TI 2005-048/2
Tinbergen Institute Discussion Paper

Zipf's Law for Integrated Economies

Harry P. Bowen[1]
Haris Munandar[2]
Jean-Marie Viaene[2,3]

[1] *Vlerick Leuven Gent Management School, Leuven,*
[2] *Erasmus University Rotterdam, and Tinbergen Institute,*
[3] *CESifo.*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

ZIPF'S LAW FOR INTEGRATED ECONOMIES

Harry P. Bowen†, Haris Munandar‡ and Jean-Marie Viaene §

June 2005

Abstract

This paper considers aspects of the growth process of countries that are members of a fully integrated economy (FIE), i.e., an economy with free mobility of goods and factors among members, and whose members share the same technology. We first demonstrate that each member's share of total FIE output and its shares of total FIE stocks of each productive factor will be equal. If economic polices are harmonized across FIE members then the equality of output and factor shares implies that the growth in any FIE member's output can be considered random. Given this, we build on Gabaix's (1999) result for the distribution of relative city sizes to show that the distribution of output and factor shares among FIE members will exhibit Zipf's law. We empirically examine for Zipf's law for the distribution of output and factor shares across two (presumably) integrated economies: the 51 US states and 14 European Union (EU) countries. Our findings support Zipf's law for US states and indicate convergence towards this law among EU countries. Our findings suggest that models of growth of members within an FIE should embody a key assumption: the normalized growth process is random and homogeneous across FIE members.

JEL Classification: E13, F15, F21, F22, O57

Keywords: Growth, economic integration, Zipf's law.

† Vlerick Leuven Gent Management School, Vlamingenstraat 83, 3000 Leuven, Belgium; e-mail: harry.bowen@vlerick.be

‡ Erasmus University Rotterdam and Tinbergen Institute, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands; e-mail: munandar@few.eur.nl

§ Erasmus University Rotterdam, Tinbergen Institute and CESifo, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands; e-mail: viaene@few.eur.nl

Zipf's Law for Integrated Economies

An extensive body of work has explored the role of international trade and of factor mobility between politically defined regions (e.g., countries) as mechanisms generating endogenous economic growth. For example, Grossman and Helpman (1991) show that trade generally enhances growth, particularly when it facilitates the international transmission of knowledge. Similarly, Rivera-Batiz and Romer (1991) show that increased trade due to economic integration may have both level and growth effects depending upon the processes by which R&D and information flow across borders. Devereux and Lapham (1994) extend Rivera-Batiz and Romer's model to show that, even without knowledge flows, the balanced growth rate when there is free trade in goods alone exceeds that in autarky, provided that initial levels of national income differ across countries.

Regarding factor mobility, Baldwin and Martin (2004) show that the relation between growth and the agglomeration of economic activity depends crucially on the extent of capital mobility between regions. Similarly, Viaene and Zilcha (2002) show that while complete capital market integration among countries has a positive effect on outputs, it does not raise long-run growth rates above autarky values. Instead, these growth rates are affected only by parameters that describe the accumulation of human capital.

Increases in trade or factor mobility can arise from greater economic integration between markets. In the limit, such integration would be represented by a fully integrated economy (FIE) in which there is free mobility of goods and factors among FIE members. While prior work has demonstrated the potentially important role of trade and factor mobility as influences on economic growth, less attention has been given to the question of how trade and factor mobility impact the distribution of output across members of a FIE, and hence how these influences affect the relative economic position of members. Apart from being simply a question of distributional consequences, an analysis of this question has important implications for models that are used to characterize the growth processes of FIE members. As we will demonstrate in this paper, the distribution of output and factor shares across FIE members can be expected to conform to a rank-share distribution that exhibit Zipf's law, which indicates a specific relationship between the ranks and values of a variable.[1] This result implies that models used to characterize the growth of members within an FIE must embody a key assumption: that the underlying normalized growth process is random and homogeneous across members.

In what follows, we first demonstrate the result of Viaene and Zilcha (2002) that each FIE member's share of total FIE output will equal its shares of the total FIE stock of each productive factor (i.e., physical capital and human capital). If economic policies are largely harmonized across members then this equal-share property implies that the growth in any member's shares of FIE output and factor stocks can be considered a random outcome. Following Gabaix's (1999), if it is assumed that the underlying distribution of growth rates is common across members then the limiting distribution of output shares (and factor shares) among FIE members will exhibit Zipf's law. We then show that if the share distributions do exhibit Zipf's law then the values of the output and factor shares are completely determined once the number of FIE members is specified.

Given the theoretical expectation of Zipf's law for output and factor shares, we empirically examine for this law within two (presumably) integrated economies: the 51

[1] Zipf's law for city sizes is an empirical regularity widely documented in the urban and regional economics literature. Interpretive surveys of the implications of rank-size distributions for urban growth include Brakman et al. (2001), Fujita et al. (1999), and Gabaix and Ioannides (2004).

US states and 14 countries of the European Union (EU). The data generally cover the period from 1965 to 2000. Our empirical results convincingly support Zipf's law for US states and they indicate convergence toward Zipf's law for EU countries.

1 Output and Factor Shares in Integrated Economies

Consider an economy that produces a single good by means of a constant return to scale production function that takes the following form:

(1) $$Y_t = F(K_t, H_t)$$

where Y_t is the level of output, K_t the level of physical capital stock and H_t the level of human capital stock, all at time t. For ease of exposition, we assume the production function takes the Cobb Douglas form:[2]

(2) $$Y_t = AK_t^{\alpha} H_t^{1-\alpha}$$

where A is a scale parameter and a is capital's share of total output. By definition, the marginal product of physical capital is:

(3) $$(F_k)_t = \alpha A \left(\frac{H_t}{K_t} \right)^{1-\alpha}$$

Combining (2) and (3) gives:

(4) $$(F_k)_t = \alpha \frac{Y_t}{K_t}$$

We now consider the marginal product of physical capital in another economy that shares the same technology:

(5) $$(F_k^*)_t = \alpha A \left(\frac{H_t^*}{K_t^*} \right)^{1-\alpha} = a \frac{Y_t^*}{K_t^*}$$

[2] The incorporation of physical and human capital has repeatedly been shown to have empirical relevance in production. In particular, the Cobb-Douglas production function provides a good fit on data for the US and other industrial countries (Mankiw et al., 1992).

where "*" indicates variables of the second economy. If physical capital is perfectly mobile between the two economies, and hence the two economies constitute a fully integrated economy (FIE), then capital will flow from the low to high rate of return economy until its rate of return is equalized. From the equality between (3) and (5) we obtain:

$$(6) \qquad \frac{H_t}{K_t} = \frac{H_t *}{K_t *} = \frac{H_t + H_t^*}{K_t + K_t^*}$$

Likewise, using (4) and(5):

$$(7) \qquad \frac{Y_t}{K_t} = \frac{Y_t^*}{K_t^*} = \frac{Y_t + Y_t^*}{K_t + K_t^*}$$

Combining (6) and (7) yields the so-called (Bowen et al. (2005)) equal-share relationship:

$$(8) \qquad \frac{Y_t}{Y_t + Y_t^*} = \frac{K_t}{K_t + K_t^*} = \frac{H_t}{H_t + H_t^*}$$

Expression (8) determines the distribution of output and the distribution of factors between the two economies. Hence, with perfect capital mobility, each economy's share of total FIE output and each economy's share of total FIE physical capital stock equals its share of total FIE stock of human capital. As discussed in Bowen et al. (2005), the equal-share relationship (8) has three important extensions. First, this relationship remains valid even if there are technological differences and differences in factor rates of return between the two economies; such differences only cause a rescaling of the original variables. Second, relationship (8) can be extended to the case of a FIE that consists of N members. If these N members are assumed to have the same technology, and there is free mobility of at least one factor (physical or human capital) among them, then the equalization of factor rates of return implies the following form of the equal-share relationship:

(9) $$\frac{Y_{it}}{\sum_{l=1}^{N} Y_{lt}} = \frac{K_{it}}{\sum_{l=1}^{N} K_{lt}} = \frac{H_{it}}{\sum_{l=1}^{N} H_{lt}}$$

or, written in terms of shares, $y_{it} = k_{it} = h_{it}$. Finally, if FIE members have harmonized economic and social policies (e.g., fiscal, education, industrial policies) then the equal-share property implies that the relative performance of any one member can be considered to be a random variable dependent on the particular state of nature at time *t*.

2 Rank-Share Distributions and Zipf's Law

A rank-share distribution is related to the concept of a rank-size distribution. The latter describes a particular relationship between the size and rank of a variable across a set of observational units. For example, let the variable be city size as measured by a city's population, and order cities in decreasing order of their size to obtain the rank of each city according to its size. A rank-size distribution for city size exists if the relationship between the natural logarithm of these two variables is linear and exhibits a negative slope. The special case of Zipf's law arises when the slope value equals -1. The existence of Zipf's law for city sizes is a widely documented empirical regularity (Brakman et al. (2001), Fujita et al. (1999), and Gabaix and Ioannides (2004)).

Several explanations have been advanced for the observed regularity of Zipf's law with respect to the distribution of city sizes. Some argue it constitutes an optimal spatial pattern that arises when congestion and urbanization externalities interact as part of the process of development and growth of cities. Such forces are usually found in core models of urban and regional growth (Eaton and Eckstein, 1997; Black and Henderson 1999; Brakman et al., 1999). Others have stressed more mechanical forces that often involve a random growth process for city size. A recent example is Gabaix

(1999), who draws on Gibrat's law[3] to assume that cities follow a random but common growth process. Normalizing city population by a country's total population, Gabaix shows (his Proposition 1) that if these population shares evolve as geometric Brownian motion (with an infinitesimal barrier) then the steady state distribution of population shares will be a rank-size distribution that exhibits Zipf's law.

As previously noted, the equal-share property for members of an FIE, together with an assumed harmonization of FIE member's economic policies, implies that the relative performance of any one FIE member can be considered a random variable. Given this, we can adopt Gabaix's (1999) specification and assume that the growth rate of the share for variable *j* (e.g., *j* = output) evolves as geometric Brownian motion, and moreover, that the distribution of such growth rates is common to all FIE members (i.e., Gibrat's law).[4] As in Gabaix (1999), this implies that the limit distribution of the shares of variable *j* across FIE members will be a rank-share distribution that exhibits Zipf's law. We now show that if the distribution of shares does conform to Zipf's law then the share values are in fact completely determined once the number of FIE members is specified.

Consider a FIE consisting of *N* members. Let S_{ij} denote member *i*'s share of the total FIE amount of variable *j* (e.g., *j* = output) and let R_{ij} denote the rank of member *i* in the ranking of the values of variable *j* across all members ($i = 1,\ldots, N$). We assume that $R_{ij} = 1$ for the member with the largest value (share) of variable *j* and that $R_{ij} = N$

[3] Gibrat's law (Gibrat, 1931) states that firm growth is independent of firm size.

[4] The equal-share relationship implies that the common mean rate of growth is zero since $\sum_{l=1}^{N} y_l = \sum_{l=1}^{N} k_l = \sum_{l=1}^{N} h_l = 1$.

for the member with the lowest value (share) of variable j. If variable j has a rank-share distribution then we can write:[5]

(10) $$S_{ij} = g_j \left(1/R_{ij}\right)^{b_j}$$

where $b_j > 0$ and $0 < g_j < 1$ is the share of variable j for the member with the highest rank (i.e., when $R_{ij} = 1$). Zipf's law corresponds to $b_j = 1$.

Let V_{ij} denote the level of variable j for member i. Now assume, without loss of generality, that member 1 has the highest value of variable j and let d_{ij} be member i's value of variable j relative to that of member 1 (i.e., $d_{ij} = V_{ij} / V_{1j}$), so that $d_{1j} = 1$. Now order the values of variable j in descending order. This ordering of the values of variable j across the $i = 1, \ldots, N$ members can then be written:

(11) $$V_{1j} > d_{2j} V_{1j} > d_{3j} V_{1j} > \ldots > d_{Nj} V_{1j}$$

Since the total FIE amount of variable j is $(1 + d_{1j} + d_{2j} + \ldots + d_{Nj})V_{1j}$, (11) implies the following relations between member ranks and shares:

(12)
$$\begin{aligned}
&Rank\ 1:\ S_{1j} = \frac{1}{1+d_{2j}+d_{3j}+\ldots+d_{Nj}} \\
&Rank\ 2:\ S_{2j} = \frac{d_{2j}}{1+d_{2j}+d_{3j}+\ldots+d_{Nj}} \\
&Rank\ 3:\ S_{3j} = \frac{d_{3j}}{1+d_{2j}+d_{3j}+\ldots+d_{Nj}} \\
&\quad\vdots \qquad\qquad\qquad \vdots \\
&Rank\ \mathrm{N}:\ S_{Nj} = \frac{d_{Nj}}{1+d_{2j}+d_{3j}+\ldots+d_{Nj}}
\end{aligned}$$

Expressions (12) indicate that the value of each share S_{ij} depends only on the number of members N. In the special case where the distribution of shares exhibits Zipf's law then it must be that $d_{2j} = 1/2$, $d_{3j} = 1/3$, $d_{4j} = 1/4$, etc. and the sequence of

[5] The literature usually expresses this as $S_{ij} = g_j \left(R_{ij}\right)^{-b_j}$. We depart from this usual form to simply our later presentation and discussion of our empirical analysis.

shares (S_{ij}) then becomes an unbounded Harmonic series. Therefore, if Zipf's law holds, the theoretical shares in (12) can be computed once the number of members (N) is specified. For example, our empirical analysis will consider two FIEs: the 51 US states and 14 EU countries. The theoretical share values for the N = 51 US states are: 0.2213, 0.1106, 0.0738, 0.0553,…, 0.0043. For the N = 14 EU countries the theoretical share values are: 0.3075, 0.1538, 0.1025, 0.0769,…, 0.0220. By the equal-share condition (9), the theoretical share values for member i are the same for output, physical capital and human capital.

Finally, we note that if the share distributions of output, physical capital and human capital shares are each assumed to exhibit Zipf's law then, since the theoretical share values depend only on the number of FIE members, the equal-share relationship derived in the preceding section must hold. Equally, it can be demonstrated that the equal-share property is also obtained if one assumes that output shares alone exhibit Zipf's law, and one further assumes that FIE members have identical, homogenous of degree one, production functions.

3 Empirical Specification

To empirically assess the hypothesis that output and factor shares have a rank-share distribution that exhibits Zipf's law we can take the natural logarithm of each side of (10) to obtain:

$$(13) \qquad \log\left(S_{ij}\right) = q_j + b_j \log\left(1/R_{ij}\right) + u_{ij} \qquad i = 1,...,N;\; j = y,k,h$$

where $q_j = \log(g_j)$ and we have appended the error term u_{ij} which is assumed to have the usual properties (i.e., i.i.d. with mean zero and constant variance). Estimates of the intercept and slope parameter in (13) can be formed by regressing the share of variable j on the inverse of the rank value across members of a given FIE.

In what follows we will separately estimate (13) for the output share (y), the physical capital share (k) and the human capital share (h). We then perform a set of tests intended to examine for evidence of rank-share distribution that exhibits Zipf's law. To examine for evidence of a rank-share distribution we test if the estimated slope parameter in each equation is significantly different from zero. To examine for Zipf's law we test if the estimated slope is significantly different from one. To examine for evidence of the equal-share relationship we test for the homogeneity of the slope estimates (i.e., if $b_y = b_k = b_h$). We further test for the equal-share relationship for the highest rank member (i.e., California for US states and Germany for EU countries) by testing homogeneity of the intercepts (i.e., if $q_y = q_k = q_h$). Finally, we examine if the distribution of shares predicted by (13) conforms to the distributions of observed and theoretical shares (computed using (12)).

We estimate (13) for each of our three variables (output, physical capital and human capital) with respect to the 51 US states and 14 EU countries. For US states, we use annual cross-section data covering the period from 1990 to 2000. For EU countries the data instead consists of cross-sections equally spaced at 5 year intervals; these data generally cover the periods from 1965 to 2000. The Appendix gives a complete description of the data.

4 Results

Table 1 reports regression estimates of (13) for the share of output, physical capital and human capital across US states. Table 2 presents such estimates for the sample of EU countries.[6] Over both set of results, the adjusted R-squares fall in the

[6] The standard errors reported in these tables are "robust" (Newey and West, 1987).

range from 0.791 to 0.945, indicating a strong relationship between the share and rank of each variable.

[Insert Tables 1 and 2 about here]

For US states, Table 1 indicates strong support for the hypotheses that the output and factor share distributions conform to a rank-share distribution; in all cases the hypothesis that the slope coefficient is zero can be strongly rejected (*p*-values < 0.001). In addition, in no case can we reject (at the 5% level) the hypothesis that the slope coefficient is significantly different from unity, indicating that each of the three share distributions exhibit Zipf's law. This is a striking empirical result, and is consistent with the finding of many studies in the urban and regional economics literature that Zipf's law holds for the distribution of city sizes.

For EU countries, Table 2 indicates strong support for the hypothesis that the output and factor share distributions conform to a rank-share distribution; in all cases we can strongly reject the hypothesis that the slope coefficient is zero (*p*-values < 0.001). However, unlike US states, the hypothesis that the rank-share distribution exhibits Zipf's law can, in some cases, be rejected at the 5% level. In particular, the hypothesis of Zipf's law can be rejected for the distribution of output shares in the early sample years (1960, 1965 and 1970) but not in later years (1975 and thereafter). A similar pattern emerges for the distribution of the human capital share: Zipf's law is rejected for 1985 and earlier years but not for the years after 1985. Finally, for physical capital, Zipf's law is rejected in three (i.e., 1985, 1995 and 2000) of the eight years. We note that the value of the slope coefficient for the output and human capital distributions appears to converge towards unity over time.

Gabaix and Ioannides (2004) have demonstrated using Monte Carlo simulation that regression estimates of rank-share distributions have an inherent bias that diminishes with the number of observational units (e.g., cities or countries). Specifically, they show that an OLS estimate of the slope parameter in (13) will be biased upward and that the estimated standard error will be biased downward. These biases would lead one to more often reject Zipf's law when it is in fact true.

Following Gabaix and Ioannides (2004), we examined for the extent of these biases in our analysis by conducting a Monte Carlo analysis of OLS estimates of (13) under the assumption that Zipf's law holds. Table 3 presents the results of this analysis conducted for five different numbers of FIE members. Three facts emerge from this analysis. First, the OLS slopes are indeed biased upward (rows 2 and 3); the upward bias is 0.081 for US states (N = 51) and 0.172 for EU countries (N = 14). Second, the OLS standard errors are biased downward relative to the true standard errors (rows 4 and 5). The true 95% confidence interval (row 6) is therefore wider compared to that based on the OLS standard error. Third, the magnitude of each bias falls the higher the number of members. These results suggest that our finding that Zipf's law holds for the distribution of output and factor shares among US states is highly robust. For EU countries, the upward bias in the estimated slope coefficient together with the downward bias in the standard error may account for the rejection of Zipf's for physical capital in some sample years.

[Insert Table 3 about here]

Table 4 reports the results of tests of the equal-share relationship. Specifically, Table 4 reports *p*-values for testing the hypothesis of intercept homogeneity and slope homogeneity across the three share distributions in each sample year.[7] For US states, data were available for all three shares only in 1990 and 2000. In neither year can we

[7] These tests were performed by establishing, in each year, a system comprising the three share equations but without initially imposing any cross-equation parameter restrictions.

reject the hypotheses of intercept equality and slope equality, supporting the equal-share relationship for US states. Except for 1965, the results also indicate support for the equal-share relationship for EU countries.[8]

[Insert Table 4 about here]

Finally, Figures 1 and 2 provide a graphical analysis of the observed shares, the theoretically expected shares (assuming Zipf's law), and the shares predicted using the estimated rank-share equation for the output share in 2000.[9] Figure 1 for US states indicates that the distribution of actual output shares in 2000 closely follows the theoretical values, except for the first observation. Figure 2 for the EU shows a similar degree of "fit" between the three sets of shares. The differences between actual and theoretical share values (results not shown) are comparable in magnitude for the US states and EU countries.

[Insert Figures 1 and 2 about here]

There are several explanations for the observed deviation in actual share values from their theoretical values. One is that the theoretical share distribution is a steady state prediction and our sample values may not represent this ideal. Another is that our model assumes that the FIE is "closed," in that goods and factor flows arise only between FIE members. In reality, there exist important trade and factor flows between US states, and EU countries, with entities that are outside each of these defined integrated economies.

[8] Bowen et al. (2005) found that the equal-share relationship held for US states and the same 14 EU countries based on annual cross-section estimates of equations that linked output and factor shares (but not their ranks).

[9] For example, in 2000, Pennsylvania ranked 6th among US states in terms of output shares; its actual share was 0.0402 while its theoretical share is 0.0369. For the EU, in 2000 the Netherlands ranked 6th in terms of output shares; its actual share was 0.0469 while its theoretical share is 0.0513.

5 Summary and Conclusions

This paper demonstrated that among members of a fully integrated economy (FIE), in which there is free exchange of goods and factors and where members share the same production technology, each member's share of total FIE output will equal its shares of total FIE physical capital and total FIE human capital. This result is called the equal-share relationship. In this setting, it was then argued that the growth in any member's share can be considered to derive from a random process. If this process is common across FIE members then the limiting distribution of each share across FIE members will take the form of a rank-share distribution that exhibits Zipf's law. Given this, it was then demonstrated that the theoretically expected share values of each FIE member are deterministic, and depend only on the number of FIE members. Finally, by the equal-share property, these theoretically expected share values would be identical for output and productive factors.

We examined empirically for evidence that the distribution of output and factor shares exhibit Zipf's law with respect to two "integrated economies": the 51 US states and 14 EU countries. Our results indicated that Zipf's law holds among US states for the distribution of output, physical capital and human capital shares, and also that these output and factor share distributions are identical, confirming the equal-share relationship for US states.

For the EU countries, the results indicated mixed support of Zipf's law. The results generally supported Zipf's law for years after, but not before, 1985. These findings suggest convergence toward Zipf's law for EU countries, perhaps reflecting the more recent efforts by EU member states to further reduced trade and factor mobility barriers among themselves.

The finding that Zipf's law holds empirically for the distribution of output and factor shares suggests a constraint on the set of admissible growth models that may be used to explain the growth experiences of members of an integrated economy. In particular, the empirical significance of the equal-share relationship implies that this relative growth performance will be largely random, and hence strongly dependent on particular states of nature. Such randomness will be more true the greater the extent of economic integration among members, as perhaps most exemplified by the integrated economy comprising US states. Hence, it is likely to be more true the more harmonized are education systems and fiscal codes, when members they do not run independent monetary policies, and when industrial policies are quickly imitated across members. Finally, while there may be several explanations for the empirical finding that the distribution of output and factor shares fit a power law, the evidence on the empirical significance of Zipf's law suggests that models of the growth of members of integrated economies should satisfy a main underlying assumption, namely, that the growth process is random and homogeneous across members.

References

Baldwin, R., Martin, P., 2004, "Agglomeration and Regional Growth," in Henderson, J.V. and Thisse, J.-F. (Eds.), *Handbook of Regional and Urban Economics-Volume IV* (Amsterdam: North-Holland), Chapter 2.

Barro, R. J., 1999, "Determinants of Democracy," *Journal of Political Economy* 107, S158-S183.

Barro, R. J., Lee, J.W., 1993, "International Comparisons of Educational Attainment," *Journal of Monetary Economics* 32, 363-394.

Barro, R. J., Lee, J.W., 1996, "International Measures of Schooling Years and Schooling Quality," *American Economic Review* 86, 218-223.

Barro, R. J., Lee, J.W., 2000, "International Data on Educational Attainment: Updates and Implications," Center for International Development Working Paper 42, Harvard University.

Black, D., Henderson, V., 1999, "A Theory of Urban Growth," *Journal of Political Economy* 107, 252-284.

Bowen, H. P., Munandar, H., and Viaene, J.-M., 2005, "The Limiting Distribution of Production in Integrated Economies: Evidence from US States and EU Countries," Tinbergen Institute Discussion Paper 05-045/2.

Brakman, S., Garretsen, H., van Marrewijk, C., 2001, *An Introduction to Geographical Economies* (Cambridge: Cambridge University Press).

Devereux, M.B., Lapham, B.J., 1994, "The Stability of Economic Integration and Endogenous Growth," *Quarterly Journal of Economics* 109(1), 299-308.

Eaton, J., Eckstein, Z., 1997, "Cities and Growth: Theory and Evidence from France and Japan," *Regional Science and Urban Economics* 27, 443-474.

Easterly, W., Levine, R., 1998, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics* 112, 1203-1250.

Fujita, M., Krugman, P.R., Venables, A. J., 1999, *The Spatial Economy: Cities, Regions, and International Trade* (Cambridge, Mass.: MIT Press).

Gabaix, X., 1999, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics* 114(4), 739-767.

Gabaix, X., Ioannides, Y. M., 2004, "The Evolution of City Size Distributions," in Henderson, J.V. and Thisse, J.-F. (Eds.), *Handbook of Regional and Urban Economics-Volume IV* (Amsterdam: North-Holland), Chapter 7.

Garofalo, G., Yamarik, S., 2002, "Regional Convergence: Evidence from a New State-by-State Capital Stock Series", *Review of Economics and Statistics* 82, 316-323.

Gibrat, R., 1931, *Les Inegalites Economiques*, (Paris: Librairie du Recueil Sirey).

Grossman, G.M., and Helpman, E., 1991, *Innovation and Growth in the Global Economy* (Cambridge: MIT Press).

Hall, R.E., Jones, C.I., 1999, "Why Do Some Countries Produce So Much More Output Per Worker than Others?," *Quarterly Journal of Economics* 114, 83-116.

Heston, A., Summers, R., 1991*a*, "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988," *Quarterly Journal of Economics* 106, 327-368.

Heston, A. Summers, R. 1991*b*, *Penn World Table Version 5.6*, Center for International Comparisons at the University of Pennsylvania (CICUP).

Heston, A., Summers, R., Aten, B., 2002, *Penn World Table Version 6.1*, Center for International Comparisons at the University of Pennsylvania (CICUP).

Mankiw, N.G., Romer, D., Weil, D.N., 1992, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* 107, 407-437.

Munnell, A., 1990, "Why Has Productivity Growth Declined? Productivity and Public Investment," *New England Economic Review,* 407-438.

Newey, W.K., West, K., 1987, "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703-708.

Rajan, R.G., Zingales, L., 1998, "Financial Dependence and Growth," *American Economic Review* 88, 559-586.

Ramey, G., Ramey, V. A., 1995, "Cross-country Evidence on the Link between Volatility and Growth," *American Economic Review* 85, 1138-1151.

River-Batiz, L., Romer, P., 1991, "Economic Integration and Endogenous Growth", *Quarterly Journal of Economics* 106(2), 531-555.

Sachs, J. D., Warner, A., 1995, "Economic Reform and the Process of Global Integration," *Brookings Papers on Economics Activity*, 1-95.

Timmer, M. P., Ypma, G., van Ark, B., 2003, "IT in the European Union: Driving Productivity Divergence?" *GGDC Research Memorandum* GD-67, University of Groningen.

US Bureau of Economic Analysis, 2002, *Fixed Assets and Consumer Durable Goods for 1925-2001.*

Viaene, J.-M., and Zilcha, I., 2002, "Capital Markets Integration, Growth and Income Distribution," *European Economic Review* 46, 301-327.

Appendix – Data Methods and Sources

The output for each of the 51 US states is measured by real gross state product as reported by the US Bureau of Economic Analysis (BEA).[10] These data were available yearly from 1990 to 2000.

Estimates of state physical capital stocks were derived from BEA (2002) estimates of the total US physical capital stock in each of nine one-digit industrial sectors comprising all economic activity.[11] These national physical capital stocks in each industry were allocated to each state by multiplying an industry's total capital stock[12] by that industry's contribution to a state's total income.[13] These industry capital stock estimates were then summed, for each state, to obtain an estimate of a state's total stock of physical capital.[14] The calculation performed for each state at each time t can be expressed algebraically as

$$k_i(t) = \sum_{j=1}^{9} \left[K_j(t) \left(y_{ij}(t) / Y_i(t) \right) \right]$$

In this equation, $k_i(t)$ is the stock of physical capital in state i, $y_{ij}(t)$ is value added by industry j in state i ($i = 1\ldots51$), $Y_i(t)$ is state i's total value added, and $K_j(t)$ is the national level stock of physical capital in industry j ($j = 1,\ldots, 9$). This procedure assumes that the capital-to-output ratio within an industry j (i.e., $k_{ij}(t)/y_{ij}(t)$) is the same across US states, that is, $k_{ij}(t)/y_{ij}(t) = K_i(t)/Y_i(t)$. In turn, this assumption implies that an

[10] Data on gross state product available at http://www.bea.doc.gov/bea/regional/gsp

[11] The sectors (BEA code) are Farming (81), Agricultural services, forestry, fishing & other (100); Mining (200); Construction (300); Manufacturing (400); Transportation(500); Wholesale and retail trade (610); Finance, insurance and real estate (700); and Services (800).

[12] Data on state physical capital stocks by industry were taken from US Fixed Assets Tables, available at http://www.bea.doc.gov/bea/dn/faweb

[13] Data on annual state personal income (value added) available at http://www.bea.doc.gov/bea/regional/spi

[14] This procedure follows that used by Munnel (1990) and Garofalo and Yamarik (2002).

industry is in a common steady state across all US states.[15] For example, the agricultural sector in Texas is in the same steady state as its counterpart in Oregon, and the manufacturing sector in Pennsylvania is in the same steady state as its counterpart in Ohio.[16] The constructed physical capital data are from 1990 to 2000, on a yearly basis.

State human capital stocks were derived from data on educational attainment in each state taken from the US Bureau of the Census.[17] Since census data on educational attainment are only available every 10 years, this limits the data on stocks of human capital to the two years 1990 and 2000.

For the countries comprising the EU, total output is measured by a country's real gross domestic product (GDP) derived from the data on real GDP per capita (base year = 1996) and population in Penn World Tables 6.1 (Heston, Summers and Aten, 2002).[18] The output data were available annually from 1960 to 2000.

Data on EU physical capital stocks were derived from Penn World Tables 5.6 (Heston and Summers, 1991and 1991b) which reports four data series for each country: (1) population, (2) physical capital stock per worker, (3) real GDP per capita and (4) real GDP per worker.[19] The physical capital stocks for each country were constructed as the product of the first three series divided by the last series. The data covers the period 1965-1990. Physical capital stock data for EU countries were also available

---

[15] If a sector is converging towards its steady state, the output-to-capital ratio would be below its steady-state value. This only poses a problem if the initial output-to-capital ratios vary across US states. If the ratios do vary, the procedure would allocate too much to those states further from steady-state and too little to those states closer to their steady state.

[16] If a sector has a different steady state, and hence a different capital-to-output ratio, the procedure will allocate too much to states with lower ratios and too little to states with higher ratios. However, this possibility is unlikely if competition lead firms in all states to adopt the best available production technology.

[17] Decennial Census Dataset available at http://factfinder.census.gov

[18] Penn World Tables 6.1 available at http://datacentre2.chass.utoronto.ca/pwt

[19] Penn World Tables 5.6 available at http://datacentre2.chass.utoronto.ca/pwt56

from Timmer et al. (2003)[20] covering period 1980-2000.[21] These data sources were combined to have physical capital stock data in each of seven years from 1965 to 2000.[22]

Each EU country's stock of human capital stock is measured by multiplying the percentage of a country's population having at least a secondary level of education with the country's total population. Data on the rate of educational attainment for each country were taken from Barro and Lee (1993, 1996, and 2000).[23] Data on a country's population were from Heston, Summers and Aten (2002). Since the data on rates of educational attainment are only available every 5 years, the data sample is limited to five-year intervals from 1960 to 2000. Following this constraint, the output and physical capital stocks were also obtained in five-year intervals.

The 14 EU countries are: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain, Sweden and United Kingdom.[24]

---

[20] Physical capital database available at http://www.ggdc.net/dseries/growth-accounting.shtml

[21] The series forms the source of the OECD productivity database. See e.g., Schreyer et al. (2003).

[22] Estimation was conducted using both sets of data for EU countries. No qualitative difference in results was found for the years in which data were available from both sources (i.e., 1980, 1985 and 1990). For these three years we therefore report only the results using the capital stock data from Timmer et al. (2003).

[23] Other studies using the Barro-Lee data include Rajan and Zingales (1998), Ramey and Ramey (1995), Barro (1999), Easterly and Levine (1998), Hall and Jones (1999) and Sachs and Warner (1995).

[24] Luxembourg is excluded for lack of data on human capital. Given the small scale of Luxembourg's economy relative to other EU countries this omission is unlikely to affect the EU results.

Table 1 - OLS Estimates of Rank-Share Relationships for US States

Variable (Share)	Year	Intercept (*q*)	Slope (*b*)	Adjusted R^2
Output (n = 51)	1990	-1.179 (0.222)	1.101 (0.073)	0.887
	1991	-1.194 (0.222)	1.093 (0.073)	0.884
	1992	-1.199 (0.227)	1.090 (0.075)	0.883
	1993	-1.207 (0.234)	1.085 (0.077)	0.881
	1994	-1.208 (0.242)	1.084 (0.079)	0.876
	1995	-1.209 (0.242)	1.083 (0.079)	0.874
	1996	-1.205 (0.242)	1.085 (0.079)	0.872
	1997	-1.192 (0.245)	1.091 (0.080)	0.868
	1998	-1.173 (0.246)	1.100 (0.081)	0.868
	1999	-1.168 (0.244)	1.103 (0.080)	0.866
	2000	-1.164 (0.238)	1.106 (0.078)	0.868
Physical Capital (n = 51)	1990	-1.199 (0.227)	1.092 (0.075)	0.892
	1991	-1.207 (0.230)	1.089 (0.076)	0.891
	1992	-1.200 (0.235)	1.092 (0.077)	0.892
	1993	-1.197 (0.239)	1.093 (0.079)	0.890
	1994	-1.196 (0.247)	1.092 (0.081)	0.884
	1995	-1.173 (0.254)	1.102 (0.083)	0.879
	1996	-1.168 (0.255)	1.105 (0.083)	0.878
	1997	-1.126 (0.261)	1.125 (0.086)	0.870
	1998	-1.126 (0.257)	1.126 (0.084)	0.876
	1999	-1.108 (0.259)	1.135 (0.084)	0.875
	2000	-1.093 (0.258)	1.143 (0.083)	0.880
Human Capital (n = 51)	1990	-1.244 (0.252)	1.064 (0.082)	0.854
	2000	-1.264 (0.268)	1.054 (0.088)	0.839

Notes: Standard error in parentheses; all intercept coefficients are significantly different from zero at 1%; all slope coefficients are not significantly different from one at 5%.

Table 2 - OLS Estimates of Rank-Share Relationships for EU Countries

Variable (Share)	Year	Intercept (*q*)	Slope (*b*)	Adjusted R^2
Output (n = 14)	1960	-0.645 (0.334)	1.461 (0.156) $^{+}$	0.908
	1965	-0.665 (0.345)	1.435 (0.165) $^{+}$	0.889
	1970	-0.699 (0.361)	1.406 (0.173) $^{+}$	0.867
	1975	-0.742 (0.458)	1.366 (0.209)	0.859
	1980	-0.755 (0.430)	1.357 (0.197)	0.870
	1985	-0.763 (0.427)	1.354 (0.195)	0.872
	1990	-0.772 (0.430)	1.346 (0.195)	0.872
	1995	-0.777 (0.420)	1.343 (0.182)	0.878
	2000	-0.857 (0.392) *	1.272 (0.171)	0.885
Physical Capital (n = 14)	1965	-0.816 (0.440)	1.293 (0.232)	0.851
	1970	-0.825 (0.402)	1.275 (0.212)	0.858
	1975	-0.836 (0.361) *	1.262 (0.195)	0.858
	1980	-0.760 (0.350)	1.332 (0.177)	0.828
	1985	-0.732 (0.289) *	1.358 (0.142) $^{+}$	0.870
	1990	-0.670 (0.435)	1.418 (0.215)	0.873
	1995	-0.632 (0.321)	1.457 (0.154) $^{+}$	0.908
	2000	-0.658 (0.388)	1.431 (0.176) $^{+}$	0.904
Human Capital (n = 14)	1960	-0.147 (0.492)	2.103 (0.302) $^{++}$	0.791
	1965	-0.343 (0.327)	1.890 (0.169) $^{++}$	0.880
	1970	-0.529 (0.213) *	1.639 (0.110) $^{++}$	0.865
	1975	-0.642 (0.177) **	1.518 (0.080) $^{++}$	0.928
	1980	-0.683 (0.182) **	1.433 (0.071) $^{++}$	0.933
	1985	-0.747 (0.133) **	1.409 (0.049) $^{++}$	0.945
	1990	-0.895 (0.235) **	1.241 (0.125)	0.912
	1995	-0.897 (0.247) **	1.225 (0.128)	0.912
	2000	-0.905 (0.237) **	1.215 (0.120)	0.919

Notes: Standard error in parentheses;
* significantly different from zero at 5%;
** significantly different from zero at 1%;
$^{+}$ significantly different from one at 5%;
$^{++}$ significantly different from one at 1%.

Table 3 - Monte Carlo Analysis of OLS Estimates of the Relationship between the Share and Rank of Shares

Statistic	Number of Integrated Economy Members (N)				
	14 (EU)	20	51 (US States)	100	200
1) OLS slope ($E(\hat{b})$)	1.172	1.143	1.081	1.054	1.034
2) Bias ($E(\hat{b}) - 1$)	0.172	0.143	0.081	0.054	0.034
3) Prob($\hat{b} > 1$)	0.629	0.632	0.634	0.629	0.619
4) Average OLS std. error	0.089	0.065	0.029	0.016	0.009
5) True std. error of $\hat{b}$	0.401	0.329	0.200	0.142	0.100
6) True 95% confidence interval for OLS slope	[0.544, 2.104]	[0.610, 1.893]	[0.734, 1.517]	[0.802, 1.354]	[0.851, 1.241]

Notes: Each column based on 100,000 Monte Carlo simulations (each with N observations) drawn from an exact power law with coefficient 1 (Zipf's Law). This involved drawing N i.i.d. variables v_i uniformly distributed in the interval [0, 1] and then constructing sizes $L_i = 1/v_i$. The L_i were then normalized into shares S_i which were then ordered and assigned a rank value R_i. 100,000 OLS regressions were then performed using the specification $\log(S_i) = q + b \log(1/R_i) + u_i$. Row 1 shows the average value of the OLS slope estimates across the 100,000 regressions for sample size N. Row 2 measures the extent of the bias in the estimated slope from its theoretical value of unity. Row 3 gives the proportion of OLS estimated slopes whose value exceeded unity; a value above 0.5 indicates an upward bias of the OLS slope estimate. Row 4 gives the average value of the OLS standard error across the 100,000 regressions. Row 5 gives the standard deviation of the 100,000 OLS slope estimates; this value estimates the true standard error of the sampling distribution of the OLS slope estimate. Row 6 shows the range that included 95% of 100,000 simulated OLS slope estimates.

Table 4 - Results Testing the Equal-Share Relationship

Integrated Economy	Year	*p*-values for testing across-equation homogeneity of	
		intercepts	slopes
US States	1990	0.9680	0.9014
	2000	0.8241	0.5964
European Union	1965	0.6063	0.0445*
	1970	0.8011	0.2797
	1975	0.8619	0.3655
	1980	0.9689	0.8461
	1985	0.9969	0.9305
	1990	0.8111	0.6034
	1995	0.7124	0.3697
	2000	0.7291	0.4072

* Cross-equation homogeneity is rejected at 5% level.

Figure 1 – Actual, Estimated and Theoretical Share-Rank Distribution of Output Shares for US States, 2000

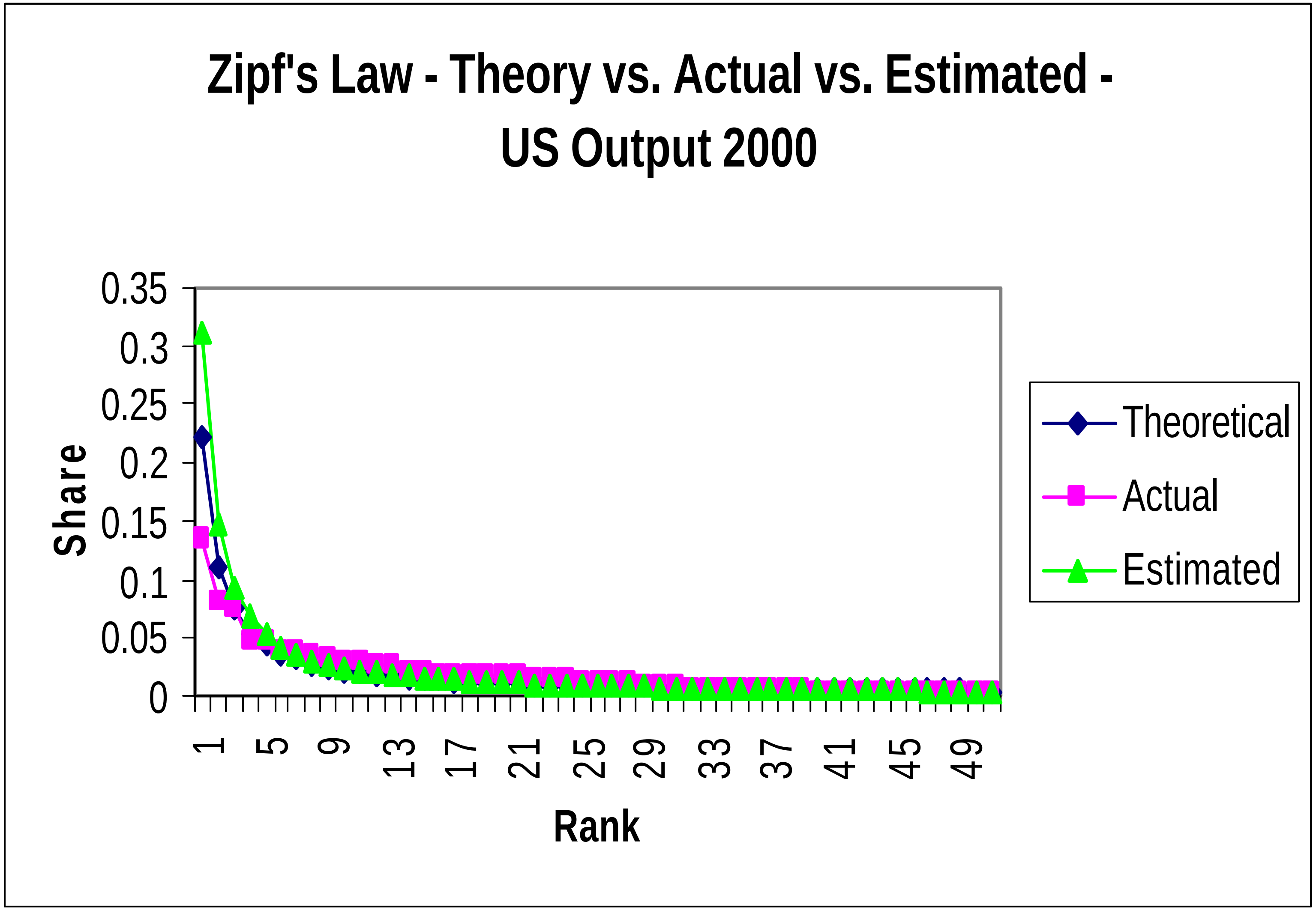

Figure 2 - Actual, Estimated and Theoretical Rank-Share Distribution of Output Shares for EU Countries, 2000