Rank-size distributions for banks:

A cross-country analysis

Roy Cerqueti*[1,2], Claudio Lupi[3], Filomena Pietrovito[3], and Alberto Franco Pozzolo[4]

[1]Sapienza University of Rome – Department of Social and Economic Sciences

[2]London South Bank University – School of Business

[3]University of Molise – Department of Economics

[4]Roma Tre University – Department of Law

Abstract

Rank-size analysis is a powerful methodological instrument for describing the main features of a unified system by moving from an available sample of its disaggregated elements. We here study the rank-size relationship of banks' total assets in a sample of 13 advanced economies. We show that a standard power-law function is unable to provide an adequate fit of the data, while a Discrete Generalized Beta Distribution provides a statistically satisfactory fit, being able to capture the behaviour of the distribution at low and high ranks. We then analyse the relationship between the parameters of the Discrete Generalized Beta Distribution estimated in each country and the main features of the country's banking industry. Our results point to a connection with the degree of development of the banking sector and with some regulatory and institutional characteristics, but no relationship with riskiness.

Keywords: Rank-size relationship; Bank assets; Discrete Generalized Beta Distribution; Bank regulation.

*Corresponding author:

Department of Social and Economic Sciences, Sapienza University of Rome. Piazzale Aldo Moro, 5 - I-00185, Rome, Italy. Email: `roy.cerqueti@uniroma1.it`

and

School of Business, London South Bank University. London SE1 0AA, UK.

1 Introduction

Rank-size analysis allows to build a theoretical system based on an observed sample of data, capable of describing the observed connections between sizes and related ranks. The overall system is modelled through a calibrated curve.

A major strength of this approach is its ability to describe empirical regularities with a very simple "law". For this reason, the framework described by the selected family of curves is a useful representation of the sample relationships only if the best-fit exercise is statistically sounding. In this case, the fitted curve is able to provide a convincing description of the system created by the observed sample – based on the functional properties of the family of curves involved in the procedure. The best-fit curve gives therefore a formal view of some specific data regularities of the empirical sample, and hopefully of the underlying population. Recent methodological contributions to the rank-size analysis can be found in Ausloos and Cerqueti (2016) and Reed (2001).

The relevance of the rank-size theory for applications mirrors in the popularity of such a framework in different fields, ranging from metrics of science and bibliometrics (see, e.g., Ausloos, 2013, 2015; Ausloos et al., 2016; Chatterjee et al., 2016; Mansilla et al., 2007) to seismology (see, e.g., Ficcadenti and Cerqueti, 2017), text analysis and linguistics (see, e.g., Ficcadenti et al., 2019, 2020; Montemurro, 2001) and, of course, economics (see, e.g., Brakman et al., 1999; Cerqueti and Ausloos, 2015; Córdoba, 2008; Dimitrova and Ausloos, 2015; Giesen and Südekum, 2011).

In the peculiar context of management, studies focusing on the rank-size relationship of firms have a very long tradition. A largely studied stylized fact in the literature on firms' size is Gibrat's law (Gibrat, 1931), according to which firm sizes follow a lognormal distribution, and the growth rate of firms is independent of their size distribution.[1] Since the seminal contribution of Gibrat (1931), a large strand of literature has devoted its attention to uncovering empirical regularities in firm size distribution.

Partly linked to Gibrat's law, another largely recognized stylized fact is that the distribution of firms according to their size is not well fitted by a lognormal distribution, but follows instead a power law. The seminal paper by Axtell (2001), for instance, shows that Zipf's distribution (i.e., a power law distribution with coefficient equal to -1) fits well the

[1]For a review of historical literature see Sutton (1997); for a more updated contribution, Coad (2009).

data above a certain minimum size threshold (a result also confirmed by Zhang et al., 2009, for Chinese firms).[2]

Following the seminal contribution by Axtell (2001), several papers have tested whether firm sizes follow a power law distribution, verifying in addition if the coefficient is equal to -1, i.e., whether Zipf's law is sustained. Results obtained so far appear to be mixed. A number of papers have confirmed that the distribution of firms according to their size is highly skewed and that the upper tail satisfies the characteristics of a Pareto distribution (see, for instance, Cefis et al., 2009; Growiec et al., 2008); in this respect, for earlier evidence, we refer to Simon and Bonini (1958). But other economic characteristics have also been found to alter significantly the firm size distribution. Gaffeo et al. (2003), for example, find only weak evidence that firm rank-size distribution follows Zipf's law, showing in addition that the distribution depends on the phase of the business cycle. Crosato and Ganugi (2007) show that truncated distributions fit better the distribution of Italian firms and, most interestingly, that the estimated coefficients are different across sectors and between innovative and non-innovative firms. Also international trade has been shown to impact on the rank-size distribution of firms: di Giovanni et al. (2011, p. 46) show that "the size distribution for exporting firms is systematically more fat-tailed than the size distribution of the non-exporting ones", as witnessed by the fact that the coefficient of the power law is lower for exporting firms than for non-exporting firms. Segarra and Teruel (2012) add an additional dimension to the previous results, showing that also the size of the sample impacts on the estimated parameters of the power law, a feature that will be crucial also in our analysis.[3]

While analyses focusing on specific industries have already been proposed (see, e.g., Ganugi et al., 2004), to the best of our knowledge no studies have focused on the banking

[2] Interestingly, a seminal theoretical contribution by Gabaix (1999) shows that the two stylized facts highlighted by Gibrat (1931) and Axtell (2001) can be reconciled if the unit of analysis – in his case the size of a city – follows a process of proportional growth; a parallel line of research has studied the relationship between the idiosyncratic shocks affecting a single economic unit and the aggregate trend of the entire economic system, as a function of the characteristics of the size distribution (see, for example, Gabaix, 2011).

[3] A parallel strand of literature has focused on entry and exit conditions and selection and imitation among new entrants and incumbents; Luttmer (2007) in particular shows that the selection mechanism tends to improve aggregate productivity at a fast rate if entry and imitation are easy and concludes, from the data analysis, that the firm size distribution is close to Zipf's law if entry is difficult.

sector. This is all the more surprising, since the financial crisis of 2007–2008 has made it evident that there are very significant differences between the largest banks and those of medium and small size, in terms of activities carried out, risk appetite, resilience in crisis situations, systemic relevance (see for example Fahlenbrach et al., 2012; Demirgüç-Kunt and Huizinga, 2013). This can obviously have a major impact also on the characteristics of the economic and financial systems of the countries where these banks operate.

In this paper, we try to fill the gap in the literature on rank-size relationships by studying the banking industry in a sample of 13 advanced economies. Our main results show that a standard power-law function is unable to provide an adequate fit of the data, with R^2s that in some cases fall below the 90% threshold, an unacceptable standard in this field of literature. We then extend our analysis using a Discrete Generalized Beta Distribution. Having one more free parameter, and controlling for the size of the sample, this distribution allows a better fit to the data, especially at lowest and highest ranks of the distribution, where the simple power law was providing the worst fit. The results are in this case encouraging, with the lowest R^2 being 0.975.

Leveraging on these results, we then provide a tentative analysis of the possible determinants of the different rank-size curves estimated in each country. Worldwide, the average size of financial intermediaries differs widely across countries (Barba Navaretti et al., 2019). A key candidate to explain these differences is obviously the role of regulation. For example, in the United States the prohibition imposed by the McFadden Act in 1927 on opening branches in States other than that of settlement of the bank effectively prevented the development of large intermediaries, and only the Riegle Neal Act of 1994 removed this restriction, partly in the face of growing competition from foreign banks. Bank size is also related to another key characteristic of the industry, that is its riskiness, obviously a primary interest after the 2007–2008 crisis (see, e.g., Laeven et al., 2016).

To study the connections between the shape of the rank-size curves and other characteristics of the banking industry we analyse the relationship between the parameters of the Discrete Generalized Beta Distribution estimated in each country and the main features of its banking industry. Our results point to a connection of the estimated curves with the degree of development of the banking sector and with some regulatory and institutional characteristics. On the contrary, we find no relationship with riskiness.

The rest of the paper is organized as follows. Section 2 presents the dataset used in our analysis and discusses its main characteristics through some descriptive statistics. Section 3 offers the analysis of the ranked data by the application of the family of the power law curves. It shows that such an analysis is not satisfactory under a statistical perspective, although some features of the results are worth a few comments. More appealing is the analysis based on the Discrete Generalized Beta Distribution, proposed and extensively discussed in Section 4. Section 5 moves from the results obtained in Section 4 and compares the fitted parameters at the country level with the main banking systems parameters. The last section offers some conclusive remarks, by tracing also lines for future directions of research.

2 Data and descriptive statistics

To analyse the distribution of bank size, we use BankScope,[4] a unique source of bank-level information for different countries provided by Bureau Van Dijk, used by many leading financial institutions and central banks for cross-country studies and policy decisions (e.g., Demirgüç-Kunt and Detragiache, 1998). The information contained in the database includes detailed spreadsheet data (balance sheet and income statements), ownership information (shareholders and subsidiaries), specialization and ratings. Although BankScope is the most largely used database for empirical studies in international banking, it does not cover the full population of banks in each nation, and the coverage itself is not uniform across countries, with possible under-representation of smaller banks. For this reason, we focus on a smaller set of countries for which BankScope collects data on a sufficiently large number of banks.[5]

For the purpose of our analysis, we are interested in descriptive individual information, such as the size, measured by the value of its total assets, the trade description, providing the type of activity carried out by the bank, the consolidation code, which indicates the level of consolidation for the different financial statements of banking groups, and the country where it is registered.

[4]Data from BankScope are not publicly accessible and the database has been used under license for this paper.

[5]Reassuringly, Duprey and Lé (2016) compare the total assets of financial intermediaries included in BankScope with those from the International Monetary Fund (IMF) and show that BankScope is highly representative for almost all European banking systems, and Barba Navaretti et al. (2019) find similar results for non-European countries.

Starting with a sample of $25,420$ banks recorded in 2016 for 33 countries, we operate a selection both at the bank and at the country-level.[6] At the bank level, we exclude holding companies integrating the statement of their subsidiaries and declaring that the unconsolidated companion is also included in the database, in order to avoid redundant observations.[7] We then select 15 types of banks, according to their trade description and excluding central banks, clearing institutions and non-banking credit institutions.[8] Finally, we exclude countries with less than a critical number of banks, that we set at $n = 100$.[9]

We are thus left with a sample including 4,722 banks from 13 countries, as shown in Table 1. In the same table we report the main descriptive statistics on banks' total assets: the country average ranges between 439 millions USD in Poland and 43.2 billion USD in Canada. The sample variability of total assets by country, as measured by the coefficient of variation, mostly ranges from 2.55 (Italy) to 6.07 (Finland), with the exception of Germany, where it reaches a peak at 14.56. Taken all together, the banks in the sample are very heterogeneous, with total assets spanning values between 0.6 millions USD (for the smallest Mexican bank) up to over 2,233 billions USD (for the largest German bank); 50% of the banks considered in this study hold assets in the range $206 - 2{,}904$ millions USD.

To analyse how the estimated parameters of the Discrete Generalized Beta Distribution depend on the regulatory and institutional characteristics of the banking sector prevailing in each country, we focus on five main sets of such features: financial development, riskiness, ownership, regulation and competitiveness.

Financial development is measured by means of two indicators, both from the financial

[6]Among all available years, we have chosen the one with the highest number of banks over the sample of countries.

[7]We make this choice because we believe that the decision to keep a subsidiary as a separate entity implies a degree of independence that justifies considering it as a separate unit in the rank-size distribution.

[8]The considered types of banks are: bank holding and holding companies (when no companion subsidiary data are available), cantonal banks, commercial banks, central cooperative banks, cooperative banks, foreign banks, investment banks, security houses, Islamic banks, private banks, Raiffeisen Banks, real estate/mortgage banks, regional banks, savings banks, shinkin banks (credit association).

[9]Bootstrap estimates of the variance and the 95% confidence intervals of the calibrated parameters of the rank-size relationship, unreported for space reasons but available upon request, show a large variability when the sample size ranges in the interval $n \in [20, \ldots, 200]$. The choice of $n = 100$ is a focal point trading-off the precision in the estimates and our aim to include a number of countries sufficient to study the cross-country variability of the estimated rank-size parameters.

Country	ISO	mean	std.dev.	c.v.	skewness	kurtosis	min	max	q25	q75	Obs.
Austria	AT	1,011	5,913	5.85	13.4	214.3	4.6	108,112	75	406	551
Canada	CA	43,236	151,804	3.51	4.4	22.5	78.7	880,592	729	7,527	102
Finland	FI	2,249	13,650	6.07	8.8	82.6	31.3	140,983	82	382	164
France	FR	42,585	220,743	5.18	6.4	45.4	28.4	1,816,054	540	5,708	122
Germany	DE	4,144	60,333	14.56	35.6	1313.9	9.8	2,233,398	266	1,962	1421
Italy	IT	1,636	4,625	2.83	6.8	56.0	9.0	45,292	236	1,277	353
Japan	JP	11,009	28,060	2.55	5.6	38.4	399.4	234,897	1,472	8,026	262
Mexico	MX	1,726	9,523	5.52	7.9	71.1	0.6	100,095	6	75	206
Poland	PL	439	1,260	2.87	4.4	25.8	11.1	9,496	33	129	131
Portugal	PT	576	1,995	3.46	6.9	54.7	27.8	17,603	100	326	105
Switzerland	CH	6,332	24,790	3.91	8.0	77.7	39.5	289,859	528	1,628	307
United Kingdom	UK	13,941	81,744	5.86	9.0	93.1	7.8	936,428	212	2,256	187
United States	US	27,743	146,288	5.27	11.3	142.0	311.3	2,082,803	2,431	9,356	811
All countries		10,039	83,742	8.34	18.9	411.0	0.6	2,233,398	206	2,904	4,722

Table 1: Banks' total assets sample statistics (values in millions USD): c.v. is the coefficient of variation; q25 and q75 are the first and third quartile, respectively; Obs. is the number of observations by country. ISO is the official ISO Alpha 2 abbreviation of the country name (also used in Figures 3–5).

structure database produced by the World Bank:[10] the share of deposit money bank assets over GDP and the share of deposits in deposit money banks over GDP. Banking system riskiness is measured by the bank Z-score, obtained by the same database.[11] Data on ownership are from Caprio et al. (2007) and include the fraction of the banking systems' assets held by banks that are 50% or more, respectively, foreign owned and government owned. Indicators on banking regulation are from the Bank Regulation and Supervision Survey (BRSS), also produced by the World Bank.[12] We consider two different characteristics: the extent to which banks may engage in securities, real estate and insurance activities, and the effectiveness of private monitoring. Higher values of these indexes indicate greater stringency. Competitiveness is measured with the Boone index, which is calculated as the elasticity of profits to marginal costs and takes higher values when competition in the banking sector is lower. Data are also in this case from the financial structure database produced by the

[10] Data are described in Beck et al. (2000) and are accessible at: .

[11] Z-score is estimated as: $\left(\frac{ROA+equity}{assets}\right) \Big/ sd(ROA)$, with $sd(ROA)$ the estimated standard deviation of ROA.

[12] Data are accessible at: https://datacatalog.worldbank.org/dataset/bank-regulation-and-supervision-survey. A comprehensive description of this database is provided by Barth et al. (2013).

World Bank. Statistics on per capita GDP, calculated in purchasing power parity (PPP) USD, and on savings ratios, the ratio between gross national savings and GDP, are produced by the IMF.[13] Finally, bank orientation is the ratio of total assets of deposit money banks and stock market capitalization, also constructed from the financial structure database.[14]

Table 2 reports the descriptive statistics of the country-level banking system indicators. Measures of financial development and riskiness have a high variability in the countries considered in the analysis. The share of deposit money bank assets over GDP is on average 111%, but ranges from 40% in Mexico to 177% in Switzerland. The ratio of deposits to GDP is on average 98%, with values between 29% in Mexico and 218% in Japan. The average value of the Z-score, our measure of riskiness, is 17%, but values range from 8% in Poland to 29% in the United States. Foreign owned banks hold on average 28% of bank assets in our sample, with values ranging between 6% in Japan to 80% in Mexico. Assets of banks owned by local or central governments are on average 10% of total bank assets, and range from 0% in Canada, Finland and the United States to 40% in Germany. Within our sample, regulations are more stringent in Poland (with an index of 11) and less so in Switzerland (with an index of 3) – in a scale ranging from 3 to 12. The strength of private monitoring shows values ranging between 6 and 11 in our sample – in a scale ranging from 0 to 12. The Boone indicator has an average value of -0.055 and ranges from -0.120 in Poland to 0.004 in Japan. Per capita GDP has an average value of 44,446 PPP USD and ranges from 28,740 USD in Mexico to 66,908 USD in Switzerland. The savings ratio is on average 22% and ranges from 12% in the United Kingdom to nearly 33% in Switzerland. Finally, the ratio of deposit money bank assets to stock market capitalization, a measure of whether a country is bank-oriented or market-oriented (see, e.g., Demirgüç-Kunt and Levine, 2004) has a large variability, ranging from 0.44 in the United States to 4.72 in Portugal, and an average value of 1.92.[15]

[13]Data on GDP are publicly accessible from the World Economic Outlook Database, available at: https://www.imf.org/en/Publications/WEO/weo-database/2020/April; data on savings ratios are accessible from the International Financial Statistics at: https://data.imf.org/?sk=4c514d48-b6ba-49ed-8ab9-52b0c1a0179b.

[14]https://www.worldbank.org/en/publication/gfdr/data/financial-structure-database.

[15]This measure is not available for Finland, Italy, the United Kingdom.

Variable	mean	std.dev.	min	max	q25	q75
Bank assets / GDP	110.938	38.572	39.935	176.660	93.414	134.521
Bank deposits / GDP	97.637	50.335	29.468	218.225	79.016	112.750
Z-score	17.406	6.184	8.471	29.083	13.002	21.240
Foreign banks	27.713	27.058	6.100	80.000	8.500	54.200
Government owned banks	9.456	12.386	0.000	39.990	0.010	11.615
Banking restrictions	6.385	2.329	3.000	11.000	5.000	8.000
Private monitoring	8.538	1.450	6.000	11.000	8.000	10.000
Boone indicator	-0.055	0.041	-0.120	0.004	-0.097	-0.023
GDP per capita	44,446	12,686	19,633	66,908	39,937	52,097
Savings ratio	22.019	5.596	12.192	32.903	19.061	26.915
Bank orientation	1.924	1.335	0.444	4.718	1.179	2.470

Table 2: Cross-country descriptive statistics of the banking system indicators. *Bank assets / GDP*, *Bank deposits / GDP*, *Z-score*, *Foreign banks*, *Government owned banks* and *Savings ratio* are percentages. *Banking restrictions*, *Private monitoring* are indexes ranging, respectively between 3–13 and 0–12. *Boone indicator* is an elasticity. *GDP per capita* is in PPP USD. *Bank orientation* is the ratio of total assets of deposit money banks and stock market capitalization.

3 The power law-based rank-size analysis

In line with the standard rank-size theory, we present first a best-fit procedure based on a power-law function of the type

$$f(r) = \frac{A}{r^{\alpha}}, \tag{1}$$

where r is the rank, whereas A and α are positive parameters to be estimated.

Function f in (1) is convex and strictly decreasing, with an asymptotic convergence to zero as $r \to +\infty$. This means that the systems described by such fitted curves are associated to distances between the sizes at consecutive ranks r and $r+1$ which decrease as r increases. Let us interpret such a statement in the considered context of the banking system of a given country. The compliance with a power-law means that the largest bank dominates the second largest one in size more than the latter dominates the one at rank $r = 3$, and the deviation between the sizes at ranks $r = 2$ and $r = 3$ is larger than that at ranks $r = 3$ and $r = 4$, and so on. Proceeding in this way, one has that banks are quite similar in size at high ranks. If a banking system departs remarkably from this regularity, then function f in equation (1) is not able to provide a statistically good description of this pattern.

The estimated parameters in equation (1) have a clear interpretation.

The parameter A gives information on the overall level of the sizes of the considered banks so that a large (small) value of A describes a power-law approximation of a banking system with large (small) banks. However, A is strongly influenced by the size of the element at the lowest rank, as approximated by the best-fit curve (i.e., the largest one, since $f(1) = A$).

The parameter α describes how the power-law curve decreases as rank r increases. In particular, a large value of α means that the curve is remarkably steep at the low ranks and relatively flat at the high ranks. Symmetrically, as the value of α becomes small, the curve flattens at low ranks and becomes steeper at high ranks.

We are now in the position of discussing the obtained findings. For each country, we use OLS to estimate the parameters of the log-linearized version of equation (1):

$$\log_{10}(TotAsset) = A - \alpha \log_{10}(r)\,. \tag{2}$$

Table 3 presents the best-fit parameters. The panoramic observation of the overall set of countries suggests that f in (1) is generally not able to approximate the scatter plots satisfactorily. This can be certified by the low values of R^2 – only Finland and the United States are above 0.97, which is a not so satisfactory a threshold for goodness-of-fit in the rank-size analysis.

Nevertheless, it is still meaningful to explore the information content of best-fit curves obtained by equation (1). Indeed, all the R^2s are around or well-above 0.9, except for Germany, which has an R^2 of $= 0.849$. Therefore, even if f in (1) fails in perfectly fitting the data, it remains worthy of describing the theoretical banking frameworks of power-law type which are statistically closer to the empirical data.

The estimated values of A range between 7.121 (Portugal) and 9.947 (United States), hence describing different realities of countries in terms of banks sizes – with particular attention to the size of the largest bank. More specifically, some countries present banking systems with very large banks and a predominance of the largest one – see e.g. Canada, France, Germany, Japan, Mexico, Switzerland, United Kingdom and Unites States, with A

Country	A	se_A	$pval_A$	α	se_α	$pval_\alpha$	R^2
Austria	8.476	0.038	0.00000	1.375	0.016	0.00000	0.931
Canada	9.849	0.086	0.00000	2.123	0.052	0.00000	0.943
Finland	8.165	0.020	0.00000	1.572	0.011	0.00000	0.992
France	9.722	0.100	0.00000	2.067	0.058	0.00000	0.912
Germany	9.556	0.042	0.00000	1.354	0.015	0.00000	0.849
Italy	8.412	0.048	0.00000	1.252	0.022	0.00000	0.899
Japan	9.150	0.034	0.00000	1.294	0.017	0.00000	0.958
Mexico	9.288	0.061	0.00000	2.557	0.032	0.00000	0.970
Poland	7.501	0.042	0.00000	1.512	0.024	0.00000	0.968
Portugal	7.121	0.036	0.00000	1.139	0.022	0.00000	0.963
Switzerland	9.087	0.036	0.00000	1.459	0.017	0.00000	0.959
United Kingdom	9.580	0.092	0.00000	2.010	0.049	0.00000	0.902
United States	9.947	0.012	0.00000	1.283	0.005	0.00000	0.989
All countries	11.897	0.040	0.00000	1.848	0.012	0.00000	0.825

Table 3: Estimated parameters of the linearized power law. se_s and $pval_s$ ($s \in \{A, \alpha\}$) are standard errors and p-values, respectively.

greater than 9 – while other ones are mainly composed by smaller banks – like Poland and Portugal, with A less than 7.

The estimated values of α can be reasonably considered as large in the context of power-law rank-size analysis, being greater than one for all the countries. However, the range among the countries is particularly wide – from 1.139 of Portugal to 2.557 of Mexico. This points to noticeable differences among the banking systems of the considered countries. Steeper behaviour of the curve at low ranks is observed in Canada, France, Mexico and the United Kingdom, with α greater than 2. This outcome means that the power-law approximations of the banking systems of such countries account for the presence of large banks which are highly heterogeneous in size – i.e., there are large differences in size at the low ranks. Differently, Italy, Japan, Portugal and the United States have $\alpha < 1.3$. Thus, they are closer to a theoretical framework related to equation (1), with more homogeneous sizes at low ranks.

The inability of f in equation (1) to fit the scatter plots is quite evident when looking at Figure 1, which reports the scatter plots and the estimated best-fit curves between total assets and the rank, represented on a semi-log scale for better visualization.

A quick visual inspection of the calibration exercises shows remarkable deviations between data and the best-fit curve. This is in agreement with the poor performance of the best-fit exercise observed above – with low values of the R^2s. Such deviations are noticeably evident in a group of peculiar cases. In particular, a large number of countries present discrepancies between the scatter plot and the best-fit curve at the high ranks – see, e.g. the remarkable deviations in Austria, France, Germany, Italy and the United Kingdom. One can argue that the smallest banks in the system form a sort of regime – i.e., there is an increasing difference between the sizes of consecutive ranks when one is above a critical rank. This means that such countries have a subsystem of small banks having very different sizes. This is undoubtedly one of the main reasons for the failure of the power law in fitting the data.

Under a different perspective, low ranks are much more regular in terms of the system described by the power law. The most noticeable deviations appear in the case of France. In this country, one can observe a small group of very large banks with similar size; then, there is a huge difference between the size of the banks at rank $r = 4$ and $r = 5$; furthermore, sizes are quite similar for ranks close to 5. This set of irregularities is responsible for the

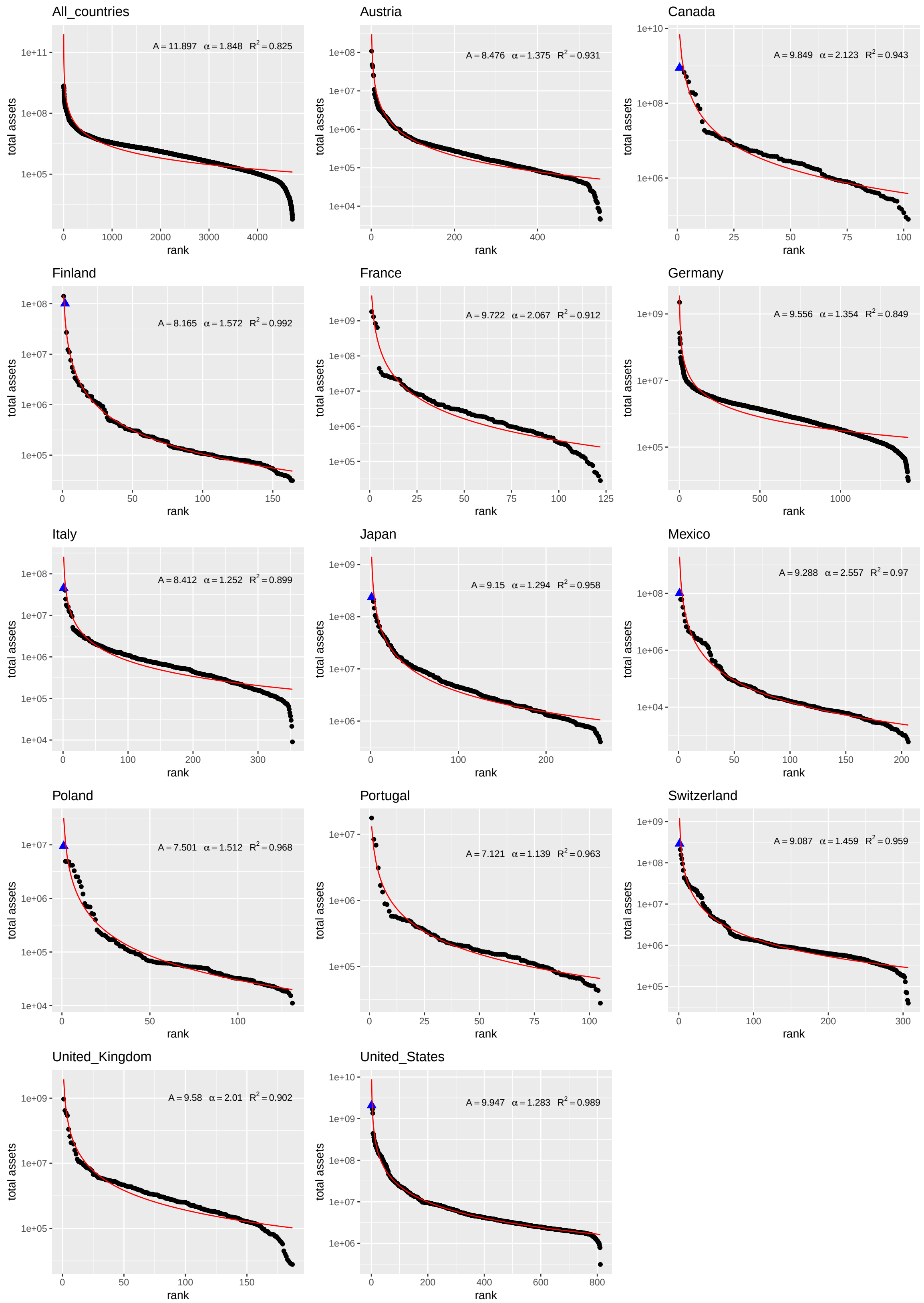

Figure 1: Actual data (dots) and estimated power laws (red curves), according to formula (2). Blue triangles represent possible outliers identified by Cook's distance with the median of an $F_{2,n-2}$ distribution as the cut-off and n the number of observations.

discrepancy between the shape of the curve and the scatter plot at low ranks. Interestingly, some countries have a quite scattered representation of the sizes of the banks at low-middle ranks. In this respect, we mention Canada, Mexico, Poland and Switzerland – whose data are above the fitting curve and exhibit a concave-shaped behaviour approximately between $r = 1$ and $r = 12$ (for Canada), $r = 10$ and $r = 30$ (for Mexico), $r = 2$ and $r = 16$ (for Poland) and $r = 10$ and $r = 35$ (for Switzerland). Italy presents a subgroup of banks at low rank which departs from the rest of the system, being of particularly large size.

4 The Discrete Generalized Beta-based rank size analysis

To overcome the modelling limitations of the power-law in equation (1), we implement an additional best-fit procedure using the following Discrete Generalized Beta Distribution (Naumis and Cocho, 2008; Martínez-Mekler et al., 2009):

$$g(r) = \frac{B(R+1-r)^{\delta}}{r^{\gamma}}, \tag{3}$$

where r is the rank, $R = \max(r)$ over the considered data sample, and B, γ and δ are positive parameters to be estimated from the data. Taking the first derivative of equation (3) with respect to r, it is easy to see that the parameter γ affects the shape of the curve at low ranks, with large (small) values associated to a steep (flat) curve when r is small. On the contrary, δ dominates at high rank values, with large (small) values associated to a steep (flat) curve at high ranks. Equation (3) thus represents a situation where there is a scaling regime for small r values, and a truncated scale regime for large r values. Interestingly, this specification also allows to control for the size of the sample through the parameter R, which Segarra and Teruel (2012) have found to be a crucial feature when estimating power law distributions.

As in the case of the standard power law represented in equation (1), the parameters B, δ and γ can be estimated by OLS on the log-linearized form of equation (3):

$$\log_{10}(TotAsset) = B + \delta \log_{10}(R+1-r) - \gamma \log_{10}(r)\,. \tag{4}$$

Table 4 presents the results of the estimates of equation (4), along with the goodness-of-fit measure R^2. The effectiveness of g in fitting the scatter plots is confirmed by the high

Country	B	se_B	$pval_B$	γ	se_γ	$pval_\gamma$	δ	se_δ	$pval_\delta$	R^2
Austria	6.609	0.022	0.00000	1.053	0.005	0.00000	0.486	0.005	0.00000	0.996
Canada	8.210	0.121	0.00000	1.698	0.041	0.00000	0.607	0.041	0.00000	0.982
Finland	7.892	0.044	0.00000	1.510	0.013	0.00000	0.091	0.013	0.00000	0.994
France	7.488	0.114	0.00000	1.517	0.037	0.00000	0.793	0.037	0.00000	0.982
Germany	6.261	0.024	0.00000	0.874	0.005	0.00000	0.732	0.005	0.00000	0.991
Italy	6.464	0.028	0.00000	0.883	0.007	0.00000	0.551	0.007	0.00000	0.994
Japan	8.039	0.034	0.00000	1.069	0.009	0.00000	0.333	0.009	0.00000	0.993
Mexico	7.967	0.111	0.00000	2.273	0.032	0.00000	0.417	0.032	0.00000	0.984
Poland	7.341	0.104	0.00000	1.474	0.033	0.00000	0.056	0.033	0.09564	0.969
Portugal	6.627	0.075	0.00000	1.012	0.025	0.00000	0.182	0.025	0.00000	0.975
Switzerland	8.171	0.067	0.00000	1.280	0.018	0.00000	0.266	0.018	0.00000	0.977
United Kingdom	6.892	0.069	0.00000	1.419	0.020	0.00000	0.865	0.020	0.00000	0.991
United States	9.497	0.022	0.00000	1.211	0.005	0.00000	0.110	0.005	0.00000	0.993
All countries	5.988	0.014	0.00000	1.130	0.002	0.00000	1.106	0.002	0.00000	0.996

Table 4: Estimated parameters of the linearized generalized beta law. se_s and $pval_s$ ($s \in \{B, \gamma, \delta\}$ are standard errors and p-values, respectively.

values of the R^2s, ranging from 0.969 of Poland to 0.996 of Austria. In all cases, the R^2s obtained from (4) are significantly higher than those from equation (2).

The ability of g in equation (4) to fit the data is clear when looking at Figure 2, which replicates Figure 1 reporting the scatter plots and the estimated best-fit curves obtained from the Discrete Generalized Beta Distribution between total assets and the rank, represented on a semi-log scale for better visualization. We only observe small deviations between the scatter plots and the curves at low ranks in some specific countries, which are associated to R^2s smaller than 0.985 (Canada, France, Mexico, Poland, Portugal and Switzerland).

The estimated values of the parameter γ show high variability, ranging from 0.874 for Germany to 2.273 for Mexico. Interestingly, aside from the case of Mexico, which shows large discrepancies between bank pairs at consecutive low ranks, two main clusters emerge: one with γ's around 1.5 (Canada, Finland, France, Poland and the United Kingdom) and the other with γ around 1 (Austria, Germany, Italy, Japan, Portugal, Switzerland and the United States).

The values of δ also have a sizeable range of variation, from 0.056 (Poland) and 0.091 (Finland) to 0.865 (United Kingdom). However, no clear clusters appear in this case.

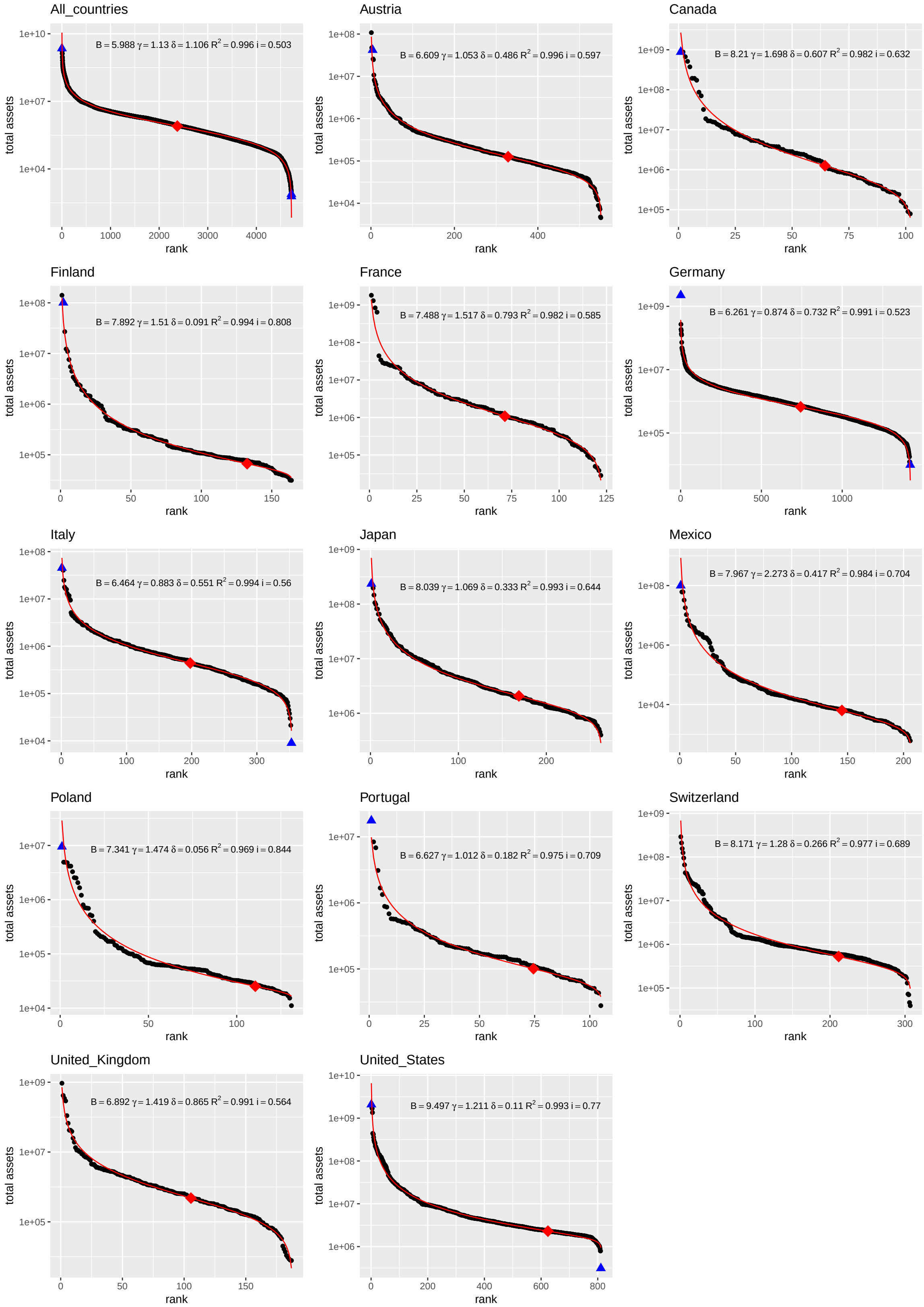

Figure 2: Actual data (dots) and estimated generalized beta laws (red curves), according to formula (4). Blue triangles represent possible outliers identified by Cook's distance with the median of an $F_{3,n-3}$ distribution as the cut-off and n the number of observations. Red diamonds represent inflexion points. The relative position of inflexion points is reported in legends as "i = ...".

5 Cross-country differences

The Discrete Generalized Beta Distribution presented in Section 4 provides a good fit of the rank-size distribution based on bank total assets, with R^2s ranging from 0.969 to 0.996, as opposed to values as low as 0.849 in the case of the traditional power law. This comes mostly from the extra flexibility conceded by having one additional parameter relative to a power-law distribution, which allows a better fit at lower as well as at higher ranks.

However, introducing an additional parameter that takes different values in each country reduces the generality of the law describing the rank-size relationship, with two crucial implications. First, the set of estimated parameters may be linked to some features of a country's banking industry, such as the size or the degree of competition. If this were the case, the general law could be interpreted as a synthetic picture of some structural characteristics of the banking sector, characterized by country specific parameters. Second, since an ample literature has shown that the banking sector is crucially shaped by the regulatory and institutional characteristics prevailing in each country, the estimated parameters may also be linked to such features (described in Section 2), that could be seen as the deep determinants of the estimated law.

In the following, we present the results of an attempt to uncover such relationships between the parameters δ and γ estimated in each country, and the features of its banking sector.

In doing so, we propose a causation analysis of the considered variables on the calibrated parameters, hence leading to intuitive economic explanations.

Figure 3 presents a set of scatter plots of the values of the coefficient γ estimated for each country using the Discrete Generalized Beta Distribution in equation (3), and the corresponding characteristics of each country's banking industry.

Despite the high dispersion of the data with respect to the fitted OLS line, also confirmed by the low values of the R^2s, some regularities emerge. First, Figure 3 shows that γ takes lower values in countries with more developed banking sectors, measured by both the ratio of bank deposits and bank total assets over GDP. Since γ affects the shape of the rank-size curve at low ranks (i.e., for larger banks), this suggests that in countries with a more developed financial industry, the difference in size between the largest banks and next ones in the rank by total assets is smaller. In other words, it appears that countries with a more

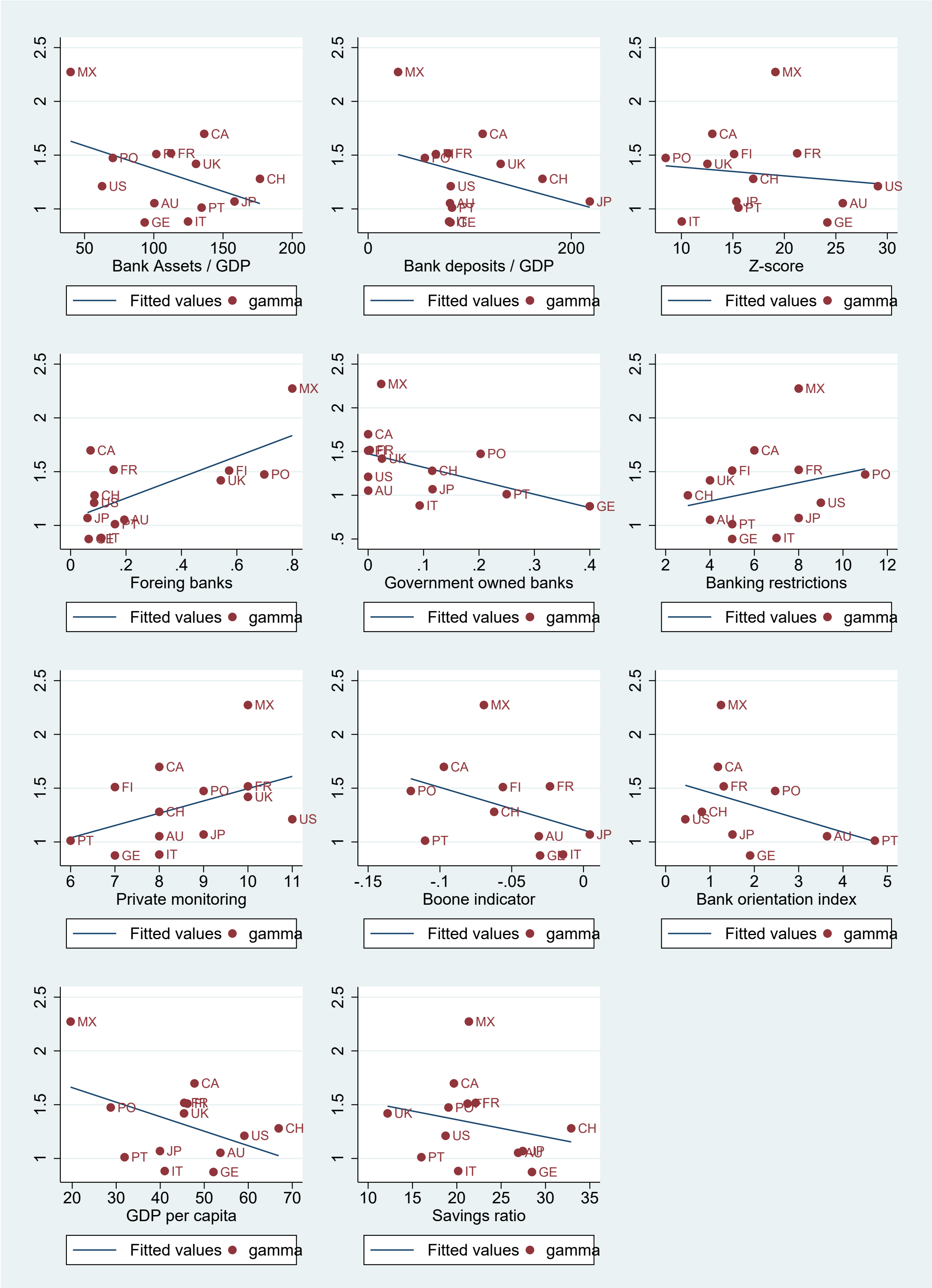

Figure 3: Estimated γ in (4) and banking industry and country characteristics

developed banking market have a relatively more uniform distribution of bank size among the largest banks. Interestingly, this is not due to the development of the entire financial sector, but to that of banks, as shown by the negative relationship between γ and the bank orientation index, that is the ratio of bank total assets to stock market capitalization.

Figure 4 provides some evidence, albeit weaker, of an opposite relationship with δ, which takes higher values in countries with a more developed banking sector. Since δ affects the shape of the rank-size curve at high ranks (i.e., for small banks), this suggests that in countries with a more developed financial industry, the difference in size between the smallest banks and the previous ones in the rank by total assets is larger. Countries with a more developed banking market thus have a less uniform distribution of bank size among the smallest banks.[16] The correlation with the Bank orientation index is in this case weakly negative.

Average riskiness does not seem to be correlated with the shape of the rank-size distribution, as suggested by both scatter plots of the average Z-score in a a country and the vales of both γ and δ.

More interestingly, we find a relatively strong evidence of a relationship between γ and the share of assets of foreign and government owned banks. Remarkably, the correlation is positive with respect to the presence of foreign banks, suggesting in this case a steeper rank-size relationship among larger banks, and negative with respect to the presence of government owned banks, suggesting instead a less steep relationship. In interpreting these results, it is important to recall two features of our sample, which includes only developed countries. First, with the exception of Mexico, foreign banks are generally small subsidiaries of large multinational groups rather than large banks controlled by foreign groups, as it is often the case in less developed countries. Second, governments generally divested from the banking sector at the end of the last century, although with the Global financial crisis some large banking groups went back under their control (e.g., NatWest in the UK, Hypo Alpe-Adria-Bank International in Austria, Allied Irish banks in Ireland). This suggests that the effect of the share of foreign and government owned banks does not depend on the actual presence of such financial intermediaries, but instead on their indirect impact on domestic, private banks. No clear relationship appears instead in the case of the shape of

[16] We have also checked whether there is a relationship between the γ's and the δ's in our sample, but the correlation coefficient is 0.001.

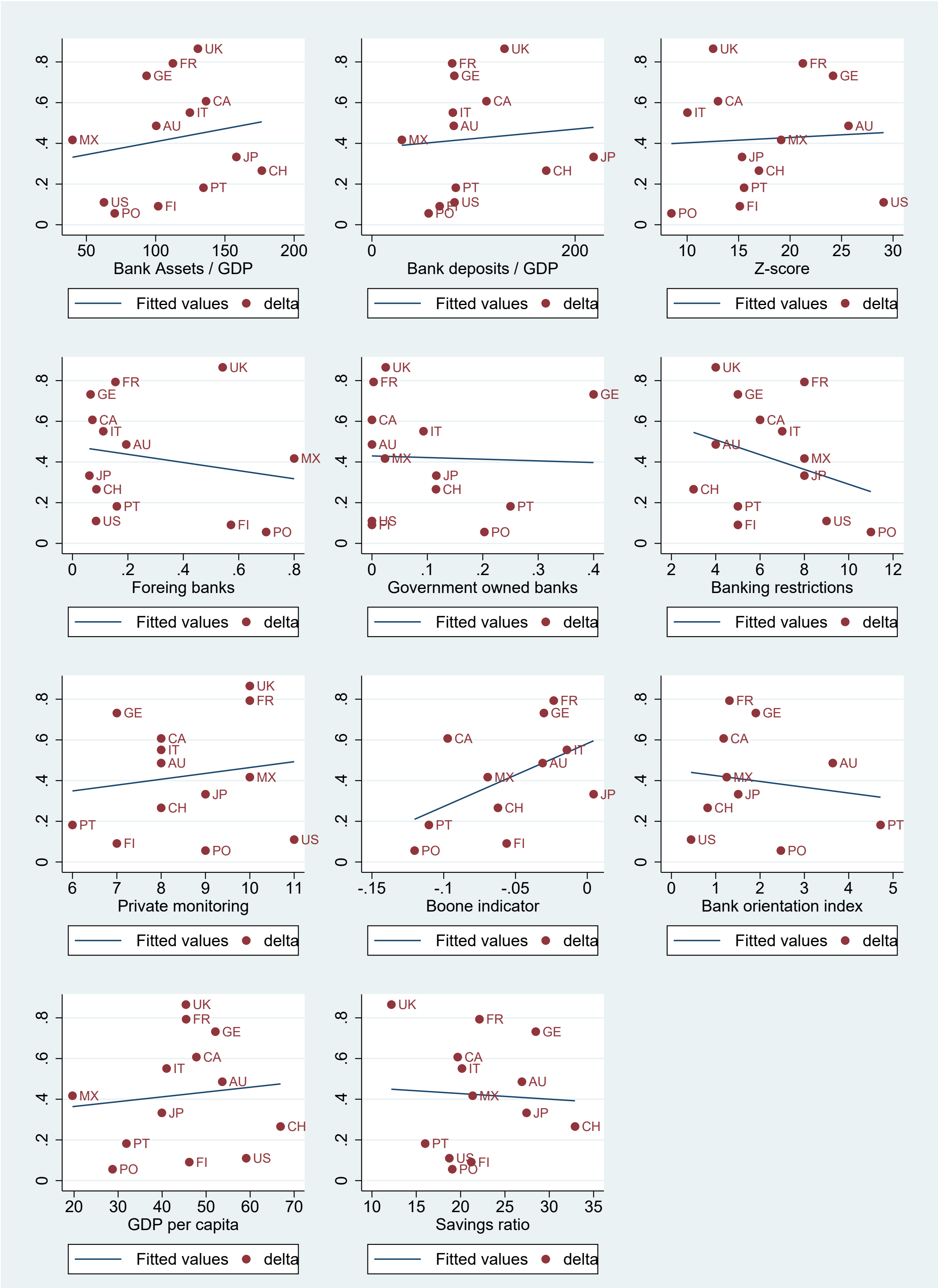

Figure 4: Estimated δ in (4) and banking industry and country characteristics.

the rank-size distribution of smaller banks, as shown in Figure 4.

Regulations also have an impact on the shape of the rank-size distribution. Quite surprisingly, Figure 3 provides evidence of a positive relationship between the value of γ and the indexes of how restrictive is banking regulation and how strong is private monitoring. Regulations thus tend to favour a less uniform distribution of bank size among the largest banks. More restrictive banking regulations have instead an opposite effect on the distribution at higher ranks, as shown by the negative relationship with δ in Figure 4. In the case of smaller banks, we find instead no clear evidence of a connection with the strength of private monitoring.

The shape of the rank-size correlation is also related to the degree of competition in the banking market, especially at higher rank values. Figure 3 shows a negative correlation between γ and the Boone indicator, suggesting that a more uniform distribution of bank size among larger banks is associated with higher competition. Figure 4 shows instead a positive correlation between δ and the value of the Boone index of banking competition, which takes higher values for less competitive banking markets. A more uniform distribution of bank size among small banks is therefore associated with lower competition.

Finally, we find some evidence that γ is also negatively related to the level of per capita GDP and the savings ratio. This result possibly mirrors the fact that richer and countries and those with higher savings ratios also have more developed banking systems, which in turn, as we have documented above, shows smaller differences in the size of the largest banks and the next ones in the rank by total assets, as we documented above. The relationships with δ are instead barely significant.

One might also be interested in understanding whether the ability of the generalized beta distribution to fit of the rank-size distribution based on bank total assets is itself affected by some characteristics of a country's banking industry. In other words, we may be interested in understanding if there are some characteristics which make the law less satisfactory under a statistical point of view in describing the rank-size relationship.

To this purpose, in Figure 5 we present a set of scatter plots of the values of the R^2 of the fit of the generalized beta distribution to the data in each country and the same features of the banking industry described above. The results provide only very weak evidence that some country features have a significant correlation with the goodness-of-fit of the regression. Indeed, only riskiness – which interestingly seemed unrelated to both β and γ

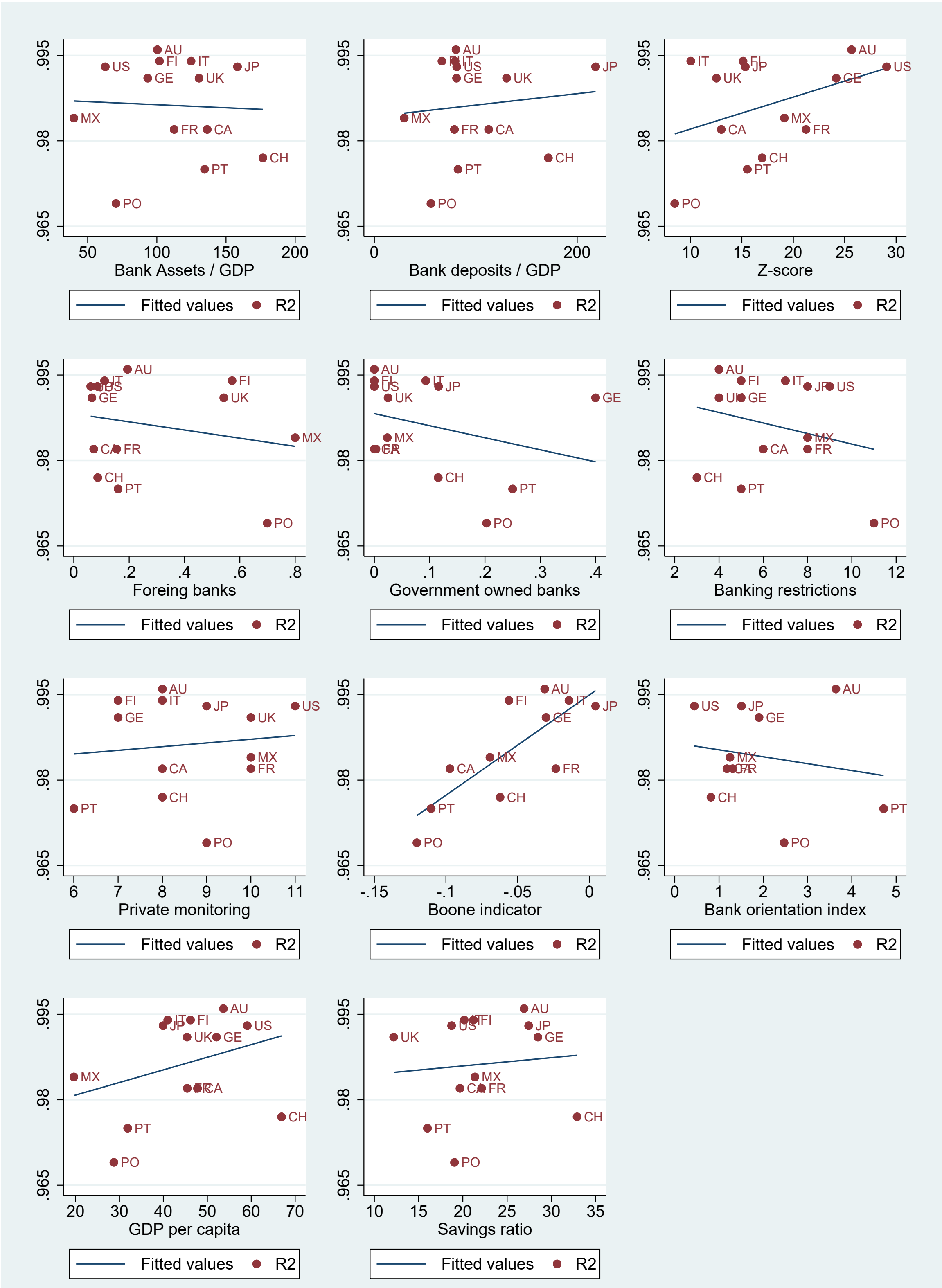

Figure 5: Estimated R^2 for the best fit with equation (4) and banking industry and country characteristics.

– the Boone index of competitiveness – which instead was positively correlated with δ and negatively with γ – and per capita GDP appear to have a positive correlation with the R^2.

Overall, these results suggest that the shape of the rank-size curves fitted by the Discrete Generalized Beta Distribution are not abstract numbers, but they are meaningfully related to the structure of the banking industry and its regulation.

6 Conclusions

This paper deals with the exploration of the banking system of a country, which is a theme of crucial interest either for policymakers and institutional economists as well as for academicians. The complexity of the financial institutions and the presence of regularities and discrepancies in their patterns let this exploration be particularly challenging. In this paper, we adopt a ranked data analysis approach by introducing suitable rank-size laws for fitting a collection of the empirical observations of banks' total assets. The available samples are taken from the banking systems of 13 developed countries, the size is the total assets and the rank is intended in descending order – so that the lowest ranks are associated to the largest banks.

We show that the Discrete Generalized Beta Distribution provides a statistically satisfactory representation of the considered samples at a country level. Such an outcome has a clear interpretation. Indeed, the analysed banking systems show a substantially homogeneous large set of banks of medium dimensions; differently, they present a high level of difformity in the clusters of the largest and the smallest banks.

A detailed discussion of how the best-fit parameters are related to the structural characteristics of the banking sectors offers several insights on the roots of similarity and difformity among the considered countries. We find that a high level of development of a banking system is associated to a substantial uniformity of the sizes of the largest banks and a quite scattered set of values for the total assets of the smallest ones. Furthermore, the roles played by regulations, competition and presence of foreign banks are also acknowledged as relevant in shaping the banking system.

Of course, this paper opens many questions, to be left for future research. Among them, we point the attention to the analysis of the connections between the fitted curves/banking systems and the macroeconomic variables of the investigated countries. In so doing, we would offer a view of the banking sector well-beyond its constitutive parameters, to shed light on the interactions between the economic health of the country and its system of financial intermediation.

References

Ausloos, M., 2013. A scientometrics law about co-authors and their ranking: the co-author core. Scientometrics 95, 895–909. doi:10.1007/s11192-012-0936-x.

Ausloos, M., 2015. Coherent measures of the impact of co-authors in peer review journals and in proceedings publications. Physica A: Statistical Mechanics and its Applications 438, 568–578. doi:10.1016/j.physa.2015.06.010.

Ausloos, M., Cerqueti, R., 2016. A universal rank-size law. PLoS ONE 11, e0166011. doi:10.1371/journal.pone.0166011.

Ausloos, M., Nedic, O., Fronczak, A., Fronczak, P., 2016. Quantifying the quality of peer reviewers through Zipf's law. Scientometrics 106, 347–368. doi:10.1007/s11192-015-1704-5.

Axtell, R.L., 2001. Zipf distribution of US firm sizes. Science 293, 1818–1820. doi:10.1126/science.1062081.

Barba Navaretti, G., Calzolari, G., Pozzolo, A.F., Trentinaglia de Daverio, M.T., 2019. Few large with many small: Banks size distribution and cross-border financial linkages. Journal of Financial Services Research 56, 229–258. doi:10.1007/s10693-019-00325-5.

Barth, J.R., Lin, C., Ma, Y., Seade, J., Song, F.M., 2013. Do bank regulation, supervision and monitoring enhance or impede bank efficiency? Journal of Banking & Finance 37, 2879–2892. doi:10.1016/j.jbankfin.2013.04.030.

Beck, T., Demirgüç-Kunt, A., Levine, R., 2000. A new database on the structure and development of the financial sector. The World Bank Economic Review 14, 597–605. doi:10.1093/wber/14.3.597.

Brakman, S., Garretsen, H., Van Marrewijk, C., Van Den Berg, M., 1999. The return of Zipf: towards a further understanding of the rank-size distribution. Journal of Regional Science 39, 183–213. doi:10.1111/1467-9787.00129.

Caprio, G., Laeven, L., Levine, R., 2007. Governance and bank valuation. Journal of Financial Intermediation 16, 584–617. doi:10.1016/j.jfi.2006.10.003.

Cefis, E., Marsili, O., Schenk, H., 2009. The effects of mergers and acquisitions on the firm size distribution. Journal of Evolutionary Economics 19, 1–20. doi:10.1007/s00191-008-0105-9.

Cerqueti, R., Ausloos, M., 2015. Evidence of economic regularities and disparities of Italian regions from aggregated tax income size data. Physica A: Statistical Mechanics and its Applications 421, 187–207. doi:10.1016/j.physa.2014.11.027.

Chatterjee, A., Ghosh, A., Chakrabarti, B.K., 2016. Universality of citation distributions for academic institutions and journals. PloS ONE 11, e0146762. doi:10.1371/journal.pone.0146762.

Coad, A., 2009. The growth of firms: A survey of theories and empirical evidence. New Perspective on the Modern Corporation, Edward Elgar Publishing, Cheltenham, UK.

Córdoba, J.C., 2008. On the distribution of city sizes. Journal of Urban Economics 63, 177–197. doi:10.1016/j.jue.2007.01.005.

Crosato, L., Ganugi, P., 2007. Statistical regularity of firm size distribution: the Pareto IV and truncated Yule for Italian SCI manufacturing. Statistical Methods and Applications 16, 85–115. doi:10.1007/s10260-006-0023-7.

Demirgüç-Kunt, A., Detragiache, E., 1998. The determinants of banking crises in developing and developed countries. IMF Staff Papers 45, 81–109. doi:10.2307/3867330.

Demirgüç-Kunt, A., Huizinga, H., 2013. Are banks too big to fail or too big to save? International evidence from equity prices and CDS spreads. Journal of Banking & Finance 37, 875–894. doi:10.1016/j.jbankfin.2012.10.010.

Demirgüç-Kunt, A., Levine, R., 2004. Financial structure and economic growth: A cross-country comparison of banks, markets, and development. MIT Press, Cambridge, MASS.

di Giovanni, J., Levchenko, A.A., Rancière, R., 2011. Power laws in firm size and openness to trade: Measurement and implications. Journal of International Economics 85, 42–52. doi:10.1016/j.jinteco.2011.05.003.

Dimitrova, Z.I., Ausloos, M., 2015. Primacy analysis in the system of Bulgarian cities. Open Physics 1, 218–225. doi:10.1515/phys-2015-0029.

Duprey, T., Lé, M., 2016. Bankscope dataset: getting started. Techreport SSRN 2191449. Banque de France. doi:https://dx.doi.org/10.2139/ssrn.2191449.

Fahlenbrach, R., Prilmeier, R., Stulz, R.M., 2012. This time is the same: Using bank performance in 1998 to explain bank performance during the recent financial crisis. The Journal of Finance 67, 2139–2185. doi:10.1111/j.1540-6261.2012.01783.x.

Ficcadenti, V., Cerqueti, R., 2017. Earthquakes economic costs through rank-size laws. Journal of Statistical Mechanics: Theory and Experiment 2017, 083401. doi:10.1088/1742-5468/aa7ab7.

Ficcadenti, V., Cerqueti, R., Ausloos, M., 2019. A joint text mining-rank size investigation of the rhetoric structures of the US Presidents' speeches. Expert Systems with Applications 123, 127–142. doi:10.1016/j.eswa.2018.12.049.

Ficcadenti, V., Cerqueti, R., Ausloos, M., Dhesi, G., 2020. Words ranking and Hirsch index for identifying the core of the hapaxes in political texts. Journal of Informetrics 14, 101054. doi:10.1016/j.joi.2020.101054.

Gabaix, X., 1999. Zipf's law for cities: An explanation. The Quarterly Journal of Economics 114, 739–767. doi:10.1162/003355399556133.

Gabaix, X., 2011. The granular origins of aggregate fluctuations. Econometrica 79, 733–772. doi:10.3982/ECTA8769.

Gaffeo, E., Gallegati, M., Palestrini, A., 2003. On the size distribution of firms: Additional evidence from the G7 countries. Physica A: Statistical Mechanics and its Applications 324, 117–123. doi:10.1016/s0378-4371(02)01890-3.

Ganugi, P., Grossi, L., Crosato, L., 2004. Firm size distributions and stochastic growth models: A comparison between ICT and mechanical Italian companies. Statistical Methods and Applications 12, 391–414. doi:10.1007/s10260-003-0073-z.

Gibrat, R., 1931. Les Inégalités. Applications: Aux Inégalités des Richesses, à la Concentration des Entreprises, Aux Populations des Villes, Aux Statistiques des Familles, etc., d'une Loi Nouvelle, La Loi de l'Effet Proportionnel. Librairie du Recueil Sirey, Paris.

Giesen, K., Südekum, J., 2011. Zipf's law for cities in the regions and the country. Journal of Economic Geography 11, 667–686. doi:10.1093/jeg/lbq019.

Growiec, J., Pammolli, F., Riccaboni, M., Stanley, H.E., 2008. On the size distribution of business firms. Economics Letters 98, 207–212. doi:10.1016/j.econlet.2007.04.031.

Laeven, L., Ratnovski, L., Tong, H., 2016. Bank size, capital, and systemic risk: Some international evidence. Journal of Banking & Finance 69, S25–S34. doi:10.1016/j.jbankfin.2015.06.022.

Luttmer, E.G.J., 2007. Selection, growth, and the size distribution of firms. The Quarterly Journal of Economics 122, 1103–1144. doi:10.1162/qjec.122.3.1103.

Mansilla, R., Köppen, E., Cocho, G., Miramontes, P., 2007. On the behavior of journal impact factor rank-order distribution. Journal of Informetrics 1, 155–160. doi:10.1016/j.joi.2007.01.001.

Martínez-Mekler, G., Martínez, R.A., del Río, M.B., Mansilla, R., Miramontes, P., Cocho, G., 2009. Universality of rank-ordering distributions in the arts and sciences. PLoS ONE 4, e4791. doi:10.1371/journal.pone.0004791.

Montemurro, M.A., 2001. Beyond the Zipf–Mandelbrot law in quantitative linguistics. Physica A: Statistical Mechanics and its Applications 300, 567–578. doi:10.1016/S0378-4371(01)00355-7.

Naumis, G.G., Cocho, G., 2008. Tail universalities in rank distributions as an algebraic problem: The beta-like function. Physica A: Statistical Mechanics and its Applications 387, 84–96. doi:10.1016/j.physa.2007.08.002.

Reed, W.J., 2001. The Pareto, Zipf and other power laws. Economics Letters 74, 15–19. doi:10.1016/s0165-1765(01)00524-9.

Segarra, A., Teruel, M., 2012. An appraisal of firm size distribution: Does sample size matter? Journal of Economic Behavior & Organization 82, 314–328. doi:10.1016/j.jebo.2012.02.012.

Simon, H.A., Bonini, C.P., 1958. The size distribution of business firms. The American Economic Review 48, 607–617.

Sutton, J., 1997. Gibrat's legacy. Journal of Economic Literature 35, 40–59.

Zhang, J., Chen, Q., Wang, Y., 2009. Zipf distribution in top Chinese firms and an economic explanation. Physica A: Statistical Mechanics and its Applications 388, 2020–2024. doi:10.1016/j.physa.2009.01.027.