

EL AUGE Y DECLIVE
DE LAS CIUDADES

DAVID CUBERES

**EL AUGE Y DECLIVE
DE LAS CIUDADES**

FUNDACIÓN RAFAEL DEL PINO

MARCIAL PONS, EDICIONES JURÍDICAS Y SOCIALES, S. A.
MADRID 2006 BARCELONA

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del «Copyright», bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo públicos.

© David Cuberes

© Fundación Rafael del Pino

© MARCIAL PONS

EDICIONES JURÍDICAS Y SOCIALES, S. A.

San Sotero, 6 - 28037 MADRID

☎ (91) 304 33 03

ISBN: 84-9768-334-X

ISBN-13: 9788497683340

Depósito legal: M-46563-2006

Diseño de la cubierta: Álvaro Reyero

Fotocomposición: JOSUR TRATAMIENTO DE TEXTOS, S. L.

Impresión: ELECÉ, INDUSTRIA GRÁFICA, S. L.

Polígono El Nogal

Río Tiétar, 24 - 28110 Algete (Madrid)

MADRID, 2006

Para Montse y Gurutze

ÍNDICE

	<u>Págs.</u>
PRESENTACIÓN	11
AGRADECIMIENTOS	13
NOTA DEL AUTOR	15
CAPÍTULO I. INTRODUCCIÓN	19
CAPÍTULO II. LITERATURA RELACIONADA	23
CAPÍTULO III. EL MODELO TEÓRICO	27
1. LA SOLUCIÓN CON MERCADOS COMPETITIVOS	28
A) Familias	28
B) Empresas	29
C) Equilibrio	31
CAPÍTULO IV. IMPLICACIONES EMPÍRICAS	35
1. DATOS	35
2. EL AUGE Y DECLIVE DE LAS CIUDADES	38
A) Descripción del hecho empírico	38
B) Evaluación del modelo	43
3. ASIMETRÍA EN LAS TASAS DE CRECIMIENTO DE LAS CIUDADES	44
A) Descripción del hecho empírico	44
B) Evaluación del modelo	47

	<u>Págs.</u>
4. EL RANGO DE LAS CIUDADES QUE CRECEN MÁS RÁPIDO	48
A) Descripción del hecho empírico	48
B) Evaluación del modelo	50
5. OTRAS PROPIEDADES.....	50
A) Declive suave de algunas ciudades	50
B) La ley de Zipf y el desarrollo económico	51
CAPÍTULO V. CONCLUSIONES	55
APÉNDICES	57
BIBLIOGRAFÍA	99

PRESENTACIÓN

Con el convencimiento de que una sociedad culta y preparada, no sólo será eficiente, sino que, a la vez, estará formada por ciudadanos solidarios y exigentes consigo mismos y con los demás, la **Fundación Rafael del Pino** tiene como principal objetivo contribuir a la mejora de los conocimientos de los dirigentes de la España del futuro.

Con esta clara vocación formativa, nacieron en 2001 las primeras becas de la Fundación dirigidas a aquellos estudiantes y profesionales españoles que, como David CUBERES, destacan por su actitud de liderazgo y por la excelencia en su actividad profesional y académica, con el objetivo último de que puedan desarrollar con todo éxito sus capacidades y transformar el conocimiento en riqueza.

En su momento, la Fundación consideró que la elección del tema principal al que David habría de dedicar sus esfuerzos durante sus estudios de postgrado, desarrollados en la Universidad de Chicago, era muy acertada. Su inquietud intelectual le llevó a plantearse la aplicación del análisis económico al estudio del auge y declive de las ciudades, en línea con la metodología de trabajo que iniciara el Premio Nobel de Economía, Gary BECKER, que visitó recientemente la Fundación. Su elección suponía, por una parte, abordar un tema que, en un contexto de intensos movimientos migratorios, alcanza una relevancia notable. Por otra parte, implicaba

aceptar el reto de aplicar el análisis económico para comprender mejor las fuerzas que dirigen el crecimiento de las ciudades.

Alejado de las enciclopédicas tesis doctorales, tan habituales como inútiles, el trabajo riguroso y pionero de David CUBERES tuvo como resultado su tesis doctoral titulada *The Rise and Decline of Cities*, defendida brillantemente en el Departamento de Economía de dicha Universidad.

La calidad de la tesis y la originalidad de los resultados alcanzados motivaron que la Fundación decidiera apoyar su publicación dentro de su Colección Economía y Empresa. El libro describe «la evolución del tamaño de las ciudades en diferentes países a lo largo del tiempo», estilizando las características generales del proceso. Conceptos como desarrollo y dinamismo urbano, tamaño óptimo de las ciudades, convergencia, costes de congestión, etc. se suceden en esta investigación que ofrece al lector un análisis profundo del dinamismo de las sociedades urbanas. Además, el análisis de David CUBERES, reforzado con una potente base teórica y con las leyes de Gibrat y Zipf en el horizonte, no limita su alcance a la reflexión teórica. Es también una vía para la acción, como demuestran las siguientes palabras: «entender la formación de las ciudades y las dinámicas de su crecimiento es crucial para diseñar políticas en países cuya población está aumentando rápidamente», como ocurre en España en estos momentos. Particular relevancia tiene el hecho de considerar las ciudades como empresas y someter el análisis a los robustos cimientos de la teoría macroeconómica.

Pocas acciones pueden tener mayor implicación con la vocación de la Fundación de contribuir a mejorar la formación de los futuros dirigentes españoles como su programa de becas de posgrado, uno de cuyos frutos es el presente libro.

Amadeo PETITBÒ
Director de la Fundación Rafael del Pino

AGRADECIMIENTOS

El 10 de junio de 2005 leía mi tesis doctoral *The Rise and Decline of Cities* en el departamento de economía de la Universidad de Chicago. Un año después, la Fundación Rafael del Pino ha hecho posible que mi trabajo de investigación pueda llegar a más personas a través de este libro.

Es difícil nombrar a todos los que de una u otra forma, han contribuido a este proyecto. El primero de la lista es Casey MULLIGAN, quien me apoyó desde el primer día en que entré en su despacho con un simple gráfico. Los datos que le mostré indicaban que las ciudades no pueden crecer indefinidamente, y que debía ser posible encontrar un mecanismo económico que racionalizara esta evidencia. Con sus incesantes críticas, Casey consiguió que mi idea se convirtiera en un texto coherente. Durante nuestras frecuentes charlas, me enseñó a pensar como un economista y me transmitió su pasión por el trabajo bien hecho. También me convenció de que la mejor forma de tratar un problema económico, tal vez la única, es combinando cuidadosamente la evidencia empírica con la modelización teórica.

Durante la elaboración de mi tesis, me beneficié enormemente de muchas conversaciones con Gary BECKER. En cada ocasión, sus comentarios hicieron que volviera a la esencia del problema económico en cuestión. Esta es, al fin y al cabo, la forma en que Gary encuadra cualquier tema que esté relacionado con la economía.

Quiero agradecer los comentarios de Roc ARMENTER, Pau CASAS-ARCE, Jan EECKHOUT, Xavier GABAIX, François GOURIO, Vernon HENDERSON, Yannis IOANNIDES, Aitor LACUESTA-GABARAIN, Robert LUCAS, Diego PUGA, Nancy STOKEY, Chad SYVERSON, Balazs SZENTES, Robert TAMURA, Robert TOWNSED, y participantes en seminarios de la Universidad de Chicago, Clemson University, Universidad de Alicante, Banco de España, Universidad de Navarra, Universidad del País Vasco, Universidad Autónoma de Barcelona, y Universidad Pompeu Fabra. Pep VILARRUBIA sugirió el uso de la principal base de datos en la parte empírica del libro.

Durante el desarrollo de este proyecto fue indispensable la ayuda financiera de la Fundación Rafael del Pino y el incondicional apoyo de su director, Amadeo PETITBÒ. También agradezco la ayuda de Beatriz LOBATÓN en la transformación de la tesis doctoral en este libro. Raquel NIETO hizo un excelente trabajo de corrección y edición del manuscrito original.

Estoy en deuda con mi familia que ha sabido apoyarme en todo momento, a pesar de la distancia. Un especial agradecimiento merece Aitor LACUESTA-GABARAIN, con el que tanto hemos compartido estos últimos años.

Finalmente, a Gurutze, por estar siempre ahí, especialmente en los momentos en que el progreso era lento.

NOTA DEL AUTOR

Como macroeconomista aplicado, existen dos cuestiones económicas que considero primordiales por su relevancia y rabiosa actualidad. En esta introducción quisiera comentar brevemente la evolución del pensamiento económico moderno respecto a estos dos temas y cuál es la contribución de mi libro a esta literatura.

La primera cuestión que me interesa destacar es la inmigración internacional. En países como Estados Unidos, el espectacular aumento del número de inmigrantes (en su mayoría provenientes de Latinoamérica) durante los últimos años, ha hecho que tanto los medios de comunicación como los estamentos políticos de ese país, hayan situado el problema de la movilidad del trabajo en el primer lugar de su lista de prioridades. La presión migratoria no es menor en Europa, donde el flujo de inmigrantes del este de Europa y el norte de África está también aumentando a gran velocidad. Entender las causas y el impacto de este enorme movimiento de personas en la economía de los países afectados es una tarea compleja que ha ocupado una enorme parte de la literatura moderna en economía. Para focalizar un poco la discusión, quiero comentar aquí los estudios que han analizado cuál es el impacto de la inmigración en la economía de destino de los inmigrantes. Como en muchas otras cuestiones económicas, no existe actualmente un consenso sobre esta cuestión. Centrándose en la inmigración de México a Estados Unidos, George BORJAS, de Harvard University, concluye

que ésta tiene un impacto negativo en la economía estadounidense, principalmente debido a que disminuye el salario de los ciudadanos con menor educación en este país. Por otro lado, David CARD, de University of California, Berkeley, muestra que no existe evidencia que las ciudades que han recibido un mayor flujo de inmigrantes mejicanos, como por ejemplo, Los Ángeles, experimente una tendencia salarial de la población con menor educación inferior a la de ciudades con poca entrada de inmigración, como Indianapolis.

El segundo problema macroeconómico que quiero destacar es, si cabe, más importante que el anterior. Se trata de tratar de explicar cómo es posible que exista una disparidad tan grande entre los ingresos per capita de los diferentes países del mundo. Como comenta Daron ACEMOGLU, del Massachusetts Institute of Technology, la dispersión en la riqueza de los países ha aumentado muchísimo desde que el famoso economista Adam SMITH se preocupó por estas diferencias, hace ya más de doscientos años. Se podría decir que fue el premio Nóbel Robert SOLOW, del Massachusetts Institute of Technology, quien, en 1956, inició la literatura moderna sobre el crecimiento económico. Su modelo y muchos otros que le siguieron destacan el importante papel de factores acumulables como capital físico y humano para fomentar el crecimiento. De acuerdo con estas teorías, la principal receta para que un país subdesarrollado aumente su tasa de crecimiento es incrementar la inversión en estos dos tipos de capital. Casi cuatro décadas más tarde, el también premio Nóbel Robert LUCAS de la University of Chicago hizo la siguiente observación: si asumimos que existen rendimientos decrecientes del capital, es difícil entender porqué éste no se desplaza de los países ricos a los pobres, donde debería ser, por definición, más productivo. Su explicación es que, en los países más desarrollados, existen externalidades positivas del capital humano agregado (o medio) hacia el capital físico, haciendo que este último sea más productivo que en economías con menor nivel educativo. Sin embargo, unos años más tarde, Robert HALL y Charles JONES de Stanford University y University of California, Berkeley, respectivamente, mostraron que los diferentes niveles de capital físico y humano no son suficientes para explicar la gran variación de ingresos entre países. Esto dio pie a una fructífera literatura que destacó la importancia del progreso tecnoló-

gico y construyó elaborados modelos en los que la investigación y el desarrollo eran las claves para el crecimiento. En los últimos años, ha ganado protagonismo una explicación alternativa a la disparidad de ingresos per capita en el mundo. La idea es que los países menos desarrollados no tienen las instituciones adecuadas para promover el crecimiento económico y que las tradicionales recetas no funcionarán a menos que las estas instituciones sean reformadas en la dirección correcta.

Los dos problemas comentados anteriormente han sido estudiados desde un punto de vista empírico usando una gran variedad de datos. Gracias al trabajo de Angus MADDISON, Scott BAIER, Robert TAMURA o Gerald DWYER entre muchos otros, hoy en día es posible contrastar muchas de las teorías de crecimiento económico utilizando valiosísimas series temporales que se remontan a más de mil años atrás. Sin embargo, la literatura sobre inmigración se ha limitado a series temporales más cortas debido a la dificultad que representa construir datos coherentes de inmigración para muchos años.

Este libro puede entenderse como una contribución a los dos problemas macroeconómicos comentados anteriormente. El primer punto en común es teórico. La cuestión que intento resolver en mi estudio es cómo crece la población y el capital instalado en las ciudades de un país, en un contexto en el que existen pequeñas diferencias iniciales entre las diferentes localidades. Estas ciudades pueden ser fácilmente interpretadas como países, en cuyo caso partimos de una situación inicial parecida a la descrita por Adam SMITH, en la que las diferencias entre países eran muy pequeñas. El modelo teórico que propongo está basado en la literatura de crecimiento que he resumido más arriba, principalmente en la idea de Paul ROMER, de Stanford University, de introducir externalidades positivas en la función de producción de cada ciudad. Mi modelo introduce algunas variaciones respecto de la literatura estándar de crecimiento para adaptarlo a un contexto urbano. La principal novedad es la presencia de costes de congestión en la acumulación de capital en una localidad.

Los datos que presento muestran cómo la desigualdad en población y capital entre ciudades aumenta con el tiempo hasta alcanzar un nivel máximo y disminuye posteriormente. Mi modelo teó-

rico explica razonablemente bien esta evolución del trabajo y el capital entre las ciudades de un país, pero no encaja cuando se aplica a países en vez de ciudades, puesto que la desigualdad en el ingreso en el mundo ha aumentado claramente en los últimos doscientos años. En este trabajo, el énfasis está puesto en entender la distribución geográfica de recursos dentro de un país, a medida que éste se desarrolla. La cuestión de cómo esta distribución geográfica de recursos afecta a su vez al desarrollo económico del país no queda resuelta y debe ser estudiada con más detenimiento en futura investigación.

La segunda conexión entre los estudios mencionados anteriormente y el presente libro es la metodología empírica que se ha utilizado. En concreto, mi estudio contribuye a expandir la literatura que utiliza largas series temporales (o paneles) de datos. En este sentido, es más cercano a la literatura de crecimiento que, como ya se ha subrayado, ha experimentado un espectacular aumento en el número de estudios que se basan en datos históricos. Mi trabajo sigue esta práctica haciendo uso de dos completísimas bases de datos que son la de Jan LAHMEYER y Thomas BRINKHOFF. Combinando estas dos fuentes de información, es posible recopilar información enormemente valiosa sobre la población de las mayores ciudades en todos los países del mundo, potencialmente para el periodo 1800-2000.

Creo firmemente que la investigación que trata de responder preguntas concretas sobre el desarrollo y crecimiento económico puede tener un fuerte impacto en el bienestar de muchas personas, especialmente en países en vías de desarrollo. A mi modo de ver, la profesión está necesitada de estudios que combinen modelos teóricos con datos (a ser posible largas series temporales o datos de panel) para entender los grandes problemas macroeconómicos del mundo en el que vivimos y tratar, en la medida de lo posible, de encontrar políticas económicas que puedan solucionarlos.

El autor

CAPÍTULO I

INTRODUCCIÓN

*David Cuberes**

The University of Chicago

La evolución del tamaño de las ciudades en la mayoría de los países verifica tres hechos empíricos inexplorados. En primer lugar, a medida que los países se desarrollan, la concentración de población en sus mayores ciudades crece y, tras alcanzar un nivel máximo, a continuación decrece. Los gráficos 1 y 2 del apéndice D muestran este patrón en los Estados Unidos y Francia, respectivamente.

* Quiero agradecer a Casey MULLIGAN su apoyo durante la realización de este proyecto. Estoy también en deuda con Gary BECKER y Chad SYVERSON por muchas productivas discusiones. Agradezco los comentarios de Roc ARMENTER, Scott BAIER, Jan ECKHOUT, Oded GALOR, Edward GLAESER, François GOURIO, Vernon HENDERSON, Michal JERZMANOWSKI, Aitor LACUESTA-GABARAIN, Diego PUGA, Curtis SIMON, Robert TAMURA, Marc TEIGNIER, dos revisores anónimos, y participantes en seminarios en the University of Chicago, Clemson University, Universidad de Alicante, Banco de España, Universidad de Navarra, Universidad del País Vasco, Universitat Autònoma de Barcelona, Universitat Pompeu Fabra, y el North American Regional Science Council 2005 meeting. Este proyecto no habría podido llevarse a cabo sin la ayuda financiera de la Fundación Rafael del Pino y la Esther and T. W. Schultz Endowment dissertation fellowship. Todos los errores son de mi responsabilidad. E-mail: cuberes@clemson.edu.

El segundo hecho estilizado es que la distribución de las tasas de crecimiento de las ciudades de un país es, en un momento dado, asimétrica hacia la derecha. En otras palabras, el proceso de crecimiento de las ciudades está caracterizado por la existencia de unas pocas ciudades que crecen mucho más deprisa que el resto. La última de las regularidades empíricas es que el rango de la ciudad que crece más deprisa en un país determinado aumenta a medida que pasa el tiempo.

Estos tres hechos son difícilmente reconciliables con la ley de Gibrat¹, que sostiene que los procesos de crecimiento de las ciudades siguen un paseo aleatorio. Esta ley ha sido la base de la mayor parte de la literatura empírica y teórica sobre ciudades, que se resume en el siguiente capítulo. Después de presentar estas regularidades, el trabajo desarrolla un modelo teórico que difiere significativamente de la lógica de la ley de Gibrat y es capaz de explicar estos hechos.

Entender la formación de ciudades y las dinámicas de su crecimiento es crucial para diseñar políticas en países cuya población está aumentando rápidamente. Los nuevos hallazgos empíricos que se presentan en este trabajo sugieren que estos patrones de crecimiento varían enormemente a lo largo del tiempo y entre diferentes localidades. Esto tiene importantes implicaciones en países en vías de desarrollo. Por ejemplo, la ayuda externa hacia estos países tenderá a localizarse en unas pocas ciudades durante un período determinado de tiempo, incluso si ésta se distribuye inicialmente de forma uniforme en el país. Esta reasignación de recursos puede de hecho aumentar la desigualdad en el país en cuestión durante el período de transición de sus ciudades hacia su estado estacionario. De forma más general, este trabajo permite analizar los diferentes patrones de urbanización en países en vías de desarrollo que están experimentando un importante proceso de emigración rural hacia las ciudades. Por otro lado, otra aplicación de este trabajo es la posibilidad de estudiar cómo los factores de producción (trabajo y capital) se movilizan en regiones que han experimentado un proceso de integración económica y política. El ejemplo más claro es el de la Unión Europea, que reciente-

¹ Véase GIBRAT (1931) para la formulación original de esta ley y SUTTON (1997) para algunas de sus aplicaciones en economía.

mente ha aceptado la entrada de diez nuevos países y está negociando la incorporación de algunos más en los próximos años. Entender esta reasignación de recursos productivos entre los diferentes países miembros de la Unión es extremadamente importante tanto para académicos como para políticos. Finalmente, analizar cómo las ciudades se organizan y evolucionan a lo largo del tiempo es necesario para predecir los cambios geográficos y económicos en regiones que han sufrido desastres naturales o conflictos bélicos. Este trabajo predice que, tanto en el caso de una unión económica y política como en el de desastres o guerras, las ciudades más pobladas en el momento del cambio deberían ser las que atrajesen la mayor parte de los recursos durante algún tiempo, hasta que sus costes de congestión alcanzasen un nivel crítico.

En el capítulo tercero del trabajo se presenta un modelo de crecimiento de equilibrio general con dos ciudades, modelizadas como empresas con tecnología Cobb-Douglas. Estas empresas tienen rendimientos crecientes de escala, pero rendimientos decrecientes respecto a cada factor de producción: capital y trabajo. Por otro lado, existen costes de congestión asociados a los stocks de capital instalados en cada ciudad. Se asume que la inversión en una ciudad es irreversible. En el modelo, los rendimientos crecientes son la fuerza que favorece la concentración de recursos, mientras que los costes de congestión la limitan. Por razones de espacio, sólo se resuelve el problema de los mercados competitivos en el texto principal. La versión equivalente para un planificador social benevolente se presenta en el apéndice B. La principal diferencia entre los dos escenarios es que la presencia de una externalidad positiva del capital urbano medio sobre las empresas que operan en ella hace que el equilibrio competitivo sea subóptimo.

El modelo es capaz de reproducir los tres hechos empíricos mencionados con anterioridad. La ciudad inicialmente mayor crece sola durante algunos períodos, generando así la parte creciente de la curva con forma de U invertida que representa la evolución de los ratios de población en las mayores ciudades a lo largo del tiempo². Eventualmente, los costes de congestión se hacen demasiado grandes en dicha ciudad, y comienza la inversión en la segunda

² De ahora en adelante se referirá a esta curva como la “U invertida”.

ciudad más grande. Dado que estos costes son muy bajos en la segunda ciudad, su tasa de inversión es estrictamente mayor que la de la ciudad de mayor tamaño, generando así la parte descendiente de la U invertida. Dependiendo de los valores asignados a los parámetros del modelo, la ciudad inicialmente más pequeña puede alcanzar a la mayor en términos de capital y trabajo, en cuyo caso, las dos ciudades crecen en la misma tasa durante algunos períodos. Otra predicción del modelo es que, en cada momento del tiempo, una ciudad crece mucho más deprisa que el resto, generando así el segundo nuevo hecho empírico, es decir, una distribución de tasas de crecimiento que es asimétrica hacia la derecha. Finalmente, en el modelo, la ciudad con una tasa de crecimiento mayor en un momento dado es la más poblada³, condicionado al hecho de que los costes de congestión aún no hayan alcanzado un nivel crítico. Por lo tanto, en el modelo, el rango de la ciudad que crece más rápido aumenta con el tiempo, tal y como sucede en los datos.

Este trabajo está organizado como se indica a continuación. En el capítulo II se resume brevemente la literatura existente sobre la distribución del tamaño de las ciudades y su crecimiento. El capítulo III describe el modelo teórico. El capítulo IV comenta la base de datos que se utiliza en el estudio y los potenciales problemas de selección muestral, así como presenta los principales hallazgos empíricos, y analiza cómo el modelo teórico puede explicarlos. Finalmente, el capítulo V concluye el estudio.

³ A lo largo del texto se refiere a la ciudad más poblada como la *ciudad mayor*. Igualmente, cuando se habla del tamaño de una ciudad, se refiere a su población, no a su extensión geográfica.

CAPÍTULO II

LITERATURA RELACIONADA

La literatura existente sobre la evolución de la distribución del tamaño de las ciudades se ha centrado principalmente en analizar propiedades del estado estacionario de éstas, prestando poca atención a su transición hacia dicho estado. Una característica general de la literatura empírica es que infravalora los serios problemas de selección muestral en los datos. Algunos de estos trabajos empíricos y teóricos se resumen seguidamente ¹.

Existen varios trabajos que estudian el crecimiento de ciudades y lo relacionan con la ley de Gibrat. HENDERSON y WANG (2003) analizan el proceso de urbanización ² en un conjunto de países durante el período 1960-2000. Estos autores presentan evidencia de ausencia de raíces unitarias ³ en el proceso de crecimiento de las ciudades en estos países, contradiciendo directamente la ley de Gibrat. IOANNIDES y OVERMAN (2003) muestran

¹ Para un exhaustivo resumen de la literatura sobre la distribución del tamaño de las ciudades, véase GABAIX y IOANNIDES (2004).

² En CUBERES (2005) se muestra que, si bien la inmigración rural-urbana es un fenómeno clave en la mayor parte de los países durante el período de tiempo analizado, este hecho no es capaz de explicar, por sí mismo, el patrón de U invertida. Un segundo proceso, el de competencia entre ciudades para atraer recursos (o primacía urbana), tiene también lugar.

³ La ley de Gibrat puede ser formulada como la presencia de una raíz unitaria en el proceso de crecimiento de las ciudades, es decir $\alpha = 1$ en la ecuación $\ln S_{it} = \alpha \ln S_{it-1} + \varepsilon_{it}$, donde S_{it} es el tamaño de la ciudad i en el período t y ε_{it} es un ruido blanco. En CUBERES (2005) se presentan diferentes estimadores del parámetro α para los Estados Unidos, confirmando la ausencia de raíces unitarias.

que las desviaciones de la ley de Gibrat no son estadísticamente significativas para la mayor parte de las áreas metropolitanas de Estados Unidos durante el período 1900-1990. ECKHOUT (2004) muestra una nube de puntos con la tasa de crecimiento de las ciudades en el período 1990-2002 y su tamaño en 1990, donde no se observa ninguna correlación entre las dos variables, lo que sugiere que la ley de Gibrat puede ser una buena descripción de los datos durante este intervalo de tiempo. EATON y ECKSTEIN (1997) argumentan que las poblaciones relativas de las 40 mayores áreas metropolitanas de Francia y Japón permanecieron constantes durante el intervalo de tiempo 1876-1990 para Francia y 1925-1985 en el caso de Japón, y sostienen que este hecho parece también verificarlo la ley de Gibrat. Otros trabajos que analizan el crecimiento del tamaño de las ciudades (de su población) sin relacionarlo con la ley de Gibrat son BLACK y HENDERSON (1999), DOBKINS y IOANNIDES (1998a), GLAESER, SCHEINKMAN y SHLEIFER (1995), GLAESER y SHAPIRO (2001), y CARLINO y CHATTERJEE (2001). El presente trabajo no contrasta directamente la hipótesis de la ley de Gibrat, pero presenta nuevas regularidades empíricas que cuestionan su validez.

Una diferente línea de investigación se centra en estudiar mecanismos que pueden generar la famosa ley de Zipf⁴. GABAIX (1999a) es el primero en demostrar que la ley de Gibrat implica la ley de Zipf. ROSSI-HANSBERG y WRIGHT (2003) presentan un modelo más elaborado en el que esta relación también se verifica. CÓRDOBA (2003) estudia esta causalidad y concluye que la ley de Gibrat no solamente es suficiente sino que también es necesaria para que se verifique la ley de Zipf. Otros autores han explorado explicaciones alternativas para la ley de Zipf. Algunos de los trabajos más relevantes en esta dirección son DURANTON (2002), AXTEL y FLORIDA (2001), FUJITA *et al.* (1999) y KRUGMAN (1996b).

Los primeros estudios que desarrollan la idea de un tamaño óptimo de ciudad son HENDERSON (1974, 1988). La principal diferencia con el planteamiento del presente trabajo es que en éste se

⁴ Véase ZIPF (1949), ROSEN y RESNICK (1981) y SOO (2003). Esta ley describe un hecho estilizado en economía urbana: el tamaño de la mayor ciudad de un país es, aproximadamente, el doble del de la segunda mayor ciudad, tres veces el de la tercera, etcétera.

modeliza la dinámica de las ciudades, mientras que el análisis de HENDERSON es estático.

La literatura existente relacionada con los tres nuevos hechos empíricos es escasa. El patrón de U invertida en la concentración de población ha sido brevemente documentado por WILLIAMSON (1965), HANSEN (1990) y otros autores. Sin embargo, la baja calidad de los datos que utilizan y su dudosa estrategia empírica hacen que las conclusiones a las que llegan sean difíciles de interpretar. HENDERSON (2004) revisa diferentes estudios sobre primacía urbana y su relación con el desarrollo económico, encontrando un patrón de U invertida como función del PIB de los países. El hecho de que la distribución de las tasas de crecimiento de las ciudades es asimétrica hacia la derecha se menciona en GLAESER y GYOURKO (2005). Sin embargo, su estudio sólo analiza el caso de Estados Unidos para el período 1920-2000. El trabajo que se presenta aquí expande su análisis a períodos de tiempo más largos y a un mayor número de países. Por otro lado, el objetivo de GLAESER y GYOURKO es mostrar que las ciudades crecen más deprisa de lo que decrecen, debido a la baja tasa de depreciación de la infraestructura urbana. En el caso del presente trabajo, se muestra que, entre las ciudades que crecen en una tasa positiva, unas pocas lo hacen mucho más rápido que el resto. Finalmente, la secuencialidad en el crecimiento de ciudades ha sido formalizada en KRUGMAN y BREZIS (1997) y en HENDERSON y VENABLES (2004). Este último trabajo es probablemente el más cercano al presentado aquí. Las principales diferencias entre los dos estudios son que, en primer lugar, HENDERSON y VENABLES no presentan evidencia empírica de esta secuencialidad. En segundo lugar, desde un punto de vista teórico, no demuestran formalmente esta característica del proceso de formación de ciudades. En mi trabajo, se presenta evidencia empírica y una demostración formal de la secuencialidad en el crecimiento de ciudades. En su escenario, que asume crecimiento positivo de la población urbana, es posible obtener predicciones sobre rentas de la tierra y el impacto de diferentes políticas gubernamentales, ausentes en mi trabajo. Una importante contribución de ambos estudios es que, al asumir irreversibilidad en la inversión, se eliminan los cambios bruscos en la población de las ciudades una vez que éstas han alcanzado su nivel óptimo, presentes en la mayor parte de la anterior literatura.

CAPÍTULO III

EL MODELO TEÓRICO

Esta sección desarrolla un marco teórico que reproduce los tres nuevos hechos empíricos sobre ciudades. Por simplicidad, el modelo consiste solamente en dos ciudades que son modelizadas como empresas con idénticas funciones de producción Cobb-Douglas¹. En cada ciudad, la producción del único bien de consumo tiene lugar usando dos factores de producción: capital y trabajo. Esta tecnología presenta rendimientos decrecientes en cada factor de producción, pero rendimientos crecientes de escala. Por otro lado, existen costes convexos de producción asociados con la cantidad de capital instalado en una misma ciudad. Estos costes de congestión son la fuerza que limita la concentración de recursos en una ciudad. Se asume que, en el período inicial, una de las dos ciudades tiene un stock de capital instalado ligeramente superior al de la otra ciudad. La inversión en el modelo es irreversible: una vez que una cantidad dada de capital está instalada en una ciudad, no es posible destruirla o consumirla, ni tampoco es posible reasignar dicho capital a otra ciudad². Por otro lado, el trabajo puede emigrar entre ciudades sin ningún tipo de coste.

Bajo este escenario, si los costes de congestión para bajos niveles de capital son suficientemente reducidos, los recursos se desplazan de forma desproporcionada hacia la ciudad con un stock inicial de capital mayor, hasta que estos costes convierten la ciudad

¹ En el apéndice C se presenta una extensión del modelo para el caso de $J > 2$ ciudades. Todos los resultados se mantienen en este caso más general.

² Éste es un supuesto razonable si se interpreta el capital físico como infraestructura (edificios, etc.) instalados en una ciudad.

inicialmente menor en un lugar más rentable para invertir. En ese momento, puesto que los costes de congestión son menores en la ciudad más pequeña, la tasa de inversión se vuelve estrictamente mayor en esa ciudad que en la de mayor tamaño, hasta que el stock de capital en las dos se iguala. Dependiendo de los parámetros del modelo, este escenario de convergencia puede no llegar a completarse totalmente. En esta parte del trabajo solamente se resuelve el problema de los mercados competitivos. El apéndice B desarrolla el correspondiente problema del planificador social benevolente.

1. LA SOLUCIÓN CON MERCADOS COMPETITIVOS

En primer lugar, se procede a describir el comportamiento de las familias y empresas en la economía y, seguidamente, se calcula su equilibrio competitivo.

A) Familias

Las fuentes de ingreso de las familias son las ganancias salariales y el retorno de los activos en los que éstas invierten. El salario competitivo es ω . Defínase Z^j como la cantidad de activos invertidos en la ciudad j , $j = A, B$. El ingreso proveniente del capital para una familia que invierte una cantidad positiva en cada uno de los dos activos viene dado $\sum_{j=A,B} r^j Z^j$, donde r^j representa el retorno bruto del capital en la ciudad j . Estos recursos pueden ser usados para consumir o invertir en activos. La restricción presupuestaria de una familia representativa es pues

$$\sum_{j=A,B} \dot{z}^j + c = \omega + \sum_{j=A,B} r^j z^j$$

donde $z^j \equiv \frac{Z^j}{N}$ y $c^j \equiv \frac{C}{N}$ representan el stock del activo j , $j = A, B$ y el consumo per cápita, respectivamente. Asumiendo utilidad logarítmica, el problema de maximización de la familia representativa es (omitiendo subíndices de tiempo):

$$\begin{aligned}
& \max \int_0^{\infty} e^{-\rho t} \ln(c) dt \\
& \sum_{j=A,B} \dot{z}^j + c = \omega + \sum_{j=A,B} r^j z^j \\
& \dot{z}^j \geq 0, \forall j = A, B \\
& z_0^j, \forall j = A, B \text{ dados}
\end{aligned}$$

donde, como se comentaba anteriormente, las familias tienen una restricción de irreversibilidad en la inversión, es decir $\dot{z}^j \geq 0$, $j = A, B$.

B) Empresas

Cada empresa dispone de una tecnología que presenta rendimientos constantes de escala. Esto garantiza que sus beneficios son nulos y que, sin pérdida de generalidad, se puede analizar el problema de una empresa representativa en cada ciudad. Esta empresa está sujeta a una externalidad positiva similar a la de ROMER (1986) que proviene del stock de capital medio instalado en la ciudad en la que opera. La función de producción de la empresa i localizada en la ciudad j viene dada por

$$Y^{ij} = (N^{ij})^\alpha (K^{ij})^{1-\alpha} (\tilde{K}^j)^\psi$$

donde Y^{ij} , N^{ij} , y \tilde{K}^j respectivamente representan producción, trabajo y capital. El parámetro ψ es estrictamente positivo y \tilde{K}^j representa el capital medio instalado en la ciudad j , esto es $\tilde{K}^j \equiv \frac{1}{M} \sum_{m=1}^M K^{mj}$, donde M es un número grande que representa el número de empresas que operan en la ciudad j . Como en el modelo de ROMER, el supuesto de que las empresas toman K^{ij} como dado, puede ser racionalizado por un modelo con un continuo de individuos³.

Las empresas contratan trabajo y capital a sus precios competitivos y venden su producto en el mercado.

³ LUCAS (1988) y TAMURA (1991) destacan la importancia del capital humano medio para el crecimiento de la economía.

Por otro lado, tienen que pagar su contribución al coste de congestión⁴ en la ciudad donde están localizadas. Se asume la forma funcional $g(K^j) = (K^j)^\sigma$, donde $K^j \equiv \sum_{m=1}^M K^{mj}$, M es el número de empresas que operan en la ciudad j , y $\sigma > 1$. Normalizando el precio del bien de consumo a uno, los beneficios de la empresa i , localizada en la ciudad j vienen dados por

$$\pi^{ij} = (N^{ij})^\alpha (K^{ij})^{1-\alpha} (\tilde{K}^j)^\psi - (r^j + \delta)K^{ij} - \frac{1}{M} g(K^j) - \omega N^{ij}$$

donde $\delta \in (0,1)$ es la tasa de depreciación del capital.

Las condiciones de primer orden de este problema son:

$$\frac{\partial \pi^{ij}}{\partial N^{ij}} = 0 \Leftrightarrow \alpha (N^{ij})^{\alpha-1} (K^{ij})^{1-\alpha+\psi} = \omega \quad (1)$$

$$\frac{\partial \pi^{ij}}{\partial K^{ij}} = 0 \Leftrightarrow (1-\alpha)(N^{ij})^\alpha (K^{ij})^{\psi-\alpha} = r^j + \delta + g'(K^{ij}) \quad (2)$$

donde se ha impuesto la condición de equilibrio simétrico. Puesto que los trabajadores pueden cambiar de empresa y ciudad sin incurrir en ningún coste, el salario es, en equilibrio, el mismo en todas las empresas y ciudades. Esto implica que la asignación de trabajo se puede resumir con la siguiente ecuación:

$$\frac{N^A}{N^B} = \left(\frac{k^A}{k^B} \right)^{\frac{\beta}{1-\alpha}}$$

donde $k^j \equiv \frac{K^j}{N}$, $j = A, B$ representa el capital per cápita en la ciudad j .

Esto implica a su vez que el producto marginal bruto (sin considerar los costes de congestión) del capital en la empresa j es:

$$f_j = (1-\alpha)\Omega^{-\alpha} (k^j)^{\frac{\beta}{1-\alpha}-1} \quad (3)$$

donde la población total ha sido normalizada a uno y $\Omega \equiv \sum_{j=A,B} (k^j)^{\frac{\beta}{1-\alpha}}$. La condición de primer orden respecto del capi-

⁴ En el apéndice A se proveen fundamentos microeconómicos que justifican el uso de este tipo de costes.

tal para una empresa representativa en la ciudad j se puede reescribir como

$$\tilde{f}_j = r^j + \delta \quad (4)$$

donde $\tilde{f}_j \equiv f_j - g'(K^{ij})$ es el producto marginal neto del capital en la ciudad j .

C) Equilibrio

Puesto que la economía está cerrada al comercio internacional y no existe gobierno, el único activo disponible es el capital físico y, por tanto, $z^j = k^j$, $\forall j = A, B$. La siguiente proposición muestra que, debido a la diferencia inicial en los stocks de capital instalado y la existencia de rendimientos crecientes de escala, el producto marginal bruto del capital es siempre estrictamente mayor en la ciudad que tiene un mayor stock de capital.

Proposición 1: $f_A > f_B$, $\forall k^A > k^B > 0$

Demostración: véase apéndice B.

Los supuestos 1 y 2 imponen restricciones en las condiciones iniciales para que el problema sea económicamente interesante. El supuesto 1 establece que, en la fecha inicial, la ciudad A tiene un stock de capital ligeramente mayor al de la ciudad B .

Supuesto 1

$$K_0^A = K_0^B + \epsilon$$

donde ϵ es un número positivo pequeño.

El supuesto 2 impone que, en la fecha inicial, los costes de congestión en la ciudad con un mayor stock de capital son relativamente bajos, comparados con las ganancias de productividad asociadas con el tamaño de ésta. Esto garantiza que la inversión en la ciudad A es más rentable que en la ciudad B en la fecha inicial. Además, es necesario asumir que la inversión es rentable en la pequeña ciudad, es decir, su producto marginal neto del capital es estrictamente mayor que la suma de la tasa de descuento intertemporal y la tasa de depreciación.

Supuesto 2

$$\tilde{f}_{A0} > \tilde{f}_{B0} > \rho + \delta$$

Con estos supuestos, el equilibrio competitivo en esta economía tiene las siguientes propiedades. Las familias empiezan invirtiendo solamente en la ciudad con un mayor stock de capital, la ciudad A , puesto que es la oportunidad de inversión más rentable. Sin embargo, a medida que el stock de capital instalado en esta ciudad aumenta, los costes de congestión crecen rápidamente, de forma que, eventualmente, la inversión empieza a ser positiva también en la ciudad B . Para que el problema sea interesante, es necesario asumir que en el período \hat{t} , en el que la inversión comienza en la ciudad B , la inversión en esa ciudad es efectivamente rentable, esto es

Supuesto 3

$$\tilde{f}_{B\hat{t}} > \rho + \delta$$

Proposición 2: La tasa de crecimiento del consumo per cápita viene dada por

$$\gamma_c \equiv \frac{\dot{c}}{c} = \begin{cases} \tilde{f}_A - \rho - \delta, \forall t \in (0, \hat{t}] \\ \tilde{f}_j - \rho - \delta, \forall t \in (\hat{t}, t^*], \forall j = A, B \\ 0, \forall t > t^* \end{cases}$$

Demostración: véase apéndice B.

Dependiendo de los valores de los parámetros del modelo, es posible que la ciudad con un menor stock inicial de capital converja a la mayor ciudad antes de que la economía alcance su estado estacionario. Llamemos a este escenario *convergencia*. Si el proceso de convergencia no se completa antes de que la economía alcance una tasa de crecimiento nula, ésta reside para siempre en un estado de *no convergencia*⁵.

⁵ Desde un punto de vista empírico, parece evidente que el escenario de no convergencia es el más interesante, ya que, en realidad, se observan ciudades de diferentes tamaños en cada país.

El caso de convergencia

Ésta es una situación en la que la brecha inicial en los stocks de capital desaparece antes de que la economía alcance su estado estacionario. Defínase \bar{t} como el período en el cual eso ocurre. La condición necesaria para este escenario es $\tilde{f}_A(\bar{t}) = \tilde{f}_B(\bar{t}) > \rho + \delta$, es decir, el producto marginal neto del capital converge antes de que la inversión deje de ser rentable. La probabilidad de estar en el caso de convergencia depende de tres tipos de parámetros: primero, depende negativamente del tamaño de la brecha inicial en los stocks de capital. En segundo lugar, cuanto mayor sea ρ , es decir, cuanto más impacientes sean las familias, más rápidamente el estado estacionario será alcanzado y, para una brecha inicial dada, menor será la probabilidad de convergencia. Finalmente, los parámetros de producción α , β y el parámetro de costes σ también afectan a la probabilidad de converger. Intuitivamente, si los rendimientos crecientes son elevados ($\alpha + \beta$ grande), mayor será la fracción de inversión en la ciudad A, reduciendo la probabilidad de convergencia. Lo contrario es cierto si σ es elevado, puesto que, en ese caso, la inversión en la ciudad inicialmente mayor termina relativamente pronto.

Defínase t^* como el período en el que $\tilde{f}_j = \rho + \delta, \forall j = A, B$, es decir, el período en el que la economía alcanza su estado estacionario. Los tres períodos \hat{t} , \bar{t} , y t^* están bien definidos porque, a diferencia del modelo neoclásico, la convergencia a un stock de capital determinado se alcanza en tiempo finito. Esto es así debido a que la condición de Inada $\lim_{k_j \rightarrow \infty} \tilde{f}_j = -\infty$ asegura que la tasa de crecimiento de la economía disminuye suficientemente rápido. El siguiente corolario resume la política de inversión que siguen las familias en el escenario de convergencia:

Corolario 1: La política de inversión es:

$$\begin{cases} i^A > i^B = 0, \forall t \in (0, \hat{t}] \\ i^A < i^B, \forall t \in (\hat{t}, \bar{t}] \\ i^A = i^B, \forall t > \bar{t} \end{cases}$$

El caso de no convergencia

Bajo este escenario, la brecha inicial en los stocks de capital no desaparece totalmente antes de que la economía alcance su estado estacionario. El corolario 2 presenta el plan de inversión que maximiza la utilidad de las familias en este caso:

Corolario 2: La política de inversión es:

$$\left\{ \begin{array}{l} i^A > i^B = 0, \forall t \in (0, \hat{t}] \\ i^A < i^B, \forall t \in (\hat{t}, t^*) \\ i^j = \delta(k^j)^*, \forall j = A, B, \forall t \geq t^* \end{array} \right.$$

La evolución de los stocks de capital en cada ciudad en el caso de convergencia se muestran en el gráfico 3 del apéndice D. Dados los supuestos que se utilizan, el stock inicial de capital es mayor en la ciudad A . Durante el intervalo de tiempo $(0, \hat{t}]$ este stock aumenta en la ciudad A y disminuye en B debido a la depreciación. Una vez que el producto marginal neto del capital se iguala en las dos ciudades, en el período \hat{t} , el stock de capital per cápita crece más rápidamente en la ciudad B que en la A . Esto sigue siendo cierto hasta que el nivel de capital de ambas ciudades converge. A partir de ese instante, los dos stocks crecen a la misma tasa y, finalmente, se mantienen constantes en el estado estacionario, que es alcanzado en el período t^* . El gráfico 4 muestra la evolución de los ratios de población en cada ciudad en esta situación. El peso relativo de la ciudad A aumenta en el intervalo $(0, \hat{t}]$, puesto que las familias solamente invierten allí. Después del período \hat{t} , el ratio de A decrece porque la inversión es mayor en la ciudad B . En el intervalo (\hat{t}, t^*) , como la inversión es idéntica en las dos ciudades y el stock de capital es el mismo, los ratios de población son también idénticos. Los gráficos 5 y 6 muestran el caso de no convergencia. La única diferencia respecto del escenario previo es que el nivel de capital per cápita y el ratio de población de la ciudad B nunca llegan a converger completamente con los de la ciudad A .

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 Brinkhoff¹ 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 caso².

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 trabajo⁴. 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 inicial⁵.

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 más 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 válida 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 estu-

dio. 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 países⁹. 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 tiempo¹⁰, 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 clara-

mente 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 una 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. Así, se puede asumir 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 Zipf¹².

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:

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

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:

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

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.

CAPÍTULO V

CONCLUSIONES

En este trabajo se estudia la evolución del tamaño de las ciudades en diferentes países a lo largo del tiempo. En él se describen tres nuevos hallazgos empíricos. Primero, a medida que los países se desarrollan, el ratio de población en sus mayores ciudades presenta una evolución de U invertida. Es decir, esta medida de concentración aumenta en las fases iniciales de desarrollo y disminuye posteriormente. Usando distintas bases de datos, se calculan los ratios de población que viven en las ciudades inicialmente mayores de un país, y se analiza cómo éstos evolucionan a través del tiempo. El hecho de centrarse en las mayores ciudades en la primera década en la que se dispone de datos, limita los problemas de selección de muestra presentes en la literatura existente sobre el tema. Por otro lado, esta metodología facilita la evaluación del modelo teórico propuesto en el capítulo III, ya que permite seguir las mismas ciudades en distintos momentos del tiempo. El segundo hecho que se documenta es que la distribución de la sección cruzada de tasas de crecimiento de las ciudades es asimétrica hacia la derecha. Finalmente, la última novedad empírica es que el rango de la ciudad que crece más rápidamente en cada década aumenta a lo largo del tiempo.

Desde un punto de vista teórico, estos hechos se interpretan en el trabajo como evidencia en favor de los modelos de tamaño óptimo de una ciudad y, por tanto, como un rechazo de la ley de Gibrat, que ha sido ampliamente utilizada en la literatura existente y sugiere que las ciudades crecen siguiendo un paseo aleatorio. En el trabajo se construye un modelo teórico que predice estos tres hechos

y analiza la transición dinámica de las ciudades hacia su estado estacionario. Este modelo de crecimiento consta de dos ciudades, modelizadas como empresas con tecnologías Cobb-Douglas. La producción en cada una de estas empresas exhibe rendimientos crecientes de escala, pero rendimientos decrecientes de capital y trabajo. Los costes de congestión vienen representados por una función de costes convexa que depende solamente del capital medio instalado en la ciudad en la que opera la empresa. Dependiendo de los valores asignados a los parámetros del modelo, la ciudad que inicialmente tiene un stock de capital menor, converge con la mayor. Por otro lado, el hecho de que los rendimientos crecientes son externos a las empresas privadas implica que el equilibrio competitivo es ineficiente.

El modelo predice un claro patrón de U invertida en la concentración de la población y una distribución de tasas de crecimiento de las ciudades que es asimétrica, con un mayor peso en unas pocas ciudades que crecen mucho más rápidamente que las demás. Asimismo, esta teoría implica que la ciudad que crece más rápidamente en cada período del tiempo es la mayor, condicionado al hecho de que sus costes de congestión no hayan alcanzado un nivel crítico.

Otras implicaciones del modelo están apoyadas por la evidencia empírica. En primer lugar, las ciudades que decrecen lo hacen lentamente. Por otro lado, las mayores ciudades de los países más desarrollados están más cercanas a la ley de Zipf que las de los países en vías de desarrollo.

APÉNDICES

APÉNDICE A

FUNDAMENTOS MICROECONÓMICOS

UN MODELO DE SHARING

Bienes Indivisibles

Siguiendo a DURANTON y PUGA (2003), un motivo por el que las ciudades exhiben rendimientos crecientes es porque permiten a los consumidores compartir bienes indivisibles. Este argumento motiva la existencia de rendimientos crecientes a escala asumiendo directamente que éstos tienen lugar a nivel agregado. En el modelo presentado en este trabajo, la variable \tilde{K}^j puede representar la cantidad media de bienes indivisibles instalados en la ciudad j . Al mismo tiempo, a medida que el tamaño de la comunidad que usa estos bienes indivisibles aumenta, aparecen los costes de congestión. Asumiendo un modelo estándar de ciudad lineal monocéntrica, estos costes vienen del hecho de que el crecimiento de la población de la ciudad fuerza a que los nuevos bienes indivisibles se instalen cada vez más lejos de su centro. En este modelo, tiene sentido pensar que estos costes son pagados por las empresas privadas: a medida que el número de edificios aumenta, los consumidores se van moviendo del centro de la ciudad (donde están situadas todas las empresas) hacia los suburbios y, por tanto, para venderles sus productos, las empresas deben pagar los correspondientes costes de transporte (recuérdese que las empresas no pueden cambiar su localización debido a la existencia de irreversibilidad en la inversión). Indivisibilidades en la provisión de un bien público es solamente una forma de motivar este tipo de modelos. Otra alter-

nativa es asumir indivisibilidades en el proceso de producción. Existe de hecho una extensa literatura que modeliza ciudades de esta forma [KOOPMANS (1957), MILLS (1967), MIRRLEES (1972)].

Mercados Comunes

Otra alternativa es interpretar K^{ij} para una empresa como infraestructura (por ejemplo, edificios y máquinas necesarios en el proceso de producción). En este caso, las empresas privadas se benefician del nivel de estructura de una ciudad, que está relacionado con lo que la literatura llama *mercados comunes* (*common marketplaces*). Cada empresa se beneficia de estar cerca de otras empresas, puesto que puede compartir con éstas un mayor mercado de trabajo. Alternativamente, si se introducen bienes intermedios en el proceso de producción (por ejemplo, usando una función de producción Dixit-Stiglitz), el hecho de situarse cerca de otras empresas reduce el coste de adquirir estos factores de producción. Considérese de nuevo el modelo monocéntrico lineal, donde ahora se asume que la única posible localización de los consumidores es el centro de la ciudad [como en WANG (1990)]. A medida que el stock de estructura urbana aumenta, las inversiones adicionales deben situarse cada vez más lejos del centro, de forma que las empresas tienen que pagar mayores costes de transporte para vender sus productos a los consumidores. Algunos modelos en esta línea son BERLIANT y WANG (1993), WANG (1993), BERLIANT y KONISHI (2000), y KONISHI (2000).

Un modelo de rentas de propiedad

Considérese una función de producción para la empresa i en la localidad j que utiliza tres factores de producción: capital (K^{ij}), trabajo (N^{ij}), y tierra (L^j). Supóngase que existen rendimientos constantes de escala pero rendimientos decrecientes en el capital y el trabajo. La oferta de tierra es limitada. Este último supuesto puede modelizarse como una oferta de tierra que se vuelve cada vez más inelástica a medida que el stock de capital y población crece (el ejemplo más sencillo es una función de producción Leontieff que requiere una cantidad mínima de tierra para operar). En este modelo, los costes de congestión provienen de forma natural de los costes

de las rentas de propiedad. Una ciudad inicialmente grande crecerá muy rápidamente pero, eventualmente, cuando no quede tierra disponible (o su precio se vuelva prohibitivo) la inversión se desplaza a la segunda ciudad, como en el modelo presentado en el texto principal. Este escenario es cualitativamente equivalente a uno en el que el factor de producción tierra es omitido y se considera solamente una forma reducida de los costes de instalación del capital. Otro modelo muy parecido a éste es aquél en el que se asume que hay un componente de productividad que inicialmente aumenta con K (o Y) y eventualmente se convierte en una función decreciente de éste. ARNOTT y STIGLITZ (1979) es un buen ejemplo de un modelo basado en las rentas de la propiedad.

Conocimiento por desbordamiento

Otra posibilidad es interpretar K en el modelo como una combinación de estructura de capital inmóvil y capital humano. Esta interpretación tiene la ventaja de que permite justificar la presencia de rendimientos crecientes debido al conocimiento por desbordamiento (*knowledge spillovers*). Por otro lado, la inversión es irreversible solamente respecto de la fracción de bienes de capital que representan infraestructura. El capital humano (incorporado en la gente) se mueve libremente entre ciudades. Los costes de congestión pueden ser modelizados como rentas de la tierra o costes de transporte que están directamente relacionados con el stock de capital físico instalado en una ciudad, como en los modelos comentados anteriormente. TAMURA (2002, 2006) analiza el papel del capital humano en la generación de rendimientos crecientes en la producción.

APÉNDICE B

EL PROBLEMA DEL PLANIFICADOR SOCIAL

El problema de un planificador social benevolente es asignar trabajo y capital entre ciudades y a lo largo del tiempo, de forma que la utilidad de los individuos es maximizada. Para simplificar notación, se omiten los subíndices de tiempo en la mayoría de las expresiones siguientes:

$$\begin{aligned} & \max \int_0^{\infty} e^{-\rho t} \ln(c) dt \\ & \sum_{j=A,B} N^j = N \\ & \sum_{j=A,B} I^j + \sum_{j=A,B} g(K^j) + C = \sum_{j=A,B} Y^j \\ & \dot{K}^j = I^j - \delta K^j, \forall j = A, B \\ & I^j \geq 0, \forall j = A, B \\ & K_0^j \text{ dado, } \forall j = A, B \end{aligned}$$

donde $Y^j = (N^j)^\alpha (K^j)^\beta$, $0 < \alpha < 1$, $0 < \beta < 1$, y $\alpha + \beta > 1$, para cualquier $j = A, B$. La constante N representa la población nacional, I^j y K_0^j respectivamente denotan la tasa de inversión y el stock inicial de capital instalado en la ciudad j , $j = A, B$, y $\delta \in (0, 1)$ es la tasa de depreciación del capital. Para solucionar este problema, se procede a dividirlo en dos etapas.

Etapa 1: Asignación óptima del trabajo

En esta etapa, tomando como dados los niveles de capital instalado, el planificador social elige la cantidad óptima de trabajo en cada ciudad. Este problema es estático y se reduce a maximizar la producción nacional. Es decir,

$$\begin{aligned} & \max \{Y^A + Y^B\} \\ & \sum_{j=A,B} N^j = N \\ & Y^j = (N^j)^\alpha (K^j)^\beta, \forall j = A, B \\ & K^j \text{ dado}, \forall j = A, B \end{aligned}$$

La solución de este problema es

$$\frac{N^A}{N^B} = \left(\frac{k^A}{k^B} \right)^{\frac{\beta}{1-\alpha}} \quad (5)$$

donde $k^j \equiv \frac{K^j}{N}$, $j = A, B$ representa el stock de capital per cápita instalado en la ciudad j .

De forma similar se definen las siguientes variables per cápita: $y^j \equiv \frac{Y^j}{N}$, $i^j \equiv \frac{I^j}{N}$. Sustituyendo esta asignación óptima en la función de producción agregada del país, se obtiene la siguiente función de valor:

$$f \equiv \sum_{j=A,B} y^j (N^{j*}) = \Omega^{1-\alpha}$$

donde la población nacional ha sido normalizada a 1 y $\Omega \equiv \sum_{j=A,B} (k^j)^{\frac{\beta}{1-\alpha}}$. Se define el producto marginal del capital en la ciudad j como

$$f_j \equiv \frac{\partial f}{\partial k^j} = \beta \Omega^{-\alpha} (k^j)^{\frac{\beta}{1-\alpha}-1} \quad (6)$$

Etapa 2: Asignación óptima del capital

En esta etapa del problema, el capital es asignado entre ciudades y a lo largo del tiempo, de forma que se maximiza el valor presente de la utilidad. Este problema dinámico se puede escribir en términos per cápita como sigue:

$$\begin{aligned} & \max \int_0^{\infty} e^{-\rho t} \ln(c) dt \\ & \sum_{j=A,B} i^j + \sum_{j=A,B} g(k^j) + c = f(k^A, k^B) \\ & \dot{k}^j = i^j - \delta k^j, \forall j = A, B \\ & i^j \geq 0, \forall j = A, B \\ & k_0^j \text{ dado}, \forall j = A, B \end{aligned}$$

Nótese que la función de valor obtenida en la etapa 1 se utiliza aquí en la parte derecha de la restricción de recursos. El Hamiltoniano de este problema es

$$H = e^{-\rho t} \ln c + \lambda \left[f - \sum_{j=A,B} i^j - \sum_{j=A,B} g(k^j) - c \right] + \sum_{j=A,B} \mu^j [i^j - \delta k^j]$$

donde λ , μ^j , $j = A, B$ son multiplicadores de Lagrange. Este problema tiene tres variables de control (c , i^A , i^B) y dos variables de estado (k^A , k^B). Las condiciones de primer orden para la ciudad j son:

$$\frac{\partial H}{\partial c} = 0 \Leftrightarrow e^{-\rho t} \frac{1}{c} = \lambda \quad (7)$$

$$\frac{\partial H}{\partial i^j} \leq 0 \Leftrightarrow -\lambda + \mu^j \leq 0 \quad (8)$$

$$\frac{\partial H}{\partial k^j} = -\dot{\mu}^j \Leftrightarrow \lambda [f_j - g'(k^j)] - \mu^j \delta = -\dot{\mu}^j \quad (9)$$

y la condición de transversalidad es

$$\lim_{t \rightarrow \infty} k_t^j \mu_t^j = 0 \quad (10)$$

Como en el problema descentralizado, es claro que, debido a la presencia de rendimientos crecientes de escala, el capital es más productivo en la mayor ciudad, es decir $f_A > f_B$, $\forall k^A > k^B > 0$.

Los supuestos 1-3 del problema descentralizado se asumen también aquí. Las únicas dos diferencias son que, en primer lugar, el período de convergencia \hat{t} es diferente del que se define en el equilibrio competitivo. En segundo lugar, como se muestra seguidamente, el producto marginal bruto en la ciudad j es diferente en los dos problemas.

Como sucedía en el modelo competitivo, dependiendo de los parámetros del modelo, el estado estacionario de la economía puede estar caracterizado por una distribución degenerada del tamaño de ciudad, donde todas las ciudades tienen los mismos niveles de población y capital (el caso de convergencia) o por una estructura con ciudades de diferente tamaño (el caso de no convergencia). La evolución cualitativa de la economía se muestra, como en el escenario con mercados competitivos, en los gráficos 3-6.

Equivalencia entre los dos problemas

Para hacer los dos problemas comparables, es necesario asumir que $\beta = \psi - \alpha + 1$. Puesto que en el problema del planificador, el parámetro β se restringe a tomar valores entre 0 y 1, esto implica la siguiente restricción en el parámetro de la externalidad ψ .

$$0 < \psi < \alpha$$

En el equilibrio competitivo, la presencia de una externalidad positiva del capital medio de una ciudad en cada una de las empresas que operan en ella implica una tasa de inversión subóptima en cada momento del tiempo. Esta tasa de inversión baja genera dos tipos de ineficiencia en la economía. La primera es la ineficiencia estándar a la ROMER (1986). Esto puede verse claramente en las expresiones del producto marginal del capital [ecuaciones (3) y

(6)]. Puesto que $1 - \alpha < \beta$, las empresas invierten demasiado poco en la economía descentralizada, lo que implica una tasa de crecimiento demasiado baja.

La segunda ineficiencia tiene que ver con los costes de congestión. Puesto que estos costes son convexos, es óptimo construir ciudades de tamaño similar. En el caso de no convergencia, está

claro que $\left(\frac{k^A}{k^B}\right)^{\text{mercado}} > \left(\frac{k^A}{k^B}\right)^{\text{planificador}}$, mientras que en el caso de convergencia, el ratio del mercado es estrictamente superior al del planificador durante al menos un período.

DEMOSTRACIONES

Proposición 1: $f_A > f_B, \forall k^A > k^B > 0$

Demostración: $f_A > f_B$ implica

$$(k^A)^{\frac{\beta}{1-\alpha}-1} > (k^B)^{\frac{\beta}{1-\alpha}-1}$$

Es decir,

$$\left(\frac{k^A}{k^B}\right)^{\frac{\psi}{1-\alpha}} > 1$$

Esto es cierto siempre que $k^A > k^B > 0$, ya que los rendimientos crecientes de escala implican $\psi > 0$ *Q.E.D.*

Proposición 2: La tasa de crecimiento del consumo per cápita viene dado por

$$\gamma_c \equiv \frac{\dot{c}}{c} = \begin{cases} \tilde{f}_A - \rho - \delta, \forall t \in (0, \hat{t}] \\ \tilde{f}_j - \rho - \delta, \forall t \in (\hat{t}, t^*], \forall j = A, B \\ 0, \forall t > t^* \end{cases}$$

Demostración: Reescribábase el problema de la familia como:

$$\begin{aligned} & \max \int_0^{\infty} e^{-\rho t} \ln(c) dt \\ & \sum_{j=A,B} i^j + c = \omega + \sum_{j=A,B} r^j z^j \\ & \dot{z}^j = i^j, \forall j = A, B \\ & i^j \geq 0, \forall j = A, B \\ & z_0^j \text{ dado}, \forall j = A, B \end{aligned}$$

donde i^j representa la tasa neta de inversión en el activo j . El Hamiltoniano de este problema es

$$H = e^{-\rho t} \ln c + v \left[\omega + \sum_{j=A,B} r^j z^j - c - \sum_{j=A,B} i^j \right] + \sum_{j=A,B} \theta^j i^j$$

Donde v y θ^j , $j = A, B$ son multiplicadores de Lagrange. Las condiciones de primer orden vienen dadas por

$$\frac{\partial H}{\partial c} = 0 \Leftrightarrow e^{-\rho t} \frac{1}{c} = \lambda \quad (11)$$

$$\frac{\partial H}{\partial i^j} \leq 0 \Leftrightarrow -v + \theta^j \leq 0 \quad (12)$$

$$\frac{\partial H}{\partial z^j} = -\dot{\theta}^j \Leftrightarrow v r^j = -\dot{\theta}^j \quad (13)$$

Supóngase que, en el intervalo de tiempo $(0, \hat{t}]$, $i^B > i^A = 0$. Esto implica que $v = \theta^B$ y, por lo tanto, de (13)

$$\frac{\dot{v}}{v} = \delta - \tilde{f}_B \quad (14)$$

Tomando logaritmos y derivadas en (11)

$$-\rho - \frac{\dot{c}}{c} = \frac{\dot{v}}{v} \quad (15)$$

Si se igualan (14) y (15) se obtiene

$$\gamma_c = \tilde{f}_B - \rho - \delta$$

Como en ese intervalo de tiempo $\tilde{f}_B < \tilde{f}_A$ esta política de inversión no puede ser óptima para los consumidores. La misma lógica implica que cualquier tasa de inversión positiva en la ciudad B durante este intervalo es subóptima. En el intervalo de tiempo $(\hat{t}, t^*]$ los dos productos marginales netos del capital se igualan y, por tanto, la tasa de crecimiento del consumo viene dada por

$$\gamma_c = \tilde{f}_j - \rho - \delta, \forall j = A, B$$

ya que la inversión neta es estrictamente positiva en ambas ciudades. Finalmente, en cualquier período $t > t^*$, el consumo es constante, ya que la economía ha alcanzado su estado estacionario.

APÉNDICE C

MÚLTIPLES CIUDADES

En esta sección se plantea un modelo con $J > 2$ ciudades, y se argumenta que el comportamiento cualitativo de la economía se caracteriza igualmente por un crecimiento secuencial de ciudades, siempre y cuando éstas difieran en tamaño en el período inicial.

FAMILIAS

La restricción presupuestaria de las familias puede escribirse como

$$\sum_{j=1}^J \dot{z}^j = \omega + \sum_{j=1}^J r^j z^j - c$$

donde $z^j \equiv \frac{Z^j}{N}$ representan la cantidad (en per cápita) de activo j que posee la familia, $j = 1, \dots, J$. Asumiendo una función de utilidad logarítmica, el problema de una familia representativa es:

$$\max \int_0^{\infty} e^{-\rho t} \ln(c) dt$$

$$\sum_{j=1}^J \dot{z}^j + c = \omega + \sum_{j=1}^J r^j z^j$$

$$\dot{z}^j = i^j, \forall j = 1, \dots, J$$

$$i^j \geq 0, \forall j = 1, \dots, J$$

$$z_0^j \text{ dado}, \forall j = 1, \dots, J$$

donde c representa el consumo per cápita y, como antes, las familias se enfrentan a las restricciones de irreversibilidad $\dot{z}^j \geq 0$, $\forall j = 1, \dots, J$.

EMPRESAS

La función de producción de la empresa i localizada en la ciudad j es idéntica a la descrita en el caso de dos ciudades. La movilidad del trabajo implica

$$\frac{N^j}{N^{j'}} = \left(\frac{k^j}{k^{j'}} \right)^{\frac{\beta}{1-\alpha}}$$

El producto marginal bruto del capital (es decir, sin considerar los costes de congestión) en la ciudad j es entonces una función x de los stocks de capital en cada ciudad

$$f_j = x(k^1, k^2, \dots, k^J)$$

donde la población total N se ha normalizado a 1.

EQUILIBRIO

Como en el caso de dos ciudades, es fácil demostrar que $f_j > f_{j'}$, $\forall k^j > k^{j'} > 0$. Considérense los siguientes supuestos sobre la distribución inicial de ciudades:

Supuesto 1'

$$k_0^1 > k_0^2 > \dots > k_0^J$$

Supuesto 2'

$$\tilde{f}_1(0) > \tilde{f}_2(0) > \dots > \tilde{f}_J(0) > \rho + \delta$$

donde $\tilde{f}_j(0) \equiv f_j(0) - g'(k_j(0))$, $\forall j = 1, \dots, J$

Con estos supuestos el único equilibrio competitivo en una economía con J ciudades tiene las mismas características que en el caso más sencillo de dos ciudades. En la fecha inicial, las familias

solamente invierten en la ciudad que tiene un mayor stock de capital, puesto que es la mejor oportunidad de inversión existente. Cuando el stock de capital instalado en esta ciudad alcanza un nivel crítico (dictado por los costes de congestión), la inversión comienza en la segunda mayor ciudad. El proceso continúa con el resto de las ciudades. Como en el modelo con dos ciudades, para asegurar que la inversión tiene lugar en todas las ciudades, es necesario asumir que la productividad del capital en todas ellas es suficientemente alta como para que ésta sea rentable:

Supuesto 3'

$$\tilde{f}_{j\hat{t}_j} > \rho + \delta, j = 1, \dots, J$$

donde \hat{t}_j representa el período en el que la ciudad $j-1$ ha alcanzado su nivel crítico de costes de congestión. La forma exacta de la distribución del tamaño de ciudades en el estado estacionario depende del valor que se asigne a los parámetros. Los dos casos extremos son: uno en el que todas las ciudades terminan teniendo el mismo tamaño (*convergencia estricta*) y otro en el que todas las ciudades tienen diferente tamaño (*no convergencia*), con un gran número de casos intermedios.

CREACIÓN DE NUEVAS CIUDADES

Si se introduce crecimiento de la población en el modelo, es necesario modificar el modelo teórico para permitir la creación de nuevas ciudades. Un trabajo reciente que modeliza esto es HENDERSON y VENABLES (2004). En el modelo presentado en este trabajo, incluso sin crecimiento de la población, es posible introducir la posibilidad de invertir en una nueva ciudad en cada período. Esta ampliación presenta dos dificultades: en primer lugar, el problema de maximización debe incluir una elección discreta (invertir en una nueva ciudad o no hacerlo) además de las decisiones ya estudiadas. La segunda dificultad reside en el hecho de que existe un problema de coordinación entre las empresas, puesto que ninguna tiene incentivos a ser la primera en invertir en una ciudad donde el stock de capital instalado es nulo. A menos que el gobierno provea a las empresas de los incentivos adecuados, es posible que no se invierta en ninguna nueva ciudad.

APÉNDICE D

Gráfico 1
 RATIOS DE POBLACIÓN EN ESTADOS UNIDOS

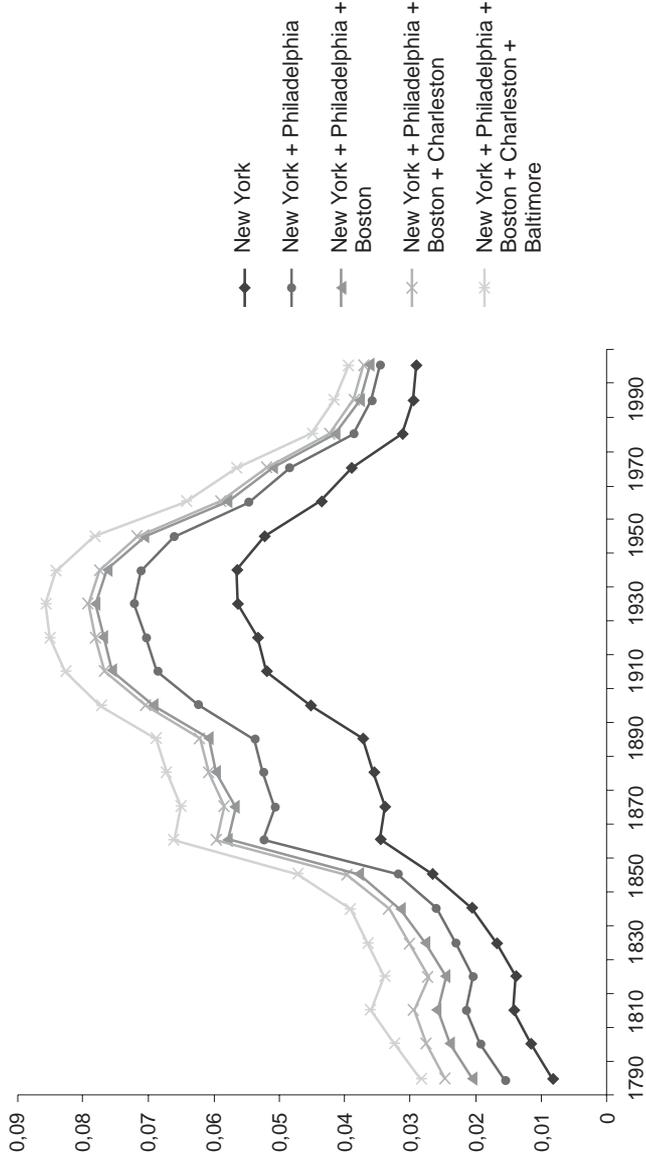


Gráfico 2
RATIOS DE POBLACIÓN EN FRANCIA

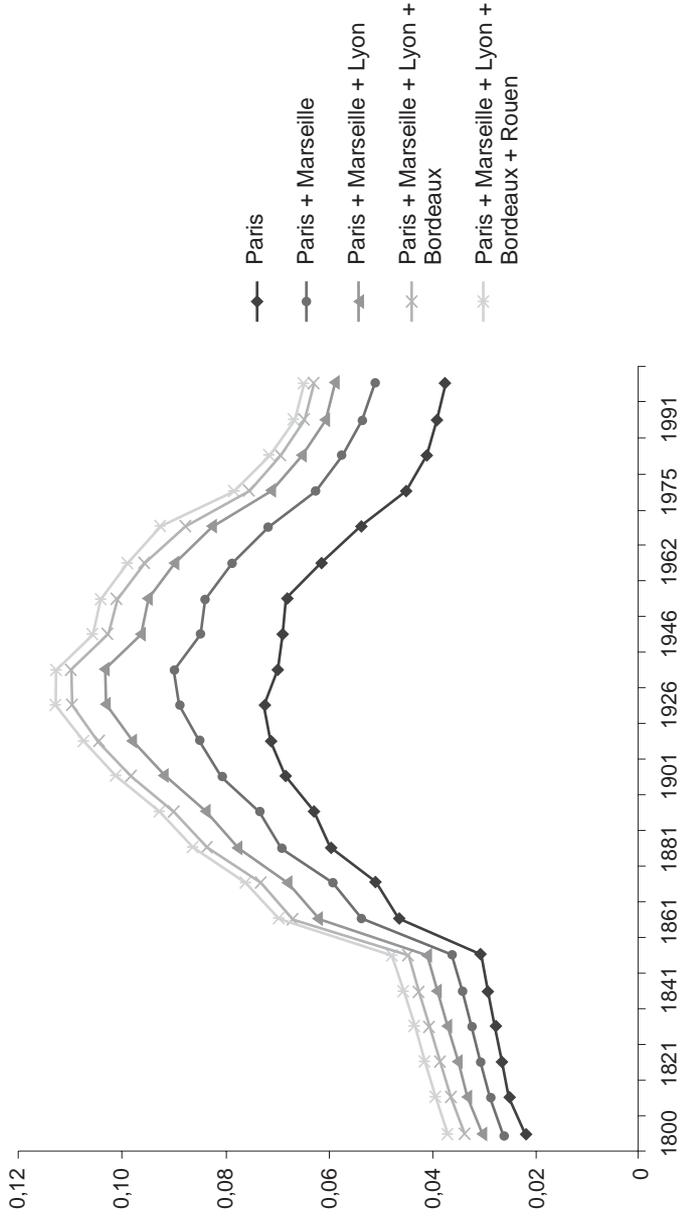


Gráfico 3
CAPITAL PER CÁPITA (CONVERGENCIA)

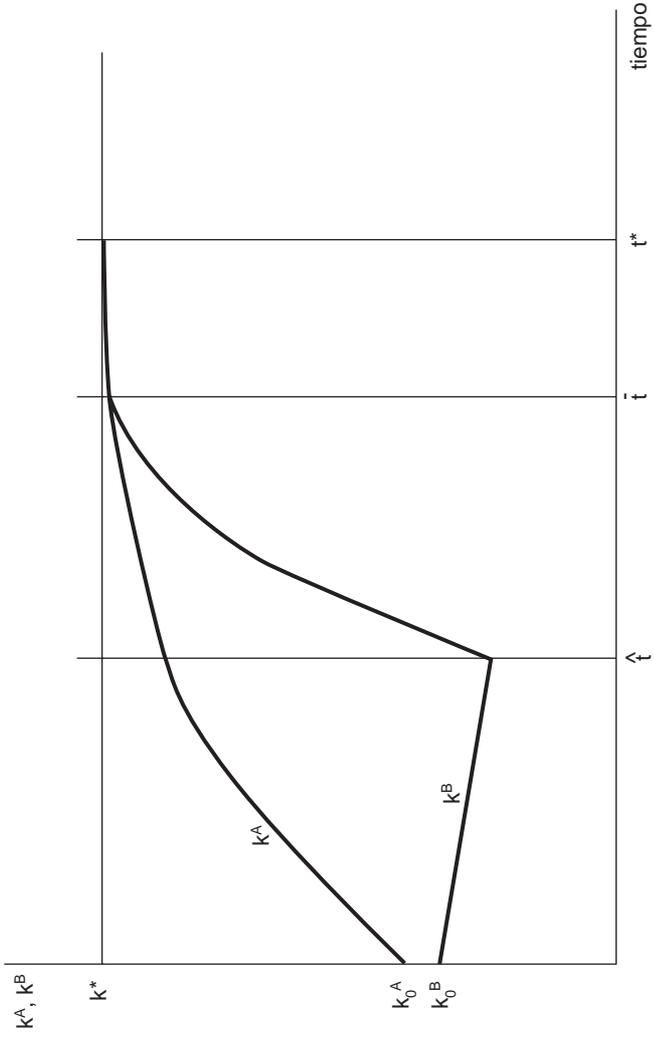


Gráfico 4
 RATIOS DE POBLACIÓN (CONVERGENCIA)

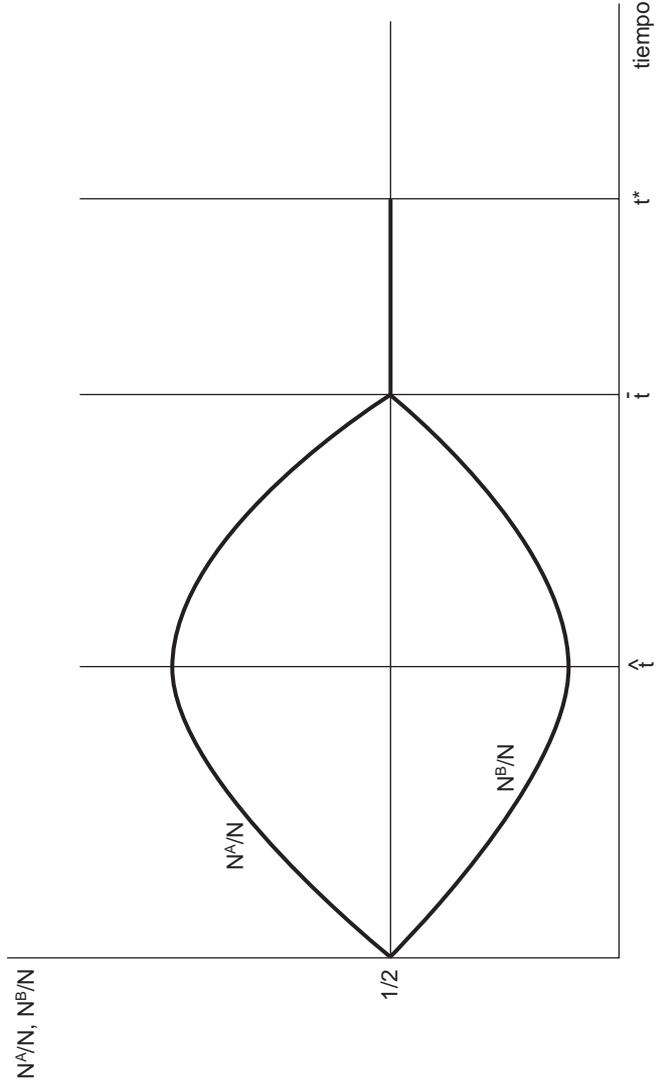


Gráfico 5
CAPITAL PER CÁPITA (NO CONVERGENCIA)

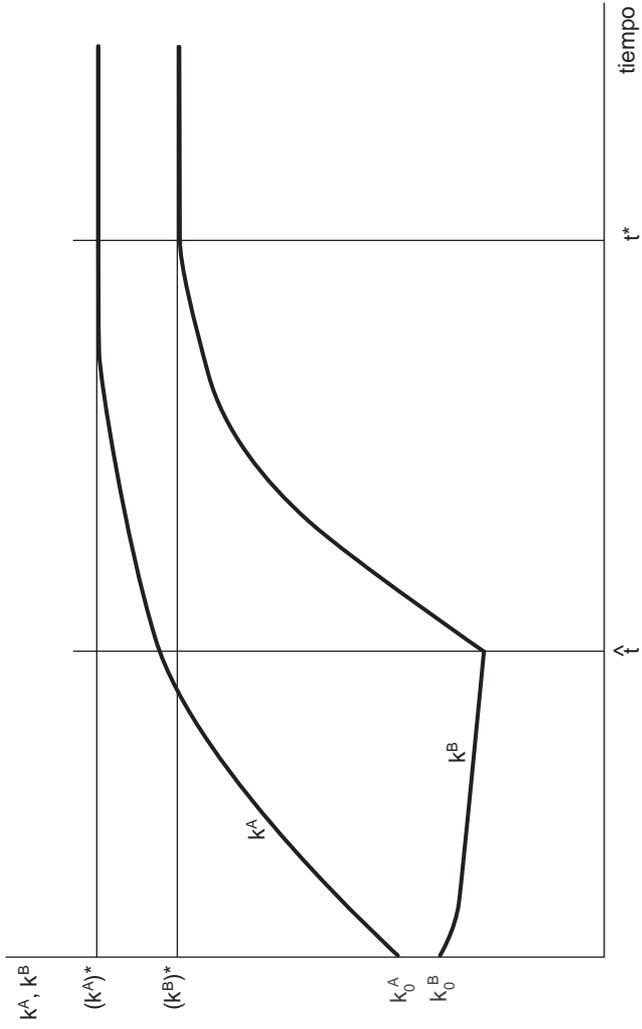


Gráfico 6
 RATIOS DE POBLACIÓN (NO CONVERGENCIA)

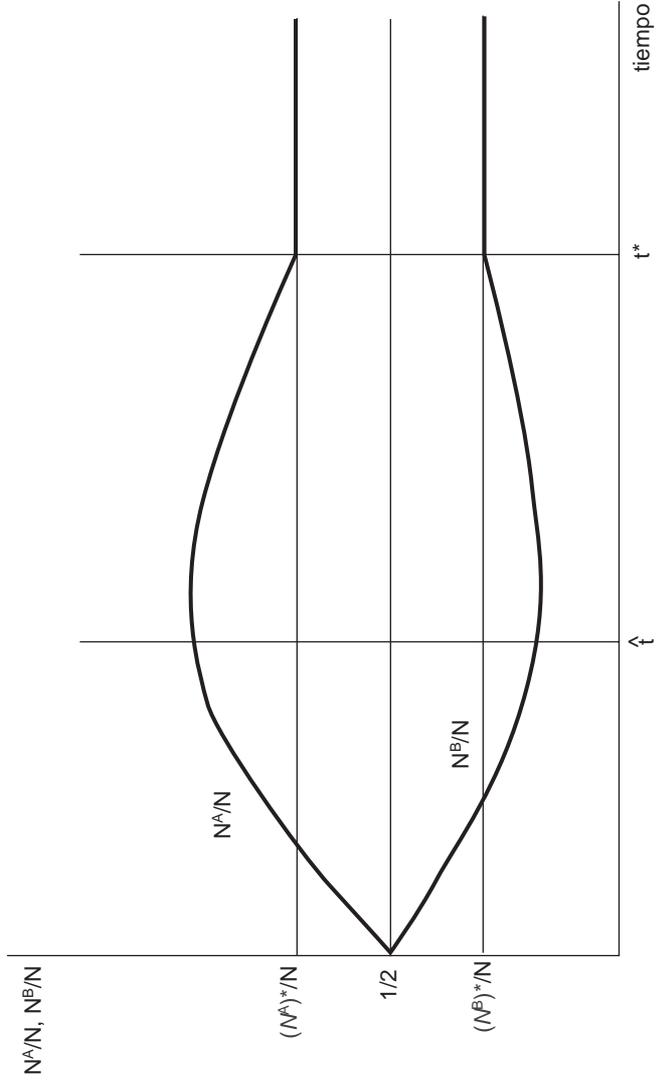


Gráfico 7
 RATIOS DE POBLACIÓN EN MÉXICO

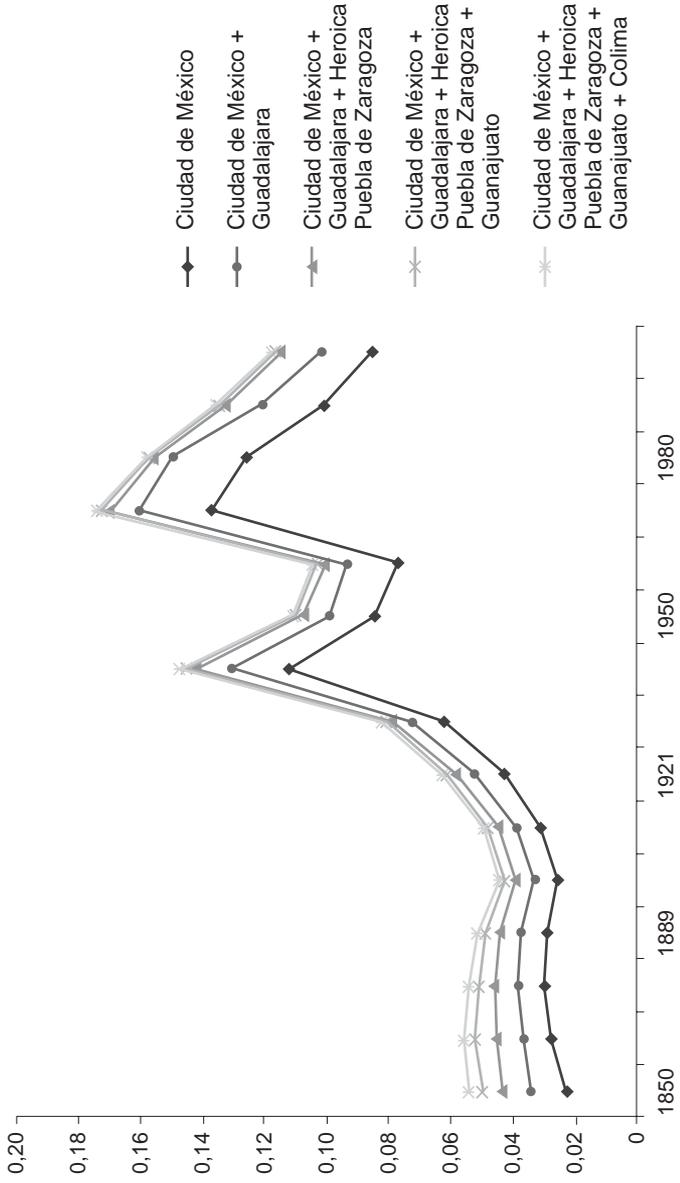


Gráfico 8
RATIOS DE POBLACIÓN EN CHINA

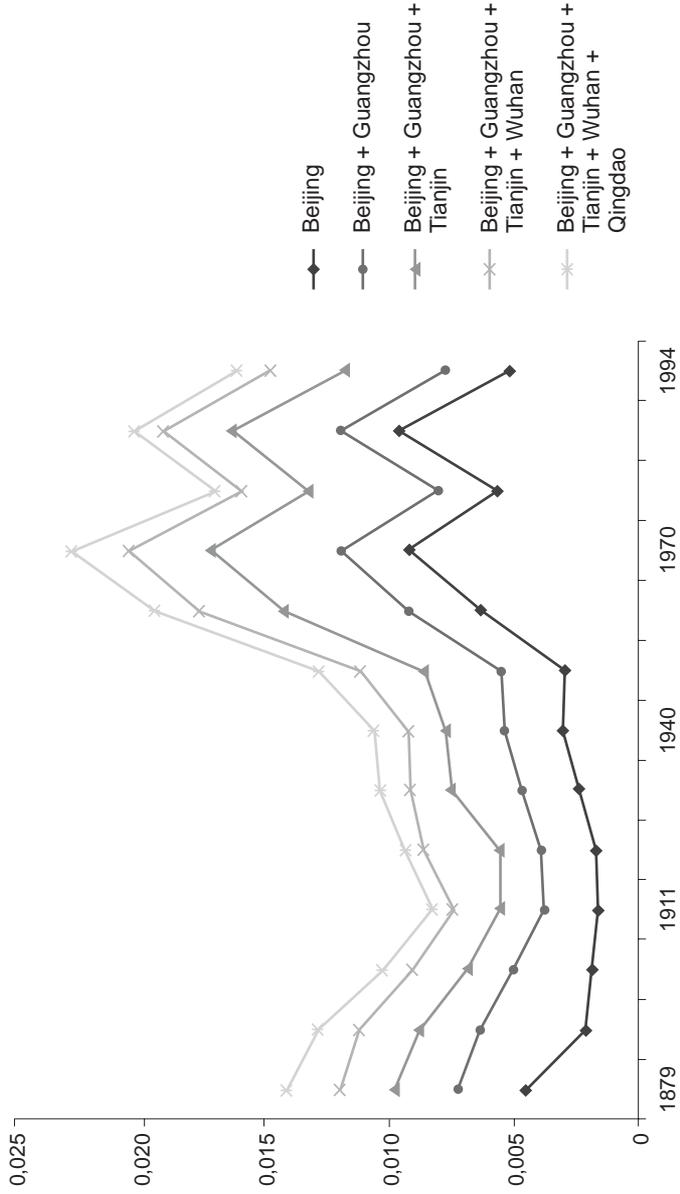


Gráfico 9
RATIO DE POBLACIÓN DE LUANDA EN ANGOLA

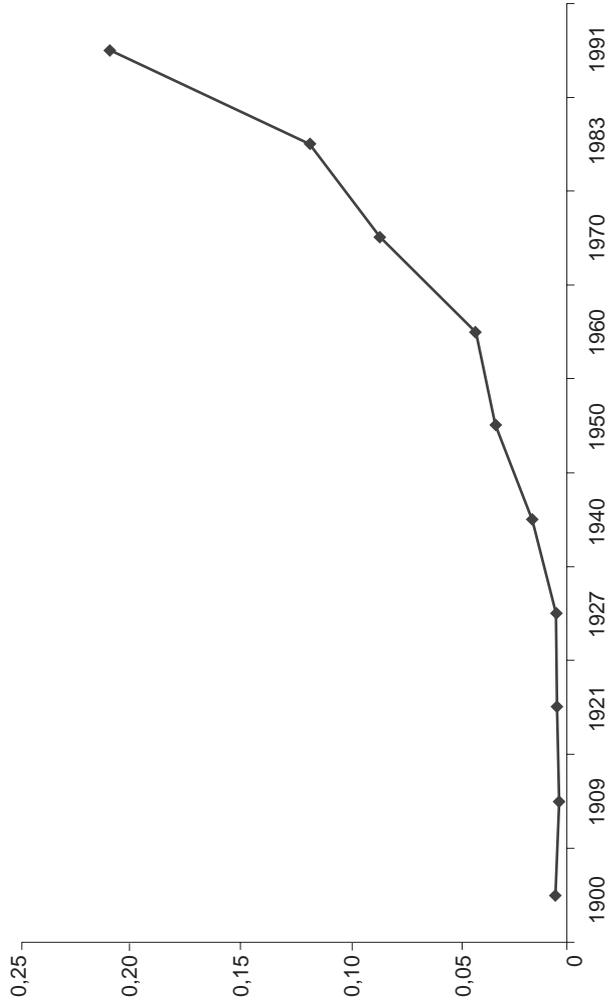


Gráfico 10
 CONTRIBUCIONES MARGINALES
 DE LAS CIUDADES DE ESTADOS UNIDOS A LOS RATIOS DE POBLACIÓN

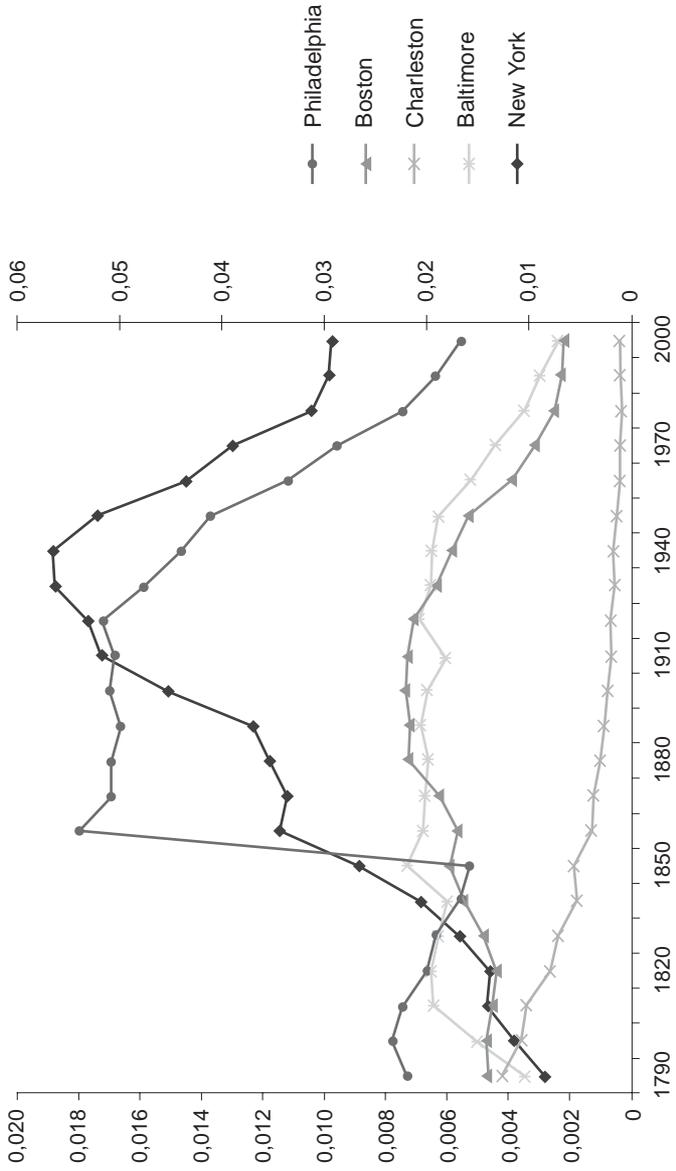


Gráfico 11
RATIOS DE POBLACIÓN DE LAS CINCO
MAYORES CIUDADES DE ESTADOS UNIDOS EN CADA DÉCADA

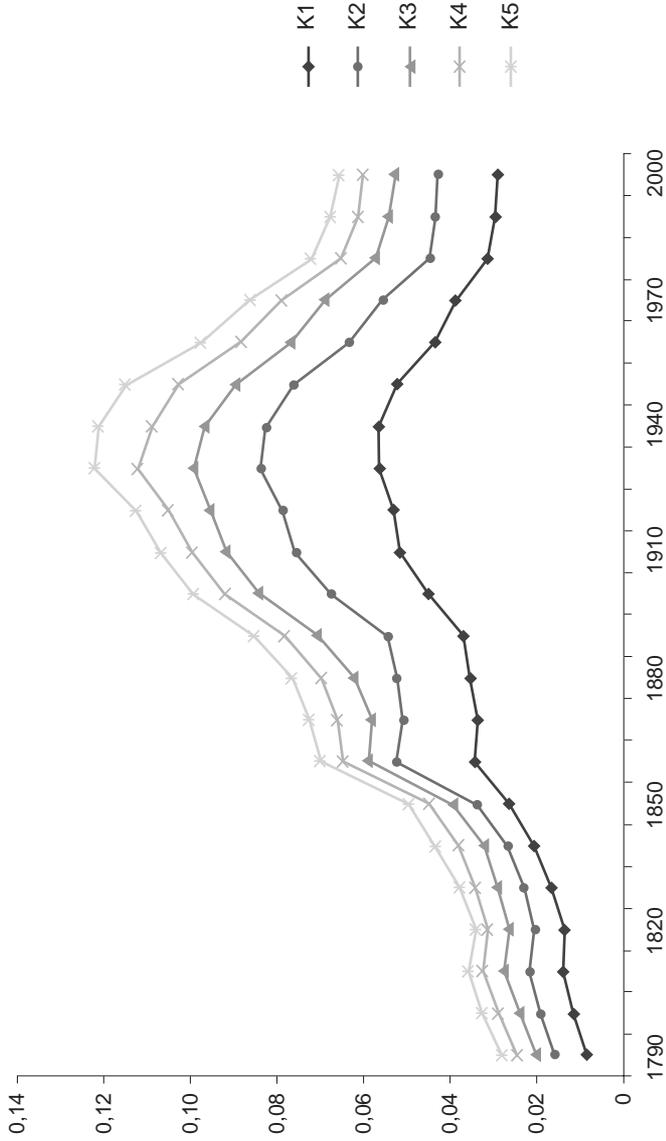


Gráfico 12
DIVERGENCIA EN LAS CIUDADES DE ESTADOS UNIDOS (1790-1840)

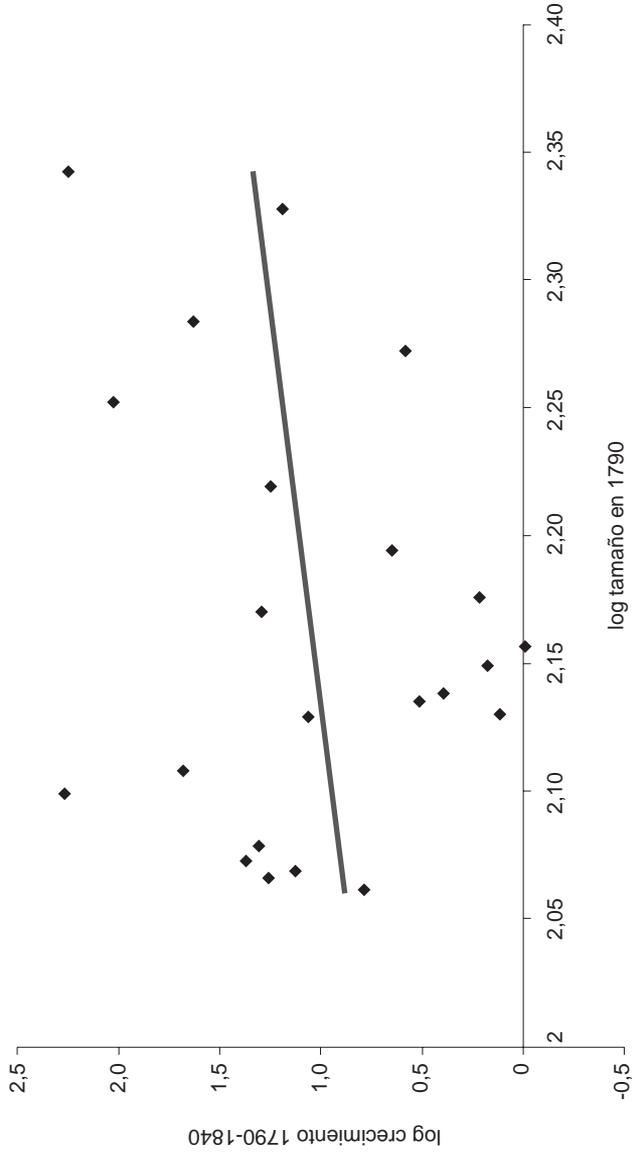


Gráfico 13
CONVERGENCIA ENTRE LAS CIUDADES DE ESTADOS UNIDOS (1940-2000)

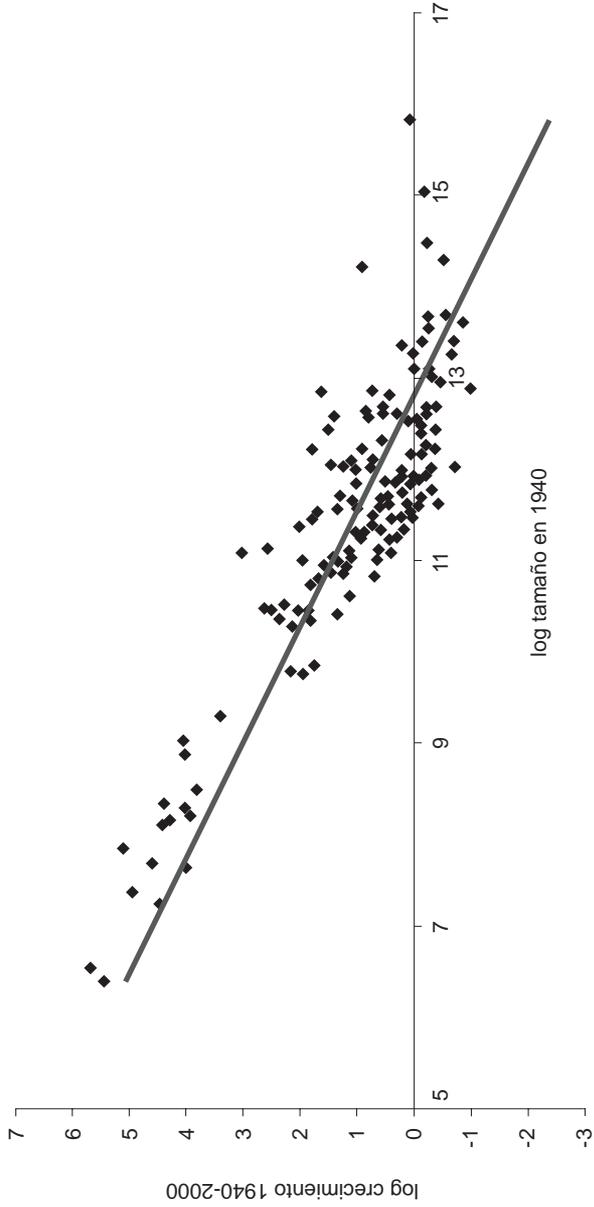


Gráfico 14
RANGO PONDERADO DE LAS CINCO CIUDADES
DE ESTADOS UNIDOS CON MAYOR CRECIMIENTO DE POBLACIÓN

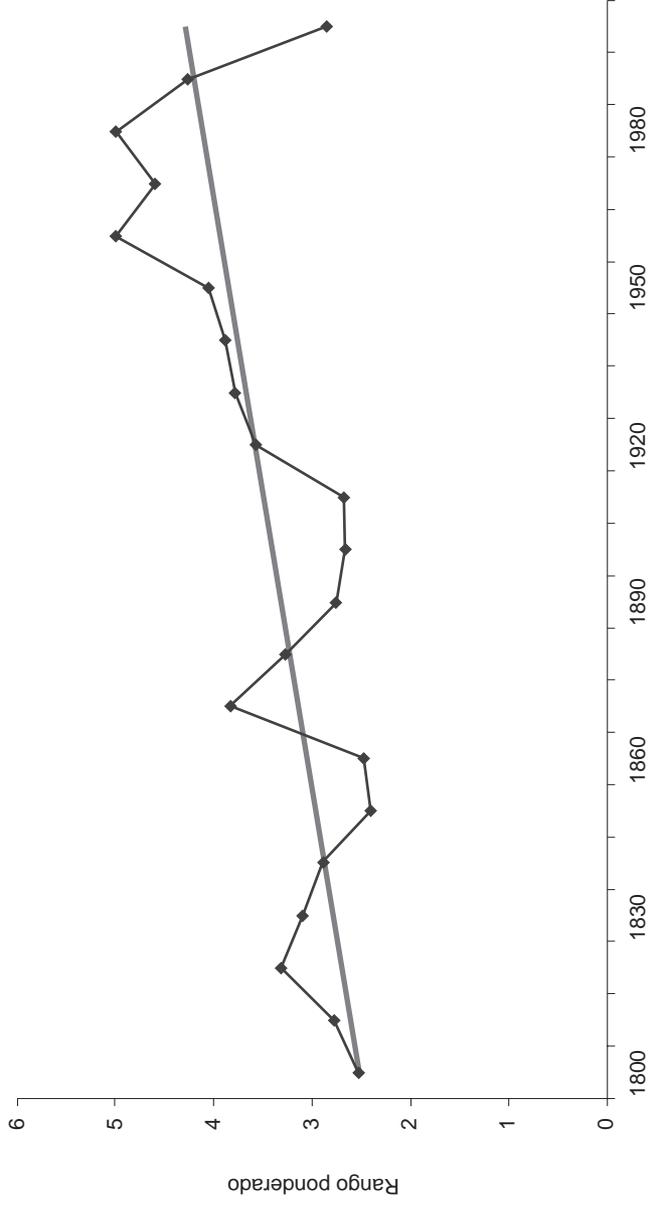


Gráfico 15
DESVIACIONES DE LA LEY DE ZIPF
EN LAS CINCO MAYORES CIUDADES DE ESTADOS UNIDOS EN 2000

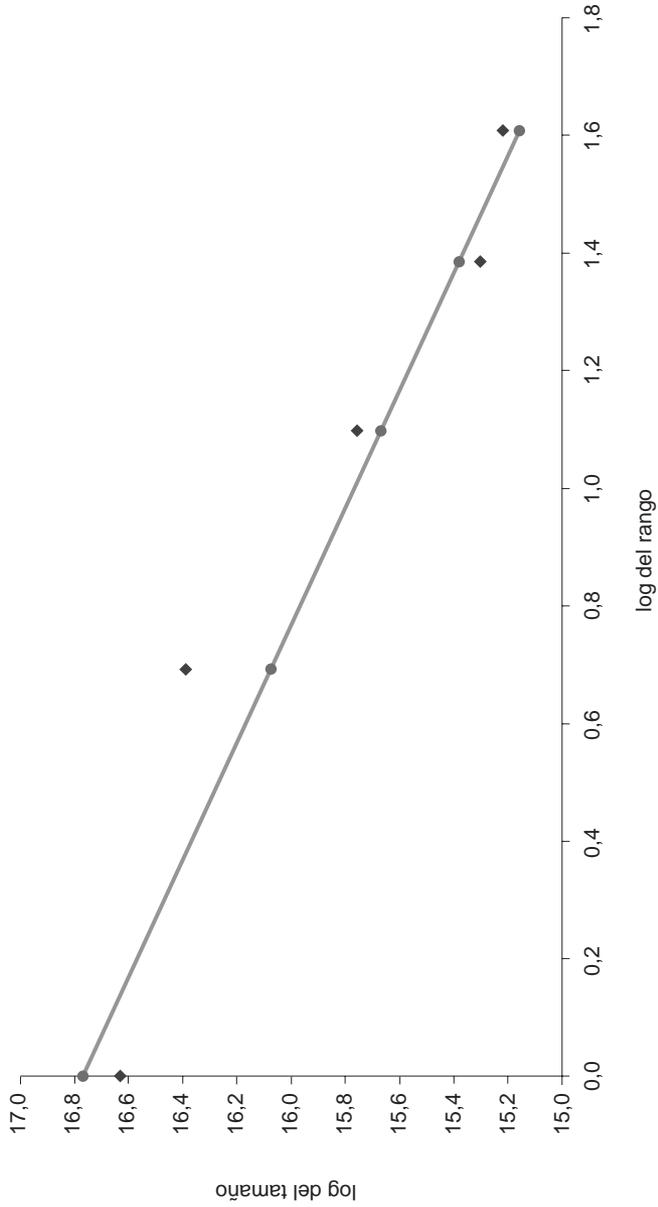


Gráfico 16
DESVIACIONES DE LA LEY DE ZIPF
EN LAS CINCO MAYORES CIUDADES DE ZAIRE EN 2000

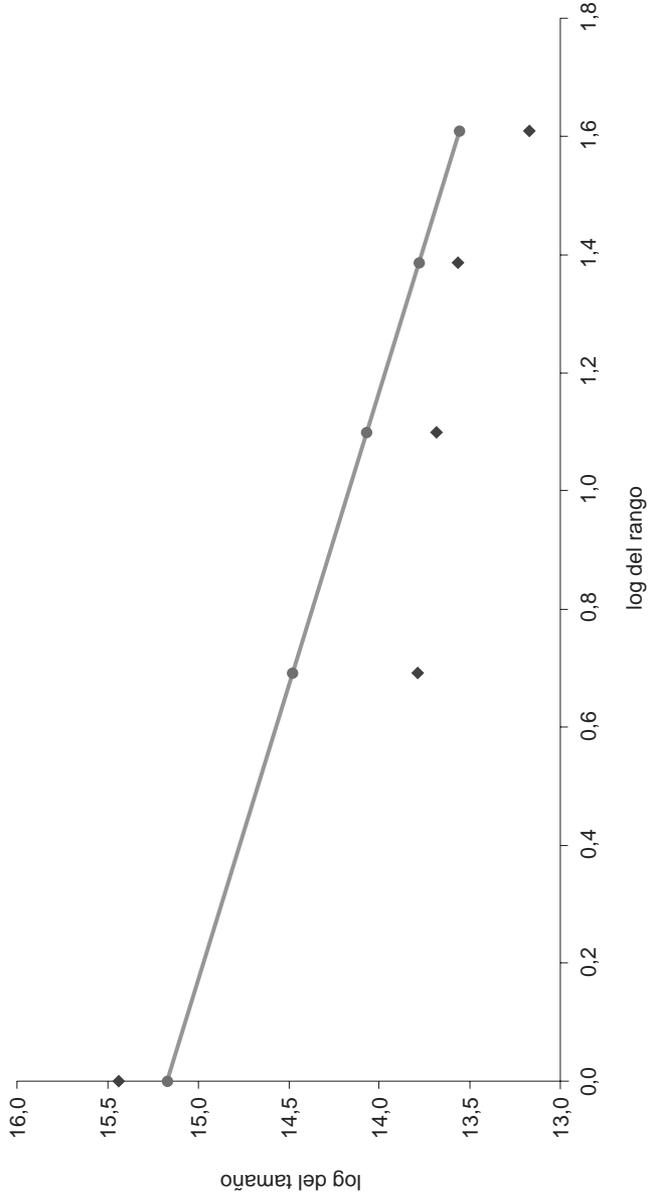


Gráfico 17
DESARROLLO ECONÓMICO Y DESVIACIONES DE LA LEY DE ZIPF

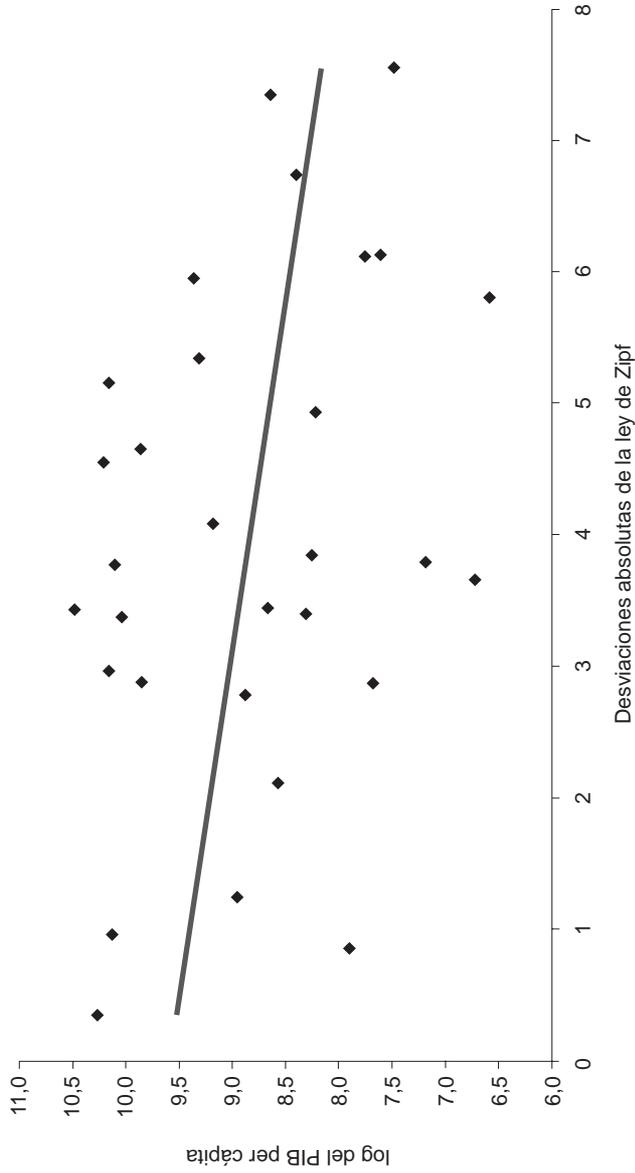


Tabla 1

<i>País</i>	<i>Años con datos</i>	<i>Año de máxima concentración</i>
AFGANISTÁN	1925-1988	1988
ALBANIA	1923-1989	1960
ANGOLA	1900-1991	1991
ARGELIA	1882-1998	1977
ARGENTINA	1780-1990	1910
AUSTRALIA	1871-1991	1981
AUSTRIA	1870-1991	1914
BANGLADESH	1891-1991	1991
BÉLGICA	1867-1999	1867
BOLIVIA	1881-2001	1985
BRASIL	1860-2000	1960
BULGARIA	1888-1990	1990
CAMERÚN	1900-1987	1987
CANADÁ	1851-1990	1931
CHINA	1879-1994	1990
COLOMBIA	1851-1993	1999
COREA DEL SUR	1911-2000	1990
DINAMARCA	1870-2000	1941
ECUADOR	1930-2001	2001
EGIPTO	1872-1996	1966
ESPAÑA	1800-2000	1970
ESTADOS UNIDOS	1790-2000	1940
ETIOPÍA	1925-1994	1974
FILIPINAS	1903-1995	1948
FINLANDIA	1860-2000	1970
FRANCIA	1800-1999	1926
GRECIA	1861-2001	1971
GUATEMALA	1921-1994	1973
HOLANDA	1750-1999	1909
HONDURAS	1901-2000	2000
INDIA	1865-1991	1981
INDONESIA	1901-1990	1990
IRÁN	1910-1996	1976
ITALIA	1800-2001	1970
JAPÓN	1872-1999	1960
KENIA	1911-1999	1999
LIBIA	1905-1988	1988
MALASIA	1901-1991	1980

Tabla 1 (Cont.)

<i>País</i>	<i>Años con datos</i>	<i>Año de máxima concentración</i>
MARRUECOS	1900-1994	1950
MÉXICO	1850-1990	1970
NEPAL	1961-2001	2001
NICARAGUA	1907-1995	1907
NIGERIA	1901-1991	1931
NORUEGA	1801-1990	1970
PAKISTÁN	1891-1981	1981
PARAGUAY	1899-1992	1972
PERÚ	1877-1981	1971
POLONIA	1851-2000	1892
PORTUGAL	1864-2001	1950
REINO UNIDO	1851-1991	1881
REP. CENTRO AFRICANA	1931-1989	1936
RUMANIA	1890-2000	1992
SENEGAL	1916-1988	1945
SUDÁN	1922-1993	1993
SUECIA	1850-1989	1960
SURÁFRICA	1911-1980	1970
TURQUÍA	1927-2000	2000
URUGUAY	1910-1996	1963
VENEZUELA	1921-1990	1981

Tabla 2

<i>País</i>	<i>Años</i>	<i>Coefficiente medio de asimetría</i>	<i>Media-mediana</i>
ALBANIA	1930-1989	0,87	0,05
ARGELIA	1892-1998	0,91	0,04
ARGENTINA	1960-1999	0,60	0,08
AUSTRALIA	1901-1991	0,34	0,02
AUSTRIA	1881-2001	0,44	0,02
BOLIVIA	1892-2001	0,89	0,04
BRASIL	1900-2000	0,51	0,03
BULGARIA	1900-1990	0,56	0,01
CANADÁ	1871-1996	0,46	0,04
CHINA	1948-1994	0,30	0,008
COLOMBIA	1911-1999	0,11	0,04
COREA DEL SUR	1931-2000	0,33	0,01
COSTA RICA	1939-2000	0,20	0,04
DINAMARCA	1931-1989	-0,07	0,02
ECUADOR	1941-2001	0,68	-0,01
EGIPTO	1966-1996	0,72	0,10
EL SALVADOR	1912-1980	0,25	-0,007
ESPAÑA	1870-2000	0,49	0,02
ESTADOS UNIDOS	1790-2000	1,05	0,06
FILIPINAS	1918-1995	0,54	0,05
FINLANDIA	1891-2000	0,52	0,03
FRANCIA	1861-1999	0,02	-0,004
GRECIA	1928-2001	0,17	0,009
HOLANDA	1811-1999	0,41	0,02
HONDURAS	1910-2000	0,11	0,02
HUNGRÍA	1869-1999	0,43	0,02
INDIA	1871-1991	-0,26	-0,002
INDONESIA	1930-1990	1,30	-0,02
IRÁN	1920-1996	0,26	0,07
IRLANDA	1901-1991	0,91	0,06
ISRAEL	1939-2000	0,19	0,08
ITALIA	1861-1970	0,67	0,01
JAPÓN	1891-1999	0,32	0,02
KENIA	1939-1999	0,75	0,07
LUXEMBURGO	1911-2001	0,67	0,06
MALASIA	1931-1991	-0,54	-0,02
MARRUECOS	1941-1982	-0,15	0,01
MÉXICO	1867-1980	0,28	0,03

Tabla 2 (Cont.)

<i>País</i>	<i>Años</i>	<i>Coficiente medio de asimetría</i>	<i>Media-mediana</i>
NORUEGA	1815-1980	0,42	0,03
NUEVA ZELANDA	1901-2001	0,24	0,02
PAKISTÁN	1901-1981	0,36	0,03
POLONIA	1862-1990	0,44	0,02
PORTUGAL	1878-2001	0,44	0,04
REPÚBLICA CHECA	1890-1991	0,83	0,02
RUMANIA	1900-2000	0,28	0,01
RUSIA	1910-1991	0,15	0,03
SUDÁN	1941-1993	0,33	0,03
SUECIA	1920-1994	0,17	0,04
SUIZA	1910-1990	0,09	0,004
TURQUÍA	1940-2000	0,68	0,03
URUGUAY	1953-1996	-0,66	-0,01
VENEZUELA	1926-1990	0,07	0,02

Tabla 3

<i>País</i>	<i>Años</i>	<i>Positivo</i>	<i>Test 1</i>	<i>Test 2</i>	<i>Test 3</i>	<i>Test 4</i>
ARGENTINA	1960-1999	0,80	0,60	0,80	0,80	0,80
AUSTRIA	1934-2001	0,71	0,43	0,43	0,57	0,71
BRASIL	1900-2000	0,91	0,64	0,70	0,70	0,70
BULGARIA	1921-1990	0,86	0,14	0,14	0,43	0,43
CANADÁ	1911-1996	1,00	0,78	0,89	0,78	0,78
ESPAÑA	1910-1999	0,70	0,30	0,30	0,40	0,40
ESTADOS UNIDOS	1800-2000	0,90	0,43	0,48	0,57	0,62
FINLANDIA	1931-2000	1,00	1,00	1,00	1,00	1,00
FRANCIA	1881-1999	0,64	0,21	0,43	0,43	0,36
GRECIA	1921-2001	0,86	0,29	0,29	0,57	0,57
HOLANDA	1951-1999	0,67	0,50	0,67	0,67	0,50
INDIA	1881-1991	0,67	0,17	0,38	0,25	0,38
ITALIA	1871-2001	0,86	0,29	0,43	0,43	0,50
JAPÓN	1891-1999	0,83	0,25	0,20	0,20	0,20
MÉXICO	1910-1980	0,88	0,50	0,625	0,625	0,75
NORUEGA	1910-1980	1,00	0,375	0,625	0,50	0,50
POLONIA	1911-2000	0,90	0,40	0,50	0,70	0,60
REPÚBLICA CHECA	1921-1991	0,86	0,57	0,57	0,57	0,57
RUMANIA	1956-2000	1,00	0,17	0,34	0,34	0,34
RUSIA	1939-1991	0,60	0,20	0,20	0,20	0,20
SUECIA	1920-1994	1,00	0,375	0,50	0,375	0,50
Media		0,84	0,41	0,5	0,53	0,54

Tabla 4

<i>País</i>	<i>Años</i>	<i>Rango</i>	<i>Rango ponderado</i>
ALBANIA	1930-1989	Creciente	Creciente
ARGELIA	1892-1998	Sin patrón	Sin patrón
ARGENTINA	1960-1999	Creciente	Decreciente
AUSTRALIA	1901-1991	Creciente	Sin patrón
AUSTRIA	1881-2001	Creciente	Creciente
BOLIVIA	1892-2001	Decreciente	Decreciente
BRASIL	1900-2000	Creciente	Creciente
BULGARIA	1900-1990	Creciente	Creciente
CANADÁ	1871-1996	Creciente	Creciente
CHINA	1948-1994	Creciente	Decreciente
COLOMBIA	1911-1999	Decreciente	Decreciente
COREA DEL SUR	1931-2000	Creciente	Creciente
COSTA RICA	1939-2000	Decreciente	Creciente
DINAMARCA	1931-1989	Decreciente	Decreciente
ECUADOR	1941-2001	Creciente	Creciente
EGIPTO	1966-1996	Creciente	Creciente
EL SALVADOR	1912-1980	Decreciente	Creciente
ESPAÑA	1870-2000	Creciente	Creciente
ESTADOS UNIDOS	1790-2000	Creciente	Creciente
FILIPINAS	1918-1995	Creciente	Creciente
FINLANDIA	1891-2000	Creciente	Creciente
FRANCIA	1861-1999	Creciente	Creciente
GRECIA	1928-2001	Creciente	Creciente
HOLANDA	1811-1999	Creciente	Creciente
HONDURAS	1910-2000	Decreciente	Sin patrón
HUNGRÍA	1869-1999	Creciente	Creciente
INDIA	1871-1991	Creciente	Creciente
INDONESIA	1930-1990	Creciente	Creciente
IRÁN	1920-1996	Decreciente	Decreciente
IRLANDA	1901-1991	Creciente	Creciente
ISRAEL	1939-2000	Creciente	Creciente
ITALIA	1864-1970	Decreciente	Decreciente
JAPÓN	1891-1999	Creciente	Creciente
KENIA	1939-1999	Creciente	Sin patrón
LUXEMBURGO	1911-2001	Creciente	Creciente
MALASIA	1931-1991	Creciente	Decreciente
MARRUECOS	1941-1982	Decreciente	Decreciente
MÉXICO	1867-1980	Decreciente	Decreciente

Tabla 4 (Cont.)

<i>País</i>	<i>Años</i>	<i>Rango</i>	<i>Rango ponderado</i>
NORUEGA	1815-1980	Creciente	Sin patrón
NUEVA ZELANDA	1901-2001	Sin patrón	Creciente
PAKISTÁN	1901-1981	Decreciente	Sin patrón
POLONIA	1862-1990	Creciente	Creciente
PORTUGAL	1878-2001	Creciente	Creciente
REPÚBLICA CHECA	1890-1991	Decreciente	Creciente
RUMANIA	1900-2000	Creciente	Creciente
RUSIA	1910-1991	Creciente	Sin patrón
SUDÁN	1941-1993	Decreciente	Decreciente
SUECIA	1920-1994	Creciente	Creciente
SUIZA	1910-1990	Sin patrón	Creciente
TURQUÍA	1940-2000	Creciente	Decreciente
URUGUAY	1953-1996	Creciente	Creciente
VENEZUELA	1926-1990	Creciente	Creciente

Tabla 5

<i>Ciudad</i>	<i>Rango</i>	<i>Tasa de crecimiento</i>
NEW YORK	1	0,60
PHILADELPHIA	2	0,36
BOSTON	3	0,67
CHARLESTON	4	0,31
BALTIMORE	5	0,14

Tabla 6

<i>País</i>	<i>Años</i>	<i>Rango ponderado</i>
ARGENTINA	1960-1999	Decreciente
AUSTRIA	1934-2001	Creciente
BRASIL	1900-2000	Creciente
BULGARIA	1921-1990	Creciente
CANADA	1911-1996	Creciente
REPÚBLICA CHECA	1921-1991	Creciente
FINLANDIA	1931-2000	Decreciente
FRANCIA	1880-1999	Creciente
GRECIA	1921-2001	Creciente
INDIA	1881-1991	Sin patrón
JAPÓN	1891-1999	Creciente
MÉXICO	1910-1980	Sin patrón
HOLANDA	1951-1999	Creciente
NORUEGA	1910-1980	Creciente
POLONIA	1911-2000	Creciente
RUMANIA	1956-2000	Decreciente
RUSIA	1939-1991	Creciente
ESPAÑA	1910-1999	Creciente
SUECIA	1920-1994	Creciente
ESTADOS UNIDOS	1800-2000	Creciente

BIBLIOGRAFÍA

- ARNOTT, R., y STIGLITZ, J. (1979), «Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size», *Quarterly Journal of Economics*, 93: 471-500.
- AXTELL, R. L., y FLORIDA, R. (2001), *Emergent Cities: a Microeconomic Explanation of Zipf's Law*, documento presentado en la Society of Computational Economics, Yale University.
- BERLIANT, M., y KONISHI, H. (2000), «The endogenous formation of a city: Population agglomeration and marketplaces in a location specific production economy», *Regional Science and Urban Economics*, 30(3): 289-324.
- BERLIANT, M., y WANG, P. (1993), «Endogenous formation of a city without agglomerative externalities or market imperfections: Marketplaces in a regional economy», *Regional Science and Urban Economics*, 23(1): 121-144.
- BLACK, D., y HENDERSON, J. V. (1999), «A Theory of Urban Growth», *Journal of Political Economy*, 107: 252-284.
- BRINKHOFF, T., *City Population*, <http://www.citypopulation.de>
- BROWN, J. D. (1996), *Testing in Language Programs*, Upper Saddle River, NJ: Prentice Hall.
- CARLINO, G., y CHATTERJEE, S. (2001), «Aggregate Employment Growth and the Deconcentration of Metropolitan Employment», *Journal of Monetary Economics*, vol. 48, núm. 3, diciembre.
- CORDOBA, J. C. (2003), «On the Distribution of City Sizes», *documento de trabajo*, Rice University.
- CUBERES, D. (2005), *Appendix to The Rise and Decline of Cities*, manuscrito, University of Chicago.
- D'AGOSTINO, R. B.; BALANGER, A., y D'AGOSTINO, R. B. Jr. (1990), «A Suggestion for Using Powerful and Informative Tests of Normality», *The American Statistician*, 44(4): 316-321.
- DOBKINS, L., y IOANNIDES, Y. (1998a), «Dynamic Evolution of the U.S. City Size Distribution» en J. M. HURIOT y J. F. THISSE

- (eds.), *The Economics of Cities*, New York: Cambridge University Press.
- DURANTON, G. (2002), *City Size Distribution as a Consequence of the Growth Process*, Department of Geography and Environment, London School of Economics.
- DURANTON, G., y PUGA, D. (2004), «Micro-foundations of urban agglomeration economies» en J. VERNON HENDERSON y J.-F. THISSE (eds.), *Handbook of Regional and Urban Economics*, vol. 4, Amsterdam: North-Holland: 2063-2117.
- EATON, J., y ECKSTEIN, Z. (1997), «Cities and Growth: Theory and Evidence from France and Japan», *Regional Science and Urban Economics*, XXVII: 443-474.
- ECKHOUT, J. (2004), «Gibrat's Law for (all) Cities», *American Economic Review*, diciembre, vol. 94, núm. 5: 1429-1451(23).
- FUJITA, M.; KRUGMAN, P., y VENABLES, A. (1999), *The Spatial Economy*, Cambridge, MA: MIT Press.
- GABAIX, X. (1999a), «Zipf's Law and the Growth of Cities», *American Economic Review Papers and Proceedings*, LXXXIX: 129-132.
- GABAIX, X., y IOANNIDES, Y. (2004), «The Evolution of City Size Distributions,» en J. V. HENDERSON, y J. F. THISSE (eds.), *Handbook of Regional and Urban Economics*, Amsterdam: North-Holland: 2341-2378.
- GIBRAT, R. (1931), *Les inégalités économiques*, Paris: Librairie du Recueil Sirey.
- GLAESER, E., y GYOURKO, J. (2004), «Urban Decline and Durable Housing», próximamente en *Journal of Political Economy*.
- GLAESER, E.; SCHEINKMAN, J., y SHLEIFER, A. (1995), «Economic Growth in a Cross-Section of Cities», *Journal of Monetary Economics*, XXXVI: 117-143.
- GLAESER, E., y SHAPIRO, J. (2001), «Is There a New Urbanism? The Growth of U.S. Cities in the 1990s», *NBER Working Paper*, núm. W8357.
- HANSEN, N. (1990), «Impacts of Small and Intermediate-Sized Cities on Population Distribution: Issues and Responses», *Regional Development Dialogue*, 11: 60-76.
- HENDERSON, J., <http://www.econ.brown.edu/faculty/henderson/worldcities.html>
- (1974), «The Sizes and Types of Cities», *American Economic Review*, LXIV: 640-656.
- (1988), *Urban Development: Theory, Fact, and Illusion*, Oxford University Press.
- (2004), «Urbanization and Growth», capítulo preparado para el *Handbook of Economic Growth*, vol. 1, P. AGHION y S. DURLAUF (eds.), North Holland.

- HENDERSON, J., y VENABLES, A. J. (2004), «The Dynamics of City Formation: Finance and Governance», *documento de trabajo*.
- HENDERSON, J., y WANG, H. G. (2005), «Urbanization and City Growth», *Journal of Economic Geography*.
- IOANNIDES, Y. M., y OVERMAN, H. C. (2003), «Zipf's Law for Cities: An Empirical Examination», *Regional Science and Urban Economics*, 33: 127-137.
- KONISHI, H. (2000), «Formation of hub cities: Transportation cost advantage and population agglomeration», *Journal of Urban Economics*, 48(1): 1-28.
- KOOPMANS, T. C. (1957), *Three Essays on the State of Economic Science*, New York: McGraw-Hill.
- KRUGMAN, P. (1996b), «Confronting the Mystery of Urban Hierarchy», *Journal of the Japanese and the International Economies*, 10: 399-418.
- KRUGMAN, P., y BREZIS, E. S. (1997), «Technology and Life Cycle of Cities», *Journal of Economic Growth*: 369-383.
- LAHMEYER, J., <http://www.library.uu.nl/wesp/jalahome.htm>
- LUCAS, R. E. Jr. (1988), «On the Mechanics of Economic Development», *Journal of Monetary Economics*, 22, julio: 3-42.
- MILLS, E. S. (1967), «An aggregative model of resource allocation in a metropolitan area», *American Economic Review Papers and Proceedings*, 57(2): 197-210.
- MIRRELES, J. A. (1972), «The optimum town», *Swedish Journal of Economics*, 74(1): 114-135.
- PENN WORLD TABLE, <http://pwt.econ.upenn.edu/4>
- ROMER, P. M. (1986), «Increasing Returns and Long-Run Growth», *Journal of Political Economy*, 94, 5, octubre: 1002-1037.
- ROSEN, K., y RESNICK, M. (1980), «The Size Distribution of Cities: an Examination of the Pareto Law and Primacy», *Journal of Urban Economics*, 8: 165-186.
- ROSSI-HANSBERG, E., y WRIGHT, M. L. J. (2003), «Urban Structure and Growth», *NBER Working Paper*, núm. W11262.
- ROYSTON, P. (1983), «A Simple Method for Evaluating the Shapiro-Franca W' Test of Nonnormality», *Applied Statistics*, 32: 297-300.
- (1991c), sg3.5, «Comment on sg3.4 and an Improved D'Agostino Test», *Stata Technical Bulletin*, 3: 23-24. Publicado en *Stata Technical Bulletin Reprints*, vol. 1: 110-112.
- SHAPIRO, S. S., y FRANCA, R. S. (1972), «An Approximate Analysis of Variance Test for Normality», *Journal of the American Statistical Association*; 67: 215-216.
- SHAPIRO, S. S., y WILK, M. B. (1965), «An Analysis of Variance Test for Normality (complete samples)», *Biometrika*, 52: 591-611.

- SOO, K. T. (2003): «Zipf's Law for Cities: a Cross Country Investigation», *Regional Science and Urban Economics*, 33 (2).
- SUTTON, J. (1997), «Gibrat's Legacy», *Journal of Economic Literature*, XXXV: 40-59.
- TABACHNICK, B. G., y FIDELL, L. S. (1996), *Using Multivariate Statistics*, 3.^a ed., New York: Harper Collins.
- TAMURA, R. (1991), «Income Convergence in an Endogenous Growth Model», *Journal of Political Economy*, 99: 522-540.
- (2002), «Human Capital and the Switch from Agriculture to Industry», *Journal of Economic Dynamics and Control*, 27: 207-242.
- (2006), «Human Capital and Economic Development», próximamente en *Journal of Development Economics*.
- U. S. CENSUS, <http://www.census.gov>
- WANG, P. (1990), «Competitive equilibrium formation of marketplaces with heterogeneous consumers», *Regional Science and Urban Economics*, 20 (2): 295-304.
- (1993), «Agglomeration in a linear city with heterogeneous households», *Regional Science and Urban Economics*, 23(2): 291-306.
- WILLIAMSON, J. (1965), «Regional Inequality and the Process of National Development: A Description of the Patterns», *Economic Development and Cultural Change*, vol. XIII, núm. 4, Parte II, julio. *World Urbanization Prospects: The 2003 Revision*.
- ZIPF, G. (1949), *Human Behavior and the Principle of Least Effort*, Cambridge, MA: Addison-Wesley.