Model choice and size distribution: a Bayequentist approach

Engler, John-Oliver; Baumgärtner, Stefan

[Link to publication](#)

Model choice and size distribution:

a Bayequentist approach

by

John-Oliver Engler and Stefan Baumgärtner

University of Lüneburg
Working Paper Series in Economics

No. 265

February 2013

Model choice and size distribution:
a Bayequentist approach

JOHN-OLIVER ENGLER* and STEFAN BAUMGÄRTNER

February 27, 2013

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 a relative ranking based on the Bayesian Akaike
Information Criterion (Step 2). We enhance these statistical criteria with the
additional criterion of microfoundation (Step 3) which is to select the size
distribution that comes with a dynamic micro model of size dynamics. A
numerical performance test of Step 1 shows that our generalization is able
to correctly rule out the distribution hypotheses unjustified by the data at
hand. We then illustrate our approach, and demonstrate its usefulness, with a
sample of commercial cattle farms in Namibia. In conclusion, 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.

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

Keywords: model choice, model selection, hypothesis testing, size distributions, Gibrat's
Law, Pareto distribution, rank-size rule, environmental risk, semi-arid rangelands, cattle
farming

---

*Corresponding author, Sustainability Economics Group, Leuphana University of Lüneburg, P.O. Box
2440, D-21314 Lüneburg, Germany, email: engler@leuphana.de

1   Introduction

The identification of theoretical size distribution models in empirical data has been a topic of considerable debate in economics since Vilfredo Pareto's (1896) seminal work on the distribution of wealth in Italy. The correct identification of theoretical size distribution models has remained a contentious issue that has continued to resurface in the economic literature ever since (Gibrat 1931, Champernowne 1953, Fisk 1961, Dagum 1977, Bandourian et al. 2002, Eeckhout 2004) and is still generating lively discussions (Lévy 2009, Eeckhout 2009). The relevance of this field comes from the fact that 'models on population dynamics and economic growth have implications on the distribution of city sizes' (Lévy 2009: 1672). Similarly, any model on income dynamics implies a certain income distribution. It has therefore been a longstanding research interest to identify empirical evidence for these theoretical models. Entities that are usually described by their size distribution over a population include: individual income or wealth in a society, population numbers of cities in a certain region (country, continent, world) or firm sizes[1] in an economy. To date, the two most influential theoretical concepts related to size distribution models are Pareto's Law (Pareto 1896) which is sometimes also referred to as rank-size rule – meaning power law behavior in the upper tail of the distribution – and Gibrat's Law of Proportionate Effect (Gibrat 1931) which implies a lognormal size distribution.

There are a number of methodological problems specific to size-distribution fitting and related model choice problems. First, as pointed out by Clauset et al. (2009) ordinary least-squares (OLS) regressions do not work reliably in the context of fitting theoretical size distribution models to empirical data, primarily because OLS regressions do not account for the crucial characteristic of a probability density function that the integral over its support is one. Second, while high values of R^2 do explain what fraction of the variance in the data is explained by the model, they cannot confirm or rule out the *hypothesis* that the data actually follow a certain distribution.[2] One cannot easily fix this by employing the Kolmogorov-Smirnov test since it has been shown to produce biased p-values in case of distribution fitting (Clauset et al. 2009, Bubelny 2011). Third, a p-value of any frequentist hypothesis test cannot be interpreted as probability that the

---

[1]This entity can be measured by business volume or staff numbers, for example.
[2]Nor do low values of sum of squared errors (SSE) or sum of absolute errors (SAE).

hypothesis actually holds true which is impractical in the case of several unrejected models from which the researcher would like to choose the 'best' one. For actual model comparison and selection, the likelihood ratio and χ^2 tests are frequentist methods that have been proposed to compare the relative performance of two models at a time (cf. Neyman and Pearson 1933, Greenwood and Nikulin 1996). However, as pointed out by Raftery (1986), the likelihood ratio and χ^2 tests are subject to the large-sample error of the first kind[3], let alone that comparing only two models at a time seems rather impractical.

In this paper, we address these three problems. Our major contribution is the formulation of a three-step statistical model-selection framework for size distributions in empirical data. Inspired by the remark by Efron (2005) who observed a division between 'Bayesians, frequentists and scientists', we offer a combination of frequentist (Step 1) and Bayesian (Step 2) statistical methods unified into one framework, together with a formalization of the notion of microfoundation (Step 3). By microfoundation, we refer to the existence of a micro model that leads to the observed overall size distribution. We will hence refer to our framework as Bayequentist.

Step 1 is a generalization of the plausibility-of-fit 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 in the second step with Akaike's Information Criterion (Akaike 1973), AIC for short, from which it is possible to calculate model weights to obtain a model ranking (Burnham and Anderson 2004) to complement the results from Step 1. In the third step, we propose to introduce an additional 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 argue that out of two candidate models that pass the minimum statistical requirements, one should prefer the one that comes with a microfoundation. 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 the question 'What can a good fit of a certain model to empirical data tell us about the system it describes?'. Altogether, our framework circumvents

---

[3]This refers to a systematic rejection of even a good model in large enough samples.

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 test, it gives a relative model ranking that cannot be inferred by comparison of p-values, and it contains a criterion that addresses the model's explanatory power.

We illustrate our Bayequentist framework, and demonstrate its usefulness, with a sample of 399 commercial cattle farms from Namibia's semi-arid rangelands (Olbrich et al. 2009; Olbrich et al. 2012). This sample is very well-suited for this purpose 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 an 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 all three steps of our framework and thus demonstrates how the framework might help to unify the discourse between Gibrat's Law and Pareto's Law that dominate the literature. (4) As for illustration of Step 3, 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. Hence, environmental risks and the farmers' risk preferences are key micro-determinants of the size distribution observed at the macro-level. There is thus a microscopic theory allowing us to test for a macroscopic model in the data.

Our paper is organized as follows. In Section 2, we contextualize our contribution in relation to the literature. In Section 3, 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 4 illustrates the proposed framework by completely executing the procedure for the Namibian commercial cattle farm data. Section 5 discusses achievements and limitations of our framework. Section 6 concludes.

2   Relation to the literature

The literature on size distributions has developed in the fields of economics and statistics. The economics literature can roughly be divided into two groups. In the first

group, contributions deal with one particular size distribution model, either presenting evidence for the validity of that model in one or several data sets, or presenting a theoretical derivation of that very size distribution from a set of economically motivated assumptions, or a mixture of both. Apart from the founding works of Pareto (1896) and Gibrat (1931), more recent examples in this group include Mansfield (1962), Hart and Oulton (1996), Gabaix (1999), Pinkovskyi and Sala-i-Martin (2009), Chotikapanich et al. (2012), González-Val et al. (2012), Toda (2012) and Tsou (2012). An extensive review of the literature in this strand concerned with Gibrat's Law is given by Sutton (1997). The second group comprises contributions that compare and discuss the goodness-of-fit of at least two – but often more – theoretical size distribution models in empirical economic data. Recent examples in this strand are Evans (1987), Bandourian et al. (2002), Eeckhout (2004) and Giesen et al. (2010) or the very recent debate between Lévy (2009) and Eeckhout (2009) about the goodness-of-fit of the lognormal and the Pareto distributions of the U.S. city size distribution. The body of statistics literature on the subject deals with the problem of model choice from both a technical and a philosophical perspective. Most relevant for this paper are the works of Kass and Raftery (1995), Burnham and Anderson (2004), Reed and Jorgensen (2004), Efron (2005) and Clauset et al. (2009).

In this paper, we aim at unifying these major strands of literature. We maintain that one possible source of the described heterogeneity and – to some extent – disagreement in the literature, originates from the methodological issues that we have summarized in the introduction. For example, the disagreement between Lévy and Eeckhout (c.f. Lévy 2009, Eeckhout 2009 following the original paper Eeckhout 2004) boils down to a lack of consensus of how to tell whether some fit is actually supporting the hypothesis sufficiently well when dealing with empirical size distributions. Here, we suggest a statistical framework to answer this crucial question.

3   Bayequentist model choice for size distributions

This section explains our proposed three-step Bayequentist statistical model-selection framework and its components. After definition of notation and introduction of the relevant size distribution models from the literature in Section 3.1, we detail the three steps

one after the other: in Section 3.2, we elaborate on our algorithm based on Clauset et al. (2009) and test its performance (Step 1). In Section 3.3, we explain how to obtain model probabilities from a Bayesian scoring method (Step 2), and we discuss the underlying philosophical principles. Section 3.4 introduces the criterion of model microfoundation that aims at the scientific significance of the models (Step 3) and thereby complements the first two steps.

3.1 Notation and candidate size distributions

To establish notation, for a data sample $\mathbf{x} = \{x_1, \ldots, x_N\}$ where x_i denotes the size of entity i and N is the total number of entities in the sample, and a candidate size distribution model $p(\Theta_1, \ldots, \Theta_M | x)$ with parameters Θ_k where $k = 1 \ldots M$, the associated likelihood function is obtained by

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

We denote the parameter values that maximize this likelihood function given the observed data $\mathbf{x}$ by $\hat{\Theta}_1, \ldots, \hat{\Theta}_M$. The logarithm of the likelihood function (Equation 1) is called loglikelihood for short and is denoted by $\mathcal{L}(\Theta_1, \ldots, \Theta_M | \mathbf{x})$. The set that contains the candidate size distribution models is referred to as $\mathcal{M}$ and its cardinality is denoted by $|\mathcal{M}|$. After any step in the selection procedure, the set of remaining models is denoted by an extra prime so that $\mathcal{M}''$ means the set of candidate size distribution models remaining after Step 2 and so on. We assume that, initially, $|\mathcal{M}| > 1$ as there would not be a model *choice* problem otherwise. However, as will be discussed throughout this section, a singleton, e.g. $|\mathcal{M}'| = 1$, or even an empty set may occur at some later point in the process.

From the literature, we identified six different size distribution models that have repeatedly been proposed to describe empirical data. Kleiber and Kotz (2003) classify these 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). In Figure 1, we give exemplary plots of these models while their explicit functional forms and interrelations are detailed in Appendix A (Table 4 and Figure 8).
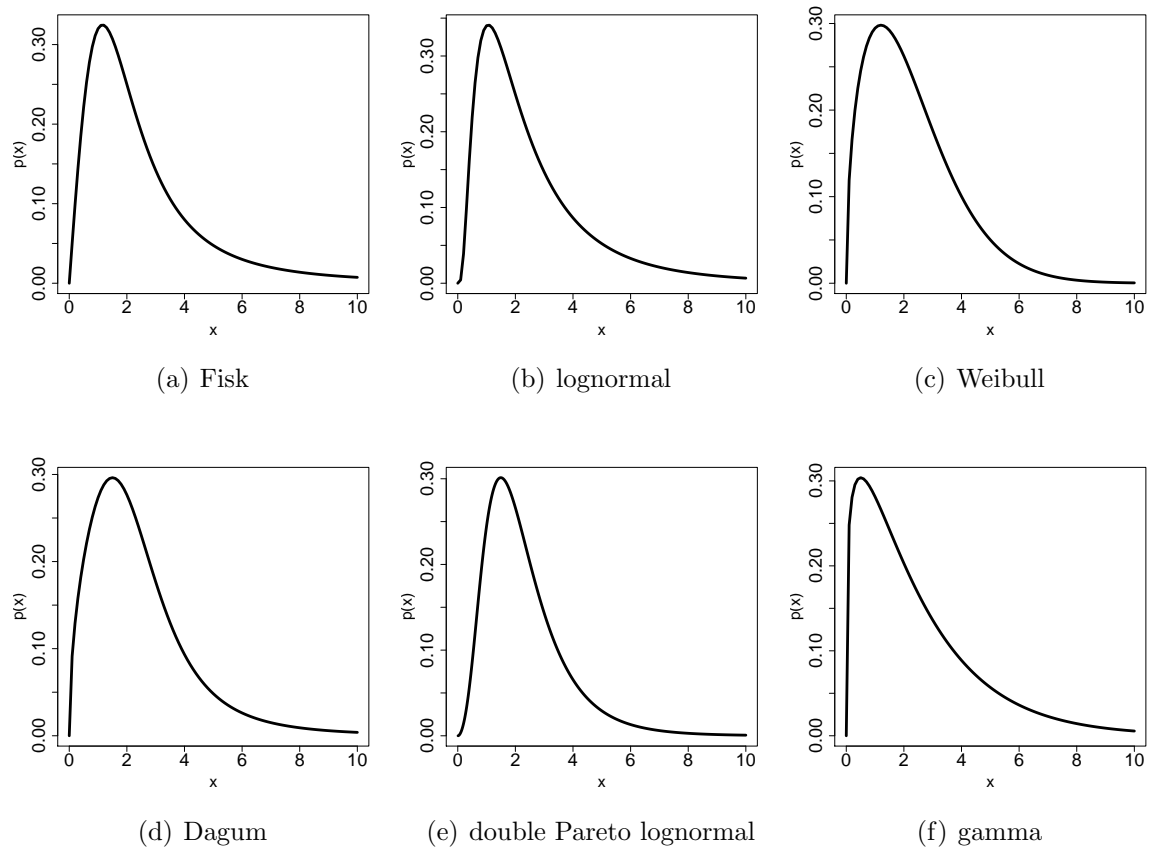

(a) Fisk   (b) lognormal   (c) Weibull

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

Figure 1: The six most common size distributions in economic literature.

3.2 Step 1: Plausibility-of-fit

We assess the statistical *plausibility* of the fit generalizing the method proposed by 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 larger 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 those ones that make the power-law hypothesis highly unplausible. Clauset et al. (2009) have proposed the following hypothesis test algorithm: After the power-law distribution is maximum-likelihood (ML) fit to the original 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 Z KS statistics of ML fits of the power law to Z 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 original 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 synthetic data sets Z as given by

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

Thus, a comparably modest effort of $Z = 2500$ 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 (cf. 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 as follows:

1. ML estimate the parameters $\hat{\Theta}_1, \ldots, \hat{\Theta}_M$ of the hypothesized distribution (H_0) based on the data sample $\mathbf{x} = \{x_1, \ldots, x_N\}$.

2. Compute the KS test statistic of the obtained fit. Jitter[4] the empirical data if ties are present.

3. Generate Z 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 Z.

5. Reject H_0 if $p < 0.1$, do not reject else.

Our algorithm departs from the one presented by Clauset et al. (2009) in stages 2 and 3. In stage 2, we have added the jittering component to avoid conservative[5] p-values that may result from tied data. In stage 3, we have removed the explicit reference to the Pareto distribution in favor of the general expression 'hypothesized distribution'. The threshold p-value of 0.1 is the value recommended by Clauset et al. (2009). Hence, this modified 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 function.[6]

For performance assessment of this generalized algorithm, we 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

---

[4]'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).

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

[6]Two major methods exist for this: the acceptance-rejection method and the inversion method (von Neumann (1951) and Devroye (1986), respectively).

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 2 where we omit nested distributions.[7] The results demonstrate that
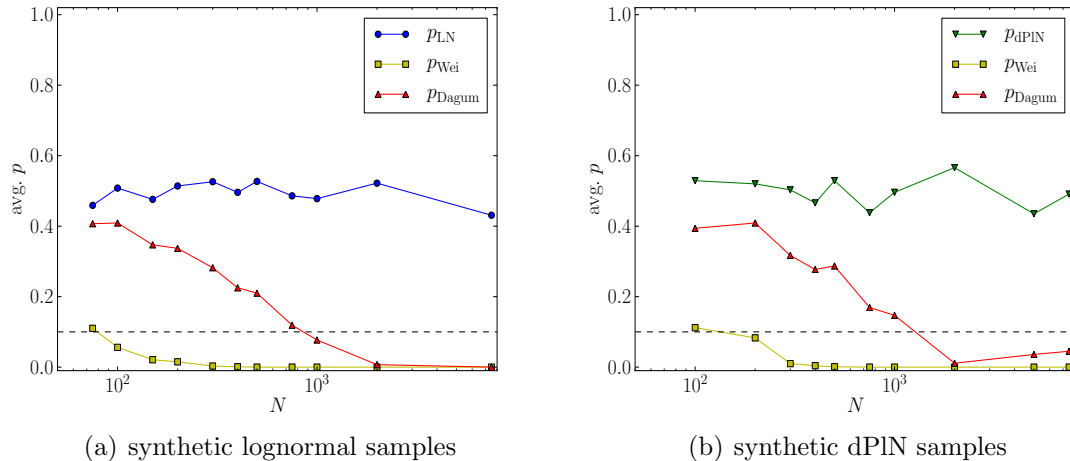

(a) synthetic lognormal samples   (b) synthetic dPlN samples

Figure 2: Average p-values of various size distribution hypotheses for synthetic data samples drawn from (a) a lognormal and (b) a double-Pareto lognormal distribution. The dashed line indicates the threshold $p = 0.1$ below which the hypothesized distribution is ruled out.

the method is in principle suitable to identify a true lognormal (Figure 2a) or true dPlN distribution (Figure 2b) in finite data samples as the average p-value remains above the threshold of $p = 0.1$ only for the distribution that the samples were actually drawn from. On the other hand, Figure 2 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, Figure 2a and b, respectively). In contrast, the three parameter Dagum distribution generally provides a much better fit to the data and is in consequence much harder to reject ($N \geq 1000$, Figure 2) 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, our numerical results suggest that the generalized version may need greater data samples – depending on the competing hypotheses (cf. Clauset et al. 2009) – to reliably rule out incorrect alternatives.

---

[7]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 2. And, by construction, the dPlN could not be ruled out for the synthetic lognormal samples.

In summary, in Step 1, each element of the candidate size distribution set $\mathcal{M}$ is checked for statistical plausibility-of-fit with the procedure presented here. Size distribution models with a p-value greater than 0.1 pass the test and become elements of $\mathcal{M}'$. Yet, it is *not* possible to infer a model ranking from the p-values obtained from Step 1 since all these p-values can tell us is how plausible H_0 is compared to mere chance, and not compared to another model. This is only possible with a Bayesian approach, which we now turn to.

3.3 Step 2: Model ranking

Suppose that Step 1 has left a candidate set $\mathcal{M}'$ with $|\mathcal{M}'| > 1$. 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 1973). To recall, the AIC relies on the likelihood function $L(\hat{\Theta}_1, \ldots, \hat{\Theta}_M)$ at the likelihood optimum and the number M of parameters of the respective candidate size distribution model:

$$\text{AIC} = -2\ln L(\hat{\Theta}_1, \ldots, \hat{\Theta}_M) + 2M. \tag{3}$$

Since low AIC scores are better in terms of less Kullback-Leibler information loss from data to model (Kullback and Leibler 1951), the criterion penalizes additional model parameters by having their number M entering Equation (3) as a positive summand. The intuition behind this is that, out of two candidate models with the same likelihood based on the data, the more parsimonious one is selected, a principle known as *Occam's razor*[8].

Based on the AIC scores, it is possible to infer a ranking of the different candidate models according to their 'Akaike weights' (cf. Burnham and Anderson 2004) which make a statement for each candidate model about the relative strength of evidence ('weight of evidence', Burnham and Anderson 2004) in favor of one specific model given the data and given the other elements of $\mathcal{M}'$. In Bayesian statistics, these weights reflect the posterior model probability, i.e. the model probability given some prior and given the actual data.

---

[8]It is not possible to clearly ascribe this term to the mind of one person: In 1852, English philosopher Sir William Hamilton (cf. Kaye 2007) 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' (Vogel Carey 2010). In fact, the roots of this principle can be traced back to the works of Ptolemy $(90 - 168)$ and Aristotle $(384 - 322$ BC$)$ (Baker 2011).

Burnham and Anderson (2004) suggest the following scheme to find these weights: take the model with the lowest AIC score and rename that score AIC_{min}. For every other candidate model j where $j = 1 \dots |\mathcal{M}'|$, calculate the number

$$\Delta_j = \text{AIC}_j - \text{AIC}_{\text{min}}. \tag{4}$$

The so-defined Δ_j are a measure for relative strength of evidence.[9] By construction, the model with the smallest AIC score has $\Delta = 0$ and, hence, the strongest support based on the data and given the other elements of $\mathcal{M}'$. From this, one obtains the Akaike weights as

$$w_j = \frac{\exp(-\Delta_j/2)}{\sum_l \exp(-\Delta_l/2)} \tag{5}$$

where $l = 1 \dots |\mathcal{M}'|$. These weights give a relative ranking of size distribution models based on the data and the other size distribution models contained in the set $\mathcal{M}'$. These model weights w_j have two essential interpretations, one from information theory (1) and one from Bayesian statistics (2): (1) the higher the numerical weight value w_j of some model j, the smaller its information loss in the Kullback-Leibler sense; (2) as demonstrated by Burnham and Anderson (2004), the Akaike weight w_j of model j is equal to the Bayesian posterior model probability $\Pr(j|\mathbf{x})$ conditional on the data $\mathbf{x}$. Either way, the Akaike weights imply a relative ranking, and we recommend to use this ranking to complement the analysis from Step 1.

Philosophically, there is a lively debate on whether one should select the more parsimonious or the more complicated out of two competing models. Methodologically, the two extremes are the Bayes Information Criterion (BIC, Schwarz 1978), which penalizes additional model parameters more drastically than the AIC[10], 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 consensus 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. Rather, we think that a

---

[9]It has been suggested to use the Δ_js as stand-alone model selection criterion using the following classification (Burnham and Anderson 2004): $\Delta_j \leq 3$: substantial support, $4 \leq \Delta_j \leq 7$: moderate support, $\Delta_j \geq 10$: essentially no support. We disagree with this view since the Δ_j values critically depend on the composition of the set of candidate size distribution models. Therefore, using only this criterion for model selection might result in a choice from a set out of which all elements are overall implausible – in the sense of Step 1 – descriptions of the original data.

[10]$\text{BIC} = -2 \ln L(\hat{\Theta}_1, \dots, \hat{\Theta}_M) + M \ln N$

more complicated model that lies in $\mathcal{M}'$ is not necessarily worse than a simpler model within $\mathcal{M}'$, because it has deeper explanatory power. In our Bayequentist framework, we therefore advocate the use of the AIC in Step 2 because it provides a formal compromise between a ranking 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.

As result of Step 2, we have the ordered set $\mathcal{M}''$. It contains the same elements as $\mathcal{M}'$, now ordered according to their relative weights of evidence which can also be interpreted as size distribution model probabilities given the data and the other elements in $\mathcal{M}''$. From a statistical point of view, the ordered set $\mathcal{M}''$ is the best we can do. The framework up to this point is designed to capture statistical significance, both in an absolute (model compared to mere chance, Step 1) and a relative sense (model compared to the other candidates, Step 2). A procedure to also capture 'scientific significance' (cf. McCloskey 1995) of the fits is introduced in the following step.

3.4 Step 3: Microfoundation

Is there any good reason *not* to stick to the statistically best-fitting model out of $\mathcal{M}''$, but to actually choose only number 2 or 3 from the ranking? Yes, we argue in this section, there is.

We maintain that a size-distribution model should achieve both, a good fit to empirical data (in the sense of Sections 3.2 and 3.3) and it should come with a plausible microfoundation. By microfoundation, we refer to any microscopic model that can be shown to generate the overall size distribution model. For example, Gibrat's Law which states that the individual size of a firm and its growth rate are independent can be analytically shown to give rise to an overall lognormal firm size distribution (cf. Sutton 1997). In fact, our criterion of microfoundation is a relaxation of a proposition already made in the context of income distributions: a good overall income size distribution model should be based on a plausible *stochastic* model (Dagum 1983, Reed and Wu 2008). This very particular understanding of microfoundation means that there exists a stochastic

differential equation[11]

$$dX_t = f(X_t,t)dt + g(X_t,t)dW_t \tag{6}$$

that describes the size evolution increment dX_t of the individual parts of the economic system in question[12] between two neighboring time instants t and $t - dt$. dW_t is the increment of a standard Wiener process with expected value $\mathbb{E}[W_t] = 0$ and variance $\text{Var}[W_t^2] = t$ 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 by the microscopic stochastic growth process undergone by the individual constituents of the population.

This stochastic differential equation understanding of microfoundation is what can be mostly found in the literature. It has the appeal of being able to analytically establish a relation between the dynamic behavior of individual parts of the system and the overall size distributive outcome. However, this is only one out of many possible microfoundations. In a broader sense, this could be any agent-based, rule-based, or other type of microscopic model. To answer the question whether a certain model not having an analytical solution provides a microfoundation for a certain size distribution model, we suggest to use the following criterion: Calculate the p-values using the plausibility-of-fit algorithm from Step 1 for sufficiently many[13] runs of the microscopic model separately. If the resulting p-values are greater than 0.1 more often than not, then it cannot reliably be ruled out that the microscopic model in question does indeed provide a statistically plausible microfoundation for a certain size distribution model.

The bottom line 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 size distribution comes with a suitable microfoundation. Our argument here is that we suggest to deselect the size distribution models that do not have a microfoundation. This may even imply that the statistically best-fitting size distribution models from $\mathcal{M}''$ get deselected and are therefore not contained in $\mathcal{M}'''$ anymore. On the other hand, the situation $|\mathcal{M}'''| > 1$ would imply to select the size distribution model from $\mathcal{M}'''$ that is ranked highest in $\mathcal{M}''$ because the relative ranking in $\mathcal{M}'''$ remains the same as

---

[11]We refer to stochastic differential equations in the Itō (and not Stratonovich) sense of stochastic integration (cf. e.g. Higham (2001)). Yet, our argument does not depend at all on this specification.

[12]These may for example be individual wages of laborers or city sizes in a country.

[13]We consider 100 runs as the absolute minimum.

in $\mathcal{M}''$ conditional on the fulfillment of the microfoundation criterion. If $\mathcal{M}'''$ is empty, the best fit is given by the ranking in $\mathcal{M}''$. However, in this case, there is no inference about the system and its constituents possible and the whole fitting exercise is devoid of any insight other than the descriptive findings.

In summary, we have the set of size distribution models $\mathcal{M}'''$ with microfoundation. i.e. with a microscopic model that gives a theoretical explanation of the observed overall size-distributive outcome. Additionally, $\mathcal{M}'''$ inherits the statistical ranking from Step 2 so that the model ranked first in $\mathcal{M}'''$ is the statistically best-fitting model from the original candidate set $\mathcal{M}$ that also has a microfoundation.

3.5 The framework at a glance

In Figure 3, we illustrate our Bayequentist framework for the identification of size distributions in empirical data. It consists of three steps which aim at different aspects we think a good size distribution fit to empirical data should fulfill. 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 3.2. Second, given the data, it should perform sufficiently well relative to the competing size distribution 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 described in Section 3.3. 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 (cf. 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. As to motivation, 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 description of the 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 fitting theory to data, it seems that the underlying question is not only 'Which model fits best?' but also 'What
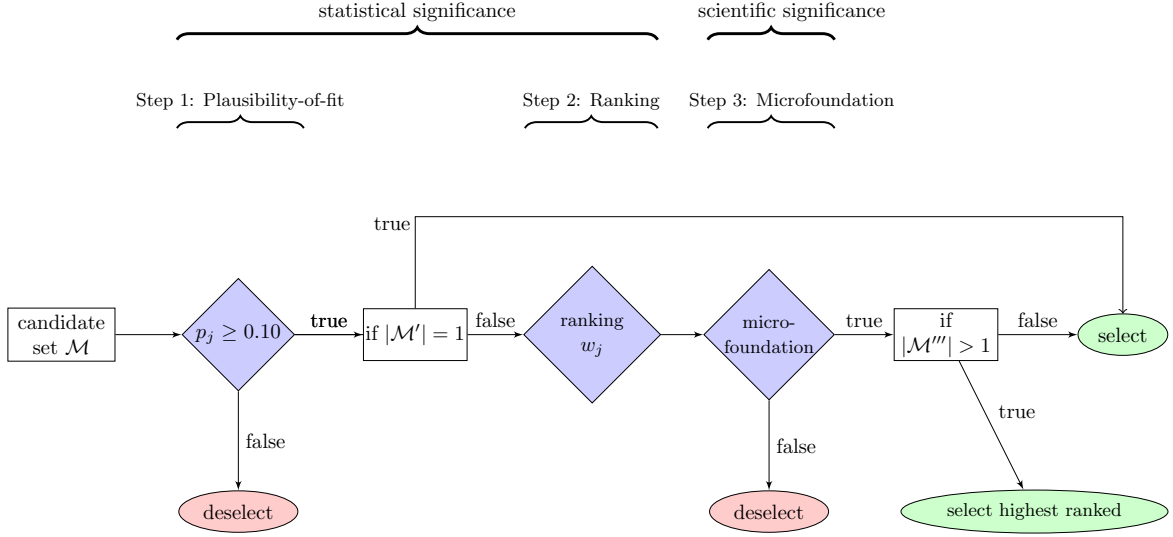
Figure 3: Our proposed Bayequentist model selection framework for size distributions in empirical data: the frequentist Step 1 checks for *absolute* goodness-of-fit in that it checks the plausibility of the hypothesis compared to mere chance. The Bayesian Step 2 checks for *relative* goodness-of-fit as it assesses the plausibility of competing hypotheses compared to each other. Steps 1 and 2 are both 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.

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. As to procedural reasons two things might happen that are – for the sake of clearness – not included in Figure 3 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. $|\mathcal{M}'''| > 1$. 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 the set $\mathcal{M}$ was misspecified, which means that one needs to reconsider the choice of the candidate size distribution models in $\mathcal{M}$ in general and possibly start over again. On the other hand, if $|\mathcal{M}'''| > 1$, we recommend to re-examine the model performances in Steps 1 and 2 and select according to the weights of evidence (w_j, Equation 5) ranking from Step 2. Ultimately, the research question sets the frame.

4   Illustration: commercial cattle farms in Namibia

In this section, we illustrate our Bayequentist framework (cf. Section 3) with size data of Namibian commercial cattle farms that have been surveyed in 2008 (Olbrich et al. 2009/2012). For illustrative purposes, we slightly deviate from the proposed procedure in that we present the results in every Step for every model in the original candidate set $\mathcal{M}$. In Section 4.1, we describe the data set before we present the results of our model selection framework in Sections 4.2 through 4.4.

4.1   Data

The data sample of 399 Namibian commercial cattle farms is described in detail by Olbrich et al. (2009/2012). It contains two specifications of 'size': the number of cattle held on the farm or area in hectares. The former has been hypothesized to be a 'proxy for wealth' (Olbrich et al. 2009: 19). Namibia's semi-arid climate with a dry and a wet season in each year causes cattle numbers on each farm to vary over the course of one year, for example due to pasture and herd management. Therefore, each record contains entries about the cattle count in November and in April. We take the average of these two values to measure the farm size in terms of cattle number, thereby correcting 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-number data in this section is based on $N = 351$ data points, and fits to the area data are 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 Figure 4. Table 1 shows descriptive statistics of the data set while Figure 4 plots histograms of the cattle-number and area distributions.

17

Table 1: Descriptive statistics of the sample of Namibian commercial cattle farms.

descriptive statistics	cattle [number]	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) cattle number

(b) area
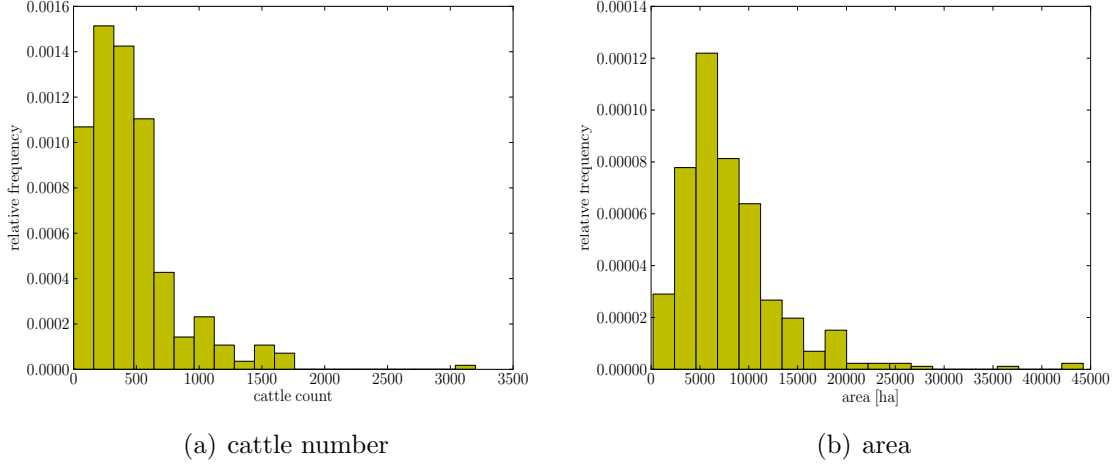
Figure 4: Distribution of commercial cattle farm sizes in Namibia as measured in cattle number (a) and hectares (b).

4.2   Step 1: Plausibility-of-fit

Table 2 shows the results of Step 1 of our framework (Section 3). Although the Pareto distribution does not have the same support as the other size distribution models and although it is nested within the dPlN distribution, we include it here as a robustness check for the plausibility-of-fit method. We base our estimations for each p-value on $Z = 2500$ synthetic data sets so that the resulting p-values are accurate up to approximately 1% (cf. Equation 2). For the cattle-number data, we find that three size distribution models pass the plausibility-of-fit test, the Dagum ($p = 0.39$), the Pareto ($p = 0.33$) and the dPlN ($p = 0.23$) while none of the candidate size distribution models pass the plausibility-of-fit test for the areal data. Because of nestedness, results for Pareto and dPlN are

akin. Regarding parameter estimation, we find similar parameters for the lognormal, the Pareto and the dPlN distributions which reflects again the nestedness of the Pareto and lognormal within the dPlN. Indeed, we find a so-called Pareto coefficient of $\alpha = 3.22$ that is matched by a tail index of the dPlN of $\alpha = 3.52$ for the cattle data. For the areal data, we find $\alpha = 3.25$ for the Pareto distribution and a tail index of the dPlN of 3.28. Similarly, the parameters describing the dPlN's lognormal main body[14] – ν and τ – match the estimates that we find for the lognormal fits (cattle number: $\nu = 6.22$ compared to $\mu = 5.80$ and $\tau = 0.42$ compared to $\sigma = 0.89$, area: $\nu = 8.96$ compared to $\mu = 8.78$ and $\tau = 0.35$ compared to $\sigma = 0.68$).

Table 2: 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%. Models with a p-value smaller than 0.1 are deselected. In spite of having a smaller support than the other candidate models, we also report the estimates for the Pareto distribution here as a robustness check. Definitions of models can be found in Appendix A.

model	cattle number 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),	0.23*	$\alpha = 3.28$ (2.23), $\beta = 2.05$ (0.82),	0.00
	$\nu = 6.22$ (0.11), $\tau = 0.42$ (0.10)		$\nu = 8.96$ (0.08), $\tau = 0.35$ (0.08)	
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),	0.39*	$a = 3.32$ (1.09), $b = 8366.58$ (1.08),	0.00
	$p = 0.45$ (1.18)		$p = 0.63$ (1.18)	
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

Thus, in this case, $\mathcal{M}'$ contains two models, namely the dPlN and the Dagum distribution for the cattle-number data while in case of the area data, $\mathcal{M}'$ is empty.

4.3   Step 2: Model ranking

In Table 3, we list the results of the second step of our Bayequentist model selection framework (Section 3.3). It does not make sense to include the Pareto distribution here because, with this method, we can only meaningfully rank size distributions that have the

---

[14]Being a crossover of double Pareto and lognormal distribution, the dPlN roughly behaves like a Pareto distribution (i.e. like a power law) in its tails and like a lognormal distribution else (Reed and Jorgensen 2004).

same support domain. Also, we include the results of the area data for illustration in spite of the negative result of Step 1. We document every important number in the ranking process from left to right, and for each farm size data set (cf. Table 3): loglikelihoods $\mathcal{L}_j$, AIC scores and resulting Akaike weights w_j. Loglikelihoods $\mathcal{L}_j$ of Dagum and dPlN are almost identical with slight advantages for the Dagum model for the cattle data ($\mathcal{L}_{\text{dPlN}} = -2462.85$ compared to $\mathcal{L}_{\text{Dagum}} = -2462.70$) and the dPlN in case of the area data ($\mathcal{L}_{\text{dPlN}} = -3817.57$ compared to $\mathcal{L}_{\text{Dagum}} = -3818.04$). The AIC scores reflect this finding as these two models rank first (Dagum, AIC $= 4931.41$ for cattle numbers and 7642.08 for area data and second (dPlN, AIC $= 4933.71$ for cattle and 7643.15 for areal data). Using the definition of the model weights w_i (Equation 5), we find substantial evidence for both of the two models (cattle number: $w_{\text{dPlN}} = 23.30\%$ compared to $w_{\text{Dagum}} = 73.66\%$, area: $w_{\text{dPlN}} = 36.00\%$ compared to $w_{\text{Dagum}} = 61.40\%$). With a cumulative Akaike weight of about 3% for both data sets, the other size distribution models play no significant role. The quality of fit of the dPlN as well as the Dagum distribution to our data is illustrated

Table 3: Results of Step 2: Loglikelihoods (logarithm of Equation 1), AIC scores (Equation 3) and Akaike weights (Equation 5). The resulting model ranking is Dagum, dPlN, Gamma, Weibull, Fisk, lognormal (cattle number) and Dagum, dPlN, Fisk, Fisk, lognormal, Weibull (area).

model	cattle number			area		
	loglikelihood	AIC	weight [%]	loglikelihood	AIC	weight [%]
lognormal	-2492.21	4988.42	<0.00	-3836.59	7677.18	<0.00
dPlN	-2462.85	4933.71	23.30	-3817.57	7643.15	36.00
Weibull	-2470.72	4945.43	0.07	-3841.32	7686.63	<0.00
Dagum	-2462.70	4931.41	73.66	-3818.04	7642.08	61.40
Fisk	-2473.82	4951.64	<0.00	-3822.21	7648.41	2.59
Gamma	-2466.91	4937.83	2.97	-3827.67	7659.35	0.01

in Figures 5, 6 and 7 where we use for computational convenience that the logarithm of a dPlN distributed variable is Normally Laplace distributed (cf. Lemma 1, Appendix B).

In summary, Step 2 of our framework leads to the following result: In case of the cattle-number data, the Dagum distribution is ranked first while the dPlN is ranked second. The area data however illustrates nicely why Step 2 alone is not sufficient for size distribution model selection: it yields a relative ranking (1.Dagum, 2.dPlN, 3.Fisk, 4.Gamma), but without any qualification of whether the fits are statistically plausible.
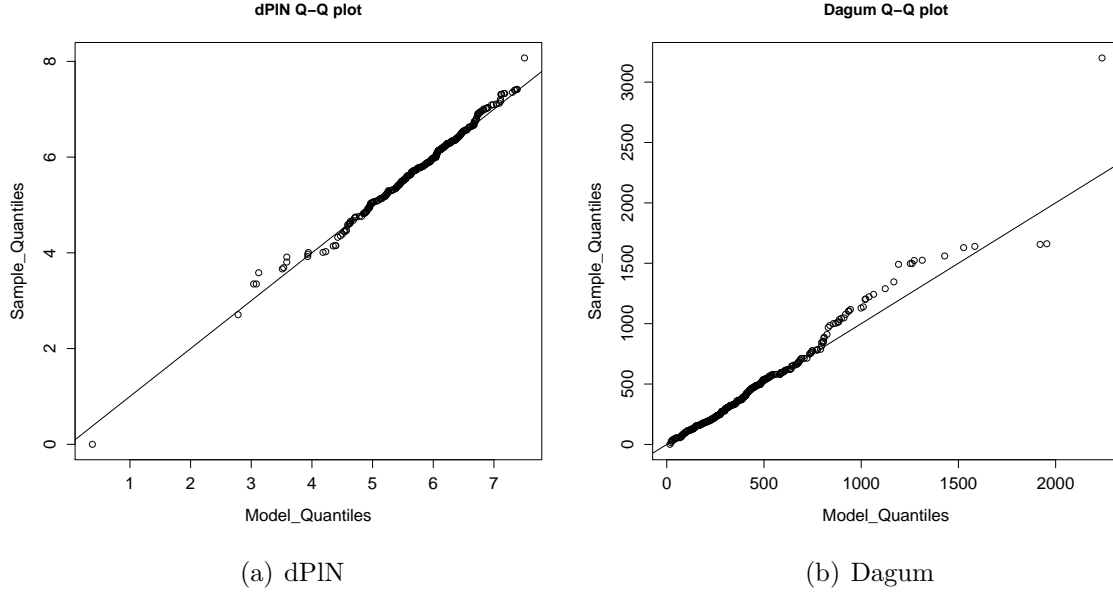
(a) dPlN

(b) Dagum

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

(a) probability densities
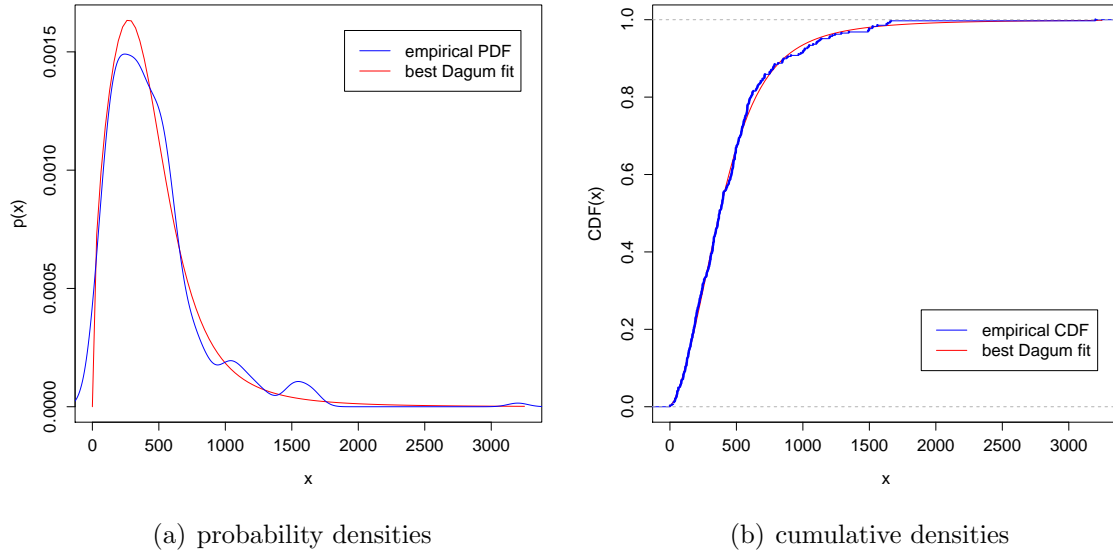
(b) cumulative densities

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

4.4 Step 3: Microfoundation

Camillo Dagum (1977) defined the distribution of his name 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 any stochastic or other kind of micro model to explain how the observed
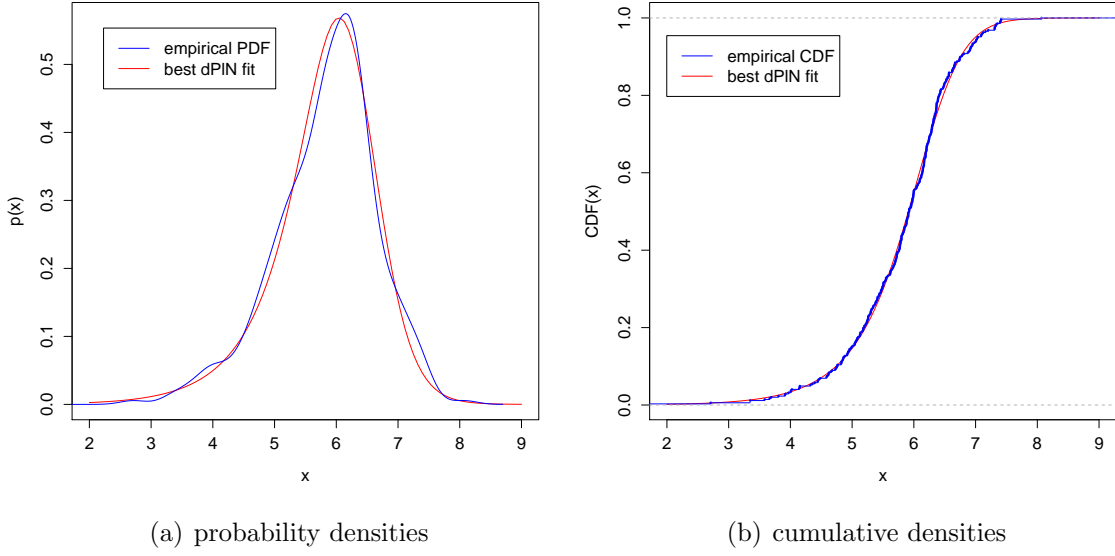
| (a) probability densities | (b) cumulative densities |

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

overall distribution might have been generated from an underlying individual microscopic evolutionary process. Therefore, it lacks 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 Gibrat's Law with two assumptions about the entities under consideration: (1) the initial size distribution is lognormal and (2) the entities under consideration are not equally old and have an overally exponential age distribution. That means, if we have a set of economic entities that follow a lognormal distribution initially *and* that have an exponential age distribution *and* that individually grow according to a geometric Brownian motion, then that set will be dPlN distributed.

Therefore, in spite of the worse relative goodness-of-fit statistics from Step 2 (cf. Section 4.3), we select the dPlN as the best size distribution model out of the six candidate models.

The finding that the dPlN model provides the best fit to the farm size (measured in cattle number) distribution of Namibian commercial cattle farms contains several messages. First, we may infer that Gibrat's Law holds in our farm sample, albeit with two modifying crucial assumptions: (1) exponentially distributed farm age structure and (2) lognormal initial farm size distribution. In general, Gibrat's Law states that the size of a

22

firm and its growth rate are independent (Gibrat 1931). Mathematically, this translates to the following stochastic differential equation:

$$\frac{\mathrm{d}S}{S} = \mu \mathrm{d}t + \sigma \mathrm{d}W, \quad S(0) = S_0, \quad t \geq 0 \tag{7}$$

where the sign of the drift term μ determines the behavior of the expected value of $S(t)$ and where $W(t)$ is the standard Wiener process as introduced in Section 3.4. Thus, if $\mu > 0$, then $\mathbb{E}\left[S(t)\right] > S_0$, and vice versa. The parameter σ models the influence of randomness on the overall process S and the larger the absolute value of σ, the higher the impact of randomness on the behavior of $S(t)$.

In the following, we give an interpretation of Gibrat's Law in the context of Namibian commercial cattle farms. The stochastic differential equation (Equation 7) is equivalent to[15]

$$S(t) = S_0 \cdot \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W(t)\right]. \tag{8}$$

That is, the finding of the dPlN in our data suggests that individual farm sizes evolve according to a stochastic exponential growth process (cf. Equation 8). S_0 is the initial size of the farm and it holds that $S(t) > 0$ for every positive t. If $\sigma = 0$, Equation 8 reduces to simple exponential growth with growth rate μ. Thus, μ reflects the deterministic growth rate of the farm which the farmer can influence through his management decisions such as buying or selling cattle, acquisition of new machinery to augment productivity, hiring or laying off staff and so forth. Consequently, the expected farm size at any positive time instant is $\mathbb{E}\left[S(t)\right] = S_0 e^{\mu t}$. 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 (Olbrich et al. 2009). The farmer cannot influence this environmental risk, it is thus externally given.

Second, the stochastic exponential growth process (Equation 8) 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

---

[15]This follows by solving the stochastic differential equation (Equantion 7) with help of the substitution $X(t) = \ln S(t)$ combined with Itō's lemma (cf. e.g. Protter 2004).

process (Equation 8) 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.

5 Discussion

In this section, we 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 3.2) and (2) the combination of frequentist and Bayesian statistical methods with a qualitative criterion of model microfoundation in an integrated three-step framework. As such, our framework may 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 3.2.

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 2, Section 3.2), 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 upon which we build our generalized version. 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 discriminative power and the the more parameters in a model the harder it is to rule out, as expected (Figure 3.2).

Justification of Step 3. Step 3 could obviously be challenged for being a somewhat broad criterion in an otherwise quite specific framework. Yet, we would rather see this as an asset than as a shortcoming 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 general way. Second, it prevents Occam's razor from shaving away not only 'all but what is necessary' but possibly from shaving away more than that.

We have already argued in Section 3.5 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[16] 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 in space and time 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 relativity per se for being overly complicated because special relativity has more explanatory power than classical mechanics. 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

---

[16]Special relativity explains the movement of bodies in space and time for velocities close to the speed of light.

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 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.

6   Conclusion

In this paper, we have proposed and illustrated a new statistical framework for identifying theoretical size distribution models in empirical data. The innovative feature of our framework is the three-step combination of frequentist and Bayesian statistical methods (Steps 1 and 2) with the criterion of microfoundation (Step 3). Step 1 is based on a frequentist plausibility-of-fit testing algorithm originally proposed in the context of testing for Pareto's rank-size rule 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. The test answers the question of how plausible some hypothesized size distribution model is compared to mere chance. To actually compare competing size distribution models directly with each other, we have proposed to combine this in Step 2 with the Bayesian method of calculating Akaike weights that can be interpreted as relative model probabilities given the data. We have argued that Step 2 answered the question of strength of evidence compared to the other candidates. Lastly, we have proposed Step 3 which demands to also take into account possible inferences from overall size distribution to individual dynamics.

Analyzing a unique and highly resolved data set of 399 Namibian commercial cattle farms (Olbrich et al. 2009), we have demonstrated both, use and usefulness 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. Thus, we

were able to infer 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 driver for farm size growth rather than just a minor parameter.

Our analysis suggests at least two 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[17] (Giesen et al. 2010, Giesen and Suedekum 2012) – then the two major strands in literature that either find Gibrat's or Pareto's Law to hold, could be reconciled. Second, it needs to be further investigated to what extent the growth dynamics discovered here can be generalized to other firms. Further investigations on this issue might have far reaching implications for the debate on the role of firm sizes for the economy and its growth.

Acknowledgments

---

[17]After completion of our manuscript, we learned of the contribution of González-Val et al. (2013). They fit city size data of 34 OECD countries to the lognormal, q-exponential, log-logistic and dPlN distributions and report that in about 87% of cases, the dPlN fits the data best. They judge the quality of their fits by their respective Kolmogorov-Smirnov and Cramér-von Mises p-values as well as AIC and BIC scores. Hence, they choose a model choice approach similar to ours and arrive at a similar conclusion. However, our approach here studies the problem of model choice from a more general perspective and goes beyond theirs in explicitly formulating the criterion of model microfoundation as a separate step.

Appendix A: mathematical representation and inter-relation 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 3.2 and how the members of these are related. The classification presented here is due to Kleiber and Kotz (2003). Table 4 provides the mathematical definition of these size distribution models.
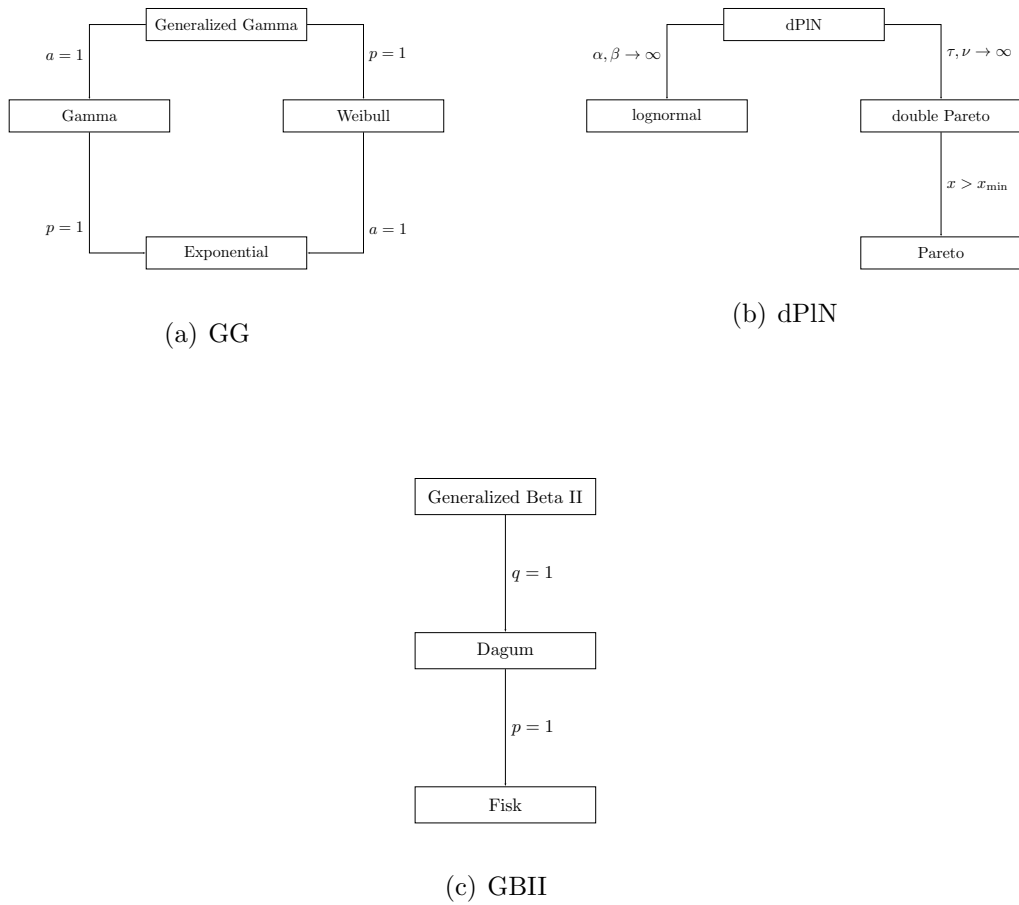
(a) GG

(b) dPlN

(c) GBII

Figure 8: Functional supergroups of size distribution models and relationships within each group, presented in hierarchical order, i.e. decreasing generality from top to bottom in each diagram. (a) The generalized gamma (GG) distribution and its siblings. (b) The double Pareto lognormal family; (c) The family originating from the generalized beta distribution of the second order (GBII).

Table 4: The size distribution models most commonly used in economics. Φ denotes the cumulative density 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	$\frac{\alpha\beta}{\alpha+\beta} x^{-\alpha-1} e^{(\alpha\nu + \frac{\alpha^2\tau^2}{2})} \Phi\left(\frac{\ln(x) - \nu - \alpha\tau^2}{\tau}\right) +$ $\frac{\alpha\beta}{\alpha+\beta} x^{\beta-1} e^{(-\beta\nu + \frac{\beta^2\tau^2}{2})} \Phi^c\left(\frac{\ln(x) - \nu + \beta\tau^2}{\tau}\right)$	Reed and Jorgensen (2004)
Pareto	2	$\frac{1}{\alpha-1}\left(\frac{x}{x_{\min}}\right)^{-\alpha}$	Pareto (1896)
Weibull	2	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^k\right)$	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	$\frac{\delta}{\gamma}\left(\frac{x}{\gamma}\right)^{\delta-1}/(1 + (\frac{x}{\delta})^\beta)^2$	Fisk (1961)
Gamma	2	$x^{\kappa-1} \frac{\exp(-\frac{x}{\Theta})}{\Theta^\kappa \Gamma(\kappa)}$	Ammon (1895)

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 (Equation (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

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

which can be simplified to

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

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

$$\tilde{P}(x) = \Phi\left(\frac{\ln x - \nu}{\tau}\right) - \frac{1}{\alpha + \beta}\left[\beta x^{-\alpha}A(\alpha,\nu,\tau)\Phi\left(\frac{\alpha\tau^2 - \ln x + \nu}{\tau}\right) + \right.$$
$$\left. \alpha x^{\beta}A(-\beta,\nu,\tau)\Phi^c\left(\frac{\ln x - \nu + \beta\tau^2}{\tau}\right)\right] = P_{\mathrm{dPlN}}(x) \quad (9)$$

because of $\Phi(-z) = 1 - \Phi(z)$. The last Equation (9) is the CDF of the double Pareto lognormal (eq. (23), Reed and Jorgensen 2004). $\square$

References

Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, *Proceedings of the Second International Symposium on Information Theory Budapest*, pp. 267–281

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

Baker, A. (2011), Simplicity, In: Zalta E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, Summer 2011 edition,

http://plato.stanford.edu/archives/sum2011/entries/simplicity/

Bandourian, R., J.B. McDonald and R.S. Turley (2002), A comparison of parametric models of income distribution across countries and over time, *Luxembourg Income Study working paper series* 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 (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

Chotikapanich, D., W.E. Griffiths D.S. Prasada Rao and V. Valencia (2012), Global income distributions and inequality, 1993 and 2000: Incorporating country-Level inequality with beta distributions, *The Review of Economics and Statistics* 94(1), 52–73

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), Income distribution models, In: Kotz, S. and N.L. Johnson (eds.), *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(4), 1676–1683

Efron, B. (2005), Bayesians, Frequentists, and Scientists, *Journal of the American Statistical Association* 100(469), 1–5

Evans, D.S. (1987), Tests of alternative theories of firm growth, *Journal of Political Economy* 85(4), 657–674

Fisk, P. (1961), The graduation of income distributions, *Econometrica* 29(2), 171–185

Gabaix, X. (1999), Zipf's law for cities: an explanation, *The Quarterly Journal of Economics* 114(3), 739–767

Gibrat, R. (1931), *Les Inégalités Economiques*, 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(2), 129–137

Giesen, K. and J. Suedekum (2012), The size distribution across all 'cities': a unifying approach, *IZA Discussion Paper No.* 6352

González-Val, R., A. Ramos F. Sanz and M. Vera-Cabello (2013), Size distributions for all cities: Which one is best? *MPRA Paper No. 44314*

González-Val, R., L. Lanaspa and F. Sanz (2012), New evidence on Gibrat's law for cities, *IEB Working paper No.* 2012/18

Greenwood, P.E. and M.S. Nikulin (1996), *A Guide to Chi-Squared Testing*, J. Wiley, New York

Hart, P.E. and N. Oulton (1996), Growth and size of firms, *The Economic Journal* 106(438), 1242–1252

Higham, D.J. (2001), An algorithmic introduction to numerical simulation of stochastic differential equations, *SIAM Review* 43(3), 525–546

Johnson, D.H. (1999), The insignificance of statistical significance testing, *Journal of Wildlife Management* 63(3), 763–772

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

Kaye, S. (2007), William of Ockham, In: Fieser, J. and B. Dowden (eds.), *IEP – The Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/ockham/#H9

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(4), 1672–1675

Mansfield, E. (1962), Entry, Gibrat's law, innovation, and the growth of firms, *American Economic Review* 52, 1023–1051

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

Neyman, J. and E. Pearson (1933), On the problem of the most efficient tests of statistical hypotheses, *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character* 231, 289–337

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

Olbrich, R., M.F. Quaas and S. Baumgärtner (2012), A survey of commercial cattle farmers in semi-arid rangelands of Namibia on risk, sustainability and management. *Journal of Applied Social Science Studies* 132(3), 463–471

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

Pinkovskyi, M. and X. Sala i Martin (2009), Parametric estimations of the world distribution of income, *The National Bureau of Economic Research Working Paper Series*, Working Paper No. 15433

Protter, P.E. (2004), *Stochastic Integration and Differential Equations*, 2nd edition. Springer, New York

Raftery, A. (1986), Choosing Models for Cross-Classifications, *American Sociological Review* 51(1), 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), New four- and five parameter models for income distributions, In: D. Chotikapanich (ed.), *Modeling Income Distributions and Lorenz Curves*

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

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

Toda, A.A. (2012), The double power law in income distribution: Explanations and evidence, *Journal of Economic Behavior & Organization* 84(1), 364–381

Tsou, W.-T. (2012), Central place theory and city size distribution, *The Economic Journal* 122, 903–932

Vogel Carey, T. (2010), Parsimony (in as few words as possible), *Philosophy Now (UK)* 81, retrieved on 10/27/2012

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

Working Paper Series in Economics

(recent issues)

No.264:   *Chiara Franco* and *John P. Weche Gelübcke*: The death of German firms: What role for foreign direct investment?, February 2013

No.263:   *Joachim Wagner*: Are low-productive exporters marginal exporters? Evidence from Germany, February 2013

No.262:   *Sanne Hiller, Philipp J. H. Schröder,* and *Allan Sørensen*: Export market exit and firm survival: theory and first evidence, January 2013

No.261:   *Institut für Volkswirtschaftslehre*: Forschungsbericht 2012, Januar 2013

No.260:   *Alexander Vogel* and *Joachim Wagner*: The Impact of R&D Activities on Exports of German Business Services Enterprises : First Evidence from a continuous treatment approach, December 2012

No.259:   *Christian Pfeifer*: Base Salaries, Bonus Payments, and Work Absence among Managers in a German Company, December 2012

No.258:   *Daniel Fackler, Claus Schnabel,* and *Joachim Wagner*: Lingering illness or sudden death? Pre-exit employment developments in German establishments, December 2012

No.257:   *Horst Raff* and *Joachim Wagner*: Productivity and the Product Scope of Multi-product Firms: A Test of Feenstra-Ma, December 2012 [published in: Economics Bulletin, 33 (2013), 1, 415-419]

No.256:   *Christian Pfeifer* and *Joachim Wagner*: Is innovative firm behavior correlated with age and gender composition of the workforce? Evidence from a new type of data for German enterprises, December 2012

No.255:   *Maximilian Benner*: Cluster Policy as a Development Strategy. Case Studies from the Middle East and North Africa, December 2012

No.254:   *Joachim Wagner* und *John P. Weche Gelübcke*: Firmendatenbasiertes Benchmarking der Industrie und des Dienstleistungssektors in Niedersachsen – Methodisches Konzept und Anwendungen (Projektbericht), Dezember 2012

No.253:   *Joachim Wagner*: The Great Export Recovery in German Manufacturing Industries, 2009/2010, November 2012

No.252:   *Joachim Wagner*: Daten des IAB-Betriebspanels und Firmenpaneldaten aus Erhebungen der Amtlichen Statistik – substitutive oder komplementäre Inputs für die Empirische Wirtschaftsforschung?, Oktober 2012

No.251:   *Joachim Wagner*: Credit constraints and exports: Evidence for German manufacturing enterprises, October 2012

No.250:   *Joachim Wagner*: Productivity and the extensive margins of trade in German manufacturing firms: Evidence from a non-parametric test, September 2012 [published in: Economics Bulletin 32 (2012), 4, 3061-3070]

No.249:   *John P. Weche Gelübcke*: Foreign and Domestic Takeovers in Germany: First Comparative Evidence on the Post-acquisition Target Performance using new Data, September 2012

No.248:   *Roland Olbrich, Martin Quaas,* and *Stefan Baumgärtner*: Characterizing commercial cattle farms in Namibia: risk, management and sustainability, August 2012

No.247:   *Alexander Vogel* and *Joachim Wagner*: Exports, R&D and Productivity in German Business Services Firms: A test of the Bustos-model, August 2012

No.246:   *Alexander Vogel* and *Joachim Wagner*: Innovations and Exports of German Business Services Enterprises: First evidence from a new type of firm data, August 2012

No.245:   *Stephan Humpert*: Somewhere over the Rainbow: Sexual Orientation Discrimination in Germany, July 2012

No.244:   *Joachim Wagner*: Exports, R&D and Productivity: A test of the Bustos-model with German enterprise data, June 2012 [published in: Economics Bulletin, 32 (2012), 3, 1942-1948]

No.243:   *Joachim Wagner*: Trading many goods with many countries: Exporters and importers from German manufacturing industries, June 2012 [published in: Jahrbuch für Wirtschaftswissenschaften/Review of Economics, 63 (2012), 2, 170-186]

No.242:   *Joachim Wagner*: German multiple-product, multiple-destination exporters: Bernard-Redding-Schott under test, June 2012 [published in: Economics Bulletin, 32 (2012), 2, 1708-1714]

No.241:   *Joachim Fünfgelt* and *Stefan Baumgärtner*: Regulation of morally responsible agents with motivation crowding, June 2012

No.240:   *John P. Weche Gelübcke*: Foreign and Domestic Takeovers: Cherry-picking and Lemon-grabbing, April 2012

No.239:   *Markus Leibrecht* and *Aleksandra Riedl*: Modelling FDI based on a spatially augmented gravity model: Evidence for Central and Eastern European Countries, April 2012

No.238:   *Norbert Olah, Thomas Huth* und *Dirk Löhr*: Monetarismus mit Liquiditätsprämie Von Friedmans optimaler Inflationsrate zur optimalen Liquidität, April 2012

No.237:   *Markus Leibrecht* and *Johann Scharler*: Government Size and Business Cycle Volatility; How Important Are Credit Contraints?, April 2012

No.236:   *Frank Schmielewski* and *Thomas Wein*: Are private banks the better banks? An insight into the principal-agent structure and risk-taking behavior of German banks, April 2012

No.235:   *Stephan Humpert*: Age and Gender Differences in Job Opportunities, March 2012

No.234:   *Joachim Fünfgelt* and *Stefan Baumgärtner*: A utilitarian notion of responsibility for sustainability, March 2012

No.233:   *Joachim Wagner*: The Microstructure of the Great Export Collapse in German Manufacturing Industries, 2008/2009, February 2012 [published in: Economics - The Open-Access, Open-Assessment E-Journal, Vol. 7, 2013-5]

No.232:   *Christian Pfeifer* and *Joachim Wagner*: Age and gender composition of the workforce, productivity and profits: Evidence from a new type of data for German enterprises, February 2012

No.231:   *Daniel Fackler, Claus Schnabel,* and *Joachim Wagner*: Establishment exits in Germany: the role of size and age, February 2012

No.230:   *Institut für Volkswirtschaftslehre*: Forschungsbericht 2011, January 2012

No.229:   *Frank Schmielewski*: Leveraging and risk taking within the German banking system: Evidence from the financial crisis in 2007 and 2008, January 2012

No.228:   *Daniel Schmidt* and *Frank Schmielewski*: Consumer reaction on tumbling funds – Evidence from retail fund outflows during the financial crisis 2007/2008, January 2012

No.227:   *Joachim Wagner*: New Methods for the Analysis of Links between International Firm Activities and Firm Performance: A Practitioner's Guide, January 2012

No.226: *Alexander Vogel* and *Joachim Wagner*: The Quality of the KombiFiD-Sample of Business Services Enterprises: Evidence from a Replication Study, January 2012 [published in: Schmollers Jahrbuch/Journal of Applied Social Science Studies 132 (2012), 3, 379-392]

No.225: *Stefanie Glotzbach*: Environmental justice in agricultural systems. An evaluation of success factors and barriers by the example of the Philippine farmer network MASIPAG, January 2012

No.224: *Joachim Wagner*: Average wage, qualification of the workforce and export performance in German enterprises: Evidence from KombiFiD data, January 2012 [published in: Journal for Labour Market Research, 45 (2012), 2, 161-170]

No.223: *Maria Olivares* and *Heike Wetzel*: Competing in the Higher Education Market: Empirical Evidence for Economies of Scale and Scope in German Higher Education Institutions, December 2011

No.222: *Maximilian Benner*: How export-led growth can lead to take-off, December 2011

No.221: *Joachim Wagner* and *John P. Weche Gelübcke*: Foreign Ownership and Firm Survival: First evidence for enterprises in Germany, December 2011 [published in: International Economics/Économie Internationale, 132(2012), 4, 117-139]

No.220: *Martin F. Quaas, Daan van Soest,* and *Stefan Baumgärtner*: Complementarity, impatience, and the resilience of natural-resource-dependent economies, November 2011

No.219: *Joachim Wagner*: The German Manufacturing Sector is a Granular Economy, November 2011 [published in: Applied Economics Letters, 19(2012), 17, 1663-1665]

No.218: *Stefan Baumgärtner, Stefanie Glotzbach, Nikolai Hoberg, Martin F. Quaas, and Klara Stumpf*: Trade-offs between justices , economics, and efficiency, November 2011

No.217: *Joachim Wagner*: The Quality of the KombiFiD-Sample of Enterprises from Manufacturing Industries: Evidence from a Replication Study, November 2011 [published in: Schmollers Jahrbuch/Journal of Applied Social Science Studies 132 (2012), 3, 393-403]

No.216: *John P. Weche Gelübcke*: The Performance of Foreign Affiliates in German Manufacturing: Evidence from a new Database, November 2011 [published in: Review of World Economics, 149(1) (2013), 151-182]

No.215: *Joachim Wagner*: Exports, Foreign Direct Investments and Productivity: Are services firms different?, September 2011

No.214: *Stephan Humpert and Christian Pfeifer*: Explaining Age and Gender Differences in Employment Rates: A Labor Supply Side Perspective, August 2011

No.213: *John P. Weche Gelübcke*: Foreign Ownership and Firm Performance in German Services: First Evidence based on Official Statistics, August 2011 [forthcoming in: The Service Industries Journal]

No.212: *John P. Weche Gelübcke*: Ownership Patterns and Enterprise Groups in German Structural Business Statistics, August 2011 [published in: Schmollers Jahrbuch / Journal of Applied Social Science Studies, 131(2011), 4, 635-647]