On the parametric description of log-growth rates of Romanian city sizes*

Irina Băncescu† Luminiţa Chivu‡ Till Massing§ Vasile Preda¶
Miguel Puente-Ajovín|| Arturo Ramos**

Abstract

We consider log-growth rates of Romanian cities' populations for all cities in the country in the period 1992-2019 on an annual basis. We construct annual, quinquennial and decennial log-growth rates and fit to them thirty-one different statistical distributions. The best results with Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling statistics are obtained by a mixture of five stretched Gaussian distributions (5sG) with some fixed parameters, and with the AIC, BIC, HQC information criteria are obtained with mixtures of three logistic distributions (3L), that may have or may have not exponential tails. Just as an illustration, we propose a generating stochastic mechanism for the 3L.

Keywords: logistic distribution, Student's t distribution, stretched Gaussian distribution, Romanian population, log-growth rates

*The work of Till Massing has been supported by *Deutsche Forschungsgemeinschaft (DFG)* Grant Number 455257011. The work of Miguel Puente-Ajovín and Arturo Ramos has been supported by the Spanish *Ministerio de Ciencia e Innovación* (PID2020-112773GB-I00) and by Aragon Government (ADETRE Reference Group S39_20R).

†Bucharest University of Economic Studies, Bucharest, Romania Email: irina.bancescu@ie.ase.ro.

‡"Costin C. Kiriţescu" National Institute of Economic Research, Bucharest, Romania. Email: chivu@ince.ro.

§Universität Duisburg-Essen, Fakultät für Wirtschaftswissenschaften, Lehrstuhl für Ökonometrie, Essen, Germany. Email: till.massing@uni-due.de.

¶"Costin C. Kiriţescu" National Institute of Economic Research, Bucharest, Romania. Email: vasilepreda0@gmail.com.

||Department of Economic Analysis, Universidad de Zaragoza, Teruel, Spain. Email: mpajovin@unizar.es.

**Department of Economic Analysis, Universidad de Zaragoza, Zaragoza, Spain. Email: aramos@unizar.es.

1 Introduction

There has been a relatively recent interest in the parametric description of the log-growth rates of the sizes of firms and of countries' outputs in their relation to Gibrat's law (Bottazzi and Secchi, 2006, 2011; Canning et al., 1998; Cottineau et al., 2019, 2017; de Wit, 2005; Delli Gatti et al., 2005; Fagiolo et al., 2008; Fu et al., 2005; Fujiwara et al., 2004; Gallegati and Palestrini, 2010; Riccaboni et al., 2008; Rozenfeld et al., 2008; Sarkar, 2019; Stanley et al., 1996). In this strand of the literature, the Laplace and the asymmetric double Laplace distribution are typically studied, instead of the normal distribution, to account for the observed exponential tails exhibited by the log-growth rates of the mentioned quantities. Later, in order to generalize the normal distribution at the same time as the asymmetric double Laplace distribution, there has also been considered the asymmetric Subbotin distribution. It is generally accepted that these distributions provide a faithful description of the log-growth rates considered, above all the description of the approximately exponential tails.

Let us mention that in Li et al. (2017) it is proposed a simple unified model to explain the growth dynamics of cities and scaling laws, where the model predicts that the size of cities grows linearly regardless of its current size.

For city sizes, the topic of growth rates has been studied, but mainly in a non-parametric way (Ioannides and Overman, 2003), see also Wu and He (2017) for a more recent article on this subject. Perhaps because of the lack of good data sets until very recently, there have been very few studies that consider the parametric description of log-growth rates of all city sizes in a country. It is not until 2004 that Eeckhout considers the sample of all USA Places in 1990 and 2000 (in the year 1990 without the Census Designated Places, CDPs) and takes the log-growth rates to be normally distributed, according to the strict fulfillment of Gibrat's law (Gibrat, 1931; Sutton, 1997). In this version of the law, the log-growth rates are i.i.d. normal variables that are added to yield the log-sizes, which at the end are described by a normal distribution (log-normal distribution for the sizes) since the convolution of normal distributions is exactly another normal distribution.

Later there have appeared a few studies that deal with the topic of the parametric description of log-growth rates of city sizes (Ramos, 2017; Schluter and Trede, 2013, 2016; Massing et al., 2020). In the second and third of these references, the Student's t distribution is proposed (for the German municipalities or *Gemeinden*) as an appropriate model. In the last one, mixtures of Student's t distribution (with fixed degrees of freedom) are proposed and the fit has been very good. The robust result that emerged is that the normal is not selected indeed by all of the statistical criteria employed. Apart from that, other alternative distributions were studied: the asymmetric double Laplace normal (adLN) of Reed (2002, 2003); Reed and Jorgensen (2004); Manas (2009), a new distribution called "double mixture exponential Generalized Beta of the second kind (dmeGB2)", and the mentioned mixtures of Student's t distributions. The latter ones provided the best fit, in terms of a number of statistical criteria, of all the proposals.

It is the main goal of the current paper to apply these ideas to a recent data set of populations of Romanian cities, covering the period 1992-2019 on an annual basis. These data sets comprise as well data for *all* Romanian cities, not only the upper tail, according to the main suggestion of Eeckhout (2004). We compare for the log-growth rates constructed out of the previous population data on periods of one, five and ten years, thirty-one different parametric models, with and without exponential tails, composite models or not, heavy tails or not. This allows us to compare different features of the parametric models presented. The study of the distribution of the log-growth rates of Romanian cities complements therefore the study of Romanian city sizes (in levels) carried out in Băncescu et al. (2019).

The rest of the paper is organized as follows. The next section of the paper presents the databases used. In Section 3 we describe the parametric distributions considered and the statistical methodology. In Section 4 we summarize the main results (that are detailed in a supplementary Excel file). In Section 5 we show a generating statistical mechanism for the model most often selected in terms of information criteria in Section 4. Finally, we offer in Section 6 some conclusions.

2 The databases

The analysis performed in this paper is based on a complete data set of the population of Romanian cities on an annual basis from the year 1992 to 2019, provided by the National Institute of Statistics of Romania in the TEMPO Online database. Romania's territory is organized into administrative-territorial units (UATs) (county, municipality, city, commune and village). Our database is composed of data population for three types of UATs, meaning municipality, city and commune, the first two being urban territories, municipalities having a larger population than cities. Moreover, municipality has a greater social and economic development than a city. Due to the economic development of the country, that occurred in the last decades, the number of municipalities has almost doubled from 56 in 1992 to 103 in 2019, while the number of cities has increased by only 12 cities. Also, the number of communes did not registered a significant increase (only 6.47%) reaching a number of 2862 communes in 2019.

This data set is the most exhaustive existing data set for all cities of Romania to date, although the city definition is administrative. This might cause slight distortions to the upper tail of city size distributions when considered as economic units (by satellite nighttime lights or other approach), but in turn we are able to describe *all cities*, not only those of the upper tail, thus it is a trade-off: a little bit less accuracy for the upper tail at the price of considering all cities, even the smallest ones. We think that this approach is more complete and the log-growth rate description will be much more accurate.

We consider, among the mentioned data samples, the possible annual, quinquennial and decennial log-growth rates, summing up a number of 68 samples.

We compute the log-growth rates of Romanian cities' populations according to the well-known formula

$$g_{i,t} = \log x_{i,t} - \log x_{i,t-1} \in (-\infty, \infty)$$

where $x_{i,t}$ is the population of city i at time t. When a fixed t is taken we will simply write $g \in (-\infty, \infty)$ for the variable of all log-growth rates of the cross-sections taken.

We show in Table 1 the descriptive statistics of the log-growth rates samples. It is to be remarked the high skewness departing from zero and the very high kurtosis in many samples.

3 Description of the distributions

In this Section we will present the distributions used in this paper.

The first distribution we will consider is the well-known normal distribution (denoted by N) for the log-growth rates g. We thus set

$$f_{\mathrm{N}}(g; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(g-\mu)^2}{2\sigma^2}\right) \tag{1}$$

where $\mu \in (-\infty, \infty)$ is the mean of g and $\sigma > 0$ is its standard deviation according to this distribution. This distribution for the log-growth rates of cities is the one postulated by the strict formulation of Gibrat's Law (see, e.g., Gibrat (1931); Sutton (1997) and references therein), so it is a natural candidate for our study. The corresponding cumulative distribution function (CDF) is

$$\Phi(g; \mu, \sigma) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{g-\mu}{\sigma\sqrt{2}}\right) .$$

where erf denotes the error function associated to the standard normal distribution.

The second distribution in our study will be the Laplace distribution (see, e.g., Johnson et al. (1995); Kotz et al. (2001) and references therein), that has been already used in studying log-growth rates of firms, cities and of GDP growth rates (Stanley et al., 1996; Canning et al., 1998; Fu et al., 2005; Riccaboni et al., 2008; Schluter and Trede,

Sample	Obs	Mean	SD	Skewness	Kurtosis	Min	Max
1992-1993	2946	-0.0098039	0.01273414	-0.071965	8.16755673	-0.1260747	0.06456707
1993-1994	2946	-0.0070921	0.01177544	0.05857588	4.63957999	-0.0716224	0.06617729
1994-1995	2946	-0.007342	0.01005216	-0.0593998	4.7682703	-0.05126	0.05971286
1995-1996	2946	-0.0058639	0.01674904	-10.737897	304.431577	-0.5047729	0.08701138
1996-1997	2948	-0.0057498	0.01047897	0.1788858	5.99923716	-0.0619667	0.07137666
1997-1998	2948	-0.0030071	0.01206456	0.12221042	5.88834214	-0.0677855	0.09092131
1998-1999	2948	-0.0032301	0.01441868	-7.0191193	185.907215	-0.3462676	0.20479441
1999-2000	2951	0.00010472	0.01023074	0.19445382	4.47713132	-0.0437228	0.05410312
2000-2001	2951	-0.0020242	0.0096344	-0.1191908	4.34721879	-0.0569144	0.03978856
2001-2002	2951	-0.0039931	0.02272786	-19.592721	511.339699	-0.6573906	0.06130622
2002-2003	2955	-0.0086764	0.05789082	-11.31199	145.678925	-1.0259175	0.05111573
2003-2004	2983	-0.0260102	0.11849713	-5.1363936	31.7338474	-1.4005525	0.20633643
2004-2005	3133	-0.007424	0.04972799	-10.971425	142.43521	-0.9581111	0.05553701
2005-2006	3164	-0.0047558	0.03304326	-18.638775	423.352837	-0.8971897	0.07016093
2006-2007	3173	-0.0023043	0.01770132	-14.85696	454.656359	-0.5710052	0.10065077
2007-2008	3175	-0.000525	0.02316591	-7.9798785	190.141417	-0.4669877	0.33135714
2008-2009	3180	-0.0032144	0.01325094	1.94304365	31.410511	-0.1108144	0.20631259
2009-2010	3180	-0.0037798	0.01582807	-11.838034	460.053986	-0.5505723	0.19923065
2010-2011	3181	-0.003323	0.0149477	2.91362537	31.9363009	-0.0624274	0.2309418
2011-2012	3181	-0.0012263	0.01522113	4.28442458	46.4009601	-0.0374637	0.24946086
2012-2013	3181	-0.0045014	0.01258893	1.92180924	22.7351012	-0.1030552	0.14408503
2013-2014	3181	-0.0038836	0.02404541	23.4397267	773.337335	-0.0491327	0.86190372
2014-2015	3177	-0.0051233	0.0206625	6.04071068	413.650826	-0.5373622	0.52813783
2015-2016	3177	-0.0012069	0.02477508	-0.962289	330.896523	-0.6129388	0.54218991
2016-2017	3181	-0.005545	0.01533976	6.65379062	122.994269	-0.0612186	0.36270474
2017-2018	3181	-0.0051518	0.01602871	4.51189999	55.0157995	-0.0753494	0.28414225
2018-2019	3180	-0.0074974	0.01766174	7.68142097	129.782131	-0.0580638	0.36257556
1992-1997	2946	-0.0358571	0.04761883	-0.4867257	9.23647579	-0.556713	0.2303925
1993-1998	2946	-0.0290642	0.04646245	-0.5944167	10.0840941	-0.5515601	0.24565254
1994-1999	2946	-0.0251996	0.0462052	-0.6920577	11.8423161	-0.5387519	0.32909213
1995-2000	2946	-0.0177804	0.04611272	-0.5539825	11.5496839	-0.5134309	0.35617731
1996-2001	2948	-0.0139353	0.04321621	-0.1292183	6.41482566	-0.2862525	0.32687984
1997-2002	2948	-0.0121831	0.04804208	-2.2245013	30.5335415	-0.6698131	0.28934735
1998-2003	2948	-0.0178806	0.07404977	-5.83271	59.1903667	-1.0216817	0.29898163
1999-2004	2951	-0.0409935	0.13758036	-3.9217263	21.0651061	-1.3919355	0.20287023
2000-2005	2951	-0.0494695	0.14681731	-3.6258514	18.1989714	-1.4069136	0.20886203
2001-2006	2951	-0.0529775	0.15115344	-3.5410182	17.4362977	-1.4287608	0.21991692
2002-2007	2955	-0.0520254	0.15188445	-3.5592833	17.7105813	-1.4406233	0.25656216
2003-2008	2983	-0.0440298	0.14398967	-3.6096085	19.2668062	-1.4305901	0.39849644
2004-2009	3133	-0.0189003	0.08204037	-3.9404439	38.7541737	-1.028188	0.47048244
2005-2010	3164	-0.014829	0.06758869	-2.452236	44.4515726	-1.0544151	0.60245739
2006-2011	3173	-0.0132755	0.06339947	1.01304885	24.7481281	-0.5862494	0.81333636
2007-2012	3175	-0.0122258	0.06380425	1.91438123	28.2761775	-0.5806276	0.88888007
2008-2013	3180	-0.0160496	0.05994642	2.75158711	33.4997184	-0.5852964	0.90117471
2009-2014	3180	-0.016721	0.0629772	3.74502548	44.3716636	-0.5873924	0.93615733
2010-2015	3177	-0.0180512	0.06673069	5.77008877	86.4452663	-0.576908	1.20548517
2011-2016	3181	-0.015942	0.06690439	6.38858738	103.863756	-0.6231758	1.3251744
2012-2017	3181	-0.0202607	0.06814663	6.13829698	94.5900755	-0.6173064	1.31256923
2013-2018	3181	-0.0209111	0.07036823	5.83264087	81.8425522	-0.6260881	1.30085177
2014-2019	3180	-0.0245564	0.066521	3.7575531	35.9692864	-0.6198958	0.80582816
1992-2002	2946	-0.0480413	0.08489051	-0.564441	7.03988526	-0.7151215	0.37514734
1993-2003	2946	-0.0469426	0.10106249	-2.2984962	18.9226407	-1.0128795	0.4159533
1994-2004	2946	-0.0663083	0.15238764	-2.7564761	14.3461829	-1.4374027	0.43762222
1995-2005	2946	-0.0673766	0.16073828	-2.6009546	12.8958979	-1.431318	0.4897087
1996-2006	2948	-0.0670229	0.16393077	-2.5896382	12.6575083	-1.4409028	0.46491913
1997-2007	2948	-0.0644575	0.16579792	-2.5543474	12.5369594	-1.4421208	0.42163004
1998-2008	2948	-0.0628889	0.16881064	-2.4621552	12.2005052	-1.4293023	0.42970008
1999-2009	2951	-0.0633149	0.17128953	-2.4012872	12.0413131	-1.4148238	0.51795164
2000-2010	2951	-0.0677247	0.1741894	-2.3109567	11.8992546	-1.4264005	0.6902605
2001-2011	2951	-0.0696457	0.17807859	-2.1481574	11.8769978	-1.4440833	0.9058947
2002-2012	2955	-0.0671958	0.17932484	-2.0281782	12.3555185	-1.4616382	1.03580198
2003-2013	2983	-0.062317	0.17195257	-1.8559197	13.5474727	-1.4708517	1.14363754
2004-2014	3133	-0.0361317	0.12675629	0.27054489	18.7197754	-1.166181	1.23050077
2005-2015	3160	-0.0331119	0.1202744	1.80289988	25.5881822	-1.2055865	1.32891346
2006-2016	3173	-0.0293488	0.11868508	2.97854972	28.2109916	-0.620937	1.42410722
2007-2017	3175	-0.0325667	0.12023121	3.26697384	30.0674225	-0.6263746	1.47489902
2008-2018	3180	-0.0369673	0.120693	3.51052265	31.6031799	-0.6305716	1.45562317
2009-2019	3179	-0.0413203	0.12205404	3.54226591	30.2648196	-0.6426147	1.33000499

Table 1: Descriptive statistics of the log-growth rates of Romanian city sizes.

2013, 2016) and is given by the probability density function

$$f_{\mathrm{Lap}}(g;\mu,\beta)=\frac{1}{2\beta}\exp\left(-\frac{|g-\mu|}{\beta}\right) \tag{2}$$

where $\mu\in(-\infty,\infty)$ is a location parameter and $\beta>0$ is a shape parameter. It is known that the Laplace distribution is peaked at the mode $x=\mu$ and that its purely exponential tails are heavier than that of the normal distribution, and that would be in principle useful for our study.

A third candidate for our study would be a distribution that have asymptotically exponential tails and that has a similar shape to the normal distribution but has "longer tails" Balakrishnan (1991), that is, the logistic distribution, which is given by a probability density function like

$$f_{\mathrm{L}}(g;\mu,\sigma)=\frac{\exp\left(-\frac{g-\mu}{\sigma}\right)}{\sigma\left(1+\exp\left(-\frac{g-\mu}{\sigma}\right)\right)^{2}} \tag{3}$$

where $\mu\in(-\infty,\infty)$ and $\sigma>0$. Its cumulative distribution function can be written as

$$\mathrm{cdf}_{\mathrm{L}}(g;\mu,\sigma)=\frac{1}{1+\exp\left(-\frac{g-\mu}{\sigma}\right)} \tag{4}$$

The fourth distribution in our study will be the asymmetric double Laplace normal distribution (denoted by adLN), introduced by Reed (2002, 2003); Reed and Jorgensen (2004) and later used, e.g., by Manas (2009), see also Giesen et al. (2010); Giesen and Suedekum (2012, 2014), that is the convolution of a Laplace and a normal distribution:

$$\begin{aligned} f_{\mathrm{adLN}}(g;\alpha,\beta,\mu,\sigma) \;=\; & \frac{\alpha\beta}{(\alpha+\beta)}\left(\exp\left(-\alpha(g-\mu)+\frac{1}{2}\alpha^{2}\sigma^{2}\right)\Phi(g;\mu+\alpha\sigma^{2},\sigma)\right. \\ & \left.+\exp\left(\beta(g-\mu)+\frac{1}{2}\beta^{2}\sigma^{2}\right)\Phi(-g;-\mu+\beta\sigma^{2},\sigma)\right) \end{aligned} \tag{5}$$

where $\mu\in(-\infty,\infty)$, $\alpha,\beta,\sigma>0$ are the four parameters of the distribution. It has the property that it approximates different exponential laws in each of its two tails: $f_{\mathrm{adLn}}(g)\approx \mathrm{e}^{-\alpha g}$ when $g\to\infty$ and $f_{\mathrm{adLn}}(g)\approx \mathrm{e}^{\beta g}$ when $g\to-\infty$. The body is approximately normal, although it is not possible to exactly delineate the switch between the normal and the exponential behaviours for the reason adduced above.

Both of the Laplace and normal distributions can be encoded on another slightly more general distribution, that has different names and parameterizations and that can be shown as a continuous scale mixture of normal distributions, being the mixing function a positive stable distribution (see, e.g., Subbotin (1923); West (1987); Luévano (2013); Xu et al. (2020) and references therein). It has been used, for example, to model the log-growth rates of crime (Alves et al., 2013). The parameterization that we have chosen for the probability density function is as follows:

$$f_{\mathrm{sG}}(g;\beta,\mu,\sigma)=\frac{\beta^{-1/\beta}}{2\sigma\Gamma\left(1+1/\beta\right)}\exp\left(-\frac{1}{\beta}\left|\frac{g-\mu}{\sigma}\right|^{\beta}\right) \tag{6}$$

where $\mu\in(-\infty,\infty)$ and $\sigma,\beta>0$ are shape parameters, and $\Gamma(\cdot)$ denotes the Gamma function. When $\beta=1$ we recover the Laplace distribution, and when $\beta=2$ we recover the normal distribution. If $\beta<2$ the tails are longer than that of the normal; on the contrary if $\beta>2$ the tails are lighter than that of the normal distribution, so this distribution will be of use as well in our study.

The preceding distribution can be generalized still a little bit more, so as to being asymmetric, defined in Bottazzi and Secchi (2011) and later considered by Bottazzi and Secchi (2006); Canning et al. (1998); Delli Gatti et al. (2005); Fagiolo et al. (2008); Fu et al. (2005); Fujiwara et al. (2004); McDonald (1991); Riccaboni et al. (2008); Rozenfeld et al. (2008); Stanley et al. (1996). We will call it the asymmetric Subbotin distribution (denoted by aSub). The density function is then

$$f_{\mathrm{aSub}}(g;a_l,a_r,b_l,b_r,\mu)=\begin{cases}\frac{1}{d}\exp\left(-\frac{1}{b_l}\left|\frac{g-\mu}{a_l}\right|^{b_l}\right) & g\leq\mu\\ \frac{1}{d}\exp\left(-\frac{1}{b_r}\left|\frac{g-\mu}{a_r}\right|^{b_r}\right) & \mu\leq g\end{cases} \tag{7}$$

where $d=a_l b_l^{1/b_l}\Gamma\left(1+1/b_l\right)+a_r b_r^{1/b_r}\Gamma\left(1+1/b_r\right)$. The quantities a_l,a_r,b_l,b_r are positive shape parameters and $\mu\in(-\infty,\infty)$ is a location parameter. Particular cases of this distribution are the (asymmetric) double Laplace distribution when $b_l=b_r=1$, the normal distribution when $a_l=a_r$ and $b_l=b_r=2$ and the previous stretched Gaussian when $a_l=a_r=\sigma$ and $b_l=b_r=\beta$. The main interest of this distribution is to model the exponential tails of the observed log-growth rates (of the output of a country, or of firm sizes) and their (slight) deviations from this behaviour when the parameters b_l,b_r depart from unity. Also, it is useful when studying the scaling properties of the log-growth rates (see the references above). It has been used before in the study of the log-growth rates of city sizes in Massing et al. (2020).

The seventh distribution is the non-standardized Student's t distribution (denoted by St) for the log-growth rates g, see, e.g., Johnson et al. (1995) and references therein, which is given by the probability density function

$$f_{\mathrm{St}}(g;\mu,\sigma,\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\sigma}\left(1+\frac{1}{\nu}\left(\frac{g-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}} \tag{8}$$

where $\mu\in(-\infty,\infty)$ (location parameter), $\sigma>0$ (scale parameter) and $\nu>0$ is the number of degrees of freedom. Particular cases of this distribution are the Cauchy distribution ($\nu = 1$) and the normal distribution ($\nu = \infty$). If $1 < \nu \leq 2$, the variance of the distribution becomes infinite. This distribution has been used to study the log-growth rates of the size of German cities Schluter and Trede (2013, 2016).

Next, we call "exponential tails normal" (denoted by ETN) to the exponentiated version of the PTLN of Luckstead and Devadoss (2017); Luckstead et al. (2017); Akhundjanov et al. (2017), and is as follows Peña et al. (2021):

$$f_{\mathrm{ETN}}(g; \alpha, \tau_l, \mu, \sigma, \tau_r, \beta) = \begin{cases} de \exp(\alpha g), & -\infty < g \leq \tau_l \\ d f_{\mathrm{N}}(g; \mu, \sigma), & \tau_l \leq g \leq \tau_r \\ dc \exp(-\beta g), & \tau_r \leq g < \infty \end{cases} \tag{9}$$

where the continuity at the thresholds' τ_l and τ_r constants are $e = \dfrac{f_{\mathrm{N}}(\tau_l; \mu, \sigma)}{\exp(\tau_l \alpha)}$, $c = \dfrac{f_{\mathrm{N}}(\tau_r; \mu, \sigma)}{\exp(-\tau_r \beta)}$, and the normalization to unity constant d is given by

$$d = \left(\frac{1}{\alpha} f_{\mathrm{N}}(\tau_l; \mu, \sigma) + \Phi(\tau_r; \mu, \sigma) - \Phi(\tau_l; \mu, \sigma) + \frac{1}{\beta} f_{\mathrm{N}}(\tau_r; \mu, \sigma) \right)^{-1}.$$

The ETN distribution has, by construction, exponential tails and normal body separated by two definite thresholds: τ_l (left tail-body) and τ_r (right tail-body).

The next distribution we consider is the exponential version of the Generalized Beta 2 (or Generalized Beta Prime) distribution (McDonald, 1984; McDonald and Xu, 1995; Kleiber and Kotz, 2003) (denoted by eGB2), that has the probability density function written as

$$f_{\mathrm{eGB2}}(g; a, b, p, q) = \frac{a \exp((g-b)ap)}{B(p,q)\,(1 + \exp(a(g-b)))^{p+q}} \tag{10}$$

and cumulative distribution function

$$\mathrm{cdf}_{\mathrm{eGB2}}(g; a, b, p, q) = \frac{1}{B(p,q)} B\left(\frac{\exp(a(g-b))}{1 + \exp(a(g-b))}, p, q \right)$$

where

$$B(z, p, q) = \int_0^z t^{p-1}(1-t)^{q-1}\, dt\,, \quad z \in [0, 1]$$

is the incomplete Beta function and $B(p,q) = B(1,p,q)$ is the Beta function. The three parameters a, p, q are positive shape parameters and $b \in (-\infty, \infty)$ is a location parameter. We use this distribution because of its very well-known flexibility and because it nests as particular cases many distributions with tails heavier than that of the normal or not (see, e.g., McDonald (1984); McDonald and Xu (1995); Kleiber and Kotz (2003) and references therein).

And a distribution that uses the eGB2 as a constituent part in a composite is the double mixture exponential GB2 (denoted by dmeGB2) distribution, that has been introduced in Ramos (2017) for the first time for studying log-growth rates of USA cities. Its brief description is as follows. Consider the previous probability density function and cumulative distribution function of the eGB2 and the exponential functions

$$\begin{aligned} u(g;\zeta) &= \exp(-\zeta g) \\ l(g;\rho) &= \exp(\rho g) \end{aligned}$$

The function $u(g;\zeta)$ will model the decreasing exponential part of the upper tail of our new distribution, where $\zeta > 0$, and $l(g;\rho)$ corresponds to the increasing exponential lower tail, with $\rho > 0$. The functions u, l are not normalized at this stage as in Ioannides and Skouras (2013). The PDF of the dmeGB2 can be written as:

$$f_{\text{dmeGB2}}(g;\rho,\epsilon,\nu,a,b,p,q,\tau,\zeta,\theta) = \begin{cases} b_2[(1-\nu)\, d_2\, f_{\text{eGB2}}(g;a,b,p,q) + \nu\, e_2\, l(g;\rho)], & \infty < g \leq \epsilon \\ b_2\, f_{\text{eGB2}}(g;a,b,p,q), & \epsilon \leq g \leq \tau \\ b_2[(1-\theta)\, c_2\, f_{\text{eGB2}}(g;a,b,p,q) + \theta\, a_2\, u(g;\zeta)], & \tau \leq g < \infty \end{cases} \tag{11}$$

where the constants (i.e., quantities that do not depend on the variable g) are given as follows:

$$\begin{aligned} d_2^{-1} &= 1-\nu+\frac{\exp(-\rho\epsilon)\,\nu\,\rho\,\text{cdf}_{\text{eGB2}}(\epsilon;a,b,p,q)\,l(\epsilon;\rho)}{f_{\text{eGB2}}(\epsilon;a,b,p,q)} \\ e_2^{-1} &= \frac{(1-\nu)\,\exp(\epsilon\rho)}{\rho\,\text{cdf}_{\text{eGB2}}(\epsilon;a,b,p,q)}+\frac{\nu\,l(\epsilon;\rho)}{f_{\text{eGB2}}(\epsilon;a,b,p,q)} \\ c_2^{-1} &= 1-\theta+\frac{\zeta\,\theta\,\exp(\tau\zeta)\,(1-\text{cdf}_{\text{eGB2}}(\tau;a,b,p,q))\,u(\tau;\zeta)}{f_{\text{eGB2}}(\tau;a,b,p,q)} \\ a_2^{-1} &= \frac{(1-\theta)\,\exp(-\tau\zeta)}{\zeta\,(1-\text{cdf}_{\text{eGB2}}(\tau;a,b,p,q))}+\frac{\theta\,u(\tau;\zeta)}{f_{\text{eGB2}}(\tau;a,b,p,q)} \\ b_2^{-1} &= e_2\,\frac{\exp(\epsilon\rho)}{\rho}+\text{cdf}_{\text{eGB2}}(\tau;a,b,p,q)-\text{cdf}_{\text{eGB2}}(\epsilon;a,b,p,q)+\frac{a_2}{\zeta\,\exp(\tau\zeta)} \end{aligned}$$

This distribution depends on 10 parameters $(\rho,\epsilon,\nu,a,b,p,q,\tau,\zeta,\theta)$.

These are the single distributions that we use. However, inspired by the original idea of Kwong and Nadarajah (2019) when studying USA city sizes, that is, to consider mixtures of single distributions, that has yielded so good results, we have applied the same idea to city sizes (Puente-Ajovín et al., 2020a,b) and its log-growth rates (Massing et al., 2020), log-returns of stock indices (Massing and Ramos, 2021) and log-growth rates of CO_2 emissions (Peña et al., 2021), for example, with a remarkable success. Also, mixtures of distributions of log-growth rates allow for sub-populations of these in the samples (McLachlan and Peel, 2003) so this allows for heterogeneity in the processes that drive the log-growth rates. In this study we therefore consider mixtures of some basic single functions defined before, dealing with them in as an equal footing as possible. We have chosen the maximal number of components as five, hopping that we would obtain parsimony given by the information criteria at a certain smaller number of components.

They are:

The 2Lap, 3Lap, 4Lap, 5Lap defined, respectively, by:

$$f_{2\mathrm{Lap}}(g;\mu_1,\beta_1,\mu_2,\beta_2,p_1) = p_1 f_{\mathrm{Lap}}(g;\mu_1,\beta_1) + (1-p_1) f_{\mathrm{Lap}}(g;\mu_2,\beta_2) \tag{12}$$

where $0 \le p_1 \le 1$,

$$f_{3\mathrm{Lap}}(g;\mu_1,\beta_1,\mu_2,\beta_2,\mu_3,\beta_3,p_1,p_2) = p_1 f_{\mathrm{Lap}}(g;\mu_1,\beta_1) + p_2 f_{\mathrm{Lap}}(g;\mu_2,\beta_2) + (1-p_1-p_2) f_{\mathrm{Lap}}(g;\mu_3,\beta_3) \tag{13}$$

where $0 \le p_1, p_2, 1-p_1-p_2 \le 1$,

$$\begin{aligned} f_{4\mathrm{Lap}}(g;\mu_1,\beta_1,\mu_2,\beta_2,\mu_3,\beta_3,\mu_4,\beta_4,p_1,p_2,p_3) &= p_1 f_{\mathrm{Lap}}(g;\mu_1,\beta_1) + p_2 f_{\mathrm{Lap}}(g;\mu_2,\beta_2) \\ &\quad + p_3 f_{\mathrm{Lap}}(g;\mu_3,\beta_3) + (1-p_1-p_2-p_3) f_{\mathrm{Lap}}(g;\mu_4,\beta_4) \end{aligned} \tag{14}$$

where $0 \le p_1, p_2, p_3, 1-p_1-p_2-p_3 \le 1$,

$$\begin{aligned} & f_{5\mathrm{Lap}}(g;\mu_1,\beta_1,\mu_2,\beta_2,\mu_3,\beta_3,\mu_4,\beta_4,\mu_5,\beta_5,p_1,p_2,p_3,p_4) \\ & = p_1 f_{\mathrm{Lap}}(g;\mu_1,\beta_1) + p_2 f_{\mathrm{Lap}}(g;\mu_2,\beta_2) + p_3 f_{\mathrm{Lap}}(g;\mu_3,\beta_3) \\ & + p_4 f_{\mathrm{Lap}}(g;\mu_4,\beta_4) + (1-p_1-p_2-p_3-p_4) f_{\mathrm{Lap}}(g;\mu_5,\beta_5) \end{aligned} \tag{15}$$

where $0 \le p_1, p_2, p_3, p_4, 1-p_1-p_2-p_3-p_4 \le 1$.

Likewise, the 2sG, 3sG, 4sG and 5sG are given, respectively, by the expressions (the p_j obey similar values):

$$f_{2\mathrm{sG}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,p_1) = p_1 f_{\mathrm{sG}}(g;1.2,\mu_1,\sigma_1) + (1-p_1) f_{\mathrm{sG}}(g;2.2,\mu_2,\sigma_2) \tag{16}$$

$$f_{3\mathrm{sG}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2) = p_1 f_{\mathrm{sG}}(g;1.2,\mu_1,\sigma_1) + p_2 f_{\mathrm{sG}}(g;2.2,\mu_2,\sigma_2) + (1-p_1-p_2) f_{\mathrm{sG}}(g;1.4,\mu_3,\sigma_3) \tag{17}$$

$$\begin{aligned} f_{4\mathrm{sG}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,\mu_4,\sigma_4,p_1,p_2,p_3) &= p_1 f_{\mathrm{sG}}(g;1.2,\mu_1,\sigma_1) + p_2 f_{\mathrm{sG}}(g;2.2,\mu_2,\sigma_2) \\ &\quad + p_3 f_{\mathrm{sG}}(g;1.4,\mu_3,\sigma_3) + (1-p_1-p_2-p_3) f_{\mathrm{sG}}(g;1.6,\mu_4,\sigma_4) \end{aligned}$$

$$\begin{aligned} & f_{5\mathrm{sG}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,\mu_4,\sigma_4,\mu_5,\sigma_5,p_1,p_2,p_3,p_4) \\ & = p_1 f_{\mathrm{sG}}(g;1.2,\mu_1,\sigma_1) + p_2 f_{\mathrm{sG}}(g;2.2,\mu_2,\sigma_2) + p_3 f_{\mathrm{sG}}(g;1.4,\mu_3,\sigma_3) \\ & + p_4 f_{\mathrm{sG}}(g;1.6,\mu_4,\sigma_4) + (1-p_1-p_2-p_3-p_4) f_{\mathrm{sG}}(g;1.8,\mu_5,\sigma_5) \end{aligned} \tag{19}$$

In a similar fashion, the 2St12,2St39,3St,4St,5St are defined, respectively, by

$$f_{2\mathrm{St}12}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,p_1) = p_1 f_{\mathrm{St}}(g;\mu_1,\sigma_1,4) + (1-p_1) f_{\mathrm{St}}(g;\mu_2,\sigma_2,12) \tag{20}$$

$$f_{2\mathrm{St}39}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,p_1) = p_1 f_{\mathrm{St}}(g;\mu_1,\sigma_1,4) + (1-p_1) f_{\mathrm{St}}(g;\mu_2,\sigma_2,39) \tag{21}$$

$$f_{3\mathrm{St}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2) = p_1 f_{\mathrm{St}}(g;\mu_1,\sigma_1,4) + p_2 f_{\mathrm{St}}(g;\mu_2,\sigma_2,12) + (1-p_1-p_2) f_{\mathrm{St}}(g;\mu_3,\sigma_3,39) \tag{22}$$

$$\begin{aligned} f_{4\mathrm{St}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,\mu_4,\sigma_4,p_1,p_2,p_3) &= p_1 f_{\mathrm{St}}(g;\mu_1,\sigma_1,4) + p_2 f_{\mathrm{St}}(g;\mu_2,\sigma_2,12) \\ &\quad + p_3 f_{\mathrm{St}}(g;\mu_3,\sigma_3,39) + (1-p_1-p_2-p_3) f_{\mathrm{St}}(g;\mu_4,\sigma_4,100) \end{aligned} \tag{23}$$

$$\begin{aligned} & f_{5\mathrm{St}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,\mu_4,\sigma_4,\mu_5,\sigma_5,p_1,p_2,p_3,p_4) \\ &= p_1 f_{\mathrm{St}}(g;\mu_1,\sigma_1,4) + p_2 f_{\mathrm{St}}(g;\mu_2,\sigma_2,12) + p_3 f_{\mathrm{St}}(g;\mu_3,\sigma_3,39) \\ &+ p_4 f_{\mathrm{St}}(g;\mu_4,\sigma_4,100) + (1-p_1-p_2-p_3-p_4) f_{\mathrm{St}}(g;\mu_5,\sigma_5,200) \end{aligned} \tag{24}$$

Note that in the mixtures of sG and St we have set *a priori* some values for the parameters governing the fat-tailedness. This have been done mainly for achieving more stability in the maximum likelihood estimation process; otherwise the estimation is very unstable perhaps because the corresponding parameters enter into the probability density functions as both an exponent of the studied variable g and an argument of one Gamma function or more. The chosen specific values account for thin and fat-tailedness in the mixtures, that is, that we allow for both types of behaviour in such mixtures. This procedure for the 2St12, 2St39, 3St has been successful, for example, in Massing et al. (2020); Massing and Ramos (2021).

In a similar fashion, the 2N,3N,4N,5N are defined, respectively, by

$$f_{2\mathrm{N}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,p_1) = p_1 f_{\mathrm{N}}(g;\mu_1,\sigma_1) + (1-p_1) f_{\mathrm{N}}(g;\mu_2,\sigma_2) \tag{25}$$

$$f_{3\mathrm{N}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2) = p_1 f_{\mathrm{N}}(g;\mu_1,\sigma_1) + p_2 f_{\mathrm{N}}(g;\mu_2,\sigma_2) + (1-p_1-p_2) f_{\mathrm{N}}(g;\mu_3,\sigma_3) \tag{26}$$

$$\begin{aligned} f_{4\mathrm{N}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,\mu_4,\sigma_4,p_1,p_2,p_3) &= p_1 f_{\mathrm{N}}(g;\mu_1,\sigma_1) + p_2 f_{\mathrm{N}}(g;\mu_2,\sigma_2) \\ &\quad + p_3 f_{\mathrm{N}}(g;\mu_3,\sigma_3) + (1-p_1-p_2-p_3) f_{\mathrm{N}}(g;\mu_4,\sigma_4) \end{aligned} \tag{27}$$

$$\begin{aligned} & f_{5\mathrm{N}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,\mu_4,\sigma_4,\mu_5,\sigma_5,p_1,p_2,p_3,p_4) \\ &= p_1 f_{\mathrm{N}}(g;\mu_1,\sigma_1) + p_2 f_{\mathrm{N}}(g;\mu_2,\sigma_2) + p_3 f_{\mathrm{N}}(g;\mu_3,\sigma_3) \\ &+ p_4 f_{\mathrm{N}}(g;\mu_4,\sigma_4) + (1-p_1-p_2-p_3-p_4) f_{\mathrm{N}}(g;\mu_5,\sigma_5) \end{aligned} \tag{28}$$

And, finally, the 2L, 3L, 4L, 5L are defined by

$$f_{2\mathrm{L}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,p_1) = p_1 f_{\mathrm{L}}(g;\mu_1,\sigma_1) + (1-p_1) f_{\mathrm{L}}(g;\mu_2,\sigma_2) \tag{29}$$

$$f_{3\mathrm{L}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2) = p_1 f_{\mathrm{L}}(g;\mu_1,\sigma_1)+p_2 f_{\mathrm{L}}(g;\mu_2,\sigma_2)+(1-p_1-p_2) f_{\mathrm{L}}(g;\mu_3,\sigma_3) \tag{30}$$

$$\begin{aligned} f_{4\mathrm{L}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,\mu_4,\sigma_4,p_1,p_2,p_3) &= p_1 f_{\mathrm{L}}(g;\mu_1,\sigma_1) + p_2 f_{\mathrm{L}}(g;\mu_2,\sigma_2) \\ &\quad + p_3 f_{\mathrm{L}}(g;\mu_3,\sigma_3) + (1-p_1-p_2-p_3) f_{\mathrm{L}}(g;\mu_4,\sigma_4) \end{aligned} \tag{31}$$

$$\begin{aligned} & f_{5\mathrm{L}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,\mu_4,\sigma_4,\mu_5,\sigma_5,p_1,p_2,p_3,p_4) \\ & = p_1 f_{\mathrm{L}}(g;\mu_1,\sigma_1) + p_2 f_{\mathrm{L}}(g;\mu_2,\sigma_2) + p_3 f_{\mathrm{L}}(g;\mu_3,\sigma_3) \\ & + p_4 f_{\mathrm{L}}(g;\mu_4,\sigma_4) + (1-p_1-p_2-p_3-p_4) f_{\mathrm{L}}(g;\mu_5,\sigma_5) \end{aligned} \tag{32}$$

3.1 The estimation procedure

In this paper we have estimated the parameters of all the distributions

$$B \in \{\mathrm{N}, \mathrm{Lap}, \mathrm{L}, \mathrm{adLN}, \mathrm{sG}, \mathrm{aSub}, \mathrm{St}, \mathrm{ETN}, \mathrm{eGB2}, \mathrm{dmeGB2}, 2\mathrm{Lap}, 3\mathrm{Lap}, 4\mathrm{Lap}, 5\mathrm{Lap}, \\ 2\mathrm{sG}, 3\mathrm{sG}, 4\mathrm{sG}, 5\mathrm{sG}, 2\mathrm{St}12, 2\mathrm{St}39, 3\mathrm{St}, 4\mathrm{St}, 5\mathrm{St}, 2\mathrm{N}, 3\mathrm{N}, 4\mathrm{N}, 5\mathrm{N}, 2\mathrm{L}, 3\mathrm{L}, 4\mathrm{L}, 5\mathrm{L}\}$$

presented in Section 3 for the log-growth rates of Romanian cities' populations by numerically maximizing the log-likelihood $\ell_B(\theta) = \sum_{i=1}^n \log f_B(x_i;\theta)$, where $\theta = (\theta_1, \ldots, \theta_k)$ is the vector of parameters, using the command tmle of the MATLAB® software package, on an equal footing for all the parameters.

3.2 The statistical criteria

The aim of this subsection is to assess which model is the best fit for the log-growth rates of Romanian city populations. Note that the number of parameters k depends on the specific model B. For the goodness-of-fit criteria, we use the well-known Kolmogorov–Smirnov (KS) (Kolmogorov, 1933), Cramér–von Mises (CM) (Cramér, 1928; von Mises, 1928) and Anderson–Darling (AD) (Anderson and Darling, 1954) tests. We generate random samples of size 100,000 out of each estimated distribution, and obtain the statistics and also the p-values by Monte Carlo simulation of the distributions of the statistics by 10,000 Monte Carlo samples. The KS and CM statistics better reflect the deviance between the empirical and the fitted distribution close to the center and the AD statistic better reflects the deviance in the tails (Razali and Wah, 2011; Cirillo, 2013). Generally, the KS, CM and AD statistics do not account for over-fitting. However, it is possible that a special or limiting case may have a lower distance due to the nature of the distances and the distributions. Anyway, the lower the test's statistics among the competitors, the better the fit in *absolute terms*, that is, without taking into account the issue of over-fitting.

To remedy the issue of over-fitting, it could be useful to employ *Information Criteria* that account for the number of parameters and minimize the mentioned risk. The ones that we use in this study are well known and are the following:

- The *Akaike information criterion (AIC)* (Akaike, 1974) is defined by

$$AIC = 2k - 2\ell_B(\hat{\theta}).$$

- The *Bayesian information criterion (BIC)* (Schwarz, 1978; Figueiredo, M. A. T. and Leitão, J. M. N. and Jain, A. K., 2000; Grünwald, 2007; Grünwald and Roos, 2020) is defined by

$$BIC = k\log(n) - 2\ell_B(\hat{\theta}).$$

- The *Hannan–Quinn Information Criterion (HQC)* (Hannan and Quinn, 1979; Burnham and Anderson, 1998, 2002, 2004) is defined by

$$HQC = 2k\log(\log(n)) - 2\ell_B(\hat{\theta}).$$

where in all these three information criteria $\ell_B(\hat{\theta})$ is the maximum log-likelihood for the model B at the ML estimator $\hat{\theta}$, k is the number of parameters of the distribution B and n is the sample size.

For each information criterion, the model with the smallest statistic is considered to be the best fit among the models investigated, in information theoretic terms. The AIC, BIC and HQC statistics adjust the log-likelihood by penalizing models that are too large to avoid over-fitting, more heavily the BIC than the AIC and the HQC.

4 Results

The detailed results of the estimation processes are shown in respective sheets of an Excel book that is offered as supplementary material. We do not offer this information in the paper because of the huge size of the tables obtained.

We have computed the KS, CM and AD tests' statistics and corresponding p-values for each of the models and empirical data samples We offer the detailed results of the KS, CM and AD statistics for each model and sample in the mentioned Excel book that is supplied as a supplementary material. However, we present briefly the results of estimations found and the number of times each model is rejected or non-rejected at the 5% level for each of the tests in Table 2. We have also added an index showing the percentage of the times is non-rejected (each one scores with a +1, maximum: 100%) or rejected (each one scores with a -1, minimum: -100%), meanwhile the non-estimated cases score with a zero in such an index.

	Non estimated	KS rejected	KS Non-rejected	CM rejected	CM Non-rejected	AD rejected	AD Non-rejected	Index
N	0	68	0	68	0	68	0	-100%
Lap	3	61	4	57	8	65	0	-84%
L	0	60	8	60	8	62	6	-78%
adLN	5	42	21	33	30	42	21	-22%
sG	0	57	11	56	12	61	7	-71%
aSub	0	45	23	41	27	51	17	-34%
St	0	34	34	27	41	47	21	-6%
ETN	0	21	47	13	55	18	50	49%
eGB2	5	45	18	39	24	44	17	-34%
dmeGB2	1	0	67	0	67	0	67	99%
2Lap	4	42	22	37	27	42	22	-25%
3Lap	2	3	63	2	64	2	64	90%
4Lap	3	0	65	0	65	0	65	96%
5Lap	19	1	48	1	48	1	48	69%
2sG	0	28	40	28	40	32	36	14%
3sG	0	0	68	0	68	0	68	100%
4sG	2	0	66	0	66	0	66	97%
5sG	7	0	61	0	61	0	61	90%
2St12	0	4	64	6	62	6	62	84%
2St39	0	11	57	10	58	12	56	68%
3St	0	0	68	0	68	0	68	100%
4St	4	0	64	0	64	0	64	94%
5St	4	0	64	0	64	0	64	94%
2N	0	31	37	29	39	31	37	11%
3N	1	4	63	1	66	1	66	93%
4N	3	0	65	0	65	0	65	96%
5N	4	0	64	0	64	0	64	94%
2L	3	1	64	2	63	6	59	87%
3L	4	1	63	1	63	1	63	91%
4L	5	0	63	0	63	0	63	93%
5L	37	0	31	0	31	0	31	46%

Table 2: Summary of the estimations found and the rejection/non-rejection status and index for each model.

It is to be remarked that the normal (N) distribution can be estimated always and is rejected *always* (index -100%.) meaning that Gibrat's Law in its strict formulation (Gibrat, 1931; Sutton, 1997) is rejected always for the log-growth rates processes' of Romanian cities in the relatively ample studied period. On the contrary, the 3sG and 3St attain a maximum of non-rejections (index 100%) so they seem to provide a very appropriate fit for the data sets of Romanian log-growth rates. There are other models with very high indexes like the dmeGB2 and the mixtures with 4 and 5 components.

With respect to the times that each model attain the lowest statistics, we provide a brief account in Table 3. In it we can observe as a remarkable model the 5sG with an average of about 26% of minimum test's statistics, and they are followed to a great distance by the 5St and 5L, with about 14% of average minimum statistics.

Thus we obtain a first result: the 5sG, 5St and 5L are the distributions that offer a best fit in terms of empirical distances of the used statistics, out of the compared distributions.

However, we have to compare the models in terms of the information criteria defined above, in order to select the model or models that offer the best information, that is, taking into account the possible over-fitting that could take place. We show in another sheet of the Excel book the detailed values of the information criteria AIC, BIC, HQC and a selection of the models by sample and criterion. Likewise, the summary

	Min KS	Min KS %	Min CM	Min CM %	Min AD	Min AD %	Total	Total %
N	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Lap	0	0.00%	0	0.00%	0	0.00%	0	0.00%
L	0	0.00%	0	0.00%	0	0.00%	0	0.00%
adLN	0	0.00%	0	0.00%	0	0.00%	0	0.00%
sG	0	0.00%	0	0.00%	0	0.00%	0	0.00%
aSub	0	0.00%	0	0.00%	0	0.00%	0	0.00%
St	0	0.00%	0	0.00%	0	0.00%	0	0.00%
ETN	0	0.00%	0	0.00%	0	0.00%	0	0.00%
eGB2	0	0.00%	0	0.00%	0	0.00%	0	0.00%
dmeGB2	0	0.00%	1	1.47%	1	1.47%	2	0.98%
2Lap	0	0.00%	0	0.00%	0	0.00%	0	0.00%
3Lap	0	0.00%	1	1.47%	0	0.00%	1	0.49%
4Lap	2	2.94%	2	2.94%	0	0.00%	4	1.96%
5Lap	1	1.47%	2	2.94%	0	0.00%	3	1.47%
2sG	0	0.00%	0	0.00%	0	0.00%	0	0.00%
3sG	0	0.00%	0	0.00%	0	0.00%	0	0.00%
4sG	2	2.94%	2	2.94%	3	4.41%	7	3.43%
5sG	19	27.94%	19	27.94%	15	22.06%	53	25.98%
2St12	0	0.00%	0	0.00%	0	0.00%	0	0.00%
2St39	1	1.47%	0	0.00%	0	0.00%	1	0.49%
3St	4	5.88%	3	4.41%	2	2.94%	9	4.41%
4St	3	4.41%	7	10.29%	6	8.82%	16	7.84%
5St	8	11.76%	6	8.82%	15	22.06%	29	14.22%
2N	0	0.00%	0	0.00%	0	0.00%	0	0.00%
3N	1	1.47%	0	0.00%	0	0.00%	1	0.49%
4N	6	8.82%	2	2.94%	4	5.88%	12	5.88%
5N	7	10.29%	6	8.82%	5	7.35%	18	8.82%
2L	1	1.47%	0	0.00%	0	0.00%	1	0.49%
3L	1	1.47%	1	1.47%	0	0.00%	2	0.98%
4L	6	8.82%	6	8.82%	5	7.35%	17	8.33%
5L	6	8.82%	10	14.71%	12	17.65%	28	13.73%
Total	68	100.00%	68	100.00%	68	100.00%	204	100.00%

Table 3: Summary of the minimum KS, CM, AD statistics for each model.

of the results is offered in Table 4. In it we can observe that the average number of minimum distances is achieved by the 3L by an ample margin (average about 26%) followed by the 2St12 (about 16%) and the 3St (about 10%).

Thus we obtain the result that the 3L, 2St12 and 3St offer the best choices in terms of information criteria, achieving in fact the parsimony in the choice of the models. Thus, the number of components needed to obtain an optimal selection in information criteria terms reduces to only 3 or 2.

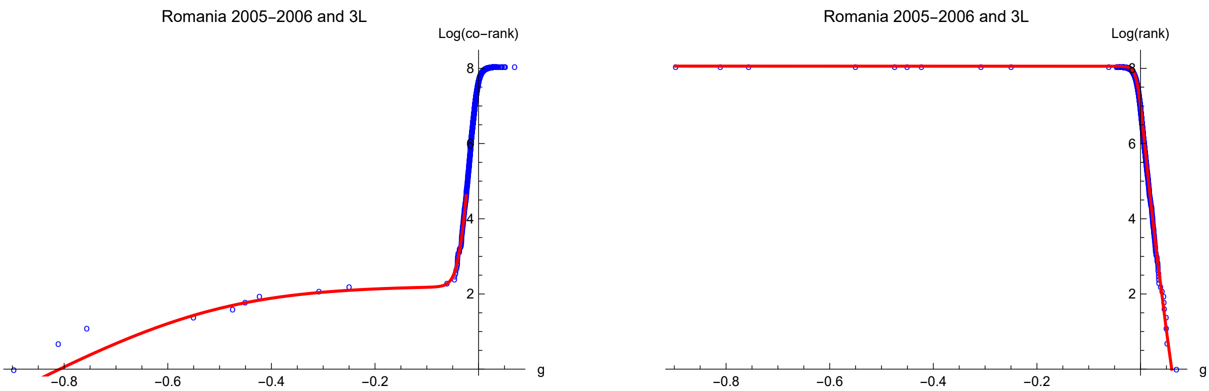
We offer as well a figure showing graphically the goodness-of-fit of an instance where the model 3L is selected by each of the three information criteria AIC, BIC, HQC, namely the interval 2005-2006, in Figure 1. Other figures are kindly provided under reasonable request. In particular, the shown fits are observed to be very good, taking moreover into account that the discrepancies at the tails are amplified by the fact of taking logarithms of the ranks or co-ranks González-Val et al. (2013). Apart from that, we show in Figure 2 the plot of the posterior probabilities, defined as

$$\begin{aligned}
\tau_{1,3\mathrm{L}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2) &= p_1 f_{\mathrm{L}}(g;\mu_1,\sigma_1)/f_{3\mathrm{L}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2)\\
\tau_{2,3\mathrm{L}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2) &= p_2 f_{\mathrm{L}}(g;\mu_2,\sigma_2)/f_{3\mathrm{L}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2)\\
\tau_{3,3\mathrm{L}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2) &= (1-p_1-p_2) f_{\mathrm{L}}(g;\mu_3,\sigma_3)/f_{3\mathrm{L}}(g;\mu_1,\sigma_1,\mu_2,\sigma_2,\mu_3,\sigma_3,p_1,p_2)
\end{aligned}$$

of the same mixture 3L and the sample of Romanian log-growth rates in the period 2005-2006. In this way we can separate to some extent the data into groups that have different probabilities to be modelled by different components in the given mixture (McLachlan and Peel, 2003).

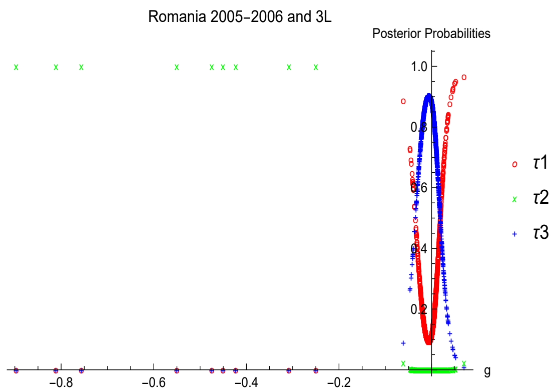
	Min AIC	Min AIC %	Min BIC	Min BIC %	Min HQC	Min HQC %	Total	Total %
N	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Lap	0	0.00%	0	0.00%	0	0.00%	0	0.00%
L	0	0.00%	1	1.47%	0	0.00%	1	0.49%
adLN	0	0.00%	0	0.00%	0	0.00%	0	0.00%
sG	0	0.00%	1	1.47%	1	1.47%	2	0.98%
aSub	0	0.00%	0	0.00%	0	0.00%	0	0.00%
St	0	0.00%	6	8.82%	0	0.00%	6	2.94%
ETN	2	2.94%	3	4.41%	3	4.41%	8	3.92%
eGB2	1	1.47%	0	0.00%	1	1.47%	2	0.98%
dmeGB2	2	2.94%	1	1.47%	2	2.94%	5	2.45%
2Lap	0	0.00%	0	0.00%	0	0.00%	0	0.00%
3Lap	0	0.00%	0	0.00%	0	0.00%	0	0.00%
4Lap	1	1.47%	0	0.00%	0	0.00%	1	0.49%
5Lap	0	0.00%	0	0.00%	0	0.00%	0	0.00%
2sG	0	0.00%	0	0.00%	0	0.00%	0	0.00%
3sG	1	1.47%	2	2.94%	4	5.88%	7	3.43%
4sG	4	5.88%	0	0.00%	0	0.00%	4	1.96%
5sG	2	2.94%	0	0.00%	0	0.00%	2	0.98%
2St12	5	7.35%	16	23.53%	11	16.18%	32	15.69%
2St39	3	4.41%	2	2.94%	3	4.41%	8	3.92%
3St	5	7.35%	7	10.29%	9	13.24%	21	10.29%
4St	7	10.29%	0	0.00%	1	1.47%	8	3.92%
5St	1	1.47%	0	0.00%	0	0.00%	1	0.49%
2N	0	0.00%	0	0.00%	0	0.00%	0	0.00%
3N	0	0.00%	0	0.00%	0	0.00%	0	0.00%
4N	2	2.94%	0	0.00%	1	1.47%	3	1.47%
5N	4	5.88%	0	0.00%	0	0.00%	4	1.96%
2L	0	0.00%	13	19.12%	3	4.41%	16	7.84%
3L	15	22.06%	16	23.53%	24	35.29%	55	26.96%
4L	12	17.65%	0	0.00%	5	7.35%	17	8.33%
5L	1	1.47%	0	0.00%	0	0.00%	1	0.49%
Total	68	100.00%	68	100.00%	68	100.00%	204	100.00%

Table 4: Summary of the minimum AIC, BIC, HQC distances for each model.



((a)) ((b))

Figure 1: Log-corank/log-rank plots for the data and the model 3L for Romania 2005-2006. The lower tail is clearly non-exponential. The upper tail is indistinguishable visually from an exponential (straight line in log plot).



((a))

Figure 2: Posterior probabilities for the data and the model 3L for Romania 2005-2006. It is seen that the extreme lower tail is modelled by the second component in the mixture, meanwhile the upper tail is modelled mainly by its first component.

5 A generating mechanism

The log-growth rates of Romanian city populations that we have at hand are collected at specific moments in time, i.e., on a yearly, quinquennial or decennial basis, and one could observe the probability density functions of the previous log-growth rates at that given times. If one uses a parametric description, the parameters of the corresponding distribution can be estimated at those moments in time. If one interpolates in a smooth way (for example, by higher order polynomials) the values of the parameters for all times, one could construct a time-dependent probability density function $f(g,t)$ that approximates the true one that could be observed if we had log-growth rates data on an approximately continuous time basis.[1] Another possibility could be to estimate a three dimensional stochastic kernel $f(g,t)$ to approximate the true one (see, e.g., Quah (1997)). Then, let us derive a stochastic process whose associated $f(g,t)$ is one of the so constructed approximations.

In order to achieve this task, recall that our log-growth variable $g \in (-\infty, \infty)$, and let us denote its dependence on time by g_t. We assume that its evolution or dynamics is governed by the Itô differential equation (see, e.g., Ord (1974); Gardiner (2004))

$$dg_t = b(g_t,t)dt + \sqrt{a(g_t,t)}dB_t \tag{33}$$

where B_t is a standard Brownian motion (Wiener process) (see, e.g.,Itô and McKean Jr. (1996); Kyprianou (2006) and references therein). The quantity $a(g_t,t)$ corresponds to the *diffusion term*, and $b(g_t,t)$ to the *drift term*. This process can be associated to the *forward Kolmogorov equation* or *Fokker-Planck equation* for the time-dependent probability density function (conditional on the initial data) $f(g,t)$ (see also Gabaix (1999, 2009)):

$$\frac{\partial f(g,t)}{\partial t} = -\frac{\partial}{\partial g}(b(g,t)f(g,t)) + \frac{1}{2}\frac{\partial^2}{\partial g^2}(a(g,t)f(g,t)). \tag{34}$$

Since the approximate probability density function $f(g,t)$ is evolving on time and perhaps there is no limiting stationary distribution, let us propose a way of solving (34) for the cited $f(g,t)$ by specifying the diffusion term and the drift term (see, e.g., Otunuga (2021) for another recent approach to time-dependent solutions of the Fokker–Planck equation). In fact, if we take $a(g,t) = s^2$, where $s > 0$ is a real constant, then by choosing

$$b(g,t) = \frac{s^2}{2f(g,t)}\frac{\partial f(g,t)}{\partial g} - \frac{1}{f(g,t)}\frac{\partial\,\mathrm{cdf}(g,t)}{\partial t} \tag{35}$$

[1] It is not inconceivable that the population data of a country like Romania would be collected in real time in the near future.

where $\mathrm{cdf}(g,t)$ is the CDF corresponding to $f(g,t)$, it is solved (34) for $f(g,t)$ Dupire (1993, 1994). However, we remark that we do not claim that the solution of (34) with this choice of $a(g,t)$ and $b(g,t)$ be unique, only that $f(g,t)$ is a solution by construction. Also, $b(g,t)$ might have bounded discontinuities in the variable g, as in the case of the ETN, in a finite number of points in the domain (Gikhman and Skorokhod, 2007). And a third remark is that we may add to the expression of $b(g,t)$ in (35) a term of the form $h(t)/f(g,t)$, where $h(t)$ is an arbitrary function of t.

With this set-up in mind, let us develop the corresponding expressions for the distribution 3L in this paper, that is selected most of the time by the information criteria. What comes next are variations of the expressions for the N, 2N, 3N that appeared first in Campolieti and Ramos (2021); Peña et al. (2021).

Thus, for the 3L, let us define the following expression:

$$k_{\mathrm{L}}(g;\mu,\sigma,s)=\dot{\mu}+(g-\mu)\frac{\dot{\sigma}}{\sigma}-\frac{s^2}{2\sigma}\tanh\left(\frac{g-\mu}{2\sigma}\right)$$

where μ is real and $\sigma>0$ are supposed to depend smoothly on t (explicit dependence is omitted for notational simplicity), the dot means derivative with respect to t, and $s>0$ is a real constant. If we denote

$$j_{3\mathrm{L}}(g,t)=f_{3\mathrm{L}}(g;\mu_1(t),\sigma_1(t),\mu_2(t),\sigma_2(t),\mu_3(t),\sigma_3(t),p_1(t),p_2(t))\,,$$

the quantities

$$\begin{aligned}\pi_{3\mathrm{L},1}(g,t)&=\frac{\mathrm{cdf}_{\mathrm{L}}(g;\mu_1(t),\sigma_1(t))-\mathrm{cdf}_{\mathrm{L}}(g;\mu_3(t),\sigma_3(t))}{j_{3\mathrm{L}}(g,t)}\\ \pi_{3\mathrm{L},2}(g,t)&=\frac{\mathrm{cdf}_{\mathrm{L}}(g;\mu_2(t),\sigma_2(t))-\mathrm{cdf}_{\mathrm{L}}(g;\mu_3(t),\sigma_3(t))}{j_{3\mathrm{L}}(g,t)}\end{aligned}$$

and the time-dependent posterior probabilities

$$\begin{aligned}\tau_{3\mathrm{L},1}(g,t)&=p_1(t)f_{\mathrm{L}}(g;\mu_1(t),\sigma_1(t))/j_{3\mathrm{L}}(g,t)\\ \tau_{3\mathrm{L},2}(g,t)&=p_2(t)f_{\mathrm{L}}(g;\mu_2(t),\sigma_2(t))/j_{3\mathrm{L}}(g,t)\\ \tau_{3\mathrm{L},3}(g,t)&=(1-p_1(t)-p_2(t))f_{\mathrm{L}}(g;\mu_3(t),\sigma_3(t))/j_{3\mathrm{L}}(g,t)\end{aligned}$$

then, if we select $a(g,t)=s^2$ for $s>0$ constant, and

$$b(g,t)=k_{\mathrm{L}}(g;\mu_1,\sigma_1,s)\tau_{3\mathrm{L},1}(g,t)+k_{\mathrm{L}}(g;\mu_2,\sigma_2,s)\tau_{3\mathrm{L},2}(g,t)+k_{\mathrm{L}}(g;\mu_3,\sigma_3,s)\tau_{3\mathrm{L},3}(g,t)-\dot{p}_1\pi_{3\mathrm{L},1}(g,t)-\dot{p}_2\pi_{3\mathrm{L},} \tag{36}$$

so that we obtain that $f(g,t)=j_{3\mathrm{L}}(g,t)$ is a solution of the corresponding Fokker–Planck equation (34). The sign of the drift term is indefinite and generalizes, amongst other things, the *mean reverting* of Uhlenbeck and Ornstein (1930); Kalecki (1945); Vasicek

(1977). It is interesting as well to note that the structure of the drift term so obtained can be understood as "the contribution of each component multiplied by each time-dependent posterior probability minus the contribution of the weighted differences of the cumulative distribution functions". In this way it is easily seen the influence of the posterior probabilities, that help to identify the sub-populations in the samples and therefore the heterogeneity in the growth processes, in the generation of the drift term of the stochastic equation. This structure is common to any time-dependent mixture like the ones studied in this paper, and can be generalized to other mixtures, starting from (35), although the actual expressions may vary and be a little bit more complicated.

6 Conclusions

We have compared thirty-one different parametric statistical distributions to fit the log-growth rates of Romanian cities in the period 1992-2019 on an annual, quinquennial and decennial basis.

The main conclusions are that the best distributions out of the previously mentioned thirty-one ones depend on the type of statistical quantities used. For the standard KS, CM, AD tests the best models out of those studied here are the 5sG, 5St and 5L. For the 5sG and 5St it is even possible in principle to improve the results by choosing other values of the parameters that control the tail-heaviness, although it might be a very difficult task. With regards to the information criteria, the three of AIC, BIC, HQC point out clearly to the simple 3L distribution (mixture of three logistic distributions). For this model we have constructed a stochastic differential equation that generates the observed time-dependent distributions solving an associated Fokker–Planck equation, like in Campolieti and Ramos (2021); Peña et al. (2021), showing in addition the contribution of the time-dependent posterior probabilities and therefore of the existing heterogeneity of the growth processes. Indeed the fact that the best models are mixtures, shows that there exists heterogeneity in the growth process of Romanian city sizes. The fit is really good for these distributions as log-rank/log-corank plots can show. In any case, the simple normal distribution is always rejected and as a byproduct is also strongly rejected the Gibrat's law in its strict sense (Gibrat, 1931; Sutton, 1997). Thus these results complement and enrich those obtained, for example, in the articles Ramos (2017); Massing et al. (2020) and we set in this way a strand of the literature with high precision fitting of log-growth rates of cities' sizes.

Author contributions

Irina Băncescu: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, supervision, validation, visualization, writing-original draft, writing-review & editing. Luminiţa Chivu: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, supervision, validation, visualization, writing-original draft, writing-review & editing. Till Massing: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, supervision, validation, visualization, writing-original draft, writing-review & editing. Vasile Preda: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, supervision, validation, visualization, writing-original draft, writing-review & editing. Miguel Puente-Ajovín: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, supervision, validation, visualization, writing-original draft, writing-review & editing. Arturo Ramos: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, supervision, validation, visualization, writing-original draft, writing-review & editing.

AI Statement

In this study no artificial intelligence (AI) has been used.

Competing interests statement

The authors declare to have no competing interests concerning the research carried out in this article.

Acknowledgments

The work of Till Massing has been supported by *Deutsche Forschungsgemeinschaft (DFG)* Grant Number 455257011. The work of Miguel Puente-Ajovín and Arturo Ramos has been supported by the Spanish *Ministerio de Ciencia e Innovación* (PID2020-112773GB-I00) and by Aragon Government (ADETRE Reference Group S39_20R).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Akhundjanov, S. B., Devadoss, S., and Luckstead, J. (2017). Size distribution of national CO_2 emissions. *Energy Economics*, 66:182–193.

Alves, L. G. A., Ribeiro, H. V., and Mendes, R. S. (2013). Scaling laws in the dynamics of crime growth rate. *Physica A: Statistical Mechanics and its Applications*, 392:2672–2679.

Anderson, T. W. and Darling, D. A. (1954). A Test of Goodness of Fit. *Journal of the American Statistical Association*, 49(268):765–769.

Balakrishnan, N. (1991). *Handbook of the logistic distribution*. CRC Press.

Bottazzi, G. and Secchi, A. (2006). Explaining the distribution of firm growth rates. *The RAND Journal of Economics*, 37(2):235–256.

Bottazzi, G. and Secchi, A. (2011). A new class of asymmetric exponential power densities with applications to economics and finance. *Industrial and Corporate Change*, 20(4):991–1030.

Burnham, K. P. and Anderson, D. R. (1998). *Model selection and inference*. New York: Springer-Verlag.

Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer-Verlag.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33:261–304.

Băncescu, I., Chivu, L., Preda, V., Puente-Ajovín, M., and Ramos, A. (2019). Comparisons of log-normal mixture and Pareto tails, GB2 or log-normal body of Romania's all cities size distribution. *Physica A: Statistical Mechanics and its Applications*, 526:121017.

Campolieti, M. and Ramos, A. (2021). The distribution of strike size: Empirical evidence from Europe and North America in the 19th and 20th centuries. *Physica A: Statistical Mechanics and its Applications*, 563:125424.

Canning, D., Amaral, L. A. N., Lee, Y., Meyer, M., and Stanley, H. E. (1998). Scaling the volatility of GPD growth rates. *Economics Letters*, 60:335–341.

Cirillo, P. (2013). Are your data really Pareto distributed? *Physica A: Statistical Mechanics and its Applications*, 392:5947–5962.

Cottineau, C., Finance, O., Hatna, E., Arcaute, E., and Batty, M. (2019). Defining urban clusters to detect agglomeration economies. *Environment and Planning B: Urban Analytics and City Science*, 46:1611–1626.

Cottineau, C., Hatna, E., Arcaute, E., and Batty, M. (2017). Diverse cities or the systematic paradox of Urban Scaling Laws. *Computers, Environment and Urban Systems*, 63:80–94.

Cramér, H. (1928). On the composition of elementary errors. *Scandinavian Actuarial Journal*, 1928(1):13–74.

de Wit, G. (2005). Firm size distributions: An overview of steady-state distributions resulting from firm dynamics models. *International Journal of Industrial Organization*, 23:423–450.

Delli Gatti, D., Di Guilmi, C., Gaffeo, E., Giulioni, G., Gallegati, M., and Palestrini, A. (2005). A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic Behavior & Organization*, 56:489–512.

Dupire, B. (1993). Pricing and hedging with smiles. In *Proceedings of AFFI Conference, La Baule, June*.

Dupire, B. (1994). Pricing with a smile. *Risk*, 7(1):271–275.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 94(5):1429–1451.

Fagiolo, G., Napoletano, M., and Roventini, A. (2008). Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. *Journal of Applied Econometrics*, 23:639–669.

Figueiredo, M. A. T. and Leitão, J. M. N. and Jain, A. K. (2000). *Energy Minimization Methods in Computer Vision and Pattern Recognition*, chapter On fitting mixture models. Springer.

Fu, D., Pammolli, F., Buldyrev, S. V., Riccaboni, M., Matia, K., Yamasaki, K., and Stanley, H. E. (2005). The growth of business firms: Theoretical framework and empirical evidence. *Proceedings of the National Academy of Sciences*, 102(52):18801–18806.

Fujiwara, Y., Di Guilmi, C., Aoyama, H., Gallegati, M., and Souma, W. (2004). Do Pareto–Zipf and Gibrat laws hold true? An analysis with European firms. *Physica A: Statistical Mechanics and its Applications*, 335:197–216.

Gabaix, X. (1999). Zipf's law for cities: An explanation. *Quarterly Journal of Economics*, 114:739–767.

Gabaix, X. (2009). Power laws in Economics and finance. *Annu. Rev. Econ.*, 2009:255–293.

Gallegati, M. and Palestrini, A. (2010). The complex behavior of firms' size dynamics. *Journal of Economic Behavior & Organization*, 75:69–76.

Gardiner, C. W. (2004). *Handbook of stochastic methods for Physics, Chemistry and the Natural Sciences (third edition)*. Springer.

Gibrat, R. (1931). *Les inégalités économiques; applications: aux inégalités des richesses, á 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.

Giesen, K. and Suedekum, J. (2012). The French overall city size distribution. *Région et Développement*, 36:107–126.

Giesen, K. and Suedekum, J. (2014). City age and city size. *European Economic Review*, 71:193–208.

Giesen, K., Zimmermann, A., and Suedekum, J. (2010). The size distribution across all cities-double Pareto lognormal strikes. *Journal of Urban Economics*, 68(2):129–137.

Gikhman, I. I. and Skorokhod, A. V. (2007). *The theory of stochastic processes III. Reprint of the 1974 edition.* Springer-Verlag.

González-Val, R., Ramos, A., and Sanz-Gracia, F. (2013). The accuracy of graphs to describe size distributions. *Applied Economics Letters*, 20(17):1580–1585.

Grünwald, P. (2007). *The Minimum Description Length Principle*. The MIT Press.

Grünwald, P. and Roos, T. (2020). Minimum description length revisited. *International Journal of Mathematics for Industry*, 11(1):1930001.

Hannan, E. J. and Quinn, B. G. (1979). The Determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B*, 41:190–195.

Ioannides, Y. M. and Overman, H. G. (2003). Zipf's law for cities: An empirical examination. *Regional Science and Urban Economics*, 33(2):127–137.

Ioannides, Y. M. and Skouras, S. (2013). US city size distribution: Robustly Pareto, but only in the tail. *Journal of Urban Economics*, 73:18–29.

Itô, K. and McKean Jr., H. P. (1996). *Diffusion processes and their sample paths (reprint of the 1974 edition)*. Springer.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions. Volume 2*. John Wiley & Sons.

Kalecki, M. (1945). On the Gibrat distribution. *Econometrica*, 13(2):161–170.

Kleiber, C. and Kotz, S. (2003). *Statistical size distributions in Economics and actuarial sciences*. Wiley-Interscience.

Kolmogorov, A. N. (1933). Sulla Determinazione Empirica di una Legge di Distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4:83–91.

Kotz, S., Kozubowski, T. J., and Podgórski, K. (2001). *The Laplace distribution and generalizations: a revisit with applications to Communications, Economics, Engineering and Finance*. Birkhäuser.

Kwong, H. S. and Nadarajah, S. (2019). A note on "Pareto tails and lognormal body of US cities size distribution". *Physica A: Statistical Mechanics and its Applications*, 513:55–62.

Kyprianou, A. E. (2006). *Introductory lectures on fluctuations of Lévy processes with applications*. Springer-Verlag.

Li, R., Dong, L., Zhang, J., Wang, X., Wang, W.-X., Di, Z., and Stanley, H. (2017). Simple spatial scaling rules behind complex cities. *Nature Communications*, 8:1841.

Luckstead, J. and Devadoss, S. (2017). Pareto tails and lognormal body of US cities size distribution. *Physica A: Statistical Mechanics and its Applications*, 465:573–578.

Luckstead, J., Devadoss, S., and Danforth, D. (2017). The size distributions of all Indian cities. *Physica A: Statistical Mechanics and its Applications*, 474:237–249.

Luévano, J.-R. (2013). Statistical features of the stretched exponential densities. *Journal of Physics: Conference Series*, 475:012008.

Manas, A. (2009). French butchers don't do quantum physics. *Economics Letters*, 103:101–106.

Massing, T., Puente-Ajovín, M., and Ramos, A. (2020). On the parametric description of log-growth rates of cities' sizes of four European countries and the USA. *Physica A: Statistical Mechanics and its Applications*, 551:124587.

Massing, T. and Ramos, A. (2021). Student's t mixture models for stock indices. A comparative study. *Physica A: Statistical Mechanics and its Applications*, 580:126143.

McDonald, J. B. (1984). Generalized functions for the size distribution of income. *Econometrica*, 52(3):647–665.

McDonald, J. B. (1991). Parametric models for partially adaptive estimation with skewed and leptokurtic residuals. *Economics Letters*, 37:273–278.

McDonald, J. B. and Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, 66:133–152.

McLachlan, G. and Peel, D. (2003). *Finite mixture models*. Wiley-Interscience.

Ord, J. K. (1974). *Statistical Distributions in Scientific Work. Volume 2–Model Building and Model Selection*, chapter Statistical models for personal income distributions. D. Reidel Publishing Company.

Otunuga, O. M. (2021). Time-dependent probability density function for general stochastic logistic population model with harvesting effort. *Physica A: Statistical Mechanics and its Applications*, 573:125931.

Peña, G., Puente-Ajovín, M., Ramos, A., and Sanz-Gracia, F. (2021). Log-growth rates of CO_2: An empirical analysis. *Physica A: Statistical Mechanics and its Applications*, page 126550.

Puente-Ajovín, M., Ramos, A., and Sanz-Gracia, F. (2020a). Is there a universal parametric city size distribution? Empirical evidence for 70 countries. *The Annals of Regional Science*, 65:727–741.

Puente-Ajovín, M., Ramos, A., Sanz-Gracia, F., and Arribas-Bel, D. (2020b). How sensitive is city size distribution to the definition of city? The case of Spain. *Economics Letters*, 197:109643.

Quah, D. T. (1997). Empirics for growth and distribution: Stratification, polarization, and convergence clubs. *Journal of Economic Growth*, 2:27–59.

Ramos, A. (2017). Are the log-growth rates of city sizes distributed normally? Empirical evidence for the USA. *Empirical Economics*, 53:1109–1123.

Razali, N. M. and Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2:21–33.

Reed, W. J. (2002). On the rank-size distribution for human settlements. *Journal of Regional Science*, 42:1–17.

Reed, W. J. (2003). The Pareto law of incomes–an explanation and an extension. *Physica A: Statistical Mechanics and its Applications*, 319:469–486.

Reed, W. J. and Jorgensen, M. (2004). The double Pareto-lognormal distribution–a new parametric model for size distributions. *Communications in Statistics-Theory and Methods*, 33(8):1733–1753.

Riccaboni, M., Pammolli, F., Buldyrev, S. V., Ponta, L., and Stanley, H. E. (2008). The size variance relationship of business firm growth rates. *Proceedings of the National Academy of Sciences*, 105(50):19595–19600.

Rozenfeld, H. D., Rybski, D., Andrade, J. S., Batty, M., Stanley, H. E., and Makse, H. A. (2008). Laws of population growth. *Proceedings of the National Academy of Sciences*, 105(48):18702–18707.

Sarkar, S. (2019). Urban scaling and the geographic concentration of inequalities by city size. *Environment and Planning B: Urban Analytics and City Science*, 46:1627–1644.

Schluter, C. and Trede, M. (2013). Gibrat, Zipf, Fisher and Tippet: City size and growth distributions reconsidered. *Working Paper 27/2013 Center for Quantitative Economics WWU Münster*.

Schluter, C. and Trede, M. (2016). Weak convergence to the Student and Laplace distributions. *Journal of Applied Probability*, 53:121–129.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.

Stanley, M. H. R., Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A., and Stanley, H. E. (1996). Scaling behaviour in the growth of companies. *Nature*, 379:804–806.

Subbotin, M. T. (1923). On the law of frequency of errors. *Matematicheskii Sbornik*, 31:296–301.

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

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the Brownian motion. *Physical Review*, 36:823–841.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188.

von Mises, R. E. (1928). *Wahrscheinlichkeit, Statistik und Wahrheit*. Julius Springer.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, 74(3):646–648.

Wu, J.-X. and He, L.-Y. (2017). How do Chinese cities grow? A distribution dynamics approach. *Physica A: Statistical Mechanics and its Applications*, 470:105–118.

Xu, W., Liang, Y., Chen, W., and Wang, F. (2020). Recent advances of stretched Gaussian distributions underlying Haussdorf fracctal distance and its applications in fitting stretched Gaussian noise. *Physica A: Statistical Mechanics and its Applications*, 539:122996.