CAPÍTULO IV
IMPLICACIONES EMPÍRICAS

En este capítulo se describe la principal base de datos utilizada en el trabajo, prestando especial atención a los potenciales problemas de selección de muestra. Seguidamente se exponen los tres nuevos resultados empíricos descritos en la introducción y se analiza cómo el modelo teórico propuesto es capaz de reproducirlos.

1. DATOS

Las dos principales unidades de estudio de ciudades son las áreas metropolitanas y las ciudades definidas administrativamente. Estas últimas (de ahora en adelante llamadas *ciudades*) son unidades políticas que están contenidas en una o varias áreas metropolitanas. De acuerdo con la definición del Census Bureau americano, un área metropolitana debe incluir, al menos, una ciudad de 50.000 o más habitantes. La diferencia entre ambos conceptos es, en algunos casos, abismal. Considérese, por ejemplo, la ciudad de Nueva York. Si se tiene en cuenta su definición como ciudad, su población en el año 2000 era de 8.008.278. Por otro lado, su área metropolitana en ese mismo año incluye Nueva York, Northern New Jersey, Long Island, las ciudades de New Haven (Connecticut), Newark y Trenton (New Jersey), así como varias pequeñas localidades en el este de Pennsylvania contando con una población de 21.199.865. Por otro lado, el error de medida puede ser considerable en este tipo de estudios. Como sugieren GLAESER y SHAPIRO (2001) «...las ciudades difieren enormemente en tamaño, a menudo incluyendo totalmente el área metropolitana, y en otros casos considerando solamente una pequeña zona del centro de la ciudad. Por lo tanto, comparar ciudades entre sí requiere una cierta aceptación de errores de medida».

Como sugiere EECKHOUT (2004), existen diferentes criterios para elegir la unidad de análisis a usar para un estudio relacionado con ciudades. El primero de ellos es definir cuál es el objetivo del proyecto de investigación. Por ejemplo, el área metropolitana es una unidad de estudio más razonable si se pretende estudiar mercados de trabajo a nivel local, o el impacto de construir una carretera o un aeropuerto en una localidad. Por otro lado, usar el concepto de ciudad es más adecuado para estudiar externalidades de capital humano que actúan en general a un nivel considerablemente local, el impacto de diferentes impuestos urbanos, o políticas de educación. El segundo criterio es más práctico: la existencia de datos de una mínima calidad.

Existen cuatro fuentes de datos internacionales sobre poblaciones de ciudades durante largos intervalos de tiempo. En todos los casos, éstos tienen una frecuencia de diez años. La primera de ellas es de Henderson y contiene datos sobre áreas metropolitanas en diferentes países durante el período 1960-2000. Otra fuente de datos es la de Brinkhoff, quien presenta información sobre la población de varias ciudades en 79 países durante el período 1970-2000. La tercera fuente de información es World Urbanization Prospects: The 2003 Revision (WUP). En este texto, se provee información sobre aglomeraciones urbanas globales, con 750.000 o más habitantes durante el período 1950-2000. Finalmente, la base de datos más completa, de Lahmeyer, contiene datos sobre el tamaño de las mayores ciudades en 225 países. En muchos casos, existen datos para el período 1790-2000. El presente trabajo combina datos de ciudades de Lahmeyer y Brinkhoff [1] con datos sobre áreas metropolitanas de Henderson y el WUP. La tabla 1 del apéndice D enumera los países estudiados y los años en los que existen datos en cada caso [2].

---

[1] De ahora en adelante se refiere a esta base de datos como Lahmeyer-Brinkhoff.

[2] Debido a la limitada disponibilidad de datos, la muestra de países y años utilizados a lo largo de este trabajo varían ligeramente dependiendo del ejercicio empírico que se realice.

Cuando se realiza un estudio sobre la dinámica de la población en un sistema de ciudades, es absolutamente crucial entender el método usado para construir la base de datos que se está utilizando, y "corregirlo" en caso de que sea necesario. El mayor problema potencial es la presencia de sesgos de selección de muestra. Considérese, como ejemplo, una base de datos que contiene datos en el período 1960-2000 para diferentes países. Si el proceso de selección de los datos es elegir las 50 ciudades mayores en el año 2000 y, posteriormente, seleccionar estas mismas ciudades en 1990, 1980, 1970, y 1960, es evidente que se habrán seleccionado solamente aquellas ciudades que han crecido más rápidamente durante este intervalo de tiempo.

La mayor parte de las bases de datos no especifican cómo han sido construidas. Sin embargo, una forma sencilla de inducir si se ha usado un criterio basado en un período $t > t_0$ (es decir, eligiendo las ciudades mayores en el período t y seleccionando esas mismas en cada período anterior) es calcular el número de ciudades inicialmente "pequeñas" que han perdido población con el paso del tiempo. Si se observa que la población de ninguna de estas ciudades ha decrecido a lo largo de un intervalo arbitrariamente largo, uno debería sospechar que éstas han sido seleccionadas en algún período $t > t_0$, puesto que, bajo esa regla de selección, solamente esas ciudades que crecieron suficientemente durante ese tiempo habrían sido elegidas. Este mecanismo obviamente elimina de la muestra aquellas ciudades que, siendo inicialmente pequeñas, han perdido población, porque no han alcanzado un tamaño mínimo para ser seleccionadas en el período [3] $t > t_0$.

El objetivo de este trabajo es entender cómo las ciudades crecen con el paso del tiempo. Idealmente, esto debería estudiarse con un panel de ciudades completo, es decir que contuviera todas las ciudades de cada país durante las décadas que cubre el estudio. Lamentablemente, esta información no está disponible para prácticamente ningún país. Una estrategia alternativa es seleccionar las ciudades mayores año por año, y seguir la evolución de su población a lo largo del tiempo. Ciertamente, se podría argumentar que la selec-

[3] La base de datos de HENDERSON y la usada en EATON y ECKSTEIN (1997) son ejemplos de datos con un pequeño porcentaje de ciudades que pierden población a lo largo del tiempo.

ción de estas ciudades en la fecha t_0 está afectada por variables inobservables (las ciudades con mayor población en 1800 pueden tener unas características determinadas que no son observables por el analista), con lo cual sigue habiendo una determinada selección de muestra. Este problema es, de algún modo, inevitable, pero es menos grave para el ejercicio que se está considerando en este trabajo [4]. Finalmente, los problemas de selección de muestra señalados anteriormente y el objetivo de establecer una comparación internacional con un número considerable de países, hace que sea necesario seleccionar solamente las J ciudades con mayor población en la década inicial [5].

2. EL AUGE Y DECLIVE DE LAS CIUDADES

A) Descripción del hecho empírico

En este apartado se calculan los ratios de población de las ciudades más grandes de los países de los que se dispone de datos, usando una metodología que reduce drásticamente los problemas de selección de muestra. Estos ratios se calculan (usando, siempre que es posible, la población total y urbana en el denominador) seleccionando las ciudades en el año inicial. El primer motivo para seleccionar las ciudades de esta forma es, como se señalaba anteriormente, solventar el potencial problema de selección de muestra. Anteriores estudios basados en el cálculo de ratios contemporáneos (es decir que seleccionan las ciudades más grandes en cada país, década por década), no consideran la posibilidad de que las J ciudades mayores de su muestra en un momento dado sean las mayores solamente por el hecho de que han sido seleccionadas como tales en algún período posterior. El segundo motivo es que, desde un punto de vista teórico, es más plausible estudiar el comportamiento de las ciudades si la identidad de las mismas no cambia con el tiempo. En otras palabras, los estudios existentes las consideran "ciudades anónimas" cuya identidad cambia frecuentemente en diferentes décadas.

---

[4] Recuérdese que el objetivo es entender cómo las tasas de crecimiento de las ciudades evolucionan *después* de esa fecha inicial.

[5] Más adelante se presentan resultados usando $J = 5$. En el caso de $J = 10$, la evolución de los ratios de población y otros estadísticos es muy parecida, pero el número de países con datos disminuye considerablemente.

Como se comentaba anteriormente, la estrategia empírica del trabajo es la siguiente. En primer lugar, se seleccionan las *J* ciudades con mayor población en una fecha inicial. Esta fecha varía entre países debido a la disponibilidad de datos. Este conjunto de ciudades es el que se utiliza en todos los años siguientes, incluso en el caso de que algunas de ellas dejen de estar entre las mayores en alguno de los períodos. En segundo lugar, usando datos de Lahmeyer-Brinkhoff y de la Penn World Table sobre población total de cada país, se calcula el ratio de la población en estas ciudades sobre la población nacional. Nótese que estos porcentajes, al usar la población nacional en el denominador, proveen de forma indirecta información sobre todas las localidades del país, incluso aquéllas para las que el ratio no se calcula. Por ejemplo, una evolución decreciente de estos ratios a lo largo del tiempo sugiere que el conjunto de localidades exhibe regresión hacia la media: si las *J* mayores ciudades representan una fracción cada vez menor de la población total (o urbana), esto implica que las ciudades pequeñas y medianas deben estar creciendo de forma mas rápida que las grandes.

En la introducción se presentaron los casos de Estados Unidos y Francia. El patrón de U invertida en los ratios de población es muy similar en otros países, como por ejemplo Canadá o Japón. Para países menos desarrollados se observa un patrón parecido aunque, en muchos casos, el pico en la concentración de población se alcanza mucho más tarde. Los gráficos 7 y 8 del apéndice D muestran los casos de México y China, respectivamente. Ciudad de México alcanza su máxima concentración en 1970, mientras que Beijing lo hace en 1990. Por otro lado, Nueva York alcanza su máximo ratio de concentración en 1940, mientras que eso sucede en 1926 en París. Esto parece sugerir que los países ricos tienden a alcanzar su ápice de concentración de población mucho antes que países en vías de desarrollo. De hecho, en algunos países extremadamente pobres, la mayor ciudad todavía no ha alcanzado ese nivel máximo. Por ejemplo, el gráfico 9 muestra el caso de Angola, país en el que su principal ciudad, Luanda (la única para la que se tienen datos fiables) está todavía aumentando su peso en el total de la población nacional. Estas observaciones también sugieren la posibilidad de que el momento de máxima concentración esté relacionado con el proceso de urbanización del país en cuestión. En CUBERES (2005) se estudia esta hipótesis en más detalle.

La definición de población urbana que se usa en este trabajo es más arbitraria que la de población total de un país. Debido a que para la mayor parte de los países existen datos oficiales solamente para el período 1950-2000, en este trabajo se usa la suma de la población de las *J* mayores ciudades, donde *J* es igual a 10, 25 o 50, dependiendo de la disponibilidad de información. Esta definición de población urbana presenta dos problemas. En primer lugar, es difícil seguir el mismo criterio para diferentes países, debido a problemas con los datos, lo que complica la interpretación de comparaciones internacionales. El segundo (y más importante) problema está nuevamente relacionado con la selección muestral: en la mayoría de países, especialmente en los menos desarrollados, el número de ciudades con datos de población en los años iniciales es mucho menor que en años posteriores, lo que sugiere que, posiblemente, estas ciudades han sido seleccionadas en base a algún período posterior. Por lo tanto, los resultados que se presentan usando población urbana deben ser interpretados con precaución. Nótese que el uso de población urbana en el denominador de los ratios de población es particularmente interesante para el subperíodo en el que se observa un aumento de la concentración de población. El motivo es que, puesto que la población urbana ha crecido más rápidamente que la población total en la mayor parte de países durante el período de tiempo estudiado, cuando la población total se usa para construir los ratios, el grado de regresión hacia la media (es decir, la parte decreciente de la U invertida) está infravalorado, mientras que la parte ascendiente de la misma, que representa el grado de divergencia, está sobrevalorado. Cuando se utiliza la población urbana, es igualmente cierto que los países menos desarrollados alcanzan su máxima concentración de población más tarde que los países ricos. Por ejemplo, París alcanza su máxima aglomeración relativa de población en 1881, mientras que Nueva York lo hace en 1850. Por otro lado, en Ciudad de México esto sucede en 1980, y en 1990 en el caso de Beijing. Por lo que respecta al patrón de U invertida, la conclusión es que, a pesar de los numerosos problemas de datos comentados con anterioridad, la evolución de los ratios de población sigue siendo bien descrita por esta curva cuando se usan datos de población urbana en el denominador.

Un problema a la hora de interpretar estos gráficos es determinar cuál es la contribución de la población individual de cada una de las ciudades en el ratio de concentración de población nacional. Por ejemplo, cuando se considera el ratio de población en Nueva York y Philadelphia sobre la población de Estados Unidos, es posible que el patrón de U invertida sea, en realidad, causado principalmente por la mayor ciudad, es decir, Nueva York. El gráfico 10 muestra las contribuciones individuales de varias ciudades estadounidenses a esta medida de concentración de población [6]. Se puede comprobar que el patrón de U invertida es seguido por la mayor parte de las ciudades, no sólo la mayor. Por otro lado, para ilustrar los resultados de estudios parecidos al realizado en este trabajo, el gráfico 11 muestra los ratios de población contemporáneos (es decir, que eligen las mayores ciudades en cada década) en las cinco ciudades mayores de Estados Unidos. El patrón de U-invertida es muy parecido al que se obtiene usando las ciudades mayores seleccionadas en el período inicial [7].

Finalmente, es interesante preguntarse si este patrón de concentración de población se mantiene cuando se consideran datos de áreas metropolitanas en lugar de ciudades. La conclusión es que, aunque menos pronunciada, la curva que se obtiene es similar a la que se observa en el caso de las ciudades. Un problema al usar este tipo de datos es que es complicado establecer comparaciones entre diferentes países debido al hecho de que la definición de lo que constituye un área metropolitana cambia sustancialmente de país en país. Una diferencia importante con el caso de las ciudades es que, al considerar este concepto de ciudad, el período de crecimiento del ratio es más prolongado. Sin embargo, en la mayor parte de los casos, la concentración de población en las áreas metropolitanas alcanza eventualmente un máximo, generando una evolución de U invertida.

¿Funciona la ley de Gibrat?

El hecho de que la concentración de población siga una evolución de U invertida a través del tiempo no implica necesariamente

[6] La escala de la derecha en este gráfico representa el ratio de población de Nueva York, mientras que el de la izquierda representa el del resto de ciudades.

[7] Las leyendas $K1,...,K5$ se refieren al peso de cada una de las cinco ciudades en la población nacional.

que la ley de Gibrat deba ser rechazada. Para contrastar rigurosamente la hipótesis expuesta por la ley de Gibrat (esto es, que las ciudades crecen siguiendo un proceso aleatorio) es necesario seguir la evolución de todas las ciudades de la muestra entre los períodos 1800 y 2000. Si eso fuera posible, se podría calcular el coeficiente de correlación entre las tasas de crecimiento de estas ciudades y su tamaño inicial. La ley de Gibrat sería entonces valida si este coeficiente no fuera significativamente diferente de cero [8].

El problema con esta estrategia es que, en el caso de que no sea posible seguir la evolución a través del tiempo de una de estas ciudades, el ejercicio resulta sesgado. A continuación se describe una forma de limitar este sesgo. En primer lugar, se divide el período a estudiar en dos partes: la primera de ellas es el subperíodo en el que la concentración de población aumenta en las J ciudades mayores. El segundo intervalo de tiempo consiste en los años que muestran un decrecimiento del ratio de concentración. Para cada uno de los subperíodos, se sigue la evolución en cada década de las ciudades inicialmente mayores. Finalmente, se calcula el coeficiente de correlación entre las tasas de crecimiento de las ciudades y su tamaño inicial.

En el caso de Estados Unidos, la medida de concentración (usando población nacional en el denominador) para las ciudades más pobladas en 1790 alcanza su máximo en 1940. Debido a la disponibilidad de datos, solamente es posible seguir la evolución de las 22 mayores ciudades durante el intervalo 1790-1840. El coeficiente de correlación es de $\rho = 0{,}21$ (gráfico 12) indicando un cierto grado de divergencia entre ciudades. Para el período 1940-1990, en el que los ratios decrecen, es posible seguir la evolución de un número mucho mayor de ciudades. En este caso, la correlación es $\rho = -0{,}87$ (gráfico 13), indicando un claro componente de regresión hacia la media.

Una segunda alternativa es limitar el análisis al comportamiento de las cinco ciudades inicialmente mayores. Este método provee una base de datos consistente, en el sentido de que prácticamente

[8] Otra alternativa es estimar el parámetro de un modelo autorregresivo de orden 1 [AR(1)] y contrastar la hipótesis de que es igual a uno. Esta metodología se ha seguido para el caso de Estados Unidos y se rechaza claramente la hipótesis de que el coeficiente es igual a uno, es decir, se rechaza la ley de Gibrat. Este resultado se muestra en CUBERES (2005).

no existen problemas de selección de muestra. Además, esta base de datos contiene las mismas ciudades a lo largo de todo el intervalo de tiempo estudiado, lo que facilita interpretar su evolución en términos del modelo propuesto en el capítulo III del trabajo. Esta segunda estrategia es la que se sigue en el resto del estudio.

La ley de Gibrat solamente podría ser compatible con el patrón de U invertida en la concentración de población si se considerasen períodos de tiempo extremadamente largos. Por ejemplo, se podría argumentar que la población de Nueva York creció más rápidamente que la de Estados Unidos durante el período 1790-1940 y, a partir de ese año, empezó a crecer por debajo de ella. Por tanto, en media (calculada a lo largo de doscientos años), se puede concluir que Nueva York ha crecido aproximadamente en la misma tasa que la población de Estados Unidos. El obvio problema con este argumento es que solamente usa información en dos momentos del tiempo, el inicial y el final, sin prestar atención a la evolución dinámica (que es claramente no lineal) de este ratio. Una segunda posibilidad es que este patrón no lineal esté principalmente causado por el número de ciudades existentes en cada década. Como muestran DOBKINS y IOANNIDES (1998a), los Estados Unidos se caracterizan por un importante aumento en el número de ciudades entre los años 1900 y 2000. La aparición de nuevas ciudades podría, en efecto, explicar la parte decreciente de la concentración de población, pero implicaría que el número de ciudades disminuyó en la parte creciente de la misma. Esto contradice directamente la observación de que la mayor parte de ciudades (al menos en el caso de Estados Unidos) aparecieron en este período de tiempo. Además, esta explicación es inválida para países como Francia y Japón, países con sistemas urbanos maduros desde el inicio de la década estudiada aquí, donde el número de ciudades ha sido muy estable a lo largo del período de tiempo analizado.

La principal conclusión de este capítulo es que el patrón de U invertida en la concentración de población es robusto al tipo de datos usados y no está causado por selección de muestra.

B) Evaluación del modelo

El patrón de U invertida en la concentración de población es probablemente el resultado empírico más sorprendente de este estudio. El modelo teórico propuesto en el capítulo III es capaz de reproducir esta evolución satisfactoriamente (ver gráficos 4 y 6). Como se explicó anteriormente, el mecanismo económico que genera esta evolución es el siguiente: la ciudad con un mayor stock inicial de capital (ciudad *A*) recibe toda la inversión durante algunos períodos. El motivo es que, para niveles bajos de capital instalado, los rendimientos crecientes de escala son más importantes que los correspondientes costes de congestión. Sin embargo, en un momento dado, estos costes (que crecen de forma convexa) superan las ventajas de producir en una ciudad grande y la inversión empieza en la ciudad inicialmente pequeña (ciudad *B*). Dado que la ciudad *B* es en ese momento mucho menor que la ciudad *A* (recuérdese que una fracción del stock de capital en la primera se ha depreciado a lo largo del tiempo), esta ciudad tiene un gran potencial de crecimiento y, por lo tanto, su tasa de inversión es estrictamente mayor que la de la ciudad *A*.

Este mecanismo claramente genera el patrón de U invertida en la ciudad *A*, pero también genera un patrón de U en la ciudad *B*. Sin embargo, en los datos no se observan ciudades donde la concentración de población sigue este patrón de U. Esto parece señalar una debilidad del modelo propuesto. No obstante, si se considera un modelo con muchas ciudades, cuando la mayor de ellas está creciendo en solitario, el ratio de concentración en el resto de ciudades decrece, pero a una tasa muy lenta en cada una de ellas, puesto que el declive se reparte entre un gran número de ciudades. En este caso, los ratios de población en estas ciudades son prácticamente planos, como es el caso en muchas ciudades de la muestra.

3. ASIMETRÍA EN LAS TASAS DE CRECIMIENTO DE LAS CIUDADES

A) Descripción del hecho empírico

El segundo hecho empírico presentado en este trabajo está relacionado con la distribución de las tasas de crecimiento de las ciudades de un país en un momento dado. Esta distribución de la sección cruzada de ciudades es, típicamente, asimétrica, con un mayor peso en la parte derecha de la distribución. En otras palabras, dicha distribución se caracteriza por la existencia de unas pocas ciudades con una tasa de crecimiento muy superior a las demás. Esta observación es claramente incompatible con la ley de Gibrat, que implica una distribución simétrica alrededor de una tasa de crecimiento nula.

Una forma de medir el grado de asimetría alrededor de la media de una función de distribución es calculando su coeficiente de asimetría. Valores positivos (negativos) de este estadístico indican que los valores positivos (negativos) de la variable en cuestión tienen mayor peso en su distribución. Una segunda posibilidad es calcular la diferencia entre la media y la mediana muestrales. De nuevo, un valor positivo (negativo) de esta medida indica asimetría hacia la derecha (izquierda). La definición del coeficiente de asimetría que se usa en este trabajo es la siguiente:

$$sk = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3$$

donde n es el número de observaciones de la muestra, x_i y $\bar{x}$ representan el valor de una observación y la media muestral respectivamente y s es la desviación estándar de la muestra.

La tabla 2 en el apéndice D muestra el coeficiente de asimetría medio de las tasas de crecimiento en las cinco ciudades mayores en cada década en 52 países[9]. En el 90 por 100 de los casos el coeficiente es positivo. La última columna de la tabla muestra la diferencia promedio entre la media y la mediana de la distribución. En este caso, el estadístico es positivo en el 86 por 100 de los países. Para poder calcular el grado de significatividad de estos estadísticos, se procede al análisis de las 20 ciudades mayores en cada década (tabla 3). El número de observaciones se elige de forma que sea posible establecer una comparación entre un grupo razonablemente grande de países. Debido a la disponibilidad de datos, el ejercicio se lleva a cabo con 21 países. La tercera columna muestra que en 84 por 100 de las décadas el coeficiente de asimetría es positivo.

[9] Para cada país, se calcula el coeficiente de asimetría de las cinco mayores ciudades en cada década y se presenta su media a través del tiempo.

A continuación se procede a contrastar el nivel de significatividad de esta asimetría en cada uno de los países usando cuatro contrastes diferentes. El primero de ellos está basado en BROWN (1996) y TABACHNICK y FIDELL (1996). Este contraste establece que una distribución es significativamente asimétrica si su coeficiente de asimetría tiene valores superiores o iguales a dos desviaciones estándar del coeficiente de asimetría, donde el error estándar del coeficiente de asimetría se define como $ses = \sqrt{\frac{6}{n}}$, siendo n el tamaño de muestra.

El segundo contraste combina un test de normalidad basado en el coeficiente de asimetría y otro basado en la kurtosis. Este contraste requiere un mínimo de ocho observaciones y la hipótesis nula es que la distribución es normal. Más detalles pueden encontrarse en D'AGOSTINO, BALANGER y D'AGOSTINO, Jr. (1990) con la corrección empírica desarrollada en ROYSTON (1991c).

El tercer y cuarto contraste son, respectivamente, el test W de SHAPIRO-WILK y el W' de SHAPIRO-FRANCIA. Ambos contrastes están basados en la normalidad de la distribución. El primero se presenta en detalle en SHAPIRO y WILK (1965), mientras que el último aparece en SHAPIRO y FRANCIA (1972) y ROYSTON (1983).

Las columnas 4-7 de la tabla 3 muestran el porcentaje de décadas en los que el coeficiente de asimetría positivo es estadísticamente significativo (al 10 por 100) en cada uno de los contrastes. Utilizando el contraste de BROWN (1996), las distribuciones están significativamente sesgadas hacia la derecha en el 41 por 100 de los casos. Este porcentaje aumenta considerablemente en los otros tres contrastes (50 por 100 de los casos son significativos).

Los resultados de estos contrastes estadísticos indican un coeficiente de asimetría hacia la derecha en la mayor parte de los casos. Una posible explicación de la falta de significatividad en una importante fracción de países es el pequeño número de observaciones disponibles. Por otro lado, es interesante destacar que, en muchos casos, se produce un claro descenso en el coeficiente de asimetría a lo largo del tiempo [10], lo que disminuye considerablemente el valor medio de este coeficiente.

[10] Véase CUBERES (2005).

B) Evaluación del modelo

El modelo teórico puede explicar también este segundo nuevo hallazgo empírico, es decir, el hecho de que la distribución de tasas de crecimiento de las ciudades es, en un período dado, asimétrica, con mayor peso de la distribución en ciudades con crecimiento positivo. Para mostrar esto de forma explícita, considérese una extensión del modelo con $J > 2$ ciudades (véase apéndice C). Si se asume que estas ciudades difieren entre sí en el stock inicial de capital, solamente la ciudad con el mayor stock va a experimentar una tasa de inversión neta positiva. Una vez los costes de congestión en esta ciudad alcancen un tamaño críticamente elevado, la inversión comenzará en la ciudad que le sigue en tamaño inicial. Puesto que el stock de capital instalado en esta segunda ciudad es menor al de la ciudad mayor, la tasa de crecimiento de la ciudad pequeña es estrictamente mayor al de la primera.

Este proceso va a continuar hasta que uno de los siguientes eventos ocurra: en primer lugar, es posible que la tercera ciudad con mayor capital inicial se convierta en una localidad provechosa para invertir antes de que la segunda ciudad haya convergido con la primera. En este caso, la tercera ciudad va a crecer más rápido que las otras dos durante algunos períodos. La otra posibilidad es que la segunda ciudad converja totalmente con la primera antes de que la inversión empiece en la tercera, lo que implica que las dos primeras ciudades crecerán en la misma tasa durante algunos períodos y, eventualmente, la tercera ciudad crecerá más rápido que ellas. En el primer escenario (no convergencia) se verifica estrictamente la predicción de que en cada período existe una ciudad que crece más rápido que el resto, generando una distribución de tasas de crecimiento asimétrica (hacia la derecha). En el caso de no convergencia, existirán períodos en los que más de una ciudad crezca en la misma tasa. Intuitivamente, si la distribución inicial de stocks de capital tiene una varianza pequeña y los costes de congestión son relativamente altos, estos intervalos tenderán a ser cortos. Se comentaba anteriormente que, a pesar de que el coeficiente de asimetría es claramente positivo en la mayor parte de países y períodos, este coeficiente no era estadísticamente significativo en bastantes casos. Una posible explicación para esta falta de significatividad es esta posibilidad teórica de que algunas ciudades crezcan en la misma tasa (es decir, verifiquen la ley de Gibrat).

4. EL RANGO DE LAS CIUDADES QUE CRECEN MÁS RÁPIDO

A) Descripción del hecho empírico

La tercera evidencia empírica relacionada con la distribución del tamaño de ciudades es que el rango de la ciudad que crece más rápidamente en un período de tiempo determinado aumenta a lo largo del tiempo en la mayor parte de países. Seguidamente se procede a estudiar la evolución del rango de las ciudades con la mayor tasa de crecimiento en cada década.

La columna 3 de la tabla 4 indica la evolución cualitativa de este estadístico durante el período de tiempo estudiado en cada uno de los países. El patrón es estrictamente creciente en el 73 por 100 de los casos. En 11 casos (21 por 100) esta medida decrece con el tiempo, mientras que la tendencia es constante en el resto de casos. En media, el rango de la ciudad que crece más rápidamente aumenta a una tasa de el 38 por 100 por año. Sin embargo, considerando solamente la ciudad que crece más rápido en cada década, genera una medida que es muy volátil y no utiliza toda la información disponible. Una forma de mejorar este estadístico es construir un índice ponderado del rango de las *M* ciudades que crecen más rápido en cada período considerado, donde los pesos se determinan por la tasa de crecimiento de cada una de estas ciudades. Considérese el ejemplo de la tabla 5, donde se presentan las tasas de crecimiento de las cinco ciudades más pobladas de Estados Unidos en el año 1790, en orden descendiente de población. La suma de las cinco tasas de crecimiento es 2,08 y el índice ponderado del rango de estas ciudades se calcula así:

$$WR = 1 * \left(\frac{0{,}6}{2{,}08}\right) + 2 * \left(\frac{0{,}36}{2{,}08}\right) + 3 * \left(\frac{0{,}67}{2{,}08}\right) + 4 * \left(\frac{0{,}31}{2{,}08}\right) + \\ + 5 * \left(\frac{0{,}14}{2{,}08}\right) = 2{,}53$$

El rango ponderado de las mayores ciudades crece a una tasa media del 3 por 100 para los países estudiados en este trabajo. El gráfico 14 muestra la evolución de este índice a lo largo del tiempo para el caso de Estados Unidos. Este nuevo estadístico es claramente menos volátil que el considerado anteriormente y hace un mejor uso de la información disponible. Cuando se usa esta medida, los resultados son muy parecidos a los obtenidos considerando solamente el rango de la ciudad con mayor tasa de crecimiento (véase la cuarta columna de la tabla 4). En el 63 por 100 de los casos, el patrón es creciente. En 12 países (23 por 100) el índice decrece con el tiempo, mientras que no hay un patrón claro para el restante 13 por 100 de casos. En algunos países (principalmente los más desarrollados), algunas ciudades tienen tasa de crecimiento negativas, especialmente en las últimas décadas de la muestra. Tasas de crecimiento negativas complican la interpretación del índice en el caso en que la tasa media de crecimiento en las ciudades consideradas en una década es negativa. El motivo es que una ciudad con crecimiento negativo tiene, en ese caso, un impacto positivo en el índice, mientras que el modelo predice que el rango de una ciudad que decrece en tamaño debería recibir un peso menor en el índice. Para evitar estas complicaciones, se excluyen las observaciones con crecimiento negativo. En cualquier caso, el porcentaje de ciudades que experimentan tasas negativas de crecimiento en alguna década es muy bajo, lo que sugiere que su omisión no debería afectar significativamente a ningún resultado.

Los cálculos descritos anteriormente utilizan solamente las cinco ciudades más pobladas en cada década para los diferentes países. Como se hizo en el análisis del coeficiente de asimetría de la distribución de tasas de crecimiento, se procede a analizar si los resultados se mantienen cuando se consideran las 20 ciudades mayores. En este caso, se calcula un índice incluyendo ciudades con tasas de crecimiento negativas y otro que las excluye. Incluir las ciudades con tasas de crecimiento negativas no supone un problema en este caso, ya que nunca ocurre que la tasa de crecimiento de población media de estas ciudades en un país es negativa [11].

La tercera columna de la tabla 6 muestra los resultados de este ejercicio. En 15 de los 21 países (71 por 100) para los que existen datos, el índice ponderado del rango aumenta con el tiempo. Para cuatro países (19 por 100), el índice es estrictamente decreciente

[11] Los resultados son muy parecidos en los dos casos. Para ahorrar espacio, se comentan solamente los correspondientes al índice que considera todas las ciudades.

y, solamente en el 9 por 100 de los casos no hay un patrón claro. Cuando se analizan veinte ciudades, la evidencia en favor de un aumento en el rango (o el rango ponderado) de las ciudades que crecen rápidamente es también muy clara.

B) Evaluación del modelo

Se ha mostrado que, en el modelo, la distribución de las tasas de crecimiento de las ciudades es claramente asimétrica hacia la derecha. Sin embargo, la teoría propuesta en este trabajo tiene un mayor poder de predicción: en cada período, el modelo implica que la ciudad que más crece es la mayor, condicionada al hecho de que los costes de congestión no hayan alcanzado aún un nivel crítico. De acuerdo con el modelo, la pendiente del rango de la ciudad que crece más rápidamente debe ser positiva, puesto que con el tiempo (es decir, a medida que los países se desarrollan), los factores de producción se asignan secuencialmente en las diferentes ciudades, con el orden de secuencialidad determinado por su tamaño inicial. Por lo tanto, se concluye que el modelo reproduce el tercer hecho empírico descrito en este trabajo.

5. OTRAS PROPIEDADES

A) Declive suave de algunas ciudades

La mayor parte de modelos teóricos que estudian la evolución del tamaño de ciudades tienen una propiedad que es radicalmente rechazada por los datos: una vez que una ciudad alcanza su tamaño óptimo, se construye una nueva ciudad, y la población de la primera se reduce a la mitad. Por ejemplo, en modelos donde el capital es perfectamente móvil, como en HENDERSON (1974), el proceso inicial de urbanización se caracteriza por enormes cambios en la población de las ciudades existentes. En estos modelos, a medida que la población aumenta, la primera ciudad crece hasta que alcanza un tamaño crítico. Inmediatamente después de que esto ocurre, una segunda ciudad aparece, y eso provoca un pérdida del 50 por 100 de la población en la primera. A continuación, la primera ciudad empieza a crecer de nuevo en la misma tasa que la segunda, hasta que, en un momento dado, una tercera ciudad aparece. La forma-ción de esta nueva ciudad hace que las dos primeras vean su población reducida en un tercio como resultado de la emigración a la tercera.

Como se comenta en HENDERSON y VENABLES (2004), este proceso es inverosímil, ya que implica, en primer lugar, enormes costes de movimientos de población y, en segundo lugar, en el caso de que exista capital o infraestructura inmóviles, requiere que las ciudades pasen de períodos con enormes cantidades de inversión a períodos con capacidad no utilizada. Estas implicaciones son claramente rechazadas por los datos, como argumentan BLACK y HENDERSON (1999). Al contrario, las ciudades que pierden población, lo hacen de forma muy gradual. En el modelo presentado en este trabajo, el motivo por el que existe un suave declive en la población de las ciudades es la presencia de inmovilidad en el capital instalado en una ciudad. Este mecanismo es parecido al de GLAESER y GYOURKO (2004).

B) La ley de Zipf y el desarrollo económico

Como se discutió en el capítulo II, la ley de Zipf es uno de los hechos empíricos más robustos en la literatura de economía urbana. Esta ley establece que, en un momento dado, la distribución del tamaño de población de las mayores ciudades (las más pobladas) de un país sigue una distribución de Pareto, con parámetro igual a uno. Una forma más gráfica de exponer este hecho es que el tamaño de la mayor ciudad de un país en un momento dado, es dos veces el tamaño de la segunda, tres veces el de la tercera, y así sucesivamente.

El modelo teórico presentado en este trabajo puede generar algunas predicciones respecto a la ley de Zipf. Asúmase por un momento que esta ley aproxima de forma más o menos adecuada la distribución del tamaño óptimo de las ciudades de un país. En ese caso, el modelo implica que las mayores ciudades de los países más desarrollados deberían estar más cercanas a la ley de Zipf que las mayores ciudades de países en vías de desarrollo. El motivo es que los primeros han estado asignando recursos de forma óptima durante un intervalo de tiempo mayor que los últimos (el período de tiempo necesario para que estos países se desarrollaran). Los países en vías de desarrollo han estado asignando la mayor parte de sus recursos nacionales solamente en la mayor ciudad, creando una brecha demasiado grande entre esa ciudad y el resto, lo que implica una violación de la ley de Zipf [12].

Esta sección contrasta esta hipótesis de la siguiente forma. En primer lugar, se presenta una nube de puntos que representan las cinco mayores ciudades de cada país en el año 2000, con el logaritmo de su tamaño inicial en el eje de ordenadas y el logaritmo de su rango en el eje de abscisas. A continuación, se comparan estos puntos con la curva teórica predicha por la ley de Zipf. Más concretamente, se seleccionan las mayores ciudades en cada país en el año 2000 (o el último año en el que existen datos) y se estima la siguiente regresión, a nivel de país:

$$\ln(\text{tamaño}) = \alpha + \beta \ln(rango) + \varepsilon$$

GABAIX (1999a) demuestra que, si la ley de Zipf se cumple exactamente, el parámetro β debe ser igual a -1. Utilizando esta información, es posible generar una "curva perfecta de Zipf" para cada país usando el estimador de mínimos cuadrados ordinarios del parámetro α como punto de referencia. Imponiendo la restricción $\beta = -1$ y ordenando las ciudades en orden decreciente, se obtiene:

$$\ln(\text{tamaño de la ciudad 1}) = \hat{\alpha}$$

$$\ln(\text{tamaño de la ciudad 2}) = \hat{\alpha} - \ln(2)$$

...

$$\ln(\text{tamaño de la ciudad } M) = \hat{\alpha} - \ln(M)$$

donde $M = 5$ [13].

Finalmente, se compara el logaritmo del tamaño verdadero de las ciudades con el que predice la ley de Zipf. Los gráficos 15-16 muestran dos ejemplos usando datos de áreas metropolitanas en diferentes países. Los puntos representados sobre la línea recta son las predicciones de la ley de Zipf, mientras que el resto de puntos

[12] Este supuesto implica que la distribución óptima del tamaño de ciudades es la misma en todos los países. En el contexto de nuestro modelo esto impone idénticas preferencias y tecnologías en todo el mundo.

[13] Resultados muy similares se obtienen si se considera $M = 10$.

representan los datos verdaderos en el año 2000. Tres patrones interesantes merecen ser resaltados:

En primer lugar, en los países más desarrollados, la mayor ciudad tiende a ser “demasiado” pequeña en relación con el resto de ciudades, mientras que en los países en vías de desarrollo, ésta tiende a ser demasiado grande. En segundo lugar, en los países donde la mayor ciudad es demasiado grande, el resto de ciudades tienden a ser pequeñas. Ésta es una implicación de equilibrio general del modelo propuesto anteriormente: las ciudades con mayor rango (es decir, con menor población) tienen que competir con las mayores ciudades para atraer población y capital y, en consecuencia, su crecimiento es muy limitado. Finalmente, el ajuste de los datos con la línea predicha por Zipf es mejor en los países más desarrollados, corroborando la intuición del modelo, esto es, que estos países han invertido durante más tiempo en estas ciudades.

Cuando se usan datos a nivel de ciudad, los resultados son parecidos. La mayor diferencia es que en ese caso, la mayor ciudad es demasiado grande en prácticamente todos los países. Sin embargo, es aún cierto que los países más desarrollados ajustan la ley de Zipf de forma más exacta que los países en vías de desarrollo. Además, también se cumple que, en los países más pobres, la segunda, tercera y cuarta ciudad son típicamente demasiado pequeñas de acuerdo con la ley de Zipf, reflejando el hecho de que deben competir con una ciudad que, en general, tiene un tamaño desproporcionadamente grande.

Para poder visualizar el tercer resultado, es decir, la correlación negativa entre desarrollo económico y distancia con la ley de Zipf en las cinco mayores ciudades, se calcula un estadístico que mide las desviaciones de la población de estas ciudades en los datos respecto de la población que predice la ley de Zipf. Si esta ley se cumple exactamente, se obtiene la siguiente relación entre ciudades:

$$e^{\alpha} = \text{Tamaño de la ciudad } \mathit{1} \text{ predicho por la ley de Zipf}$$

Entonces,

$$\text{Tamaño de la ciudad 2 predicho por la ley de Zipf } = \frac{1}{2}e^{\alpha}$$

...

$$\text{Tamaño de la ciudad } M \text{ predicho por la ley de Zipf } = \frac{1}{M}e^{\alpha}$$

Estas relaciones pueden usarse para construir la medida de dispersión del tamaño de ciudades. En concreto, se calculan la desviación absoluta y cuadrática de los datos respecto de la ley de Zipf. Finalmente, se presenta la correlación entre el logaritmo del PIB per capita de cada país en el año 2000 y la desviación de este país respecto de la ley de Zipf. Con las dos medidas de dispersión, calculadas tanto a nivel de área metropolitana como de ciudad, esta correlación es significativamente negativa, como se puede observar en el gráfico 17. El coeficiente de correlación es $\rho = -0{,}4$ ($\rho = -0{,}35$ con los datos de ciudades), y es significativo al 5 por 100.

La discusión de este capítulo ha asumido que la distribución óptima del tamaño de las ciudades de un país está representada por su ley de Zipf. Si este supuesto es correcto, cabría esperar que, a medida que los países se desarrollan, sus mayores ciudades convergen a su particular ley de Zipf. En CUBERES (2005) se presentan gráficos que ilustran la evolución de esta dispersión en tres distintos momentos de la historia de Estados Unidos y otros países. En el caso estadounidense se puede observar que la mayor ciudad (Nueva York) tiende a convergir hacia la ley de Zipf. Resultados parecidos se obtienen en otros países. Sin embargo, los resultados de convergencia a la ley de Zipf a lo largo del tiempo son considerablemente más débiles que los de convergencia en la sección cruzada de países. Por lo tanto, es necesario estudiar este fenómeno con más detalle usando estimadores y métodos más adecuados para concluir de forma no ambigua que el supuesto de que la ley de Zipf representa el tamaño óptimo de las ciudades de un país es válido.