Model Choice and Size Distribution: A Bayequentist Approach

John-Oliver Engler* and Stefan Baumgärtner

this version: September 5, 2012

Department of Sustainability Sciences & Department of Economics
Leuphana University of Lüneburg, Germany

Abstract

We propose a new three step model selection framework for size distributions in empirical data. It generalizes a recent frequentist plausibility-of-fit analysis (Step 1) and combines it with the Bayesian Akaike Information Criterion (Step 2). We enhance these statistical criteria with the additional criterion of model microfoundation (Step 3) which is to select the model that comes with a dynamic micro model of size dynamics. We perform a numerical performance test of Step 1 which shows that our generalization is able to correctly rule out the distribution hypothesis unjustified by the data at hand. We then illustrate use and usefulness of our approach by investigating the size distribution of 399 commercial cattle farms in Namibia's semi-arid rangelands. We infer that Gibrat's 'law of proportionate effect' holds for our sample, albeit with two modifying assumptions: (1) farm ages are exponentially distributed and (2) the initial farm size distribution was lognormal. We conclude that the framework proposed here has the potential to reconcile the ongoing debate about size distribution models in empirical data, the two most prominent of which are the Pareto and the lognormal distribution. It might therefore provide a basis for the unification of these two major literature strands.

JEL Classification: C12, C52, D30, D31, O44

Keywords: Model choice; model selection; hypothesis testing; size distributions; Gibrat's law; Pareto distribution; rank-size rule.

*Corresponding author, email: engler@leuphana.de

1 Introduction

The identification of theoretical size distribution models in empirical data is on the agenda of economics since Vilfredo Pareto's seminal work on the distribution of wealth in Italy (Pareto 1896). The issue has remained on the research agenda over the years (Gibrat 1931, Champernowne 1953, Fisk 1961, Dagum 1977, Bandourian et al. 2002, Eeckhout 2004) and still is being discussed lively (Lévy 2009 ,Eeckhout 2009). To date, the two most influential strands of literature are on Pareto's law which is somethimes also referred to as rank-size rule – meaning power law behavior in the upper tail of income distributions – and Gibrat's law of proportionate effect (Gibrat 1931) which implies a lognormal size distribution. Yet, there are some methodological problems specific to size distribution fitting: First, as pointed out by Clauset et al. (2009) ordinary least squares (OLS) regressions do not work reliably in the context of distribution fitting, primarily because they do not respect the crucial characteristic of a probability density function that the integral over its support is one. Second, a successful fit alone does not actually confirm or rule out the hypothesis that the data follow a certain distribution. One cannot easily fix this by employing the Kolmogorov-Smirnov test since it is known to produce biased p-values in case of distribution fitting (Clauset et al. 2009, Bubelny 2011). When it comes to actual model comparison and selection the likelihood ratio and χ^2 tests are viable alternatives that compare two models at a time which may lead to complications in case of a larger number of models. Third however, both tests are subject to the large-sample error of the first kind[1](Raftery 1986). Fourth, a p-value of such frequentist hypothesis tests cannot be interpreted as probability of the hypothesis which is unsatisfactory in the case of several unrejected models. In this paper, we address these problems with a suitable statistical model selection framework.

The major contribution of this paper is the formulation of a new three step statistical framework for the detection of theoretical size distribution models in (but not limited to) empirical economic data that offers solutions to the aforementioned methodological problems. Inspired by the works of Clauset et al. (2009) and Efron (2005) who observed a division between 'Bayesians, frequentists and scientists', we offer a combination of frequentist (Step 1) and Bayesian (Step 2) methods unified into one framework. We

[1]systematic rejection of a correct null hypothesis in large samples

add a formalization of the notion of explanatory depth that – given sufficient statistical goodness-of-fit – aims at selecting the size distribution model that can be shown to result from a micro model (Step 3). We will refer to the latter property as *model microfoundation* which we introduce as an integral part of our analysis. Step 1 is a generalization of the algorithm by Clauset et al. (2009) which they have proposed and tested in the context of identifying Pareto's law in empirical data. We take up their algorithmic structure, generalize it to the case of arbitrary size distributions and test its performance with synthetic data drawn from a known population. We propose to combine this plausibility-of-fit test in a second step with the Akaike information criterion, AIC for short (Akaike 1977) from which it is possible to calculate model weights to obtain a model ranking (Burnham and Anderson 2002) and a confidence set which complement the results from the frequentist test. In the last step, we propose to introduce a third criterion into the model selection process that goes beyond purely statistical criteria ('not just the numbers', Burnham and Anderson, 2004). This third criterion asks the 'So what?' question, namely what additional information a good fit has to offer other than being descriptively precise. We shall argue that out of two canidate models that pass the minimum statistical requirements, we prefer the one that comes with a micro model that leads to the observed distribution. Our scheme draws inspiration from the fact that we typically do not only want an answer to the question 'Which model fits best?', but also to 'What can a good fit of a certain model to empirical data tell us about the system it describes?' ('So what?'). Altogether, our framework circumvents the problems of OLS regressions by resorting to the method of maximum likelihood, it provides a general hypothesis test algorithm that gives more accurate p-values than the Kolmogorov-Smirnov goodness-of-fit test , it answers the question of model probability that cannot be answered by comparison of p-values and it contains a criterion that addresses the model's explanatory power.

We illustrate our 'Bayequentist' framework with a sample of 399 commercial cattle farms from Namibia's semi-arid rangelands (Olbrich et al. 2009; Olbrich et al., in press). It may seem as a very peculiar choice but we have chosen this sample for several reasons: (1) the sample of commercial cattle farms in Namibia is a sample of Namibian *firms* and as such, its analysis contributes to the literature on firm size, firm size distribution and firm growth from a very original and fresh perspective. (2) The data set is unique and of excellent detail. (3) It happens to be a very illustrative example for the functioning of our

framework and thus demonstrates how the framework might help to unify the discourse between Gibrat's law and the power law paradigm in the literature. There is also a minor reason for a closer investigation of that particular data set: Namibian commercial cattle farming is a rain fed business which means that the high variability in rain fall is the farm manager's main source of income risk. We regard environmental risks and the farmers' risk preferences as microscopic key determinants of the macroscopically observed size distribution. It will thus be interesting to see whether we can establish this micro-macro connection in the data.

Our paper is organized as follows: In Section 2, we explain in detail our Bayequentist framework for the detection of a peculiar size distribution in empirical data and complement the explanations with numerical performance tests. Section 3 illustrates the proposed framework by concretely discussing the procedure for the Namibian commercial cattle farm data. Section 4 discusses acievements and limitations of our framework before Section 5 draws conclusions.

2 Bayequentist model choice for size distributions

Our wording 'Bayequentist' refers to the combination of both, frequentist and Bayesian statistical methods in order to detect a theoretical size distribution model in empirical data. This section explains our proposed three step Bayequentist statistical model selection framework and its components. After definition of notation in Section 2.1, we detail the three steps separately: in Section 2.2, we elaborate on our algorithm based on Clauset et al. (2009) and test its performance. In Section 2.3, we recapitulate how to obtain model probabilities from a Bayesian scoring method and we discuss the underlying philosophical principles. Section 2.4 introduces the criterion of model microfoundation that aims at the scientific significance of the models and thereby complements the first two steps.

2.1 Notation

To establish notation, for a data sample $\mathbf{x} = \{x_1, \ldots, x_N\}$ and a candidate probability distribution $p(\Theta_1, \ldots, \Theta_m|x)$ with parameters Θ_i where $m = 1 \ldots j$, the associated likelihood

function is obtained as

$$L(\Theta_1,\ldots,\Theta_m|\mathbf{x}) = \prod_{i=1}^{N} p(\Theta_1,\ldots,\Theta_m|x_i) \tag{1}$$

We denote the set of parameters that maximizes this likelihood function given the observed data by $\hat{\Theta}_1,\ldots,\hat{\Theta}_m$. The set that contains the candidate models is referred to as $\mathcal{M}$, its cardinality will be denoted by $|\mathcal{M}|$. After any step in the selection procedure, the set of remaining models will be symblized by an extra prime superscribed so that $\mathcal{M}''$ means the set of candidate models remaining after Step 2 and so on. We assume that, initially, $|\mathcal{M}| > 1$ as there would not be a model *choice* problem in the original sense of the word otherwise. However, as will be discussed throughout this section, a singleton, meaning $|\mathcal{M}'| = 1$, or even an empty set may occur at some point of the process.

2.2 Step 1: plausibility-of-fit

We assess the statistical *plausibility* of the fit generalizing the method proposed in Clauset et al. (2009) for the identification of power law (i.e. Pareto) distributions in empirical data. The intuition behind their method is as follows: any sample randomly drawn from a power law distribution will feature deviations from a true power law distribution. Moreover, the smaller the sample size N, the more drastic the expected deviations. Hence, even if we knew for sure the population followed a power law distribution, we would nonetheless find deviations from a true power law in any finite random sample from this population. Thus, the question is how to distinguish plausible deviations from substantial ones that make the power law hypothesis highly unplausible. From these considerations, Clauset et al. (2009) have proposed the following hypothesis test algorithm: After the power law distribution is ML fit to the data sample of length N which yields $\hat{\Theta}_1,\ldots,\hat{\Theta}_m$, the corresponding Kolmogorov-Smirnov (KS) test statistic is computed. The next step is then to calculate the M KS statistics of ML fits of the power law to M synthetic data samples of length N drawn from a true power law with parameters $\hat{\Theta}_1,\ldots,\hat{\Theta}_m$. The p-value of this test is then simply the fraction of synthetic data samples that have a KS test statistic greater than the real-world sample that is subject to the power law hypothesis H_0. The absolute error estimate of the obtained p-value decreases with the number of generated synthetic

data sets M as given by

$$p_{\text{true}} = p \pm \sqrt{\frac{1}{4M}} \tag{2}$$

Thus, a comparably modest effort of $M = 2500$ synthetic data sets would bring the absolute error estimate down to ± 0.01. Clauset et al. (2009) have shown that the procedure will – on average – be able to correctly rule out the exponential and lognormal distributions as alternative hypotheses for samples larger than 200 data points.

Clauset et al. (2009) have suggested and numerically investigated this test procedure with the power law distribution specifically in mind. They demonstrate that – given large enough samples – their method can be expected to reject the power law hypothesis for samples drawn from the lognormal or exponential distributions and that the method is not prone to large-sample errors of the first kind (c.f. Clauset et al., 2009, Figure 4.1). Yet, the power law (Pareto) distribution is a very special model . While they remark that their test should in principle be suitable for any distribution – as long as there exist methods to create random numbers from that distribution – a generalized version of their plausibility-of-fit test has not been used or investigated so far. In this paper, we take up this point and generalize the algorithm to the following form:

1. ML estimate the parameters $\hat{\Theta}_1, \ldots, \hat{\Theta}_m$ of the hypothesized distribution based on the data sample $\mathbf{x} = \{x_1, \ldots, x_N\}$.

2. Compute the KS test statistic of the obtained fit. Jitter[2] the empirical data if ties are present.

3. Generate M synthetic data samples of length N drawn from the hypothesized distribution with parameters $\hat{\Theta}_1, \ldots, \hat{\Theta}_m$ and calculate their respective KS statistic.

4. Obtain the p-value of H_0 as fraction of the number of synthetic data samples that have a KS statistic greater than the data sample $\mathbf{x}$ and the number of synthetic data sets M.

Our version departs from the version presented in Clauset et al. (2009) in steps 2 and 3. In step 2, we have added the jittering component to avoid conservative[3] p-values that

[2] 'Jittering' is a standard procedure to break ties in empirical data samples. It refers to adding very small random numbers from a uniform distribution with very small support symmetric to the origin to each sample element (e.g. Mease and Buja, 2007).

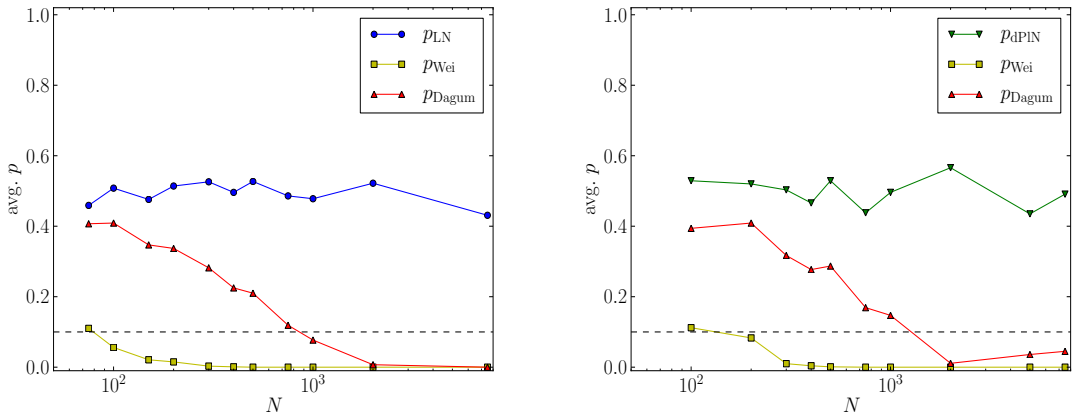
[3] Here, 'conservative' means that the p-values returned by the test are too optimistic, hence implying erroneous results.

may result from tied data. In step 3, we have removed the explicit reference to the Pareto distribution in favor of the general expression 'hypothesized distribution'. Hence, this test can in principle be applied to any statistical size distribution model since it is always possible to construct random numbers based on a known probability density frunction[4].

For performance assessment of this generalized algorithm, we first identify the relevant distributions that such a test should be capable of dealing with. Kleiber and Kotz (2003) classify the size distribution models commonly found in the economics literature into three functional superforms: the generalized beta distribution of the second kind (GBII), the generalized Gamma distribution (GG) and the lognormal group. GBII contains the Dagum (Dagum, 1977) and the Fisk (Fisk, 1961) distribution, GG contains the Weibull (Bartels and van Metelen 1975) and the Gamma distribution (Ammon 1895) and the lognormal supergroup contains the lognormal, the Pareto (Pareto, 1896) and the double Pareto lognormal distribution (dPlN, Reed and Jorgensen, 2004). We then run two numerical tests: First, we draw numerical random samples from a true lognormal distribution with randomly changing parameters at each draw. We then calculate the p-values of each sample for the following hypothesized distributions: lognormal, dPlN, Weibull and Dagum and take the average of all p-values separately for each synthetic lognormal test sample. Second, we repeat this procedure for synthetic samples from true dPlN distributions, again with randomly selected parameters at each iteration. We plot the results of the testing procedure in Figure 1 where we omit nested distributions[5]. The results demostrate that the method is in principle suitable to identify a true lognormal (Figure 1, panel a) or true dPlN distribution (Figure 1, panel a) in finite data samples as the average p-value remains above the theshold of $p = 0.1$ recommended by Clauset et al. (2009) only for the distribution that the samples were actually drawn from. On the other hand, Figure 1 also reflects the fact that it gets generally much harder to rule out a distribution the more parameters it has. The two parameter Weibull distribution is easily rejected in both cases since the average p-value remains slightly above 0.1 for very small samples only ($N = 75$ for synthetic lognormal data, $N = 100$ for synthetic dPlN data, c.f. Figure 1). In contrast, the three parameter Dagum distribution generally provides a much

[4]Two major methods exist for this: the acceptance-rejection method and the inversion method, the details can be found in von Neumann (1951) and Devroye (1986), respectively.

[5]We cannot expect – on average – to be able to rule out nested hypotheses. For example, because the lognormal is contained in the dPlN, the average p-value of both hypotheses would behave quite similar in Figure 1. And, by construction, the dPlN could not be ruled out for the synthetic lognormal samples.



(a) synthetic lognormal samples

(b) synthetic dPlN samples

Figure 1: Average p-values for synthetic data samples drawn from a lognormal (a) and a double-Pareto lognormal (b). The dashed line indicates the threshold $p = 0.1$ below which the hypothesized distribution is ruled out.

better fit to the data and is in consequence much harder to reject ($N \geq 1000$, Figure 1) although the sampling population does not follow a Dagum distribution. Positively put, the true distribution is not ruled out, even for larger samples ($N \geq 1000$). Compared to the original version of the test specifically for power law testing, our numerical results suggest that the generalized version may need greater data samples – depending on the competing hypotheses (c.f. Clauset et al., 2009) – to safely rule out incorrect alternatives.

2.3 Step 2: weights of evidence

Suppose that Step 1 has left a non-empty candidate set with more than one plausible size distribution model. To infer a model ranking based on the sample data, we propose to make use of the 'weights of evidence' (Burnham and Anderson, 2004) which are based on the Akaike information criterion (AIC Akaike, 1977). To recall, the AIC relies on the likelihood function $L(\hat{\Theta}_1, \ldots, \hat{\Theta}_m)$ at the likelihood optimum and the number of parameters j of the model:

$$\mathrm{AIC} = -2 \ln L(\hat{\Theta}_1, \ldots, \hat{\Theta}_m) + 2j \tag{3}$$

Since low AIC scores are considered better[6] the criterion penalizes additional model parameters with the model dimension j entering Equation (3) as positive summand. The

[6]Low scores reflect low information loss from data to model in the sense of Kullback-Leibler information (Kullback and Leibler, 1951)

intuition behind this is that, out of two candidate models with the same likelihood based on the data, the more parsimonous one is selected, a principle known as *Occam's razor*[7].

Based on these scores, it is possible to infer a ranking of the different candidate models according their 'Akaike weights' (c.f. Burnham and Anderson, 2004) which make a statement for each candidate model about the probability that the model is true given the data. Burnham and Anderson (2004) suggest the following scheme: find the model with the lowest AIC score and rename that score $\mathrm{AIC}_{\mathrm{min}}$. For every other candidate model i, calculate

$$\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\mathrm{min}} \tag{4}$$

and classify the relative support of model i according to the scale laid out in Table 1:

Table 1: Classification of the evidence for a model based on the data: models within $\Delta_i \leq 3$ constitute the confidence set of models.

$\Delta_i \leq 3$	substantial support
$4 \leq \Delta_i \leq 7$	moderate support
$\Delta_i \geq 10$	essentially no support

By construction, the model with the smallest AIC score[8] has $\Delta = 0$ and the strongest support based on the data. From this, one obtains the Akaike weights as

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum_r \exp(-\Delta_r/2)} \tag{5}$$

These weights are sometimes also referred to as *weights of evidence* as they reflect the probability that model i is correct given the data. It is thus possible to assign probabilities to multiple competing hypotheses simultaneously.

Philosphically, there is a lively debate on whether one should select the more parsimonous or the more complicated out of two competing models. Methodologically, the two extremes are the Bayes Information Criterion (BIC, Schwarz, 1978) which penalizes

[7] It is not possible to clearly ascribe this term to the mind of one person: In 1852, English philosopher Sir William Hamilton coined the term after 14th century English logician William of Ockham although Ockham never actually wrote the well-known phrase 'shave away all but what is necessary' . Instead, the roots of this principle can be traced back to the works of Ptolemy ($90 - 168$) and Aristotle ($384 - 322$ BC).

[8] i.e. with the least information loss from data to model in the sense of Kullback-Leibler information (Kullback and Leibler, 1951)

additional model parameters more drastically than the AIC[9] and the likelihood criterion (as advocated by Edwards, 1972) which favors the model with the highest likelihood, hence generally the one with most parameters. While there is no final consent yet, the literature on model selection seems to favor the 'principle of parsimony' over the 'principle of diversity' (Leibniz 1968 [1710]). We do not share this view in that we think that a more complicated model that lies within the confidence set of candidate models is not automatically worse than a simpler model within the confidence set *if* it has deeper explanatory power. In our 'Bayequentist' framework which we illustrate schematically in Figure 2, we therefore advocate the use of the AIC in Step 2 because it provides a formal compromise between selection based solely on the likelihood and the BIC. In other words, the AIC formally strikes a balance between the principle of parsimony and the principle of diversity.

2.4 Step 3: model microfoundation

We maintain that a size distribution model should achieve both, a good fit to empirical data (in the sense of Sections 2.2 and 2.3) and it should be based on a plausible micro model that can be shown to generate the overall distribution. In fact, this criterion is a relaxation of a proposition already made in the context of income distributions that a good income distribution model be based on a plausible *stochastic* model (Dagum 1983, Reed and Wu 2008). This particular understanding of model microfoundation means that there exists a stochastic differential equation [10]

$$\mathrm{d}X_t = f(X_t,t)\mathrm{d}t + g(X_t,t)\mathrm{d}W_t \tag{6}$$

that describes the size evolution $\mathrm{d}X_t$ of the individual parts of the economic system in question[11] between two neighboring time instants t and $t - \mathrm{d}t$. $\mathrm{d}W_t$ is the increment of a standard Wiener process with $\langle W_t \rangle = 0$ and $\langle W_t^2 \rangle = t$ for $t > 0$. Thus, the entirety of these individual parts constitute the overall distribution at any time instant t. The actually observed size distribution in a random sample from a population can be explained

[9] $\mathrm{BIC} = -2 \ln L(\hat{\Theta}_1, \ldots, \hat{\Theta}_m) + j \ln N$

[10] We refer to stochastic differential equations in the Itō sense of stochastic integration in this paper. Yet, the argument does not depend at all on the choice between Itō or Stratonovich stochastic integration.

[11] These may for example be individual wages of laborers in the United States or city sizes in Germany

by the microscopic stochastic growth process undergone by the individual constituents of the population.

What can be found in literature is focussed very much on the stochastic differential equation understanding of model microfoundation. However, this is only one out of many possibilities for a model microfoundation. In a broader sense, we do not see a good reason to exclude agent-based, rule-based or other models per se. The point here is that the identification of a particular size distribution in empirical data allows for inference about the underlying individual growth dynamics *if and only if* the distribution comes with a suitable microfoundation. Our argument is that – given, say, two models that pass Step 1 – it is well justified to change the relative model ranking from Step 2 *if* only one of these models comes with a microfoundation and *if* the researcher aims at explaining rather than just describing. Hence, Step 2 may be overridden in favor of the model with microfoundation as long as Step 1 is passed.

2.5 The framework at one glance

In Figure 2, we illustrate our Bayequentist framework for the identification of size distributions in empirical data. It contains several dimensions of 'goodness-of-fit'. First and foremost, in an absolute sense, a good fit should describe the data plausibly (Step 1) which is assessed by the frequentist method laid out in Section 2.2. Second, it should lie within the confidence set of models, that is, given the data it should perform sufficiently well relative to the competing models (Step 2), something that cannot well be assessed by direct comparison of p-values. For this second step, we propose to use the Bayesian approach recapitulated in Section 2.3. Third, a good fit of a size distribution to empirical data should go beyond the descriptive finding alone by allowing inferences about the underlying process that generated the observed distribution (Step 3). In short, Steps 1 and 2 test for statistical significance and descriptive power, Step 3 explores explanatory power which we see as a proxy for scientific significance (c.f. McCloskey 1995).

There are also situations where alterations of our scheme might become necessary which may be caused by the researcher's motivation or by procedural reasons. For example, it may not be of interest by what mechanism the size distribution was actually generated on the micro level at all. Instead, only the best possible theoretical description of the

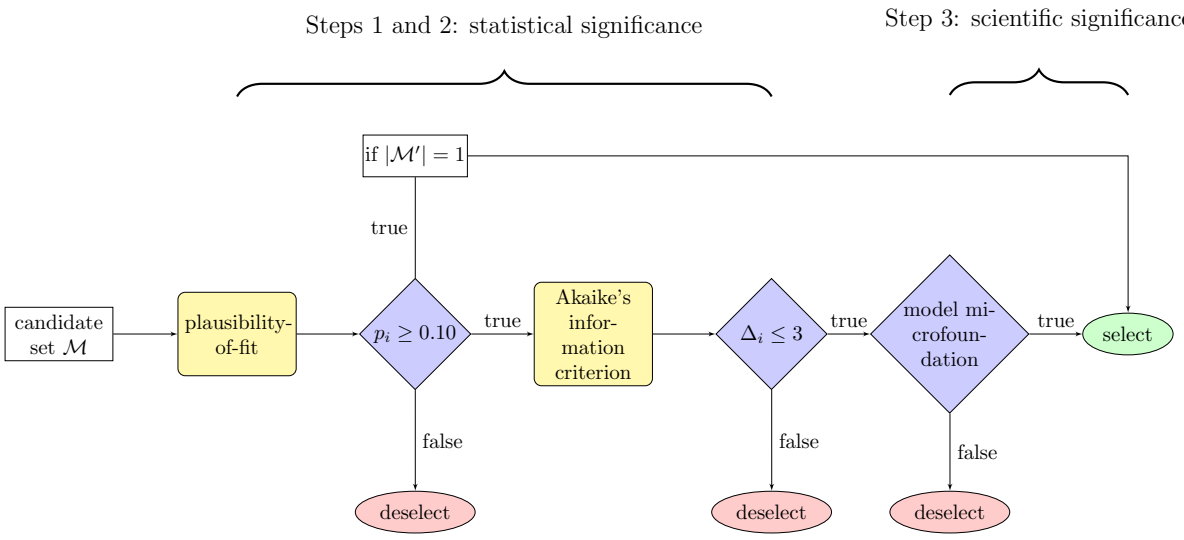


Figure 2: Our proposed Bayequentist model selection framework for size distribution fitting to empirical data: the frequentist Step 1 checks for *absolute* goodness-of-fit and the Bayesian Step 2 checks for *relative* goodness-of-fit, both being measures for statistical significance. Step 3 takes the analysis beyond a merely statistical level and checks for *explanatory power*, which we see as a proxy for scientific significance.

data might be relevant. In this case where the interest in the data is of purely descriptive origin, Steps 1 and 2 are sufficient. However, most of the time when fiting theory to data, it seems that the underlying question is not only 'Which model fits best?' but also 'What does a good fit tell us most likely about the population that the sample was drawn from?'. In this case, Step 3 needs to be included into the framework. Ultimately, the research question sets the frame. As to procedural reasons two things might happen that are – for the sake of clearness – not included in Figure 2 which are both related to Step 3: (1) no model may be left in the end meaning $|\mathcal{M}'''| = 0$, or (2) more than one model passes Step 3, i.e. $|\mathcal{M}'''| \geq 2$. As to the case $|\mathcal{M}'''| = 0$, one can either resort to the result of Steps 1 and 2 combined having the best possible fit in statistical terms then or one needs to reconsider the choice of the candidate models in $\mathcal{M}$ in general and possibly start over again. On the other hand, if $|\mathcal{M}'''| \geq 2$, we recommend to re-examine the model performances in Steps 1 and 2 and select according to the weights of evidence ranking from Step 2.

3 Illustration: commercial cattle farms in Namibia

This section illustrates the Bayequentist framework that we have proposed in the previous section with size data of Namibian commercial cattle farms that have been surveyed in 2008/09 (Olbrich et al. 2009; Olbrich et al., in press). We describe the data in the following subsection before we present the results of our model selection framework in Sections 3.2 through 3.4.

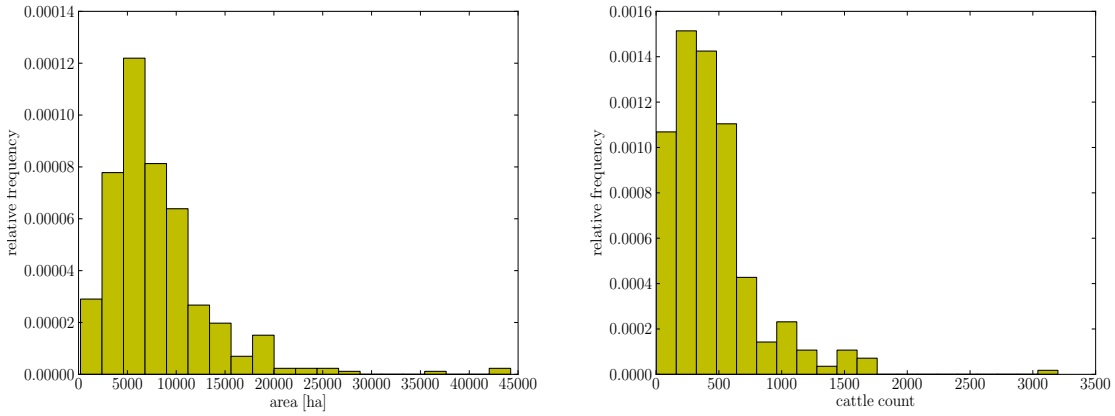
3.1 Data and candidate models

The data sample of 399 commercial Namibian cattle farms uses two specifications of 'size'. Thus, whenever we say 'farm size', we mean either areal farm size in hectares or the number of cattle held on the farm. The latter has been hypothesized as being a 'proxy for wealth' (Olbrich et al., 2009). Namibia's semi-arid climate with dry and wet season causes the cattle numbers to vary on each farm in the course of one year, for example due to pasture management. Therefore, each record contains entries about the cattle count in November and in April. We mean the average of these two values when we refer to the farm size as measured by cattle head count so that we also correct for seasonal effects.

Deviations from the total sample size of $N = 399$ occur in our analysis are due to incomplete data records. For example, the fits of different size distribution models to the cattle data in this section is based on $N = 351$ data points whereas fits to the area data have been based on $N = 391$ data points. We used the R programming language (version 2.13.0) for statistical data analysis and visualization as well as Python(x,y) (version 2.6.5) for the graphs in Figures 1 and 3. Table 2 shows descriptive statistics of the data set while Figure 3 plots histograms of the empirical herd size and area distributions.

Table 2: Descriptive statistics of the sample of Namibian commercial cattle farms.

descriptive statistic	cattle [head count]	area [ha]
sample size	351	391
minimum value	1	200
maximum value	3200	42244
mean	453	7969
median	374	6800
standard deviation	361	5504
skewness	2.37	2.50
kurtosis	10.5	11.0
Gini coefficient	0.391	0.336

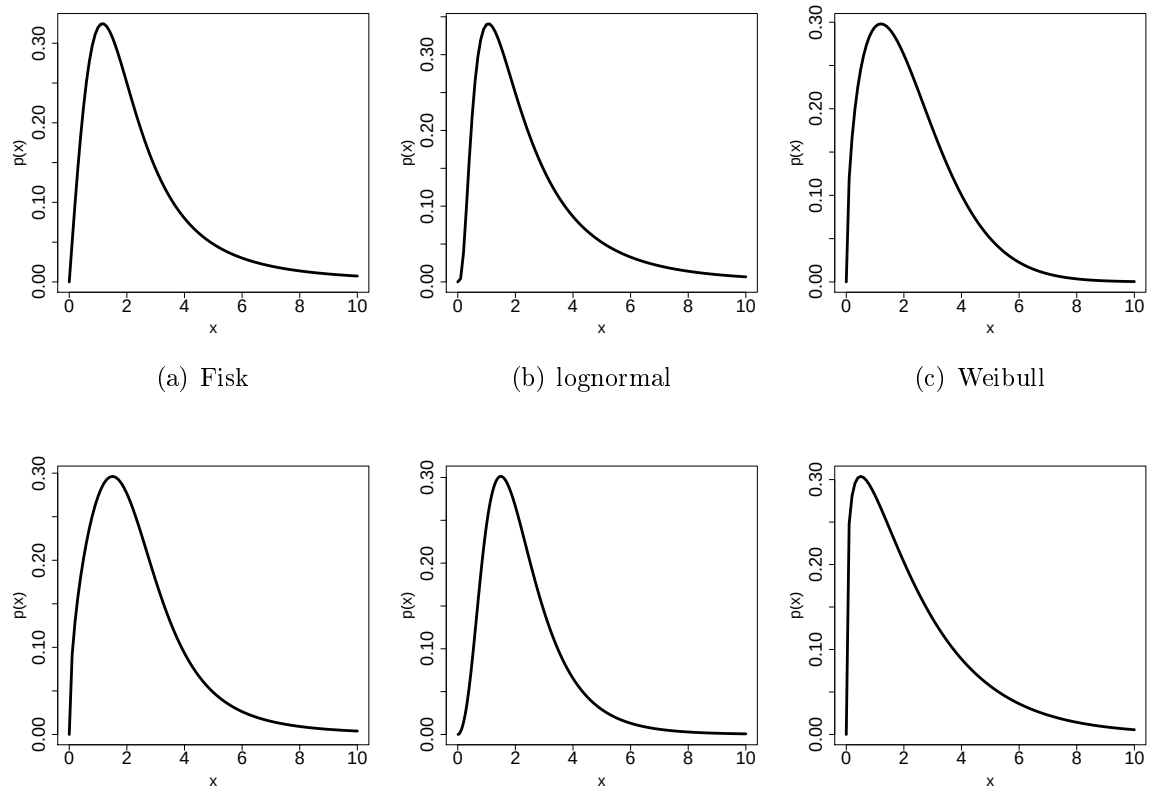


(a) area (b) cattle

Figure 3: Distribution of commercial cattle farm sizes in Namibia as measured in hectares (panel a) and cattle count (panel b)

Candidate models

From the literature on size distributions in economics which is focussed especially on the distribution of wealth, income and city sizes, we identify six size distribution models that have been identified in empirical data repeatedly. These models are the Fisk distribution, the lognormal, the double Pareto lognormal, the Weibull, the Dagum and the Gamma. In Figure 4, we give a plot of these models while their explicit functional forms and interrelations are detailed in Appendix A.



(a) Fisk (b) lognormal (c) Weibull

(d) Dagum (e) double Pareto lognormal (f) gamma

Figure 4: The six most common size distributions in economic literature.

3.2 Step 1: plausibility-of-fit

Table 3 shows the results of Step 1 of our framework. We base our estimations for the $p-$values on $M = 2500$ synthetic data sets each so that the resulting $p-$values are accurate up to approximately 1%. For the cattle data, we find that two models pass the plausibility-of-fit test, the Dagum ($p = 0.39$) and the dPlN ($p = 0.23$) while none of the candidate models pass the plausibility-of-fit test for the areal data. Regarding parameter estimation, we find similar parameters for the lognormal, the Pareto and the dPlN distributions which reflects the nestedness of the Pareto and lognormal within the dPlN. Indeed, we find a Pareto coefficient of 3.22 that is matched by a tail index of 3.52 for the cattle data. For the areal data, we find a Pareto coefficient of 3.25 and a tail index of 3.28. Conversely, the parameters describing the dPlN's lognormal body – ν and τ – match the estimates that we find for the lognormal fits (cattle: $\nu = 6.22$ vs. $\mu = 5.80$ and $\tau = 0.42$ vs. $\sigma = 0.89$, area: $\nu = 8.96$ vs. $\mu = 8.78$ and $\tau = 0.35$ vs. $\sigma = 0.68$).

Table 3: Results of step 1: Maximum likelihood estimates of the model parameters along with p-values. p-values significant at the 10% level are marked with * and accurate up to 1%. Definitions of models can be found in Appendix A.

model	cattle parameter estimates	p	area parameter estimates	p
lognormal	$\mu = 5.80$ (0.05), $\sigma = 0.89$ (0.03)	0.00	$\mu = 8.78$ (0.04), $\sigma = 0.68$ (0.02)	0.00
dPlN	$\alpha = 3.52$ (3.76), $\beta = 1.41$ (0.17), $\nu = 6.22$ (0.11), $\tau = 0.42$ (0.10)	0.23*	$\alpha = 3.28$ (2.23), $\beta = 2.05$ (0.82), $\nu = 8.96$ (0.08), $\tau = 0.35$ (0.08)	0.00
Pareto	$\alpha = 3.22$ (0.39), $x_{\min} = 460$ (49)	0.33*	$\alpha = 3.25$ (0.39), $x_{\min} = 7000$ (746)	0.02
Weibull	$k = 1.36$ (0.05), $\lambda = 494.85$ (20.63)	0.00	$k = 1.59$ (0.06), $\lambda = 8965.36$ (301.18)	0.00
Dagum	$a = 3.16$ (1.10), $b = 559.65$ (1.10), $p = 0.45$ (1.18)	0.39*	$a = 3.32$ (1.09), $b = 8366.58$ (1.08), $p = 0.63$ (1.18)	0.00
Fisk	$\gamma = 351.45$ (1.04), $\delta = 2.14$ (1.04)	0.04	$\gamma = 6665.40$ (1.04), $\delta = 2.74$ (1.03)	0.02
Gamma	$\Theta = 254.1$ (20.28), $\kappa = 1.77$ (0.12)	0.00	$\Theta = 3000.00$ (215.43), $\kappa = 2.64$ (0.17)	0.00

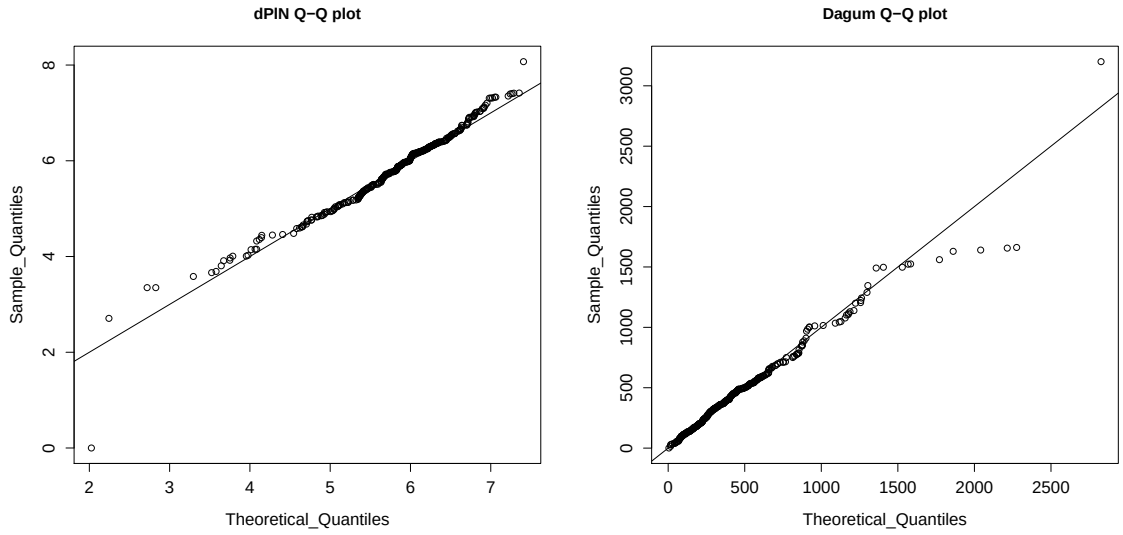
3.3 Step 2: weights of evidence

In Table 4, we list the results of the second step of our Bayequentist model selection framework (Figure 2). Loglikelihoods $\mathcal{L}_i$ are almost identical with slight advantages for the Dagum model for the cattle data ($\mathcal{L}_{\text{dPlN}} = -2462.85$ vs. $\mathcal{L}_{\text{Dagum}} = -2462.70$) and the dPlN in case of the area data ($\mathcal{L}_{\text{dPlN}} = -3817.57$ vs. $\mathcal{L}_{\text{Dagum}} = -3818.04$). The AIC scores reflect this finding as these two models rank first (Dagum, AIC = 4931.41 for cattle and 7642.08 for areal data and second (dPlN, AIC = 4933.71 for cattle and 7643.15 for areal data). Using the classification from Table 1 together with the definition of the model weights w_i (Eqn. 5), we find substantial evidence for both of the two models (cattle: $w_{\text{dPlN}} = 23.30\%$ vs. $w_{\text{Dagum}} = 73.66\%$, area: $w_{\text{dPlN}} = 36.00\%$ vs. $w_{\text{Dagum}} = 61.40\%$). With a cumulative weight of about 3% for both data sets, the other models play no significant role.

Table 4: Results of step 2: Loglikelihoods, AIC scores, model weights and model evidence.

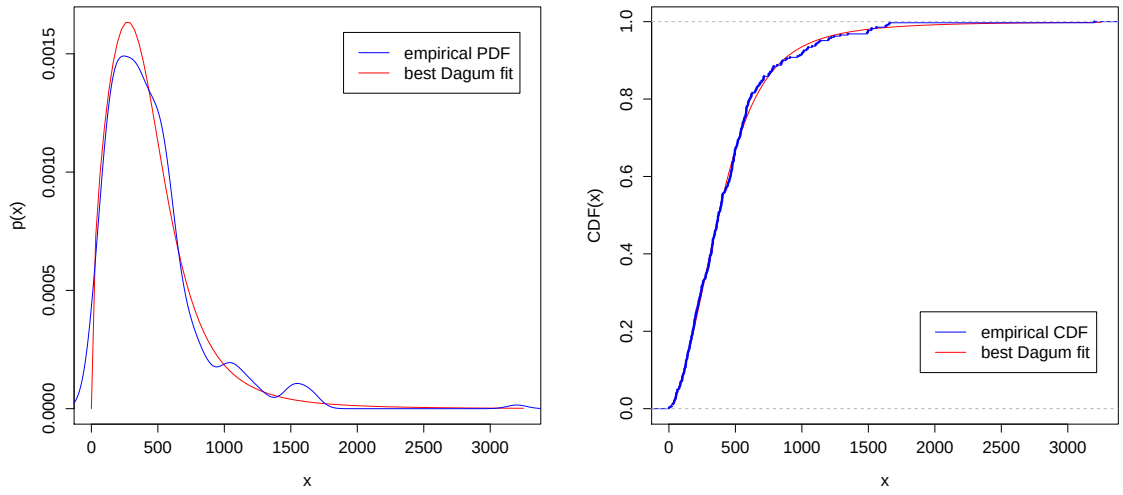
	cattle				area			
model	loglikelihood	AIC	weight [%]	evidence	loglikelihood	AIC	weight [%]	evidence
lognormal	-2492.21	4988.42	<0.00	none	-3836.59	7677.18	<0.00	none
dPlN	-2462.85	4933.71	23.30	substantial	-3817.57	7643.15	36.00	substantial
Weibull	-2470.72	4945.43	0.07	none	-3841.32	7686.63	<0.00	none
Dagum	-2462.70	4931.41	73.66	substantial	-3818.04	7642.08	61.40	substantial
Fisk	-2473.82	4951.64	<0.00	none	-3822.21	7648.41	2.59	moderate
Gamma	-2466.91	4937.83	2.97	moderate	-3827.67	7659.35	0.01	none

The quality of fit of the dPlN as well as the Dagum model to our data is illustrated in Figures 5, 6 and 7 where we use for computational convenience that the logarithm of a dPlN distributed variable is Normally Laplace distributed (c.f. Lemma 1, Appendix B).



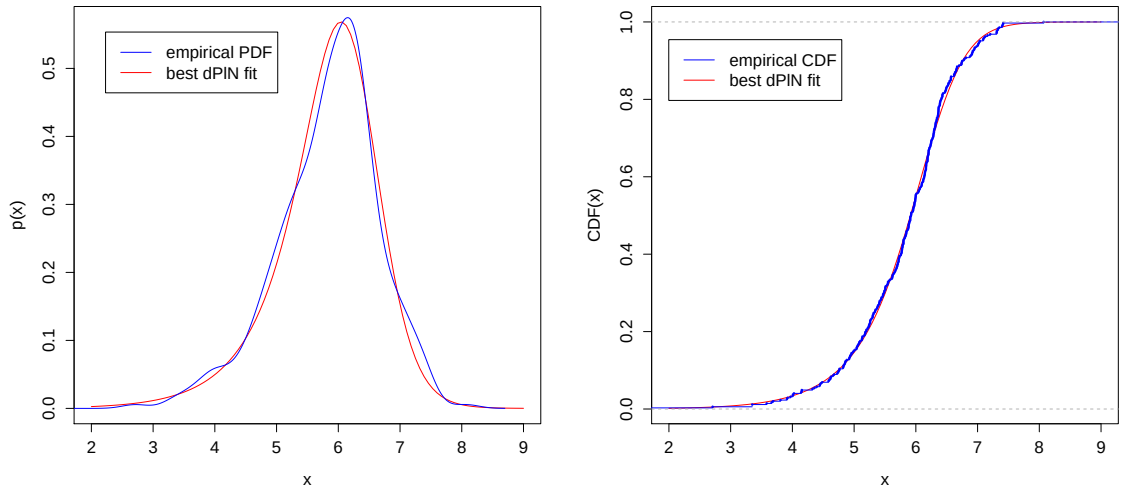
(a) dPlN (b) Dagum

Figure 5: Q-Q plots of dPlN (a.) and Dagum (b.) distributions



(a) probability densities (b) cumulative densities

Figure 6: Best fitting Dagum distribution versus kernel density regression of the mean cattle data.



(a) probability densities

(b) cumulative densities

Figure 7: Best fitting Normal Laplacian distribution versus kernel density regression of the mean cattle data.

3.4 Step 3: model microfoundation

Camillo Dagum defined his distribution by an ordinary differential equation which covers many of the empirically observed characteristics of income and wealth distributions. Insofar, it is aimed at describing these distributions well but does not come with a stochastic model to explain how the observed distribution might have been generated from an underlying stochastic growth process. Therefore, it lacks stochastic microfoundation and we deselect it.

On the other hand, there is a stochastic microfoundation for the double Pareto lognormal (dPlN) distribution. As Reed and Jorgensen (2004) have shown, the dPlN results from a combination of three assumptions about the enitities under consideration: (1) the underlying growth process is a geometric Brownian motion, (2) the initial size distribution is lognormal and (3) the entities under consideration are not equally old and have an overally exponential age distribution.

Therefore, in spite of worse relative goodness-of-fit statistics, we select the dPlN as the best fitting distribution out of the seven candidate models.

Interpretation and inference

The finding that the dPlN model provides the best fit to the herd size distribution of Namibian commercial cattle farms contains several messages. First, because of the argumentation from Step 3 of our framework, we may infer that Gibrat's law holds, albeit with two modifying crucial assumptions: (1) exponentially distributed farm age structure and (2) lognormal initial farm size distribution. On the individual farm level, the validity of Gibrat's law implies that farm size $S(t)$ at time instant $t > 0$ evolves according to

$$S(t) = S(0) \cdot \exp\left[\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W(t)\right] \tag{7}$$

where $S(t) > 0 \quad \forall \quad t > 0$ and where $W(t)$ is the standard Wiener process with $\mathbb{E}[W(t)] = 0$ and $\mathrm{Var}[W(t)] = t$. The parameter μ reflects the deterministic growth rate of the farm, in other words the growth rate that the farmer implicitly sets through his management decisions such as buying or selling cattle, acquisition of new machinery to augment productivity, hiring or decapitating staff and so forth. Consequently, the expected farm size at $t > 0$ is $\mathbb{E}[S(t)] = S(0)\mathrm{e}^{\mu t}$. The absolute value of σ thus determines the influence of randomness on the overall growth process. For the semi-arid rangelands in Namibia, a huge part of this randomness comes from variation in annual rainfall (c.f. Olbrich et al., 2009). Second, the stochastic exponential growth process (Equation 7) and the underlying stochastic differential equation imply two things: (1) larger farms do not grow faster or slower than smaller farms and (2) the growth (rate) of a farm in one period is independent of the growth (rate) in the preceding period. While the first conclusion is the core of Gibrat's law, the second one is a direct consequence of the fact that the stochastic process (Equation 7) is a Markov chain. Thus, our results suggest that an economic policy concerned with optimally fast job generation in Namibia's large agricultural sector should not focus on certain subgroups of farmers as our results suggest that small farms do not grow significantly faster than already large ones and vice versa.

4 Discussion

In this section, we will summarize what we have established in this paper before we critically reflect on limitations of and possible objections to our framework. Particularly, we will discuss our method's robustness, the justification of Step 3 and other possible boundaries of our approach.

Achievements. We close an existing methodological gap in the literature concerning the identification of theoretical size distribution models in empirical data. The three step statistical framework that we have proposed here contains two original contributions which are: (1) the generalization and test of an algorithm that was previously introduced only in the context of testing empirical data for the Pareto distribution (c.f. Step 1, Section 2.2) and (2) the combination of frequentist, Bayesian statistical methods with a qualitative criterion of model microfoundation in an integrated three step framework. As such, our framework may have the potential to reconcile the debate of validity between Pareto's and Gibrat's law (Eeckhout 2009, Lévy 2009) in the case of U.S. city and settlement sizes. Our illustrative analysis of the size distribution of commercial cattle farms in Namibia at least suggests that the dPlN distribution which unifies both laws is a promising candidate to also unify the debate, as already remarked by Giesen and Suedekum (2012). Yet, we formally expand the analysis by Giesen and Suedekum (2012) in that we admit more candidate models and replace the biased p-value obtained from the Kolmogorov-Smirnov test with the better one from the algorithm described in Section 2.2.

Robustness. In our paper, we understand 'robustness' in the context of errors of first and second kind and will discuss these for Step 1 as the effectivity of Steps 2 and 3 critically depends on the success of Step 1. As Kass and Raftery (1995) remark, frequentist hypothesis testing often suffers from the occurrence of large-sample errors of the first kind which means that any correct hypothesis will be rejected at some point if one only chooses a large enough sample. In our numerical simulations that we have carried out for performance assessment (Figure 1, Section 2.2), we do not find any support for a greater error of the first kind with larger samples as the test's average p-values remain at a level of about 0.5 for all sample sizes considered ($N = 75 \ldots 10000$). This is in agreement with Clauset et al. (2009) who find similar results for their power law specific test which we

build our generalized version upon. As to the test's error of the second kind, sample size and number of parameters of the hypothesized distributions determine the discriminative power of the test. In a nutshell, we find the following relations: the greater the sample the better the discriminiative power and the the more parameters in a model the harder it is to rule out, as expected (Figure 2.2). **Justification of Step 3.** Step 3 could obviously be challenged for being a too qualitative feature in an otherwise quantitative framework. It is true that it does not yield a nice quantitative criterion as Steps 1 and 2 do. Yet, we would rather see this as an asset than as a flaw of our framework. First and foremost, it incorporates the notion of 'scientific significance' (c.f. McCloskey 1995, Johnson 1999) as an additional feature into our framework and therefore complements the statistical concepts from Steps 1 and 2 naturally aiming at statistical significance. The So what? question is however not touched by asking about statistical significance alone. Step 3 of our framework addresses this problem in a qualitative way. Second, it prevents Occm's razor from shaving away not only 'all but is necessary' but possibly from shaving away more than that. We have already argued in Section 2.5 that we prefer the AIC to the BIC for providing a better quantitative compromise between the principle of parsimony and the principle of diversity. The same argument seems even more striking for Step 3. Consider the relationship between theory of special relativity and classical mechanics in physics. The former is far more complicated than the latter while they have the same subject matter, the movement of bodies under the influence of external forces. Describing and understanding these movements, classical mechanics will do for the most part. However, it would be wrong to reject special relatvity per se for being overly complicated because special relativity has more explanatory power than classical mechanics[12]. Hence, while Occam's razor may in general be a justified principle of science, it may only be applied to situations where several theories or models are on a par in terms of explanatory power. Third, as Steps 1 and 2 are purely statistical criteria, it inevitably suffers from the limitations common to this approach (we have discussed them in the paragraph 'Robustness'). As qualitative criterion, Step 3 circumvents these limitations and tackles the problem from a completely different angle, effectively lowering the danger of systematic misjudgments. In summary, Step 3 serves at least three purposes: (1) it integrates the notion of scientific significance into the selection process, (2) it prevents

[12]Special relativity explains the movement of bodies for velocities close to the speed of light.

Occam's razor from becoming too sharp and (3) it does not face the same limitations as the quantitative methods from Steps 1 and 2 and therefore can serve as corrective for these shortcomings.

5 Conclusion

In this paper, we have proposed and illustrated a new statistical framework for detecting size distribution models in empirical data. The special feature of our framework is the three step combination of frequentist with Bayesian statistical methods. Step 1 is based on a frequentist plausibility-of-fit testing algorithm originally proposed for the detection of Pareto's law in empirical data (Clauset et al. 2009). Here, we have generalized this algorithm and demonstrated it to work for the case of size distributions other than Pareto's law. We have proposed to combine this with the Bayesian relative goodness-of-fit criterion AIC (Akaike Information Criterion) in Step 2 and a criterion of explanatory power (Step 3). Analyzing a unique and highly resoluted data set of 399 Namibian commercial cattle farms (Olbrich et al. 2009), we have demonstrated both, use and fruitfulness of this new framework. We have found that the Dagum and dPlN distributions fit the data best in terms of statistical plausibility and relative goodness-of-fit (Steps 1 and 2), yet in Step 3 we have selected the dPlN model because of its superiority in explanatory depth. In turn, we have inferred that commercial cattle farms in Namibia follow a stochastic exponential growth process which implies that Gibrat's *law of proportionate* effect holds and that environmental risk is a key force in farm size growth much rather than just a supporting element.

Our analysis yields to promising fields for future research: First and foremost, we suggest to revisit the debate in the literature concerning the (non-)validity of Gibrat's law and the rank-size rule (i.e. the occurrence of a Pareto/power law in the upper tail) in empirical data of city and settlement sizes, distribution of wealth and incomes and the firm size distribution in an economy. Our three step Bayequentist framework should enable researchers to make confident statements regarding size distributions in these data through its robust model selection procedure. Should the dPlN prove to fit to these data as well – and there is first evidence (Giesen et al. 2010, Giesen and Suedekum 2012)

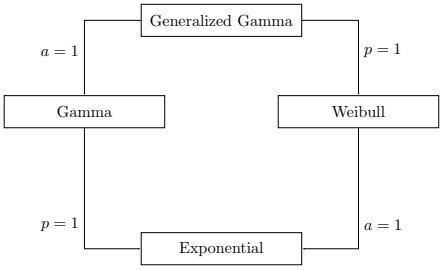
– then the two major strands in literature that either find Gibrat's Law to hold or the Pareto law of incomes, could be reconciled. Moreover, it needs to be further investigated to what extent the growth dynamics discovered here can be generalized to other firms as this could have far reaching implications for policy making in case of the debate on the role of firm sizes for growth in industrial organization.

Appendix A: mathematical representation and interrelation of the most common size distributions

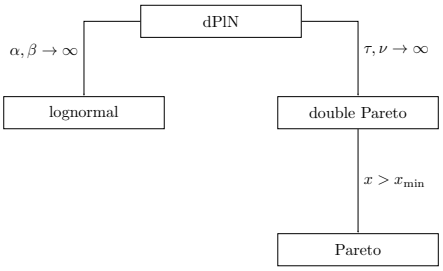
This appendix provides the the mathematical representation of the size distribution models used in this paper. Figure 8 presents the three functional supergroups that we briefly mentioned in Section 2.2 and how the members of these are related. The classification presented here is due to Kleiber and Kotz (2003). Table 5 provides the mathematical definition of these size distribution models.

Table 5: The size distribution models most commonly used in economics. Φ denotes the cumulative probability function of the standard normal distribution and $\Phi^c(x)$ its complementary function $1-\Phi(x)$. For ease of reading, each parameter has been given a distinct name.

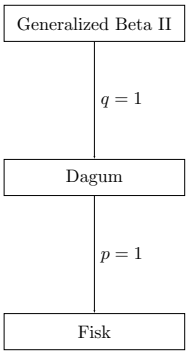
model	no. of parameters	probability density function	first used in
lognormal	2	$\frac{1}{\sigma x\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$	Gibrat (1931)
dPlN	4	$\frac{\alpha\beta}{\alpha+\beta}x^{-\alpha-1}\mathrm{e}^{(\alpha\nu+\frac{\alpha^2\tau^2}{2})}\Phi\left(\frac{\ln(x)-\nu-\alpha\tau^2}{\tau}\right)+$ $\frac{\alpha\beta}{\alpha+\beta}x^{\beta-1}\mathrm{e}^{(-\beta\nu+\frac{\beta^2\tau^2}{2})}\Phi^c\left(\frac{\ln(x)-\nu+\beta\tau^2}{\tau}\right)$	Reed and Jorgensen (2004)
Pareto	2	$\frac{1}{\alpha-1}(\frac{x}{x_{\min}})^{-\alpha}$	Pareto (1896)
Weibull	2	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}\exp(-(\frac{x}{\lambda})^k)$	Bartels and van Metelen (1975)
Dagum	3	$\frac{ap}{x}\left(\frac{(x/b)^{ap}}{((\frac{x}{b})^a)+1)^{p+1}}\right)$	Dagum (1977)
Fisk	2	$\frac{\delta}{\gamma}(\frac{x}{\gamma})^{\delta-1}/(1+(\frac{x}{\delta})^\beta)^2$	Fisk (1961)
Gamma	2	$x^{\kappa-1}\frac{\exp(-\frac{x}{\Theta})}{\Theta^\kappa\Gamma(\kappa)}$	Ammon (1895)



(a) GG



(b) dPlN



(c) GBII

Figure 8: The generalized gamma (GG) distribution and its sibblings (panel a) and the double Pareto lognormal family (panel b) and the family originating from the generalized beta distribution of the second order (GBII, panel c).

Appendix B: relationship of dPlN and Normal Laplacian

Throughout our analysis, we have used the relationship of dPlN and Normal Laplace distribution which is formulated in Lemma 1. This appendix provides the associated technicalities.

Lemma 1. *The logarithm of a double Pareto lognormally distributed variable is Normal Laplace distributed, i.e. if* $X \sim \mathrm{dPlN}(\alpha,\beta,\nu,\tau)$ *then* $Y := \ln X \sim \mathrm{NL}(\alpha,\beta,\nu,\tau)$.

Proof. We start from the CDF of the Normal Laplacian and show that $P_{\mathrm{NL}}(\alpha,\beta,\nu,\tau) = P_{\mathrm{dPlN}}(\alpha,\beta,\nu,\tau)$ under a change of variable $y = \ln x$. As Reed and Jorgensen (2004) have

shown, the CDF of the Normal Laplacian reads (Eqn. (15) in Reed and Jorgensen (2004))

$$P_{\mathrm{NL}}(y) = \Phi\left(\frac{y-\nu}{\tau}\right) - \phi\left(\frac{y-\nu}{\tau}\right)\frac{\beta R(\alpha\tau - (y-\nu)/\tau) + \alpha R(\beta\tau + (y-\nu)/\tau)}{\alpha+\beta}$$

where $R(z) = (1-\Phi(z))/\phi(z)$ with $\Phi(z) = \int_{-\infty}^{z}\phi(t)\mathrm{d}t$ and $\phi(z) = \frac{1}{\sqrt{2\pi}}\exp(-z^2/2)$. A change of variable $y = \ln x$ yields

$$\tilde{P}(y=\ln x) = \Phi\left(\frac{\ln x-\nu}{\tau}\right) - \frac{1}{\alpha+\beta}\left[\beta\frac{1-\Phi\left(\frac{\alpha\tau^2-\ln x+\nu}{\tau}\right)}{\phi\left(\alpha\tau-\frac{\ln x-\nu}{\tau}\right)}\phi\left(\frac{\ln x-\nu}{\tau}\right) + \alpha\frac{1-\Phi\left(\frac{\beta\tau^2+\ln x-\nu}{\tau}\right)}{\phi\left(\beta\tau+\frac{\ln x-\nu}{\tau}\right)}\phi\left(\frac{\ln x-\nu}{\tau}\right)\right]$$

which can be simplified to

$$\tilde{P}(x) = \Phi\left(\frac{\ln x-\nu}{\tau}\right) - \frac{1}{\alpha+\beta}\left[\beta\left(1-\Phi\left(\frac{\alpha\tau^2-\ln x+\nu}{\tau}\right)\right)\exp\left(\frac{(\alpha\tau)^2}{2}-\alpha(\ln x-\nu)\right) + \alpha\Phi^c\left(\frac{\beta\tau^2+\ln x-\nu}{\tau}\right)\exp\left(\frac{(\beta\tau)^2}{2}+\beta(\ln x-\nu)\right)\right]$$

Introducing the function $A(\theta,\nu,\tau) = \exp(\theta\nu + (\alpha\tau)^2/2)$, we find that

$$\tilde{P}(x) = \Phi\left(\frac{\ln x-\nu}{\tau}\right) - \frac{1}{\alpha+\beta}\left[\beta x^{-\alpha}A(\alpha,\nu,\tau)\Phi\left(\frac{\alpha\tau^2-\ln x+\nu}{\tau}\right) + \alpha x^{\beta}A(-\beta,\nu,\tau)\Phi^c\left(\frac{\ln x-\nu+\beta\tau^2}{\tau}\right)\right] = P_{\mathrm{dPlN}}(x) \quad (8)$$

because of $\Phi(-z) = 1-\Phi(z)$. The last Equation (8) is the CDF of the double Pareto lognormal (eq. (23), Reed and Jorgensen (2004)). □

References

Akaike, H. (1977): Chap. On Entropy Maximization Principle. In *Applications of Statistics*. Amsterdam: North Holland, 27–41

Ammon, O. (1895): *Die Gesellschaftsordnung und ihre Natürlichen Grundlagen*. Jena

Bandourian, R., J.B. McDonald and R.S. Turley (2002): *A Comparison of Parametric Models of Income Distribution Across Countries and Over Time*. http://ssrn.com/abstract=324900, Luxembourg Income Study, working paper No. 305

Bartels, C.P.A. and H. van Metelen (1975): *Alternative Probability Density Functions of Income*. Vrije University Amsterdam: Research Memorandum 29

Bubelny, P. (2011): *Some notes on biasedness and unbiasedness of two-sample Kolmogorov-Smirnov test*. http://arxiv.org/pdf/1106.5598

Burnham, K.P. and D.R. Anderson (2002): *Model Selection and Multimodel Inference: A Practical Information Theoretical Approach*. 2nd edition. Springer, ISBN 0–387–95364–7

Burnham, K.P. and D.R. Anderson (2004): *Multimodel Inference - Understanding AIC and BIC in Model Selection*. Sociological Methods & Research 33, 261–304

Champernowne, D.G. (1953): *A Model of Income Distribution*. Economic Journal 63 (250), 318–351

Clauset, A., C.R. Shalizi and M.E.J. Newman (2009): *Power-Law Distributions in Empirical Data*. SIAM Review 51 (4), 661–703

Dagum, C. (1977): *A New Model of Personal Income Distribution: Specification and Estimation*. Economie Appliquée 30, 413–437

Dagum, C. (1983): entry on *Income Distribution Models*. Encyclopedia of Statistical Sciences 4

Devroye, L. (1986): *Non-Uniform Random Variate Generation*. Springer, New York

Edwards, A.W.F. (1972): *Likelihood*. Cambridge University Press, London

Eeckhout, J. (2004): *Gibrat's Law for (All) Cities*. American Economic Review 94 (5), 1429–1452

Eeckhout, J. (2009): *Gibrat's Law for (All) Cities: Reply*. American Economic Review 99, 1676–1683

Efron, B. (2005): *Bayesians, Frequentists, and Scientists*. Journal of the American Statistical Association 100, 1–5

Fisk, P. (1961): *The Graduation of Income Distributions*. Econometrica 29, 171–185

Gibrat, R. (1931): *Les inégalités économiques*. Librairie du Recueil Sirey, Paris

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

Giesen, K. and J. Suedekum (2012): *The Size Distribution Across All "Cities": A Unifying Approach*. http://ssrn.com/abstract=2010947, IZA Discussion Paper No. 6352

Johnson, D.H. (1999): *The Insignificance of Statistical Significance Testing*. Journal of Wildlife Management 63, 763–772

Kass, R.E. and A.E. Raftery (1995): *Bayes Factors*. Journal of the American Statistical Association 90 (430), 773–795

Kleiber, C. and S. Kotz (2003): *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley & Sons, New Jersey

Kullback, S. and R.A. Leibler (1951): *On Information and Sufficiency*. Annals of Mathematical Statistics 22, 79–86

Leibniz, G.W. (1968): *Die Theodizee*. Felix Meiner Verlag, Hamburg

Lévy, M. (2009): *Gibrat's Law for (All) Cities: Comment*. Amercican Economic Review 99, 1672–1675

McCloskey, D. (1995): *The Insignificance of Statistical Significance*. Scientific American, 32–33

Mease, D., A.J. Wyner and A. Buja (2007): *Cost-Weighted Boosting with Jittering and Over/Under-Sampling: JOUS-Boost*. Journal of Machine Learning Research 8, 409–39

Neumann, J. von (1951): *Various techniques used in connection with random digits. Monte Carlo methods.* Nat. Bureau Standards, volume = 12, 36–38

Olbrich, R., M.F. Quaas and S. Baumgärtner:

Olbrich, R., M.F. Quaas and S. Baumgärtner (2009): *Sustainable use of ecosystem services under multiple risks - a survey of commercial cattle farmers in semi-arid rangelands in Namibia*. University of Lüneburg Working Paper Series in Economics No. 137

Pareto, V. (1896): *Cours d'Economie Politique*. Geneva, Droz

Raftery, A. (1986): *Choosing Models for Cross-Classifications*. American Sociological Review 51, 145–146

Reed, W.J. and M. Jorgensen (2004): *The double Pareto-lognormal distribution - A new parametric model for size distribution*. Communications in Statistics 34, 1733–1753

Reed, W.J. and F. Wu (2008): Chap. New Four- and Five-Paramter Models for Income Distributions. In*Modeling Income Distributions and Lorenz Curves*. Volume 5,III, Springer, 211–223

Schwarz, G. (1978): *Estimating the Dimension of a Model*. The Annals of Statistics 6, 461–464