

Regularidades empíricas de la estructura industrial mexicana: Evidencia de la Ley de Zipf

Empirical patterns of the mexican industrial structure: Evidence of Zipf's Law

Francisco Javier Benita Maldonado¹, Júnior Alfredo Martínez Hernández²

Benita Maldonado, F. J.; Martínez Hernández, J. A., Regularidades empíricas de la estructura industrial mexicana: Evidencia de la Ley de Zipf, *Investigación y Ciencia de la Universidad Autónoma de Aguascalientes*, 52, 21-26, 2011.

RESUMEN

Recientemente han sido encontrados, en el ámbito de las ciencias sociales, una gran cantidad de patrones y regularidades empíricas tradicionalmente asociadas con las ciencias naturales. Particularmente, en la ciencia económica, se ha encontrado evidencia empírica en torno a la existencia de fenómenos económicos gobernados por un tipo de regularidad conocida como Ley de Potencia. Fenómenos de agregación económica, como la distribución del ingreso y la riqueza, el crecimiento de las ciudades, los retornos de activos financieros y los volúmenes de transacción bursátiles, están especialmente condicionados por una Ley de Potencia, en particular, llamada Ley de Zipf. Este artículo presenta evidencia empírica suficiente para afirmar la vigencia de la Ley de Zipf para el caso de la estructura industrial en México. Utilizando la clasificación del SCIAN, se analiza la estructura de 21 subsectores industriales. Mediante el uso de cuatro métodos, se encuentra que el exponente de potencia tiende a su valor teórico esperado.

Palabras clave: Ley de Potencia, Ley de Zipf, agregación económica, organización industrial, Econometría.

Key words: Power Law, Zipf's Law, economic aggregation, industrial organization, Econometrics.

Recibido: 14 de Marzo de 2011, aceptado: 14 de Junio de 2011

¹ Facultad de Economía, Universidad Autónoma de Nuevo León, francisco_benita@hotmail.com.

² Maestría en Economía, Departamento de Métodos Cuantitativos, Universidad de Guadalajara, jr_alfredo@hotmail.com.

ABSTRAC

Recently, in the field of social sciences, a lot of empirical patterns have been detected that are traditionally engaged with the natural sciences. Particularly, Economic Science has found empirical evidence about some phenomena ruled by special regularities known as Laws of Power. Such economic processes as income and wealth distribution, the size of cities, financial asset returns and stock market trading volume are specially conditioned by a kind of power law known as Zipf's Law. This article presents enough evidence in order to assess the presence of Zipf's Law in the Mexican industrial structure. Using SCIAN data, we analyzed the structure of 21 industrial subsectors. Through a conventional econometric specification, we found that the power law exponent tends to its theoretical expected value.

INTRODUCCIÓN

La investigación científica, en particular la proveniente de las ciencias exactas, ha mantenido siempre un especial interés por el estudio de las regularidades y patrones que gobiernan una gran cantidad de fenómenos naturales. Sin embargo, dicho afán se ha trasladado recientemente hacia el ámbito de las ciencias sociales. Particularmente, es en la ciencia económica donde se han encontrado interesantes patrones de funcionamiento para una buena cantidad de fenómenos económicos.

En México, uno de los temas más importantes es el impacto del proceso de industrialización so-

bre el crecimiento y desarrollo. Si bien, los saldos del proceso no han sido del todo favorables, es importante replantear el tema a la luz del intenso proceso de reestructuración económica que se vive en todo el mundo. Proceso que está dominado, en gran medida, por la inversión que realizan los países en ciencia, tecnología y capital humano. A la par de estos factores, el proceso de desarrollo industrial está condicionado por aspectos como la distribución espacial de insumos y mercados.

Profundizar en el análisis de estos factores y en la comprensión de la estructura industrial subyacente permitirá estar en mejores condiciones para evaluar políticas y estrategias para el sector. Un examen superficial revela uno de los hechos estilizados más importantes acerca de las actividades industriales en nuestro país: existe un sesgo en la composición sectorial. Dicho sesgo se expresa en la coexistencia de pocas actividades productivas que concentran la mayor parte de los empleados, mientras que un gran número de actividades aglutina una menor cantidad de trabajadores.

Un análisis más detallado revela la existencia de un patrón empírico subyacente en la estructura industrial de nuestro país. Este patrón emerge cuando se analiza la relación entre el tamaño de las actividades industriales acorde a su número de trabajadores. Esta relación está gobernada por una regularidad estadística llamada Ley de Potencia.

De acuerdo con Gabaix (2008), una Ley de Potencia es la forma que adoptan una gran cantidad de regularidades en la ciencia económica. Se trata de una relación funcional denotada por la expresión $Y = kX^a$ o $Y = kX^a$, donde X y Y son variables de interés, k es una constante y a es conocida como el exponente de la Ley de Potencia. Según Gabaix (2009), diversas leyes económicas toman la forma de una Ley de Potencia, en particular, los procesos de agregación macroeconómica, la distribución del ingreso y la riqueza, el tamaño de las ciudades y de las empresas, así como la distribución de variables financieras, como los retornos de activos y los volúmenes de transacción.

Siguiendo a Gabaix (2009), se puede argumentar que la Ley de Potencia más frecuente es la llamada Ley de Pareto. Esta ley expresa que la probabilidad de que alguna variable X asuma un valor mayor que x es proporcional a $1/x^\zeta$ para algún número positivo ζ . Un caso especial es la Ley

de Zipf, la cual es un caso particular de la Ley de Pareto con exponente $\zeta \approx 1$.

Para el caso de la economía mexicana, existen pocos referentes académicos en donde se haya planteado el estudio de la regularidad que nos ocupa. Los trabajos encontrados se centran en analizar la pertinencia de la Ley de Zipf para la distribución del tamaño de ciudades y el crecimiento demográfico.

Frecuentemente utilizada en el estudio de las poblaciones, la hipótesis planteada por Zipf, indicaba que el tamaño de las ciudades evoluciona con respecto al tamaño de la ciudad más grande. De los Cobos *et al.* (2009) afirman que esta ley se cumple para la distribución del tamaño de las ciudades en México entre los años de 1990 y 2005. Sin embargo, dentro de la literatura también se encuentran trabajos, como el de Urzúa (2001), que ha demostrado que para el caso de la distribución del tamaño de ciudades en México, la Ley de Zipf no puede ser cierta, excepto para un determinado tamaño de la muestra.

En el presente artículo se presenta la evidencia empírica que permite afirmar, bajo la metodología empleada, que la estructura industrial en México, consignada por el tamaño de los sectores industriales acorde a su número de trabajadores, está gobernada por la Ley de Zipf. Concretamente, se analiza la relación existente entre el tamaño de los subsectores según la cantidad de empleados que laboran en él. Las estimaciones presentadas muestran que los subsectores siguen, en la mayoría de los casos, el patrón de esta ley. En otras palabras, el valor del parámetro a de la Ley de Potencia es de alrededor de 1, pudiendo, con esto, afirmar que la actividad industrial en México tiende a distribuirse de acuerdo a lo planteado por la Ley de Zipf.

MATERIALES Y MÉTODOS

La intuición explicativa detrás de una Ley de Potencia, en particular de la Ley de Zipf, es un proceso generador de datos. De acuerdo con Simon (1995), el proceso detrás de la Ley de Zipf es el crecimiento proporcional aleatorio. La forma clásica de aproximarse a la Ley de Zipf es a través de la llamada regla del tamaño de rango. Se trata de una regla determinista basada en la percepción de que la segunda más grande de las observaciones posee la mitad del tamaño de la primera más grande, la tercera observación más

grande posee un tercio del tamaño de la primera más grande y así sucesivamente.

Lo expresado por Zipf es primordialmente el cumplimiento de una regularidad empírica. El planteamiento de la Ley de Zipf no proviene de algún modelo teórico, sino más bien del fenómeno observado, a través del cual la distribución del tamaño poblacional tiende a ser la misma en todos los casos.

Son varios los métodos más utilizados para probar la validez de esta ley. Este estudio hace uso de cuatro de ellos, fijando para todas las metodologías un nivel de significancia del 5% o el cual puede ser interpretado como la probabilidad (5%) de rechazar el cumplimiento de la Ley de Zipf cuando en verdad se está cumpliendo. En general, en este tipo de estudios, este nivel de significancia es conveniente.

El primero y más usado es el método de regresión lineal por Mínimos Cuadrados Ordinarios, el cual establece que para n subsectores ordenados, según el tamaño de su personal ocupado $S_{(1)} \geq S_{(2)} \geq \dots \geq S_{(n)}$, la regresión de Zipf se ajusta a una recta de Mínimos Cuadrados Ordinarios del logaritmo de la jerarquía (rango) de i sobre el logaritmo del tamaño del subsector S_i :

$$\ln i = A - \zeta_n \ln S_{(i)} \quad (1)$$

donde A es el intercepto resultante. De acuerdo con Gabiax y Ioannides (2003), dicho procedimiento es el más utilizado en la literatura empírica. Para un tamaño grande de n se puede demostrar que el coeficiente ζ_n tiende, con probabilidad cercana a 1, al valor realizado de ζ . La ventaja de este método es que proporciona una bondad de ajuste visual para la Ley de Potencia. Para comprobar la hipótesis nula de que la Ley de Zipf se cumple, se procedió a calcular el valor correspondiente del estadístico t que prueba a un nivel de significancia que el coeficiente del logaritmo de la población debe ser igual a -1.

El segundo método fue una corrección del anterior. La estimación del exponente de Pareto por Mínimos Cuadrados Ordinarios es el método más común para probar la existencia de la Ley de Zipf. Sin embargo, Gabaix y Ibragimov (2011) señalan que dicho método es extremadamente sesgado para muestras pequeñas. En aras de remediar el sesgo, los autores proponen una corrección para la especificación econométri-

ca habitual. Se sugiere que al correr la regresión convencional de Mínimos Cuadrados Ordinarios, se debe usar como insumo estadístico una jerarquía de rangos corregida de la siguiente forma:

$$\ln(i - 0.5) = A - \zeta_n \ln S_{(i)} \quad (2)$$

Los autores demuestran, mediante métodos numéricos, que la corrección de la jerarquía ($i - 0.5$) es óptima dado que reduce el sesgo de una clasificación ordenada. Aquí también se realizó la prueba t donde se espera que el estimado no sea diferente de -1.

El tercer método es el estimador de máxima verosimilitud de Hill, el cual es un proceso de estimación alternativo en donde se estima ζ , que es el exponente de la distribución de Pareto. El estimador de Hill se obtiene a partir de la siguiente ecuación:

$$\hat{\zeta} = \frac{n-1}{\sum_{i=1}^{n-1} [\ln S_{(i)} - \ln S_{(n)}]} \quad (3)$$

Este estadístico posee las propiedades de un estimador de máxima verosimilitud. Para encontrar el error estándar se usó:

$$\sigma_n = \hat{\zeta}^2 \left(\frac{\sum_{i=1}^{n-1} (\tau_i / \hat{\zeta})^2}{n-2} \right)^{1/2} (n-1)^{-1/2} \quad (4)$$

donde $\tau_i = i(\ln S_{(i)} - \ln S_{(i+1)})$. Así, si $1/\hat{\zeta} \gg \sigma_n(1/\hat{\zeta})$ aceptamos la hipótesis de que la pendiente de ζ es igual a 0.

Por último, se empleó el estadístico LMZ propuesto por Urzúa (2001):

$$LMZ = 4n[z_1^2 + 6z_1 z_2 + 12z_2^2] \quad (5)$$

donde los valores de Z_1 y Z_2 se calculan a partir de lo siguiente:

$$z_1 = 1 - \frac{1}{n} \sum_{i=1}^n \ln \frac{S_i}{S_{(n)}} \quad \text{y} \quad z_2 = \frac{1}{2} - \frac{1}{n} \sum_{i=1}^n \frac{S_{(n)}}{S_{(i)}} \quad (6)$$

En la expresión 6, n es el tamaño de la muestra de los subsectores industriales, S_i es el tamaño del i -ésimo subsector industrial y S_n es el tamaño del subsector más pequeño en la muestra. Urzúa (2001) señala que el error de las regresiones anteriores no se distribuye normal, puesto que el logaritmo del rango es una variable discreta. En este sentido, el estadístico

LMZ permite efectuar un análisis estadístico más robusto, pues prueba la validez de la ley en términos probabilísticos. Así, este estadístico sigue asintóticamente, una distribución χ^2 con dos grados de libertad, y para probar la hipótesis nula de que dicha ley se cumple basta, dado un nivel de significancia asignado, comparar el valor de *LMZ* con el valor crítico correspondiente.

Para realizar el presente análisis de la dinámica industrial en México, se toma como base la división de los 21 subsectores manufactureros. La información se obtuvo a través de los Censos económicos de 1999, 2004 y 2009 que hacen referencia a datos de un año anterior. Se trabajó con el personal ocupado total para los 21 subsectores manufactureros que componen el Siste-

ma de Clasificación Industrial para América del Norte (SCIAN).

RESULTADOS

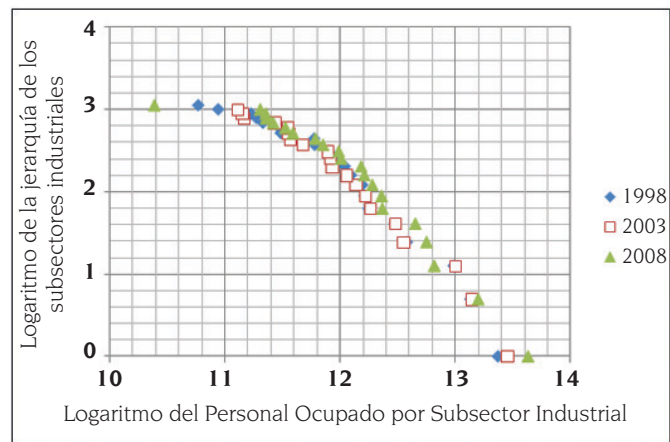
En la gráfica 1 se aprecian visualmente los resultados obtenidos por la regresión sin ajuste. En esta inspección se observa la relación entre el rango de la jerarquía de los subsectores industriales, ordenados por tamaño de trabajadores, y el logaritmo del número de trabajadores empleados en cada subsector industrial.

Los resultados de la regresión tradicional de Mínimos Cuadrados Ordinarios, sin ajuste y con ajuste, fueron los siguientes:

Tabla 1. Coeficientes de las regresiones

REGRESIÓN CONVENCIONAL DE MÍNIMOS CUADRADOS ORDINARIOS						
Año	Constante	Error	Pendiente	Error	Bondad de Ajuste	Estadístico t
1998	15.65	0.09	-1.13	0.08	0.92	1.70
2003	15.63	0.77	-1.13	0.07	0.94	1.94
2008	14.76	1.03	-1.05	0.09	0.89	0.54
REGRESIÓN DE MÍNIMOS CUADRADOS ORDINARIOS CON CORRECCIÓN DE JERARQUÍA						
Año	Constante	Error	Pendiente	Error	Bondad de Ajuste	Estadístico t
1998	17.45	1.25	-1.29	0.10	0.89	2.74
2003	17.47	1.10	-1.29	0.09	0.91	3.14
2008	16.48	1.33	-1.20	0.11	0.86	1.78

Fuente: Elaboración propia con datos obtenidos del INEGI (2001, 2005 y 2010).



Gráfica 1. Inspección de la Ley de Zipf (Relación Rango/Tamaño).

Fuente: Elaboración propia con datos obtenidos del INEGI (2001, 2005 y 2010).

Si se presta especial atención a la columna con las pendientes de las regresiones para cada año examinado, se puede observar que en ambos modelos la relación propuesta por la Ley de Zipf, no sólo se cumple de manera invariable, sino que se fortalece para cada año que transcurre. Esto hace sospechar que para muestras más grandes es posible pensar que el valor del exponente de Pareto efectivamente converge a $\zeta \simeq 1$.

Tras calcular el valor del correspondiente estadístico t , no se rechaza el cumplimiento de la Ley de Zipf en Mínimos Cuadrados sin ajuste para los tres años con el nivel de significancia que se asignó. Para el caso de la regresión con ajuste, los estimados no son estadísticamente diferentes a uno únicamente para el año de 2008, es decir, se observa que para 1998 y 2003 esta ley parece no cumplirse. Los resultados arrojados por el tercer método se sintetizan en la tabla 2:

Tabla 2. Coeficientes por máxima verosimilitud

Año	Estimador de Hill (ζ)	Error (σ_n)
1998	0.80	0.35
2003	0.78	0.40
2008	0.57	0.56

Fuente: Elaboración propia con datos obtenidos del INEGI (2001, 2005 y 2010).

Los coeficientes obtenidos por el estimador de Hill se alejan drásticamente del valor teórico esperado para la Ley de Zipf. Sin embargo, esto es normal, puesto que el estimador de Hill es construido por el criterio de máxima verosimilitud, lo que implica que es más eficiente cuando la muestra examinada es más grande. Aún así, aplicando una prueba de hipótesis empleando el error estándar no se rechaza la Ley de Zipf para los tres años.

Finalmente, con el estadístico LMZ , se encuentran valores de 6.80, 7.71 y 17.24 para 1998, 2003 y 2008, respectivamente. Se puede argumentar que tras realizar el contraste del valor crítico con una χ^2 con dos grados de libertad, se cumple $H_0: \zeta = 1$ únicamente para el año de 1998. Esto hace sospechar que aunque tras la estimación econométrica convencional y el uso de un estimador de máxima verosimilitud, hay evidencia estadística suficiente para negar esta ley en términos probabilísticos para los años 2003 y 2008.

Como muestran los resultados, dependiendo de la metodología utilizada, es posible o no, no rechazar el cumplimiento de esta ley. Además, se observa que las metodologías más robustas son más estrictas y sólo aceptan la Ley de Zipf para años específicos.

DISCUSIÓN

Los resultados presentados con anterioridad, aunque no pueden validar de manera general el cumplimiento de esta ley, sí arrojan evidencia empírica al hecho de que la estructura industrial de los 21 subsectores industriales de México tiende a distribuirse a lo largo del tiempo, acorde a lo planteado por Zipf.

Mediante el método menos ortodoxo (regresión convencional sin ajuste), se puede aceptar su cumplimiento para cada uno de los años, sin embargo, al hacer uso de metodologías más estrictas (regresión con ajuste y el estadístico de prueba LMZ) es posible rechazar esta hipótesis con el nivel de significancia asignado *a priori* para ciertos años. Es por ello que para Urzúa (2010) no es tan obvio tratar de probar dicha ley a través de regresiones, y recomienda reportar los resultados probando diferentes métodos para enriquecer el análisis y añadirle rigurosidad estadística.

Por otra parte y partiendo del hecho de aceptar el cumplimiento de esta ley para el caso de la estructura industrial, implicaría que este patrón de agrupamiento opera no sólo para el crecimiento de ciudades, sino para una buena parte de fenómenos de aglomeración, susceptibles de ser clasificados jerárquicamente.

La composición del empleo industrial en México, según el enfoque planteado en esta investigación, está sesgada hacia la industria alimentaria, la fabricación de prendas de vestir y la fabricación de equipo de transporte (en ese orden). Estas son las actividades industriales en las cuales México se ha especializado en los últimos 10 años, mientras que, por otro lado, la fabricación de productos derivados del petróleo, confección de productos textiles, las industrias metálicas básicas y la industria de la madera son subsectores que concentran escasa proporción de los empleados.

Es interesante resaltar que la Ley de Potencia, dependiendo de la metodología utilizada, se cumplió para el caso de un agregado no natural, sino contable, pues las categorías hacen

alusión a una tipología de contabilidad nacional. Aunque se realizaron diversas simulaciones analizando distintas variables para las actividades industriales, como la producción bruta total, el número de unidades económicas o el total de activos fijos a nivel estatal, municipal, por ramas, subramas y clases, en todos los casos se rechazaba con un amplio margen esta ley.

Finalmente, hay que reconocer las limitaciones del trabajo, pues aún no existe un consenso entre investigadores acerca de cuál es la mejor metodología para probar esta ley; es por ello que este estudio hace uso de varios de ellos para abonar a la discusión sobre si se cumple o no esta regularidad empírica.

CONCLUSIONES

Presentar evidencia empírica sobre la validez de la Ley de Zipf no es, en sí mismo, un propósito, es tan sólo el punto de partida que abre un debate acerca de la explicación que está detrás del crecimiento proporcional aleatorio. Pese a que la Ley de Zipf ha sido validada en una gran cantidad de estudios, no existe un consenso respecto a las causas que la originan. En este punto es donde se encuentra el futuro de este debate.

Futuras extensiones del presente trabajo residen en el reto de encontrar las causas que originan la Ley de Zipf.

Al respecto, se dejará sobre la mesa tan sólo algunas intuiciones preliminares. Es razonable suponer que la Ley de Zipf pudiera tener sus causas en dos fuerzas que operan al interior de la estructura industrial y que afectan el crecimiento de las industrias y, por ende, la distribución de sus trabajadores. La primera es el valor esperado de la tasa de crecimiento de sus trabajadores empleados, la cual puede ser una función directa del nivel de producción y de la tecnología empleada al interior del sector. La segunda fuerza reside en las características propias de cada subsector y los determinantes de concentración del empleo, como la intensidad en el uso de la mano de obra, las diferencias tecnológicas, el tamaño y acceso a los mercados e, inclusive, el carácter multinacional de las firmas.

Los elementos aquí expuestos podrían ser el insumo para elaborar un modelo teórico que explique las fluctuaciones laborales al interior de un sector industrial dado. Esto podría constituir un primer intento para dar respuesta a las causas que originan la dinámica enunciada por la Ley de Zipf.

LITERATURA CITADA

- DE LOS COBOS, S.; GODDARD, M.; GUTIÉRREZ, A., La Ley de Zipf en el crecimiento demográfico de México, presentado en XV SIGEF Congress, *Economic and Financial Crisis: New Challenges and Perspectives*, Lugo, España, 170-179, 2009.
- GABAIX, X., *Power Laws, The New Palgrave Dictionary of Economics*, 2nd Edition Eds. Steven N. Durlauf and Laurence E. Blume. Palgrave Macmillan, New York, 2008.
- GABAIX, X.; IOANNIDES, Y., The Evolution of City Size Distributions. *Handbook of Regional and Urban Economics*. North-Holland. Vol. 4, Chapter 53, pp. 2341-2378, 2004.
- GABAIX, X., Power Laws in Economics and Finance. *Annual Review of Economics*. Vol. 1, pp. 255-293, 2009.
- GABAIX, X.; IBRAGIMOV, R., Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business Economics and Statistics*. Vol. 29, No. 1, pp. 24-39, 2011.
- INEGI., XV Censo Industrial. *Censo Económico 1999*. Aguascalientes: INEGI, 2001.
- INEGI, XVI Censo Industrial. *Censo Económico 2004*. Aguascalientes: INEGI, 2005.
- INEGI, XVII Censo Industrial. *Censo Económico 2009*. Aguascalientes: INEGI, 2010.
- SIMON, H., On a class of skew distribution functions. *Biometrika*. Vol. 44, pp. 425-440, 1955.
- URZÚA, C., Las ciudades mexicanas no siguen la ley de Zipf, *Estudios demográficos y urbanos*. Vol. 16, pp. 661-669, 2001.
- URZÚA, C., Testing for Zipf's law: A common pitfall. EGAP. Documentos de trabajo, 2010.