



**Universidad**  
Zaragoza

# Trabajo Fin de Grado Administración y Dirección de Empresas

Análisis Empírico de la ley de Zipf en las empresas  
españolas

Empirical Analysis of Zipf's law in Spanish firms

Autora

Marina Pilar Bosque Bori

Director

Marcos Sanso Navarro

Facultad de Economía y Empresa  
2023

**Resumen:**

La ley de Zipf es una ley empírica utilizada para describir la distribución del tamaño de una población. Para el caso del tamaño empresarial, el cumplimiento de esta ley implica que la empresa más grande duplica el tamaño de la segunda, triplica el de la tercera, cuatriplica el de la cuarta y así sucesivamente. Este Trabajo de Fin de Grado tiene como objetivo comprobar si se cumple esta ley para el caso del tamaño empresarial de España en 2019. Observando a las 139.557 empresas españolas que tienen más de 8 empleados primero, se rechaza su validez. Al segmentar por comunidades autónomas esta muestra, también se rechaza esta ley en todas ellas. Después, se observa a las 20.743 empresas más grandes del país segmentando por sectores de actividad y, aunque en la mayoría se rechaza, en algunos sí que se acepta. Por último, se observa que cuanto más grande es la muestra, el parámetro de la distribución es mayor.

*Palabras clave: tamaño empresarial, análisis empírico, Ley de Zipf, España.*

**Abstract:**

The Zipf's Law is an empirical law used to describe the distribution of a population's size. To the firm-size case, the fulfilment of this law implies that the larger firm doubles the size of the second, triples the size of the third, quadruples the size of the fourth and so on. This Degree Final Dissertation aims to review if this Law describes the firm-size of Spain in 2019. Observing the 139.557 firms that have more than 8 employees the law is rejected. When segmenting this sample by Autonomous Regions, this law is rejected in all of them, too. Later, when the 20.743 largest firms are segmented by activity sectors, it is observed that although in most sectors it is rejected, in some of them is accepted. Lastly, it is observed that the larger the sample, the larger is the parameter of the distribution.

*Keywords: firm-size, empirical analysis, Zipf's Law, Spain.*

## INDICE

1. Introducción.....	4
2. Marco conceptual .....	7
2.1. <i>Leyes Potenciales y ley de Zipf</i> .....	8
2.2. <i>Algunos estudios realizados sobre la ley de Zipf</i> .....	10
3. Análisis empírico.....	12
3.1. <i>Tejido empresarial de España</i> .....	12
3.2. <i>Objetivo del análisis</i> .....	15
3.3. <i>Muestra seleccionada</i> .....	16
4. Estudio empírico en España .....	23
4.1. <i>Estudio por comunidades autónomas</i> .....	24
4.2. <i>Estudio por sectores de actividad</i> .....	26
4.3. <i>Diferencias de los resultados según el tamaño de la muestra</i> .....	30
5. Conclusiones.....	30
6. Bibliografía.....	33

## 1. Introducción

El estudio del crecimiento de las empresas es un tema muy relevante en la economía. De hecho, existe una infinidad de modelos que intentan describir estadísticamente la distribución del tamaño empresarial, tratando de predecir la dinámica que llevarán las empresas en cuanto a su tamaño. Algunos de los modelos destacados en la literatura que estudian este fenómeno son los de (Gibrat, 1931), (Levy & Solomon, 1996), (Sutton, 1997) y (Gabaix, Zipf's Law for Cities: An Explanation, 1999). El tamaño de una empresa es un factor condicionante a la hora de obtener ayudas públicas, el acceso a determinadas fuentes de financiación, su relevancia en el mercado y hasta en la sociedad, también influye en los tipos de gravámenes aplicables en determinados impuestos. Además, la condiciona a la hora de compararla con el resto de las compañías de similar dimensión al relativizar sus datos y posicionarlas en el agregado en el que se encuadre. Es decir, el tamaño de una empresa parece que condiciona enormemente la situación económico-financiera de esta. Aunque, este es un tema controvertido y contradictorio muchas veces, porque tal y como se explica en el trabajo de Gómez-Miranda y Rodríguez-Ariza (2004), no está muy clara la relación absoluta entre el tamaño de las empresas y sus características económico-financieras.

La ley de Zipf es una ley empírica que fue postulada por el lingüista y filólogo estadounidense George Kingsley Zipf en la década de 1940, en su obra *Human behavior and the principle of least effort*. (Zipf, 1949). El lingüista de la Universidad de Harvard utilizaba herramientas estadísticas para analizar la frecuencia de uso de las palabras de varias lenguas. Con estos análisis estadísticos observó que la frecuencia de aparición de las palabras en una lengua sigue una distribución  $P_n \sim 1/n^a$ , donde  $P_n$  es la palabra  $n$ -ésima más frecuente y  $a$  es un número real positivo. Esto quiere decir que la palabra más usada se utiliza el doble de veces que la segunda, el triple que la tercera, cuatro veces más que la cuarta y así sucesivamente. Este fenómeno resulta bastante peculiar, siendo que las lenguas son tan subjetivas y complejas, es extraño que la frecuencia de uso de estas se distribuya de una manera tan pronosticable. El hecho de que se cumpla esta ley supone que para aprender una lengua se necesite saber muy pocas palabras para poder utilizarlo, como ya se ha demostrado bastantes veces con la ley de Pareto. Con el 20% de las palabras más utilizadas, se puede entender aproximadamente el 80% de un idioma. Estas leyes se cumplen para muchos casos. La extrañeza del caso ha hecho que se pruebe en muchos más ámbitos como lo son el tamaño de las ciudades de un país

(Gabaix, 1999) o el tamaño de las empresas de un país (Wit, 2005), que es el caso de estudio a realizar en este trabajo.

Varios estudios han concluido que la ley de Zipf proporciona una buena aproximación a la distribución del tamaño empresarial si se utilizan los estimadores adecuados (Bottazzi, Pirino, & Tamagni, 2015). Así que, para estudiar el crecimiento de las empresas, parece que resulta ser una buena idea utilizarla. El presente trabajo pretende demostrar el cumplimiento o no de la ley de Zipf para las empresas más grandes de España, utilizando los datos que facilita la plataforma SABI sobre el tamaño empresarial en el año 2019. Hay que destacar que la razón por la que se ha escogido este año ha sido con la intención de eliminar los efectos de la pandemia por COVID-19. A continuación, se han extraído los datos de las 10 empresas más grandes de España de esta base de datos, con la intención de proponer un simple ejemplo (Tabla 1-1) para ilustrar esta ley de una manera intuitiva. Como se puede ver, no se cumple para las 10 empresas más grandes de España. Excepto por Mercadona, el número de empleados del resto de empresas es bastante similar.

*Tabla 1-1 Ejemplo simple e intuitivo de la ley de Zipf para las 10 empresas más grandes de España. Fuente: Elaboración propia a partir de datos de SABI.*

Nombre	Comunidad autónoma	Número empleados 2019	% Respecto a la más grande	% Según la ley de Zipf
<b>Mercadona S.A.</b>	Comunidad Valenciana	95.142	<b>100%</b>	<b>100%</b>
<b>El Corte Inglés S.A.</b>	Madrid	58.385	<b>61%</b>	<b>50%</b>
<b>Sociedad Estatal Correos y Telégrafos S.A. S.M.E.</b>	Madrid	53.041	<b>56%</b>	<b>33%</b>
<b>Amrest Holdings S.E.</b>	Madrid	48.273	<b>51%</b>	<b>25%</b>
<b>Clece S.A.</b>	Madrid	47.562	<b>50%</b>	<b>20%</b>
<b>Ranstad Empleo</b>				
<b>Empresa de Trabajo Temporal S.A.</b>	Madrid	47.107	<b>50%</b>	<b>17%</b>
<b>Centros Comerciales Carrefour S.A.</b>	Madrid	33.918	<b>36%</b>	<b>14%</b>
<b>Eulen S.A.</b>	País Vasco	28.599	<b>30%</b>	<b>13%</b>
<b>ISS Facility Services S.A.</b>	Madrid	28.344	<b>30%</b>	<b>11%</b>
<b>Caixabank S.A.</b>	Comunidad Valenciana	27.404	<b>29%</b>	<b>10%</b>

Según el Directorio Central de Empresas (DIRCE), el número de empresas activas en España el día 1 de enero de 2020 era de 3.404.428. En el Gráfico 1-1 se observa el número de empresas que hay en cada comunidad autónoma, siendo las comunidades en las que más empresas hay, Cataluña, Comunidad Valenciana, Madrid y Andalucía. En cambio, según el Gráfico 1-2 se observa que la densidad de empresas sigue siendo mayor en Cataluña y Madrid, pero la Comunidad Valenciana y Andalucía desaparecen de la cumbre. Es destacable ver que Andalucía tiene muchas empresas, pero una densidad muy baja. El caso contrario es Extremadura, donde existen pocas empresas en términos absolutos, pero por cada 10.000 habitantes, es de las más pobladas. En el apartado 3.1 se profundizará en el tejido empresarial español.

*Gráfico 1-1 Explotación Estadística del Directorio Central de Empresas, Comunidades y Ciudades Autónomas (DIRCE), Total CNAE, Total, 2020*

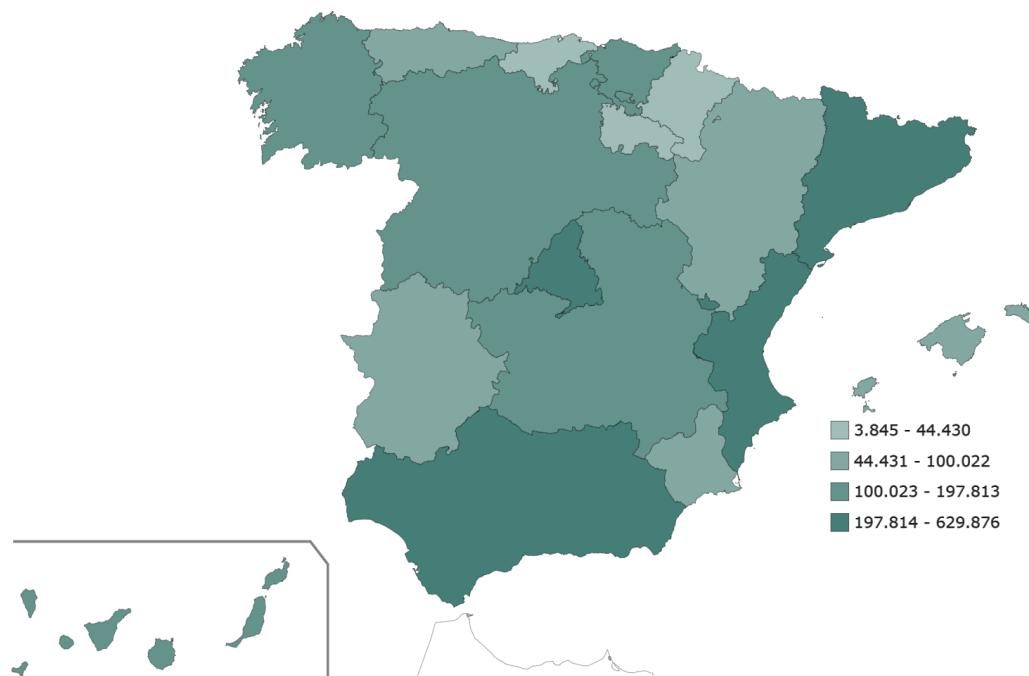
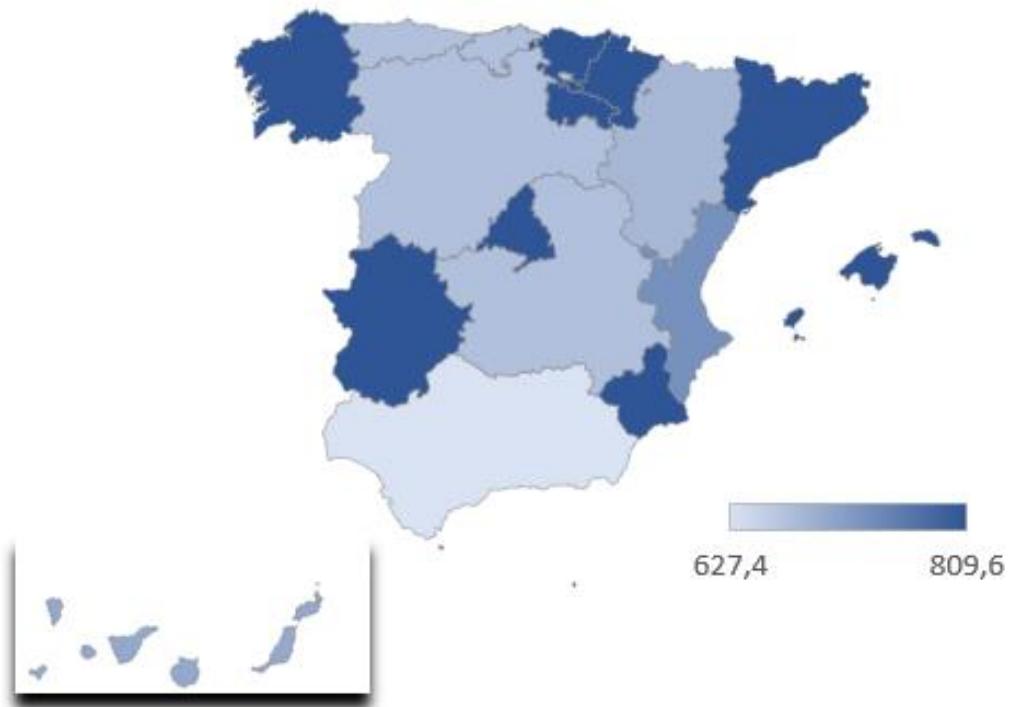


Gráfico 1-2 Número de empresas por cada 10.000 habitantes, por Autonomías. Fuente: Elaboración propia a partir de datos del DIRCE a 1 de enero de 2020.



En la última parte del trabajo antes de las conclusiones, se introducirán las herramientas econométricas utilizadas y los resultados obtenidos del análisis realizado. Sin embargo, se va a adelantar que en la base de datos SABI sólo hay disponibles datos para 1.884.137 empresas de las cuales sólo 893.620 empresas están activas. Después del apartado anterior se puede ver que hay mucha diferencia con el número real de empresas en España. Esto hace que la muestra utilizada sea considerada pequeña, sobre todo porque al tener limitaciones de extracción de datos, no es posible utilizar tampoco los datos de esas 893.620 empresas. Se ha tratado de conseguir el mayor número de datos sin sobrepasar estas limitaciones utilizando el número de empleados como variable a utilizar. Pero se tendrá que aplicar igualmente, una corrección de medio punto a la fórmula a aplicar (Gabaix & Ibragimov, 2007).

## 2. Marco conceptual

El tamaño de las empresas que se crean, que existen actualmente y las empresas que han existido ofrece un cúmulo de ejemplos que se pueden estudiar empíricamente para ofrecer diferentes modelos con el fin de explicar el ciclo de vida que han tenido las

empresas que ya no existen y así poder predecir la entrada, crecimiento y salida de las empresas existentes. Además, otra de sus utilidades más importantes, es que permite explicar la diferencia entre empresas y entre sectores. Hay disponibles una infinidad de modelos y de resultados de distribuciones de tamaño empresarial, por lo que hay que tenerlos en cuenta a todos ellos para poder utilizar el más conveniente según la situación respecto al momento de ciclo de vida de la empresa y el sector al que pertenezca. Como ya se ha mencionado, estos resultados ofrecen una información muy valiosa sobre los futuros cambios que experimentará la empresa. Además de ofrecer una explicación al tamaño de las empresas actuales (Wit, 2005).

En la literatura, algunos de los resultados que más se han utilizado son los de (Gibrat, 1931), (Levy & Solomon, 1996), (Sutton, 1997) y (Gabaix, 1999). Al compararlos, se ha observado que obtienen similares conclusiones, pero también se contradicen. Esto quiere decir que la literatura actual tiene fallos y que aún quedan vacíos que llenar en esta materia tan compleja. En los últimos años se han realizado varios estudios empíricos que llegan a la conclusión de que, la ley de Zipf, se aproxima bastante bien en la cola alta de la distribución del tamaño de las empresas (Bottazzi, Pirino, & Tamagni, 2015). Es decir, para las empresas más grandes, la aproximación resulta decentemente correcta. Estos estudios han seguido la influencia del trabajo de (Axtell, 2001).

## 2.1. Leyes Potenciales y ley de Zipf

Hace más de un siglo, el filósofo e ingeniero italiano Vilfredo Federico Pareto, (1848-1923), descubrió las leyes potenciales. En su libro *Cours d'economie politique* publicado en 1896 observó que, generalmente, el 20% de los esfuerzos, produce el 80% de los resultados. Para comprobar que se cumplía su teoría, realizó un estudio empírico de la propiedad de la tierra en Italia, llegando a la conclusión de que el 20% de los propietarios poseían el 80% de las tierras, mientras que el resto, el 80% de los propietarios, poseían solamente, el 20% de las tierras italianas (Pareto, 1896). Desde su formulación por Pareto, muchos economistas y científicos han comprobado que esta relación 80/20 se cumple en una variedad de ámbitos de la sociedad. Lo que supone el cumplimiento de esta ley es que sólo es necesario enfocarse en una pequeña parte de una población para conseguir la mayoría de los resultados. Como fue Pareto quien descubrió esta ley, se la llama también *distribución de Pareto* en su honor.

Las leyes potenciales aparecen constantemente en física, biología, geología y en astronomía, así como en economía y finanzas, informática, ingeniería, demografía y ciencias sociales. Parece que están presentes en todos los ámbitos de nuestra vida. La distribución del tamaño de las ciudades (Gabaix y Ibragimov 2007; Gabaix 1999; Lanaspa et al. 2004), de los terremotos (Gutenberg & Richter, 1944), de los cráteres lunares (Neukum & Ivanov, 1994), hasta de las guerras (Roberts & Turcotte, 1998) y de la fortuna de las personas (Pareto, 1896) parecen seguir una ley potencial. El origen del comportamiento de estas leyes ha sido tema de debate en la comunidad científica durante más de un siglo. Por lo que existe un gran número de evidencias empíricas para demostrar su existencia, así como de teorías que tratan de explicarlas (Newman, 2005).

Una ley potencial, o distribución de Pareto, en una relación funcional entre dos cantidades, es decir, un cambio relativo en una de ellas resulta en un cambio relativo en la otra proporcional a la potencia de cambio. Siendo que, independientemente del tamaño inicial de las cantidades, una cantidad varía como un potencial de la otra:  $y = ax^b$ , siendo  $a$  y  $b$  números reales e  $y$  y  $x$ , cantidades sobre las que se mide la relación;  $a$  es la constante de proporcionalidad y  $b$  el exponente de la potencia. Si consideramos el tamaño empresarial de España como  $T$ , su distribución se ajustará a una ley potencial si:

$$R = aT^b \quad (1)$$

donde  $R$  es el rango o posición dentro del conjunto de empresas,  $T$  es su tamaño y  $a$  y  $b$  son parámetros que hay que estimar. El parámetro  $b$  es el que expresa las diferencias que existen entre las empresas. Cuanto más grande sea  $b$ , mayor diferencia existe entre los tamaños de estas. Cuando se obtiene que  $b$  es igual a la unidad, se obtiene la regla rango-tamaño o ley de Zipf. Esta ley es un caso particular de la distribución de Pareto o ley potencial y tiene una implicación bastante destacable. Si se cumple, la empresa más grande será el doble de grande que la segunda, el triple que la tercera, el cuádruple que la cuarta, etc.

La ecuación anterior (1), se puede expresar en términos lineales utilizando una transformación logarítmica que va a permitir la estimación de los dos parámetros de una manera bastante sencilla. Así, tomando logaritmos a ambos lados de la igualdad, se puede expresar la ley potencial de la siguiente manera:

$$\log R = \log a - b \log T + \varepsilon_{(R)} \quad (2)$$

Esta transformación del modelo permite la estimación de los parámetros asociados a la ecuación por Mínimos Cuadrados Ordinarios (a partir de ahora, OLS). Este método de estimación es ampliamente aceptado y utilizado debido a su sencillez y robustez, sin embargo, presenta un sesgo para muestras pequeñas. Esto quiere decir que para muestras muy grandes funciona bien, pero para muestras pequeñas puede presentar una estimación de los parámetros que no sean óptimas. Para solucionar este problema, basta con añadir una corrección de medio punto al rango (3). (Gabaix & Ibragimov, 2007). Este punto será ampliado más adelante, a la hora de realizar el análisis.

$$\log(R - 0,5) = \log a - b \log T + \varepsilon_{(y)} \quad (3)$$

## 2.2. Algunos estudios realizados sobre la ley de Zipf

Esta interesante ley es aplicable a un vasto número de sucesos. Por ello, existen multitud de estudios de la ley de Zipf en la literatura mundial. Se ha aplicado a multitud de casos, desde su original, la frecuencia de aparición de las palabras de una lengua en libros; pasando por el caso más reiterado, el tamaño de las ciudades según el número de habitantes, hasta el tamaño de las empresas. A continuación, se va a mostrar algunos ejemplos de estudios realizados a lo largo de los años que aceptan o rechazan esta ley.

El primer estudio que se va a presentar es sobre el caso original, la frecuencia de las palabras en la lengua inglesa. En 2016, investigadores del Centre de Recerca Matemática (CRM) analizaron una colección de textos en inglés del proyecto Gutenberg, que es una base de datos pública y gratuita con más de 30.000 obras escritas en inglés. Previamente no se había utilizado más de una docena de textos para comprobar la ley de Zipf. Según este análisis, al ignorar las palabras que no aparecen más que una o dos veces en cada libro, el 55% de los textos se ajustan perfectamente a la ley de Zipf. Al tener en cuenta todas las palabras se obtuvo que el 40% la cumplían. Lo que más les sorprendió a los investigadores es que la frecuencia de aparición de las palabras esté determinada por una fórmula con un solo parámetro libre. (Moreno-Sánchez, Font-Clos, & Corral, 2016). La aplicación de la ley de Zipf para la población de las ciudades de un país es el primer ejemplo económico de esta ley (Zipf, 1949). Probablemente es el caso más estudiado debido a la relativa facilidad existente para

conseguir los datos del tamaño de las ciudades. La variable *tamaño* en estos estudios es el número de habitantes de cada territorio. Gabaix y Ibragimov (2007), estudian el caso de las 135 áreas metropolitanas alistadas en el Abstracto Estadístico de los Estados Unidos en el año 1991, al igual que hizo anteriormente Gabaix (1999). Este incluye todas las aglomeraciones con tamaño superior a 250.000 habitantes. Los resultados que obtiene son que los estimadores del índice de la cola no son diferentes a 1 con un 10% de error, así que la ley de Zipf para ciudades se confirma para este caso. De la misma manera, (Lanaspa, Perdiguero, & Sanz, 2004) hacen un estudio sobre la población de los municipios más grandes de España, en relación con esta ley, concluyen que para España no se cumple, por lo que para el caso del tamaño de las ciudades españolas no existe evidencia de que se cumpla.

En cuanto a estudios sobre el tamaño empresarial, se presentan los siguientes ejemplos y sus resultados. Axtell en el año 2001 publicó un trabajo sobre la distribución de la ley de Zipf para el tamaño de las empresas de Estados Unidos. Utilizó los datos de la población total de compañías que pagan impuestos en el país durante varios años y para varias definiciones de tamaño empresarial, entre ellas el número de empleados, que es el criterio que se va a utilizar en este trabajo. Observó que las empresas experimentaron unos cambios significativos en sus ingresos y número de trabajadores desde 1992 a 1997. Además, empresas individuales migraron por encima y por debajo de la distribución de Zipf, pero considera que las *fuerzas económicas* mantuvieron las desviaciones en un nivel muy bajo, pues la distribución del tamaño empresarial general apenas se vio modificada. Para este caso, Axtell concluye que la distribución de Zipf describe la distribución de las empresas estadounidenses y que podría cumplirse para otros países también. En este trabajo veremos si se cumple para el caso de España.

En un análisis pre y post-fusiones del sector compuesto por las cajas de ahorros españolas Golpe, Iglesias, & Martín (2013) utilizan la ley de Zipf y la ley de Gibrat para avalar las medidas propuestas por el Banco de España acerca de la sobredimensión de las Cajas de Ahorro y la necesidad de reducir su capacidad. Este crecimiento de capacidad fue debido a que durante la fase expansiva hasta 2007, se generó desequilibrios en el sistema financiero español, concretamente, en las Cajas de Ahorros. El gobierno español puso en marcha un plan de reestructuración que desencadenó un proceso extremo de fusiones entre las cajas, pasando de ser 45 entidades en 2009 a 18

en 2010. Con este análisis demuestran que es cierto que se produjo un sobredimensionamiento comercial al cumplirse la ley de Zipf. En el caso de los depósitos no se cumplía, queriendo decir que este incumplimiento suponía que las cajas de ahorro sobredimensionaron su negocio y su capacidad sin tener en cuenta la captación de depósitos. Lo que supone que esta ley se puede utilizar también para comprobar el funcionamiento de medidas financieras.

### **3. Análisis empírico**

En el transcurso de los apartados anteriores se ha contextualizado el análisis que se va a realizar a continuación mencionando una parte de la literatura existente por el momento sobre la ley de Zipf y se ha introducido la metodología con la que se va a realizar este análisis empírico. En este apartado se va a reunir y ampliar, en su caso, toda esa información.

#### **3.1. Tejido empresarial de España**

Antes de realizar el análisis empírico, conviene describir el tejido empresarial de este país con el fin de entender los resultados que se van a obtener en este estudio. Según el Directorio Central de Empresas (DIRCE), a 1 de enero de 2020 existían en España 3.404.428 empresas, de las cuales sólo el 0,1% de estas eran grandes empresas. Esto quiere decir que casi todas las empresas españolas (99,9%) son pequeñas y medianas empresas, PYME. Todas estas empresas tienen menos de 250 empleados, el 56% no tiene ningún asalariado y más del 95% tienen menos de 10.

Si se observa la Tabla 3-1, podemos ver que sólo hay 4.826 empresas grandes en todo el país. Hay 21.543 con entre 50 y 249 asalariados (empresas medianas), 127.399 con entre 10 y 49 asalariados (pequeñas empresas) y 1.338.650 con entre 1 y 9 asalariados (microempresas). Esto hace un total de 1.492.418 empresas con empleados. Esto, junto a que la forma jurídica predominante es la de persona física, (55,9%), nos muestra que la mayoría de las empresas no son empresas realmente, sino trabajadores autónomos. Además, la variación interanual en esta forma jurídica es positiva, ha aumentado un 2,8% en 2019.

En cuanto a la distribución por sectores de actividad, la mayoría de las empresas se sitúan en el sector servicios, el 81,9%. La industria supone un 5,7% de las empresas y el sector de la construcción alberga el 12,3%. Respecto a las variaciones observadas, el número de empresas ha aumentado un 1,2% en el año 2019. Además, las empresas que han crecido en mayor medida son las grandes (5,1%) y las medianas (4,7%). Las pequeñas empresas también aumentaron, en un 2,3% y las microempresas sin asalariados también (1,6%). En el caso de las variaciones en los sectores, únicamente creció el de *Resto de servicios* un 3,1%. En los demás se destruyeron empresas, siendo en el comercio un 2% y en la industria un 1,7%. Hay que destacar que en este año las sociedades con forma jurídica de sociedad anónima disminuyeron un 6,1%.

Tabla 3-1 Tejido empresarial de España en 2019. Fuente: Elaboración propia a partir de datos del DIRCE a 1 de enero de 2020.

	Empresas	Empresas por cada 10.000 habitantes	%	Variación Interanual
<b>EMPRESAS POR TAMAÑO</b>				
<b>Total</b>	<b>3.404.428</b>	717,5	100,0%	1,2%
<b>PYME (0-249)</b>	<b>3.399.602</b>	716,4	99,8%	1,2%
<b>PYME sin asalariados</b>	<b>1.912.010</b>	402,9	56,2%	1,6%
<b>PYME con asalariados</b>	<b>1.487.592</b>	313,5	43,7%	0,8%
<b>Microempresas (1-9 asalariados)</b>	<b>1.338.650</b>	282,1	39,3%	0,6%
<b>Pequeñas (20-49 asalariados)</b>	<b>127.399</b>	26,8	3,7%	2,3%
<b>Medianas (50-249 asalariados)</b>	<b>21.543</b>	4,5	0,6%	4,7%
<b>Grandes (250 o más asalariados)</b>	<b>4.826</b>	1,0	0,1%	5,1%
<b>EMPRESAS POR SECTORES</b>				
<b>Total</b>	<b>3.404.428</b>	717,5	100,0%	1,2%
<b>Industria</b>	<b>195.615</b>	41,2	5,7%	-1,7%
<b>Construcción</b>	<b>420.118</b>	88,5	12,3%	-0,5%
<b>Comercio</b>	<b>726.573</b>	153,1	21,3%	-2,0%
<b>Resto de servicios</b>	<b>2.062.122</b>	434,6	60,6%	3,1%
<b>EMPRESAS POR CONDICIÓN JURÍDICA</b>				
<b>Total</b>	<b>3.400.257</b>	716,6	100,0%	1,1%
<b>Persona física</b>	<b>1.899.810</b>	400,4	55,9%	2,8%
<b>Sociedad Anónima</b>	<b>63.456</b>	13,4	1,9%	-6,1%
<b>Sociedad Limitada</b>	<b>1.143.625</b>	241,0	33,6%	-0,9%
<b>Comunidad de bienes</b>	<b>116.504</b>	24,6	3,4%	0,1%
<b>Sociedad cooperativa</b>	<b>21.111</b>	4,4	0,6%	1,3%
<b>Otras formas jurídicas</b>	<b>155.751</b>	32,8	4,6%	-0,3%

Siguiendo con la Tabla 3-2, casi el 62% de las PYME de España están situadas en sólo cuatro comunidades autónomas. En la cabeza, Cataluña, con 628.819 empresas que suponen el 18,50% de todas las PYME españolas. En segundo lugar, se sitúa la Comunidad de Madrid con 550.538, el 16,19%. La tercera comunidad es Andalucía, que tiene 530.599 empresas (15,61%) y, en cuarto lugar, se posiciona la Comunidad Valenciana con el 10,89% del tejido de PYME español, lo que son 370.222 empresas. La suma de estas cuatro comunidades es de 2.080.178 empresas.

Tabla 3-2 PYME en España por Comunidades Autónomas. Fuente: Elaboración propia a partir de datos del DIRCE a 1 de enero de 2020.

Ámbito territorial	PYME 0-249	PYME 0-249
<b>España</b>	<b>3.399.602</b>	<b>100%</b>
<b>Andalucía</b>	<b>530.599</b>	<b>15,61%</b>
<b>Aragón</b>	<b>90.569</b>	<b>2,66%</b>
<b>Asturias</b>	<b>68.367</b>	<b>2,01%</b>
<b>Baleares</b>	<b>99.920</b>	<b>2,94%</b>
<b>Canarias</b>	<b>152.589</b>	<b>4,49%</b>
<b>Cantabria</b>	<b>38.836</b>	<b>1,14%</b>
<b>Castilla y León</b>	<b>160.065</b>	<b>4,71%</b>
<b>Castilla-La Mancha</b>	<b>128.632</b>	<b>3,78%</b>
<b>Cataluña</b>	<b>628.819</b>	<b>18,50%</b>
<b>Comunidad Valenciana</b>	<b>370.222</b>	<b>10,89%</b>
<b>Extremadura</b>	<b>67.303</b>	<b>1,98%</b>
<b>Galicia</b>	<b>197.644</b>	<b>5,81%</b>
<b>Madrid</b>	<b>550.538</b>	<b>16,19%</b>
<b>Murcia</b>	<b>96.644</b>	<b>2,84%</b>
<b>Navarra</b>	<b>44.343</b>	<b>1,30%</b>
<b>País Vasco</b>	<b>142.758</b>	<b>4,20%</b>
<b>La Rioja</b>	<b>22.679</b>	<b>0,67%</b>
<b>Ceuta y Melilla</b>	<b>9.075</b>	<b>0,27%</b>

### 3.2. Objetivo del análisis

El objetivo principal de este trabajo va a ser comprobar si se cumple, para el caso de España, la ley de Zipf, igual que se cumple para los Estados Unidos de América en el trabajo de Axtell (2001) o para las Cajas de Ahorros de España en el periodo de 2002 a 2009 en Golpe, Iglesias, Martín (2013). Para ello, se va a hacer la estimación para una muestra de 139.557 empresas. Después, se comprobará, de la misma manera, segmentando esta muestra en comunidades autónomas. Luego, se procederá una vez más, pero para una muestra más pequeña de 20.743 empresas segmentándola por sectores de actividad. Al hacer uso de dos muestras, se comparará también la diferencia de resultados para el caso de España para probar si existe una diferencia significativa al utilizar una muestra más pequeña (Segarra & Teruel, 2012).

### 3.3. Muestra seleccionada

La pandemia por COVID-19 ha supuesto una crisis a nivel mundial en todos los aspectos socioeconómicos existentes, por esa misma razón, el análisis objeto de este trabajo podría verse afectado, ya que muchas empresas quebraron y cerraron durante este suceso. Teniendo lo anterior en mente, se ha decidido utilizar los datos de 2019 para poder eliminar estos efectos del análisis. Además, se ha elegido la variable *número de empleados* para describir el tamaño de las empresas. Se ha decidido utilizar esta variable porque es la más sencilla de extraer teniendo en cuenta que se pretendía conseguir la muestra más grande posible y que la base de datos SABI no permite una extracción tan numerosa. A parte, es la variable más comprensible para la población, ya que la cantidad de trabajo que da una empresa es el valor más significativo para la sociedad. Sobre todo, en España, donde las tasas de paro son bastante elevadas. Para 2019, este dato era igual al 13,8% en contraste con la tasa de desempleo de la UE-28 que era del 6,4%, el valor más bajo registrado desde que se inició la serie mensual de desempleo de la UE-28, en enero del año 2000. Para entender mejor estos números, la tasa de paro más baja en España fue en el año 2007 con un 8,57% (EPA, INE).

Debido a las limitaciones mencionadas de la base de datos que se va a utilizar, SABI, ha sido necesario limitar la muestra de empresas. Por ello, se van a utilizar dos muestras. Primero, se va a hacer el análisis extrayendo los datos de todas las empresas con información disponible que tenían 9 o más de 9 empleados en 2019. Dejando la primera muestra en un total de 139.557 empresas. Con esta muestra, se realizará el análisis del cumplimiento de la ley de Zipf para toda España y también para cada comunidad y ciudad autónomas. Después, con las empresas de esa muestra que tienen 50 o más trabajadores, esta es la segunda muestra, se va a realizar el mismo análisis, pero para cada sector de actividad. Aprovechando la utilización de estas dos muestras se va a estudiar si cambia mucho el resultado dependiendo del tamaño de la muestra, para el caso de España en general. El tamaño de ambas muestras se considera pequeña porque en España existían más de 3 millones de empresas en 2019 (3.404.428). Esto supone que la muestra es el 4% del número total de empresas españolas en ese año. No obstante, el tamaño muestral utilizado no debería suponer ningún problema, ya que se va a tener en cuenta la corrección para muestras pequeñas explicada anteriormente para mitigar el error posible (Gabaix & Ibragimov, 2007).

En los siguientes párrafos, se va a observar la diferencia entre las muestras utilizadas y los datos reales. En el Gráfico 3-1 se muestra la distribución del número de empresas de la primera muestra, la de 139.557 empresas, extraída de SABI. En él se puede observar que la comunidad autónoma de Cataluña es la que más empresas alberga. Seguida de Madrid, Andalucía y de la Comunidad Valenciana. Esto coincide con lo mencionado en el apartado 3.1 sobre las PYME, también. Debajo, se puede observar el Gráfico 3-2 que representa la distribución de las empresas de España en el año 2019 por comunidades autónomas. Este Gráfico coincide en el orden de las cuatro comunidades con más empresas. Aunque la muestra desnivela la diferencia que existe entre el número de empresas que hay en Madrid respecto a las de Andalucía. Este suceso se debe a que el número de empresas grandes que hay en Andalucía es bastante más bajo que el de Madrid. El resto de los territorios presenta gráficamente una diferencia parecida en ambos Gráficos.

En la Tabla 3-2, se tienen los datos sobre las empresas activas por sectores y por comunidades y ciudades autónomas a 1 de enero de 2020. Esta Tabla es la equivalente a la Tabla 3-3, que representa los mismos datos, pero para la segunda muestra, compuesta por 20.743 empresas, que se va a utilizar en el análisis sectorial. Para comprobar si la muestra es representativa, se han calculado las diferencias de proporción entre la muestra utilizada y los datos que ofrece el Directorio Central de Empresas (DIRCE) según el CNAE, para cada uno de los sectores según las comunidades y ciudades autónomas. Estas diferencias se pueden observar en la Tabla 3-4. La mayoría de las diferencias observables son de menos del 1% por lo que no se consideran significativas. Aunque en términos totales, el sector *Resto de Servicios*, presenta una diferencia de un 12,9%. Significando que la muestra considera que este sector tiene una proporción menor que la realidad. En concreto, en los datos del DIRCE, este sector tiene una representación del 60,6% del total de empresas, mientras que en la muestra representa el 47,6%. En cambio, en el sector *Industria*, ocurre lo contrario, la muestra valora un 20,8% más de lo que lo hace la realidad. Según el DIRCE, este sector representa únicamente el 5,7% del tejido empresarial español, mientras que, en la muestra, el 26,6% de las empresas pertenecen a este sector. La explicación a este fenómeno es que normalmente las empresas del sector industrial ofrecen mucho más trabajo y son más grandes debido a su naturaleza que las empresas de servicios, donde la mayor parte de las veces, estas últimas empresas son profesionales autónomos. Como la variable

utilizada para describir el tamaño es el número de empleados, la mayoría de las empresas industriales está contenida en la muestra, mientras que las del sector servicios no. La diferencia de cantidad de empresas entre los dos sectores es debido a que, como todo país desarrollado, el sector servicios ocupa un lugar más importante que el industrial. Esta diferencia engloba también al sector comercial que, de la misma manera, se observa que su representación en la muestra es menor a la realidad. Para una ilustración gráfica de la distribución de la muestra y de la de los datos extraídos del DIRCE, con la intención de visualizar mejor la diferencia entre ambas, se facilita el Gráfico 3-3.

*Tabla 3-3 Empresas activas según sector económico, por Comunidades y Ciudades Autónomas a 1 de enero de 2020.*  
*Fuente: Elaboración propia a partir del DIRCE.*

	<b>TOTAL</b>	<b>Industria</b>	<b>Construcción</b>	<b>Comercio</b>	<b>Resto de Servicios</b>
<b>TOTAL</b>	<b>3.404.428</b>	<b>195.615</b>	<b>420.118</b>	<b>726.573</b>	<b>2.062.122</b>
<b>Andalucía</b>	<b>531.045</b>	28.395	57.938	133.421	311.122
<b>Aragón</b>	<b>90.682</b>	6.532	11.787	17.953	54.410
<b>Asturias</b>	<b>68.433</b>	3.495	8.133	14.010	42.795
<b>Baleares</b>	<b>100.022</b>	4.834	17.291	15.884	62.013
<b>Canarias</b>	<b>152.756</b>	5.677	16.951	32.570	97.558
<b>Cantabria</b>	<b>38.880</b>	2.086	5.537	7.722	23.535
<b>Castilla y León</b>	<b>160.199</b>	11.477	22.875	36.175	89.672
<b>Castilla-La Mancha</b>	<b>128.713</b>	11.174	19.696	31.785	66.058
<b>Cataluña</b>	<b>629.876</b>	35.339	75.555	121.585	397.397
<b>Comunidad Valenciana</b>	<b>370.645</b>	24.680	45.521	84.418	216.026
<b>Extremadura</b>	<b>67.336</b>	4.787	8.384	19.093	35.072
<b>Galicia</b>	<b>197.813</b>	12.100	28.030	45.249	112.434
<b>Madrid</b>	<b>552.027</b>	20.965	60.387	93.648	377.027
<b>Murcia</b>	<b>96.764</b>	7.107	12.537	24.641	52.479
<b>Navarra</b>	<b>44.430</b>	3.833	6.282	9.208	25.107
<b>País Vasco</b>	<b>143.028</b>	10.700	19.745	30.706	81.877
<b>La Rioja</b>	<b>22.700</b>	2.233	2.775	4.903	12.789
<b>Ceuta</b>	<b>3.845</b>	73	294	1.431	2.047
<b>Melilla</b>	<b>5.234</b>	128	400	2.171	2.535

Tabla 3-4 Distribución por sectores de la segunda muestra. Fuente: Elaboración propia a partir de datos extraídos de SABI.

	TOTAL	Industria	Construcción	Comercio	Resto de Servicios
TOTAL	20.743	5.509	1.460	3.895	9.879
<b>Andalucía</b>	<b>2.233</b>	577	216	469	971
<b>Aragón</b>	<b>604</b>	263	40	97	204
<b>Asturias</b>	<b>320</b>	91	40	43	146
<b>Baleares</b>	<b>596</b>	24	56	88	428
<b>Canarias</b>	<b>883</b>	75	68	207	533
<b>Cantabria</b>	<b>193</b>	73	18	26	76
<b>Castilla y León</b>	<b>634</b>	275	33	87	239
<b>Castilla-La Mancha</b>	<b>520</b>	218	63	74	165
<b>Cataluña</b>	<b>4.115</b>	1.190	187	901	1.837
<b>Comunidad Valenciana</b>	<b>2.090</b>	720	113	455	802
<b>Extremadura</b>	<b>251</b>	92	21	49	89
<b>Galicia</b>	<b>990</b>	355	97	141	397
<b>Madrid</b>	<b>4.969</b>	560	357	891	3.161
<b>Murcia</b>	<b>650</b>	236	37	183	194
<b>Navarra</b>	<b>372</b>	199	26	37	110
<b>País Vasco</b>	<b>1.158</b>	480	78	125	475
<b>La Rioja</b>	<b>140</b>	80	7	20	33
<b>Ceuta</b>	<b>16</b>	1	2	1	12
<b>Melilla</b>	<b>9</b>	0	1	1	7

Tabla 3-5 Diferencias de proporción para cada sector-autonomía entre la muestra y los datos del DIRCE.

	Diferencia TOTAL	Diferencia Industria	Diferencia Construcción	Diferencia Comercio	Diferencia Resto de Servicios
<b>TOTAL</b>	<b>0,0%</b>	<b>-20,8%</b>	<b>5,3%</b>	<b>2,6%</b>	<b>12,9%</b>
<b>Andalucía</b>	4,8%	-1,9%	0,7%	1,7%	4,5%
<b>Aragón</b>	-0,2%	-1,1%	0,2%	0,1%	0,6%
<b>Asturias</b>	0,5%	-0,3%	0,0%	0,2%	0,6%
<b>Baleares</b>	0,1%	0,0%	0,2%	0,0%	-0,2%
<b>Canarias</b>	0,2%	-0,2%	0,2%	0,0%	0,3%
<b>Cantabria</b>	0,2%	-0,3%	0,1%	0,1%	0,3%
<b>Castilla y León</b>	1,6%	-1,0%	0,5%	0,6%	1,5%
<b>Castilla-La Mancha</b>	1,3%	-0,7%	0,3%	0,6%	1,1%
<b>Cataluña</b>	-1,3%	-4,7%	1,3%	-0,8%	2,8%
<b>Comunidad Valenciana</b>	0,8%	-2,7%	0,8%	0,3%	2,5%
<b>Extremadura</b>	0,8%	-0,3%	0,1%	0,3%	0,6%
<b>Galicia</b>	1,0%	-1,4%	0,4%	0,6%	1,4%
<b>Madrid</b>	-7,7%	-2,1%	0,1%	-1,5%	-4,2%
<b>Murcia</b>	-0,3%	-0,9%	0,2%	-0,2%	0,6%
<b>Navarra</b>	-0,5%	-0,8%	0,1%	0,1%	0,2%
<b>País Vasco</b>	-1,4%	-2,0%	0,2%	0,3%	0,1%
<b>La Rioja</b>	0,0%	-0,3%	0,0%	0,0%	0,2%
<b>Ceuta</b>	0,0%	0,0%	0,0%	0,0%	0,0%
<b>Melilla</b>	0,1%	0,0%	0,0%	0,1%	0,0%

Gráfico 3-1 Distribución gráfica de la muestra de empresas utilizada en el estudio. Fuente: Elaboración propia a partir de datos obtenidos de SABI.

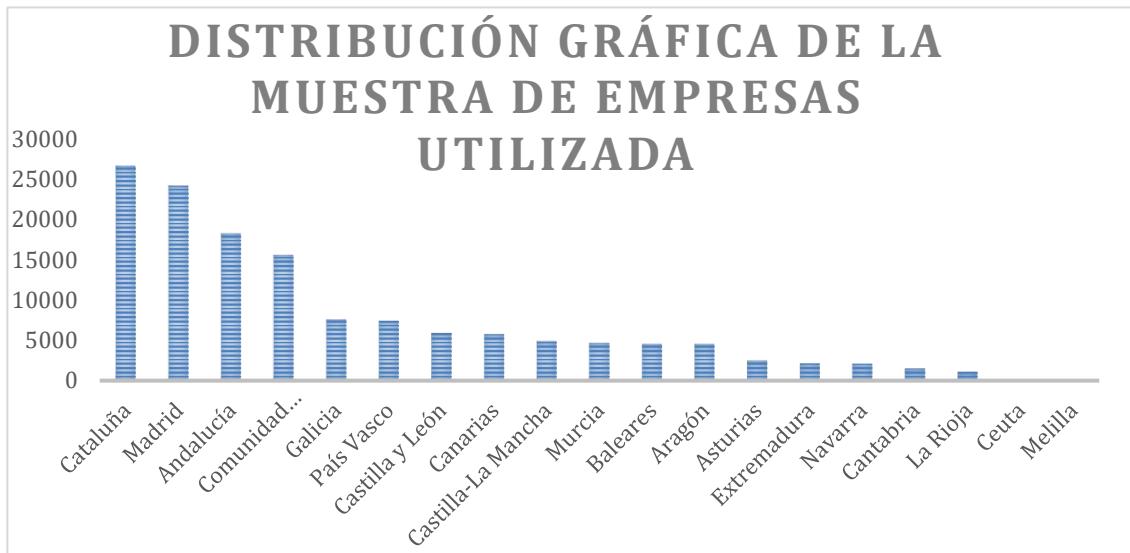


Gráfico 3-2 Distribución gráfica de las empresas españolas por comunidades autónomas. Fuente: Elaboración propia a partir de datos del DIRCE a 1 de enero de 2020.

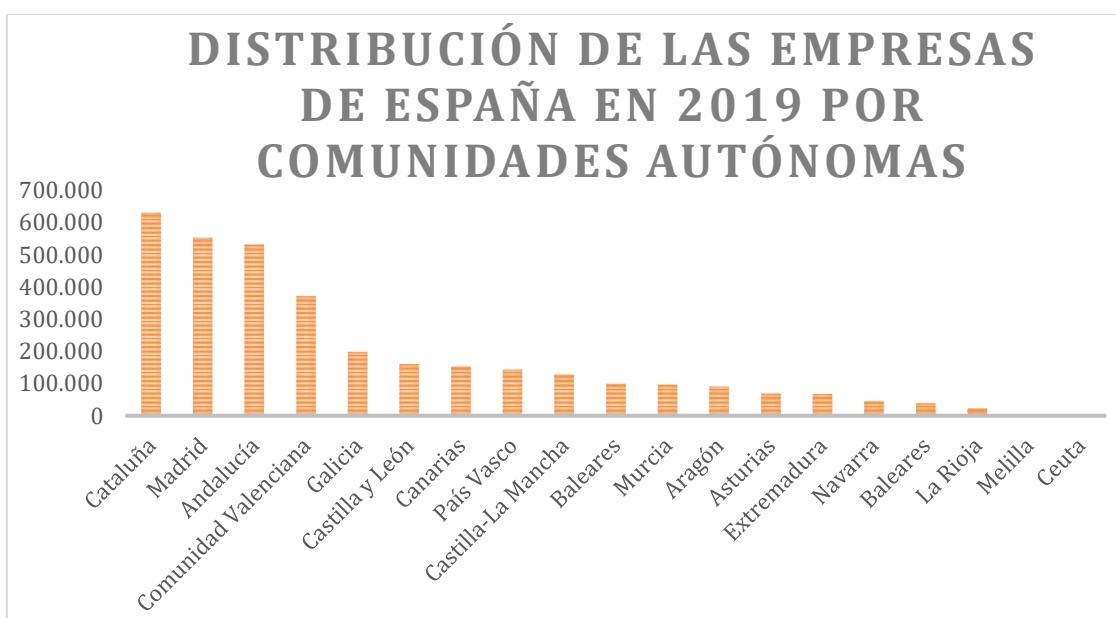
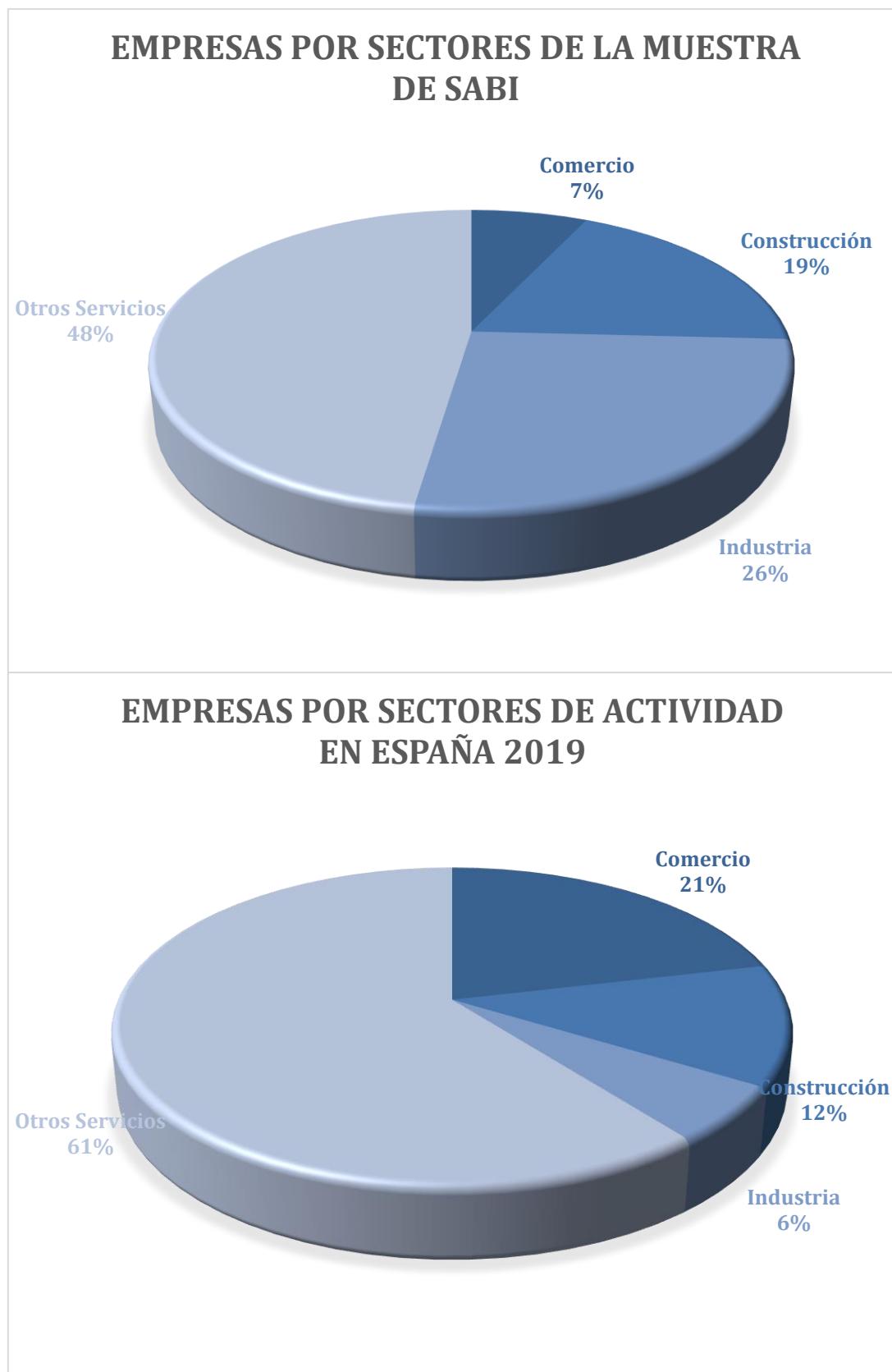


Gráfico 3-3 Proporción de empresas por sector de la muestra de 20.743 obtenida de SABI vs Proporción de empresas por sector en España en el 2019. Fuente: Elaboración propia a partir de datos de SABI y del DIRCE.



#### 4. Estudio empírico en España

Para estudiar la distribución de Zipf sobre el tamaño empresarial español, recordamos que se ha definido la variable *tamaño* como el número de empleados. Después, se han ordenado las empresas de la más grande a la más pequeña, siendo la que tiene rango 1 la que mayor número de empleados tiene y la más pequeña, con menor número de empleados, rango  $n$ . En el caso de la primera muestra,  $n = 139.557$ , denotando sus tamaños de esta manera:

$$T_{(1)} \geq T_{(2)} \geq \dots \geq T_{(R)} \geq \dots \geq T_{(n-1)} \geq T_{(n)} \quad (4)$$

de este modo  $T_{(1)}$  es la empresa con mayor número de empleados, con mayor tamaño y  $T_{(n)}$  es la de menor magnitud. Luego,  $R$  es el símbolo del rango o posición que ocupa cada empresa en la muestra.

Según lo expuesto en líneas anteriores, la ley de Zipf se formula a partir de la distribución de Pareto que se muestra en (1). Concretamente, la primera ley se cumple en el caso de que el *coeficiente de Pareto* tome la unidad como valor. Así, si se sustituye en (1) el postulado de la ley de Zipf ( $b = 1$ ), se obtiene lo siguiente:

$$RT_{(R)} = a \quad (5)$$

Dados  $R$  y  $T_{(R)}$ , el producto de ambas debe ser constante o aproximarse a un valor constante para todo  $R$ . Para el caso de la empresa más grande,  $R$  es igual a 1, por lo que se llega a que  $T_{(R)} = a$ , esto significa que  $a$  representa el tamaño de la empresa con mayor tamaño de España.

La modelización más utilizada en la literatura para contrastar si un conjunto de observaciones se distribuye o no de acuerdo con la ley de Zipf consiste en plantear una versión lineal de la ley potencial de Pareto:

$$\log R = a - b \log T_{(R)} + \varepsilon_{(R)} \quad (6)$$

Esto es para posteriormente estimar los parámetros mediante OLS y estudiar si es posible o no aceptar la hipótesis nula  $H_0: \beta_2 = -1$ . No obstante, en Gabaix e Ibragimov (2007), se evidencian los sesgos que presentan las estimaciones de los parámetros a

analizar mediante OLS. Los autores demuestran, de este modo, que el desplazamiento del rango en 0,5 en (6) resulta óptimo al minimizar los sesgos que presenta OLS en muestras pequeñas. De esta manera, al seguir la especificación expuesta en Gabaix e Ibragimov (2007), la regresión a estimar es:

$$\log (R - 0,5) = a - b \log T_{(R)} + \varepsilon_{(R)} \quad (7)$$

Es preciso destacar que el error estándar asintótico para el exponente de Pareto  $b$ , asciende a  $\sqrt{\frac{2}{n}} b_n$ . Por lo que, aunque se corrija la heterocedasticidad y la autocorrelación, el error estándar que proporcionan los programas de regresión es incorrecto (Goerlich & Mas, 2010).

Estos procedimientos son utilizados en múltiples estudios empíricos (véase, por ejemplo, Lanaspa et al., 2004; Rose, 2006; González-Val y Navarro, 2010; Goerlich y Mas, 2008).

Una vez estimados los parámetros mediante OLS para la primera muestra compuesta por 139.557 empresas, se ha obtenido la siguiente estimación:

$$\log (R - 0,5) = 14,2694 - 1,10380 \log T_n + \varepsilon_{(R)} \quad (8)$$

Al valor de  $b$  hay que restarle el error estándar para el índice de la cola dado por  $\sqrt{\frac{2}{n}} b_n$  (8) con lo que se obtiene que el error estándar es 0,00412. Por lo que los estimadores del índice de la cola son estadísticamente diferentes de 1. Esto quiere decir que no se acepta la hipótesis nula  $-b=1$  por lo que podemos afirmar que el tamaño empresarial de España no sigue una distribución de Zipf. Estos resultados implican que el tamaño de las empresas españolas no corresponde con su posición en el mercado. Aunque, el valor de  $b$  no está demasiado lejos de la unidad.

#### 4.1. Estudio por comunidades autónomas

De la misma manera, se ha comprobado si se cumple la ley de Zipf segmentando la muestra para estudiar su cumplimiento o no para las comunidades autónomas y las dos ciudades autónomas que componen España. Se ha estimado el parámetro  $b$  y el

parámetro  $a$  para el tamaño muestral correspondiente a cada territorio de la muestra de 139.557 empresas, igual que el de su error estándar. Esta información está reunida en la Tabla 4-1. Como se puede observar, casi todas las estimaciones se encuentran pasada la unidad, es decir, para el intervalo de confianza en todos los casos, el valor 1 para  $b$  no se encuentra contenido en ellos. Por lo que se rechaza la hipótesis nula y se puede afirmar que para el caso de las comunidades y ciudades autónomas no se cumple la ley de Zipf.

El no cumplimiento de la ley supone que el tamaño de las empresas no está relacionado con su posición en el mercado en cuanto a que la primera empresa no es el doble de grande que la segunda, el triple que la tercera, etc. El valor de  $b$ , cuanto más lejos de la unidad esté, expresa que la diferencia entre el tamaño de las empresas es mayor. En la mayor parte de los territorios analizados el valor del parámetro  $b$  se encuentra entre 0,8 y 1,2 o un poco más, por lo que no se considera que estén demasiado lejos (Gabaix, 2009). Esto quiere decir que no hay una diferencia tan grande entre el tamaño de las empresas de cada comunidad autónoma. No obstante, en el caso de las dos Ciudades Autónomas, Ceuta y Melilla, son las que más sobrepasan el valor de 1,2. Ceuta tiene un valor de 1,36, que no está tan alejado de, por ejemplo, Castilla-La Mancha (1,31). El de Melilla es de 1,5, resultando ser el territorio con el valor más alto con diferencia de todas las autonomías. Lo que quiere decir que en Melilla existe una diferencia de tamaños muy grande. Por otro lado, las que menos se alejan del valor 1, que simboliza el equilibrio, son Navarra (1,06) y el País Vasco (1,09). Por lo que, aunque no cumplan la ley de Zipf, están bastante cerca de hacerlo.

Tabla 4-1 Tabla resumen de los resultados obtenidos por comunidades autónomas.

Comunidad autónoma	Tamaño de la muestra	Valor de $a$	Valor de $b$	Error Estándar	Se cumple/ No se cumple
<b>Andalucía</b>	18.217	12,525	-1,22929	0,0129	<b>No se cumple</b>
<b>Aragón</b>	4.561	10,9953	-1,17076	0,0245	<b>No se cumple</b>
<b>Asturias</b>	2.495	10,3454	-1,16471	0,0329	<b>No se cumple</b>
<b>Baleares</b>	4.568	11,0352	-1,18087	0,0247	<b>No se cumple</b>
<b>Canarias</b>	5.745	11,2487	-1,15396	0,0215	<b>No se cumple</b>
<b>Cantabria</b>	1.515	9,98782	-1,19443	0,0434	<b>No se cumple</b>
<b>Castilla y León</b>	5.922	11,526	-1,2915	0,0237	<b>No se cumple</b>
<b>Castilla La Mancha</b>	4.911	11,3879	-1,31414	0,0265	<b>No se cumple</b>
<b>Cataluña</b>	26.604	12,6709	-1,11747	0,0097	<b>No se cumple</b>
<b>Ceuta</b>	108	7,85602	-1,35914	0,185	<b>No se cumple</b>
<b>Comunidad Valenciana</b>	15.586	12,2918	-1,19014	0,0135	<b>No se cumple</b>
<b>Extremadura</b>	2.123	10,4704	-1,28317	0,0394	<b>No se cumple</b>
<b>Galicia</b>	7.600	11,509	-1,177	0,0191	<b>No se cumple</b>
<b>La Rioja</b>	1.110	9,88506	1,26986	0,0539	<b>No se cumple</b>
<b>Madrid</b>	24.158	12,0436	-0,899855	0,0082	<b>No se cumple</b>
<b>Melilla</b>	101	8,11984	-8,11984	0,212	<b>No se cumple</b>
<b>Murcia</b>	4.697	11,0087	-1,155	0,0238	<b>No se cumple</b>
<b>Navarra</b>	2.110	10,0188	-1,06128	0,0327	<b>No se cumple</b>
<b>País Vasco</b>	7.418	11,3444	-1,09089	0,0179	<b>No se cumple</b>

#### 4.2. Estudio por sectores de actividad

En este apartado se va a comprobar si se cumple la ley de Zipf para cada sector de actividad al que pertenecen las empresas más grandes de España. Estas empresas son las que tenían contratados 50 empleados o más en el año 2019. Se va a utilizar esta

muestra debido a la dificultad para clasificar y manejar los datos de la muestra utilizada para el caso de las comunidades autónomas. El tamaño de la muestra de las empresas más grandes de España con 50 o más empleados es de 20.743. Como aclaración importante, se ha tenido en consideración que el sector servicios está constituido por una gran variedad de actividades que difieren enormemente entre sí, así que para realizar el análisis de este sector se ha decidido desglosarlo en actividades más específicas, dividiéndolas en grupos más similares entre sí.

A continuación, se introducen los resultados obtenidos en este segundo análisis, esta vez segmentando por sectores de actividad. Al igual que en el análisis por autonomías, se han estimado los parámetros  $a$  y  $b$  para el tamaño muestral correspondiente a cada sector de la segunda muestra, compuesta por 20.743 empresas, así como el de su error estándar. Esta información está reunida en la Tabla 4-2. Como se puede observar, igual que ocurría en el anterior caso mencionado, en casi todas las estimaciones el valor 1 para  $b$  no se encuentra contenido en sus respectivos intervalos de confianza. Por lo que se rechaza la hipótesis nula y se puede afirmar que, para el caso de la industria, la construcción, el comercio, los transportes, la hostelería, las actividades profesionales y técnicas, el sector editorial y para las telecomunicaciones se rechaza la ley. Por otro lado, para los sectores de la gestión de residuos, los servicios personales, el sector informático y el audiovisual, si que está contenido el valor 1 en sus intervalos de confianza por lo que, para estos cuatro sectores de actividad, se puede afirmar, que se cumple la ley de Zipf.

Tabla 4-2 Tabla resumen del análisis de la ley de Zipf para sectores de actividad.

Sector	Tamaño de la muestra	Valor de $a$	Valor de $b$	Error Estándar	Se cumple/ No se cumple
<b>Industria</b>	5509	13,7658	-1,28822	0,0245	No se cumple
<b>Construcción</b>	1460	12,401	-1,32484	0,049	No se cumple
<b>Comercio</b>	3895	12,6949	-1,13733	0,0258	No se cumple
<b>Gestión de residuos</b>	513	10,2287	-1,00622	0,0628	Se cumple
<b>Transportes</b>	1322	11,6009	-1,12019	0,0436	No se cumple
<b>Hostelería</b>	1497	12,0179	-1,20864	0,0442	No se cumple
<b>Actividades profesionales y técnicas</b>	1760	11,9194	-1,11974	0,0337	No se cumple
<b>Servicios personales</b>	3767	12,1502	-0,996149	0,023	Se cumple
<b>Editorial</b>	119	10,9728	-1,49851	0,1942	No se cumple
<b>Informática</b>	669	10,5675	-1,01796	0,0557	Se cumple
<b>Audiovisual</b>	127	8,98404	-1,02764	0,1255	Se cumple
<b>Telecomunicaciones</b>	105	8,07151	-0,845161	0,1166	No se cumple

El cumplimiento de la ley en los cuatro sectores mencionados tiene la implicación de que estos sectores están, si no muy cerca, en el equilibrio y su tamaño está relacionado con la posición que ocupan en el mercado español. En cambio, el sector editorial es el que tiene un valor más alto y alejado de la unidad. Es casi el equivalente a Melilla en el análisis del apartado anterior porque su valor estimado de  $b$  es de 1,498, que es aproximando de 1,5, igual que el de Melilla. Esto supone que el sector editorial es heterogéneo en cuanto a la diferencia de tamaño entre las empresas. En el Gráfico 4-2, se observa que la empresa situada en primera posición del sector editorial está muy alejada del resto y presenta varios escalones más entre las más grandes. Con el objetivo de visualizar la diferencia entre que se cumpla la ley de Zipf y que no lo haga, se considera conveniente comparar el Gráfico 4-2 con el Gráfico 4-3 que representa la distribución del sector audiovisual. La diferencia clara que existe entre

ellos es que, en el Gráfico del sector editorial, la línea desciende de manera más uniforme, en cambio, la del sector editorial desciende a trozos más abruptamente.

Gráfico 4-1 Distribución del sector editorial. Fuente: elaboración propia a partir de datos obtenidos de SABI.

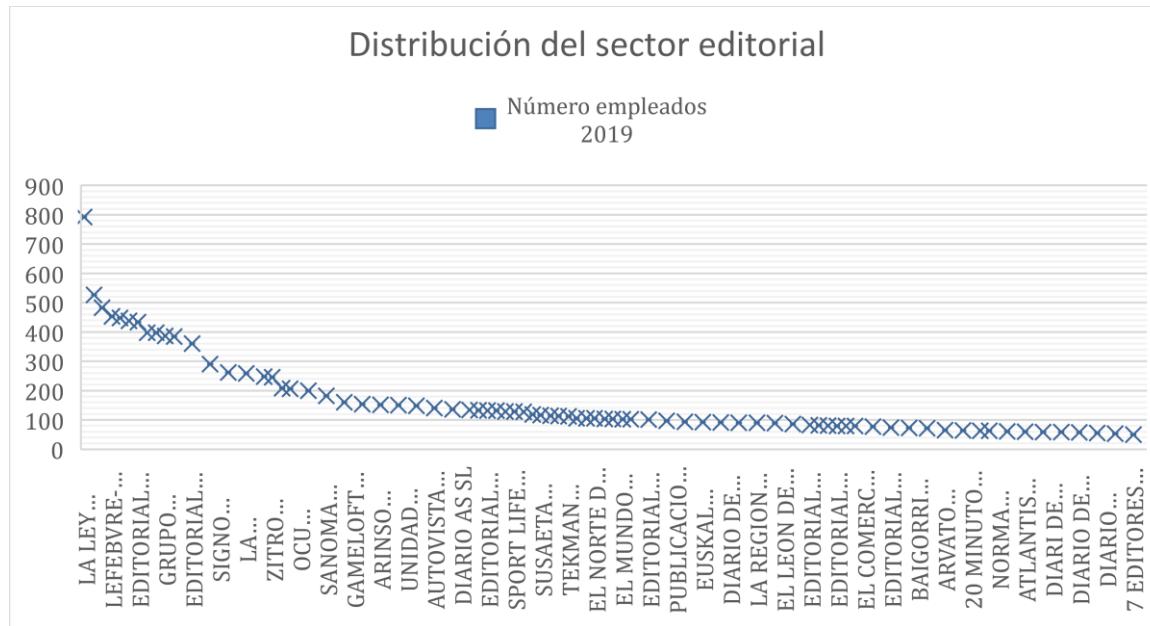
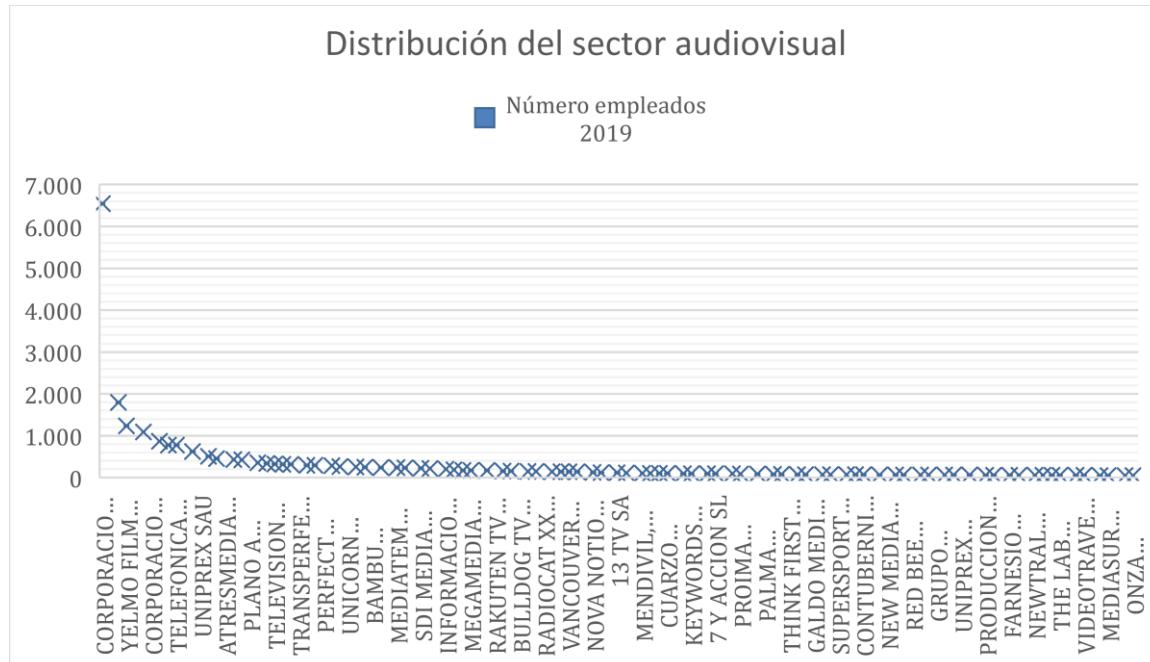


Gráfico 4-2 Distribución del sector editorial. Fuente: Elaboración propia a partir de datos obtenidos de SABI.



#### 4.3. Diferencias de los resultados según el tamaño de la muestra

Como se han utilizado dos muestras, se va a comprobar si existe una diferencia muy significativa al cambiar de muestra. Así, si utilizamos la segunda muestra, la de las empresas con más de 50 empleados, el valor de  $b$ , para España, es de -1,13909 y su error estándar es 0,0112. En la primera muestra,  $b$  era 1,1038 y su error, 0,00412. En ambos casos no se cumple la ley de Zipf, pero en la muestra más pequeña, el valor de  $b$  se aleja más de 1 y el error estándar también es más alto. Si esta misma estimación se realiza para únicamente las empresas consideradas grandes (más de 249 empleados), el valor de  $b$  es de -1,21087. Por lo que, en base a este caso, el tamaño de la muestra sí que es relevante para la distribución del tamaño empresarial, pues el parámetro de la ley potencial  $b$  disminuye al añadir empresas más pequeñas, como demuestran Segarra y Teruel (2012).

## 5. Conclusiones

El objetivo planteado por este trabajo ha sido comprobar si se cumple o no la ley de Zipf para el caso de las empresas españolas. Para conseguirlo, se han planteado varias perspectivas. Primero, se ha llevado a cabo el estudio para el caso de las empresas españolas con más de 8 empleados y para cada una de sus comunidades y ciudades autónomas. En este primer planteamiento, se ha podido observar que la ley de Zipf no se cumple para el caso de las empresas de España ni para ninguna de sus autonomías. Después, se ha procedido a realizar el estudio sobre una segunda muestra más pequeña dividiéndola por sectores. De la misma manera, no se ha podido aceptar en la gran mayoría de los casos. Por lo que se concluye que no se cumple la ley de Zipf para las empresas de España, al menos para las muestras utilizadas. Es posible que, para una muestra más grande de empresas, o al incluir todas las empresas de España, se pueda cumplir esta ley. Se supone esta posibilidad debido a que se ha visto que, al añadir más empresas pequeñas, el parámetro  $b$  se aproximaba más a la unidad. Aun así, como se ha mencionado varias veces, esta ley suele describir bien la cola alta de las distribuciones, lo que quiere decir es que, si se cumple, debería cumplirse para las empresas más grandes de España y esto no ha sido así.

Las razones por las que puede ser que no se haya cumplido para el tejido empresarial español, cuando sí que se cumplen para el tejido empresarial

estadounidense, por ejemplo, son inciertas. A continuación, se va a ofrecer alguna posible razón por la que ha podido ocurrir. La primera justificación factible, es que el tejido empresarial español es muy diferente al estadounidense. En Estados Unidos existe un porcentaje de empresas grandes bastante mayor al que existe en España. Este porcentaje mayor es aplicable al número de empresas total. Como segunda justificación, se tienen los principales fundamentos para estas diferencias, que son la forma de apoyar a las empresas por parte del gobierno y por parte de la ciudadanía. Estados Unidos es conocido por ser la incubadora de empresas por excelencia, su gobierno ofrece uno de los sistemas más sólidos y efectivos de apoyo a los pequeños negocios creados. Es decir, ayuda a que las pequeñas empresas crezcan. Acompañado de su gobierno, como no podría ser de otra manera, ya que el gobierno es una extensión del pueblo, la mentalidad estadounidense es la de emprender y explotar las ideas de cada individuo. Por otro lado, en España, ocurre todo lo contrario. Las empresas cuentan con ayudas insuficientes y un apoyo mínimo por parte del gobierno español y, por consiguiente, de su ciudadanía. Existe una reticencia a emprender y sobre todo a crecer. Las empresas se encuentran con techos impositivos y restrictivos casi imposibles de atravesar, por lo que muchos negocios que podrían ser grandes acaban por hundirse.

La ley de Zipf es una ley que pretende explicar un fenómeno natural, cómo se distribuyen naturalmente las poblaciones. Por ejemplo, las palabras que usamos no están reguladas, el lenguaje va cambiando con el paso del tiempo, pero suele seguir esta distribución. En inglés la palabra más utilizada es *the* y la segunda es *of*. Es bastante razonable que sean las palabras más usadas en esa lengua. En España, la ciudad más poblada es Madrid, tiene sentido que lo sea porque es la capital ya que, normalmente, es en las capitales de los países donde todo el mundo quiere situarse por ser el nexo con el resto del mundo y con el resto del país. Como el sector empresarial no está igual de restringido o liberado en todos los países, puede que sea esperable que en algunos países se cumpla esta ley y en otros no, predominando su cumplimiento en los países que tengan un desarrollo más natural y libre, con menos restricciones.

En la segunda parte del análisis, cuando se ha estudiado la muestra de empresas con más de 50 empleados con el fin de estudiar sus sectores de actividad, ha resultado que en unos pocos sectores sí que se ha cumplido la ley. Esto crea más incógnitas sobre

el estudio planteado. Siguiendo en la línea de las conclusiones anteriores, podría ser que algunas actividades gocen de más libertad de crecimiento y desarrollo que otras.

Finalmente, hay que mencionar que queda abierta la línea de trabajo con una muestra mayor y escogiendo otros términos para definir el tamaño como son el EBITDA, la cifra de ventas o cualquier otro que contenga una información comparable a todas las empresas. Además de estudiarla para un periodo de tiempo en vez de para sólo un año. El estudio de la ley de Zipf para el tamaño empresarial es un campo muy interesante y amplio que merece ser estudiado.

## 6. Bibliografía

- Axtell, R. (2001). Zipf Distribution of U.S. Firm Sizes. *Science*, 293, 1818-1820.
- Bottazzi, G., Pirino, D., & Tamagni, F. (2015). Zipf Law and the Firm Size Distribution: a critical discussion of popular estimators. *Journal of Evolutionary Economics*, 25(3), 585-610.
- Gabaix, X. (1999). Zipf's Law for Cities: An Explanation. *The Quarterly Journal of Economics*, 114(3), 739-767.
- Gabaix, X. (2009). Power Laws in Economics and Finance. *The Annual Review of Economics*, 1, 255-293.
- Gabaix, X., & Ibragimov, R. (2007). Rank – 1 / 2: A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business & Economics Estatistics*, 29, 24-39.
- Gibrat, R. (1931). *Les inégalités économiques: applications: aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc: d'une loi nouvelle: la loi de l'effet proportionnel*. Paris: Recueil Sirey.
- Goerlich, F. J., & Mas, M. (2008). *Sobre el tamaño de las ciudades en España. Dos reflexiones y una regularidad empírica*. Fundación BBVA - Documentos de Trabajo 06.
- Goerlich, F. J., & Mas, M. (2010). La distribución empírica del tamaño de las ciudades en España 1900-2001. ¿Quién verifica la ley Zipf? *Revista de Economía Aplicada*, 54, 133-159.
- Golpe, A. A., Iglesias, J., & Martín, J. M. (2013). Análisis pre y post-fusiones del sector compuesto por las cajas de ahorros españolas: el tamaño importa. *Desarrollo Regional Sostenible en tiempos de crisis*, 2, 896-891.
- Gómez-Miranda, M. E., & Rodriguez-Ariza, L. (2004). Evicencia Empírica en Torno al Estudio del Factor Tamaño como Condicionante Empresarial. *Revista de Contabilidad*, 7(13), 167-197.
- González-Val, R., & Sanso-Navarro, M. (2010). Gibrat's law for countries. *Journal of Population Economics*, 23, 1371-1389.
- Gutenberg, B., & Richter, R. F. (1944). Frequency of earthquakes in California. *Bulletin of Seismological Society of America*, 34, 185-188.
- INE. (2020). *Explotación Estadística del Directorio Central de Empresas, Comunidades y Ciudades Autónomas, Total CNAE*.
- INE. (s.f.). *Encuesta de Población Activa (EPA). Serie histórica (Datos en miles de personas)*.

- Lanaspa, L. F., Perdiguero, A. M., & Sanz, F. (2004). La distribución del tamaño de las ciudades en España, 1900-1999. *Revista de Economía Aplicada*(12(34)), 5-16.
- Levy, M., & Solomon, S. (1996). Power Laws Are Logarithmic Boltzmann Laws. *International Journal of Modern Physics C*, 7(04), 595-601.
- Mediana, D. G. (2021). *Estructura y dinámica empresarial en España a 1 de enero de 2020*.
- Moreno-Sánchez, I., Font-Clos, F., & Corral, Á. (2016). Large-Scale Analysis of Zipf's Law in English Texts. *PLoS ONE*, 11(1), e0147073.
- Neukum, G., & Ivanov, B. A. (1994). Crater size distributions and impact probabilities on Earth from lunar, terrestrial, planet, and asteroid cratering data. *Hazards Due to Comets and Asteroids, Space Science Series*, 359-416.
- Newman, M. (2005). Power Laws, Pareto Distributions and Zipf's Law. *Contemp Phys*, 46, 323-351.
- Pareto, V. (1896). *Cours d'Économie Politique*. Geneva: Droz.
- Roberts, D. C., & Turcotte, D. L. (1998). Fractality and Self-Organized Criticality of Wars. *Fractals*, 6(4), 351-357.
- Rose, A. K. (2006). Cities and countries. *Journal of Money, Credit and Banking*, 38(8), 2225-2246.
- Segarra, A., & Teruel, M. (2012). An appraisal of firm size distribution: Does sample size matter? *Journal of Economic Behavior & Organization*, 82, 314-328.
- Sutton, J. (1997). Gibrat's Legacy. *Journal of Economic Literature*, 35(1), 40-59.
- Wit, G. (2005). Firm size distribution: An overview of steady-state distributions resulting from firm dynamics models. *International Journal of Industrial Organization*, 23(5-6), 423-450.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge: Addison-Wesley.