permite mostrar que los resultados de las simulaciones tienen como consecuencia distribuciones de rango-tamaño similares a las que se obtuvieron empíricamente. BGMB usa argumentos históricos y estática comparativa para explicar las variaciones del parámetro ζ . Al igual que BGMB, este estudio busca dar explicaciones históricas a los cambios del coeficiente de Zipf para Colombia a lo largo del siglo veinte.

¹³ El valor del coeficiente para este caso es de 0.97, muy cercano a la ley de Zipf.

El coeficiente de Zipf a lo largo de la historia colombiana

Es posible replicar el ejercicio realizado para el censo de 1993 en los censos de 1918, 1928, 1938, 1951, 1964, 1973 y 1985¹⁴. La siguiente gráfica ilustra el comportamiento de los coeficientes estimados:



La ley de Zipf no se cumple para ninguno de los censos¹⁵. Esto implica que en ningún caso la ciudad de mayor tamaño es exactamente k veces tan grande como la k -ésima ciudad de mayor tamaño¹⁶.

La siguiente gráfica muestra la estimación del coeficiente de Zipf cuando se suprime el 10%, 50%, 90% y 95% de los municipios de menor tamaño¹⁷. La tendencia de ζ a lo largo

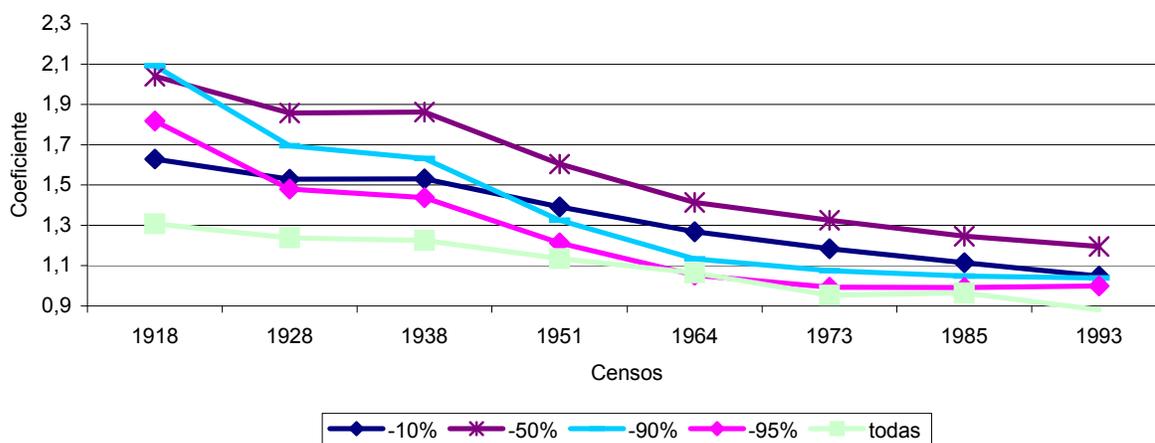
¹⁴ También se corrieron regresiones con las áreas metropolitanas del país y se obtuvieron resultados similares (ver anexo 3).

¹⁵ En todos los casos se rechaza la hipótesis de la ley de Zipf al 1% de significancia.

¹⁶ En todas las estimaciones el R cuadrado es superior a 0.88.

del período, independientemente del tamaño de la muestra, es decreciente. Este hecho se mantiene inclusive para los municipios considerados como ciudades¹⁸.

Figura 3
Coeficiente de Zipf con variaciones en la muestra



Cabe subrayar que las fallas en la medición de los datos de los censos de población pueden sesgar las estimaciones realizadas hasta ahora. Estas fallas son causadas por factores como alteraciones del orden público, condiciones climáticas adversas o problemas culturales de comunicación¹⁹. En el caso en que los datos de población estén subestimados para los municipios de menor tamaño, el coeficiente de Zipf estimado sería inferior al de las cifras reales. Lo contrario ocurre si dichos datos estuvieran sobreestimados. A pesar de que no es posible determinar la magnitud y la dirección del sesgo en las estimaciones, según el

¹⁷ También se realizó el ejercicio quitando el 5%, 20%, 30%, 40%, 60%, 70% y 80% de los municipios más pequeños y se obtuvo en todos los casos la misma tendencia.

¹⁸ Se excluyeron los municipios con población menor a 100.000 habitantes debido a que los restantes constituyen las áreas urbanas del país para el censo de 1993 (Cfr. DANE, 2001). Estas áreas urbanas representan alrededor del 5% de los municipios más grandes de la muestra.

¹⁹ Por ejemplo, en el censo de 1973, las serias inundaciones en el mes del censo afectaron las operaciones de recolección en algunas zonas rurales como el sur del departamento de Magdalena. En 1985, algunas áreas rurales de no más de 30 municipios en Antioquia, Caquetá, Magdalena, Valle del Cauca, Arauca y Amazonas se quedaron sin censar por motivos de orden público (DANE, 2001).

DANE (2001), el país cuenta con una buena base de datos censales que permiten y facilitan análisis demográficos y socioeconómicos de grandes dimensiones²⁰.

La disminución del exponente de la distribución del tamaño de los municipios en Colombia se debe principalmente a la migración poblacional que ocurrió por factores sociales, económicos y políticos a lo largo del siglo veinte. A comienzos del siglo veinte la población colombiana no se concentraba en sus más importantes centros urbanos como ocurre actualmente. El porcentaje de población rural superaba el 70% (Bejarano et. Al., 1989).

Antes de 1920 la columna vertebral de las comunicaciones era el Río Magdalena, complementada únicamente por una exigua red de ferrocarriles que solo superaba a Haití en número de kilómetros por cada 100.000 habitantes (Phanor, 2001). Una posible explicación consistente con la disminución del coeficiente de Zipf desde 1918 es la evolución del sistema de transportes, en especial la construcción y ampliación de ferrocarriles, y la construcción de carreteras desde la década de los 30 (Safford y Palacios, 2002). Estas vías interconectaron varias regiones del país y facilitaron la integración regional de los mercados y la movilidad de la población. Entre 1914 y 1934 la red ferroviaria en uso pasó de 1163 kms a 3262 kms (Bejarano et. Al., 1989). Este proyecto de construcción se

²⁰ Las tasas de cobertura calculadas por el DANE (2001) para la segunda mitad del siglo veinte son: 98.2% (1964), 92.8% (1973), 92.685% (1985) y 88.3% (1993).

financió con 25 millones de dólares correspondientes a la indemnización por la separación de Panamá, y con 160 millones de dólares por préstamos externos (Meisel y Bonet, 2001)²¹.

La red de caminos que se había concebido hacia finales de los años treinta estaba casi culminada para mitad de siglo e integraba a las zonas productivas más importantes del país. Esto favoreció cambios estructurales en la distribución de la población colombiana que se movilizaba en mayor número hacia nacientes centros de producción industrial²². Autores como Meisel y Bonet (2001), Jiménez y Sideri (1985) y Safford y Palacios (2002), coinciden en afirmar que la mejora en la infraestructura vial de Colombia durante el siglo veinte redujo los costos de transporte e incentivó las actividades industriales.

El desarrollo de la industria tuvo como resultado economías de escala y la incorporación de una creciente fuerza productiva, lo cual es consistente con la caída en el coeficiente de Zipf, ya que las actividades industriales incentivan el desplazamiento de la población hacia los grandes centros urbanos. Entre 1930 y 1945 la participación de la industria en el producto interno bruto aumentó en un 241% (Bejarano et. Al., 1989). Entre 1945 y 1973 la participación de la industria en el producto aumenta nuevamente del 14.4% al 22.8% (Safford y Palacios, 2002). Las economías de escala generadas por la industria y la reducción en los costos de transporte, tuvieron como resultado una reducción en los costos de los sectores productivos en general y de la industria en particular.

²¹ En el período citado, la mayoría de los kilómetros de la red ferroviaria se concentraba en los ferrocarriles de Antioquia, Cúcuta, La Dorada, Girardot, Pacífico y Tolima. Éstos comunicaban a las zonas cafeteras del país, en las cuales la red de ferrocarriles pasó de 783 a 1642 kms en uso.

²² Meisel y Bonet (2001), reconocen a Bogotá, Medellín, Cali y Barranquilla como los primeros centros de producción industrial a mediados del siglo veinte.

Entre 1973 y 1985 se detiene la caída del valor del coeficiente de Zipf. Una posible explicación es la caída de la participación de las actividades industriales en este período. Entre la segunda mitad de la década de los 70 y la primera década de los 80, muchas de las actividades productivas se ven desplazadas por actividades de bajo costo y alta rentabilidad como las relacionadas con el narcotráfico (Safford y Palacios, 2002). Adicionalmente, el sector industrial pierde participación en el PIB con el sector de los servicios²³. Ambos factores reducen la concentración de la población en los principales centros urbanos.

Un factor adicional que puede explicar la interrupción en la caída del coeficiente de Zipf es la bonanza cafetera ocurrida entre 1970 y 1979. Las actividades productivas relacionadas con el café, que se desarrollan principalmente en áreas rurales, atrajeron un mayor número de pobladores. Esto es consistente con la caída de la tasa de urbanización que registra el DANE (2001) para el período 1973-1985²⁴.

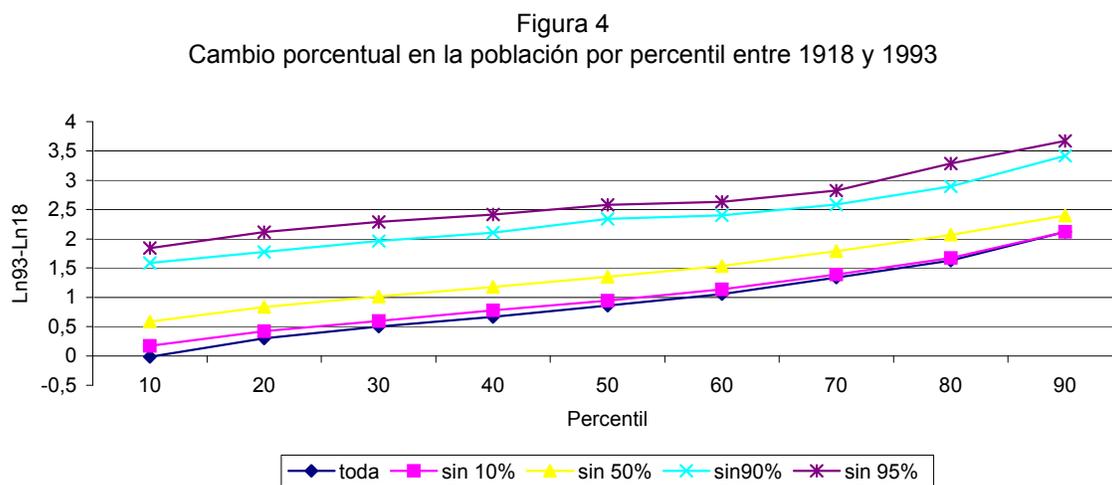
Entre los censos de 1985 y 1993 el coeficiente de Zipf muestra nuevamente una tendencia decreciente. No existe evidencia para explicar la disminución del coeficiente de Zipf en este período con argumentos como mejoras sustanciales en los sistemas de transporte o en la estructura industrial de las grandes ciudades. Por tal razón, se buscan posibles explicaciones como el final de la bonanza cafetera que ocurrió hacia 1979 y el desplazamiento forzado

²³ La participación promedio anual en el PIB para el sector industrial entre 1976 y 1980 fue del 22.8% y para el sector de los servicios del 49.4%. Para 1993 la participación del sector industrial fue del 18.1%, y la del sector de los servicios fue del 55.6% (Safford y Palacios, 2002).

²⁴ La tasa de urbanización cayó del 14.9% al 8.8% en el período citado.

causado por la agudización del conflicto armado en las zonas rurales del país. Alrededor de 642.000 personas se movilizaron entre 1985 y 1993 a causa de la violencia rural (CODHES, 2005). Esta migración mostró una tendencia creciente entre 1985 y 1989.

A continuación, se argumenta que la caída del coeficiente de Zipf a lo largo del siglo veinte se debe principalmente a la migración poblacional hacia los principales centros urbanos²⁵. La siguiente gráfica muestra la tasa de crecimiento de algunos percentiles para diferentes tamaños de muestra entre 1918 y 1993:

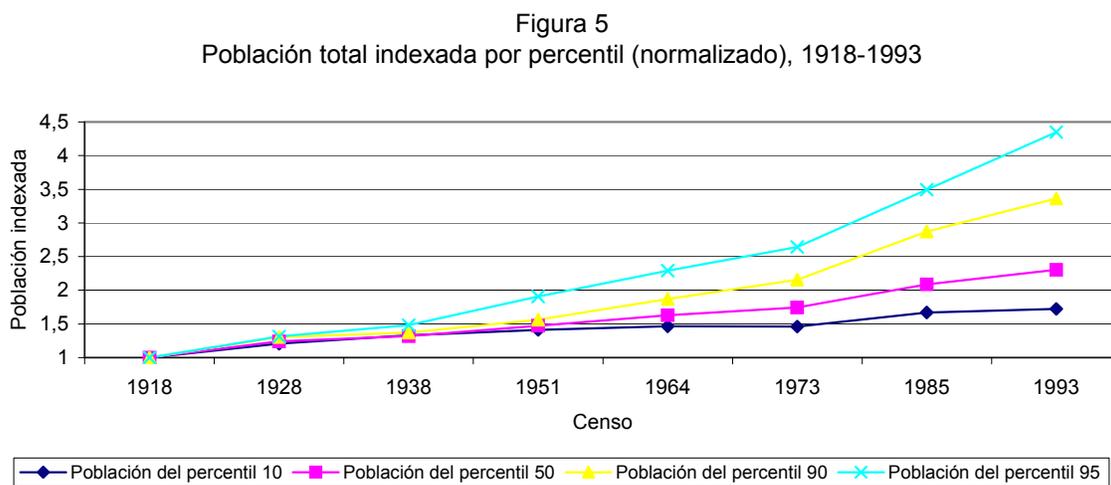


Como se observa en la gráfica, cada percentil creció más entre 1918 y 1993 que los demás percentiles de menor tamaño independientemente de la muestra. Esto implica que en los censos del siglo veinte los municipios de mayor tamaño crecieron más que los municipios

²⁵ Otro factor que podría explicar la caída del coeficiente de Zipf es la tasa de natalidad en los municipios de mayor tamaño. Sin embargo, si el aumento poblacional de la ciudad más grande del país (Bogotá), por ejemplo, se debiera a la tasa de natalidad, ésta hubiera tenido que ser del 4.96% anual entre 1951 y 1993. Este valor es casi tres veces superior al verdadero valor de la tasa de natalidad en dicho período, que fue de 1.78%.

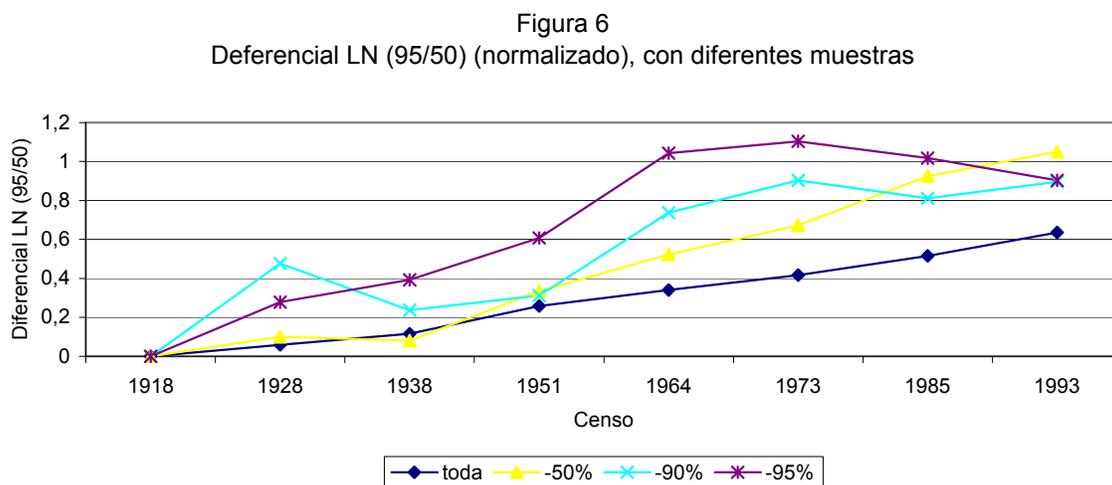
medianos y pequeños, y que los factores históricos que incentivaron la aglomeración en los grandes centros urbanos son consistentes con la evidencia empírica. En el anexo 4 se encuentran las gráficas de las tasas de crecimiento para diferentes períodos y para diferentes muestras, y en todos los casos los municipios de mayor tamaño mostraron el mayor crecimiento.

La siguiente gráfica, muestra la evolución de diferentes percentiles a lo largo de los censos de población:



En la gráfica se observa que la población de cada percentil creció más que los demás percentiles de menor tamaño. Este es otro indicador que es consistente con la migración hacia los grandes centros urbanos. En el anexo 5 se encuentra la evolución de los mismos percentiles para diferentes muestras.

La siguiente gráfica muestra la diferencia entre los percentiles 95 y 50 para diferentes tamaños de muestra a lo largo de los censos de población. La diferencia entre los percentiles es mayor cuando se toman los municipios de mayor tamaño, lo cual también es consistente con el desplazamiento de la población hacia los principales centros urbanos. En el anexo 6, se encuentran diferencias para distintos percentiles que muestran una tendencia similar



Conclusiones

La hipótesis de la ley de Zipf no se cumple para Colombia en ninguno de los censos de población del siglo veinte, y se rechaza para todos los casos con un nivel de significancia del 1%. Los valores del coeficiente de Zipf reflejan una tendencia decreciente salvo una leve interrupción entre 1973 y 1985, y son consistentes con hechos ocurridos a lo largo del siglo veinte que explican la aglomeración en las grandes ciudades. El valor del coeficiente de Zipf cae entre 1918 y 1973, lo cual probablemente se explica por mejoras en los costos

de transporte y por el aumento en la participación de las actividades industriales en el producto, que generó economías de escala y mayor concentración de la población en los grandes centros urbanos. Entre 1973 y 1985 la caída del coeficiente de Zipf se detiene, lo cual se explica por la bonanza cafetera de la década de 1970, y por el desplazamiento de las actividades industriales por actividades de bajo costo y alta rentabilidad como el narcotráfico y por el sector de los servicios. Entre 1985 y 1993 el coeficiente de Zipf cae nuevamente, y dado que no se encuentra evidencia para argumentar mejoras en los sistemas de transporte o aumentos en la participación del sector industrial en la producción nacional, dicha caída se explica posiblemente por el desplazamiento forzado a causa de la violencia, y por el final de la bonanza cafetera. La evidencia empírica es consistente con los argumentos a favor del desplazamiento de la población hacia los grandes centros urbanos. Inclusive, si se argumenta la existencia de factores de congestión en los principales centros urbanos, la magnitud de estos factores no fue suficiente para que hubiera un movimiento de población significativo de las grandes ciudades a los municipios de menor tamaño.

Referencias Bibliográficas

Bejarano, Jesús Antonio, Bernardo Tovar, Carlos Esteban Posada, Juan José Echavarría, Juan Felipe Gaviria, Guillermo Perry, Juan Manuel Santos, Charles Bergquist, Alberto Mayor Mora y José Olinto Rueda Plata, *Nueva Historia de Colombia* (1989), Colombia, Planeta.

Brakman, Steven, Harry Garretsen, Charles van Marrewijk y Marianne van den Berg, “The return of Zipf: towards a further understanding of the rank-size distribution”, *Journal of Regional Science* (1999) 39(1):183-213.

Consultoría para los derechos humanos y el desplazamiento (CODHES), “Comportamiento del desplazamiento 1985-2005” (2005), http://www.codhes.org.co/cifra/GraficoTendencias1985_2005.jpg.

DANE, *Homologación de los microdatos censales colombianos. 1964-1993* (2001), Bogotá, CIDS, Universidad Externado de Colombia.

DAPD, *Estadísticas históricas Santa Fe de Bogotá D.C. 1950-1999* (2000), Bogotá, Alcaldía Mayor de Santa Fe de Bogotá.

Di Guilmi, Corrado, Edoardo Gaffeo y Mauro Gallegati, “Power law scaling in the world income distribution”, *Economics Bulletin* (2003) 15(6):1-7.

Gabaix, Xavier, “Zipf’s law for cities: an explanation”, *The Quarterly Journal of Economics* (1999) 114(3):739-767.

Gabaix, Xavier, “Zipf’s law and the growth of cities”, *The American Economic Review* (1999) 89(2): 123-132.

Gabaix, Xavier y Yannis Ioannides, “The evolution of city size distributions”, *Handbook of Regional and Urban Economics, Volume 4* (2002), Amsterdam, Elsevier Science.

Hill, Bruce M., “A simple general approach to inference about the tail of a distribution”, *Annals of Statistics* (1975) 3(5): 1163-1174.

Jiménez, Margarita y Sandro Sideri, *Historia del desarrollo regional en Colombia* (1985), Bogotá, Fondo editorial CEREC.

Kali, Raja, “The city as a giant component: a random graph approach to Zipf’s law”, *Applied Economic Letters* (2003) 15:717-720.

Knudsen, Thorbjorn, “Zipf’s law for cities and beyond: the case of Denmark”, *American Journal of Economics and Sociology* (2001) 60(1): 123-146.

McGreevey, William Paul, *Historia económica de Colombia* (1975), Colombia, Tercer Mundo.

Meisel Roca, Adolfo y Jaime Bonet Morón, “La convergencia regional en Colombia: una visión de largo plazo, 1926-1995”, *Regiones, ciudades y crecimiento económico en Colombia* (2001) : 11-56.

Meisel Roca, Adolfo y Luis Armando Galvis, “El crecimiento de las ciudades colombianas y sus determinantes, 1973-1998”, *Regiones, ciudades y crecimiento económico en Colombia* (2001) : 57-90.

Overman, Henry y Yannis Ioannides, “Zipf’s law for cities: an empirical examination”, *CEP discussion paper* (2000), London School of Economics and Political Science.

Phanor James, Eder, *Colombia* (2001), Cali, Manuelita.

Safford, Frank Robinson y Marco Antonio Palacios, *Colombia: fragmented land, divided society* (2002), New York, Oxford University.

Simon, H., “On a class of skew distribution functions”, *Biometrika*, *XLII* (1955) : 425-440.

Steindl, J., *Random processes and the growth of firms* (1965), New York: Hafner.

Tong Soo, Kwok “Zipf’s law for cities: a cross country investigation”, *CEP discussion paper* (2001), London School of Economics and Political Science.

Apéndice

Anexo 1

Prueba F para verificar si el coeficiente de Zipf es estadísticamente diferente de 1				
Censo	Coeficiente	Valor de la		Probabilidad
1918	-1,30683006	234,0548	**	0,00000
1928	-1,23819583	149,1202	**	0,00000
1938	-1,22516335	122,7936	**	0,00000
1951	-1,1344969	34,86682	**	0,00000
1964	-1,06404665	28,93184	**	0,00000
1973	-0,95325797	94,97853	**	0,00000
1985	-0,96526523	146,0656	**	0,00000
1993	-0,8810181	428,0442	**	0,00000

** se rechaza la hipótesis nula (coeficiente de Zipf=1) al 1% de nivel de significancia

Anexo 2

Resultados de las regresiones de la ley de Zipf								
Censo	Obs	Coeficiente de correlación múltiple	R cuadrado	R cuadrado ajustado	Intercepto	Error típico del intercepto	Coeficiente	Error típico del coeficiente
1918	701	0,92377081	0,85335251	0,85314271	16,8888481	0,17822193	-1,30683006	0,02049055
1928	729	0,929175149	0,86336646	0,86317852	16,6132963	0,16308471	-1,23819583	0,01826851
1938	739	0,926289388	0,85801203	0,85781937	16,5948232	0,1651506	-1,22516335	0,01835859
1951	815	0,936437505	0,8769152	0,87676381	16,0333599	0,13620178	-1,1344969	0,0149067
1964	941	0,945930551	0,89478461	0,89467256	15,6484422	0,11012553	-1,06404665	0,01190716
1973	1029	0,932920097	0,87033991	0,87021366	14,7670934	0,10688179	-0,95325797	0,01148111
1985	1024	0,958913059	0,91951426	0,9194355	15,0615361	0,08490628	-0,96526523	0,00893308
1993	1053	0,946893927	0,89660811	0,89650973	14,3906683	0,08881064	-0,8810181	0,00922838

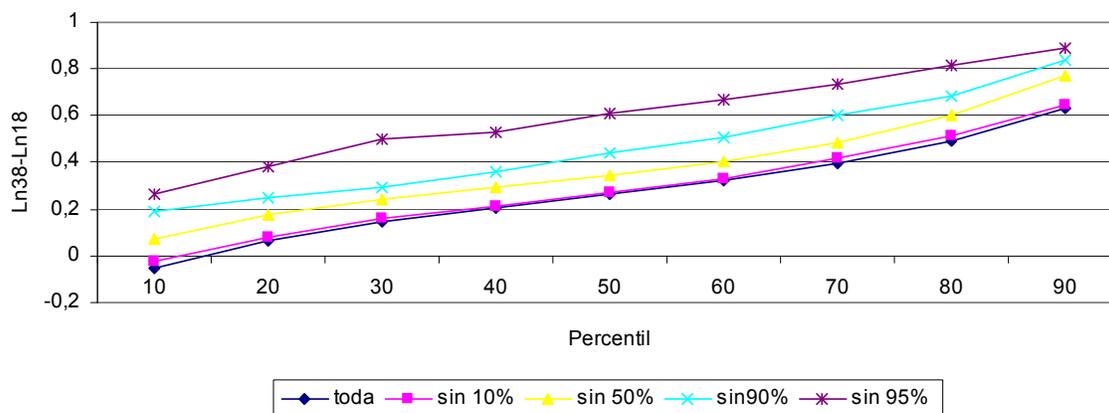
Resultados de las regresiones de la ley de Zipf sin el 10% de las observaciones								
Censo	Obs	Coeficiente de correlación múltiple	R cuadrado	R cuadrado ajustado	Intercepto	Error típico del intercepto	Coeficiente	Error típico del coeficiente
1918	631	0,972332784	0,94543104	0,94534429	19,7926833	0,13765767	-1,62869767	0,01560174
1928	656	0,976557902	0,95366534	0,95359449	19,2996746	0,11929175	-1,52803866	0,01317045
1938	665	0,978416076	0,95729802	0,95723361	19,4421581	0,11458026	-1,53041782	0,01255318
1951	733	0,983596192	0,96746147	0,96741696	18,466334	0,08749817	-1,39032913	0,00943064
1964	847	0,985974835	0,97214638	0,97211341	17,6227219	0,06938359	-1,26840787	0,00738594
1973	926	0,986076267	0,9723464	0,97231648	17,0049424	0,06220016	-1,18344351	0,00656564
1985	922	0,988107508	0,97635645	0,97633075	16,5660732	0,05530032	-1,11541292	0,00572261
1993	948	0,986213888	0,97261783	0,97258889	16,0845156	0,05603423	-1,04840256	0,00571933

Anexo 3

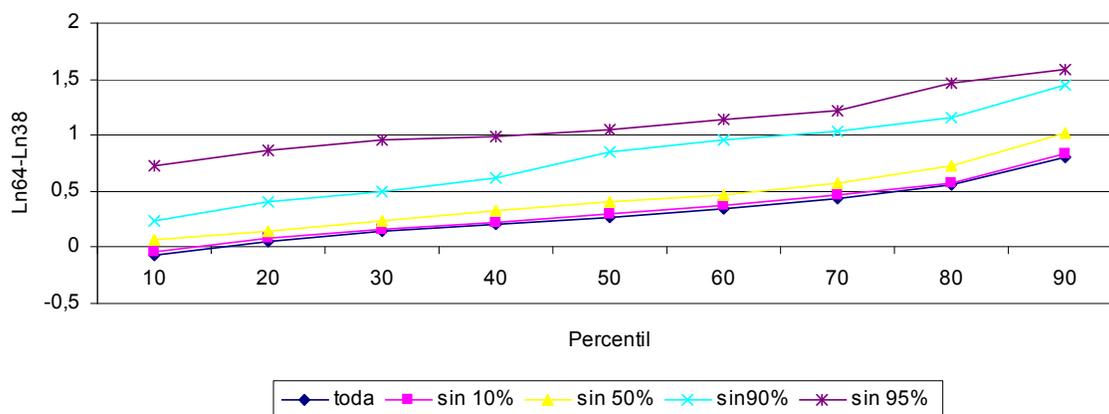
Resultados de las regresiones de la ley de Zipf - Áreas metropolitanas								
Censo	Obs	Coefficiente de correlación múltiple	R cuadrado	R cuadrado ajustado	Intercepto	Error típico del intercepto	Coefficiente	Error típico del coeficiente
1918	687	0,928332682	0,86180157	0,86159982	16,7261215	0,17174136	-1,2905286	0,0197456
1928	715	0,932000215	0,8686244	0,86844014	16,4792306	0,15932609	-1,22537987	0,01784709
1938	724	0,92977568	0,86448281	0,86429512	16,4454081	0,16050914	-1,21115149	0,01784633
1951	799	0,938807347	0,88135923	0,88121037	15,9345745	0,13369417	-1,12678986	0,01464382
1964	923	0,9465575	0,8959711	0,89585815	15,6069648	0,11024629	-1,06340441	0,01193984
1973	1011	0,932224669	0,86904283	0,86891304	14,7638924	0,10862347	-0,95695205	0,01169469
1985	1006	0,957374891	0,91656668	0,91648358	15,1118927	0,08800127	-0,97511303	0,00928487
1993	1035	0,944838672	0,89272012	0,89261626	14,4516608	0,09228379	-0,89190231	0,00961986

Anexo 4

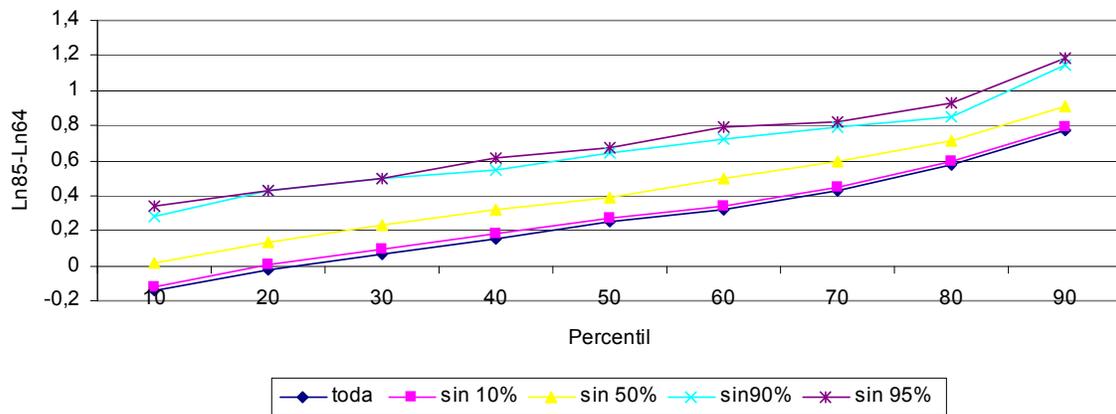
Cambio porcentual en la población por percentil entre 1918 y 1938



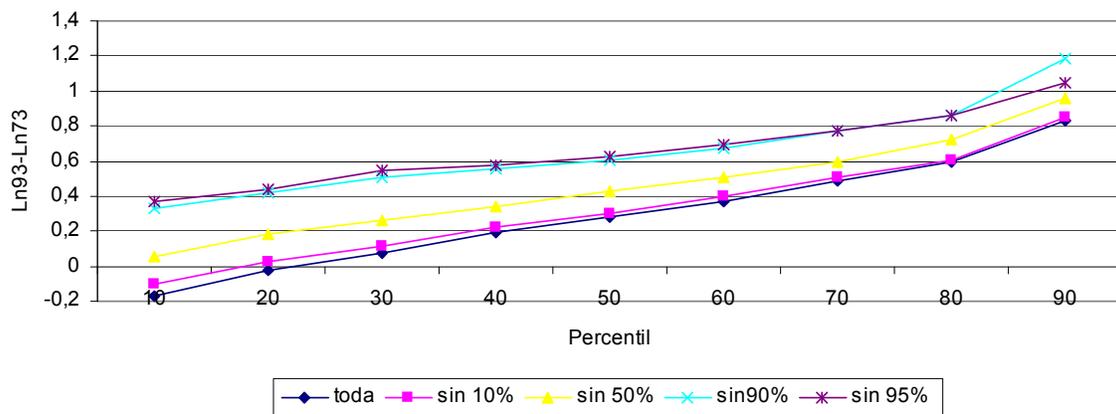
Cambio porcentual en la población por percentil entre 1938 y 1964



Cambio porcentual en la población por percentil entre 1964 y 1985

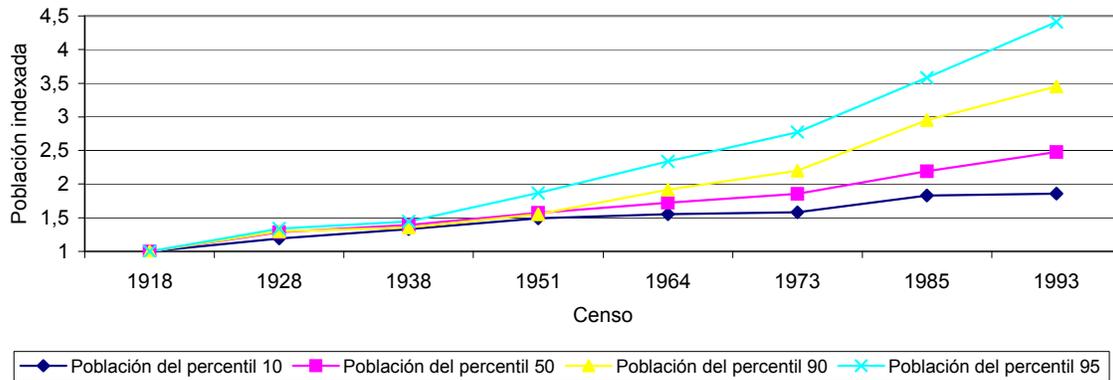


Cambio porcentual en la población por percentil entre 1973 y 1993

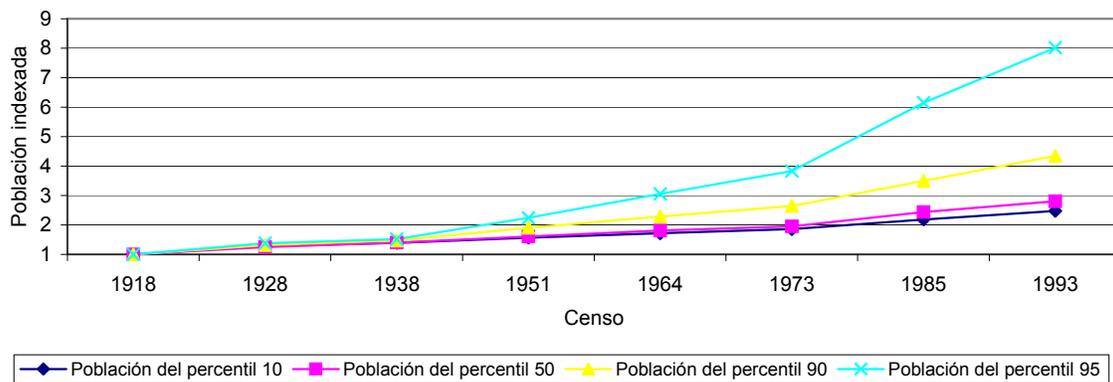


Anexo 5

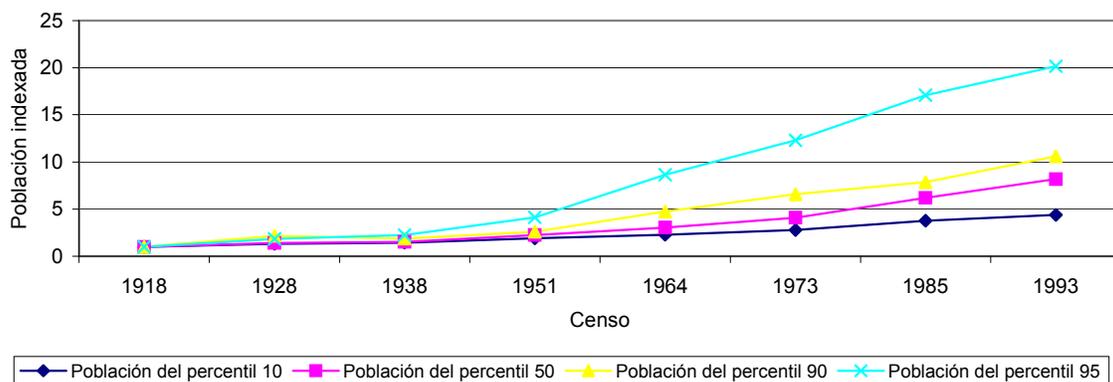
Población indexada por percentil sin el 10% de la muestra (normalizado)
1918-1993



Población indexada por percentil sin el 50% de la muestra (normalizado)
1918-1993

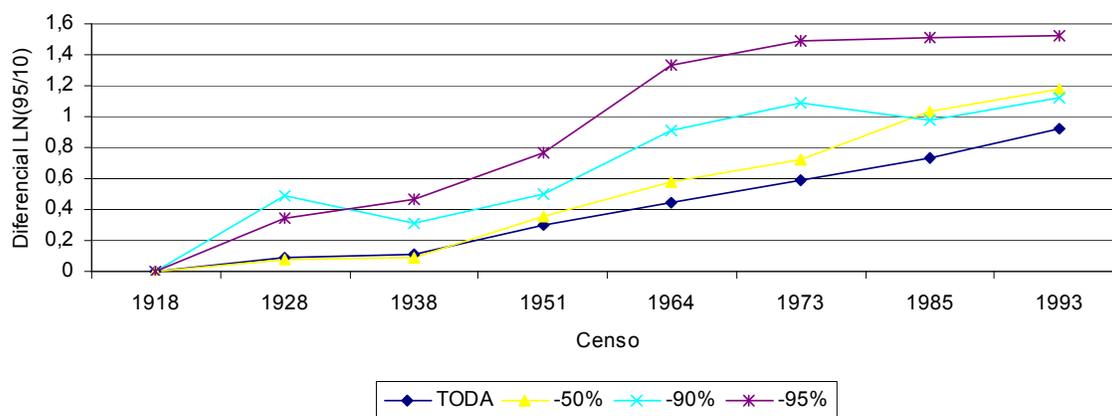


Población indexada por percentil sin el 95% de la muestra (normalizado)
1918-1993



Anexo 6

Diferencial LN(95/10) (normalizado), con diferentes muestras



Diferencial LN(50/10) (normalizado), con diferentes muestras

