



Universidad
Zaragoza

Trabajo Fin de Grado

La ley de Zipf para (prácticamente) todas las
ciudades del mundo

Autor/es

Vasile Adrian Vlad

Director/es

Guillermo Peña Blasco

Facultad de Economía y Empresa
2026

Título del trabajo: **La ley de Zipf para (prácticamente) todas las ciudades del mundo**

Autor del trabajo: **Vasile Adrian Vlad**

Director del trabajo: **Guillermo Peña Blasco**

Project title: **Zipf's law for (almost) all the world's cities**

Titulación: **Graduado en Economía**

RESUMEN:

En el presente trabajo se analiza el cumplimiento de la ley de Zipf para 117.054 ciudades de todo el mundo con más de 1.000 habitantes, tanto de forma agregada, como en función de la renta y la ubicación geográfica. El análisis se realiza empleando la metodología de regresiones rodantes para estimar los coeficientes de Pareto de forma progresiva, primero para todos los datos en conjunto, y después agrupándolos en función del área geográfica de pertenencia y de su renta. Los resultados se presentan de forma gráfica y en tablas, y se explican, observando un cumplimiento parcial de la ley de Zipf en ciertos grupos de urbes, pero no como una ley universal. Posteriormente, se realiza una explicación de las razones subyacentes detrás de los resultados obtenidos, desde una perspectiva económica que abarca el desarrollo de las regiones, o las políticas aplicadas.

ABSTRACT:

This paper examines compliance with Zipf's law for 117,054 cities worldwide with more than 1,000 inhabitants, considering aggregate, income-based and geographic perspective. The analysis is conducted using the rolling regressions methodology to progressively estimate Pareto coefficients, first for the entire dataset, then grouped by geographic region and income level. Results are presented graphically and in tables, revealing a partial compliance with Zipf's law in certain groups of cities, but not as a universal law. Finally, the underlying reasons for the obtained results are explained, from an economic perspective that considers both regional development and applied policies.

ÍNDICE DE CONTENIDOS

ÍNDICE DE CONTENIDOS	3
ÍNDICE DE ECUACIONES	4
1. INTRODUCCIÓN.....	5
2. REVISIÓN DE LITERATURA.....	7
3. METODOLOGÍA	11
3.1. ANTECEDENTES A LA LEY DE ZIPF: LA DISTRIBUCIÓN DE PARETO	11
3.2. ESTRUCTURA DE LA LEY DE ZIPF	12
3.3. CORRECCIONES APLICADAS A LA LEY DE ZIPF	13
3.4. DE LA LEY DE ZIPF A LA LEY DE GIBRAT	14
3.5. APLICACIÓN DE LA LEY DE ZIPF	14
3.5.1. OBTENCIÓN Y PREPARACIÓN DE LOS DATOS.....	14
3.5.2. CÁLCULO Y ESTIMACIÓN DEL CUMPLIMIENTO DE LA LEY DE ZIPF.....	16
3.5.3. CUMPLIMIENTO DE LA LEY DE ZIPF.....	17
4. RESULTADOS.....	17
4.1. RESULTADOS AGREGADOS.....	18
4.2. RESULTADOS POR REGIONES DE RENTA.....	19
4.2.1. RENTA ALTA.....	19
4.2.2. RENTA MEDIA-ALTA.....	21
4.2.3. RENTA MEDIA-BAJA.....	23
4.2.4. RENTA BAJA.....	24
4.3. RESULTADOS POR REGIÓN.....	26
4.3.1. ÁFRICA.....	26
4.3.2. ESTE ASIÁTICO Y PACÍFICO.....	27
4.3.3. EUROPA Y ASIA CENTRAL.....	29
4.3.4. LATINOAMÉRICA Y EL CARIBE.....	30
4.3.5. ORIENTE MEDIO Y NORTE DE ÁFRICA.....	31
4.3.6. SUR DE ASIA.....	32
5. DISCUSIÓN DE RESULTADOS.....	33
5.1. DISCUSIÓN DEL RESULTADO GLOBAL	34
5.2. ANÁLISIS POR GRUPOS DE RENTA.....	35
5.3. ANÁLISIS POR REGIÓN GEOGRÁFICA.....	36
5.4. UTILIDAD GENERAL DE LA LEY DE ZIPF	38
6. CONCLUSIONES.....	39
7. BIBLIOGRAFÍA.....	42
8. ANEXOS.....	44
8.1. ANEXO I.....	44
8.2. ANEXO II.....	45

ÍNDICE DE ECUACIONES

Ecuación 3-1. Distribución de Pareto	11
Ecuación 3-2. Fórmula de la ley de Zipf.....	12
Ecuación 3-3. Fórmula ley de Zipf expresada con constantes.	12
Ecuación 3-4. Transformación lineal de la ley de Zipf (Clauset et al.,2009).....	12
Ecuación 3-5. Transformación lineal de la ley de Zipf con una constante.	12
Ecuación 3-6. Ley de Zipf con reducción de sesgo (Gabaix e Ibragimov, 2011)	13
Ecuación 3-7. Ecuación de la ley de Zipf para la estimación.....	16
Ecuación 3-8. Desviación estándar del estimador.	17
Ecuación 3-9. Intervalo de confianza del estimador de Pareto.....	17

ÍNDICE DE TABLAS

Tabla 4-1. Valores puntuales clave para el análisis conjunto.....	18
Tabla 4-2. Valores puntuales clave para los países de renta alta.....	20
Tabla 4-3. Región de cumplimiento de Zipf para los países de renta alta.	21
Tabla 4-4. Valores puntuales clave para los países de renta media-alta.	21
Tabla 4-5. Región de cumplimiento de Zipf para los países de renta media-alta.....	22
Tabla 4-6. Valores puntuales clave para los países de renta media-baja.	23
Tabla 4-7. Región de cumplimiento de Zipf para los países de renta media-baja.....	24
Tabla 4-8. Región de cumplimiento de Zipf para los países de renta baja.	24
Tabla 4-9. Valores puntuales clave para los países de África.	26
Tabla 4-10. Valores puntuales clave para los países del Este asiático y Pacífico.	27
Tabla 4-11. Valores puntuales clave para los países de Europa y Asia Central.....	29
Tabla 4-12. Valores puntuales clave para los países de Latinoamérica y el Caribe.	30
Tabla 4-13. Valores puntuales clave para los países de Oriente Medio y el Norte de África.....	31
Tabla 4-14. Valores puntuales clave para los países del sur de Asia.....	32

ÍNDICE DE GRÁFICAS

Gráfica 4-1. Evolución del coeficiente de Pareto para el total de la muestra.....	18
Gráfica 4-2. Evolución del coeficiente de Pareto para los países de renta alta	20
Gráfica 4-3. Evolución del coeficiente de Pareto para los países de renta media-alta. ..	22
Gráfica 4-4. Evolución del coeficiente de Pareto para los países de renta media-baja. .	23
Gráfica 4-5. Evolución del coeficiente de Pareto para los países de renta baja.	25
Gráfica 4-6. Evolución del coeficiente de Pareto para los países de África.	26
Gráfica 4-7. Evolución del coeficiente de Pareto para los países del Este asiático y Pacífico.....	28
Gráfica 4-8. Evolución del coeficiente de Pareto para los países de Europa y Asia Central.	29
Gráfica 4-9. Evolución del coeficiente de Pareto para los países de Latinoamérica y el Caribe.	30
Gráfica 4-10. Evolución del coeficiente de Pareto para los países de Oriente Medio y el Norte de África.	32
Gráfica 4-11. Evolución del coeficiente de Pareto para los países del sur de Asia.	33

1. INTRODUCCIÓN

A lo largo de las últimas décadas, el mundo ha experimentado un proceso de transformación urbana sin precedentes. Mientras que en 1.960 apenas 1.000 millones de personas residían en ciudades, frente a 2.000 millones en zonas rurales, en 2.020 la población urbana superaba holgadamente a la rural, con 4.360 millones de personas residiendo en urbes frente a 3.400 millones en el entorno rural (Restrepo Betancur, 2024). La vida en las ciudades se ha convertido en la norma, cuando históricamente siempre ha sido la excepción. Autores como Puyol (2012) afirman que la tendencia esboza un futuro en el que el mundo será un mundo de ciudades, proyectando un declive en la población rural de 400 millones de personas hasta 2.050, que hará que la población rural llegue a 3.000 millones para mediados de siglo.

Este proceso de rápida reorganización del hábitat humano plantea una cuestión fundamental para urbanistas, economistas o geógrafos: ¿existe algún patrón regular que explique la distribución de la población en las ciudades? Una de las respuestas más interesantes surge de la ley de Zipf, ley originalmente procedente de la literatura, y posteriormente aplicada a multitud de áreas, entre ellas, el urbanismo. La ley de Zipf predice que, en un sistema que sigue una distribución ideal, la ciudad más poblada tendrá el doble de población que la segunda, el triple que la tercera, y así sucesivamente, siguiendo una relación inversa entre rango y tamaño. Esta ley se ha vinculado con la ley de Gibrat, que afirma que la tasa de crecimiento de las ciudades es independiente de su tamaño inicial, de modo que ambas leyes proponen un marco teórico sólido en el cual la evolución urbana se entiende como un proceso estocástico que, bajo una serie de condiciones, llevaría a una distribución de Pareto con exponente unitario (Gabaix, 1999). No obstante, la literatura de Zipf ha llegado a la conclusión de que esta ley no es universal mostrando sensibilidad a factores como el nivel de desarrollo económico o la muestra escogida.

El objetivo del presente trabajo es analizar empíricamente el cumplimiento de la ley de Zipf para todas las ciudades a nivel mundial, buscando mejorar el grado de comprensión de la distribución de estas, así como entender mejor la estructura y las dinámicas territoriales a nivel global. Para ello, se aplicará la metodología de regresiones rodantes con la corrección de sesgo propuesta por Gabaix e Ibragimov (2011), estimando el coeficiente de Pareto en una muestra de 117.054 ciudades de más de 1.000 habitantes de todo el mundo. A continuación, con la finalidad de considerar las posibles

particularidades socioeconómicas o geográficas, se realizará el análisis por grupos de renta (países de renta alta, media-alta, media-baja, baja), y por regiones geográficas (África, Este asiático y Pacífico, Europa y Asia Central, Latinoamérica y el Caribe, Oriente Medio y Norte de África, Sur de Asia). A continuación, se identificarán las ventanas de cumplimiento de la ley de Zipf para cada caso, determinando los rangos de tamaño urbano en que se cumple ($\beta \approx 1$), así como los rangos de mayor aglomeración urbana ($\beta > 1$) y de mayor homogeneidad ($\beta < 1$). Con todo esto, se busca contribuir a la literatura empírica sobre la ley de Zipf con una perspectiva actualizada, global y metodológicamente robusta, que permita evaluar el grado de utilidad de esta ley como herramienta para la comprensión de la distribución urbana.

Los resultados obtenidos por este trabajo confirman a escala global, para 117.054 ciudades, que la ley de Zipf no es una ley universal, sino que se trata de un fenómeno que se manifiesta en ventanas muy específicas de núcleos urbanos. Dichas ventanas de cumplimiento variarán en función del nivel de renta o de la región geográfica, poniendo de manifiesto la alta sensibilidad de la ley de Zipf al grado de desarrollo económico y al contexto geográfico, y validando así la tesis de Eeckhout (2004), quien afirmó que la aparente validez de Zipf es un efecto de la selección muestral aplicada. La principal innovación de este trabajo es, precisamente, ser la comprobación más exhaustiva del fenómeno a nivel global, ofreciendo unos resultados que permiten identificar, de forma detallada, dónde y en qué condiciones se cumple la realidad zipfiana, así como la primera comprobación exhaustiva de la ley de Eeckhout.

Este trabajo se estructura del siguiente modo. Tras la actual introducción, se procede, en la Sección 2, a revisar la literatura teórica empírica sobre la ley de Zipf, y su vinculación con la ley de Gibrat; a continuación, en la Sección 3 se detalla la metodología empleada, incluyendo desde la construcción de la base de datos hasta el procedimiento de estimación; posteriormente, en la Sección 4, se presentan los resultados obtenidos, tanto agregados como desglosados por renta y región geográfica; más adelante, en la Sección 5, se discuten dichos resultados, interpretándolos y extrayendo sus implicaciones. Por último, en la sección 6, el trabajo concluye sintetizando los principales hallazgos, señalando las limitaciones del presente trabajo, y sugiriendo futuras líneas de investigación.

2. REVISIÓN DE LITERATURA

La ley de Zipf, formulada inicialmente por George Kingsley Zipf en la década de 1940, es una ley de potencia, esto es, una relación donde se produce una relación inversa entre el tamaño y la frecuencia de un evento. Afirma que, en una gran diversidad de fenómenos, la frecuencia de un elemento es inversamente proporcional a su posición en una lista ordenada por frecuencias. Inicialmente propuesta para la literatura, afirma que la palabra más repetida en una obra literaria tiende a repetirse el doble de veces que la segunda, el triple que la tercera, y así sucesivamente. Si bien el descubrimiento resultó intrigante en el ámbito lingüístico, el salto conceptual de aplicar Zipf a las sociedades humanas reveló un potencial explicativo más ambicioso. En este sentido, a lo largo de los años, se han llevado a cabo diversos estudios para evaluar el grado de cumplimiento de la ley de Zipf.

La búsqueda de un mecanismo teórico capaz de explicar este fenómeno se materializó en el trabajo llevado a cabo por Gabaix (1999). Su modelo propone que, bajo un modelo de crecimiento que cumpla la ley de Gibrat (el crecimiento es un proceso independiente del tamaño inicial), la distribución de los tamaños de los núcleos urbanos converge de forma necesaria hacia una distribución Zipf con exponente unitario. Con este estudio, Gabaix ofreció una explicación fundamentada, estableciendo la hipótesis de regularidad en torno a la distribución de Zipf como un elemento universal en las distribuciones urbanas. Sin embargo, ya desde finales de los años setenta, estudios como el de Rosen y Resnick (1979), mostraron que, para una muestra de 44 países, el exponente Pareto promedio era de 1,136, mostrando que la distribución poblacional tendería a ser más igualitaria que lo esperado con un exponente unitario. Además, Rosen y Resnick detectaron que la clásica línea recta de Pareto se desviaba, mostrando una concavidad hacia arriba, siendo indicativo de un sistema donde el tamaño inicial condiciona el crecimiento, y desafiando así la ley de Gibrat, y revelando retornos crecientes a escala en la urbanización. Además, en el plano muestral, anticiparon la elevada sensibilidad metodológica que otros autores detectarían posteriormente, al hallar que, al pasar de usar datos de ciudades a datos de áreas metropolitanas, el exponente de Pareto disminuye notablemente, así como al variar la muestra seleccionada.

En este sentido, es crucial el trabajo de Eeckhout (2004), quien, utilizando datos censales de todos los municipios de Estados Unidos, de un total de 25.359 municipios, demostró que la distribución del tamaño de las ciudades es lognormal, y no sigue una ley de potencia (Pareto), tal como postula Zipf. Además, logra validar la ley de Gibrat, ley que

afirma que el crecimiento poblacional de los núcleos urbanos es proporcional e independiente del tamaño inicial de la ciudad. De este modo, concluye que el gran cumplimiento de la ley de Zipf que se creía generalizado para todas las poblaciones no era tal, sino que dependía de una selección muestral sesgada, demostrando la elevada sensibilidad del coeficiente de Pareto estimado al punto de corte de la muestra. Así, cuanto menor sea el número de municipios seleccionados, mayor será el coeficiente estimado.

Para que este cumplimiento tenga lugar, Eeckhout (2004) desarrolla su modelo bajo una serie de condiciones clave: asume libre movilidad de trabajadores entre ciudades, donde asume que los individuos al establecerse en un lugar lo hacen buscando maximizar su utilidad. Con ello, predice una fuerte tendencia a la aglomeración urbana, debido a las externalidades positivas generadas por la misma. En este sentido, la distribución lognormal que encuentra empíricamente es resultado de un equilibrio entre las fuerzas de aglomeración y las fuerzas de congestión y costes de vida (beneficios y costes de la aglomeración).

Otros autores como Soo (2005), han investigado empíricamente la ley de Zipf para ciudades en 73 países, utilizando los métodos de Mínimos Cuadrados Ordinarios y el Estimador de Hill, mediante el cual se calcula el exponente de una distribución de ley de potencia a partir de los datos observados, y demostrando que la ley de Zipf es frecuentemente rechazada. Desafía de este modo la universalidad propuesta por Gabaix, mostrando que el patrón de Zipf no es necesariamente una ley universal, sino que puede verse afectado por elementos externos de índole económica o política. En estudios posteriores, otros autores buscaron enfocarse en casos concretos para ver si en la práctica la ley de Zipf se cumplía. Autores como Peng (2010), quien estudió el cumplimiento de la ley de Zipf para las ciudades chinas, utilizando el método de regresiones de muestra rodante (Rolling sample regression), obteniendo una serie de conclusiones a destacar: el cumplimiento de la ley de Zipf no es igual para las ciudades más grandes y para las más pequeñas, cumpliéndose en mayor medida para las urbes más grandes, y demostrando que, en función de la metodología empleada para la estimación, Zipf se cumplirá en mayor o menor medida (en función de decisiones como el punto de corte de la muestra).

Estudios más recientes han puesto a prueba en distintos contextos urbanos la hipótesis de sensibilidad muestral mencionada. En este sentido, Peña y Sanz-Gracia (2021), aplicaron la misma metodología de regresiones de muestra rodante aplicada por Peng (2010) para

los cinco países de los BRICS (Brasil, Rusia, India, China y Sudáfrica), obteniendo resultados diversos. Si bien en Brasil, Rusia y Sudáfrica el exponente de Pareto decrece de forma sistemática al añadir ciudades más pequeñas, en India y China se observan secciones de la distribución donde ese decrecimiento no ocurre, habiendo tramos donde se produce un crecimiento, y reforzando así la idea de la alta sensibilidad del exponente de Zipf tanto al tamaño como a la composición de la muestra, y además, confirmando el cumplimiento de la ley de Zipf con exponente igual a 1 para la zona superior de la distribución, siendo rechazada en el resto. Con estos resultados, los autores sugieren el empleo de distribuciones híbridas para modelar la distribución del tamaño de las ciudades, usando distribuciones como la combinación de lognormal y Pareto.

Más allá del problema de alta sensibilidad en función de la muestra seleccionada, existen otros problemas derivados de la metodología. En este sentido, Gabaix e Ibragimov (2011), demostraron que la estimación por MCO presenta cierto sesgo, especialmente cuando la muestra empleada es escasa. En este sentido, al usar la regresión estándar $\log(Rango)=a-b\log(Tamaño)$ las estimaciones de Pareto obtenidas presentan sesgo para muestras pequeñas. Los autores demostraron que, realizando una sustitución en el primer elemento, y convirtiéndolo a $\log(rango-1/2)$, se obtenían unas estimaciones de Pareto con una reducción sustancial del sesgo. De este modo, con este trabajo, no solo cuestionaron la validez de los estudios realizados anteriormente, sino que ofrecieron una herramienta de importancia capital para estudios venideros.

Hasta ahora, todos los análisis de Zipf se han llevado a cabo usando datos de sección cruzada, sin hacer análisis con series temporales. En este sentido, y con una perspectiva regional de largo plazo, González-Val et al. (2024), llevan a cabo el análisis del cumplimiento de la ley de Zipf y la ley de Gibrat para Estados Unidos desde el año 1900 y hasta el año 2010, obteniendo como resultados que, a nivel de Estados, siendo especialmente relevante el caso de los estados del sur del país, de carácter más agrícola, se cumple la ley de Zipf aproximadamente la mitad del siglo XX, mientras que para el conjunto nacional la ley deja de validarse en los años 60. En cuanto a la ley de Gibrat, los resultados muestran un panorama en el que la ley se cumple para la mayoría de las ciudades, así como para 27 Estados, pero la varianza del crecimiento sí es dependiente del tamaño de la urbe. En conclusión, para llevar a cabo el análisis, es preciso considerar la región analizada, así como el periodo temporal, siendo ambas leyes sensibles a los valores elegidos.

No obstante, este estudio de Zipf a largo plazo también se ha llevado a cabo para países en vías de desarrollo, países donde el comportamiento de Zipf difiere del que tiene en países desarrollados. Un ejemplo es el estudio llevado a cabo por Bernal y Nieto (2006) para la Colombia del siglo XX. Tras analizar datos para las urbes colombianas entre 1918 y 1993 utilizando el método de Mínimos Cuadrados Ordinarios, encontraron no solo que la ley de Zipf es rechazada para todos los años, sino los fenómenos que logran explicar las fluctuaciones del coeficiente de Pareto, vinculando estas fluctuaciones a eventos críticos para la historia del país. Sin embargo, es preciso señalar que estos resultados tienen cierto sesgo, dado que en para este análisis se emplea la regresión estándar que sería corregida posteriormente por Gabaix e Ibragimov (2011).

A nivel nacional, destaca el estudio llevado a cabo por Lanaspá et al. (2004), quienes analizaron la distribución poblacional española para el siglo XX, prestando especial atención a la selección muestral, al repetir el análisis para las 100, 300, 500 y 700 ciudades más pobladas del país. Mediante el uso de Mínimos Cuadrados Ordinarios, obtuvieron el rechazo de la ley de Zipf de forma estricta, mediante métodos de estimación no lineales, como la especificación de Rosen y Resnick (1980) y la de Fan y Casetti (1994), obtuvieron una distribución del exponente de Pareto con forma de “U”, presentando el punto de inflexión hacia 1970-1980. Concluyen de este modo que el crecimiento hasta los años 70 fue divergente, con una desigualdad mayor entre polos urbanos, y una convergencia posterior, mostrando de este modo la tendencia experimentada por el país, primero hacia una urbanización acelerada y concentrada en unos pocos polos, y después hacia una urbanización más igualitaria, con las ciudades pequeñas y medianas del país ganando protagonismo.

En conclusión, a lo largo de la historia de los estudios sobre la ley de Zipf en sistemas urbanos, la conclusión observable es que dicha ley no es una verdad inmutable. Su comportamiento presenta grandes variaciones en función de elementos como el punto de corte seleccionado en las muestras, o en función de las técnicas de estimación empleadas. De este modo, la ley de Zipf no surge como una ley natural, sino como una aproximación descriptiva condicionada por la región analizada, el periodo histórico o decisiones metodológicas de los investigadores. No obstante, y a pesar de lo anterior, la capacidad explicativa de Zipf, especialmente al comparar distribuciones reales con la esperada por Zipf (exponente igual a uno), es muy elevada, permitiendo detectar patrones en las

distribuciones de población y facilitando así la evaluación del impacto de políticas económicas.

3. METODOLOGÍA

La revisión de literatura llevada a cabo ha puesto de manifiesto la complejidad y madurez del debate académico en torno a la ley de Zipf aplicada a sistemas urbanos. Se han identificado las principales teorías explicativas, métodos aplicados para la obtención de coeficientes, los principales resultados obtenidos, así como las principales críticas. No obstante, para poder llevar a cabo un correcto análisis empírico, es preciso explicar, de forma sistemática, los principales conceptos clave en torno a la ley de Zipf y de Gibrat, con la finalidad de concretar el marco matemático que será la base del diseño metodológico de este trabajo. Así, en este apartado, se busca explicar el origen de la ley de Zipf y su funcionamiento, la relación que guarda con la ley de Gibrat, así como la aplicación llevada a cabo de esta metodología para el presente trabajo.

3.1. ANTECEDENTES A LA LEY DE ZIPF: LA DISTRIBUCIÓN DE PARETO

Los fundamentos matemáticos de la regularidad fruto de este estudio se remontan al trabajo llevado a cabo por el economista y sociólogo Vilfredo Pareto (1896) quien, al llevar a cabo un análisis de la distribución de la riqueza y los ingresos en diferentes sociedades, observó un patrón común, mediante el cual un segmento reducido de la población concentraba una alta proporción del total de recursos. Esta observación, formalizada posteriormente como el Principio de Pareto, o la regla 80-20 (el 20% de la población posee el 80% de la tierra), tiene su expresión matemática en la distribución de Pareto, una distribución de potencias, que afirma que la probabilidad de encontrar valores elevados decae siguiendo una ley de potencias. Formalmente, la distribución de Pareto afirma que la probabilidad de que una variable aleatoria X supere un umbral x es inversamente proporcional a una potencia de x :

$$P(X > x) = \left(\frac{x_m}{x}\right)^\alpha \text{ para } x \geq x_m$$

Ecuación 3-1. Distribución de Pareto

Donde $x_m > 0$ es el valor mínimo posible y $\alpha > 0$ es el exponente de Pareto. Un valor de α menor indica una distribución más desigual, con una probabilidad mayor de valores extremos. Con esta distribución, Pareto ofreció la primera descripción matemática de la

desigualdad en sistemas socioeconómicos, sentando las bases para futuros estudios de distribuciones asimétricas en áreas como la sociología, la geofísica o la economía (Persky, 1992).

3.2. ESTRUCTURA DE LA LEY DE ZIPF

Tal como se ha revisado, la ley de Zipf, originada en el análisis lingüístico (Zipf, 1949), establece una relación potencial inversa entre el rango de un elemento y su frecuencia. Aplicada al análisis de la población de los núcleos urbanos, establece que la población P_r de una ciudad en la posición r (ordenadas de mayor a menor tamaño), se relaciona con su rango mediante la siguiente expresión:

$$P_r = \frac{P_1}{r^\beta}$$

Ecuación 3-2. Fórmula de la ley de Zipf.

Con P_1 siendo la ciudad de mayor tamaño de la distribución, y β es el exponente de Pareto (o exponente de Zipf). De forma más general, la relación se puede expresar en torno a una constante C , obteniendo la siguiente expresión:

$$P_r = \frac{C}{r^\beta}$$

Ecuación 3-3. Fórmula ley de Zipf expresada con constantes.

En la ley de Zipf original, para su estimación empírica mediante regresión lineal, la ecuación se transforma tomando logaritmos, obteniendo la siguiente expresión (Clauset et al., 2009):

$$\ln(P_r) = \ln(P_1) - \beta \ln(r)$$

Ecuación 3-4. Transformación lineal de la ley de Zipf (Clauset et al., 2009)

Esta expresión, para el caso de la ecuación con un valor constante en el numerador, se expresa como:

$$\ln(P_r) = C - \beta \ln(r)$$

Ecuación 3-5. Transformación lineal de la ley de Zipf con una constante.

De este modo, al llevar a cabo la representación gráfica de la $\ln(P_r)$ frente a $\ln(r)$ para todas las ciudades analizadas, si se cumple la ley de Zipf, los puntos tenderán a alinearse en torno a una recta con pendiente $-\beta$.

En la ley de Zipf, el parámetro central es el exponente de Pareto, o exponente de Zipf, cuyo valor tiene una interpretación económica directa. De este modo, un valor del exponente de Zipf igual a 1 es un indicador de la expresión pura de la ley de Zipf. Implica que cada ciudad tiene un valor inversamente proporcional a su rango, cumpliéndose así lo descrito por Zipf (la segunda ciudad tiene la mitad de población que la primera, la tercera un tercio...). Un valor del exponente superior a 1 es indicativo de un sistema urbano donde la concentración es mayor, de modo que la urbe más grande predomina por encima de lo establecido inicialmente por Zipf. Del mismo modo, un valor del exponente inferior a 1 sugiere un sistema urbano más homogéneo, donde hay una menor desigualdad en el tamaño de los núcleos urbanos.

3.3. CORRECCIONES APLICADAS A LA LEY DE ZIPF

Si bien hasta ahora la formulación de la ley de Zipf ha proporcionado el marco teórico básico, la aplicación de esta empleando la regresión lineal de forma directa conlleva un problema metodológico importante.

Tal como demostraron Gabaix e Ibragimov (2011), la estimación por el método de Mínimos Cuadrados Ordinarios (MCO) del exponente β mediante la ecuación estándar obtenida aplicando logaritmos a la ecuación original: $\ln(P_r) = C - \beta \ln(r)$, la estimación del parámetro de Pareto obtenida será sesgada, siendo este problema especialmente relevante al analizar muestras reducidas. Este sesgo es importante, dado que puede distorsionar la interpretación del parámetro, así como las conclusiones sobre el cumplimiento de la ley obtenidas.

La solución propuesta por Gabaix e Ibragimov (2011), es sencilla, consistiendo en realizar un ajuste al valor de la posición, utilizando para la estimación, en vez del valor r , el valor $r - 0,5$, obteniendo así la siguiente especificación:

$$\ln(P_r) = C - \beta \ln\left(r - \frac{1}{2}\right)$$

Ecuación 3-6. Ley de Zipf con reducción de sesgo (Gabaix e Ibragimov, 2011)

Con esta corrección, se consigue reducir de forma sustancial el sesgo en la estimación del parámetro de Zipf, proporcionando intervalos de confianza más precisos, y mejorando los resultados para el análisis, especialmente al trabajar con muestras pequeñas.

3.4. DE LA LEY DE ZIPF A LA LEY DE GIBRAT

La ley de Gibrat, formulada por Robert Gibrat en su obra *Les Inégalités Economiques* (Las Desigualdades Económicas) (1931), es una de las principales hipótesis acerca del crecimiento de las unidades económicas. Inicialmente propuesta como una hipótesis para el crecimiento de las empresas, ha sido extrapolada con éxito a áreas como el urbanismo. Gibrat planteó que la tasa de crecimiento de una empresa es independiente del tamaño inicial, lo que implica un proceso estocástico natural donde, si bien las diferencias iniciales pueden verse amplificadas, no es debido a una tendencia sistemática hacia las unidades más grandes o pequeñas.

La relación entre la ley de Zipf y la ley de Gibrat fue propuesta por Gabaix (1999), quien en su modelo demostró de forma matemática que, si el modelo de crecimiento de las urbes sigue un proceso capaz de satisfacer la ley de Gibrat, y dentro de ese modelo existe un elemento que ejerza de “barrera reflectante”, impidiendo la desaparición total de cualquier ciudad del análisis, la distribución de las ciudades analizadas converge necesariamente, a largo plazo, hacia una distribución de Pareto con exponente unitario ($\beta=1$).

De este modo, la teoría propuesta por Gabaix convierte al exponente β en la herramienta de diagnóstico de la distribución analizada anteriormente, con un exponente de 1 indicando crecimiento neutral, cumpliendo la ley de Gibrat, un exponente mayor indicando un crecimiento más concentrado, y un exponente menor indicando mayor homogeneidad en el crecimiento. Por tanto, el valor de β y su evolución son herramientas especialmente útiles para inferir el tipo de crecimiento y sus cambios en el largo plazo.

3.5. APLICACIÓN DE LA LEY DE ZIPF

3.5.1. OBTENCIÓN Y PREPARACIÓN DE LOS DATOS

El presente análisis se centra en analizar el cumplimiento de la ley de Zipf para todas las ciudades de 1.000 habitantes o más del mundo. Este análisis se va a realizar tres veces. El primero de los análisis se llevará a cabo con todas las ciudades de forma agregada; el segundo, clasificándolas en función del nivel de renta de los países a los que pertenecen; y el tercero, clasificándolas en función de la región geográfica a la que pertenecen.

Los datos utilizados para llevar a cabo el presente estudio proceden de la base de datos “Geonames All Cities with a Population ≥ 1.000 ”, publicada a través de la plataforma

OpenDataSoft (Opendatasoft, 2024). En esta base de datos se encuentra información armonizada sobre la ubicación geográfica y el número de residentes de todos los núcleos de población de más de 1.000 habitantes existentes a nivel mundial.

Con el fin de que la base de datos sea eficiente y útil para el propósito del análisis, se ha llevado a cabo un proceso de limpieza y ordenación de los datos originales. Este proceso ha sido de una importancia capital, puesto que los datos base se encontraban desagregados de forma desigual, con una gran cantidad de núcleos de población (especialmente los más poblados, con mayor densidad de población o con mayor superficie) divididos por barrios o distritos.

Para resolver este problema, y obtener una base de datos en la que todos los datos se encuentren agregados de forma homogénea, se ha utilizado la herramienta Power Query de Microsoft Excel, que, entre sus funciones, permite la agrupación de datos en base a columnas comunes. Con el fin de evitar que se produzcan agrupaciones erróneas entre localidades de igual nombre, se ha realizado la agrupación en base a dos datos: por un lado, el país del enclave; por otro, el huso horario de pertenencia del enclave. Como consecuencia de este proceso de filtrado, la base de datos pasa de tener 147.044 elementos a 117.036 elementos, evidenciándose de este modo la magnitud del problema de desagregación existente.

Una vez se ha llevado a cabo el proceso de limpieza de los datos, el paso final para poder llevar a cabo el análisis es el de la organización. Para poder estudiar Zipf, los datos se deben ordenar de mayor a menor población, y se debe añadir una columna de datos adicional, donde estos núcleos se numerarán de forma ordenada.

Con este proceso de construcción, limpieza y organización de la base de datos, se puede proceder al análisis del cumplimiento de la ley de Zipf en base a los tres enfoques descritos, para lo cual será preciso separar los datos. Así, en función del análisis:

- Para el análisis del cumplimiento entre todos los núcleos de población a nivel mundial, no es preciso hacer ninguna modificación en los datos.
- Para el análisis del cumplimiento en función al nivel de renta, se utilizarán los criterios de clasificación utilizados por el Banco Mundial, con cuatro grupos de países: High (Renta Alta), Low (Renta Baja), Lower Middle (Renta Media-Baja), Upper Middle (Renta Media-Alta) (ver Anexo I).

- Para el análisis del cumplimiento por región geográfica, también se utilizarán los criterios de clasificación del Banco Mundial, con las siguientes clasificaciones: África, Este Asiático y Pacífico, Europa y Asia Central, Latinoamérica y el Caribe, Oriente Medio y Norte de África, Sur de Asia (ver Anexo II).

3.5.2. CÁLCULO Y ESTIMACIÓN DEL CUMPLIMIENTO DE LA LEY DE ZIPF

La totalidad del proceso de estimación ha sido llevado a cabo en Microsoft Excel, aprovechando sus funciones de análisis de datos. El análisis ha sido llevado a cabo adoptando un enfoque econométrico de regresiones rodantes (Rolling sample regressions). Se ha elegido este método porque permite observar el comportamiento del coeficiente de Pareto a medida que se añaden ciudades más pequeñas de forma progresiva.

Para llevar a cabo ese procedimiento en Microsoft Excel, se comienza seleccionando una muestra inicial de las 20 ciudades más grandes, con la que se realizará la estimación del coeficiente de Pareto por MCO. Posteriormente, se seguirá un enfoque incremental, de modo que la siguiente estimación incluirá 21 ciudades, la siguiente 22, y así sucesivamente hasta incluir todos los elementos de la base de datos.

Además, con la finalidad de obtener unos resultados más fiables, se ha añadido la corrección de sesgo propuesta por Gabaix e Ibragimov (2011), mediante la cual al rango de cada elemento se le resta $\frac{1}{2}$. De este modo, y siguiendo la especificación de Peña y Sanz-García (2021), se estima la siguiente ecuación para Zipf:

$$\ln(r - 0,5) = \alpha - \beta \ln(S)$$

Ecuación 3-7. Ecuación de la ley de Zipf para la estimación.

Donde r es el rango de la ciudad, S es su población, α es la constante del modelo y β es el exponente de Pareto. Es importante señalar que el valor del exponente de Pareto es negativo, dada la relación inversa existente entre el rango de una ciudad y su población. No obstante, y con la finalidad de permitir la interpretación, se considerará el valor absoluto del coeficiente, entendiendo $\beta = 1$ como un cumplimiento estricto de la ley de Zipf.

3.5.3. CUMPLIMIENTO DE LA LEY DE ZIPF

Una vez se ha llevado a cabo el proceso de estimación de los coeficientes de Pareto para toda la base de datos, se ha llevado a cabo un proceso de evaluación del cumplimiento estadístico de la ley de Zipf que, recordemos, se cumple cuando $\beta = 1$. Para llevar a cabo este proceso, se ha calculado la desviación estándar del estimador, mediante la siguiente fórmula:

$$\sigma_{\hat{\beta}} = |\hat{\beta}| \times \sqrt{\frac{2}{N + 1/2}}$$

Ecuación 3-8. Desviación estándar del estimador.

Una vez obtenida la desviación estándar para cada uno de los estimadores de Pareto, se puede proceder al cálculo de los intervalos de confianza para un nivel de confianza del 95%, cuya expresión es la siguiente:

$$IC(\hat{\beta}) = (\hat{\beta} \mp 1,96 \times \sigma_{\hat{\beta}})$$

Ecuación 3-9. Intervalo de confianza del estimador de Pareto.

Donde se utilizará el valor crítico $z = 1,96$, que aproxima la distribución t de Student. De este modo, si el valor teórico óptimo para el parámetro de Zipf ($\beta = 1$) se encuentra dentro del intervalo de confianza calculado, no se rechazará la hipótesis nula de cumplimiento de la ley de Zipf. En caso contrario, se concluirá que existe evidencia estadística suficiente para afirmar que la distribución de la población difiere de forma significativa de la predicha por la ley de Zipf.

4. RESULTADOS

A continuación, se presentan los resultados empíricos obtenidos al aplicar la metodología anteriormente descrita. Estos resultados se presentan en tres apartados: en primer lugar, los resultados obtenidos al aplicar el análisis del cumplimiento de la ley de Zipf a la totalidad de las ciudades en conjunto; en segundo lugar, los resultados obtenidos de aplicar el análisis para las principales regiones de renta del mundo; finalmente, los resultados obtenidos de aplicar el análisis para las principales regiones geográficas. Para cada uno de los apartados, se va a examinar la evolución del coeficiente de Pareto (β) a medida que la muestra aumenta, y se va a evaluar el cumplimiento de la ley de Zipf mediante el empleo de intervalos de confianza con un nivel de confianza del 95%. Tal

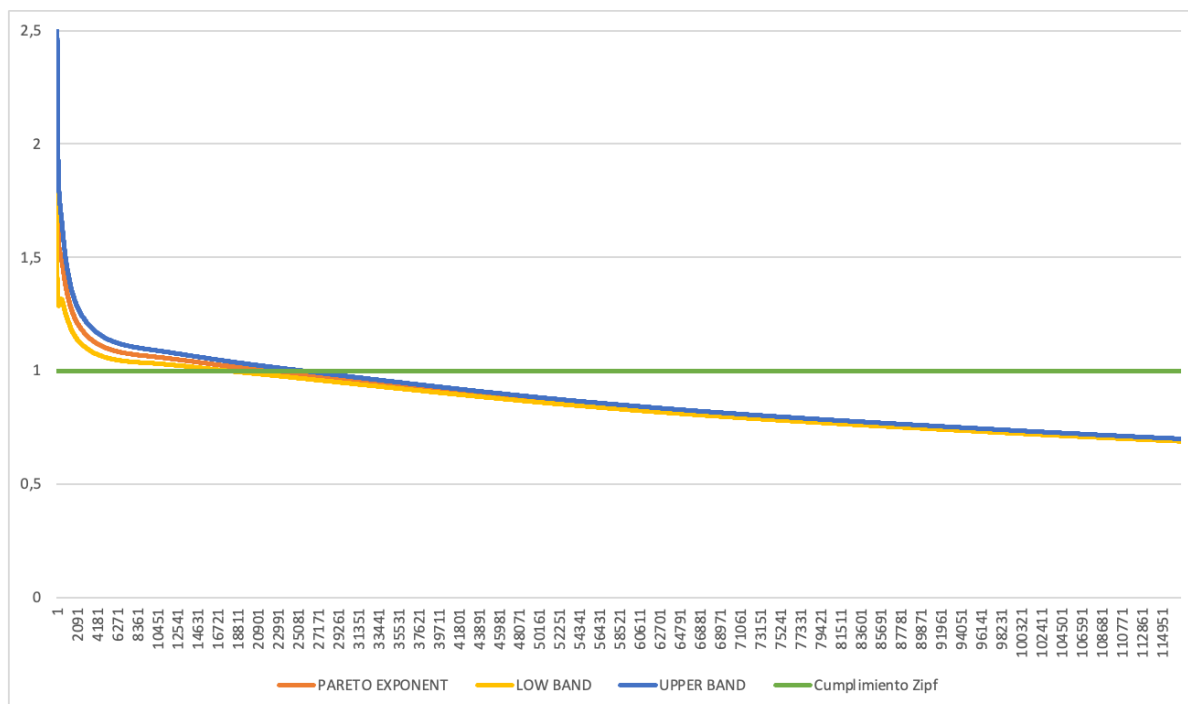
como se ha mencionado en el apartado de metodología, los resultados obtenidos corresponden a la metodología con la corrección aplicada por Gabaix e Ibragimov (2011).

4.1.RESULTADOS AGREGADOS

La aplicación de regresiones rodantes al conjunto de 117.054 ciudades revela una dinámica compleja en la distribución de tamaños a nivel mundial. En este sentido, la Tabla 4-1 presenta una serie de valores puntuales clave, mientras que la Figura 4-1 muestra la evolución del coeficiente de Pareto junto con su intervalo de confianza al 95% para la totalidad de los datos.

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	3,7989	1,1866	1,4732	6,1247	No
100	1,9588	0,2763	1,4172	2,5004	No
500	1,5006	0,0949	1,3147	1,6865	No
1.000	1,3678	0,0612	1,2480	1,4877	No
5.000	1,1032	0,0221	1,0599	1,1464	No
10.000	1,0621	0,0150	1,0327	1,0915	No
50.000	0,8721	0,0055	0,8613	0,8829	No
100.000	0,7292	0,0033	0,7228	0,7356	No
117.054	0,6940	0,0029	0,6883	0,6996	No

Tabla 4-1. Valores puntuales clave para el análisis conjunto.



Gráfica 4-1. Evolución del coeficiente de Pareto para el total de la muestra.

Con estos datos, se pueden extraer una serie de conclusiones relevantes. En primer lugar, es de destacar la curva que traza el exponente de Pareto, el cual comienza en un valor muy elevado, de 3,80 para las 20 primeras muestras, y termina alcanzando un valor de 0,69 para la muestra completa, suponiendo así una caída del 82%. Estos valores se

reflejan en la gráfica, donde el exponente comienza con una pendiente muy elevada que va suavizándose a medida que se añaden observaciones. A efectos poblacionales, esta caída tan abrupta para el coeficiente de Pareto es indicativo de una distribución muy desigual, así como de una existencia de no linealidad de la distribución urbana global. Estos resultados son consistentes con la ley de Eeckhout (2004), mostrando un comportamiento característico de la distribución log-normal.

En cuanto al cumplimiento de la Ley de Zipf, se observa una pequeña ventana de cumplimiento, que abarca desde la observación 17.661 (Prairieville, EEUU, con 26.895 habitantes), y hasta la observación 25.522 (Holzkirchen, Alemania, con 16.770 habitantes). En este sentido, se pueden establecer tres ventanas de cumplimiento de Zipf:

- **Ventana de sobrerrepresentación:** cuando $N < 17.661$, con N siendo el tamaño muestral, los valores obtenidos para el coeficiente de Pareto son superiores al valor ideal de Zipf ($\beta=1$), indicando una mayor concentración en torno a las grandes urbes.
- **Ventana de Zipf:** cuando $17.661 \leq N \leq 25.522$, los valores obtenidos para el coeficiente de Pareto se ajustan estadísticamente al valor idea de Zipf.
- **Ventana de homogeneidad:** para $N > 25.522$, el valor del coeficiente de Pareto obtenido es inferior al valor de cumplimiento de la ley de Zipf, indicando una mayor homogeneidad en la distribución que la predicha por Zipf.

4.2. RESULTADOS POR REGIONES DE RENTA

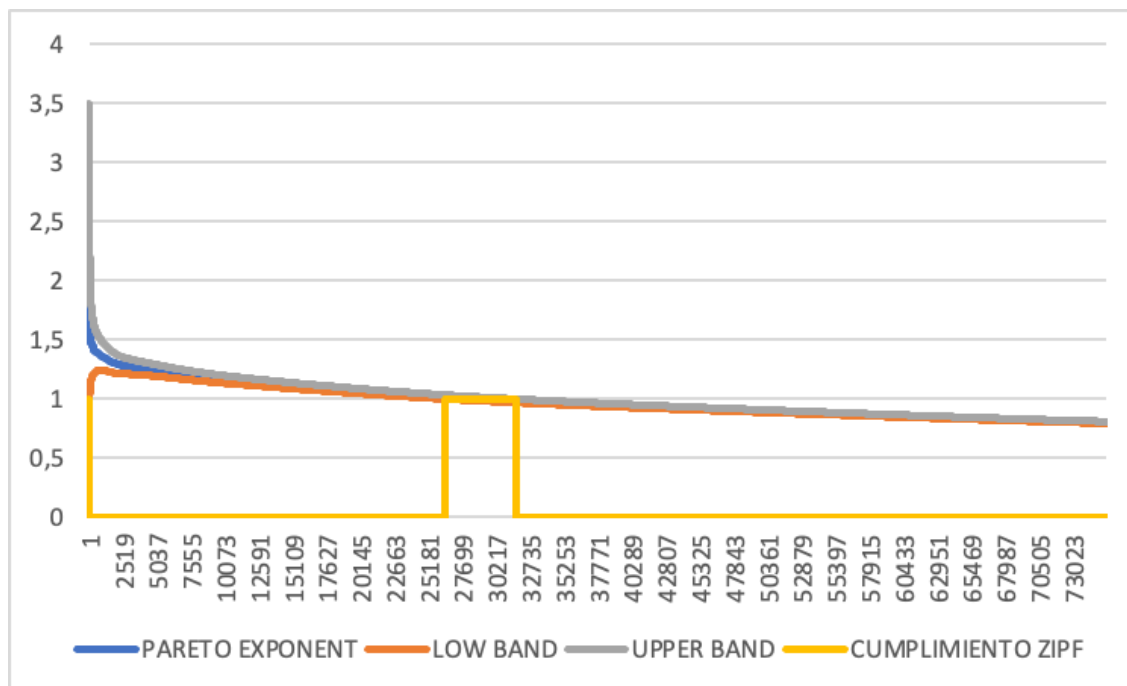
Una vez llevado a cabo el análisis de la distribución global, se procede a examinar el análisis del cumplimiento de Zipf por regiones de renta. Siguiendo la clasificación del Banco Mundial, se identifican los siguientes cuatro grupos de renta: Alta, Media-Alta, Media-Baja, Baja.

4.2.1. RENTA ALTA

Las regiones de renta alta aportan al total de la muestra un total de 75.530 ciudades. Aplicando el análisis de regresiones rodantes a este grupo, se obtienen los siguientes datos, presentados en la Tabla 4-2, donde se muestran la evolución del coeficiente de Zipf, el error estándar y el intervalo de confianza al 95% de confianza para los principales puntos de la muestra, mientras que la Figura 4-2 muestra la evolución del coeficiente a lo largo de la totalidad de la muestra:

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	2,1708	0,6780	0,8418	3,4998	Sí
100	1,5699	0,2215	1,1358	2,0039	No
500	1,4023	0,0886	1,2286	1,5761	No
1.000	1,3599	0,0608	1,2407	1,4791	No
10.000	1,1667	0,0165	1,1344	1,1991	No
25.000	1,0273	0,0092	1,0093	1,0453	No
75.530	0,8018	0,0041	0,7937	0,8099	No

Tabla 4-2. Valores puntuales clave para los países de renta alta.



Gráfica 4-2. Evolución del coeficiente de Pareto para los países de renta alta

En la Figura 4-2 se muestra una línea amarilla, que cumple la función de indicador del cumplimiento de Zipf. Adopta el valor 1 cuando el resultado obtenido se ajusta a las condiciones de cumplimiento de la ley de Zipf (el valor $\beta=1$ se encuentra dentro de intervalo de confianza estimado), y adopta el valor 0 cuando el resultado no se ajusta a Zipf. De este modo, para las regiones de renta alta se observan dos regiones donde la ley de Zipf se cumple.

En primer lugar, se observa que el indicador de cumplimiento muestra el cumplimiento de la ley de Zipf para los primeros valores de la muestra. No obstante, esto es así debido a la elevada amplitud del intervalo de confianza obtenido para estos valores, comprendidos entre $20 \leq N \leq 48$, donde no se puede rechazar estadísticamente $\beta=1$. Sin embargo, el alto coeficiente de Pareto obtenido para $N=20$, de 2,17, es un indicador claro de una mayor concentración en estas ciudades.

En segundo lugar, para $26.390 \leq N \leq 31.667$ (donde $N=26.390$ corresponde a Venta de Baños, España, con 5.918 habitantes, y $N=31.667$ corresponde a Outara, Tahití, con 4.609

habitantes), se observa una región de alto cumplimiento estadístico, donde los valores superior e inferior del intervalo de confianza obtenido se ajustan estrechamente alrededor del valor $\beta=1$ de cumplimiento de Zipf, tal como se puede observar en la Tabla 4-3:

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
26.390	1,0174	0,0089	1,0000	1,0347	Sí
31.667	0,9847	0,0078	0,9693	1,0000	No

Tabla 4-3. Región de cumplimiento de Zipf para los países de renta alta.

Más allá de esta ventana intermedia de cumplimiento de Zipf, se observa que el coeficiente de Pareto desciente de forma progresiva a medida que ciudades más pequeñas son añadidas, siguiendo un patrón consistente con la distribución log-normal propuesta por Eeckhout (2004). En consecuencia, para los países de renta alta, se observan las siguientes 3 ventanas de cumplimiento de Zipf:

- **Ventana de sobrerrepresentación:** cuando $N < 26.390$, el valor del coeficiente obtenido será superior a 1, indicando una mayor concentración de población en las grandes urbes.
- **Ventana de Zipf:** cuando $26.390 \leq N \leq 31.667$, los valores obtenidos se ajustan estadísticamente al valor ideal de Zipf.
- **Ventana de homogeneidad:** para $N > 31.667$, el valor del coeficiente de Pareto será menor indicando un mayor nivel de homogeneidad.

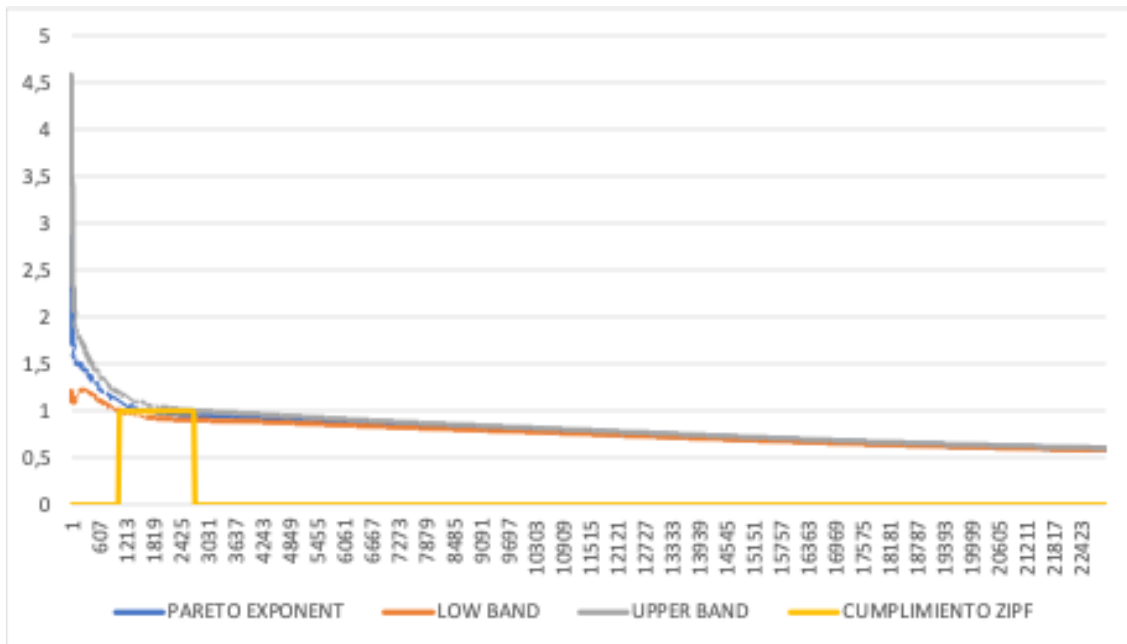
En conclusión, para países de altos ingresos, la ley de Zipf se cumplirá únicamente para un rango muy específico de ciudades intermedias.

4.2.2. RENTA MEDIA-ALTA

Los países de renta media-alta aportan un total de 23.010 ciudades a la muestra. Aplicando la metodología de regresiones rodantes a estas ciudades, se obtienen los datos de la Tabla 4-4 para los principales puntos de la muestra, y en la Figura 4-3 se observa la evolución del coeficiente de Pareto:

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	2,8410	0,8874	1,1017	4,5802	No
100	1,5404	0,2173	1,1145	1,9664	No
500	1,3227	0,0836	1,1588	1,4866	No
1.000	1,1122	0,0497	1,0148	1,2097	No
10.000	0,8010	0,0113	0,7788	0,8232	No
20.000	0,6261	0,0063	0,6139	0,6384	No
23.010	0,5929	0,0055	0,5820	0,6037	No

Tabla 4-4. Valores puntuales clave para los países de renta media-alta.



Gráfica 4-3. Evolución del coeficiente de Pareto para los países de renta media-alta.

En el caso de los países de renta media-alta, se observa un coeficiente de Zipf que disminuye de manera muy notable a medida que el tamaño muestral aumenta, pasando de un coeficiente de 2,841 con $N=20$ a un coeficiente de 0,5929 para un coeficiente de $N=23.010$. Además, al observar la variación del error estándar, se observa una caída muy notable entre el primer coeficiente estimado, con un error estándar de 0,8874 a un error de 0,0055 para el último coeficiente. El efecto de esta variación es la estimación de intervalos de confianza cada vez más estrechos, reflejando una precisión cada vez mayor a medida que la muestra aumenta.

El cumplimiento de la ley de Zipf tiene lugar de forma consistente entre la estimación 1.087 (Curug, Indonesia, con 191.406 habitantes) y la estimación 2.761 (Picheng, China, con 64.585 habitantes), para las cuales los valores obtenidos son los siguientes (Tabla 4-5):

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
1.087	1,0917	0,0468	0,9999	1,1834	Sí
2.761	0,9499	0,0256	0,8998	1,0000	Sí

Tabla 4-5. Región de cumplimiento de Zipf para los países de renta media-alta.

De este modo, la ventana de cumplimiento de Zipf para los países de renta media-alta se establece entre los valores 1.087 y 2.761, donde el intervalo de confianza para el nivel de confianza del 95% se ajusta con mayor precisión en torno al valor óptimo de Zipf ($\beta=1$). Una vez superado este rango, el coeficiente de Pareto desciende por debajo de la unidad,

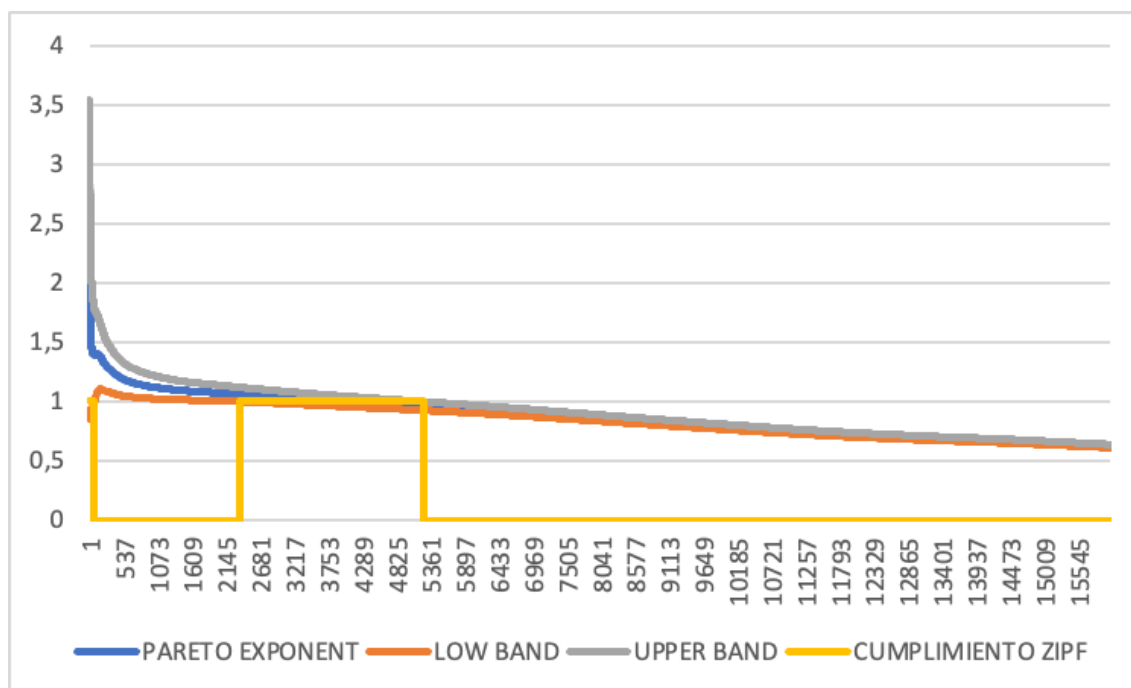
en un comportamiento que se alinea con lo esperado bajo una distribución log-normal, en línea con la ley de Eeckhout (2004).

4.2.3. RENTA MEDIA-BAJA

Los países situados en el grupo de países de renta media-baja, aportan al total de la muestra un total de 16.070 observaciones. Así, al aplicar las regresiones rodantes, se obtienen los siguientes datos, por un lado, de los principales puntos de la muestra (Tabla 4-6), y, por otro, de la evolución del coeficiente de Pareto a lo largo de toda la muestra (Figura 4-4):

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	2,1956	0,6858	0,8515	3,5398	Sí
100	1,3928	0,1965	1,0077	1,7780	No
500	1,2035	0,0761	1,0544	1,3527	No
1.000	1,1238	0,0502	1,0253	1,2223	No
10.000	0,7865	0,0111	0,7647	0,8083	No
15.000	0,6527	0,0075	0,6379	0,6675	No
16.070	0,6249	0,0070	0,6113	0,6386	No

Tabla 4-6. Valores puntuales clave para los países de renta media-baja.



Gráfica 4-4. Evolución del coeficiente de Pareto para los países de renta media-baja.

De la misma manera que en el caso de los países de renta alta, se observan dos regiones de cumplimiento de la ley de Zipf. La primera de ellas tiene lugar entre los primeros coeficientes estimados, concretamente entre $20 \leq N \leq 96$. No obstante, este cumplimiento se debe, de igual modo que en los países de renta alta, a una gran amplitud del intervalo de confianza, derivada del elevado error estándar obtenido. Este hecho es observable al

analizar el valor del coeficiente de Pareto obtenido para $N=20$, el cual es de 2,1956, indicativo de una gran concentración entre las urbes más habitadas para los países de renta media-baja.

Por otro lado, se da el cumplimiento de la ley de Zipf de forma amplia para los datos comprendidos entre 2.387 (Biswan, India, con 52.516 habitantes) y 5.369 (Arukutti, India, con 17.944 habitantes), valores entre los cuales el ajuste al valor ideal para la ley de Zipf es mayor, tal como puede observarse en la Tabla 4-7:

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
2.387	1,0601	0,0307	1,0000	1,1203	Sí
5.269	0,9632	0,0188	0,9265	1,0000	Sí

Tabla 4-7. Región de cumplimiento de Zipf para los países de renta media-baja.

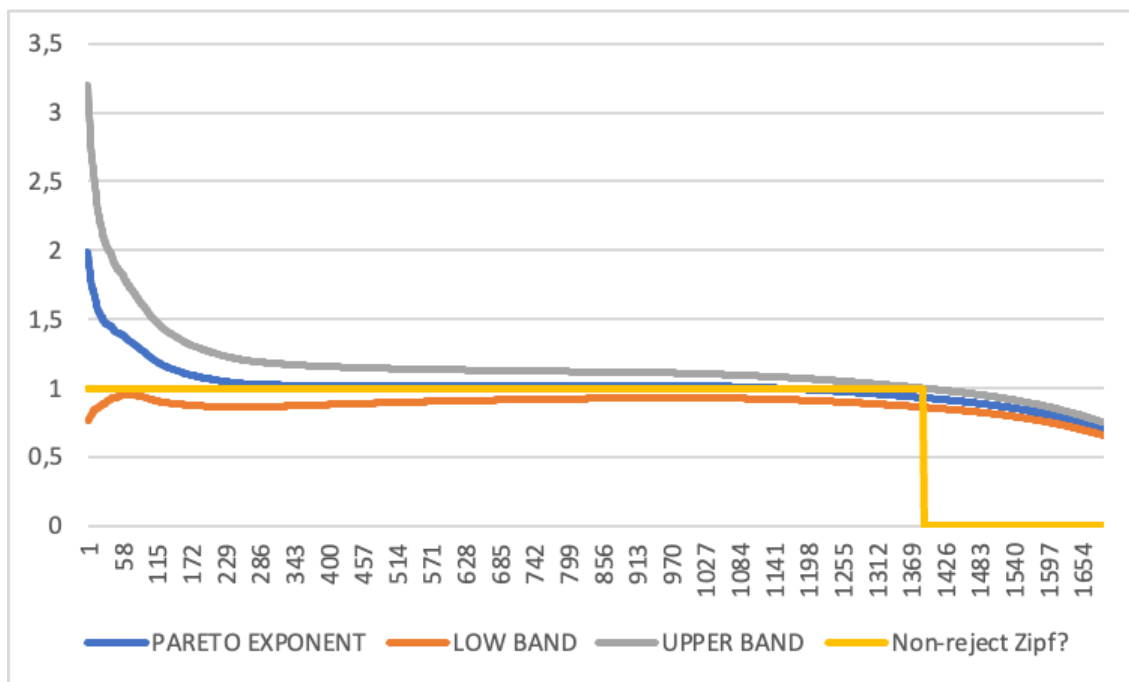
Así, se pueden establecer tres divisiones. Por un lado, hasta $N=2.387$ se establece una distribución donde el coeficiente de Zipf indica un mayor nivel de concentración. En los valores comprendidos entre 2.387 y 5.269 se cumple la ley de Zipf, con un $\beta \approx 1$. A partir del valor 5.269, el coeficiente de Zipf tenderá a ser menor que el de cumplimiento de Zipf, indicando una distribución más homogénea entre las poblaciones de menos habitantes, comportamiento característico de una distribución log-normal según la ley de Eeckhout (2004).

4.2.4. RENTA BAJA

Los países de renta baja aportan la menor cantidad de datos a la muestra, aportando únicamente 1.710 datos. Aplicando a estos datos las regresiones rodantes, se obtienen los siguientes datos para los principales puntos de la muestra (Tabla 4-8), así como la distribución del coeficiente de Pareto para la totalidad de la distribución (Figura 4-5):

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	1,9829	0,6193	0,7690	3,1968	Sí
100	1,3093	0,1847	0,9473	1,6713	Sí
500	1,0191	0,0644	0,8928	1,1453	Sí
1.000	1,0205	0,0456	0,9311	1,1100	Sí
1.500	0,8911	0,0325	0,8273	0,9548	No
1.710	0,7045	0,0241	0,6573	0,7518	No

Tabla 4-8. Región de cumplimiento de Zipf para los países de renta baja.



Gráfica 4-5. Evolución del coeficiente de Pareto para los países de renta baja.

Para los países de renta baja, el cumplimiento de Zipf es mayoritario. No obstante, es importante destacar el mismo problema que acontece en los países de renta alta y en los países de renta media-baja, donde Zipf se cumple para los primeros coeficientes fruto de un intervalo de confianza de mucha amplitud para un nivel de confianza del 95%, derivado de un alto error estándar de 0,6193. En este caso, el coeficiente de Pareto obtenido es menor que para los países de renta media-baja y renta alta, pero sigue siendo elevado, de 1,9829.

A medida que la muestra aumenta, el error disminuye y el intervalo de confianza se hace más preciso en torno al valor de cumplimiento de la ley de Zipf ($\beta=1$). El punto de corte, donde Zipf dejará de cumplirse, se encuentra en el punto $N=1.410$. Cabe destacar que, a diferencia de lo observado en otros grupos de renta, en los países de renta baja el coeficiente de Pareto no desciende de forma tan significativa como en los otros grupos de renta, no mostrando de forma significativa una distribución log-normal como la de la ley de Eeckhout (2004), probablemente debido a la menor madurez de la red urbana.

4.3. RESULTADOS POR REGIÓN

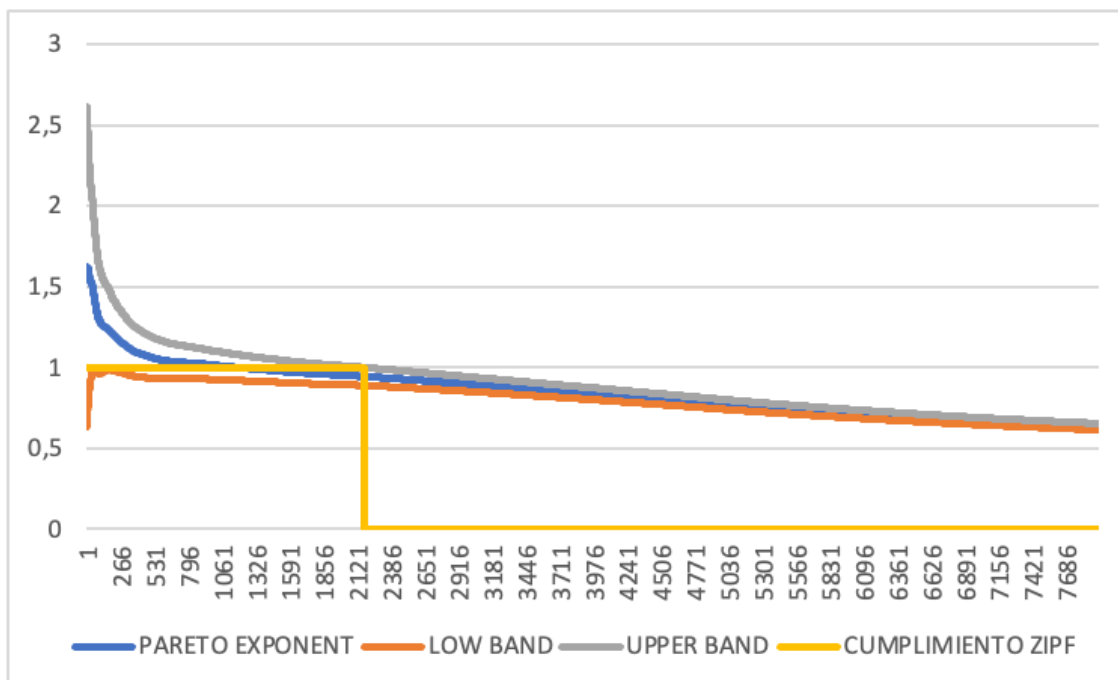
A continuación, se exponen los resultados obtenidos tras la aplicación de la ley de Zipf a las siguientes regiones globales, clasificadas por el Banco Mundial: África, Este Asiático y Pacífico, Europa y Asia Central, Latinoamérica y el Caribe, Oriente Medio y Norte de África, Sur de Asia.

4.3.1. ÁFRICA

Los países contenidos en este grupo aportan al análisis un total de 7.962 datos, para los cuales se ha llevado a cabo el análisis de regresiones de series rodantes, obteniéndose los siguientes resultados. Por un lado, se muestran los resultados para puntos clave (Tabla 4-9), así como la evolución del coeficiente de Zipf para la totalidad de los datos analizados (Figura 4-6):

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	1,6218	0,5066	0,6290	2,6147	Sí
100	1,3384	0,1888	0,9683	1,7085	Sí
500	1,0681	0,0675	0,9357	1,2004	Sí
1.000	1,0152	0,0454	0,9263	1,1042	Sí
5.000	0,7723	0,0154	0,7420	0,8026	No
7.962	0,6334	0,0100	0,6137	0,6531	No

Tabla 4-9. Valores puntuales clave para los países de África.



Gráfica 4-6. Evolución del coeficiente de Pareto para los países de África.

Para los países de África, la evolución de coeficiente de Pareto estimado comienza presentando un valor elevado, de 1,6218 para N=20 (Port Harcourt, Nigeria, con 2.120.000 habitantes) que va decreciendo a medida que se añaden valores a la muestra,

hasta alcanzar un valor de 0,6334 para la totalidad de la muestra. Para el conjunto de estos países, se observa el cumplimiento de la ley de Zipf al 95% de confianza hasta el valor $N=2192$ (Rano, Nigeria, con 17.972 habitantes), valor a partir del cual el coeficiente de Zipf ideal de $\beta=1$ deja de estar incluido en el intervalo de confianza.

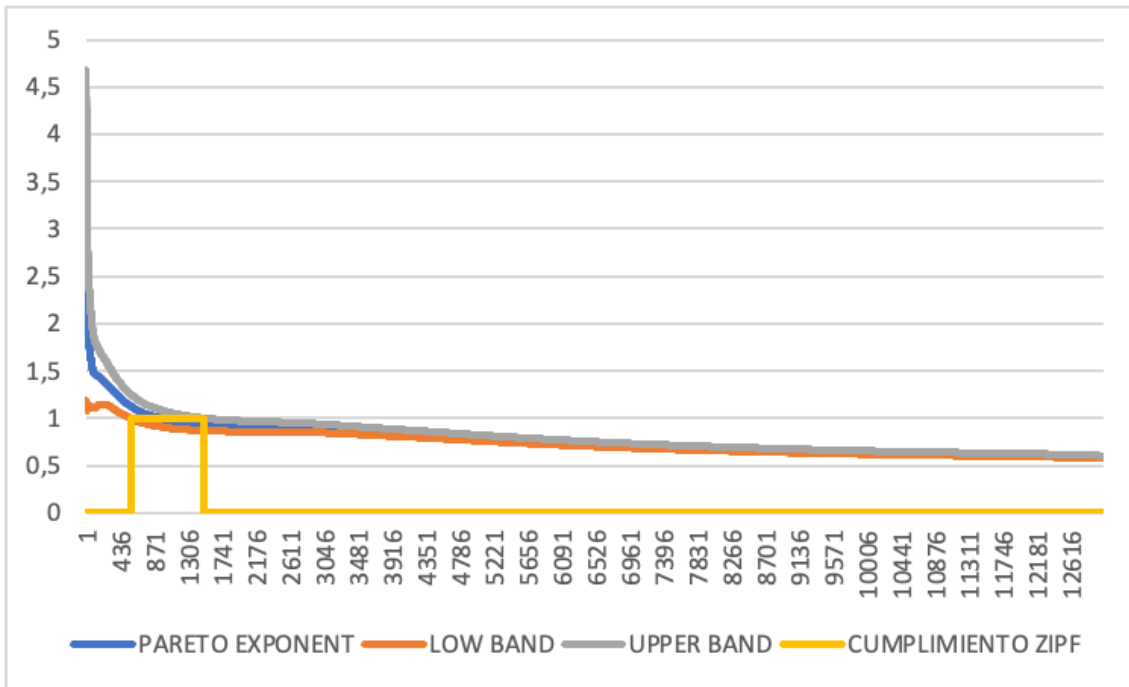
Es importante destacar que, para los primeros coeficientes estimados, el cumplimiento de Zipf no se da por un ajuste del intervalo de confianza en torno al valor unitario, sino que se debe a que, debido a un alto error estándar, el intervalo de confianza obtenido tiene una gran amplitud, indicativo de que se produce una gran concentración de población en torno a las primeras $N=20$ urbes. Una vez superada la ventana de cumplimiento de Zipf, el coeficiente de Pareto desciende por debajo de la unidad, hasta alcanzar un valor de 0,6334 para la muestra completa, en un comportamiento consistente con la distribución log-normal propuesta por Eeckhout (2004).

4.3.2. ESTE ASIÁTICO Y PACÍFICO

Este grupo aporta un total de 13.069 datos a la muestra total. Tras hacer el análisis de estos datos, se obtienen los siguientes resultados para puntos clave (Tabla 4-10), así como la evolución del coeficiente de Zipf para toda la muestra (Figura 4-7):

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	2,8996	0,9057	1,1245	4,6747	No
100	1,5331	0,2163	1,1092	1,9570	No
500	1,1818	0,0747	1,0354	1,3283	No
1.000	0,9941	0,0444	0,9070	1,0812	Sí
5.000	0,7952	0,0159	0,7640	0,8264	No
10.000	0,6382	0,0090	0,6206	0,6559	No
13.069	0,5893	0,0073	0,5750	0,6036	No

Tabla 4-10. Valores puntuales clave para los países del Este asiático y Pacífico.



Gráfica 4-7. Evolución del coeficiente de Pareto para los países del Este asiático y Pacífico.

Para este grupo de países, se observa un valor inicial del coeficiente de Pareto muy elevado, comenzando en 2,8996 para N=20 (Chongqing, China, con 7.457.599 habitantes), indicativo de una gran concentración en torno a los núcleos más poblados, e incumpliendo Zipf. El decrecimiento del coeficiente a partir de N=20 es muy elevado, alcanzando un valor de 0,5893 para toda la muestra. De los elementos clave, Zipf se cumple únicamente para N=1000, donde el intervalo de confianza se ajusta más acotadamente en torno al valor unitario del coeficiente de Pareto.

Gráficamente, se observa que la ley de Zipf se cumple para los valores comprendidos entre $588 \leq N \leq 1.533$ (con N=588 siendo Cileungsir, Indonesia, con 288.347 habitantes, y n=1.533 siendo Hede, China, con 89.107 habitantes). Entre estos valores, el intervalo de confianza se ajusta en torno al valor de $\beta=1$, incluyendo al mismo en el intervalo de confianza para un nivel de confianza del 95%.

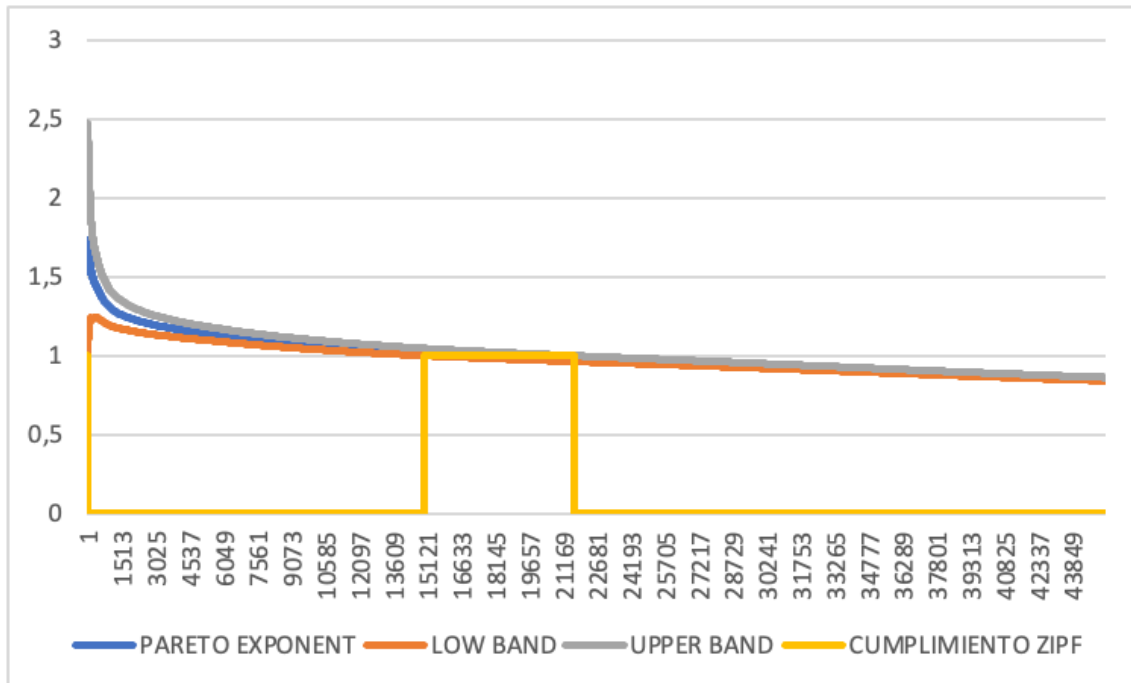
Se establecen de este modo tres grandes grupos entre los datos: en primer lugar, los coeficientes correspondientes a $N < 588$, los cuales son superiores al valor unitario, indicando una mayor concentración entre estos núcleos de población, en segundo lugar, los valores ya mencionados donde Zipf se cumple; y finalmente, los valores correspondientes a $N > 1.533$, valores para los cuales el coeficiente será menor que 1, indicando una mayor homogeneidad en la distribución poblacional, un comportamiento característico de la distribución log-normal propuesta por Eeckhout (2004).

4.3.3. EUROPA Y ASIA CENTRAL

Estos países aportan a la muestra un total de 45.339 datos. Aplicando el mismo análisis, se obtienen los siguientes datos para puntos clave (Tabla 4-11), así como la evolución del coeficiente de Zipf para toda la muestra (Figura 4-8):

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	1,5305	0,4780	0,5935	2,4675	Sí
100	1,6554	0,2335	1,1977	2,1132	No
500	1,4129	0,0893	1,2378	1,5879	No
1.000	1,3063	0,0584	1,1918	1,4207	No
10.000	1,0709	0,0151	1,0412	1,1006	No
25.000	0,9639	0,0086	0,9470	0,9808	No
45.357	0,8515	0,0057	0,8404	0,8626	No

Tabla 4-11. Valores puntuales clave para los países de Europa y Asia Central.



Gráfica 4-8. Evolución del coeficiente de Pareto para los países de Europa y Asia Central.

Para el caso de Eurasia, se observa el mismo problema ya mencionado anteriormente. Se da el cumplimiento de Zipf para $N=20$ (Viena, Austria, con 1.691.468 habitantes), pero es debido a un error estándar elevado, que provoca que el intervalo de confianza para un nivel de confianza del 95% tenga una amplitud muy elevada, que incluye al coeficiente de Pareto. Para el resto de los valores clave, el intervalo de confianza irá disminuyendo de amplitud, así como el error estándar. En cuanto a la evolución del coeficiente de Pareto, este decrece de 1,5305 para $N=20$ hasta 0,8515 para $N=45.357$, indicando de este modo una distribución más homogénea que la existente en otras regiones para la totalidad de la muestra.

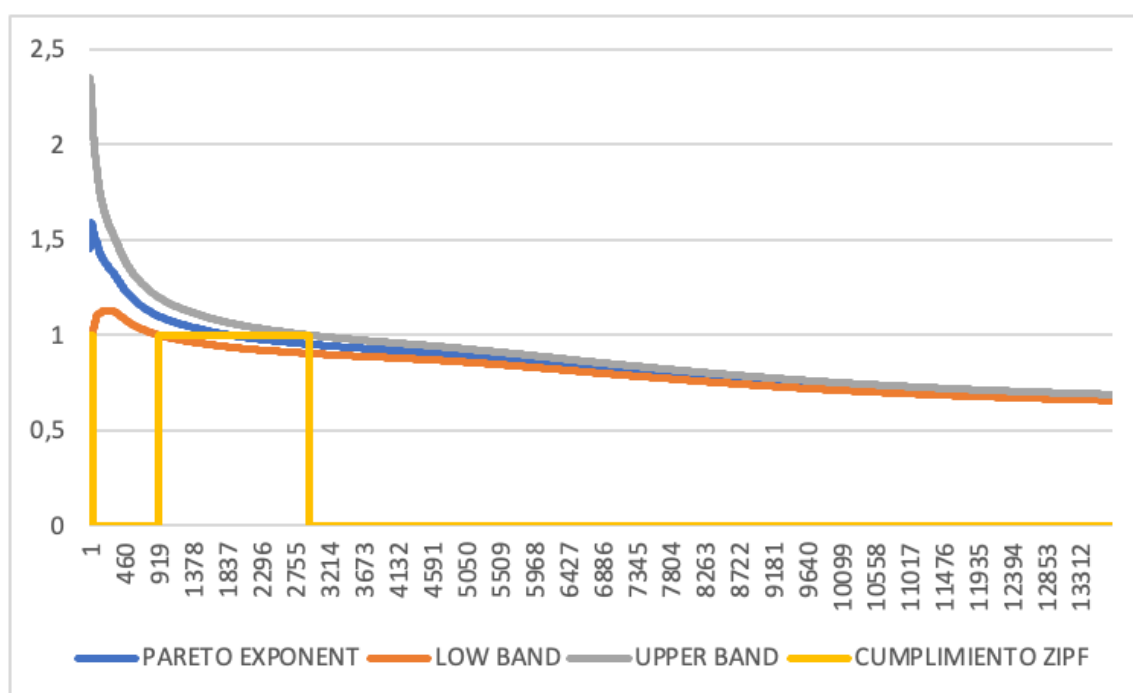
Gráficamente, se observa el cumplimiento de la ley de Zipf en dos regiones: por un lado, la región inicial (hasta $N=45$), donde la amplitud del intervalo de confianza incluye el valor unitario del coeficiente de Pareto, y, por otro lado, los valores comprendidos entre $15.018 \leq N \leq 21.675$ (con $N=15.018$ siendo Ølstykke, Dinamarca, con 5.348 habitantes y $N=21.675$ siendo Lunca Cetățuii, Rumanía, con 3.386 habitantes), región en la que el intervalo de confianza se ajusta en torno a 1 de forma más eficiente. El valor final para toda la muestra, de 0,8515, sugiere una tendencia a la homogeneización en la cola inferior de la distribución, compatible con la tendencia hacia una distribución log-normal como la propuesta por Eeckhout (2004).

4.3.4. LATINOAMÉRICA Y EL CARIBE

El conjunto de estos países aporta un total de 13.772 datos a la muestra. Tras analizarlos se obtienen los siguientes para los puntos clave (Tabla 4-12), así como la evolución del coeficiente de Zipf para toda la muestra (Figura 4-9):

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	1,4583	0,4555	0,5655	2,3511	Sí
100	1,4994	0,2115	1,0848	1,9140	No
500	1,2331	0,0779	1,0803	1,3859	No
1.000	1,0902	0,0487	0,9947	1,1857	Sí
5.000	0,8953	0,0179	0,8602	0,9304	No
10.000	0,7326	0,0104	0,7123	0,7529	No
13.772	0,6727	0,0081	0,6568	0,6886	No

Tabla 4-12. Valores puntuales clave para los países de Latinoamérica y el Caribe.



Gráfica 4-9. Evolución del coeficiente de Pareto para los países de Latinoamérica y el Caribe.

En esta región ocurre el mismo problema en los parámetros clave: Zipf se cumple para $N=20$ no debido a un ajuste eficiente del intervalo de confianza, sino debido a su gran amplitud que abarca el valor unitario del coeficiente de Pareto. No obstante, el valor estimado es de 1,4583, indicando una mayor concentración que la esperada si se cumpliera la ley de Zipf. Al igual que en el resto de los casos, a medida que se amplía el número de elementos de la muestra, disminuye la amplitud del intervalo de confianza.

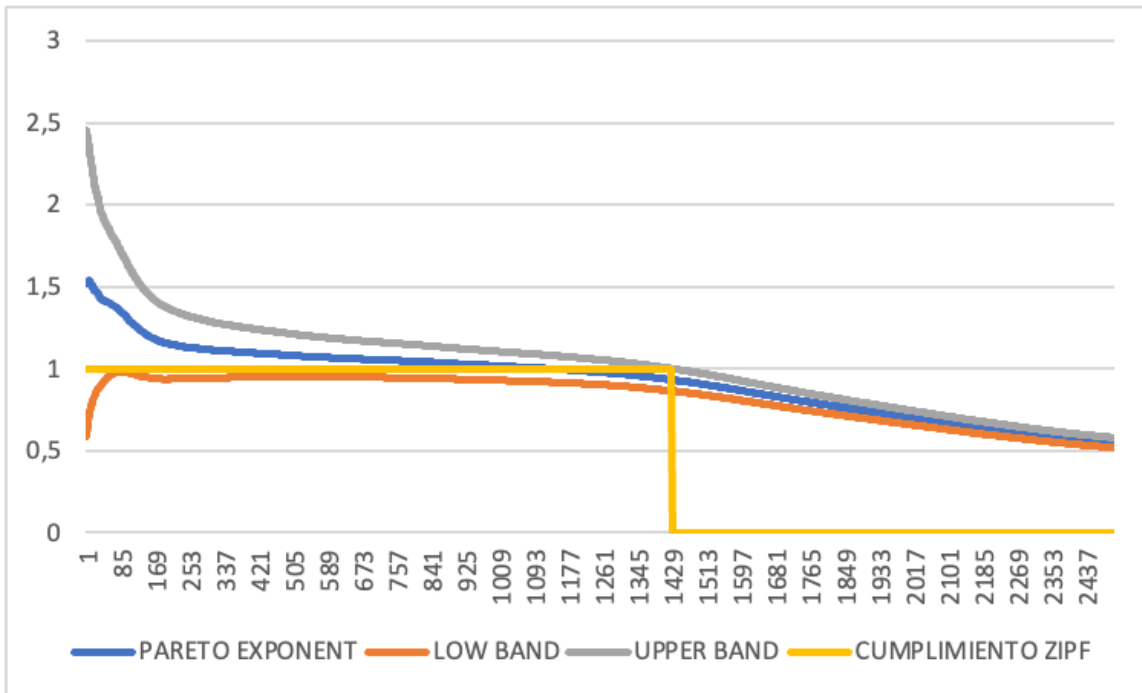
Gráficamente, se observa una región de cumplimiento de Zipf comprendida entre $950 \leq N \leq 2.974$ (siendo $N=950$ Chacao, Venezuela, con 64.609 habitantes, y $N=2.974$ siendo São Félix do Xingu, Brasil, con 16.477 habitantes), región donde los valores del intervalo de confianza se ajustan de forma más adecuada al coeficiente de Pareto. A partir del valor $N > 2.974$, el coeficiente estimado es inferior al valor unitario esperado para el cumplimiento de Zipf, indicando una mayor homogeneidad para los núcleos de población con menos habitantes. Este comportamiento es consistente con una distribución log-normal como la propuesta por la ley de Eeckhout (2004).

4.3.5. ORIENTE MEDIO Y NORTE DE ÁFRICA

El conjunto de estos países aporta un total de 2.524 datos a la muestra que, tras analizarse por la metodología de regresiones rodantes, revelan los siguientes datos para elementos clave (Tabla 4-13), así como la siguiente evolución para el coeficiente de Zipf (Figura 4-10):

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	1,5246	0,4762	0,5912	2,4579	Sí
100	1,3561	0,1913	0,9811	1,7311	Sí
500	1,0845	0,0686	0,9501	1,2188	Sí
1.000	1,0203	0,0456	0,9309	1,1097	Sí
1.500	0,9138	0,0334	0,8484	0,9792	No
2.000	0,7125	0,0225	0,6683	0,7567	No
2.524	0,5488	0,0154	0,5185	0,5791	No

Tabla 4-13. Valores puntuales clave para los países de Oriente Medio y el Norte de África.



Gráfica 4-10. Evolución del coeficiente de Pareto para los países de Oriente Medio y el Norte de África.

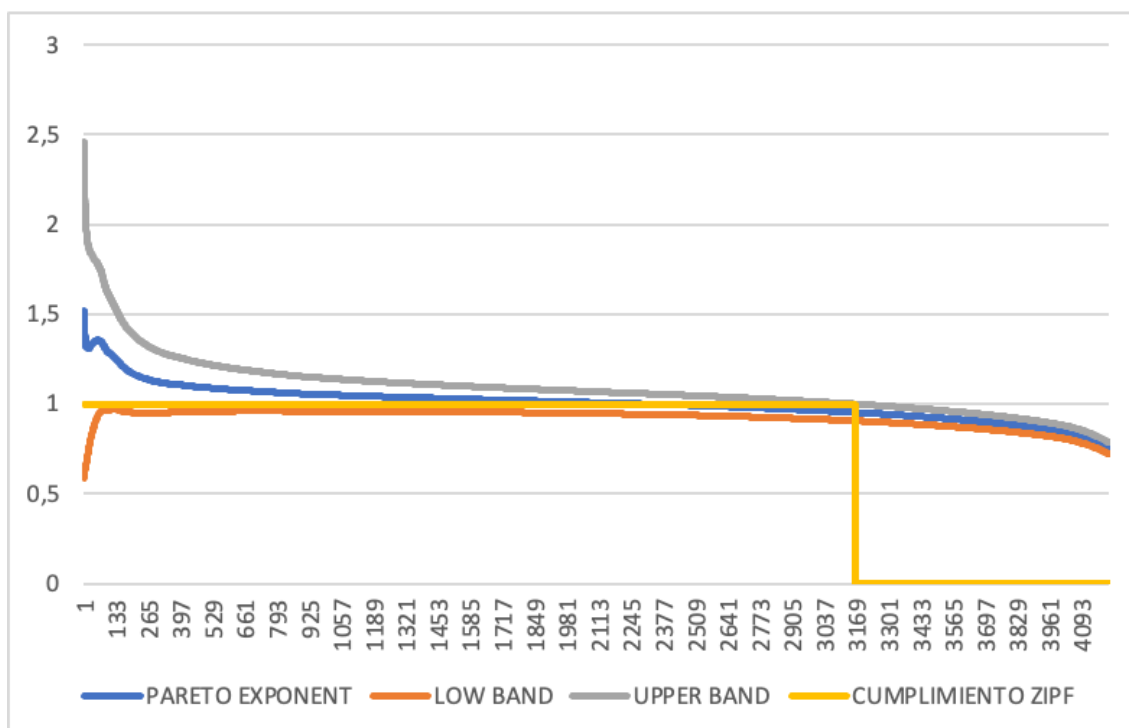
Para este caso, el cumplimiento de la ley de Zipf es total entre $20 \leq N \leq 1.446$. No obstante, ocurre igual que en casos anteriores, donde el cumplimiento para los primeros valores responde a una elevada amplitud del intervalo de confianza, y no a un valor acotado eficientemente en torno al valor unitario. Esto es observable en los valores del coeficiente estimados para $N=20$ y $N=100$, superiores al valor esperable si Zipf se cumpliera, de $\beta=1$. Gráficamente, se observa un cumplimiento de Zipf hasta el ya mencionado valor de 1.446 (Azor, Israel, con 12.957 habitantes). A partir de ese valor, el decrecimiento del exponente de Pareto es muy notable, hasta alcanzar un valor de 0,5488, indicativo de una gran homogeneidad entre los núcleos de menor población. Este comportamiento es consistente con el patrón log-normal descrito por Eeckhout (2004).

4.3.6. SUR DE ASIA

Esta región global aporta 4.227 datos a la muestra total, que, tras haber sido analizados, ofrecen los siguientes datos para los puntos clave (Tabla 4-14), así como la siguiente evolución del coeficiente de Zipf (Figura 4-11):

Nº MUESTRA	BETA ESTIMADO	ERROR ESTÁNDAR	IC95% INFERIOR	IC95% SUPERIOR	CUMPLE ZIPF ($\beta=1$)
20	1,5231	0,4757	0,5907	2,4555	Sí
100	1,3289	0,1875	0,9615	1,6963	Sí
500	1,0939	0,0692	0,9584	1,2295	Sí
1.000	1,0527	0,0471	0,9604	1,1449	Sí
2.000	1,0123	0,0320	0,9496	1,0751	Sí
4.227	0,7521	0,0164	0,7200	0,7842	No

Tabla 4-14. Valores puntuales clave para los países del sur de Asia.



Gráfica 4-11. Evolución del coeficiente de Pareto para los países del sur de Asia.

Para el caso del sur de Asia, se observa un cumplimiento elevado de Zipf, que abarca los valores $20 \leq N \leq 3.183$. No obstante, para los primeros valores vuelve a darse el problema de la elevada amplitud de los intervalos de confianza, mientras que la estimación del coeficiente de Pareto indica un incumplimiento de Zipf, con unos coeficientes de Pareto muy elevados para $N=20$ y $N=100$, que muestran una mayor concentración en torno a las urbes más pobladas del ideal si se diera el cumplimiento total de Zipf. Es de destacar que la caída del coeficiente de Pareto estimado es de los menores, junto con el caso de Eurasia, con una disminución de aproximadamente el 50%. El valor final, de 0,7521, se mantiene relativamente elevado en comparación a otras regiones, sugiriendo una homogeneización moderada en la parte inferior, compatible con una distribución log-normal como la propuesta por Eeckhout (2004).

5. DISCUSIÓN DE RESULTADOS

Este trabajo ha analizado empíricamente el cumplimiento de la ley de Zipf y la validez del modelo de Eeckhout (2004) a nivel mundial, así como segmentando los datos en función de renta y región geográfica, utilizando regresiones rodantes y la corrección sugerida por Gabaix e Ibragimov (2011). Los resultados obtenidos sugieren que la ley de Zipf no es una ley de cumplimiento universal, sino que su cumplimiento se registra en ventanas acotadas de tamaño urbano, cuyo tamaño varía según la región geográfica o

según su nivel de renta. En cuanto a la ley de Eeckhout (2004), los resultados demuestran su pertinencia, observándose una distribución log-normal para todos los grupos de ciudades analizados, tal como predijo su modelo de equilibrio entre aglomeración y congestión bajo libre movilidad. Cabe destacar que este estudio se distingue por el empleo de una base de datos única y armonizada, que abarca todas las ciudades del mundo de más de 1.000 habitantes. Con este enfoque se consigue homogeneizar criterios metodológicos, ofreciendo una perspectiva comparativa robusta.

5.1.DISCUSIÓN DEL RESULTADO GLOBAL

A nivel global, se ha observado que, para un total de 117.054 ciudades analizadas, la ley de Zipf se cumple únicamente para un total de 7.861 ciudades, con poblaciones comprendidas entre 26.895 habitantes (Prairieville, EEUU), y 16.770 habitantes (Holzkirchen, Alemania). Así, al analizar todas las ciudades en conjunto, se detecta un patrón en el que existen tres grupos de poblaciones:

En primer lugar, se encuentran aquellas urbes con una población superior a 26.895 habitantes. Para estas poblaciones, el coeficiente de Pareto es superior a uno de forma consistente ($\beta > 1$), siendo un elemento clave para identificar economías de aglomeración. De este modo, se revela un panorama en el cual las ciudades más grandes del mundo ejercen de polos de atracción, atrayendo tanto población como capital, generando entornos dinámicos en los cuales los beneficios de la gran densidad poblacional, tales como grandes mercados laborales, oportunidades y acceso a servicios especializados, superan en gran medida a los costes derivados de la congestión, creando círculos virtuosos o viciosos, que dependerán en gran medida de cómo se estructura la ciudad y de los planes de ordenación urbana seguidos para asimilar una población cada vez mayor.

En segundo lugar, se encuentran aquellas urbes que sí cumplen la ley de Zipf. Son las urbes comprendidas entre 16.770 y 26.895 habitantes, un total de 7.861 urbes a nivel global. Son aquellas ciudades donde se ha encontrado el punto óptimo entre los beneficios de la aglomeración y sus costes. Son urbes que brindan oportunidades a sus habitantes, a la vez que limitan los potenciales costes derivados de la aglomeración, situándose así en un punto óptimo en un análisis coste-beneficio.

En tercer lugar, se encuentran las urbes con población inferior a 16.770 habitantes. Estas ciudades suponen la mayor parte de las poblaciones del mundo, para las cuales el coeficiente de Pareto es inferior a 1 de forma consistente ($\beta < 1$), presentando una

distribución más homogénea de lo esperado. Por debajo de este umbral de población, gana sentido otra lógica poblacional, basada en elementos como la economía rural, o la dependencia de recursos naturales locales para la subsistencia, elementos que presentan un mayor peso para estas urbes que la integración en redes urbanas globales. Esta estructura tripartita global encuentra su fundamento teórico en el modelo de Eeckhout (2004), que predice una distribución log-normal de las ciudades como resultado del equilibrio entre las fuerzas de aglomeración y congestión bajo libre movilidad. Como señalan Peña y Sanz-Gracia (2021), Eeckhout establece este resultado en un contexto donde la libre movilidad genera una fuerte tendencia hacia la aglomeración urbana, vinculada a un aumento en la productividad de los trabajadores. Los hallazgos empíricos son consistentes con este marco: la "ventana de Zipf" (segundo grupo) representa el rango donde el equilibrio es óptimo, flanqueado por la aglomeración dominante (primer grupo) y la homogeneidad dispersa (tercer grupo) que el propio modelo anticipa para las colas de la distribución.

No obstante, esta estructura supone un promedio a nivel global que oculta realidades profundamente dispares. Con el fin de poder comprender las características específicas que el promedio oculta, se lleva a cabo el análisis del cumplimiento de la ley de Zipf, separando por regiones de renta y por regiones geográficas.

5.2. ANÁLISIS POR GRUPOS DE RENTA

El análisis para las ciudades de renta alta presenta una ventana de cumplimiento real de Zipf (sin considerar las primeras ciudades de la distribución, donde el cumplimiento se deriva de un alto error estándar) de 5.277 urbes, comprendidas entre 5.918 y 4.609 habitantes. Estos resultados pueden tener su origen en un modelo de desarrollo donde los costes de la congestión superan a los beneficios, propiciando la aparición de núcleos periféricos en torno a las principales metrópolis. Estos núcleos, si bien no forman parte de las grandes urbes, están conectados con las mismas y tienen servicios de similar calidad disponibles, favoreciendo el establecimiento de la gente en los mismos, y mejorando la calidad de vida de la población en general. En este sentido, el resultado observado en los países de renta alta indica la existencia de un mapa urbano equilibrado y repartido, en el que, si bien hay una serie de grandes núcleos urbanos, existe también una gran extensión de urbes de diferentes tamaños.

El panorama cambia de forma radical al analizar las urbes de los países considerados de renta media. En estos países, el cumplimiento de la ley de Zipf se desplaza hacia núcleos urbanos notablemente mayores, entre 64.585 y 191.406 habitantes para los países de renta media-alta, y entre 17.944 y 52.516 habitantes para los países de renta media-baja. En estos países es notable el efecto que genera la urbanización acelerada típica de los países en desarrollo, donde la tendencia es hacia la atracción a las ciudades de grandes masas de población. Junto a estas grandes ciudades, se observa la existencia de una ventana de equilibrio en las ciudades de tamaño intermedio, urbes que presentan un dinamismo suficiente para atraer inversión y talento, evitando los costes de aglomeración de las grandes ciudades. En definitiva, esta ventana de Zipf presenta los núcleos que descentralizan el crecimiento y edificando un sistema urbano más equilibrado.

Para el caso de los países de renta baja, se observa una estructura urbana en la que hay un cumplimiento prácticamente total de la ley de Zipf. No obstante, este resultado es engañoso, dado que las primeras urbes presentan un coeficiente de Pareto elevado, indicando una atracción desproporcionada de unas pocas grandes urbes, mientras que el resto del territorio se compone de núcleos pequeños y asentamientos rurales, no existiendo un escalón intermedio como en los casos anteriores, de urbes de tamaño mediano. Así, la evolución de Pareto para los países de renta baja refleja una situación de fragmentación entre los núcleos urbanos.

5.3. ANÁLISIS POR REGIÓN GEOGRÁFICA

Además del nivel de renta, la ubicación geográfica y el contexto histórico también tienen una huella única en la distribución urbana. En este sentido, el análisis por regiones muestra cómo factores como la integración política, los modelos de crecimiento o las barreras físicas ejercen influencia en los patrones de distribución poblacional.

Para el continente africano se observa que el cumplimiento de la ley de Zipf alcanza las 2.192 ciudades más pobladas, hasta Rano, Nigeria, con 17.972 habitantes. Con los resultados observados, es posible afirmar que en África existe una realidad multipolar, caracterizada por la coexistencia de una serie de polos urbanos que actúan como imanes regionales, con una clara tendencia a la concentración, que conviven con una gran cantidad de asentamientos pequeños y dispersos, típicos de regiones arraigadas en lo rural y con escasa integración económica. En este sentido, se puede observar que la tendencia

de África es similar a la que presenta en el análisis regional el grupo de países de renta baja.

La región del este asiático y el Pacífico muestra un perfil radicalmente diferente, caracterizado por una concentración urbana masiva, dominada por ciudades como Chongqing. No obstante, ley de Zipf sí encuentra una ventana de cumplimiento en las ciudades comprendidas entre los 89.107 y los 288.347 habitantes. Este patrón de cumplimiento es indicativo de un modelo de desarrollo muy concreto, caracterizado por una industrialización acelerada, repartida entre unos polos manufactureros y financieros, que se complementa con una serie de ciudades industriales secundarias y terciarias. Finalmente, una vez pasada la capa de ciudades industriales, se observa la existencia de una gran base de núcleos homogéneos, vinculados a la producción manufacturera o a economías locales. En definitiva, este modelo de urbanismo refleja una estructura construida en torno al papel de estos países de fábrica global.

Europa y Asia central, por otro lado, muestran una situación más equilibrada, con una ventana de cumplimiento de Zipf situada en ciudades notablemente más pequeñas, comprendidas entre 3.386 y 5.348 habitantes. En este caso, se observa una red urbana más integrada, y con menos diferencias entre sí que el resto de dinámicas. Este patrón tiene su explicación en varios factores, entre los que se incluyen un proceso de urbanización temprano, que ha permitido la aparición de varios núcleos de población cercanos entre sí. Además, para el caso de los países de la Unión Europea, las políticas de integración regional han sido esenciales para fortalecer regiones y poblaciones periféricas, evitando de este modo una elevada concentración en las ciudades. De este modo, se ha conseguido construir un sistema en el cual ciudades de pequeño tamaño están conectadas con grandes urbes sin dejar de disponer por ello de los servicios y de las oportunidades de las mismas.

Para el caso de Latinoamérica y el Caribe se observa un caso particular de urbanización temprana y rápida en países de renta media. Se observa una dinámica en la que ocurre un fenómeno de concentración inicial, con las capitales acaparando recursos y población, junto con una ventana de cumplimiento de Zipf para las urbes con entre 16.477 y 64.609 habitantes. Estas urbes, de tamaño intermedio, reflejan una tendencia clara hacia la aparición de un sistema urbano diversificado, formado por ciudades secundarias, capitales de Estados o regiones y pequeños polos económicos que crecen por elementos externos a los de las grandes ciudades, como pueden ser los recursos naturales o el

turismo. Esta tendencia puede interpretarse bajo la óptica del modelo de Eeckhout (2004), en el cual la libre movilidad y la búsqueda de maximización de utilidad lleva a la aglomeración en las grandes capitales, mientras que, una vez superado el equilibrio entre beneficios y costes de la aglomeración, emergen fuerzas centrífugas que favorecen la aparición de núcleos de población intermedios. Se observa así una distribución urbana en transición, en la que, si bien las grandes urbes tienen peso, hay una tendencia creciente a la diversificación urbana que, con las políticas adecuadas podría acelerarse en el futuro.

En cuanto a Oriente Medio y el norte de África, se observa un patrón en el que se cumple la ley de Zipf para todas las urbes hasta 12.957 habitantes, pero, a la vez, los coeficientes de Pareto superiores a 1 para las primeras urbes indican una aglomeración superior a la esperada en torno a las principales ciudades de este grupo de países, compuestas por ciudades petroleras y capitales nacionales. Mientras tanto, en el resto se observa una tendencia a la homogeneidad, con un coeficiente de Pareto que desciende hasta alcanzar valores de 0,54. En este sentido, y dados los países que conforman este grupo, se puede afirmar que esta homogeneidad no responde a factores como los mencionados para otras regiones geográficas, sino que responden a la escasez de recursos naturales, que obliga a generar este estilo de distribución homogénea y dispersa en torno a la costa, a los oasis o a los ríos. De este modo, para estas regiones, existe una clara disparidad entre grandes urbes y pequeños núcleos urbanos establecidos en torno a recursos hídricos.

Finalmente, en el sur de Asia, se observa un caso extremo de elevada densidad demográfica. Se observa un cumplimiento de la ley de Zipf para prácticamente la totalidad de la muestra, reflejando una situación en la que la gran cantidad de población residente en esta región hace que existan una cantidad de asentamientos suficiente para forzar una transición suave de la curva de Zipf del valor de aglomeración para las primeras urbes hasta el valor de homogeneidad de las últimas. En definitiva, en esta zona el cumplimiento de la ley de Zipf se da debido a la cantidad tan elevada de alojamientos existentes, fruto de una de las mayores densidades poblacionales a nivel global.

5.4.UTILIDAD GENERAL DE LA LEY DE ZIPF

En definitiva, la principal utilidad de este estudio ha sido la de transformar la ley de Zipf de una mera curiosidad estadística en una herramienta práctica de análisis territorial. En este sentido, la ley de Zipf es especialmente útil, dado que, conociendo la ventana de equilibrio donde se da el cumplimiento de la ley, se puede estimar el grado de madurez

de un sistema urbano, permitiendo, de este modo, identificar núcleos estratégicos con el fin de hallar las urbes de tamaño óptimo y diseñar estrategias de descongestión; para poder detectar y prevenir el declive demográfico de regiones y municipios; y para poder tener una medida cuantitativa que evalúe el grado de cumplimiento de las políticas de cohesión territorial. Cabe destacar también que la ley de Zipf se cumple con mayor amplitud en los países de renta baja, debido a la distribución bipolar existente en estos países, caracterizada por la existencia de unas pocas grandes urbes junto a una gran cantidad de urbes pequeñas, sin apenas ciudades de rango medio, haciendo que la distribución se parezca más a Zipf en un rango mayor de ciudades.

6. CONCLUSIONES

En este trabajo se ha llevado a cabo un análisis empírico del cumplimiento de la ley de Zipf en el sistema urbano mundial. Con esta finalidad, se construyó una base de datos original con 117.054 ciudades de más de 1.000 habitantes de todo el mundo, a la que se le aplicó un análisis de regresiones rodantes con la corrección de sesgo de Gabaix e Ibragimov (2011), estimándose coeficientes de Pareto a medida que se incorporaban más ciudades. El análisis se llevó a cabo tanto de forma global como desagregando en función del nivel de renta y de la ubicación geográfica.

Con este trabajo, se ha buscado aportar un nuevo nivel de análisis de la ley de Zipf. Frente a estudios previos, que realizaban el análisis en países concretos, o en muestras específicas de grandes ciudades, este trabajo propone una comprobación sistemática y global de la ley de Zipf, trasladando el debate teórico de Eeckhout (2004) al terreno mundial. Su innovación no se limita al aspecto cuantitativo de considerar todas las ciudades de más de 1.000 habitantes del mundo, sino que también es cualitativa, al ofrecer una imagen del cumplimiento de Zipf para diferentes regiones y grupos de renta.

Los resultados revelan que la ley de Zipf no es una ley universal, sino que es más bien un fenómeno localizado en ventanas muy concretas del espectro urbano, validando con ello la crítica central de la sensibilidad muestral. Asimismo, dichos resultados demuestran el cumplimiento en todos los casos de la ley de Eeckhout (2004), confirmando de este modo, que la distribución de los coeficientes de Pareto sigue una distribución log-normal. Se identifican tres umbrales claros: un régimen de aglomeración en las ciudades de mayor tamaño de todas las regiones ($\beta > 1$); una región de equilibrio zipfiano en las ciudades de tamaño intermedio ($\beta \approx 1$), cuyos límites varían en función de la renta y la región (desde

5.000 habitantes para ciudades de renta alta, y hasta 200.000 en países de renta media-alta); y una región de homogeneidad ($\beta < 1$), donde la distribución es más igualitaria de lo predicho por una ley de potencia pura.

Los objetivos planteados por el trabajo se han cumplido, consiguiendo comprender la distribución global de las ciudades de forma detallada, y desvelando la existencia de la ya mencionada estructura tripartita. En segundo lugar, se ha aplicado y comprobado la ley de Zipf en diferentes submuestras, demostrando la importancia del contexto económico y geográfico en sus resultados. Finalmente, se ha logrado ofrecer una nueva perspectiva empírica con una perspectiva actualizada y una metodología robusta, demostrando que la ley de Zipf es útil precisamente por su capacidad de diagnosticar las dinámicas específicas de aglomeración y equilibrio de cada sistema urbano.

No obstante, y a pesar de lo anterior, este trabajo presenta una serie de limitaciones, derivadas, principalmente, de su análisis global. En primer lugar, la base de datos, si bien se ha depurado y organizado para este trabajo, procede de una fuente secundaria, y de un proceso de agregación que puede no tener en cuenta realidades urbanas complejas. En segundo lugar, este análisis es un análisis de sección cruzada, donde no se tiene en cuenta la evolución temporal de los coeficientes obtenidos. Finalmente, el análisis por grupos de renta y geográficos, a pesar de su utilidad, puede no tener en cuenta heterogeneidades internas dentro de cada grupo,

Las limitaciones mencionadas abren sendos caminos para investigaciones futuras. Se pueden llevar a cabo análisis de validación empleando fuentes primarias, con la finalidad de obtener datos más fieles, que permitan una comprensión todavía mayor de la ley de Zipf. Por otro lado, otra línea de investigación podría abarcar el análisis de series temporales para comprobar la evolución de los coeficientes. Asimismo, se podrían evaluar los aspectos concretos que afectan al cumplimiento de la ley de Zipf en cada región de renta o geográfica, atendiendo a factores políticos, históricos o económicos). Finalmente, se puede llevar a cabo una ampliación del enfoque. Investigando la relación del tamaño óptimo propuesto por la ley de Zipf con indicadores de nivel de vida o de sostenibilidad, ampliando el grado de utilidad de la ley de Zipf de una ley de ordenación urbana a una herramienta de políticas públicas.

En definitiva, este trabajo muestra que la distribución del tamaño de las ciudades a nivel global es un fenómeno complejo que no obedece a una simple ley de potencia. Más que una ley universal, Zipf es una potente herramienta de diagnóstico del grado de madurez,

concentración, y equidad territorial de un sistema urbano, ofreciendo un marco sólido para el diseño de políticas de planificación territorial en un mundo cada vez más urbanizado.

7. BIBLIOGRAFÍA

Bernal, G.L., Nieto, C.E. (2006): “Evolución del coeficiente de Zipf para Colombia en el siglo XX”. Pontificia Universidad Javeriana.

Clauset, A., C. R. Shalizi, M. E. J. Newman (2009): "Power-law distributions in empirical data." *SIAM Review*, Vol. 51, N° 4, páginas 661-703.

Eeckhout, J. (2004): "Gibrat's Law for (All) Cities." *American Economic Review*, Vol. 94, N° 5, páginas 1429-1451.

Fan, C. C., Casetti, E. (1994): "The Spatial and Temporal Dynamics of US Regional Income Inequality, 1950-1989." *The Annals of Regional Science*, Vol. 28, páginas 177-196.

Gabaix, X. (1999): "Zipf's Law and the Growth of Cities." *American Economic Review*, Vol. 89, N° 2, páginas 129-132.

Gabaix, X., Ibragimov, R. (2011): “Rank – 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents”. *Journal of Business & Economic Statistics*, Vol. 29, N° 1, páginas 24–39.

González-Val, R., Ximénez-De-Embún, D. P., Sanz-Gracia, F. (2024): “A long-term, regional-level analysis of Zipf's and Gibrat's laws in the United States”. *Cities*, Vol. 149, Artículo 104946.

Lanaspa, L.; Perdiguero, A. M.; Sanz, F. (2004): “La distribución del tamaño de las ciudades en España, 1900-1999”. *Revista de Economía Aplicada*, Vol. XII, N° 34, pp. 5-16.

Opendatasoft. (2024): Geonames - All Cities with a population > 1000 [recurso en línea]. Portal de datos públicos de Opendatasoft. Disponible en: https://public.opendatasoft.com/explore/dataset/geonames-all-cities-with-a-population-1000/export/?disjunctive.country&sort=population&rows=50&disjunctive.cou_name_en

Peña, G., Sanz-Gracia, F. (2021): “Zipf's exponent and Zipf's law in the BRICS: a rolling sample regressions approach”. *Economics Bulletin*, 41, 2543-2549.

Peng, G. (2010): “Zipf's law for Chinese cities: Rolling sample regressions”. *Physica A: Statistical Mechanics and its Applications*, Vol. 389, N° 18, pp. 3804-3813.

Persky, J. (1992): "Retrospectives: Pareto's Law". *Journal of Economic Perspectives*, Vol. 6, N° 2, pp. 181-192.

Puyol, R. (2012): "Del campo a la ciudad. El proceso de urbanización en el mundo". *Nueva Revista*, N° 138.

Restrepo Betancur, L. F. (2024): "Crecimiento poblacional urbano a nivel mundial en las últimas seis décadas". *Anales de Geografía de la Universidad Complutense*, Vol. 44, N° 1, pp. 277-295.

Rosen, K., Resnick, M. (1979): "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy". UC Berkeley Fisher Center Working Papers. Berkeley, CA.

Soo, K. T. (2005): "Zipf's Law for cities: An empirical examination". *Regional Science and Urban Economics*, Vol. 35, N° 3, páginas 239-263.

Zipf, G. K. (1949): *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge.

8. ANEXOS

8.1. ANEXO I

Dentro de cada grupo de renta, se incluyen los siguientes países:

- **Renta Alta:** Samoa Americana, Andorra, Antigua y Barbuda, Aruba, Australia, Austria, Bahamas, Baréin, Barbados, Bélgica, Bermudas, Islas Vírgenes Británicas, Brunéi Darussalam, Bulgaria, Canadá, Islas Caimán, Islas del Canal, Chile, Costa Rica, Croacia, Curazao, Chipre, Chequia, Dinamarca, Estonia, Islas Feroe, Finlandia, Francia, Polinesia Francesa, Alemania, Gibraltar, Grecia, Groenlandia, Guam, Guyana, Hong Kong RAE, China, Hungría, Islandia, Irlanda, Isla de Man, Israel, Italia, Japón, Corea del Sur, Kuwait, Letonia, Liechtenstein, Lituania, Luxemburgo, Macao RAE, China, Malta, Mónaco, Nauru, Países Bajos, Nueva Caledonia, Nueva Zelanda, Islas Marianas del Norte, Noruega, Omán, Palaos, Panamá, Polonia, Portugal, Puerto Rico (EE.UU.), Catar, Rumania, Federación Rusa, San Marino, Arabia Saudita, Seychelles, Singapur, San Martín (Parte Neerlandesa), República Eslovaca, Eslovenia, España, San Cristóbal y Nieves, San Martín (Parte Francesa), Suecia, Suiza, Trinidad y Tobago, Islas Turcas y Caicos, Emiratos Árabes Unidos, Reino Unido, Estados Unidos, Uruguay, Islas Vírgenes (EE.UU.).
- **Renta Media-Alta:** Albania, Argelia, Argentina, Armenia, Azerbaiyán, Bielorrusia, Belice, Bosnia y Herzegovina, Botsuana, Brasil, Cabo Verde, China, Colombia, Cuba, Dominica, República Dominicana, Ecuador, El Salvador, Guinea Ecuatorial, Fiyi, Gabón, Georgia, Granada, Guatemala, Indonesia, Irán, Irak, Jamaica, Kazajistán, Kosovo, Libia, Malasia, Maldivas, Islas Marshall, Mauricio, México, Moldavia, Mongolia, Montenegro, Macedonia del Norte, Paraguay, Perú, Samoa, Serbia, Sudáfrica, Santa Lucía, San Vicente y las Granadinas, Surinam, Tailandia, Tonga, Turquía, Turkmenistán, Tuvalu, Ucrania.
- **Renta Media-Baja:** Angola, Bangladés, Benín, Bután, Bolivia, Camboya, Camerún, Comoras, Congo, Costa de Marfil, Yibuti, Egipto, Suazilandia, Ghana, Guinea, Haití, Honduras, India, Jordania, Kenia, Kiribati, Kirguistán, Laos, Líbano, Lesoto, Mauritania, Micronesia, Marruecos, Myanmar, Namibia, Nepal, Nicaragua, Nigeria, Pakistán, Papúa Nueva Guinea, Filipinas, Santo Tomé y Príncipe, Senegal, Islas Salomón, Sri Lanka, Tayikistán, Tanzania, Timor

Oriental, Túnez, Uzbekistán, Vanuatu, Vietnam, Cisjordania y Gaza, Zambia, Zimbabue.

- **Renta Baja:** Afganistán, Burkina Faso, Burundi, República Centroafricana, Chad, República Democrática del Congo, Eritrea, Gambia, Guinea-Bisáu, Corea del Norte, Liberia, Madagascar, Malawi, Malí, Mozambique, Níger, Ruanda, Sierra Leona, Somalia, Sudán del Sur, Sudán, Siria, Togo, Uganda, Yemen.

8.2. ANEXO II

Dentro de cada grupo geográfico, se incluyen los siguientes países:

- **África:** África, Burkina Faso, África Central, República Democrática del Congo, Guinea Ecuatorial, Gabón, Guinea-Bisáu, Madagascar, Mauricio, Nigeria, Seychelles, Sudán del Sur, Togo, Zimbabue, Angola, Burundi, República Centroafricana, República del Congo, Eritrea, Gambia, Kenia, Malawi, Mozambique, Ruanda, Sierra Leona, África Austral, Uganda, Benín, Cabo Verde, Chad, Costa de Marfil, Esuatini, Ghana, Lesoto, Malí, Namibia, Santo Tomé y Príncipe, Somalia, Sudán, África Occidental, Botsuana, Camerún, Comoras, África Oriental, Etiopía, Guinea, Liberia, Mauritania, Níger, Senegal, Sudáfrica, Tanzania, Zambia.
- **Este asiático y Pacífico:** Australia, Fiyi, Corea del Sur, Estados Federados de Micronesia, Papúa Nueva Guinea e Islas del Pacífico, Filipinas, Taiwán (China), Tuvalu, Camboya, Indonesia, República Democrática Popular Lao, Mongolia, Islas del Pacífico, Samoa, Tailandia, Vanuatu, China, Japón, Malasia, Myanmar, Palaos, Singapur, Timor Oriental, Vietnam, Asia Oriental y el Pacífico, Kiribati, Islas Marshall, Nueva Zelanda, Papúa Nueva Guinea, Islas Salomón, Tonga
- **Europa y Asia Central:** Albania, Armenia, Austria, Azerbaiyán, Bielorrusia, Bélgica, Bosnia y Herzegovina, Bulgaria, Asia Central, Croacia, Chipre, República Checa, Dinamarca, Estonia, Europa y Asia Central, Finlandia, Francia, Georgia, Grecia, Hungría, Islandia, Irlanda, Italia, Kazajistán, Kosovo, República Kirguisa, Letonia, Lituania, Luxemburgo, Moldavia, Montenegro, Países Bajos, Macedonia del Norte, Noruega, Polonia, Portugal, Rumania, Federación Rusa, Serbia, República Eslovaca, Eslovenia, Europa Sudoriental y Balcanes, España, Tayikistán, Turquía, Turkmenistán, Ucrania, Uzbekistán

- **Latinoamérica y el Caribe:** Antigua y Barbuda, Argentina, Bahamas, Barbados, Belice, Bolivia, Brasil, Caribe, América Central, Chile, Colombia, Costa Rica, Dominica, República Dominicana, Ecuador, El Salvador, Granada, Guatemala, Guyana, Haití, Honduras, Jamaica, América Latina, Mercosur, México, Nicaragua, Panamá, Paraguay, Perú, San Cristóbal y Nieves, Santa Lucía, San Vicente y las Granadinas, Surinam, Trinidad y Tobago, Uruguay, Venezuela.
- **Oriente Medio y Norte de África:** Argelia, Yibuti, Egipto, República Árabe de, Irán, República Islámica de, Irak, Israel, Jordania, Líbano, Malta, Oriente Medio y Norte de África, Marruecos, Omán, República Árabe Siria, Túnez, Cisjordania y Gaza, Yemen.
- **Sur de Asia:** Afganistán, Bangladés, Bután, India, Maldivas, Nepal, Pakistán, Asia Meridional, Sri Lanka.