Environmental Uncertainty and Economic Heterogeneity

Von der Fakultät Nachhaltigkeit

der Leuphana Universität Lüneburg

zur Erlangung des Grades

Doktor der Wirtschafts-, Sozial- und Politikwissenschaften

Dr. rer. pol.

genehmigte Dissertation

von

John-Oliver Engler

geboren am 27. September 1983 in Emmendingen

Eingereicht am: 15. September 2014

Betreuer und Gutachter: Prof. Dr. Stefan Baumgärtner

Gutachter: Prof. Dr. Martin F. Quaas

Gutachter: Prof. Dr. Henrik von Wehrden

Tag der Disputation: 18. Dezember 2014

Die einzelnen Beiträge des kumulativen Dissertationsvorhabens sind oder werden wie folgt in Zeitschriften veröffentlicht:

> Engler, J.-O. und S. Baumgärtner (2015). Model choice and size distribution: a Bayequentist approach. *American Journal of Agricultural Economics* 97(3), 978–997, doi: 10.1093/ajae/aau034, Datum der Online-Veröffentlichung: 25.06.2014.
>
> Engler, J.-O., H. von Wehrden und S. Baumgärtner (2014). Farm size, environmental risk and risk preferences: the case of Namibian commercial cattle farming. Einreichung geplant für 2015.
>
> Baumgärtner, S. und J.-O. Engler (2014). An axiomatic foundation of preferences under Knightian uncertainty. Einreichung geplant für 2015.

Einzelne Beiträge sind in den formalen Vorgaben des jeweiligen Publikationsmediums gehalten. Der Autor bittet etwaige Unregelmäßigkeiten zu entschuldigen.

Contents

List of Figures

List of Tables

Acknowledgments

My thesis greatly benefited from discussion, critical feedback, inspiration, and support by quite a few people. First and foremost, I am indebted to Prof. Dr. Stefan Baumgärtner for giving me the chance to pursue my Ph. D. project at his chair under his supervision. I very much appreciated his creativity, openness towards new ideas and cooperation on many occasions. Secondly, I thank Prof. Dr. Martin F. Quaas and Prof. Dr. Henrik von Wehrden for agreeing on refereeing my thesis and for critical discussion of some of the ideas worked out in this thesis.

Some of my colleagues during my four years at Leuphana's Chair of Sustainability Economics and Institute of Economics deserve a special mention: I spent countless hours with Dr. Nikolai Hoberg discussing ideas and thoughts, some of which were thesis-related, laughing, and cheering us up. Dr. Nik, thank you for an unforgettable time! Dr. Dave Abson, Lena Biehler, Dr. Joachim Fünfgelt, Dr. Tanvir Hussain, Felix John, Moritz Meyer, Dr. Roland Olbrich, Dr. Sebastian Strunz, Klara Stumpf and Dr. John P. Weche-Gelübke all contributed to making my time at Leuphana a very pleasant experience, and certainly one that I will always remember.

I would like to express my gratitude to my parents Ingrid and Günther for being supportive to me and my aspirations throughout my career, and up to this point. Special thanks to Hanne and Marta for constantly being curious about my work. Finally, to my soulmate and longtime companion in life Judith, I would like to simply express my gratitude for her being there.

Lüneburg, September 2014

Introduction

Almost any decision we take involves uncertainty regarding future states of the world. If we consider buying a season pass for the open air swimming pool, there is a chance of a very cool and rainy summer that leaves us going to the spa rather than the swimming pool. Or consider a farmer who needs to decide what crops to grow and where, how many livestock to put on his pasture next year, and whether or not to buy insurance. All such decisions are uncertain on multiple dimensions: weather conditions, political circumstances, the price of labor, fodder, insurance, and so forth. Uncertainty is ubiquitous in everyday life, and understanding uncertainty is of considerable importance in good decision making, for individuals as well as for companies and countries. Uncertainty is also a big issue in sustainability science, where much of the focus is on uncertain future states of the world.

Because economics deals with human behavior, uncertainty is thus one of the pivotal issues in economic theory. In recent years, this role has become more and more evident in environmental and resource economics in general (e.g., Brown 2000, Weitzman 2007) and 'sustainability economics' (Baumgärtner and Quaas 2010)[1] in particular. Some of the most urgent questions in these fields involve massive uncertainties, and there has been relatively little progress in dealing with these uncertainties in recent years (Stern 2013, Gollier et al. 2014). These uncertainties relate both to physical changes in the environment and to the impact on human welfare of those changes. For example, the latest report of the Intergovernmental Panel on Climate Change (IPCC) stated that 'no best estimate for equilibrium climate sensitivity can now be given because of a lack of agreement on values across assessed lines of evidence and studies.' (IPCC 2013: 16). Closely related is the debate about the social cost of carbon, which struggles with uncertainty in at least three different ways: (1) uncertain impact of atmospheric CO_2 level on the economy,

[1] As Baumgärtner and Quaas have put it, 'Any study in the field of sustainability economics has to take uncertainty seriously' (Baumgärtner and Quaas 2010: 449).

globally as well as locally, (2) the question of which discount rate to use, which involves somehow numerically quantifying uncertainty on the one hand, and choosing between different ethical approaches to pick the 'right' discount rate on the other, (3) the question of modeling these uncertainties in a balanced and analytically tractable way. This quantification of uncertainty is further complicated by the heterogeneity of the systems under investigation.

The concept of heterogeneity[2] refers to the nonuniformity in a set. In colloquial language, heterogeneity and uncertainty are used quite differently. Usually, any kind of uncertainty is perceived as rather negative, for example by focusing on 'downside risks' and the like while heterogeneity is perceived rather positive, for example in relation with liberal societies and individual freedom. However, under closer scrutiny the distinction between uncertainty and heterogeneity may not be so clear. First of all, even though most people generally dislike uncertainty, they still tend to react very differently when confronted with it. Uncertainty preferences are very individual and thus heterogeneously distributed in any sample. On the other hand, if one thinks about the nature of uncertainty, its source is heterogeneity: it is the heterogeneity of things that might happen and of options to act that we have to choose from that creates uncertainty. Seen this way, we may even say that uncertainty is a cause for heterogeneity as much as it is a consequence of it. This knotty relationship is further complicated by the potential to use heterogeneity – diversification – as a means of reducing uncertainty in economic decision making (Markowitz 1952), typified by the popular phrase 'don't put all your eggs in one basket'. The relationship between these two concepts is thus a very intriguing one.

In this thesis, I discuss uncertainty and heterogeneity in economic theory from different theoretical perspectives and at at different levels of abstraction. The research questions of this thesis develop from two main foci: First, I consider the methodological implications of the heterogeneity of scientific theories when these are confronted with empirical data (Paper 1). Second, I consider different forms of economic uncertainty created by heterogeneity: environmental risk (Paper 2) and Knightian uncertainty (Paper 3). Origin of the first question, that is dealt with in Paper 1, is the seemingly unspectacular observation that there is usually a wide selection, if not a plethora, of theories to explain any given

[2]The term 'heterogeneity' is derived from the two Greek words 'heteros' ('other', 'another', 'different'), and 'genos' ('kind').

empirical finding. The resulting heterogeneity of theoretical models to describe empirical data creates uncertainty regarding the question 'Which model fits best?', and therefore the need for a framework that helps answering this question. I identify and discuss a key example of the lack of such a framework in the academic literature: the debate about which model best describes the distribution of city sizes in urban economics. The first question I address in this thesis takes this up in a very general way: Is there a general framework based on statistics, economics and philosophy of science for choosing between models in a systematic and informationally optimal manner? As a side result, the paper establishes a link between the rather abstract concept of a heterogeneous model set and heterogeneity on the subject level – in Paper 1, this is the economic heterogeneity of individual farms and the environmental risk they are exposed to. From this rather general approach concerning heterogeneity in scientific theories and its implications, the other two questions take the natural step to consider individual decision making in the face of uncertainty. The second question is concerned with individual decision making in the presence of environmental risk: What are the relationships between farm size, environmental risk and individual risk preferences in a sample of 399 commercial cattle farmers in Namibia's semi-arid rangelands? The final question of this thesis deals with individual decision making under Knightian uncertainty (Keynes 1921, Knight 1921), that is, when deciding between different possible acts when possible future states of the world are known, but their probabilities are unknown. Specifically, the question of which fundamental axioms a binary Knightian uncertainty preference relation must satisfy to ensure the existence of a numerical representation of these preferences is investigated. The paper provides definitions of the notions of uncertainty aversion and uncertainty premium, which makes interpersonal comparison of uncertainty attitudes possible. Finally, it provides some illustration of the concepts developed and compares them to other approaches commonly found in the literature.

By investigating these research questions, this thesis makes a contribution to the following research foci of sustainability economics (cf. Baumgärtner and Quaas 2010: 449): relationship between humans and nature, long-run uncertainty, and system understanding and management. It also contributes to all three levels of abstraction of inter- and transdisciplinary research on sustainability problems as defined by Baumgärtner et al. (2008): (1) it works out, refines and reflects on economic concepts, such as rational choice under

Knightian uncertainty and goodness-of-fit, (2) it develops and discusses abstract models of human preferences under Knightian uncertainty and farm growth under environmental risk, and (3) it uses data on Namibian commercial cattle farming to test some hypotheses from the literature (Figure 1).

The following section gives an overview of the three papers that my thesis is comprised of. Each paper is reviewed in the following manner: first, I will briefly summarize key questions, the methodological approach and central results. Second, I discuss limitations and possible further steps. In section 2, I draw specific and overarching conclusions from this research.

1 Research papers

This thesis contains three research papers that look at heterogeneity and uncertainty from different theoretical, methodological and empirical perspectives and at different levels of abstraction (cf. Figure 1). In Paper 1, we develop a new approach to model choice problems, i.e. the problem of selecting one specific theoretical model from a set of multiple candidates, given the actual data. We combine aspects from statistics, economics and philosophy of science to obtain a framework that is able to tackle the heterogeneity of candidate models, while also establishing a productive link to the individual entities studied. We illustrate our framework with a data set of Namibian commercial cattle farms. In Paper 2, we take up this illustration and delve more into the aspect of economic heterogeneity of these farms, and study its relationship to environmental risk. Particularly, we investigate the interdependencies of farm size, environmental risk and individual risk preferences in the data set. In Paper 3, we focus on the special case of decision making under Knightian uncertainty. We show which assumptions on preferences under Knightian uncertainty warrant existence of a numerical preference representation, how different attitudes towards Knightian uncertainty can be modeled, and discuss differences and connections to existing approaches in economic theory.

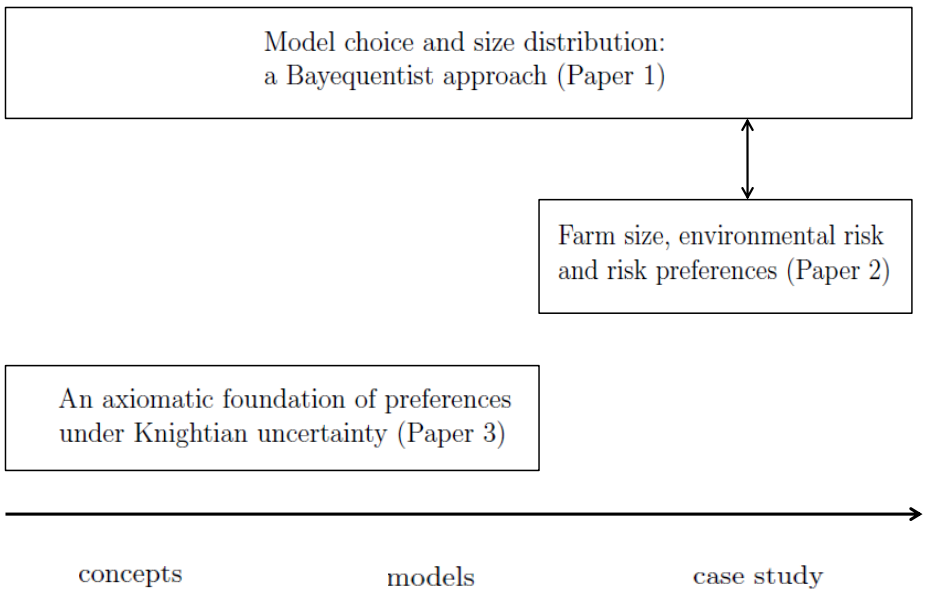


Figure 1: Relationship between papers and levels of abstraction as defined by Baumgärtner et al. (2008). Paper 1 deals with the concept of goodness-of-fit and how one can use it for inference to economic models. Illustration of the method uses case study data from Namibia. Paper 2 tests various hypotheses regarding farm management for the case of Namibian commercial cattle farming. Finally, Paper 3 deals with the concept of rational choice under Knightian uncertainty and provides the concrete model of a Rényi decision maker.

1.1 Paper 1: Model choice and size distribution: a Bayequentist approach

In the paper *Model choice and size distribution: a Bayequentist approach*, we provide a new methodological framework for statistical model choice problems that combines statistics with economics and philosophy of science concepts. In statistics, model choice refers to the problem of selecting the best fitting model from a heterogeneous set of candidates given some empirical data set. In the paper, we look at the specific model choice problem of fitting theoretical size distribution models to empirical size distribution data. Heterogeneity in the model set creates the model choice problem as well as the further underlying question of how the observed economic heterogeneity of sizes can be explained with the candidate models. In economics, this very problem has been a longstanding source of

controversy. For example, many different competing models have been proposed over the years to describe and explain the distribution of city sizes in a country or region (e.g. Auerbach 1913, Zipf 1935, Zipf 1949), the distribution of income and wealth (e.g. Pareto 1897b, Champernowne 1953, Fisk 1961, Dagum 1977, Reed and Jorgensen 2004) and the distribution of firm sizes in an economy (e.g. Gibrat 1931, Mansfield 1962, Evans 1987). Quite recently, starting with a paper by Eeckhout (2004) that reported evidence for the lognormal distribution in the 2000 U.S. city size distribution, a rather vivid debate has been ongoing about the 'best' model for city size distributions, most prominently between proponents of the lognormal and the Pareto distributions (Lévy 2009, Eeckhout 2009, Bee, Riccaboni and Schiavo 2013, Ioannides and Skouras 2013). The economic question underlying this debate is one about the most plausible microeconomic model given the observed macroscopic outcome.

Inspired by this discussion, we propose a novel three-step 'Bayequentist' model-selection framework that aims at reconciling this debate. We argue that the very specific problem of fitting size distributions to empirical data has some peculiarities that have not been addressed so far, and that require a methodological fix. Beyond these methodological issues, we also provide a discussion of philosophical principles underlying any model selection procedure and how some of these principles are incorporated into our framework. Specifically, we first generalize a frequentist hypothesis test for probing data for the Pareto distribution (Clauset, Shalizi and Newman 2009) to accommodate any size distribution hypothesis and demonstrate that it works with suitable numerical tests. We complement this first step of our framework with the standard Bayesian method to calculate model weights based on Akaike's information criterion (AIC, Akaike 1973, Schwarz 1978). We argue that, among all statistical measures, the AIC is the preferred option, because it strikes a balance between two well-known principles in philosophy of science: the principles of parsimony and diversity. In the last step, we introduce the criterion of model microfoundation, which is, all other things being equal, to select the model that comes with an economic micromodel that explains, from the perspective of the individual constituent, the genesis of the observed overall size distribution. Finally, we illustrate our framework with size distribution data on commercial cattle farms in Namibia (Olbrich, Quaas and Baumgärtner 2012). We find that a combination of lognormal and Pareto

known as double-Pareto lognormal (Reed and Jorgensen 2004) performs best according to our 'Bayequentist' framework.

The method proposed and illustrated in the paper has at least two limitations. First of all, compared to any of the other methods used so far, it requires more preliminary work in terms of software implementation, number of steps, choice of an appropriate candidate set, and qualified judgment of results. It is certainly not a "'black box' device", where the scientist puts some data in on one side and a tailor-made result pops out on the other. While some will see this as a serious limitation, this paper was specifically written to provide an alternative to exactly these unsatisfactory 'black box' one-step methods that have created the confusing situation laid out above and in the paper. Secondly, our framework is generally very selective. That is, situations where no candidate distribution passes all tests might occur. However, we think that this reflects that fitting size-distribution models to empirical data is not an easy task to accomplish, and inference from such fits should not be done prematurely. While we would not go as far as saying that premature inference is the 'source of all evil'[3], it has caused quite some confusion for sure. Our framework might offer a remedy to such premature inference. For example, if applied to the city size data sets that are so abundant in very good quality nowadays, and which was a precondition for the debate about the 'right' distribution and the 'right' theory of city growth in the urban and regional economics community, we think that our framework might help to unify the major strands in the literature, that, at first sight, seem very hard to reconcile.

1.2 Paper 2: Farm size, environmental risk and risk preferences: the case of Namibian commercial cattle farming

The paper *Farm size, environmental risk and risk preferences: the case of Namibian commercial cattle farming* takes up the data analysis started in Paper 1. It studies the correlations of different forms of economic heterogeneity – farm size and risk preferences – and environmental risk. The commercially grazed semi-arid cattle rangelands of Namibia form a perfect subject for a study on how people practically deal with environmental risk. The main reason for this is that environmental risk, which comes in the form of inter-

[3]In programming, there is the famous bonmot that 'premature optimization is the source of all evil'.

annual variability in rainfall, is generally the farmers' main source of income risk. This income risk greatly varies with farm location and is thus spread very hetereogeneously. Of course, each farmer has his personal risk preferences, which have been elicited in the survey by Olbrich, Quaas and Baumgärtner (2012). Based on these risk preferences, we conduct statistical group comparisons with regard to the key farm parameters such as the stocking rate, the coefficient of variation of inter-annual precipitation, and regarding inequalities in farm sizes.

In agricultural economics, there are quite a few theories concerning optimal rangeland management that can be tested with our data. For example, various works suggest the stocking rate[4] (e.g., McArthur and Dillon 1971, Karp and Pope 1984, Torell, Lyon and Godfrey 1991) to play a crucial role as management parameter. Rodriguez and Taylor (1988) suggest that high stocking rates may be optimal if the farmer is risk neutral, and Quaas et al. (2007) stress the importance of land acquisition in the presence of large rainfall variability. However, empirical tests of these theories are largely missing so far, due to the fact that, for reasons of privacy protection, commercial cattle farming data is usually very hard to obtain. Apart from the description of patterns that we find in the data, another aim of the paper is thus to take a look at evidence for these theories in our sample. Lastly, we also take a look at the implicit hypothesis from urban economics that the larger economic entities – the ones in the 'Paretian tail' of the overall distribution – are, in some way, special or different from the rest. We take this hypothesis as starting point for further analysis.

The paper contains a rich variety of findings. First, using the method proposed by Clauset, Shalizi and Newman (2009), we find that the Pareto distribution is a statistically plausible description of herd size distribution, but not of stocking rate and area distributions. The group comparisons based on this fit yield that larger farms in the sample are, on average, exposed to less environmental risk than the smaller ones, and farmers of larger farms choose, on average, significantly smaller stocking rates. There is, however, no difference in average risk attitude. With regard to the group comparison based on risk preferences, we do not find differences in key farm parameters such as size characteristics and precipitation, but observe that a more risk-loving attitude comes with larger inequality in the distribution of farm sizes in the sample. Overall, our findings support the central

[4]The stocking rate is the ratio of cattle and rangeland area, its unit is usually heads per hectare.

role of the stocking rate as a crucial management parameter because we consistently find the largest correlations when stocking rates are involved. However, we do not find any evidence for the hypothesis that risk-neutral farmers actually choose higher stocking rates than others. There is also no evidence that larger environmental risk comes with larger farm area.

Our practical and data-driven approach naturally comes with a few limitations. First and foremost, participation in the survey that resulted in the database we used was voluntary. Therefore, the sample is self-selected, however, we have no reason to believe that any self-selection criteria should be related to the variables relevant to our study. Second, the nature of the data did not allow to study spatial diversification strategies or 'opportunistic grazing' strategies (Beukes, Cowling and Higgins 2002), which would have also been interesting hypothesis tests. In general, the paper illustrates that 'theoretically preferable' strategies for farm management are not necessarily the ones employed in practice. For future research, it would be essential to collect additional data sets like the one utilized in this research, to find and attempt to explain further areas of deviation between theory and practice.

1.3 Paper 3: An axiomatic foundation of preferences under Knightian uncertainty

In the paper *An axiomatic foundation of preferences under Knightian uncertainty*, we contribute to one of the most central questions of economics: what to do when confronted with a choice between fundamentally uncertain options? Here, heterogeneity manifests itself in three different ways: (1) there are at least two mutually exclusive options to act, (2) there are at least two mutually exclusive possible future states of the world, whose probabilities are unknown, and (3) decision makers have different preferences regarding choice under Knightian uncertainty. All three aspects together constitute the decision problem, while (2) can be seen as an abstract form of environmental uncertainty, as opposed to the concrete form of environmental risk seen in Paper 2.

Having its roots in early probability theory (Bernoulli 1738, Laplace 1820), choice under risk and uncertainty has been on the research agenda for a long time. While Keynes

(1921) and Knight (1921) expressed the distinction between *risk* and *uncertainty*, it was only in 1944 that John von Neumann and Oscar Morgenstern succeeded in providing an axiomatic treatment of Bernoulli's solution to the St. Petersburg Paradox[5]. von Neumann and Morgenstern's success triggered a new interest in economics regarding choice under risk and uncertainty (e.g. Savage 1954, Anscombe and Aumann 1963). Experiments suggest that people tend to prefer completeness to incompleteness of probabilistic information (Ellsberg 1961), i.e. risk to ambiguity. This finding has led to the development of the so-called 'ambiguity aversion literature' (Al-Najjar and Weinstein 2009: 249), but also to an increased messiness with respect to conceptual clarity and descriptive or normative direction of contributions (cf. e.g. Al-Najjar and Weinstein 2009, Gilboa 2010). Against this background, we go 'back to the roots' in our paper in that we do not assume any availability of probabilistic information to the decision maker. Specifically, we provide conceptual clarifications and ask which axioms are required to warrant a self-contained decision-making framework in situations of Knightian uncertainty, and how uncertainty-averse decision makers can be modeled.

The results of our paper are manifold. First, we show how the notions of uncertainty aversion and uncertainty premium can be defined within our non-expected utility setting. Based on these definitions, we establish the possibility of comparing uncertainty attitudes interpersonally. Second, we use a recent result from theoretical thermodynamics on the concept of entropy (Lieb and Yngvason 1999) to provide and discuss a set of seven axioms on preferences under Knightian uncertainty. We subsequently show how these axioms imply the existence of a function from the set of Knightian acts to the real numbers that represents uncertainty preferences. We show that this function is monotonous on the subset of acts with fixed positive sum of payoffs over all possible states of the world (referred to as fixed 'payoff volume'), additive, extensive and unique up to linear-affine transformations. Finally, we illustrate one possible such function – Rényi's generalized entropy (Renyi´1961) – with the problem of deciding between three sample Knightian acts, and compare the result to well-established approaches from the literature. In this example, the resulting ranking of acts coincides with a risk-averse expected utility maximizer with

[5]The St. Petersburg Paradox refers to the following: Suppose a fair coin is tossed k times, and you are offered 2^{k-1} monetary units the first time it shows heads. Expected payoff of this game diverges to infinity, since $\mathbb{E}(y) = \sum_{k=1}^{\infty} p(k) \cdot y(k) = \sum_{k=1}^{\infty} 2^{-k} \cdot 2^{k-1} = \infty$. However, nobody would actually pay an infinite amount of money to participate in this game, which obviously cannot be explained by the expected payoff rationale.

uniform subjective prior, while the most preferred option coincides with the maximin rule (Wald 1949) and a pessimistic Hurwicz individual (Arrow and Hurwicz 1977).

The very general approach that we pursue in this paper comes at the price of some limitations. The class of functions that can represent preferences under Knightian uncertainty in our approach are analytically not as tractable as the ones from other approaches. This might result in rather cumbersome arithmetics in actual optimization applications, which are for the moment left for future investigation. Closely related is the question about incompleteness of preferences. In our paper, this implies that acts may in general only be compared for one specific and fixed payoff volume. This is indeed rather restrictive, but it can be seen as a direct consequence of Knightian uncertainty, which is informationally much more restrictive than ambiguity. In a nutshell, it means that less knowledge about a situation implies less clear statements overall. The approach to Knightian uncertainty developed here as a number of potentially useful applications. For example, the approach could be applied to concrete economic models from the literature to see how and where it gives different results than other methods. Concretely, one might think of an extension and generalization of the concept of economic insurance value (Baumgärtner 2007, Baumgärtner and Strunz 2014) with our approach. On the behavioral side, it is possible to think of studies that investigate uncertainty attitudes experimentally in a multitude of settings. The stage for such endeavors is all set.

2 Conclusion

This thesis consists of three research papers that explore and illustrate some of the consequences of the complex and multifaceted interrelationship between heterogeneity and uncertainty in economics. The first paper, *Model choice and size distribution: a Bayequentist approach*, demonstrates how heterogeneity of theories can be dealt with in the context of size distributions in economics. It establishes a link between the concept of 'goodness-of-fit' of a theoretical size distribution model and economic models describing the size distribution constituents. For the specific case of Namibian commercial cattle farming, we learn how environmental risk might be a major factor in explaining the observed economic heterogeneity. The second paper, *Farm size, environmental risk and risk pref-*

erences: the case of Namibian commercial cattle farming explores further the aspect of individual farms in this data set and tests some theories that have been brought up in the context of farm management under environmental risk that have not been empirically tested so far. The paper provides empirical evidence for the crucial role of environmental risk in shaping economic heterogeneity. Lastly, the paper titled *An axiomatic foundation of preferences under Knightian uncertainty* leaves the realm of quantifiable risks and provides a way to model human preferences regarding choice under Knightian uncertainty, based on a set of seven axioms. It develops the notions of uncertainty aversion and uncertainty premium, so that uncertainty attitudes can be interpersonally compared.

As Baumgärtner and Quaas (2010: 449) have put it, 'any study in the field of sustainability economics has to take uncertainty seriously'. While this is most certainly true for long-term issues such as sustainability problems that involve intergenerational issues and complex human-nature relationships, the need to take uncertainty seriously applies to economics in general. The inconvenient fact about this insight is that it implies more complex theoretical methods and concepts, and it makes firm statements and general conclusions more difficult to arrive at. Both of these effects can be observed in the papers in this thesis. At the same time, the present thesis also provides methods and tools to productively wrestle with different forms of uncertainty (Papers 1 and 3), and puts theories regarding behavior under risk to empirical test (Paper 2). Such 'productive wrestling' may be cumbersome, and we may not yet fully understand everything, but this is no reason to worry. As Frank Herbert[6] once said, 'The beginning of knowledge is the discovery of something we do not understand'. And so, in light of the gradual process of science towards insight and knowledge, the papers in this thesis may help set the stage for further ambitious endeavors regarding uncertainty and heterogeneity in economics and beyond.

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

[6]Frank Patrick Herbert, 1920–1986, American writer

Al-Najjar, N.I. and J. Weinstein (2009), The ambiguity aversion literature: a critical assessment, *Economics and Philosophy* 25, 249–284

Anscombe, F.J. and R.J. Aumann (1963), A definition of subjective probability, *The Annals of Mathematical Statistics* 34, 199–205

Arrow, K. (1964), The role of securities in optimal allocation of risk-bearing, *Review of Economic Studies* 31, 91–96

Auerbach, F. (1913), Das Gesetz der Bevölkerungskonzentration, *Petermanns Geographische Mitteilungen* 49, 73–76

Baumgärtner, S., C. Becker, K. Frank, B. Müller and M.F. Quaas (2007), Relating the philosophy and practice of ecological economics. The role of concepts, models, and case studies in inter- and transdisciplinary sustainability research, *Ecological Economics* 67(3), 384–393

Baumgärtner, S. (2007), The insurance value of biodiversity in the provision of ecosystem services, *Natural Resource Modeling* 20(1), 87–127

Baumgärtner, S. and M.F. Quaas (2010), What is sustainability economics? *Ecological Econmics* 69, 445-450

Baumgärtner, S. and S. Strunz (2014), The economic insurance value of ecosystem resilience, *Ecological Economics* 101, 21–32

Bee, M., M. Riccaboni, and S. Schiavo (2013), The Size Distribution of U.S. Cities: Not Pareto, Even in the Tail, *Economics Letters* 120(2), 232–237

Bernoulli, D. (1738), Specimen theoria novae de mensura sortis, Commentarii Academiae Scientarum Imperialis Petropolitanae 5, 175–192, English translation by L. Sommer in *Econometrica* 12, 1954

Beukes, P.C., R.M. Cowling and S.I. Higgins (2002), An ecological economic simulation model of a non-selective grazing system in the Nama Karoo South Africa, *Ecological Economics* 42, 221–242

Brown, G. (2000), Renewable natural resource management and use without markets, *Journal of Economic Literature* 38(4), 875–914

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 ersonal income distribution: specification and estimation, *Economie Appliquée* 30, 413–437

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

Ellsberg, D. (1961), Risk, ambiguity, and the savage axioms, *Quarterly Journal of Economics* 75, 643–669

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

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

Gilboa, I. (2010), Questions in decision theory, *Annual Review of Economics* 2, 1–19

Gollier, C., G. Heal, R. Tol, N. Stern and M. Weitzman (2014), Understanding climate impacts and the social cost of carbon, Policy Session at the 17th World Congress of the *International Economic Association*, Dead Sea, Jordan, June 8, 2014

Intergovernmental Panel on Climate Change (2013), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, In: Summary for Policymakers

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

Karp, L. and A. Pope (1984), Range management under uncertainty, *American Journal of Argicultural Economics* 66, 437–446

Keynes, J.M. (1921), *A Treatise on Probability*, MacMillan, London

Knight, F. (1921), *Risk, Uncertainty, and Profit*, 1st edition, Houghton Mifflin Company, The Riverside Press, Cambridge

Laplace, P.S. (1820), *Theorie Analytique des Probabilitées*, Mme Ve Courcier, Paris

Lieb, E. and J. Yngvason (1999), The physics and mathematics of the second law of thermodynamics, *Physics Reports* 310, 1–96

Lévy. M. (2009), Gibrat's law for (all) cities: comment, *American Economic Review* 99(4), 1672–1675

Mansfield, E. (1962), Entry, Gibrat's law, innovation, and the growth of firms." *American Economic Review* 52(5), 1023–1051

Markowitz, H. (1952), Portfolio selection, *The Journal of Finance* 7(1), 77–91

McArthur, D. and J.L. Dillon (1971), Risk, utility and stocking rate, *Australian Journal of Agricultural Economics* 15(1), 20–35

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. (1897), *Cours d'Economie Politique*, Macmillan, London

Quaas, M.F., S. Baumgärtner, C. Becker, K. Frank and B. Müller (2007), Uncertainty and sustainability in the management of rangelands, *Ecological Economics* 62, 251–266

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

Rényi, A. (1961), On measures of entropy and information, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, 547–561

Rodriguez, A. and R.G. Taylor (1988), Stochastic modelling of short-term cattle operations, *American Journal of Agricultural Economics* 70, 121–132

Savage, L.J. (1954), *The Foundations of Statistics*, Wiley and Sons, New York, 2nd revised and enlarged edition (1972), Dover Publications, New York

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

Stern, N. (2013), The structure of economic modeling of the potential impacts of climate change: Grafting gross underestimation of risk onto already narrow science models, *Journal of Economic Literature* 51(3), 838–859

Torell, L.A., K.S. Lyon and E.B. Godfrey (1991), Long-run versus short-run planning horizons and the rangeland stocking rate decision, *American Journal of Agricultural Economics* 73(3), 795–807

von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, 3rd edition. Princeton University Press

Wald, A. (1949), Statistical decision functions, *The Annals of Mathematics and Statistics* 20(2), 165–202

Weitzman, M. (2007), A review of the Stern review on the economics of climate change, *Journal of Economic Literature* 45(3), 703–724

Zipf, G.K. (1935), *The Psycho-Biology of Language*, The M.I.T. Press, Cambridge, MA

Zipf, G.K. (1949), *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*, Addison-Wesley Press, Cambridge, MA

Paper 1

Model choice and size distribution: a Bayequentist approach

Model choice and size distribution: a Bayequentist approach*

John-Oliver Engler† and Stefan Baumgärtner

Department of Sustainability Science and 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: 12, C52, D30, D31, O44

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

*This is a pre-copy-editing, author-produced PDF of an article accepted for publication in *American Journal of Agricultural Economics* following peer review. The definitive publisher-authenticated version has digital object identifier (doi) 10.1093/ajae/aau034, and is available online at: http://ajae.oxfordjournals.org/content/97/3/978.full?etoc

†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 size distribution models in empirical data has been a topic of considerable debate in economics since Vilfredo Pareto's seminal works on the distribution of wealth in Italy (Pareto 1895; Pareto 1897). The correct identification of size distribution models has remained a contentious issue that has continued to resurface in the economic literature (Gibrat 1931; Champernowne 1953; Fisk 1961; Dagum 1977; Bandourian, McDonald, and Turley 2002; Eeckhout 2004) and is still generating lively discussions (Lévy 2009; Eeckhout 2009; Ioannides and Skouras 2013; Bee, Riccaboni, and Schiavo 2013). The relevance of this field comes from the fact that any theory of, for example, income or firm size dynamics implies a certain income or firm size distribution. It has therefore been a longstanding research interest to identify empirical evidence for these theories. 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 1895), which implies a power law distribution, and Gibrat's Law of Proportionate Effect (Gibrat 1931), which implies a lognormal distribution.

There are a number of methodological problems specific to size-distribution fitting and related model choice problems. First, as pointed out by Clauset, Shalizi, and Newman (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 normalized to 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, Shalizi, and Newman 2009; Bubeliny 2011). Third, a p-value of any frequentist hypothesis test cannot be interpreted as probability that the hypothesis actually holds true which is impractical

[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).

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, Shalizi, and Newman (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 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 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.

We illustrate our Bayequentist framework with a sample of 399 commercial cattle farms from Namibia's semi-arid rangelands (Olbrich, Quaas, and Baumgärtner 2009).

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

This sample is very well-suited for this purpose for several reasons: (1) It 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 dominates the literature. (4) As for illustration of Step 3, Namibian commercial cattle farming is a rain-fed business making the high variability in annual rain fall the farmer's main source of income risk. Hence, environmental risks and the farmer's risk preference 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.

2 Relation to the literature

One finds two types of papers concerned with size distributions in the literature. The first type of paper deals with one particular size distribution model, either presenting empirical evidence for the validity of that model, or presenting a theoretical derivation of the steady-state size distribution from plausible economical assumptions, or a mixture of both. The second type of paper focuses on comparing the goodness-of-fit of at least two - and often more - theoretical size distribution models in empirical economic data. We start by reviewing works from the former group.

The works of Vilfredo Pareto (on wealth distribution) and Robert Gibrat (on firm sizes), have subsequently been a foundation for further research. Mansfield (1962) investigated the validity of Gibrat's Law on firm size using U.S. firm size data over different time periods and concluded that Gibrat's law did overall not hold up well under empirical scrutiny. To the contrary, Hart and Oulton (1996) found Gibrat's Law to hold well for U.K. firm data which led them to conclude that the role of random stochastic events was a major factor in firm growth. Studies with U.S. manufacturing data (Evans 1987; Dunne, Roberts, and Samuelson 1988) focused on the age-size independence implied by Gibrat's Law and reported evidence that larger firms grew more slowly than smaller firms.

In contrast to the work on firm sizes, contributions from urban economics – beginning with Auerbach (1913) and Zipf (1935, 1949) – have mainly focused on overall distributions, typically of city sizes in a country or region. Ever since, there has been an implicit rivalry between the Pareto distribution, Zipf's 'rank-size rule'[4] and the lognormal distribution as predominant models of city size distributions.[5] Lately, Gabaix (1999) resurrected this long running debate with his theory of city growth behind Zipf's law. In response, Eeckhout (2004) presented his model of city growth leading to the lognormal distribution, graphically substantiating this theory with plots of model predictions against actual U.S. city size data. Lévy (2009) challenged Eeckhout's conclusions on the grounds of a graphical re-examination of Eeckhout's logarithmic rank-size plots, a Q-Q plot and a χ^2-test of the lognormal hypothesis for the largest 150 U.S. cities. He concluded that the upper tail was Pareto while the rest of the distribution was lognormal. In his response, Eeckhout (2009) maintained that his conclusions were still correct stressing again the importance of considering the whole data set rather than just the upper tail while also pointing out the problems of purely graphical reasoning in model selection. Nonetheless, it seems that the debate is still far from being resolved (see, for example, Ioannides and Skouras 2013; Bee, Riccaboni, and Schiavo 2013).

Papers proposing models specifically for income and wealth distributions for one or more countries are numerous. While Champernowne (1953) and Mandelbrot (1961) introduced a discrete Markov chain model leading to the Pareto distribution, Fisk (1961), Singh and Maddala (1976), Dagum (1977), McDonald and Xu (1995) and McDonald and Ransom (2008) all introduced models that differed from the established ones and were well received. More recent examples that look specifically at empirical evidence for one model are Chotikapanich et al. (2012), Toda (2012), and the study on world income distribution over time by Pinkovskyi and Sala-i-Martin (2009) that draws its conclusions based on fits of the lognormal to world income data sets from 1970 to 2006.

Regarding the goodness-of-fit model comparison papers, perhaps the most encompassing study of this kind for income is Bandourian, McDonald, and Turley (2002) which compares eleven distributions models. As for city size, recent model fit comparisons include Giesen, Zimmermann, and Suedekum (2010) and Giesen and Suedekum (2012) who

[4]This rule is actually a special case of the Pareto distribution.

[5]Interestingly, this is the case even though, for example, Richardson (1973) remarked that 'the three distributions are so similar that it is difficult to choose between them' (Richardson 1973: 240).

report a relative underperformance of the lognormal compared to the double Pareto lognormal (dPlN, cf. Reed and Jorgensen 2004) in fitting city size distribution data of several European countries and the U.S. Recently, González-Val et al. (2013) find that the dPlN outperforms the Fisk, q-exponential and lognormal in 87% of cases when investigating Italian city size data from 1900 to 2010 and data sets from the U.S. and Spain.

Virtually all these papers can be criticized for their lack of proper hypothesis testing and model selection, i.e. they lack a common statistical framework that does not suffer from the problems laid out in the introductory section and that is able to tell whether some theoretical distribution describes the data plausibly and which does so best. Clearly, the Lévy-Eeckhout debate boils down to a lack of such a framework. Similarly, the papers that compare several models compare only their relative goodness-of-fit, which falls short of actually investigating the much more crucial question of whether the hypotheses underlying the models are statistically reasonable. We believe that one possible source of the disagreement in the literature are the methodological issues that we have summarized in the introductory section. In this paper, we propose a remedy for these issues which could possibly unify some of the major literature strands.

3 Bayequentist model choice for size distributions

After definition of notation and introduction of the relevant size distribution models from the literature in the next section, we detail the three steps of our proposed statistical model-selection framework for size distributions.

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 seven 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 generalized beta distribution of the second kind (McDonald 1984), 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 1895), the double Pareto lognormal and the generalized double Pareto lognormal distribution (GdPlN, Reed and Wu 2008). We do not include the GdPlN in our analysis since there is no closed-form expression of its density functions which would add to computation time considerably if the ML fitting problem could be solved at all. A minor reason is that, unlike the dPlN, the GdPlN is not part of the literature discourse and therefore certainly not 'commonly found'. In figure 1, we give exemplary plots of these models while their explicit functional forms are detailed in table 1. Kleiber and Kotz (2003) show the interrelations of the distributions.

3.2 Step 1: Plausibility-of-fit

We assess the statistical *plausibility* of the fit generalizing the method proposed by Clauset, Shalizi, and Newman (2009). The intuition behind their method is as follows: any sample randomly drawn from a power-law distribution will feature deviations from a true

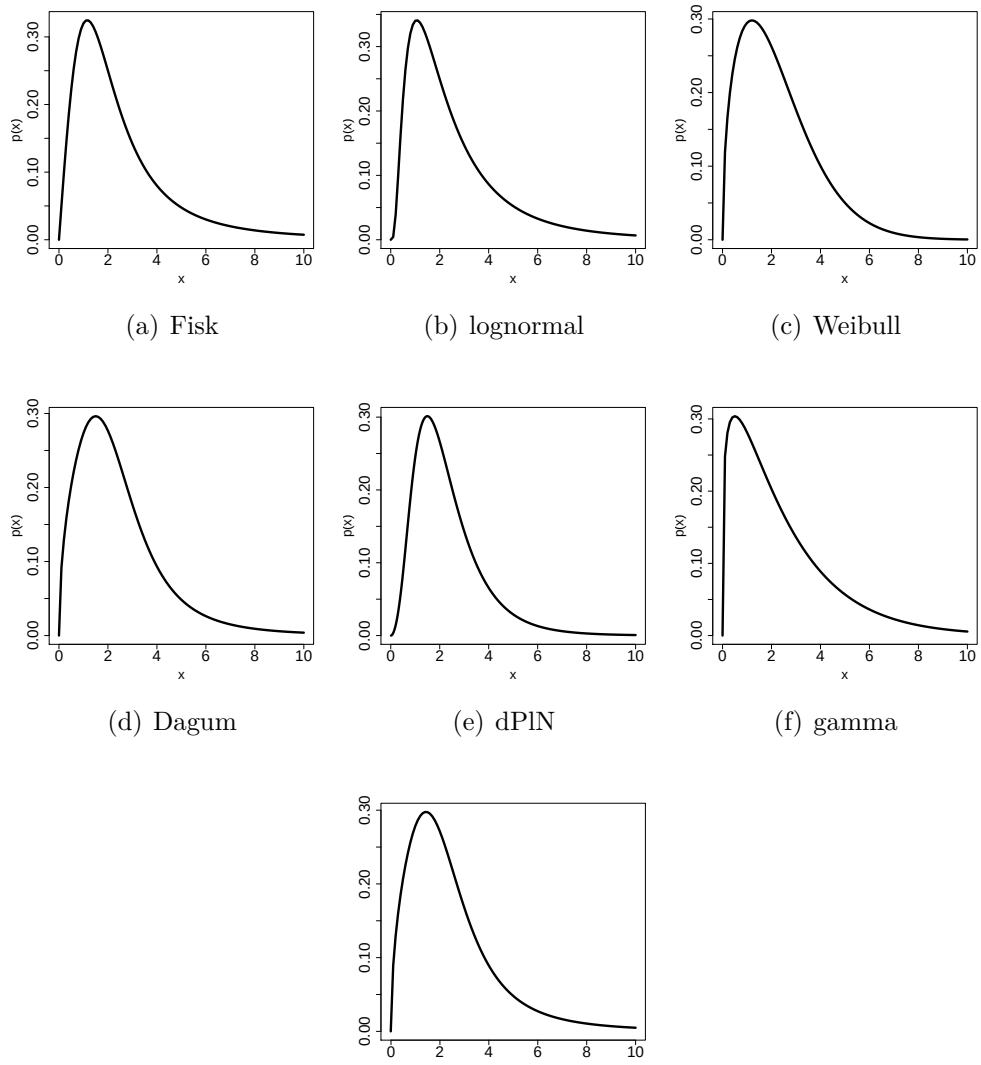
Table 1: The Size Distribution Models Most Commonly Used in Economics

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 (1895)
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	$\frac{\delta}{\gamma}(\frac{x}{\gamma})^{\delta-1}/\left[1+(\frac{x}{\gamma})^{\delta}\right]^2$	Fisk (1961)
Gamma	2	$x^{\kappa-1}\frac{\exp(-\frac{x}{\Theta})}{\Theta^{\kappa}\Gamma(\kappa)}$	Ammon (1895)
GBII	4	$\frac{ax^{ap-1}}{b^{ap}B(p,q)\left[1+(\frac{x}{b})^a\right]^{p+q}}$	McDonald (1984)

Note: Φ denotes the cumulative density function of the standard normal distribution and $\Phi^c(x)$ its complementary function $1-\Phi(x)$. $B(p,q)$ is the incomplete beta function which is defined as $B(p,q)=\int_0^x u^{p-1}(1-u)^{q-1}\mathrm{d}u$ for $0\leq x\leq 1$. For ease of reading, each parameter has been given a distinct Greek symbol.

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 these 'natural' deviations from those that make the power-law hypothesis highly unplausible.

Clauset, Shalizi, and Newman (2009) have suggested and numerically tested their procedure with the power-law distribution specifically in mind. 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:



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

(d) Dagum (e) dPlN (f) gamma

(g) GBII

Figure 1: The seven most common size distributions in economic literature

Note: Parameter values: (a) $\delta = 2$, $\gamma = 2$, (b) $\mu = 0.7$, $\sigma = 0.8$, (c) $\lambda = 1.5$, $k = 2.5$, (d) $a = 3$, $b = 3$, $p = 0.5$, (e) $\alpha = 10$, $\beta = 3$, $\nu = 1$, $\tau = 0.5$, (f) $\Theta = 2$, $\kappa = 1.5$, (g) $a = 3$, $b = 2.75$, $p = 0.5$, $q = 0.85$

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[6] the empirical data if ties are present.

[6]'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, Wyner, and Buja 2007).

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 using distribution parameter values that have been ML re-estimated for each synthetic sample.

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.

The absolute error estimate of the obtained p-value decreases with the number of synthetic data sets Z as given by

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

Our algorithm departs from the original one in stages 2 and 3. In stage 2, we have added the jittering component to avoid conservative[7] 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'. Note that it is crucial for this method to ML re-estimate in Step 3 the parameter values for each synthetic data set to avoid conservative p-values (cf. Capasso et al. 2009). It is also in principle not a problem to test for truncated data as long as one formulates the hypotheses accordingly. 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, truncated or not.[8]

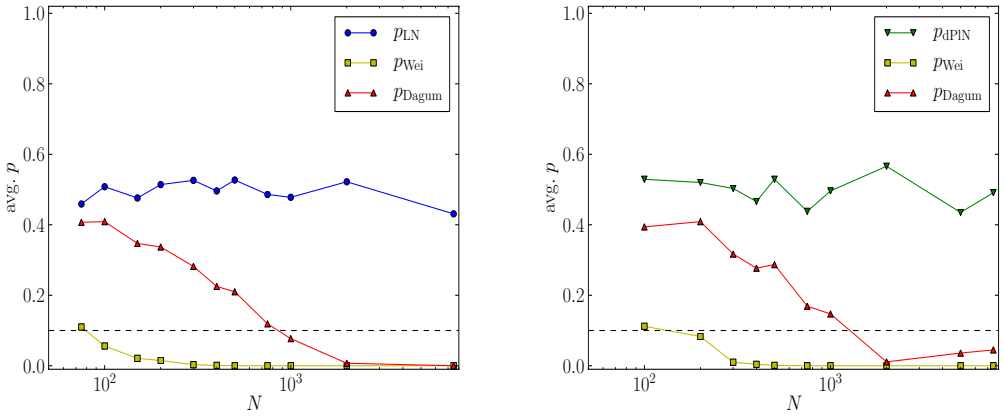
The significance level of the test should of course not be carved in stone. The threshold p-value of 0.1 is the value recommended by Clauset, Shalizi, and Newman (2009), and we take up their suggestion here. Of course, one could also think of other levels, like for example 5% which is usually the standard value in the literature. However, the nature of the test should be kept in mind: we are asking 'Given that the data actually follow distribution X, how likely is it to observe an outcome at least as extreme as observed?'.

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

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

Hence, a lower p-value threshold would actually be a less powerful criterion – and not a more strict one – when testing for H_0 and this is also the reason why we stick to $p = 0.1$ here.

For performance assessment of this generalized algorithm, we run two numerical tests: First, we draw 100 numerical random samples of constant length N from a lognormal distribution with randomly changing parameters at each draw. For each such sample, we test the following distribution hypotheses: lognormal, dPlN, Weibull and Dagum, so that we end up with four arrays of 100 p-values. For each array, we calculate the average p-value. We then repeat the same exercise for another length value of synthetic samples N. The procedure has thus a 'loop in a loop' structure where we vary N between 75 and 10'000. Second, we repeat the very same procedure for synthetic samples from dPlN distributions. We plot the results of the testing procedure in figure 2 semi-logarithmically, and we omit nested distributions.[9] The reason for this is that we cannot expect – on average – to be able to rule out such 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 hence, by construction, one could on average not expect the dPlN to be ruled out for synthetic lognormal samples.



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

Figure 2: Results of numerical tests

The results demonstrate that 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

[9] That is, we omit the curve for the dPlN hypothesis in figure 2a, and the curve for the lognormal hypothesis in figure 2b.

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 and the lower the sample size.[10] 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 – to reliably rule out incorrect alternatives.

3.3 Step 2: Model ranking

Suppose $|\mathcal{M}'| > 1$. 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:

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

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*.[11]

[10] There is of course always a possibility that we get a close resemblance to other distributions than the 'true' one by pure chance. Such issues become more likely, the smaller the sample. On the other hand, if an hypothesis is rejected in spite of a small sample, that is very valuable information, too.

[11] 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 (cf. Kaye 2007) although Ockham never actually wrote the well-known phrase 'shave away all but what is

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. 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 $j = 1 \ldots |\mathcal{M}'|$, 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.[12] 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 $l = 1 \ldots |\mathcal{M}'|$. 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. As result, we get the ordered set $\mathcal{M}''$. It contains the same elements as $\mathcal{M}'$, now ordered according to their relative weights of evidence.

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).

[12] 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.

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,[13] 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 more complicated model that lies in $\mathcal{M}'$ is not necessarily worse than a simpler model within $\mathcal{M}'$, because it could have 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. Moreover, using numerical simulations, it has been argued in the statistical literature that there are considerable performance advantages of the AIC over the BIC (Burnham and Anderson 2004; Yang 2005).

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 Steps 1 and 2) 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, in the steady state, 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

[13] $\mathrm{BIC} = -2 \ln L(\hat{\Theta}_1, \ldots, \hat{\Theta}_M) + M \ln N$

very particular understanding of microfoundation means that there exists a stochastic differential equation[14]

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

that describes the size evolution increment $\mathrm{d}X_t$ of the individual parts of the economic system in question[15] between two neighboring time instants t and $t - \mathrm{d}t$. $\mathrm{d}W_t$ is the increment of a standard Wiener process with expected value $\mathbb{E}[W_t] = 0$ and variance $\mathrm{Var}[W_t^2] = t$ 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 as the steady-state size distribution resulting from 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 while also accounting for stochasticity. 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 that generates the *macroscopic* distribution. For example, coming back to income distributions in economics, there are microscopic stochastic models such as in Champernowne (1953) or Reed (2003) and deterministic models such as in Parker (1999) and one can sensibly disagree about their merits. We should not and will not engage in this debate here since which microfoundation is 'good' and which 'bad' depends on the system one considers, on the research question and on one's preferences.[16] Therefore, this question can naturally not be resolved on this general level, it is on the researcher to make a well-informed judgment. Our point is that – given statistical significance – a size distribution model with microfoundation should be selected rather than one without.

[14] 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.

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

[16] While Dagum maintained that any good income distribution model should be stochastic (Dagum 2006), others criticized stochastic models for not being derived from basic economic principles as 'ad hoc' (Sahota 1978).

To answer the question whether a certain model not having an analytical closed-form 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[17] 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. If $\mathcal{M}'''$ is empty, the best fitting model is the one ranked highest 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. On the other hand, if $|\mathcal{M}'''| > 1$, there are two possible positions. From a strictly mechanistic viewpoint, one could 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 in $\mathcal{M}''$ conditional on the fulfillment of the microfoundation criterion. However, there are good reasons to follow Burnham and Anderson (2004) in arguing that, rather than selecting one particular model, it might be better scientific practice to keep all the models alive and engage in 'multimodel inference'.[18] We agree with them since it is clearly unrealistic that there exists one single 'true' model. Rather, there is data evidence that supports one or several models sufficiently plausible and this evidence allows for inferences.

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

[18] They write '[to select only one model] ... begins a flawed inference scenario; in particular, the implicit assumption that inference must be based on a single model is not justified by any philosophy or mathematics' (Burnham and Anderson 2004: 298).

3.5 The framework at a glance

We provide a graphical summary of our three-step Bayequentist model selection framework in figure 3. It addresses three key aspects of goodness-of-fit: (1) statistical plausibility (Step 1), (2) performance relative to the competing size distribution models (Step 2) and (3) model microfoundation allowing for inferences about the underlying process that generated the observed distribution (Step 3). Overall, frequentist Step 1 and Bayesian Step 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).

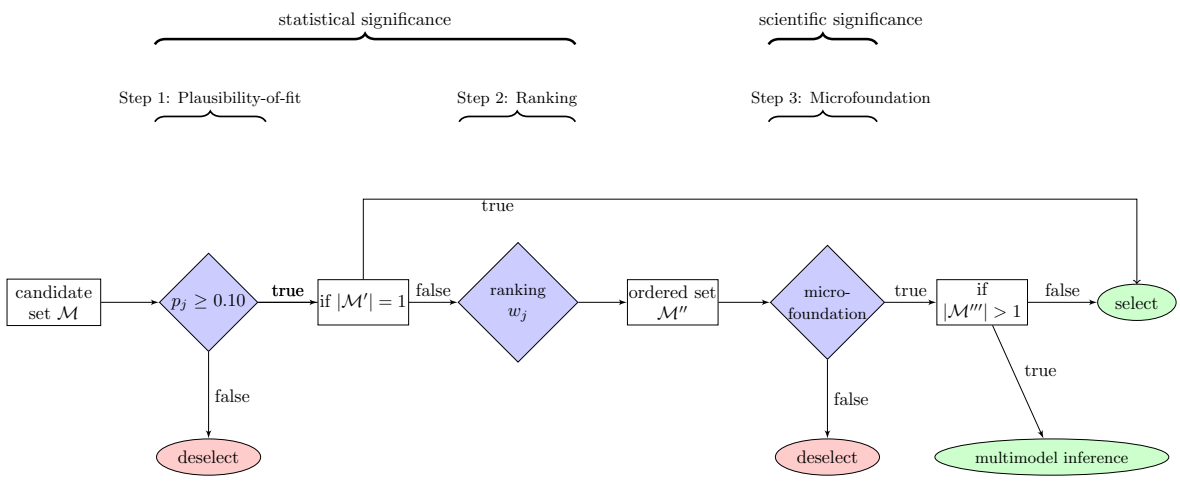


Figure 3: The proposed Bayequentist model selection framework

There are also situations where alterations of our scheme might become necessary. This may be due to 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.[19] 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 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.

[19]Such an interest may accrue from the need to (re-)evaluate effectiveness of some policy, or to assess possible extreme events based on a given data sample through extrapolation, and so forth.

As to procedural reasons, something might happen that is – for the sake of clearness – not included in figure 3 concerned with Step 3: no model may be left in the end, $|\mathcal{M}'''| = 0$. In that case, one can either resort to the ranking from Step 2 which yields at least to the best possible fit in statistical terms. It could however also be the case that the set $\mathcal{M}$ was misspecified, which means that one would need to reconsider the choice of the candidate size distribution models in $\mathcal{M}$ in general and possibly start over again.

While it may seem challengeable to place such a fundamental principle like the criterion of microfoundation last rather than first, we advocate the succession of steps proposed in figure 3, and we do so for several reasons. First, placing what is Step 3 here as first step would mean to only admit models with microfoundation into the statistical analysis. However, this would introduce a strong bias towards models with microfoundation since models without it would not be admitted to subsequent analysis. Second, it would neglect from the outset the case that the researcher is only interested in descriptive results rather than microfoundation, i.e. in the genesis of the observed outcome and its general dynamics. However, an as-good-as-possible description without microfoundation might just be what the researcher is looking for. Lastly and most importantly, we would, in general, produce a less-than-optimal result in terms of information. Consider the case that we propose here where we admit both, models with and without microfoundation, into the candidate set $\mathcal{M}$. If, at the end of the three-step procedure, it turned out that none of the models with microfoundation fits the data well, but one without does, then this finding would be a valuable one, because it would show that none of the existing theories was supported by the data at hand. This could subsequently become even more useful if this finding was consistent with other data as well. Yet, all this potential information would be discarded if Step 3 was conducted first rather than last, and there would not be any starting grounds for the germination of new theories.

4 Illustration: commercial cattle farms in Namibia

We illustrate our Bayequentist framework with size data of Namibian commercial cattle farms (Olbrich, Quaas, and Baumgärtner 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 $\mathcal{M}$.

4.1 Data

We use a 2008 data sample of 399 Namibian commercial cattle farms (Olbrich, Baumgärtner and Quaas 2012).[20] As to the operationalization of farm sizes, it contains two specifications of 'size': the number of cattle held on the farm and area in hectares. The former has been hypothesized to be a 'proxy for wealth' (Olbrich, Quaas, and Baumgärtner 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 and 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.[21] Table 2 shows descriptive statistics of the data set while figure 4 plots histograms of the cattle-number and area distributions.

4.2 Step 1: Plausibility-of-fit

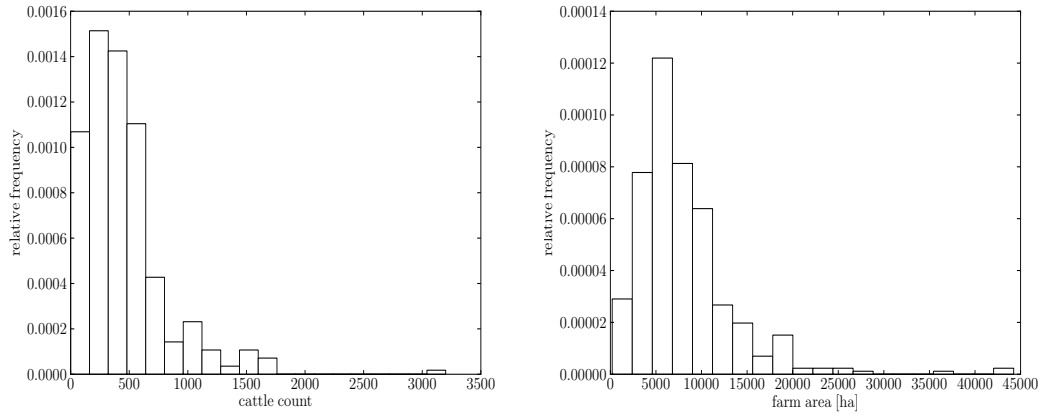
Table 3 shows the results of Step 1 of our framework including standard errors. Although the Pareto distribution does not have the same support as the other size distribution models and although it is nested within the dPlN distribution, we include it here as a robustness check for the plausibility-of-fit method. We base our estimations for each p-

[20] The data set, and its generation from a survey that consisted of a mail-in questionnaire sent to some 2100 commercial cattle farmers, is described in detail by Olbrich, Baumgärtner and Quaas (2009). They also discuss methodological issues and data quality. Based on their discussion and the comparatively high response rate of the survey, there may be a self-selection bias, but most likely no systematic truncation, and particularly no truncation in distributions.

[21] 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. For ML estimation of the dPlN, we employed the R code provided by Lu and King (2009), for the Pareto distribution, we used the original Python code of Clauset, Shalizi and Newman (available online at `http://tuvalu.santafe.edu/~aaronc/powerlaws/`.)

Table 2: Descriptive Statistics of The Namibian Commercial Cattle Farm Sample

descriptive statistics	cattle [number]	area [ha]
sample size	351	391
minimum value	1	200
maximum value	3200	42244
mean	450	7970
median	369	6800
standard deviation	361	5504
skewness	2.37	2.50
kurtosis	10.48	11.00
Gini coefficient	0.394	0.336



(a) cattle number (b) area

Figure 4: Sample farm size distributions

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 four size distribution models pass the plausibility-of-fit test, the dPlN ($p = 0.15$), the Pareto ($p = 0.33$), the Dagum ($p = 0.39$) and the GBII ($p = 0.30$) 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 resembles the tail index of the dPlN of $\alpha = 3.53$ 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[22] – ν and τ – resemble 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 3: Results of Step 1

model	cattle number parameter estimates	p	area parameter estimates	p
lognormal	$\mu = 5.80$ (0.10), $\sigma = 0.89$ (0.06)	0.00	$\mu = 8.78$ (0.08), $\sigma = 0.68$ (0.04)	0.00
dPlN	$\alpha = 3.53$ (3.46), $\beta = 1.41$ (0.18), $\nu = 6.22$ (0.11), $\tau = 0.42$ (0.10)	0.15*	$\alpha = 3.28$ (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
GBII	$a = 3.25$ (1.46), $b = 554.24$ (1.16), $p = 0.433$ (1.57), $q = 0.957$ (1.76)	0.30*	$a = 3.39$ (1.41), $b = 8303.62$ (1.11), $p = 0.95$ (1.55), $q = 0.35$ (1.66)	0.01

Note: Maximum likelihood estimates of the model parameters are given with standard errors in brackets and 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 table 1.

Thus, in this case, $\mathcal{M}'$ contains three models, namely the dPlN, the Dagum and the GBII 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 4, we list the results of Step 2 of our Bayequentist model selection framework. We document every important number in the ranking process from left to right, and for each farm size data set (cf. table 4): loglikelihoods $\mathcal{L}_j$, AIC scores and resulting Akaike weights w_j. Loglikelihoods $\mathcal{L}_j$ of Dagum, dPlN and GBII are almost identical with slight advantages for the Dagum and GBII models for the cattle data ($\mathcal{L}_{\text{dPlN}} = -2462.85$ compared to $\mathcal{L}_{\text{Dagum}} = \mathcal{L}_{\text{GBII}} = -2462.70$) and the dPlN in case of the area

[22] 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).

data ($\mathcal{L}_{\text{dPlN}} = -3817.57$ compared to $\mathcal{L}_{\text{Dagum}} = \mathcal{L}_{\text{GBII}} = -3818.04$). The AIC scores reflect this finding as these three two models rank first (1.Dagum, 2.GBII, 3.dPlN for cattle numbers and 1.Dagum, 2.dPlN, 3.GBII for area data). Using the definition of the model weights w_i (Equation 5), we find substantial evidence for all three of these models (cattle number: $w_{\text{dPlN}} = 18.3\%$, $w_{\text{Dagum}} = 57.9\%$, $w_{\text{GBII}} = 21.4\%$, area: $w_{\text{dPlN}} = 29.4\%$, $w_{\text{Dagum}} = 50.1\%$, $w_{\text{GBII}} = 18.4\%$). With a cumulative Akaike weight of about 3% for both data sets, the other size distribution models play no role. Note that the area data nicely illustrates why Step 2 alone is not sufficient: it yields a relative ranking, but without any qualification of whether the fits are statistically plausible. The quality of fit of the

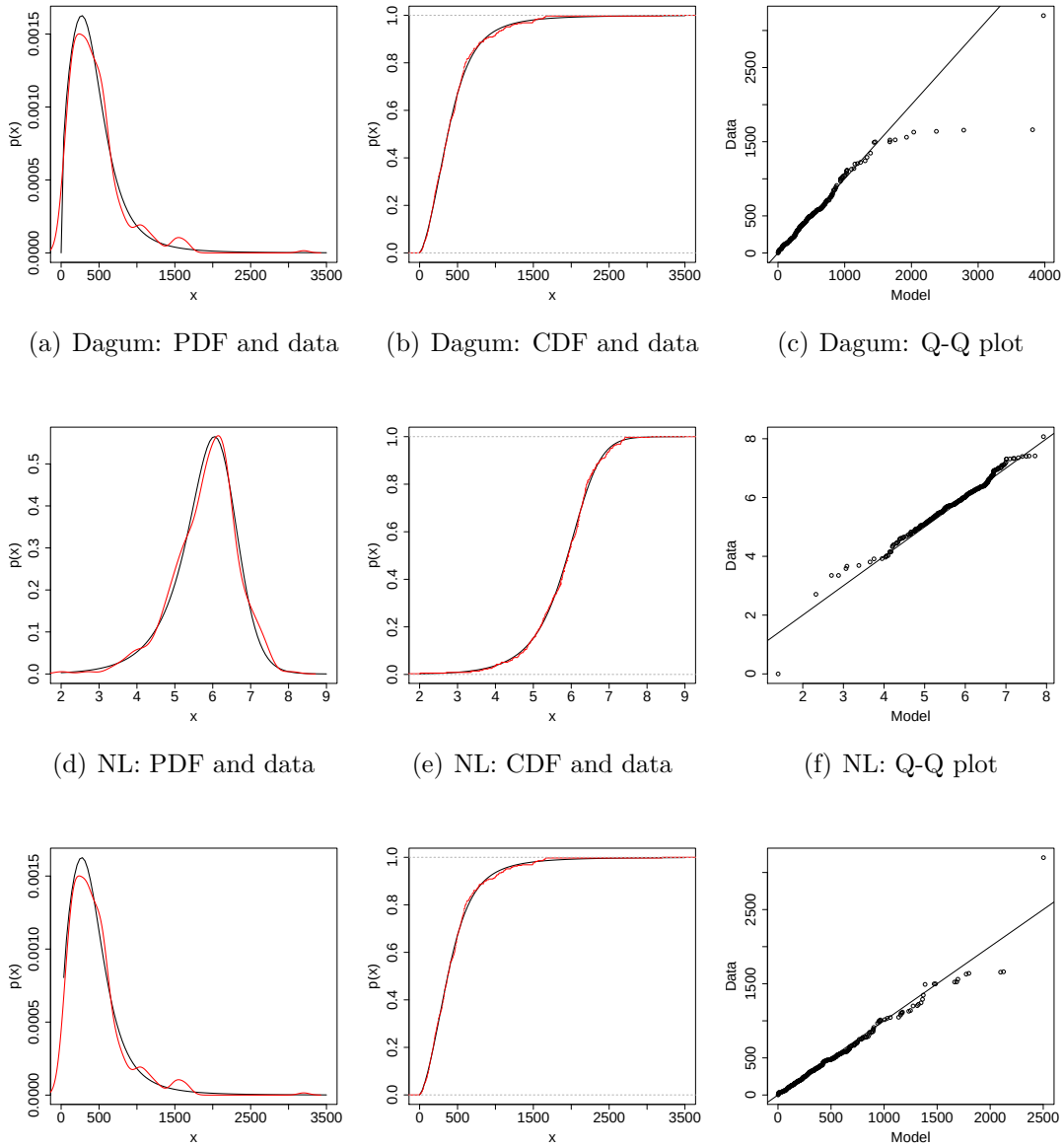
Table 4: Results of Step 2

	cattle number			area		
model	loglikelihood	AIC	weight [%]	loglikelihood	AIC	weight [%]
lognormal	-2492.21	4988.42	<0.00	-3836.59	7677.18	<0.00
dPlN	-2462.85	4933.71	18.31	-3817.57	7643.15	29.39
Weibull	-2470.72	4945.43	0.04	-3841.32	7686.63	<0.00
Dagum	-2462.70	4931.41	57.90	-3818.04	7642.08	50.13
Fisk	-2473.82	4951.64	<0.00	-3822.21	7648.41	2.12
Gamma	-2466.91	4937.83	2.33	-3827.67	7659.35	0.01
GBII	-2462.70	4933.40	21.40	-3818.04	7644.09	18.35

Note: Loglikelihoods (logarithm of Equation 1), AIC scores (Equation 3) and Akaike weights (Equation 5) are displayed. The resulting model ranking is Dagum, GBII, dPlN, Gamma, Weibull, Fisk, lognormal (cattle number) and Dagum, dPlN, GBII, Fisk, lognormal, Weibull (area).

Dagum, the dPlN and the GBII distribution to our data is illustrated in figure 5, where we use for computational convenience that the logarithm of a dPlN distributed variable is Normally Laplace (NL) distributed (cf. Reed and Jorgensen 2004).[23]

[23]Note that – exploiting Reed and Jorgensen (2004) – we plot the Normal Laplacian over $\ln x$ and not the dPlN over x in figure 5. For the sake of readability, we do not explicitly state this in the graph but we remark here that these plots feature logarithmic distortion, i.e. the data are compressed and outliers therefore seem less drastic, hence the apparent differences in fit quality (see also Eeckhout (2009) for a longer discussion of log-log plotting). Along the way, this nicely illustrates the danger of judging the statistical goodness-of-fit by the look of the plot.



(a) Dagum: PDF and data (b) Dagum: CDF and data (c) Dagum: Q-Q plot

(d) NL: PDF and data (e) NL: CDF and data (f) NL: Q-Q plot

(g) GBII: PDF and data (h) GBII: CDF and data (i) GBII: Q-Q plot

Figure 5: Best fitting theoretical distributions (black) versus kernel density regressions (red), and Q-Q plots, for the Dagum (top), NL (center) and GBII (bottom) distributions

4.4 Step 3: Microfoundation

Camillo Dagum (1977) defined the distribution of his name by an ordinary differential equation which features the income elasticity of the cumulative distribution function as parameter. The solution function of this equation is the cumulative density function of the Dagum distribution. 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

overall distribution might have been generated from an underlying individual microscopic evolutionary process. Dagum did provide Kolmogorov forward equations describing the dynamic behavior of the overall probability density function of incomes under certain assumptions about the instantaneous mean and variance at time t. However, this is not a microfoundation in the sense that we define it here and therefore we deselect it.[24]

On the other hand, there exist microfoundations for the GBII and dPlN distributions. The GBII can be shown to result from a model in which there is a representative firm that maximizes profits under certainty (Parker 1999). The firm is assumed to have a production function $F(n(y),\psi(y))$ where $n(y)$ is the number of workers earning income y and $\psi(y)$ the level of human capital available at y so that the firm's total output is $\int_0^\infty F(n(y),\psi(y))\mathrm{d}y$. Writing total labor cost at y as $c(y)$, the firm's problem is then to maximize $\int_0^\infty [F(n(y),\psi(y)) - c(y)n(y)]\,\mathrm{d}y$ with respect to $n(y)$. If one assumes a standard cost function, constant elasticity of income returns with respect to human capital and $F = \psi(y) \cdot n(y)^\alpha$ with $\alpha \in (0,1)$, the GBII distribution solves the firm's optimal control problem. As to the dPlN, Reed and Jorgensen (2004) have shown that it 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 these elements will be dPlN distributed.

Therefore, in spite of the worse relative goodness-of-fit statistics from Step 2, we select the dPlN and the GBII as the best size distribution models out of the seven candidates. In light of the research question about the influence of environmental risk on the overall distribution, Parker's (1999) model is however not very satisfying. Since variation in annual precipitation plays a key role for farms in our sample, it seems therefore unrealistic to infer from a model that does not incorporate any kind of risk or uncertainty. In the following, we will therefore concentrate on possible inferences from the dPlN microfoundation.

[24]Dagum promoted an understanding of 'model foundation' that is quite different to our understanding of model *micro*foundation: 'By *model foundation*, we understand the extent to which the mathematical form of an ID [income distribution] model is derived from realistic elementary assumptions. [...] It has *formal stochastic foundation* when the mathematical form of an ID model is the outcome of an a priori set of probability assumptions that are advanced by pure reason (rationalism) without being substantiated by factual observations.' (Dagum 2006: 3382, remark in brackets added by the authors).

The good fit of the dPlN to the farm size data 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 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 above under 'Step 3: Microfoundation'. 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. Equation 7 is equivalent to[25]

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

That is, the finding of the dPlN in our data suggests that individual farm sizes evolve according to a stochastic exponential growth process. 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, hiring or laying off staff and so forth. Consequently, the expected farm size at any positive time instant is $\mathbb{E}\left[S(t)\right] = S_0 \mathrm{e}^{\mu t}$. 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 (Olbrich, Quaas, and Baumgärtner 2009). Other sources of randomness mainly include (Olbrich 2012: 23–29): diseases, cattle theft, bush fires,

[25] 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.

wildlife predation, price volatilities and labor legislation. The farmer cannot influence any of these risks, they are thus externally given.

Does it matter, one might object, theoretically or practically for this explanation that cattle inventories are integer figures whereas land area holdings are real numbers? Although biological dynamics and cattle fecundity do play a role for the farmer in Namibian commercial cattle farming since cattle reproduces on farm (cf. Olbrich 2012), this fact does not imply that changes in cattle numbers on a particular farm are smooth. First and foremost, it is common practice for a farmer to buy and sell cattle at one of the numerous auctions, 'hundreds of which take place year around at various locations across Namibia' (Olbrich 2012: 21). Furthermore, there are several serious risks that also foster the possibility of sudden lumpy changes in cattle numbers that have been given in the previous paragraph. On a more theoretical remark, we have Donsker's Invariance Principle (Donsker 1951). It establishes that any random walk in discrete time converges under certain conditions to a standard Wiener process, which is what we use in our stochastic model of farm growth. This result means that any random walk process - which is integer in both, domain and codomain - converges to a standard Wiener process - which is not integer but real in domain and codomain. Hence, one can in general always rescale a random walk so as to converge to a Wiener process and this is why stochastic differential equation models such as Equation 8 are valid.

Overall, our findings 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 8) is a Markov chain. If a policy was concerned with optimally fast job generation in Namibia's large agricultural sector, a recommendation to focus on certain subgroups of farmers would not follow from our analysis.

5 Discussion and conclusion

We critically reflect on limitations of, and possible objections to, our framework before we conclude.

Robustness. In our paper, we understand 'robustness' in the context of errors of first and second kind. 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, we do not find any support for a greater error of the first kind with larger samples (figure 2). On the other hand, with small or very small samples, there is a natural tendency towards making an error of the second kind (i.e. erroneous acceptance of H_0) with any frequentist test that we know of, and the one proposed here is no different. In such an instance, the Bayesian Step 2 of our framework still allows for valid relative model comparison, since a Bayesian test compares relative goodness-of-fit of competing models irrespective of sample size (Kass and Raftery 1995; Burnham and Anderson 2004). With small samples, one might thus not be able to rule out many models in Step 1 but one still gets a relative ranking of hypotheses in Step 2. There is hence no reason to assume that sample size might systematically alter the results obtained with our framework.

Justification of Step 3. Step 3 could obviously be challenged for being a somewhat vague criterion in an otherwise quite specific framework. Yet, we would rather see this as an asset than as a shortcoming. First and foremost, it incorporates the notion of 'scientific significance' (cf. 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. Furthermore, Step 3 leaves the possibility of conducting a multimodel inference, as advocated by Burnham and Anderson (2004).

We have stated above 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[26] and classical mechanics in physics. The former is far more

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

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 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.

Conclusion. In this paper, we have proposed and illustrated a new statistical framework for identifying theoretical size distribution models in empirical data. Its main innovation is the three-step combination of frequentist and Bayesian statistical methods (Steps 1 and 2) with the criterion of microfoundation (Step 3). We have generalized the frequentist test for Pareto's rank-size rule by Clauset, Shalizi, and Newman (2009) to arbitrary distributions and numerically demonstrated its functioning. For direct comparison of competing size distribution hypothesis, we have combined this in Step 2 with the Bayesian method of calculating Akaike weights that can be interpreted as relative model probabilities given the data. Lastly, we have proposed to include Step 3 which demands to also take into account possible inferences from the overall size distribution to individual dynamics. Altogether, we have argued that our framework captures, in its three steps, the following three key aspects of goodness-of-fit: statistical plausibility compared to mere chance (Step 1), relative outperformance of other size distribution models in the candidate set (Step 2) and possibility for inferences about the underlying process that generated the observed outcome (Step 3).

We have illustrated our framework analyzing a unique data set of 399 Namibian commercial cattle farms. We have found that the Dagum, dPlN and GBII distributions

fit the data best in terms of statistical plausibility and relative goodness-of-fit. Yet, we have selected the dPlN model because of its superiority in explanatory depth, which is – to the best of our knowledge – the first finding of the dPlN in firm data.[27] 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 exogenous risk is a key driver for farm size growth rather than just a minor parameter.

Acknowledgments

We thank Dave Abson, Carlo Andrea Bollino, Carsten Dormann, Nikolai Hoberg, Roland Olbrich, Martin F. Quaas, Boris Schröder, and two anonymous reviewers for critical comments and discussion.

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*. Jena: G. Fischer.

Auerbach, F. 1913. "Das Gesetz der Bevölkerungskonzentration". *Petermanns Geographische Mitteilungen* 49:73–76.

[27] 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 (Fisk) 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 the respective Kolmogorov-Smirnov and Cramér-von Mises p-values as well as AIC and BIC scores. Hence, they choose an approach similar to ours and arrive at a similar conclusion. However, we think that the irrefutable problem with using 'raw' Kolmogorov-Smirnov (or Cramér-von Mises) p-values is that they are conservative, for the reason laid out by Capasso et al. (2009). To illustrate this, let us consider an example: while the Fisk distribution is rejected for the cattle data by our procedure ($p = 0.04$, table 3), it is clearly not by the KS test ($p_{\mathrm{KS}} = 0.51$) or the Cramér-von Mises test ($p_{\mathrm{CM}} = 0.19$). Similar results hold for the area data ($p = 0.02$, table 3 compared to $p_{\mathrm{KS}} = 0.42$ and $p_{\mathrm{CM}} = 0.15$).

Baker, A. 2011. "Simplicity". In: Zalta, E.N. (ed.), *The Stanford Enyclopedia 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* 29:1–30.

Bee, M., M. Riccaboni, and S. Schiavo. 2013. "The Size Distribution of U.S. Cities: Not Pareto, Even in the Tail". *Economics Letters* 120(2):232–237.

Bubeliny, P. 2011. "Some Notes on Biasedness and Unbiasedness of Two-Sample Kolmogorov-Smirnov Test". Unpublished, Dept. of Mathematics and Physics, Charles University, Prague. 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.

Capasso, M., L. Alessi, M. Barigozzi, and G. Fagiolo. 2009. "On Approximating the Distributions of Goodness-of-Fit Test Statistics Based on the Empirical Distribution Function: the Case of Unknown Parameters". *Advances in Complex Systems* 12:157–167.

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. New York: John Wiley and Sons.

Dagum, C. 2006. "Income Distribution Models". In: N. Balakrishnan, C.B. Read and B. Vidakovic (eds.), *Encyclopedia of Statistical Sciences*. 2nd Edition. Vol. 5. pp. 3378–3386.

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

Donsker, M.D. 1951. "An Invariant Principle for Certain Probability Limit Problems." *Memoirs of the American Mathematical Society* 6:1–10.

Dunne, T., M.J. Roberts, and L. Samuelson. 1988. "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries." *Rand Journal of Economics* 19(4):495–515.

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

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*. Paris: Librairie du Recueil Sirey.

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. Forthcoming. "Size Distributions for All Cities: Which One is Best?" *Papers in Regional Science*, in press.

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

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.

Ioannides, Y., and S. Skouras. 2013. "U.S. City Size Distribution: Robustly Pareto, But Only in the Tail." *Journal of Urban Economics* 73:18–29.

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 – 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.* New Jersey: John Wiley & Sons.

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

Leibniz, G.W. 1968 [1710]. *Die Theodizee.* Hamburg: Felix Meiner.

Lu, C., and D.R. King. 2009. "An Investigation into the Population Abundance Distribution of mRNAs, Proteins, and Metabolites in Biological Systems." *Bioinformatics* 25(16):2020–2027.

Lévy. M. 2009. "Gibrat's Law for (All) Cities: Comment." *American Economic Review* 99(4):1672–1675.

Mandelbrot, B. 1961. "Stable Paretian Random Functions and the Multiplicative Variation of Income." *Econometrica* 29:517–543.

Mansfield, E. 1962. "Entry, Gibrat's Law, Innovation, and the Growth of Firms." *American Economic Review* 52(5):1023–1051.

McCloskey, D. 1995. "The Insignificance of Statistical Significance." *Scientific American*, April, pp. 32–33.

McDonald, J.B. 1984. "Some Generalized Functions for the Size Distribution of Income." *Econometrica* 52:647–663.

McDonald, J.B., and M. Ransom. 2008. "The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality." In: C. Chotikapanich, ed. *Modeling Income Distributions and Lorenz Curves*, pp. 147–166.

McDonald, J.B., and Y.J. Xu. 1995. "A Generalization of the Beta Distribution with Applications." *Journal of Econometrics* 66:133–152.

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* 8:409–439.

Neyman, J., and E. Pearson. 1933. "On the Problem of the Most Efficient Test 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.

Olbrich, R. 2012. *Environmental Risk and Sustainability: The Case of Commercial Livestock Farming in Semi-Arid Rangelands.* Ph.D. dissertation, Leuphana University of Lüneburg. Available online at http://opus.uni-lueneburg.de/opus/volltexte/2012/14208/.

Pareto, V. 1895. "La Legge della Domanda." *Giornale degli Economisti* 10:59–68, English translation in *Rivista di Politica Economica 87*. 1997.

Pareto, V. 1897. *Cours d'Economie Politique*. London: Macmillan.

Parker, S.C. 1999. "The Generalized Beta as a Model for the Distribution of Earnings." *Economics Letters* 62:197–200.

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* No. 15433.

Raftery, A. 1986. "Choosing Models for Cross-Classifications." *American Sociological Review* 51(1):145–146.

Reed, W.J. 2003. "The Pareto Law of Incomes – an Explanation and an Extension." *Physica A: Statistical Mechanics and its Applications* 319:469–486.

Reed, W.J., and M. Jorgensen 2004. "The Double Pareto-Lognormal Distribution – A New Parametric Model for Size Distributions." *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* 211–223.

Richardson, H.W. 1973. "Theory of distribution of city sizes: Review and prospects." *Regional Studies* 7(3):239–251.

Sahota, G. 1978. "Theories of Personal Income Distribution: A Survey." *Journal of Economic Literature* 16:1–55.

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

Singh, S.K., and G.S. Maddala. 1976. "A Function for Size Distribution of Incomes." *Econometrica* 44:963–970.

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 and Organization* 84(1):364–381.

Vogel Carey, T. 2010. "Parsimony (in as Few Words as Possible)." *Philosophy Now (UK)* 81, January/February. Retrieved on 10/27/2012 from http://philosophynow.org/issues/81/.

von Neuman, J. 1951. "Various Techniques Used in Connection with Random Digits. Monte Carlo Methods." *National Bureau of Standards* 12:36–38.

Yang, Y. 2005. "Can the Strengths of AIC and BIC Be Shared?" *Biometrika* 92:937–950.

Zipf, G.K. 1935. *The Psycho-Biology of Language.* Cambridge, MA: The M.I.T. Press.

Zipf, G.K. 1949. *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology.* Cambridge, MA: Addison-Wesley Press.

Paper 2

Farm size, environmental risk and risk preferences: the case of Namibian commercial cattle farming

Farm size, environmental risk and risk preferences: the case of Namibian commercial cattle farming

John-Oliver Engler[a,c*], Henrik von Wehrden[c], and Stefan Baumgärtner[a,b]

[a] Department of Sustainability Science and Department of Economics, Leuphana University of Lüneburg, Germany

[b] Department of Environment and Natural Resources, University of Freiburg, Germany

[c] Department of Sustainability Science, Center for Methods, Leuphana University of Lüneburg, Germany

Abstract: Utilizing a data set of 399 Namibian commercial cattle farmers, we test several hypotheses from the literature regarding self-selection according to risk preferences and optimal farm management under environmental risk. We focus on the relations between inter-annual variability in rainfall (environmental risk), risk preferences, farm size and stocking rate. We demonstrate that the Pareto distribution – which separates the distribution into two parts – is a statistically plausible description of the empirical farm size distribution when 'farm size' is operationalized by herd size, but not by rangeland area. A statistical group comparison based on the two parts of the Pareto distribution shows that large farms are on average exposed to significantly lower environmental risk. Regarding risk preferences, we do not find any significant differences in mean risk attitude between the two branches. Our analysis confirms the central role of the stocking rate as farm management parameter, and shows that environmental risk and the farmer's gender are key variables in explaining stocking rates in our data.

JEL-Classification: D22, Q12, Q56, R11, R12

Keywords: risk preferences, environmental risk, semi-arid rangelands, cattle farming, stocking rate, farm size, Pareto distribution, range management

*Corresponding author: Center for Methods, Leuphana University of Lüneburg, Scharnhorststr. 1, D-21335 Lüneburg, Germany, phone: +49 4131 677-2715, email: engler@leuphana.de.

1 Introduction

One third of our planet's land surface is covered by semi-arid regions. Central climatic characteristic of these regions is low precipitation combined with typically high inter-annual variability in rainfall. The predominant land use in semi-arid regions is livestock farming, which often is the only economically sensible use for these areas (Quaas et al. 2007), and therefore often provides the only livelihood for the local populations. What is more, significant parts of the world's commercial livestock farming also take place in these regions (Millennium Ecosystem Assessment 2005). Grazed semi-arid rangelands are tightly coupled ecological-economic systems (Perrings and Walker 1997, Janssen, Anderies and Walker 2004, Olbrich, Quaas and Baumgärtner 2009) as precipitation levels directly influence the farmer's income, because they determine the amount of forage available for livestock farming. Consequently, precipitation variation may be seen as main income risk to the farmer (Rodriguez and Taylor 1988, Quaas et al. 2007, Olbrich, Quaas and Baumgärtner 2009). One can thus treat the coefficient of variation[1] (C_v) of inter-annual precipitation at a specific farm location as a proxy for environmental risk at that location, which also links to rangeland health (von Wehrden et al 2012). Other factors, such as long-term variability trends (Miehe et al. 2010) may add complexity to the pattern, yet our understanding of these dynamics is smaller compared to the studies examining inter-annual precipitation variability. Variability of precipitation during the growing season might lead to more realistic patterns of precipitation variability (von Wehrden et al J Arid Enviro), yet no coherent data set on these dynamics exists to date.

Precipitation data is often available from official local and national meteorological offices and where it is not, there are reliable methods to generate such data based on calibrated regional climate models (e.g. Jacob and Podzun 1997). In contrast, it is considerably harder to obtain high resolution commercial livestock farming data. If data policy allows at all, national statistics offices can usually only provide classified and anonymized data which does not allow for any detailed analysis of the relationship of farm sizes and environmental risk. Consequently, empirical studies on the role of environmental risk in commercial livestock farming in semi-arid regions are, to the best of our knowledge, largely

[1]The coefficient of variation of a series of numbers is defined as ratio of standard deviation σ and mean μ of the series: $C_v = \frac{\sigma}{\mu}$.

missing from the literature. Moreover, because of this lack of empirical data, there are plenty of theoretical studies regarding optimal farm management (see Section 3.2), but largely no empirical tests. Since rangeland management is highly relevant for rangeland health as well (Quaas et al. 2007, Miehe et al 2010), our analysis also aims at informing long-term sustainable management.

The major contribution of the present paper is threefold. One, we provide a detailed empirical analysis of the interrelationships of farm size, stocking rates, environmental risk and the farmer's risk preferences using a unique and highly-detailed 2008 data set of 399 Namibian commercial cattle farmers (Olbrich, Quaas and Baumgärtner 2009, 2012). Commercial cattle farming in Namibia takes place almost exclusively in its semi-arid rangelands (Olbrich 2012), and these are among the ones with the most inter-annually-variable rainfall conditions worldwide (Figure 1). This first part of our contribution is about exploring the following general key questions: (1) What is the correlation between farm size and environmental risk, (2) Are there significant differences between different subgroups of farmers (risk-loving compared to risk-averse and 'small' compared to 'large') regarding these correlations, and (3) Are there significant differences in mean values of key variables when comparing these subgroups? These questions are inspired by long-standing discussions in both, economics and ecology, about the role of environmental risk in land management (McArthur and Dillon 1971, Torell, Lyon and Godfrey 1991, Illius and O'Connor 1999, Vetter 2005, Quaas et al. 2007). Two, we put some of the theories concerning optimal farm management that have developed over the years to an empirical test. And three, we provide and test a wide array of econometric models aimed at explaining the role of various personal, legal and environmental characteristics of farm and farmer for the choice of the actual stocking rate.

Methodologically, we use an innovative twist to distinguish between 'small' and 'large' farms: we employ fits of the Pareto distribution (Pareto 1895). The Pareto distribution is widely used for description of wealth and income in economics (cf. Engler and Baumgärtner 2015), and a wide variety of natural phenomena (Sornette 2003, Clauset, Shalizi and Newman 2009). For our purposes, it seems like a natural approach for the following reasons: (1) It has economic microfoundation (Champernowne 1953, Mandelbrot 1961) in a sense that these models treat the evolution of wealth over time as a stochastic process, i.e. it has a random component. Because of the characteristics of the ecological-

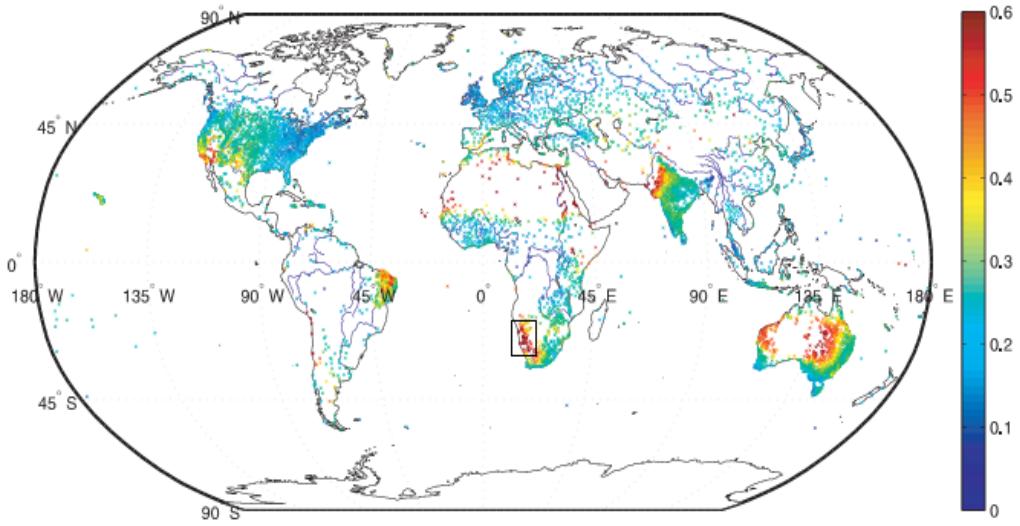


Figure 1: The global map of inter-annual variation of precipitation (Namibia highlighted), re-printed from Fatichi, Ivanov and Caporali (2012).

economic system laid out above and the hypothesis that farm size is a proxy for wealth (Olbrich, Quaas and Baumgärtner 2009), this formal incorporation of randomness is a very good fit; (2) It contains a parameter, the cutoff value $x_{\min}$, that yields a non-random division of the data set into two parts. It is a longstanding and often implicit hypothesis in the literature that economic entities described by the Pareto distribution (i.e. $x \geq x_{\min}$) are different from the rest, or special in some way (e.g. Auerbach 1913, Champernowne 1953, Mandelbrot 1961). Another question here is thus whether we find any evidence for this hypothesis in our sample by joining results from the distribution analysis with the continuous environmental and economic variables (see research questions (2) and (3) above).

Our paper is organized as follows. In Section 2, we provide a condensed overview on commercial cattle farming in Namibia. Section 3 describes the data set and its key variables as well as a short description of the Pareto distribution. In Section 4, we present our results before we discuss them in Section 5. Section 6 conludes and gives a brief outlook.

2 Namibian commercial cattle farming

At the beginning of the South African administration in Namibia (1920–1990), the commercial cattle farming system was put in place that has basically persisted until today

(Mendelsohn 2006). In the following condensed description of the commercial cattle farming sector in Namibia, we follow Olbrich (2012). Commercial cattle farming contributes 37% to Namibia's agricultural output and approximately 1–2% to its GDP. Consequently, farmlands cover close to one fifth of the country's surface and cattle is farmed extensively. There are an estimated 2'500 commercial cattle farmers that typically run their farm in one of the following three production systems: (1) rearing of calves resulting from on-farm reproduction up to the age of eight months before selling them as so-called 'weaners' at auctions; (2) Further rearing of weaners to ages between 18 and 24 months with subsequent sale to the slaughterhouse (as oxen) and (3) buying weaners at the age of eight months and raising them for about 10 to 16 months before selling them to the slaughterhouse. Auctions take place frequently all over Namibia. Although, of course, output prices may vary over time, this risk is spread homogeneously over the farmers since there are practically monopsonies in the Namibian market: almost all oxen are purchased by MeatCo of Namibia and almost all weaners sold at auctions go to a very limited number of South African buyers which are feedlot corporations.

In extensive cattle farming, the sheer area of farmland is absolutely crucial. Because Namibia's commercial cattle farming takes place in its semi-arid regions almost exclusively, precipitation risk (i.e. the inter-annual variation of precipitation), which is effectively a 'grass production risk' (cf. Obrich 2012: 23), is by far the most prominent income risk to the farmer (Quaas et al. 2007). Farming practices such as large-scale irrigation of rangelands or chemical fertilizing do not play a role in Namibia, in sharp contrast to commercial cattle production systems in the U.S. or Australia, which feature otherwise very similar climatic characteristics. Moreover, due to the lack of sufficient data, the precipitation risk is moreover not financially insurable and thus, farmers have to manage this risk through means other than financial insurance. On-farm risk management includes – but is not limited to – spatial diversification of farmland, choice of stocking rate and production system and herd organization (cf. Obrich 2012: 30ff). For example, it is standard farming practice to divide the area of farmland into small paddocks that are grazed for short periods (10 to 14 days) and subsequently rested for a minimum of two months (rotational grazing). One focus of this paper will thus be to investigate the interrelation of stocking rates and environmental risk since the choice of stocking rate has been established as the farmer's central element of risk management in the literature

(McArthur and Dillon 1971, Karp and Pope 1984, Rodriguez and Taylor 1988, Torell, Lyon and Godfrey 1991 and Quaas et al. 2007).

In summary, Namibia features a few peculiarities that make it unique compared to other extensive commercial cattle farming regions, such as, for example, some regions in Australia and the United States. For example, irrigation or chemical fertilizing is absent, which makes production technology far less important than in other commercial cattle farming regions. Lastly, outlet markets are practically monopsonies, so the usual assumption of perfect competition does not hold.

3 Methods

We review the methods used in this paper. We start by introducing the data set and its peculiarities in Section 3.1. Section 3.2 lists the hypotheses we extracted from the literature for a test against our data. In Section 3.3, we briefly review the Pareto distribution and the intricacies involved when fitting it to empirical data, before we explain the setup of our regression analysis in Section 3.4.

3.1 Data

We use a unique and highly detailed data set of 399 Namibian commercial cattle farmers that have been surveyed in 2008 and 2009 by a mail-in questionnaire and field experiments. Approximately 19% of the 840,000 (cf. Olbrich, Quaas and Baumgärtner 2012) heads of cattle commercially farmed in Namibia belonged to farmers that participated in our survey. In terms of rangeland area, our survey covers 21.5% of all rangeland that is officially designated as commercial cattle farming region.[2] The details of the complete survey and its methodology, including the data acquisition process and limitations, can be found in Olbrich, Quaas and Baumgärtner (2012). A complete copy of the survey questionnaire can be found in Olbrich, Quaas and Baumgärtner (2009). In the following, we give a brief overview of those variables of the survey relevant to the present paper.

[2]These figures are based on numbers given in Olbrich (2011)

Farm size is operationalized in two ways, number of cattle held on the farm and area in hectares. Each cattle number record consists of two numbers, one for the number of cattle at the beginning of dry season (April/May) and one for cattle number at the beginning of wet season (November). We refer to the average value of these two numbers when we speak of 'herd size' or 'cattle number' which serves the purpose of correcting for possible seasonal effects. In addition, we constructed the stocking rates from the so-obtained herd sizes and rangeland areas as stated in the questionnaire answers by the farmers.

In Namibia's semi-arid regions, meteorological stations are rare at best and records often have considerable gaps (Olbrich 2012). Therefore, the data set uses simulated precipitation data from the calibrated REMO model[3] (Jacob 1997, Jacob and Podzun 2001). These simulated data contain rainy-season (November–April) precipitation in millimeters and its standard deviation at the farm location as 30-year average from 1978–2008. We take the coefficient of (inter-annual) variation following from these data as a measure for the environmental risk unique to each farm. We illustrate the spatial distribution of the coefficient of inter-annual variation in precipitation in Figure 2.

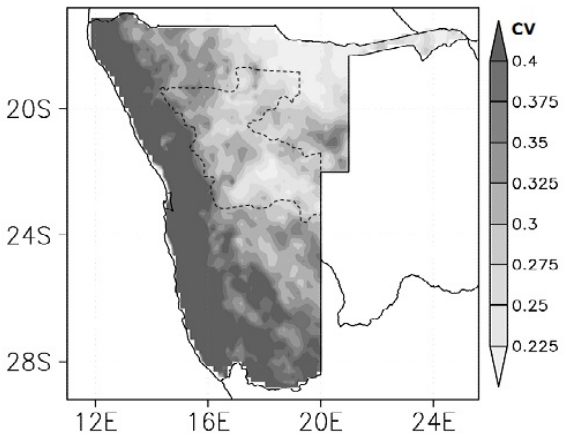


Figure 2: Spatial distribution of the inter-annual coefficient of variation of total rainy season precipitation for the period 1978–2008 and Namibia's main commercial cattle farming area (dashed line), re-printed from Olbrich (2012).

Risk preferences were elicited using the well-established adapted price list format (cf. Binswanger 1980, Holt and Laury 2002, Andersen et al. 2006). Concretely, Olbrich, Quaas and Baumgärtner (2009) offered each farmer the hypothetical choice between participating in a cattle auction with uncertain payoffs and selling a certain number (50 weaners) of their cattle to a trader who offered secure payments in each round. The auction scenario

[3]This is a model developed particularly for simulation of regional climates.

was to sell all of the 50 weaners for either N$ 90'000 (1'800 N$ per head) or N$ 130'000 (2'600 N$ per head) with equal probabilities $p_1 = p_2 = 0.5$ each, so that expected payoff was N$ 110'000 (2'200 N$ per head).[4] Conversely, the trader scenario consisted of six secure offers for the cattle that started at N$ 100'000 and increased in steps of N$ 2'500 to bring the last offer to N$ 112'500. The actual item from the questionnaire that the farmers had to consider can be seen in Figure 3. Essentially, the setup means that the 'switchpoint' at which each farmer switched from selling at the auction to selling to the trader characterizes his risk attitude: the later he switches from the uncertain auction to the trader, the larger his certainty equivalent given the lottery, and the more risk loving he is. In this paper, we use the following conversion from switchpoint to risk attitude: 1 through 4 – risk-averse, 5 – risk-neutral, 6 and 7 – risk-loving, where a switchpoint of 7 means that the trader was never chosen and hence the auction always preferred. In other words, a switchpoint from one to four means a positive risk premium, at five, the risk premium is zero, and six or seven (no switchpoint) means a negative risk premium.

Scenario	Auction	Trader
1: The trader offers you **N$ 100 000**. What would you prefer?	Sell at auction ☐	or sell to trader ☐
2: The trader offers you **N$ 102 500**. What would you prefer?	Sell at auction ☐	or sell to trader ☐
3: The trader offers you **N$ 105 000**. What would you prefer?	Sell at auction ☐	or sell to trader ☐
4: The trader offers you **N$ 107 500**. What would you prefer?	Sell at auction ☐	or sell to trader ☐
5: The trader offers you **N$ 110 000**. What would you prefer?	Sell at auction ☐	or sell to trader ☐
6: The trader offers you **N$ 112 500**. What would you prefer?	Sell at auction ☐	or sell to trader ☐

Figure 3: The six scenarios of the lottery choice experiment. Re-printed from Olbrich, Quaas and Baumgärtner (2012)

Not every data record is complete in every variable. Depending on the research question, we have the following population sizes: In Section 4.2, we have $N = 391$ for the fit of the Pareto distribution to the area data, $N = 351$ for the fit of the cattle data and $N = 347$ for the fit of the stocking rates. In Section 4.3, we base our calculations on 258 records containing complete data for cattle count as well as environmental risk and on $N = 156$ whereever switchpoints are involved. Finally, we have $N = 244$ in Section 4.4.

Despite the very good quality and unique contents, a few facts limit the extent of the analysis presented here. Because participation in the survey was on a voluntary basis, our

[4] On 2008 average, N$ 1'000 corresponded to US$ 121.07 (World Bank 2013), and the average per kilogram price for weaners was 11.88 N$ (The Namibian 2010). Hence, a 180 kg weaner could roughly fetch 2'140 N$.

sample is self-selected. However, we have no reason to believe that criteria for self-selection should be related to the variables used in this analysis. Another limitation is caused by the nature of our precipitation data. Precipitation data is based on 30-year period of precipitation in the rainy season from 1978–2008 while cattle number, rangeland areas and therefore stocking rates are a snap-shot of basically two points in time (November 2007 and April 2008). This naturally limits the potential in the data to investigate whether farmers use the 'opportunistic grazing' strategy as recommended by Beukes, Cowling and Higgins (2002), because one would need data on actual precipitation in the rainy season 2007/08. Lastly, we add that cattle and rangeland figures in the data set are aggregates, possibly over multiple different locations (cf. Olbrich, Quaas and Baumgärtner 2009: 29).

3.2 Hypotheses from the literature

We present the hypotheses that we extracted from the relevant literature on farm management from agricultural and ecological economics in Table 1. In Section 5, we will discuss these hypotheses in light of our data.

Table 1: Hypotheses regarding optimal farm management from the literature.

source	hypothesis
McArtur and Dillon (1971)	The crucial parameter for farm management is the stocking rate.
Rodriguez and Taylor (1988)	Risk-neutrality implies optimality of higher stocking rates.
Olbrich (2012)	Risk-averse farmers self-select by operating farms with lower environmental risk.
Quaas et al. (2007)	Myopic and sufficiently risk-averse farmers will on average choose low stocking rates.

3.3 The Pareto distribution

The Pareto distribution is one of the most well-known theoretical approaches to model empirical size distributions in economics, and also one of the oldest. Italian economist, engineer and sociologist Vilfredo Pareto introduced it in several works as a possible theoretical description of the wealth distribution in Italy (Pareto 1895, Pareto 1896, Pareto 1897a and Pareto 1897b). Although many alternatives have been proposed over the years (see e.g. Kleiber and Kotz 2003 for an overview on this), it is still frequently used today (Reed 2001, Reed 2003, Lévy 2009, Ioannides and Skouras 2013). The key characteristic of the Pareto distribution, is that there is no 'typical' value around which observations

cluster. That is, a mean value can be computed for a Pareto-distributed sample, but this mean value conveys less information than in the case of a normally-distributed variable. Moreover, the sample mean value converges much slower to the expected value of the sample than in the normal case (Boisot and McKelvey 2010).

The probability density function (PDF) of the Pareto distribution reads

$$p(x) = \frac{1}{\alpha - 1} \left(\frac{x}{x_{\min}} \right)^{-\alpha} \tag{1}$$

with $x > 0$. The strictly positive parameter α is called Pareto coefficient, and describes the behavior of the graph of the PDF for large x, which is usually referred to as 'the tail' of the distribution. It holds that the smaller the value of α, the more observations will be found for large x, a situation also referred to as 'long' or 'heavy' tail. $x_{\min} > 0$ is the cutoff value above which the distribution actually follows a power law, i.e. a Pareto distribution such that a double-logarithmic plot of power-law data follows a straight line.

The straight line appearance of power laws in double-logarithmic plots has quite often been used as a graphical diagnostic criterion. However, telling whether empirical data follow a certain distribution in a more satisfactory way is a pretty tricky question (cf. Clauset, Shalizi and Newman 2009, Engler and Baumgärtner 2015). Here, we use the method proposed by Clauset, Shalizi and Newman (2009) to fit the Pareto distribution to our data and to see whether it provides a plausible description of the it. This method has three advantages: (1) It employs the method of maximum likelihood, which performs better than the method of least squares (White, Enquist and Green 2008, Clauset, Shalizi and Newman 2009), (2) It is a valid test for the power-law hypothesis where other tests such as Kolmogorov-Smirnov or Cramér-von-Mises have been known to run into problems (cf. Bubeliny 2011, Engler and Baumgärtner 2015), and (3) at the same time, it provides an objective method to determine the cutoff value $x_{\min}$.

3.4 Regression analysis

In the last step of our analysis, we analyze the response of the stocking rate to three groups of variables related to different aspects of the farm using generalized linear models with Gaussian error structure. Based on our data set, we hypothesize that different site-

specific variables might explain the choice of stocking rate, and we therefore group our explanatory variables as follows: personal characteristics of the farmer (P); environmental factors (E); and farm characteristics (F).

Variables grouped as 'personal characteristics of the farmer' include his experience in years in commercial cattle farming, which includes experience as operator or manager as well as other experience such as being hired on a farm as worker. Other variables in this group are education level, gender and risk preference. Environmental variables include the inter-annual coefficient of precipitation variation and the average on-farm precipitation as described above in Section 3.1. Lastly, variables included in the group 'farm characteristics' are legal status of farm and farmer and the pasture quality as given by self-assessment in the questionnaire.

Prior to statistical modelling, we normalize variables where appropriate to guarantee optimal possibility to interpret regression coefficients. For model selection, we use the information theoretic approach (Burnham and Anderson 2004), which is implemented in the R package 'AICmodavg'. From the three groups of explanatory variables and the null model, we construct eight candidate models (Null, P, E, P+E, F, F+P, F+E, F+P+E). To correct for a possible small-sample bias, we base our model selection on AICc values and AICc weights rather than just the AIC.

4 Results

This section lists and discusses our results. In Section 4.1, we provide descriptive statistics for the data set, along with some illustrative comparisons. Section 4.2 gives a short introduction to the Pareto distribution and how we can use it as a starting ground for further analysis and group comparisons, before actually displaying our fitting results. Section 4.3 then reports the results of the group comparisons.

4.1 Descriptive statistics

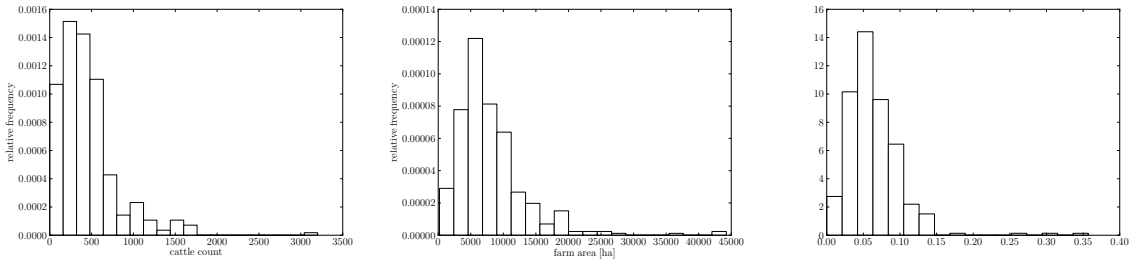
Tables 2 and 3 list some descriptive statistics of the data with Table 2 focusing on farm size, cattle numbers and the stocking rate, which is the ratio of these two, and Table 3 on

precipitation and risk attitude data. The average farm from our sample had 450 cattle on farm with a standard deviation (SD) of 369, and a size of 7970 ha (SD = 5504 ha). For comparison, in 2010, the average U.S. cattle farm contained 44 cows on a land area of 169 hectares (U.S. Department of Agriculture 2010). The U.S. state with the largest average acreage per farm, Wyoming, reported an average of 1112 hectares per farm (U.S. Census 2012). The largest farm from our sample has with 42244 hectares (approx. 422 km^2) roughly the same size as Barbados while the median farm is with 68 km^2 still roughly as big as San Marino (61 km^2). The smallest farm in our sample is still as big as Monaco (2 km^2). The values for skewness and kurtosis indicate that the distributions are all comparably right-skewed and leptokurtic. The Gini coefficients in the present sample for the variables cattle number ($G = 0.39$) and area ($G = 0.34$) are roughly half the size of the numbers reported by Eastwood, Lipton and Newell (2010) for the distribution of farm sizes in the United States in the 1990s ($G = 0.78$) and slightly smaller than what they report for Sub-Saharan Africa ($G = 0.50$). Figure 4 presents histograms of overall distributions for all three variables.

Table 2: Descriptive statistics of the overall sample of Namibian commercial cattle farms for the two size characteristics cattle number and area, and for the stocking rate. LSU = livestock units.

descriptive statistic	cattle [number]	area [ha]	stocking rate [LSU/ha]
sample size	351	391	347
minimum value	1	200	0.00013
maximum value	3200	42244	0.357
mean	450	7970	0.063
median	369	6800	0.058
standard deviation	361	5504	0.037
skewness	2.37	2.50	2.84
kurtosis	10.48	11.00	17.08
Gini coefficient	0.394	0.336	0.282

The average annual precipitation in the Namibian sample is 271 mm (Table 3) which compares to cities such as Phoenix, AZ (211 mm) or San Diego, CA (274 mm). The driest farm in the sample received only 63 mm of rain per year which roughly corresponds to the driest city in the U.S., Yuma, AZ. On the other hand, the wettest farm in the sample received on average 460 mm of precipitation per year. The very high variability



(a) cattle number (heads) (b) area (ha) (c) stocking rate (heads/ha)

Figure 4: Distribution of commercial cattle farm sizes in Namibia as measured in cattle number (a), rangeland area (b) and stocking rates (c).

Table 3: Descriptive statistics of the data set concerning precipitation and risk attitudes. Statistics are reported along with 95% confidence intervals in brackets where applicable.

descriptive statistic	value
average precipitation [mm]	270.70 (10.76)
Gini (precipitation)	0.179
median precipitation [mm]	281.54
minimum/maximum precipitation [mm]	63.35 / 460.02
average variation coefficient	0.285 (0.005)
Gini (variation coefficient)	0.073
median variation coefficient	0.280
minimum/maximum variation coefficient	0.213 / 0.475
average switchpoint	4.79 (0.14)
Gini (switchpoints)	0.117
median switchpoint	5

in inter-annual precipitation is reflected by an average coefficient of variation of 0.285 over all farms. In contrast, most places in North America and Europe have a climate characterized by C_v values from 0.10 to 0.15 (Figure 1).

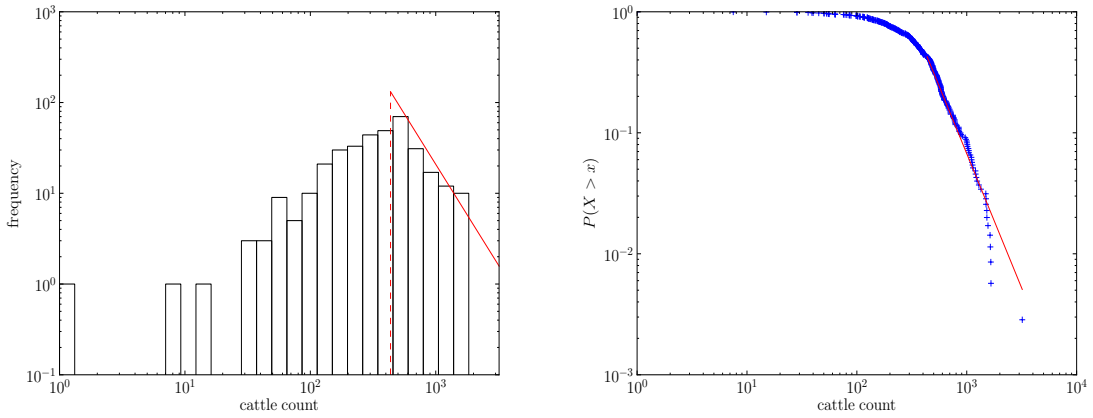
4.2 Pareto distribution

In Table 4, we list the results of the fitting procedure and the hypothesis test. We have used the originally proposed significance level, which is 10%, and estimates are based on 2500 Monte Carlo replications, so that p-values are ± 0.01 accurate (cf. Clauset, Shalizi and Newman 2009). The test rejects the power-law hypothesis for the farm area

and stocking rate data ($p = 0.01$ each), but does not reject it for the average cattle data ($p = 0.33$). In what follows, we will focus on the average cattle number data set. Overall, the test suggests that the Pareto distribution is a statistically plausible description for 42.5% of the data (149 out of 351 farms) with an estimated minimum farm size of $\hat{x}_{\min} = 436$. Figure 5 illustrates these findings with a double-logarithmic histogram and the corresponding survival function, along with the best Pareto fit in each figure.

Table 4: Fitting results of the Pareto distribution to our data (from left to right): data set, sample size N, the cutoff value of the Pareto distribution $\hat{x}_{\min}$, estimated 'tail index' $\hat{\alpha}$ along with 95% confidence interval, number of farms in Pareto branch of distribution $n \geq x_{\min}$ and p-values. p-values significant at the 10% level are marked with an asterisk.

data set	N	$\hat{x}_{\min}$	$\hat{\alpha}$	$n \geq x_{\min}$	p-value
avg. cattle count	351	436	3.22 (0.02)	149	0.33*
rangeland area	391	7000	3.25 (0.02)	190	0.01
stocking rate	347	0.051	3.42 (0.02)	215	0.01



(a) histogram (b) survival function

Figure 5: Best fit of the Pareto distribution to the average herd size data: panel (a) shows a log-log histogram along with the best fitting Pareto distribution including an extra marking of the cutoff value $x_{\min}$ while panel (b) shows the so-called survival function and the best Pareto fit.

While the Pareto distribution is not the only plausible probability distribution for this data set (cf. Engler and Baumgärtner 2015 for a detailed take on this issue), it does separate the farms into two groups with respect to the cutoff value $x_{\min}$. Since this distinction seems fairly neutral and objective, we consider this as one possible starting point that calls for further investigation concerned with the relation of environmental risk

and the farmers' risk attitudes in Section 4.3.2. Moreover, we will refer to farms in the Pareto branch of the distribution ($x \geq x_{\min}$) as 'Paretian' and accordingly, we will call the other farms 'non-Paretian' ($x \leq x_{\min}$).

Based on our findings so far, we organize our investigations into three sections that reflect the farmer's affiliation to the following groups: (1) overall sample, (2) affiliation to the two branches of the distribution (Paretian, non-Paretian) and (3) affiliation to the risk preference group. We report the findings in the following section.

4.3 Farm size, environmental risk and risk preferences

We take up the results obtained so far in the following manner: Section 4.3.1 investigates the correlations between precipitation data and farm sizes for the overall sample, before we look at the same question for the two branches of the distribution resulting from the fit of the Pareto distribution in Section 4.3.2. Finally, we compare the subgroups of farmers that result from their risk preferences in Section 4.3.3.

4.3.1 Overall sample

In Table 5, we present the Spearman correlations between herd size and precipitation data (upper section of table), mean farm size with precipitation data (middle section of the table) and stocking rates with precipitation data (lower section of table). The effect sizes classify the strength of effect for each pair of correlatives after Cohen's scale (cf. Cohen 1988: 82). For further illustration, the associated scatter plots can be found in Appendix A, and a principal component analysis of the full data set in Appendix B.

While we do not find any significant correlation between area and any of the precipitation variables ($\rho = -0.01$, $p = 0.81$ and $p = 0.89$), we do find significant correlations between herd size and mean annual precipitation ($\rho = 0.24$, $p < 0.001$) and herd size and variation coefficient ($\rho = -0.29$, $p < 10^{-5}$), albeit with small effect sizes. We find the strongest correlation coefficients when looking at the stocking rate: there is again a positive correlation with mean annual precipitation ($\rho = 0.34$) and a negative one with variation coefficients ($\rho = -0.34$). Both correlations have a medium effect size and are statistically highly significant as $p < 10^{-7}$.

Table 5: Spearman correlation coefficients of the different size measures with precipitation variables. p-values are reported for the null hypothesis 'true ρ is equal to zero', and effect sizes are classified according to Cohen's scale : $0.1 \leq \rho < 0.30$: small, $0.30 \leq \rho < 0.50$: medium, $\rho \geq 0.50$: large.

herd size correlated with	Spearman's ρ	p-value	effect size
mean precipitation	0.24	< 0.001	small
coefficient of variation	-0.29	$< 10^{-5}$	small
area correlated with			
mean precipitation	-0.01	0.81	none
coefficient of variation	-0.01	0.89	none
stocking rate correlated with			
mean precipitation	0.34	$< 10^{-7}$	medium
coefficient of variation	-0.34	$< 10^{-7}$	medium

4.3.2 Distribution branches

Table 6 lists the correlation coefficients for the same pairs of correlatives as in the previous section, but this time ordered according to the fitting results of the Pareto distribution to the data. This is to say that we compare farms larger than the cutoff value $x_{\min}$ with the ones smaller than this cutoff value. We base this distinction on the sole statistically plausible fit of the Pareto distribution to our data, which was to the average herd size data. Thus, we use $x_{\min} = 436$ in what follows.

In terms of direction of correlations, their strengths and significance, we roughly recover the results from the overall sample. Because of reduced sample sizes ($N = 110$ and $N = 148$ as compared to $N = 258$ in the previous section), effect sizes are slightly smaller. Overall, there are no significant correlations in the 'Paretian' subgroup (i.e. for farms with herd size $x \geq 436$). However, we do find significant correlations for all pairs of correlatives in the 'non-Paretian' subgroup (i.e. for farms with herd size $x < 436$), the strongest one being again the one between stocking rate and variation coefficient ($\rho = -0.42$, $p < 10^{-6}$). All correlations in this branch are very unlikely to result from pure chance as p-values are consistently less than or equal to 0.01. These results suggest that there is indeed a difference between Paretian and non-Paretian farms, but the question remains what kind of difference that might be. We look at this in more detail in Table 7.

Table 6: Spearman correlation coefficients of size with precipitation variables: Comparison of the two branches suggested by the fit of the Pareto distribution. The alternative hypothesis H_1 was 'true ρ is not equal to zero'.

Paretian farmers ($N = 110$)			
correlatives	Spearman's ρ	p-value	effect size
mean herd size – precipitation	0.00	0.99	none
mean herd size – coefficient of variation	−0.01	0.91	none
stocking rate – precipitation	0.13	0.17	none
stocking rate – coefficient of variation	−0.06	0.53	none
non-Paretian farmers ($N = 148$)			
mean herd size – mean precipitation	0.24	< 0.01	small
mean herd size – coefficient of variation	−0.23	< 0.01	small
stocking rate – mean precipitation	0.43	$< 10^{-7}$	medium
stocking rate – coefficient of variation	−0.42	$< 10^{-6}$	medium

Table 7 lists our results concerning mean values of switchpoints (i.e. risk attitudes), farm sizes and coefficient of variation of inter-annual precipitation. Since we do not make any assumption about the distribution of these variables, we used the Mann-Whitney U-Test to test for significant differences in these mean values. Because average herd size and rangeland area differ significantly in the two branches by construction, we do not report p-values for differences in these variables. We find that farmers in the two branches have roughly the same risk preferences (mean of non-Paretian farmers: 4.65 as compared to 4.83 for Paretian farmers) as 95% confidence intervals overlap quite a bit. Consequently, we cannot reject the possibility that the data would realize as observed by mere chance ($p = 0.20$). However, we find that Paretian farmers face quite different environmental risks (in terms of C_v-values) than non-Paretian farmers and this difference is very unlikely to result from pure chance as neither the 95% nor the 99% confidence intervals of the average coefficients of variation overlap 0.274 (0.006) for Paretians compared to 0.294 (0.007) for non-Paretians), in addition to a p-value smaller than 10^{-4} and a medium effect size of 0.55. Moreover, average stocking rates are higher for Paretian farms than for non-Paretian ones (0.073 compared to 0.047), an effect which not only is highly significant ($p < 10^{-13}$) but also features a large effect size (Cohen's $d = 0.96$), while Paretian farmers face

environmental conditions with significantly ($p = 0.003$) more mean annual precipitation (296.2 mm compared to 263.9 mm), which is a moderately strong effect (Cohen's $d = 0.37$).

Table 7: Paretian and non-Paretian branches of the empirical distribution in comparison. p-values have been calculated with the Mann-Whitney U Test with H_0 being 'true shift in location is equal to 0'. Where meaningful, we provide 95% confidence intervals in brackets and effect sizes (Cohen's d) to supplement the results of the statistical hypothesis tests. Classification of effect sizes follows Cohen's original suggestion ($0.2 \leq d < 0.5$: small; $0.5 \leq d < 0.8$: medium; $d \geq 0.8$: large).

variable	Paretian	non-Paretian	p	Cohen's d (effect size)
average switchpoint	4.65 (0.26)	4.83 (0.26)	0.20	0.15 (none)
average herd size	752 (73)	236 (18)	N/A	N/A
average rangeland [ha]	10845 (1091)	5517 (430)	N/A	N/A
average stocking rate [LSU/ha]	0.073 (0.020)	0.047 (0.033)	$< 10^{-13}$	0.96 (large)
average precipitation [mm]	296.2 (14.0)	263.9 (15.4)	0.003	0.37 (small)
average variation coefficient	0.274 (0.006)	0.294 (0.007)	$< 10^{-4}$	0.55 (medium)

4.3.3 Risk preference groups

In Table 8, we list the variables describing the risk-preference subgroups together with 95% confidence intervals and effect sizes (upper part), together with Spearman correlations of size and precipitation variables (lower part). Overall, we find that the risk-preference groups are fairly similar in terms of the key characteristics such as herd size, rangeland area, stocking rates and coefficient variation of variation of inter-annual precipitation. In fact, 95% confidence levels always overlap (Table 8). Regarding correlations of size and precipitation variables (lower part of Table 8), we find slightly negative correlations of herd sizes and variation coefficients, but effect sizes are small or medium by only a very slim margin ($\rho = -0.34$ for risk-neutral and $\rho = -0.23$ risk-loving subgroup). We do not find any correlation at all for the risk-averse subgroup regarding herd size-related correlations. These values are very similar to the effects observed in the overall sample (cf. Table 5). The correlations of stocking rates and variation coefficients have the same direction, but are slightly stronger which gives small effect sizes for the risk-averse ($\rho = -0.29$) and the risk-loving subgroups ($\rho = -0.44$). As to correlations of herd size or stocking rate with mean annual precipitation, the only subgroup with statistically significant correlations

is the risk-loving one, except for herd size and mean precipitation, where we also find a statistically significant value ($\rho = 0.30$) with medium effect size. The risk-loving subgroup features a moderately positive correlation of herd size and precipitation ($\rho = 0.34$) and a large positive correlation between stocking rates and precipitation ($\rho = 0.58$).

Table 8: Key characteristics of risk preference groups. Effect size classification and 95% confidence intervals are given in brackets. Spearman correlation coefficients are additionally marked with an asterisk when statistically significant at the 5% level.

	risk preference group		
variable	risk-averse	risk-neutral	risk-loving
sample size	52	64	40
average switchpoint	3.50 (0.26)	5.00 (0.00)	6.08 (0.08)
average herd size	507 (99)	434 (74)	461 (122)
median herd size	451	392	368
average rangeland [ha]	8705 (1448)	7804 (1070)	7412 (1253)
median rangeland [ha]	8000	6824	6535
average stocking rate [LSU/ha]	0.066 (0.009)	0.058 (0.006)	0.063 (0.010)
average precipitation [mm]	267.8 (26.2)	284.3 (19.3)	269.5 (27.6)
average variation coefficient	0.286 (0.012)	0.280 (0.008)	0.286 (0.013)
Spearman's ρ (herd size – precipitation)	0.00 (none)	0.30* (medium)	0.34* (medium)
Spearman's ρ (herd size – variation coeff.)	0.00 (none)	-0.34* (medium)	-0.23* (small)
Spearman's ρ (stocking rate – precipitation)	0.22 (small)	0.24 (small)	0.58* (large)
Spearman's ρ (stocking rate – variation coeff.)	-0.29* (small)	-0.18 (small)	-0.44* (medium)

Table 9 compiles the results of the Mann-Whitney U test for mean value comparisons for the relevant variables from Table 8. We do not find a single pair of subgroups where differences in mean values of any of these variables are somewhat close to the 5% level.

4.4 Regression analysis

We report the results of our model selection procedure in Table 10. The model P+E, which combines the farmer's personal characteristics (P) and environmental factors (E) performs best by far (model weight = 89.53%), and there is no other model within the range of ΔAICc ≤ 2, which is usually recommended as maximum acceptable threshold in the literature (Burham and Anderson 2004). The model parameters of the single best-

Table 9: p-values for mean-value comparisons of the risk preference groups, which have been calculated with the Mann-Whitney U Test with H_0 being 'true shift in location is equal to 0'. The leftmost column shows which groups are compared, while the top row of the table specifies the variable for which the comparison is done. We do not find significant differences in any of these mean values.

groups compared	... with respect to mean values in ...				
	herd size	area	stocking rate	precipitation	variation coeff.
risk-averse – risk-neutral	0.373	0.488	0.311	0.257	0.629
risk-loving – risk-averse	0.312	0.277	0.697	0.747	0.829
risk-neutral – risk-loving	0.794	0.543	0.591	0.432	0.572

ranked model are given in Table 11. Based on this analysis, the only parameters that seem to have a significant influence on stocking rates are gender of the farmer ($p = 0.024$), where the negative regression coefficient indicates that the stocking rate is smaller if the farmer is female, and the variation coefficient of inter-annual precipitation ($p < 0.001$). The response of the stocking rate to larger variation in precipitation is negative. All other variables do not have a significant impact on the stocking rate.

Following up on these findings, we also provide the results of an analysis of covariance (ANCOVA) to look at these effects in more detail using precipitation variation and gender as explanatory variables (Figure 6). The two regression lines for males (blue) and females (red) are clearly not coinciding, which means that the two gender groups are statistically distinct (gender effect, $p = 0.017$). Moreover, there is a negative response of the stocking rate to an increase in the variation coefficient, and this effect is highly significant (risk effect, $p \ll 0.001$).

Table 10: Full summary of the eight different possible model specifications. Best-ranked models (ΔAICc ≤ 2) are marked with *. Model terms are variables grouped according to the following scheme: (P) farmer's personal characteristics; (E) environmental factors; (F) farm characteristics.

model	parameters	AICc	Δ AICc	model weight [%]	log-likelihood	R^2
P+E*	11	−1085.32	0	89.53	554.23	0.205
E	4	−1080.32	5.00	7.35	544.25	0.137
F+P+E	20	−1077.61	7.71	1.90	560.69	0.246
F+E	16	−1076.73	8.59	1.22	555.56	0.213
P	9	−1058.12	29.20	0.00	537.44	0.088
null	2	−1048.50	36.82	0.00	526.28	0.000
F+P	18	−1045.74	39.58	0.00	542.39	0.124
F	11	−1040.53	44.80	0.00	531.83	0.045

Table 11: Model coefficients of the best-fitting generalized linear model from the candidate set, together with standard errors and p-values. Levels of significance are indicated in the usual way, i.e. *** for $p < 0.001$, ** for $0.001 \leq p < 0.01$, and * for $0.01 \leq p \leq 0.05$.

variable	group	variable estimate	std. error	p-value
(Intercept)	none	0.0594992***	0.0016755	$\ll 0.001$
experience with this particular farm	P	0.0026932	0.0029359	0.360
experience as owner of a farm	P	0.0008763	0.0031591	0.782
experience as manager of a farm	P	−0.0033087	0.0018446	0.074
other experience in farm business	P	0.0013796	0.0017301	0.426
level of eductation	P	−0.0029965	0.0017726	0.092
gender	P	−0.0175704*	0.0077243	0.024
risk preference	P	−0.0008752	0.0016982	0.607
variation coefficient	E	−0.0075055***	0.0021398	< 0.001
average precipitation	E	0.0030896	0.0021592	0.154

5 Discussion

We discuss our results in the same order as just reported: Section 5.1 looks at the results for the overall sample, Section 5.2 discusses the results for the distribution branches, and Section 5.3 looks at the group comparisons based on risk preferences. Finally, Section 5.4 interprets our results from the regression analysis.

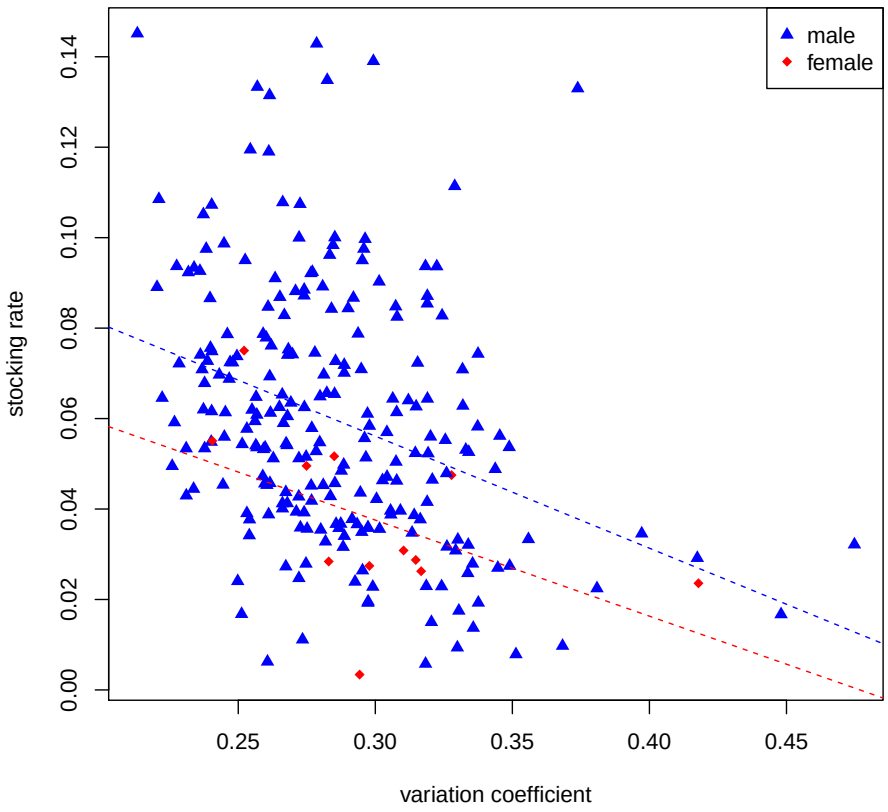


Figure 6: ANCOVA graph using gender and the variation coefficient as explanatory variables for the stocking rate. The analysis confirms that there is a highly significant effect of gender on stocking rate implemented on a farm when controlling for the variation coefficient, $F(1,241) = 8.49$, $p = 0.004$. There is no interaction effect of gender and variation coefficient ($p = 0.845$), i.e. there are two non-coinciding regression lines with distinct negative slopes.

5.1 Overall sample

The correlations that we find for the precipitation and herd size pairs of variables are in agreement with intuition: The more precipitation a farm gets on average per year, the more cattle the farm can support because there is higher grass production resulting from this higher precipitation. The positive and highly significant correlation with small effect size provide solid evidence for this feedback possibly underlying our data. On the other hand, the more uncertain the precipitation, the more uncertain it is how much cattle the land can support. One possible strategy to deal with this situation would be to choose a lower amount of cattle to be reared on the farm. Based on our data, it seems that this is what farmers predominantly do when faced with high precipitation risks. The same

effect can be seen in the correlations of stocking rates and precipitation variables, which substantiates the crucial role of the stocking rate in farm and rangeland management as already suggested in various other works (McArthur and Dillon 1971, Karp and Pope 1984, Rodriguez and Taylor 1988, Westoby, Walker and Noy-Meir 1989, Torell, Lyon and Godfrey 1991).

There are no correlations whatsoever for the area-precipitation pairs of variables. This is surprising because it would have been possible that farmers confronted with high inter-annual variation of precipitation and low mean annual precipitation try to gain access to larger areas of land that are also spatially diversified (cf. Quaas et al. 2007). This would be possible because there is a quite well-developed and functioning land market in Namibia (Olbrich 2012). Thus, we would have expected a positive correlation between area and coefficient of variation of inter-annual precipitation. However, there is no evidence for such an effect showing in our data. We find this striking because a strategy that seeks to hedge and diversify downside risks resulting from excessive variability in precipitation would be optimal at least for risk-averse farmers, given reasonable transaction costs.

5.2 Distribution branches

In Table 7, we have seen that Paretian farmers face on average significantly less inter-annual variability in rainfall, i.e. environmental risk. One possible explanation for our these findings could be that more stable environmental conditions reduce the Paretian farmers' income risks so that less resources and financial capital had to be spent on financial insurance, self-insurance and possibly self-protection. This capital could then be invested in additional cattle to be reared on farm. As a result of these effects – more certainty regarding forage production and general rangeland quality leading to less financial strain on the farmer – farms had, ceteris paribus, more opportunities to augment cattle production. On the other hand, confronted with greater environmental uncertainty, non-Paretian farmers might have a greater need for diversification of this larger risk. They might either tend to opt for various small farms at different locations or choose to become part-time farmers more often than the Paretian farmers. Both of these mechanisms would foster smaller herd sizes and less rangeland area in the non-Paretian branch.

The data give us no reason to assume that risk preferences might be distributed inhomogeneously over the two groups. This seems surprising because two ways of arguing seem a priori plausible: On one hand, one could have reasoned that non-Paretian farmers were less willing to take risks than their Paretian colleagues and so would tend to refrain from expanding their businesses, because expansion usually comes with higher financial risks. On the other hand, it would have also been imaginable that a 'no risk no gain' mentality could have led to the numbers that we found because it is possible that some of the farmers owning farms smaller than $x_{\min}$ today could have been forced to sell substantial parts of their herd in the aftermath of a risky strategy gone wrong at some earlier point in time. Both of these explanatory models would suggest at least some inhomogeneity regarding risk preferences in the two subgroups. However, it seems that our data does not support this view.

The strongest effect based on our mean value comparison between the two subgroups from the Pareto distribution is the difference in stocking rate mean value. Smaller farms (in terms of average herd size) have on average considerably (51%) more cattle per hectare than the larger ones. The average stocking rate of the non-Paretian farms is 0.077 LSU/ha which is only slightly below what Olbrich, Quaas and Baumgärtner (2013) report as average grazing capacity (0.080 LSU/ha[5]) for the sample. We interpret this as a possible fixed-costs effect: small farms in terms of herd size usually also have less rangeland. In order to be profitable in any given year, farms need to have a certain minimum herd size and this minimum herd size might be larger than actually suggested by grazing capacity. In the short term, this could be a viable strategy which might explain the much higher average stocking rates with smaller farms. It has also been suggested that risk-neutrality might imply optimality of higher stocking rates (Rodriguez and Taylor 1988). While there is no significant difference in average switchpoints here, we will consider this theory again in the next section.

As robustness checks, we performed multiple ANCOVAs to also account for other possible branching approaches in the distribution (Appendix C). Particularly, we checked three different distinctions between 'small' and 'large' farms, the sample mean, sample median and cutoff value of the Pareto distribution $x_{\min}$ to see whether the branching method has any impact on the results. Our analysis shows that the general result remains

[5]This value is based on the information provided by the farmers in the questionnaire.

the same when branching according to the sample mean value instead of the cutoff value of the Pareto distribution, but not when branching is done according to the sample median. Given that sample mean and cutoff value are almost identical, which yields to almost identical branches (only 4 farms 'switch' branches when going from Pareto cutoff value to sample mean as threshold value), this is not surprising. Statistical significance of the variation coefficient as explanatory variable for the stocking rate of large farms was practically identical ($p = 9 \cdot 10^{-6}$ and $p = 6 \cdot 10^{-6}$). Large farms according to these two definitions do not significantly react to larger variation coefficients ($p = 0.073$ and $p = 0.076$). For the sample median distinction, we do not recover this effect as farms in both parts of the distribution show the same regression coefficient (-0.181 and -0.182) and both are statistically significant at or even below the 1% level.

Apart from the statistical evidence, we do maintain that fitting the Pareto distribution is methodologically preferable to using the sample mean value as a criterion to distinguish between small and large farms. There are two main reasons for this: (1) the sample mean value has to be handled with care when the data are clearly not normally distributed (Clauset, Shalizi and Newman 2009, Boisot and McKelvey 2010); and (2) it is certainly more robust to consider a parameter from a distribution that provides a statistically plausible fit to the data than one from a distribution that can be rejected beyond any reasonable doubt (cf. Engler and Baumgärtner 2015 for explicit tests of normality in our data).

5.3 Risk preference groups

We did not find any detectable difference in any of the average size and precipitation variables between the three risk preference groups. Our finding contradicts the hypothesis from Olbrich (2012) according to which more risk-averse farmers tend to self-select by operating farms with on average lower environmental risk. This apparently missing self-selection might be explained by several factors. If a farmer buys or leases an already existing farm from someone else, the location is externally given and there is nothing the farmer can do except for accepting or not accepting the offer of buying or leasing that particular farm at the price named. That means, the most important parameters that characterize the farm (location, rangeland area, environmental conditions and the like)

are preset. Furthermore, factors such as proximity to hometown or close relatives might play a large role in these decisions, but these are not captured in the data. It is also not uncommon that farms are passed on within the family. Lastly, we have used the full sample regarding switchpoints here ($N = 156$), while Olbrich (2012) had to exclude certain farmers for econometric reasons reducing their sample size to $N = 99$.[6]

There is no difference in mean stocking rates between the three risk preference groups (Table 9). In the light of existing theories of rangeland management under uncertainty, this finding seems especially striking. For example, in their 2007 model, Quaas et al. show that a myopic and sufficiently risk-averse farmer will have a very conservative strategy – i.e. low stocking rates – as optimum. For our data, this would imply that our risk-averse subgroup should have at least the smallest stocking rate among all subgroups if they actually were risk-averse non-satiated expected utility maximizers. For risk-neutral farmers, Rodriguez and Taylor (1988) suggest that high stocking rates may have larger expected net present values than low stocking rates. However, we cannot find evidence for the presence of such effects in our data. Altogether, it seems that risk preferences do not make a statistically detectable difference in any of the size and precipitation variables.

It is quite striking that we find the strongest effects in terms of correlations of size and precipitation variables in all but one case in the risk-loving subgroup (cf. Table 8). It seems that a good part of the correlations observed for the overall sample (cf. Section 4.3.1) comes from the risk-loving farmers. Particularly the stocking rates are strongly correlated with precipitation ($\rho = 0.58$) and inter-annual coefficient of variation of precipitation ($\rho = -0.44$). We find the latter observation puzzling, because we would have expected to see the strongest negative correlation between size variables and coefficient of variation in the risk-averse subgroup, not in the risk-loving one. On the other hand, the risk lovers should in theory have featured a positive correlation, not a negative one. Again, there is no evidence for self-selection of farmers into locations that suit their risk preference.

[6]Olbrich's study focused on the determinants of risk preferences of the farmers in the sample. One key hypothesis was that risk preferences are critically influenced by life-history factors such as environmental risk experienced prior to age 18 and number of years living on the current farm. Thus, they had to remove all those farmers for which these data were not complete.

5.4 Regression analysis

The hypothesis that farm characteristics such as the quality of pasture land and legal status of farm and farmer might play a role in explaining the choice of actual stocking rate is not supported by our analysis (Table 10). Instead, it is the individual farmer's personal characteristics and environmental factors that matter. Surprisingly, within the group of variables describing the farmer's personal characteristics, neither the farmer's experience in commercial cattle farming nor his level of education or his risk preference have a significant impact on stocking rate (Table 11). That being said, it is also true that the explanatory variables 'level of education' and 'experience as manager of a farm' are close to the standard significance threshold level of 5%, and also contributed to the best model as identified by our information theoretical approach. The variables that clearly make a significant contribution to explaining stocking rate are 'variation coefficient' and 'gender', and these effects seem to be pretty robust as shown by our analysis of covariance (Figure 6). The negative impact of the variation coefficient on the stocking rate is as expected from our previous analyses, and well in agreement with intuition.

On the other hand, the actual choice of stocking rate seems to be strongly influenced by the farmer's gender (Table 11). The central result here is that female farmers choose a smaller stocking rate than their male colleagues at any level of the variation coefficient (Figure 6). That is, they seem to be more conservative regarding their farm management than men, and this conservative management cannot be explained by differing risk preferences. Gender differences in commercial farming have been reported particularly for the U.S. (Zeuli and King 1998, Hoppe and Korb 2013, van Rieper 2013), while most other studies concerned with gender seem to focus on smallholders rather than commercial farmers (e.g. Moock 1976, Doss and Morris 2000, Adeleke et al. 2008, Koru and Holden 2008, Puspitawati 2013) or use 'dated' data (Petermann, Behrmann and Quisumbing 2014: 145). Particularly, Zeuli and King (1998) report some gender-related differences in a sample of 2888 U.S. commercial farmers (112 of which were female) such as an on-average higher acreage in farms operated by women, they do not report numbers for stocking rates in cattle farming, nor can they be inferred from what they report. We also tested whether women in our sample had a higher acreage here as well, which might have been one explanation for their observed lower stocking rates all other things being

equal, but had to reject this ($p = 0.136$). Hence, while we clearly find a gender effect in the observed stocking rates, we have to leave the question regarding the cause of this effect for future research.

Concerning robustness of our results, we checked for multicollinearity in our variables using the variance inflation factor, which did not return any problematic values as all factor values were well below 5. In addition, because of the obvious potential of correlations between the variable 'experience with this particular farm' and the other experience-related explanatory variables in the model, we also double-checked our results with model specifications that explicitly excluded these sources of multicollinearity. While there were some slight changes regarding AICc weights and in the general ranking, the best-fitting model remained always the same. Moreover, there also were no changes in terms of parameter significance and direction or strength of relationships. Finally, we maintain that the small size of the female subsample ($N = 12$) does very likely not impair our results regarding the gender effect. In order to check whether sample size might pose a problem here, we conducted a t-test, which is suitable and powerful for small samples, with the stocking rates of male and female farmers, which confirmed the effect ($p = 0.002$). Therefore, based on our analysis, we maintain that the evidence for the central roles of environmental risk and the farmer's gender for the choice of stocking rate is fairly robust.

6 Summary and conclusion

In this study, we have investigated the relationship between farm size, environmental risk and risk preferences using a sample of 399 Namibian commercial cattle farmers. With the help of a recent statistical test, we have demonstrated that the Pareto distribution is a statistically plausible description of the tail of the herd size distribution, but not of stocking rate and area distributions. We have used this finding to check for differences between Paretian and non-Paretian farms, and have found the only significant difference to be the average inter-annual variability in rainfall, which is smaller for Paretian farms. Comparing the risk preference groups of farmers, we have not found any supporting evidence for the hypothesis that farmers generally self-select according to their risk preference, i.e. that more risk-averse farmers operate less risky farms. Moreover, we have found that, on aver-

age, size and precipitation characteristics of the farms are evenly distributed over the risk preference groups. Overall, we found correlations to be consistently strongest if stocking rate is a correlative, which supports its importance as key parameter in farm management, in agreement with the literature. However, we had to reject several other hypotheses from the rangeland management literature. Our regression analysis has clearly fleshed out that inter-annual precipitation variability is central to explaining stocking rates in our sample as is the farmer's gender, while other factors such as a farmer's experience or level of education do not seem to play a major role.

Altogether, we have provided solid evidence for the crucial role of environmental risk in extensive commercial cattle farming in semi-arid rangelands. Moreover, our analyses have shown that the stocking rate is indeed the central farm management parameter as suggested by many theoretical contributions in agricultural economics, while other theories do clearly not hold up that well, at least not in this particular case. The central other parameter explaining the stocking rate in our sample is the farmer's gender. This finding suggests that future research should explicitly focus on gender as an important explanatory variable for farm management and possibly also for other economic variables. Most research – and the present paper is no exception – has focused on the role of education, experience or risk and time preferences as explanatory variables for management decisions. Our results here suggest that we will be well-advised to broaden this focus in future studies.

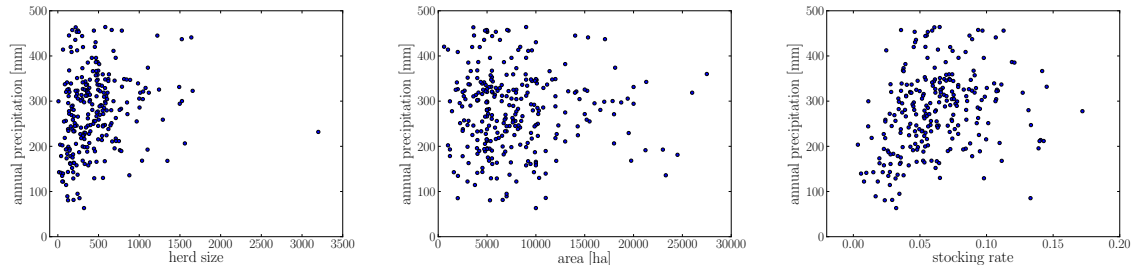
Acknowledgments

The authors would like to thank David J. Abson and Roland Olbrich for very helpful discussion of various ideas and for providing critical feedback.

Appendix

We provide scatter plots in Appendix A and the principal component analysis together with a correlation table for the full data set in Appendix B. Appendix C provides a detailed statistical test of alternative distinctions between 'small' and 'large' farms, as discussed in the main text.

A Scatter plots



(a) $\rho = 0.24$ (b) $\rho = -0.01$ (c) $\rho = 0.34$

(d) $\rho = -0.29$ (e) $\rho = -0.01$ (f) $\rho = -0.34$

Figure 7: Scatter plots for all pairs of variables listed in Table 5, based on $N = 258$ (cf. Section 3.1). The upper row of graphs (Figs. a through c) shows scatter plots of annual precipitation and all size variables herd size, area and stocking rate. The lower row of graphs (Figs. d through f) shows scatter plots of coefficient of variation of inter-annual precipitation and all size variables.

B Principal component analysis

We provide the biplot associated with the principal component analysis of our data set in figure 8. The two principal components together explain 33.7% of total variance in the data. The corresponding full correlation table can be found in Table 12. Because only data records complete in every variable can be used here ($N = 244$), some values may slightly differ from those reported in Section 4.3.1, where we could use a larger data set ($N = 258$).

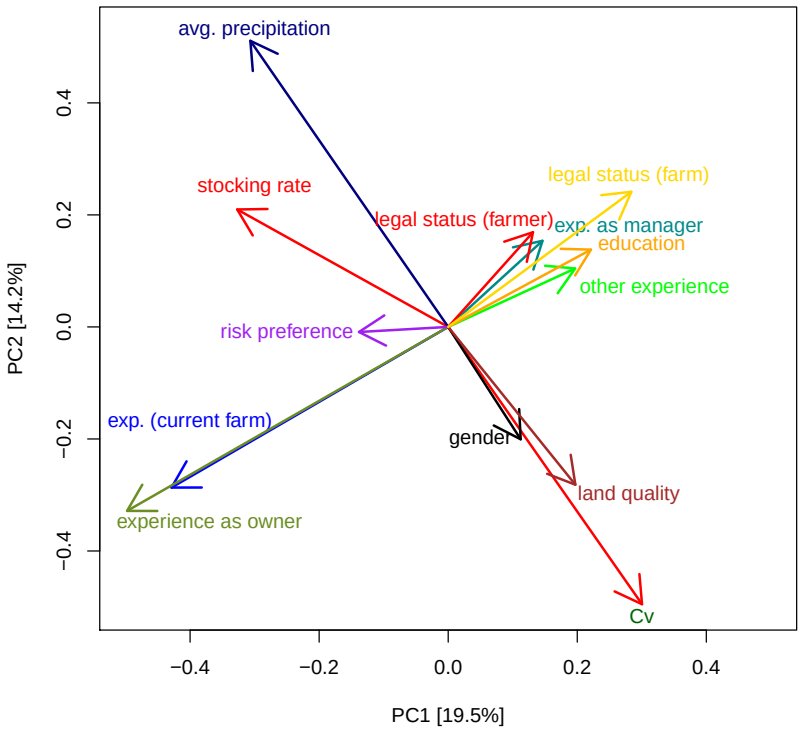


Figure 8: PCA biplot for our data set based on $N = 244$ data points, colors are introduced for better readability only and do not convey additional information.

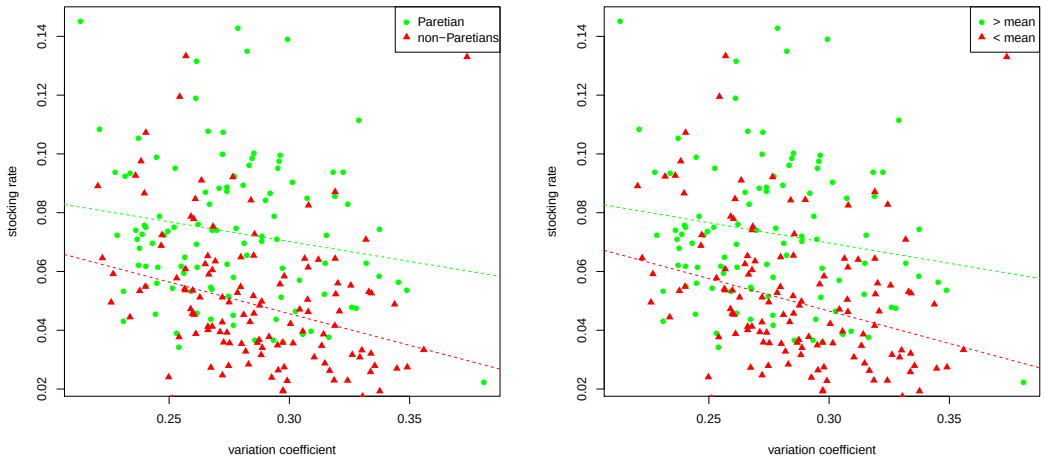
Table 12: Correlation table of the Namibian commercial cattle farmer survey using all variables relevant for the present paper.

	stocking rate	exp. (this farm)	exp. (manager)	exp. (other)	experience (owner)	C_v	avg. precip.	edu. level	gender	risk pref.	land q.	legal (farmer)	legal (farm)
stocking rate	1.00	0.18	−0.12	0.01	0.19	−0.36	0.30	−0.14	−0.17	0.01	−0.05	−0.15	−0.12
experience (this farm)	0.18	1.00	0.01	−0.07	0.79	−0.07	0.07	−0.29	−0.03	0.06	−0.04	−0.09	−0.20
experience (manager)	−0.12	0.01	1.00	−0.01	−0.25	0.03	0.07	−0.01	0.03	−0.09	0.00	0.06	0.21
experience (other)	0.01	−0.07	−0.01	1.00	−0.22	0.08	−0.06	0.10	−0.06	−0.09	0.10	0.06	0.24
experience (owner)	0.19	0.79	−0.25	−0.22	1.00	−0.09	0.08	−0.29	−0.07	0.16	−0.10	−0.13	−0.36
variation coefficient	−0.36	−0.07	0.03	0.08	−0.09	1.00	−0.63	0.00	0.09	−0.10	0.20	−0.05	0.01
avg. precip.	0.30	0.07	0.07	−0.06	0.08	−0.63	1.00	−0.11	−0.13	0.03	−0.31	0.02	−0.03
education level	−0.14	−0.29	−0.01	0.10	−0.29	0.00	−0.11	1.00	−0.09	0.11	0.01	−0.03	0.00
gender	−0.17	−0.03	0.03	−0.06	−0.07	0.09	−0.13	−0.09	1.00	−0.04	0.13	−0.03	0.02
risk preference	0.01	0.06	−0.09	−0.09	0.16	−0.10	0.03	0.11	−0.04	1.00	−0.12	−0.11	−0.06
land quality	−0.05	−0.04	0.00	0.10	−0.10	0.20	−0.31	0.01	0.13	−0.12	1.00	−0.02	0.04
legal status (farmer)	−0.15	−0.09	0.06	0.06	−0.13	−0.05	0.02	−0.03	−0.03	−0.11	−0.02	1.00	0.17
legal status (farm)	−0.12	−0.20	0.21	0.24	−0.36	0.01	−0.03	0.00	0.02	−0.06	0.04	0.17	1.00

C Test of alternative distribution branches

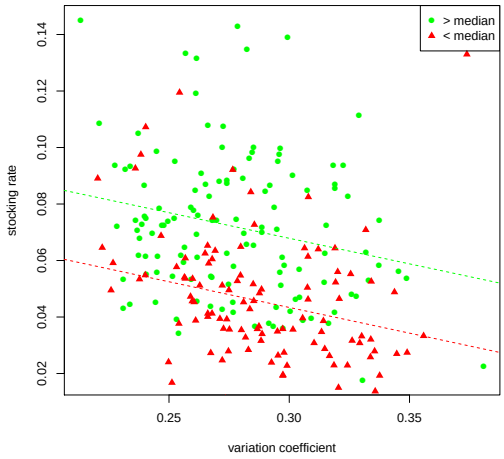
For comparison of the different possible thresholds to distinguish between 'small' and 'large' farms, we conduct one-way ANCOVAs to determine whether there is a statistically significant difference between the distribution branches on stocking rate controlling for the variation coefficient. Specifically, we already know that there is such an effect when looking at the branches of the Pareto distribution (Paretian and non-Paretian farms) from our main analysis, and we now test if this result changes using other branching thresholds. The results are illustrated in Figure 9.

Overall, we find a significant effect of distribution branch on stocking rate after controlling for the variation coefficient with every branching method. Specifically, Pareto branching yields a slightly higher significance level ($F(1,241) = 73.96$, $p = 10^{-15}$) than branching according to the sample mean value ($F(1,241) = 63.77$, $p = 6 \cdot 10^{-14}$). Regarding effect sizes, results for Pareto and sample mean branches are almost identical: there are significant differences between small and large farms in both cases in that stocking rates of large farms are not significantly explained by the variation coefficient (Pareto: $p = 0.073$, mean: $p = 0.076$), but small farms are ($p \approx 10^{-6}$ for both). In contrast, when we use the sample median as threshold, there is no significant difference between small and large farms in terms of how well the variation coefficient explains the stocking rate, we only recover a significant and equally negative impact of variation coefficient on stocking rates in both branches. This can be seen in Figure 9: in panels (a) and (b), there are two distinct slopes and intercepts, but two intercepts and one slope in panel (c), i.e. regression lines are parallel.



(a) Pareto

(b) mean



(c) median

Figure 9: ANCOVA using different ways to separate small and large farms (a) Pareto, sample mean (b) and sample median (c).

References

Adeleke, O.A., O.I. Adesiyan, O.A. Olaniyi, K.O. Adelalu, and H.M. Matanmi (2008), Gender differentials in the productivity of cereal crop farmers: a case study of maize farmers in Oluyole local government area of Oyo State, *Agricultural Journal* 3(3), 193–198

Andersen, S., G.W. Harrison M.I. Lau and E.E. Rutström (2006), Elicitation using multiple price list formats, *Experimental Economics* 9, 383–405

Auerbach, F. (1913), Das Gesetz der Bevölkerungskonzentration, *Petermanns Geographische Mitteilungen* 49, 73–76

Beukes, P.C., R.M. Cowling and S.I. Higgins (2002), An ecological economic simulation model of a non-selective grazing system in the Nama Karoo South Africa, *Ecological Economics* 42, 221–242

Binswanger, H.P. (1980), Attitudes towards risk: Experimental measurement in rural India, *American Journal of Agricultural Economics* 62, 395–407

Boisot, M. and B. McKelvey (2010), Integrating modernist and postmodernist perspectives on organizations: a complexity science bridge, *Academy of Management Review* 35(3), 415–433

Bubeliny, P. (2011), Some notes on biasedness and unbiasedness of two-sample Kolmogorov-Smirnov test, Unpublished, Dept. of Mathematics and Physics, Charles University, Prague. 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

Clauset, A., C.R. Shalizi, and M.E.J. Newman (2009), Power-law distributions in empirical data, *SIAM Review* 51(4), 661–703

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd edition. Lawrence Erlbaum Associates, Hillsdale

Doss, C.R. and M.L. Morris (2000), How does gender affect the adoption of agricultural innovations, *Agricultural Economics* 25(1), 27-39.

Eastwood, R., M. Lipton and A. Newell (2010), Chapter 65: Farm Size, In: Gardner, B.L. and G.C. Rausser (Eds.), *Handbook of Agricultural Economics*, Vol. 4, 3323–3397

Engler, J.-O. and S. Baumgärtner (2015), Model choice and size distribution: a Bayequentist approach, *American Journal of Agricultural Economics* 97(3), 978–997

Fatichi, S., V.Y. Ivanov and E. Caporali (2012), Investigating inter-annual variability of precipitation at the global scale: Is there a connection with seasonality? *Journal of Climate* 25, 5512–5523

Holt, C.A. and S.K. Laury (2002), Risk aversion and incentive effects, *American Economic Review* 92(5), 1644–1655

Hoppe, R.A. and P. Korb (2013), *Characteristics of Women Farm Operators and Their Farms*, U.S. Department of Agriculture, Economic Research Service

Illius, A.W. and T.G. O'Connor (1999), On the relevance of non-equilibrium concepts to arid and semi-arid grazing systems, *Ecological Applications* 9, 798–813

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

Jacob, D. (2001), A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin, *Meteorology and Atmospheric Physics* 77, 61–73

Jacob, D. and R. Podzun (1997), Sensitivity studies with the regional climate model REMO, *Meteorology and Atmospheric Physics* 63, 119–129

Janssen, M.A., J.M. Anderies and B.H. Walker (2004), Robust strategies for managing rangelands with multiple stable attractors, *Journal of Environmental Economics and Management* 47, 140–162

Karp, L. and A. Pope (1984), Range management under uncertainty, *American Journal of Agricultural Economics* 66, 437–446

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

Koru, B. and S. Holden (2008), Difference in maize productivity between male- and female-headed households in Uganda, Ethiopian Development Research Institute.

Lévy. M. (2009), Gibrat's law for (all) cities: comment, *American Economic Review* 99(4), 1672–1675

Mandelbrot, B. (1961), Stable Paretian random functions and the multiplicative variation of income. *Econometrica* 29, 517–543

McArthur, D. and J.L. Dillon (1971), Risk, utility and stocking rate, *Australian Journal of Agricultural Economics* 15(1), 20–35

Miehe, S., J. Kluge, H. von Wehrden, and V. Retzer (2010), Long-term degradation of Sahelian rangeland detected by 27 years of field study in Senegal, *Journal of Applied Ecology* 47, 692–700

Millennium Ecosystem Assessment (2005), Chapter 22: Dryland systems, In: Hassan, R., R. Scholes and N. Ash (Eds.), *Ecosystem and human well-being. Current state and trends*, Vol. 1

Mendelsohn, J. (2006), Farming systems in Namibia, http://www.agris.fao.org

Moock, P.R. (1976), The efficiency of women as farm managers, *American Journal of Agricultural Economics* 58(5), 831–835

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

Olbrich, R. 2012. *Environmental Risk and Sustainability: The Case of Commercial Livestock Farming in Semi-Arid Rangelands.* Ph.D. dissertation, Leuphana University of Lüneburg. Available online at http://opus.uni-lueneburg.de/opus/volltexte/2012/14208/

Olbrich, R., M.F. Quaas and S. Baumgärtner (2013), Characterizing commercial cattle farms in Namibia: risk, management and sustainability, *University of Lüneburg Working Paper Series in Economics* No. 248, revised version (July 2013)

Pareto, V. (1895), La legge della domanda, *Giornale degli Economisti* 10, 59–68, English translation in *Rivista di Politica Economica* 87, 1997

Pareto, V. (1896), La courbe de la repartition de la richesse, Reprinted in G. Busoni (ed.): Oeuvres complètes de Vilfredo Pareto, Tome 3: Ecrits sur la courbe de la

repartition de la richesse, 1965, Librairie Droz, Geneva, 1–15, English translation in *Rivista di Politica Economica* 87, 1997

Pareto, V. (1897a), Aggiunta allo studio della curva delle entrate, *Giornale degli Economisti* 14, 15–26, English translation in *Rivista di Politica Economica* 87, 1997

Pareto, V. (1897b), *Cours d'Economie Politique*, Macmillan, London

Perrings, C. and B.H. Walker (1997), Biodiversity, resilience and the control of ecological-economic systems: the case of fire-driven rangelands, *Ecological Economics* 22, 73–83

Peterman, A., J.A. Behrman and A.R. Quisumbing (2014), *A review of empirical evidence on gender differences in nonland agricultural inputs, technology, and services in developing countries* Springer Netherlands

Puspitawati, E. (2013), *Contracts and Marketing Decisions of Indonesian Potato Growers*, Doctoral dissertation, The University of Adelaide

Quaas, M.F., S. Baumgärtner, C. Becker, K. Frank and B. Müller (2007), Uncertainty and sustainability in the management of rangelands, *Ecological Economics* 62, 251–266

Reed, W.J. (2001), The Pareto, Zipf and other power laws, *Economics Letters* 74, 15–19

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

Rodriguez, A. and R.G. Taylor (1988), Stochastic modeling of short-term cattle operations, *American Journal of Agricultural Economics* 70, 121–132

Sornette, D. (2003), *Critical Phenomena in Natural Sciences*, 2nd edition. Springer, Berlin

The Namibian (2010), Namibia: Weaner Price Recovers Drastically, *The Namibian*, Windhoek, November 2, retrieved from http://allafrica.com/stories/201011020141.html on 08/13/2014

Torell, L.A., K.S. Lyon and E.B. Godfrey (1991), Long-run versus short-run planning horizons and the rangeland stocking rate decision, *American Journal of Agricultural Economics* 73(3), 795–807

United States Census (2012), Section 17: Agriculture, available online at http://www.census.gov/prod/2011pubs/12statab/agricult.pdf, retrieved on 11/21/2013

United States Department of Agriculture (2010), Farms, Land in Farms and Livestock Operations 2010 Summary, available online at http://www.nass.usda.gov/Publications/TodaysReports/reports/fnlo0211.pdf, retrieved on 11/21/2013

Vetter, S. (2005), Rangelands at equilibrium and at non-equilibrium: recent developments in the debate, *Journal of Arid Environments* 62, 321–341

van Riper, L. (2013), Women as collaborative leaders on rangelands in the Western United States, *Rangelands* 35(6), 47–57

von Wehrden, H., J. Hanspach, P. Kaczensky, J. Fischer and K. Wesche (2012), Gobal assessment of the non-equilibrium concept in rangelands, *Ecological Applications* 22(2), 393–399

von Wehrden, H., J. Hanspach, K. Ronnenberg and K. Wesche (2010), Inter-annual rainfall variability in Central Asia – A contribution to the discussion on the importance of environmental stochasticity in drylands, *Journal of Arid Environments* 74(10), 1212–1215

Westoby, M., B.H. Walker and I. Noy-Meir (1989), Opportunistic management for rangelands not at equilibrium, *Journal of Range Management* 42, 266–274

White E.P., B.J. Enquist and J.L. Green (2008), On estimating the exponent of power-law frequency distributions, *Ecology* 89(4), 905–912

World Bank (2013), Official currency exchange rates, available online at http://data.worldbank.org/indicator/PA.NUS.FCRF, retrieved on 11/04/2013

Zeuli, K.A. and R.P. King (1998), Gender differences in farm management, *Review of Agricultural Economics* 20(2), 513–529

Paper 3

An axiomatic foundation of preferences under Knightian uncertainty

An axiomatic foundation of preferences under Knightian uncertainty

Stefan Baumgärtner[a,b,*] and John-Oliver Engler[a]

[a] Department of Sustainability Science and Department of Economics, Leuphana University of Lüneburg, Germany

[b] Department of Environment and Natural Resources, University of Freiburg, Germany

Abstract: We develop a non-expected-utility approach to decision making under Knightian uncertainty which circumvents some of the conceptual problems of existing approaches. We understand Knightian uncertainty as income lotteries with known payoffs but unknown probabilities in each outcome. Based on seven axioms, we show that there uniquely (up to linear-affine transformations) exists an additive and extensive function from the set of Knightian lotteries to the real numbers that represents uncertainty preferences on the subset of lotteries with fixed positive sum of payoffs over all possible states of the world. We define the concept of uncertainty aversion such that it allows for interpersonal comparison of uncertainty attitudes. Furthermore, we propose Rényi's (1961) generalized entropy as a one-parameter preference function, where the parameter measures the degree of uncertainty aversion. We illustrate it with a simple decision problem and compare it to other decision rules under uncertainty (maximin, maximax, Laplacian expected utility, minimum regret, Hurwicz).

Keywords: Knightian uncertainty, deep uncertainty, decision making, non-expected utility, environmental decisions, ambiguity, ambiguity aversion

JEL-Classification: D81, H30

*Corresponding author: Chair of Environmental Economics and Resource Management, University of Freiburg, Tennenbacher Str. 4, D-79106 Freiburg, Germany, phone: +49 761 203-3753, email: stefan.baumgaertner@ere.uni-freiburg.de.

1 Introduction

In 1921, John Maynard Keynes and Frank Knight simultaneously coined the distinction between situations of *risk* and situations of *uncertainty* (Keynes 1921, Knight 1921).[1] The distinction is based on knowledge (risk) or nescience (uncertainty) of probabilities of possible future states of the world. For risk, there is the well-studied and established von Neumann-Morgenstern approach leading to the expected utility model (von Neumann and Morgenstern 1944). Nonetheless, there are compelling reasons to care about Knightian uncertainty as well, for it may be outright impossible to assign probabilities to outcomes at all. For example, the system creating the outcomes may be too complex[2] and the time horizon involved too long to warrant any reasonable probabilistic assessment. An arguably striking example is our planet's climate where we do not even fully understand every single part of the system yet, let alone all feedback loops contained (Mehta et al. 2009). As a matter of fact, recent climate predictions have been remarkably off (Fyfe, Gillett and Zwiers 2013). Moreover, the fundamental disagreement of expert groups on a certain issue alone, for whatever reason, might invoke situations of Knightian uncertainty (Feduzi and Runde 2011). In such cases, one might be tempted to attach subjective probabilities ('beliefs') to the scenarios, but there are catches: (1) existence of such probabilities cannot be guaranteed (Ellsberg 1961[3], Halevy 2007), even when experts are asked for their educated guesses (Millner et al. 2013), and (2) especially experts tend to be overconfident regarding their results (Alpert and Raiffa 1982), which introduces yet another intricacy. And, aside from all this, it seems also justified to ask – at least from time to time – whether any probability is really better than no probability.

A number of decision criteria have been suggested, non-probabilistic and probabilistic, and there are irrefutable problems with both. First and foremost, the probabilistic ones

[1] The term 'Knightian uncertainty' has prevailed in the literature though.

[2] Here, 'complex' refers to the system consisting of many interacting parts that are interconnected via multiple nonlinear feedback loops (cf. Sornette 2003).

[3] Ellsberg's anomaly or paradox refers to the following: Assume there is an urn that contains 120 balls in total, 40 of which are blue (b) and the other 80 yellow (y) and red (r), but with unknown color frequency ratio. Experiment participants are offered the following two bets: f_1 = 'win 10$ if the ball drawn from the urn is blue' or f_2 = 'win 10$ if the ball drawn from the urn is red' and f_3 = 'win 10$ if the ball drawn from the urn is either blue or yellow' or f_4 = 'win 10$ if the ball drawn from the urn is either red or yellow'. Ellsberg's finding was that the vast majority of participants preferred f_1 to f_2 and f_4 to f_3 which would imply for a probability measure underlying these choices that $P(b) > P(r)$ and $P(r) + P(y) > P(b) + P(y)$, a direct contradiction.

like the 'maxmin expected utility' approach (Gilboa and Schmeidler 1989), the 'smooth ambiguity' model (Klibanoff, Marinacci and Mukerji 2005) or 'variational preferences' model (Maccheroni, Marinacci and Rusticchini 2006) require probabilistic information which – as we have just argued – may be unavailable or misleading. Second, these papers rationalize Ellsberg choices by incorporating an axiom of 'ambiguity aversion' into their framework. Yet, it is doubtable whether Ellsberg choices are a desirable feature of any theory of rational decision making, because they imply things such as aversion to information, updating of information based on taste or sensitivity to sunk costs (Al-Najjar and Weinstein 2009). Moreover, as Halevy (2007) has shown in his experimental re-examination of Ellsberg's findings, whether a test person expresses ambiguity aversion is correlated with that person's incapability to apply basic probability calculus. Hence, while these models are successful from a descriptive point of view, they are normatively unsatisfactory. Non-probabilistic models, which seem more in line with Keynes and Knight, naturally tend to be minimalistic in terms of informational requirements, but this comes at the price of taking a somewhat narrow perspective on possible states of the world. For example, the maximin criterion (Wald 1949) only focuses on the worst outcome and evaluates actions accordingly, the Hurwicz rule (Arrow and Hurwicz 1977) evaluates actions according to weighted average of worst and best possible outcome, clearly an unsatisfying limitation as Gravel, Marchant and Sen (2012) have pointed out. Other rules like the principles of minimum regret (Niehans 1948, Savage 1954) and insufficient reason (Laplace 1820) take all possible states of the world into account, but they lack a formal concept of 'uncertainty aversion'.

The contribution of the present paper is to address these points by providing a conceptually new and original approach to decision making under Knightian uncertainty. In the spirit of Knight (1921), we do not a priori incorporate the concept of probability. Instead, we focus on Knightian income lotteries – or, equivalently, Knightian acts – which are distributions of monetary payoffs over different possible outcomes with completely unknown objective probabilities. We start from a set of seven axioms about the preference relation $\succeq$ over Knightian acts, and prove that there exists a real-valued, additive and extensive function which numerically represents the uncertainty preference relation. Moreover, this function is unique up to linear-affine transformations. We propose Rényi's generalized entropy (Rényi 1961) as one possible such function. It contains a positive

real-valued parameter which we interpret as the decision maker's degree of uncertainty aversion. Conceptually, this implies that uncertainty aversion is a measure of how strong an individual dislikes spreads, i.e. unevenness, in monetary payoffs. We do not assume from the start that individuals are uncertainty averse, but much rather, it turns out, a very cautious attitude towards Knightian uncertainty is a natural consequence of our axiomatization. Moreover, unlike most approaches so far, we do not relax the Sure Thing Principle[4] (Savage 1954), instead we show that a Knightian version of it follows from the set of our base axioms.

Our paper is organized as follows: Section 2 clarifies some key concepts and points to open questions from the literature. In Section 3.1, we explain setting and notation. In Section 3.2, we state the seven base axioms and the main result, which is an existence and uniqueness of a numerical uncertainty utility index (cf. Proposition 1). In Section 4.1, we propose a particular function – Rényi's generalized entropy (Rényi 1961) – as one possible functional representation of the preference relation $\succeq$ on Knightian acts before we illustrate this utility index in Section 4.3 with a choice problem between three Knightian acts. We also compare the result with other decision rules that have been proposed in the context of Knightian uncertainty (cf. Polasky et al. 2011): maximin, maximax, the Hurwicz criterion, Laplace's principle of insufficient reason and the principle of minimum regret. Section 5 briefly discusses our findings and concludes.

2 Conceptual clarifications

We present some critical reflections on key concepts of decision theory and related literature in the following.

Risk, uncertainty and ambiguity. Zweifel and Eisen (2012) state that 'the risk of an activity is represented by the probability density $p(x)$ defined over possible consequences x' where consequences may mean utility levels or monetary payoffs. $p(x)$ may be exogenously specified or scientifically calculable objective probabilities (cf. Machina and Rothschild 2008). If we apply this definition to the Keynes-Knight definition of uncertainty, *Knightian uncertainty* would then just amount to the non-existence of such an

[4]The Sure Thing Principle is sometimes also referred to as 'independence of irrelevant alternatives' and states that $A + C \succeq B + C$ implies $A \succeq B$.

objective probability density function (PDF), whereas *ambiguity* would imply that there is at least incomplete knowledge concerning chances of outcomes or more than one PDF, and the decision maker is not sure about the 'true' distribution. In other words, if information concerning probabilities is partly missing or if there are several non-identical PDFs over consequences, possibly even weighted by some subjective weighting factors (2nd order probability distributions), then we face an ambiguous situation (cf. Gravel, Marchant and Sen 2012). It is worth noting that even though ambiguity and Knightian uncertainty are in principle very distinct concepts, they are often used interchangeably.

According to Machina and Rothschild (2008), there are two major theory strands concerning choice under Knightian uncertainty: the state-preference approach (Debreu 1959, Arrow 1964, Hirshleifer 1965, Hirshleifer 1966, Yaari 1969) and the hypothesis of probabilistic sophistication. The state-preference approach starts from a set of states of the world $\mathcal{S} = \{s_a, \ldots, s_n\}$ and constructs a theory of choice with state-payoff bundles $(c_1, \ldots, c_n)$ as objects of choice. Individuals are assumed to have preferences over state-payoff bundles just like regular commodity bundles. L.J. Savage's 1954 contribution was to define an 'act' as a mapping from states to consequences and that there exists a subjective belief, derived from preferences, which substitutes for objective probabilities. Much later, this framework was fortified by the hypothesis of probabilistic sophistication (Machina and Schmeidler 1992) which clarified the notion of subjective probabilities. It states that individuals entertain subjective probabilities which take the form of additive subjective probability measures $\mu(\cdot)$ over the state space $\mathcal{S}$. Anscombe and Aumann (1963) refined Savage's framework by assuming consequences to be risky lotteries rather than simple outcomes. Within this Anscombe-Aumann framework, Bewley (2002) introduced the assumption that individuals may assert that two alternatives are incomparable and that they may only accept an alternative when it is actually preferred to their current status quo. Bewley thus assumes that Knightian preferences are incomplete, an idea that we will revisit in this paper.

The literature strand that has spawned from the impact of the Ellsberg experiment has been subsumed under the umbrella term 'ambiguity aversion literature' (Al-Najjar and Weinstein 2009). Gilboa and Schmeidler (1989) and Schmeidler (1989) both expanded the Anscombe-Aumann framework to accommodate Ellsberg-type behavior, i.e. the preference of risk to uncertainty. Gilboa and Schmeidler (1989) provided an axiomatic foundation

of 'maximin expected utility' (MEU) with multiple priors over the state space, so that the utility of an act is the minimal expected utility resulting from the priors. Schmeidler (1989) introduced the mathematical concepts of capacities and Choquet integration to model ambiguity aversion. Finally, Klibanoff, Marinacci and Mukerji (2005) and Nau (2006) modeled ambiguity via second order probability distributions and ambiguity aversion over the concavity of some second order utility function, i.e. a utility of expected utilities. It is especially this approach that has been applied frequently in climate change economics and related policy analyses (Millner, Dietz and Heal 2010, Traeger 2011, Heal and Millner 2013).

Descriptive and normative decision theory. There is, it seems, a dichotomy of approaches in decision theory. On the one hand, there are descriptive approaches that try to incorporate behavioral findings into existing theories to 'bring theory closer to reality' (Gilboa 2010: 4). Such approaches will be helpful whenever one is interested in descriptive prediction of behavior under ambiguity or uncertainty. On the other hand, the normative approaches treat behavioral peculiarities in conflict with some of their axioms such as famously reported by Allais (1953) and Ellsberg (1961) as errors of human reasoning. The ultimate aim is thus to 'bring reality closer to theory' (ibid.) by pointing out these errors of reasoning to decision makers to make better decisions possible in the future. Such theories can help determining what ought to be done, given that the decision maker agrees with the theory's premises. Obviously, one can have problems with the inherent paternalism of such theories. On the other hand, proponents usually stress that only normative decision theories can help overcoming human thinking biases that irrefutably exist.

Risk and probability. Closely related to the notions of risk and ambiguity are issues of measurability of risk ('riskiness') and probability. These notions are ubiquitous in economics, and yet it seems that their usage can be problematic. Consider the decision of investing in asset A or asset B. The standard way of arguing here is that investment A is said to be riskier than B if the standard deviation of its market price trajectory is larger. This does not seem entirely convincing. As an illustrative example, assume that the choice is between Microsoft stocks and Greek state bonds. Following the standard argument would lead to the conclusion that Microsoft stocks are riskier than Greek state

bonds. From recent history, this statement seems questionable. The underlying issue however is whether and to what extent risk can be quantified, possibly even objectively.

If one worries about risk quantification, it entails thinking about probability quantification as well. Many approaches in economic theory require at some point the existence of probabilities that are objectively 'true'. Philosophical details with the concept of truth aside, such probabilities are unlikely to exist in most practical applications. And even if data is abundant, de Finetti's circularity critique[5], which argued that in order to define probabilities the classical or frequentist way, one needs to know the meaning of 'equally probable' first, seems valid (ibid.). De Finetti made these arguments in favor of Bayesian statistics, which states that in principle 'any uncertainty can and should be quantified' (cf. Gilboa 2010: 6). While this is arguably the predominant paradigm in economics to date, there is a catch here as well: Bayesian reasoning requires priors, which are highly subjective, leading to highly subjective results. As Feduzi and Runde (2011) have pointed out, the consequence might be to face a decision problem under Knightian uncertainty.

3 Uncertainty preferences and uncertainty aversion

In this section, we introduce our setting and notation and provide some basic definitions (Section 3.1), before we state a set of seven base axioms that can be shown to constitute an axiomatic foundation of a numerical preference index under Knightian uncertainty (Section 3.2). In Section 3.3, we provide a definition of uncertainty aversion and clarify how we can compare two persons' uncertainty attitudes. Finally, we provide some illustration for the special case of just two possible future outcomes in Section 3.4.

3.1 Setting, notation and basic definitions

We employ the framework of Savage (1954), where a simple act $f : X \to Y$ maps one particular vector of n states of the world $x \in X$ to monetary consequences $y \in Y \subseteq \mathbb{R}^n_+$.[6] Let $\mathcal{F}$ denote the set containing all simple acts for given n. Moreover, denote by $y^f =$

[5] "Therefore, these two ways of defining probability [...] are airy-fairy, unless one states beforehand what 'equally probable' means" (de Finetti 2008 [1979]: 4).

[6] If X is measurable, we have a situation of *risk*, if it is not, we have a situation of (Knightian) *uncertainty*.

$(y_1^f, \ldots, y_n^f) \in \mathbb{R}_+^n$ the monetary payoff distribution resulting from a specific act $f \in \mathcal{F}$ and let $\overline{y}^f = \sum_{i=1}^n y_i^f$ be the total payoff volume associated with f. Then, we denote by $\mathcal{G}(\overline{y})$ the subset of $\mathcal{F}$, the elements of which have a sum of payoffs over all states of the world of $\overline{y} > 0$, so that $\mathcal{G}(\overline{y}) = \left\{ f \in \mathcal{F} | \sum_{i=1}^n y_i^f = \bar{y} \right\} \subseteq \mathcal{F}$. We denote by f^c a constant act which yields the same payoff in every state of the world. Moreover, $s_i^f = y_i^f / \overline{y}^f$ is the payoff share act f yields in state of the world i with respect to the total payoff volume $\overline{y}^f$ so that, by construction, $0 \leq s_i^f \leq 1$ for all $i = 1 \ldots n$ and $\sum_{i=1}^n s_i^f = 1$ for any given $f \in \mathcal{F}$. Furthermore, denote by $S^n \subseteq [0,1]^n$ the set containing all possible such distributions s over n states of the world i, with any particular element $s = (s_1, \ldots, s_n) \in S^n$. Lastly, denote by $\underline{1}$ the vector $(1, \ldots, 1) \in [0,1]^n$, so that $\frac{1}{n} \cdot \underline{1}$ is the uniform distribution of shares over n states of the world.

We are concerned with decision problems under Knightian uncertainty here. Precisely, in our setting, an act f is equivalent to a Knightian income lottery since a vector of states of the world is assigned a vector of monetary payoffs y^f via f where the probabilities p_i of the outcomes y_i are unknown. In the literature, the terms 'act' and '(Knightian) lottery' are often used interchangeably. Here, we will stick to 'act' in the following. Furthermore, we employ the usual notation for preferences over acts: $f \succeq g$ where $\succeq$ means 'is at least as good as'. If $f \succeq g$ and $g \succeq f$, then we write $f \sim g$ and say that 'f is as good as g' (indifference). Lastly, if $f \succeq g$ and not $g \succeq f$, we write $f \succ g$ and say that 'f is strictly preferred to g'. Before we state axioms about the preference relation $\succeq$ on uncertain income lotteries in the following Section 3.2, we introduce some basic definitions.

Definition 1 (scaled Knightian act)**.** *For all $\beta > 0$ and all Knightian acts $f \in \mathcal{F}$ with $y^f = (y_1^f, \ldots, y_n^f)$, we define the corresponding scaled Knightian act βf as the act with payoffs $\beta y^f = (\beta y_1^f, \ldots, \beta y_n^f)$.*

The positive scaling parameter β is hence a scalar multiplier that modifies the payoffs in each state of the world. The scaled act βf features the same relative distribution of payoff shares as f, $s^f = s^{\beta f}$, while it holds for the total payoff volumes that $\overline{y}^{\beta f} = \beta \overline{y}^f$.

One of the work horses of decision theory under risk is the idea of a compound act. Usually, compounding is interpreted as a multi-stage act, i.e. an act that has again acts as payoffs, such as flipping a coin to determine which act to play next. In this paper, we

provide a different definition similar to a concept that was introduced by Luce (1972) as 'concatenation'.

Definition 2 (compound Knightian act)**.** *For any two simple acts f,$g \in \mathcal{F}$ with $y^f = (y_1^f, \ldots, y_n^f)$, $y^g = (y_1^g, \ldots, y_n^g) \in Y$ and associated s^f, $s^g \in S^n$ the compound act $f \oplus g$ is defined as the act with $\overline{y}^{f \oplus g} = \overline{y}^f + \overline{y}^g$ and $s^{f \oplus g} = \left(s_1^f s_1^g, \ldots, s_1^f s_n^g, s_2^f s_1^g, \ldots, s_2^f s_n^g, \ldots, s_n^f s_1^g, \ldots, s_n^f s_n^g\right) \in S^{nn}$.*

We write a $\oplus$ sign between two acts to be compounded to distinguish the compounding operation from standard addition. Note that our compounding operation implies two things: (1) pathwise multiplication of payoff shares, and (2) addition of payoff volumes of acts that are compounded. Our definition of a compound act therefore deviates from the usual definition, because we cannot use probabilities, which are a key ingredient when compounding risky acts.

While the algebraic structure resulting from Definition 2 mirrors the algebraic structure resulting from the classic compounding definition under risk, the economic interpretation is quite different. Classically, compounding is interpreted as a two-stage act, where prizes of the first stage – for example flipping a coin – are again acts – for example drawing a colored ball from one out of two different urns, and receiving prize money according to the urn and color of the ball drawn. In other words, the outcome of the first stage – for example 'heads' or 'tail' – determines which act is played in the second stage – for example, urn 1 or urn 2. In contrast, here, the compound act $f \oplus g$ represents a situation where both simple acts f and g are played for sure one after the other, but prize money is paid out only after the second act is played. Moreover, prize money payoffs over compound states are determined by adding up total payoff volumes of both acts, and distributing them over the set of compound states of nature, the shares of which are obtained by multiplying shares of the simple acts pathwise.[7] While the compound states have formally the same Cartesian product structure as in the standard compounding operation in risk theory, the interpretation of how prize money is paid out is different here.

[7]Technically, it is not imperative that the dimensions of the acts that are compounded are equal. Definition 2 could easily be generalized to admit for $y^f \in Y^f \subseteq \mathbb{R}_+^n$ and $y^g \in Y^g \subseteq \mathbb{R}_+^m$ with $n \neq m$.

3.2 Axioms and main result

Our formulation here makes use of what is known as the Lieb-Yngvason formulation of statistical thermodynamics (Lieb and Yngvason 1999). We transfer their main result and its proof and reinterpret it within a setting of decision-making under Knightian uncertainty. This leads to a proposition on the existence and uniqueness of a general preference representation under Knightian uncertainty.

In the following, we state the axioms that we impose on the preference relation $\succeq$ on the set of acts $\mathcal{F}$. Recall that $\mathcal{F}$ denotes the set of acts mapping from X to $Y \subseteq \mathbb{R}^n_+$. Hence, when we talk about the preference relation '$\succeq$', this is a preference relation on (some) acts in a particular dimension n. This entails that the formalism presented here is only applicable within one specific set of acts $\mathcal{F}$ where the dimension of $\mathcal{F}$ is necessarily inherited to its subset $\mathcal{G}(\overline{y})$.

We henceforth understand a decision maker faced with Knightian uncertainty as 'rational' as someone agreeing on the following seven axioms. We give explanations and illustrations where we deviate from what can safely be regarded as standard in economic theory (cf. e.g. Savage 1954, Anscombe and Aumann 1963, Gilboa and Schmeidler 1989).

Axiom 1 (Reflexivity)**.** *For all $f \in \mathcal{F}$, $f \sim f$.*

Axiom 2 (Transitivity)**.** *For all f, g, $h \in \mathcal{F}$, $f \succeq g$ and $g \succeq h$ implies $f \succeq h$.*

Axiom 3 (Completeness on $\mathcal{G}(\overline{y})$)**.** *For all $\overline{y} > 0$ and all f, $g \in \mathcal{G}(\overline{y})$, either $f \succeq g$ or $g \succeq f$, or both.*

The completeness axiom might look at first sight more innocent and standard than it actually is. In fact, by assuming completeness of the preference relation $\succeq$ on the set $\mathcal{G}(\overline{y})$ rather than the entire set $\mathcal{F}$, we in fact implicitly *drop* the completeness assumption for the whole act set, and only assume completeness for the subset $\mathcal{G}(\overline{y})$, a notion of 'bounded rationality' in the presence of Knightian uncertainty already put forward by Bewley (2002). Of course, this axiom does *not* state that a decision maker is generally unable to express his relative preference of any two acts from $\mathcal{F}$. For example, if $a \in \mathcal{F}$ dominates $b \in \mathcal{F}$, the decision maker will certainly be able to express his preference of a over b.

Axiom 4 (Consistency)**.** *For all f,f',g,$g' \in \mathcal{F}$, $f \succeq f'$ and $g \succeq g'$ implies $f \oplus g \succeq f' \oplus g'$.*

The consistency assumption means that preferences over simple acts carry over to their respective compounds. That is, if some act f is weakly preferred to f' and another act g is weakly preferred to g', then the compound act $f \oplus g$ is also weakly preferred to the compound act $f' \oplus g'$. Hence, the compound act obtained from two acts that are weakly preferred individually over two other acts will also be weakly preferred to the compound act of these other two acts.

Axiom 5 (Scaling invariance)**.** *For all $\beta > 0$ and all f, $g \in \mathcal{F}$, $f \succeq g$ implies $\beta f \succeq \beta g$.*

While the consistency assumption (Axiom 4) refers to consistency in terms of compound acts, the scaling invariance assumption refers to consistency over scales. This is to say that a statewise proportional change in payment levels and, hence, total payoff volumes does not alter the preference ordering. For example, if the act (1$,2$) is at least as good as (0.50$,2.50$) to an individual, then, taking for example $\beta = 50$, the very same individual should also prefer (50$,100$) to (25$,125$).

Axiom 6 (Splitting and recombination)**.** *For all $f \in \mathcal{F}$ and $0 < \beta < 1$:* $f \sim \beta f \oplus (1 - \beta) f$.

Hence, it should not matter to the rational decision maker whether she plays some act f or the compound act consisting of some scaled-down versions of f, where the scaling factors add up to one. In terms of economic intuition and in light of how we have defined compounding (Definition 2), this means that we assume the decision maker to be indifferent between a two-stage Knightian act where prizes are scaled versions of some Knightian act f, with scaling factors adding up to one, and the one-stage act f itself.

Axiom 7 (Continuity)**.** *For all f,g,h_0,$h_1 \in \mathcal{F}$ and a sequence ε_k with $\lim_{k\to\infty} \varepsilon_k = 0$,*

$$f \oplus \varepsilon_k h_0 \succeq g \oplus \varepsilon_k h_1 \quad \textit{for } k \to \infty \textit{ implies} \quad f \succeq g.$$

The continuity assumption guarantees that there are no discontinuities in the preference relation which means that the presence of 'perturbation acts' with small scales (in the sense of Axiom 5) tending to zero does not induce some sort of spontaneous preference reversal. The formulation here is essentially the same as given in Rubinstein (2006), for example.

A notable point of our axiomatic framework is that we do *not* assume what is commonly referred to as the 'Sure Thing Principle' (Savage 1954). We establish this, that is to say its Knightian equivalent, in the following lemma.

Lemma 1 (Knightian independence)**.** *Let f,g,$h \in \mathcal{F}$. Then it follows from Axioms 1,2 and 4 through 7 that*

$$f \oplus h \succeq g \oplus h \quad \textit{implies} \quad f \succeq g \ .$$

Proof. See Appendix A. □

Lemma 1 states that six out of our seven axioms – i.e. all but completeness – imply Knightian independence, but not vice versa. In other words, our axioms here are more restrictive than the independence assumption alone. Lemma 1 could be interpreted as 'things occurring anyway do not affect one's preference ordering'. This property is generally considered a very important feature of theories of rational choice. However, it has been pointed out by Al-Najjar and Weinstein (2009) to pose a problem to the ambiguity aversion literature as a normative theory, since ambiguity averse agents can be shown to violate this principle. On the descriptive level, people have been shown to systematically violate this principle under certain conditions which is known as Allais paradox (Allais 1953).

We have now gathered all ingredients to establish our main result in Proposition 1.[8]

Proposition 1 (Existence and uniqueness of an uncertainty preference index)**.** *Let $\succeq$ be a binary relation on $\mathcal{F}$. Then, for all $\overline{y} > 0$ and for all $f_1 \dots f_N, f'_1 \dots f'_M \in \mathcal{G}(\overline{y})$ the following statements are equivalent:*

1. *$\succeq$ satisfies Axioms 1–7.*
2. *There exists a continuous function $H^n_{\overline{y}} : S^n \to \mathbb{R}$ that characterizes $\succeq$ in the following sense: for all $N \geq 1$, $M \geq 1$, $\beta_i \geq 0$ and $\beta'_j \geq 0$ with $\beta_1 + \ldots + \beta_N = \beta'_1 + \ldots + \beta'_M$,*

$$\beta_1 f_1 \oplus \ldots \oplus \beta_N f_N \succeq \beta'_1 f'_1 \oplus \ldots \oplus \beta'_M f'_M \tag{1}$$

[8] On a technical side remark, Lieb and Yngvason (1999) argue that the completeness axiom can be shown to follow from a number of more basic principles. However, this would mean that completeness would have to be established in a separate line of argument, and using additional axioms, first. The number of axioms needed altogether would then come to 15 instead of just the seven presented above.

holds if and only if

$$\sum_{i=1}^{N} \beta_i H_{\overline{y}}^n(s^{f_i}) \geq \sum_{j=1}^{M} \beta'_j H_{\overline{y}}^n(s^{f'_j}) \ . \tag{2}$$

$H_{\overline{y}}^n$ is unique on $\mathcal{G}(\overline{y})$ up to linear-affine transformations $\hat{H}_{\overline{y}}^n(s) = aH_{\overline{y}}^n(s) + b$ where $a, b \in \mathbb{R}$ and $a > 0$.

Proof. See Appendix B. □

The imposition of Axioms 1 through 7 on the preference relation $\succeq$ hence implies the existence of a function that maps from the set of relative payoff share distributions S^n associated with $\mathcal{G}(\overline{y})$ to the real numbers such that the most preferred Knightian act out of $\mathcal{G}(\overline{y})$ is assigned the greatest real number (Appendix B, Lemma 6 – 8), and this function is unique up to linear-affine transformations (Appendix B, Lemma 9). We therefore call the function $H_{\overline{y}}^n : \mathcal{S}^n \to \mathbb{R}$ the uncertainty utility index on $\mathcal{G}(\overline{y})$ of the decision maker. Note that such an index exists for all sets $\mathcal{G}(\overline{y})$. In the following, we state some additional properties of every function $H_{\overline{y}}^n$ that follow immediately from Proposition 1.

Corollary 1 (Properties of uncertainty utility index)**.** *Every function $H_{\overline{y}}^n$ in the sense of Proposition 1 has the following properties for all f,f',$g \in \mathcal{G}(\overline{y})$, $s^f, s^{f'}, s^g \in S^n$:*

1. *Simple act representation: $f \succeq f'$ if and only if $H_{\overline{y}}^n(s^f) \geq H_{\overline{y}}^n(s^{f'})$.*
2. *Additivity: $H_{\overline{y}^{f\oplus g}}^{nn}(s^{f\oplus g}) = H_{\overline{y}^f}^n(s^f) + H_{\overline{y}^g}^n(s^g)$.*
3. *Extensitivity: $H_{\overline{y}^{f\oplus f}}^{nn}(s^{f\oplus f}) = H_{\overline{y}^f}^n(s^f) + H_{\overline{y}^f}^n(s^f) = 2H_{\overline{y}^f}^n(s^f)$.*
4. *Maximality: $H_{\overline{y}}^n(\frac{1}{n}\underline{1}) > H_{\overline{y}}^n(s^f)$ for all $f \neq f^c$.*

Proof. Monotonicity: follow directly from Proposition 1 by setting $N = M = 1$ and $\beta = \beta'$. Additivity and extensitivity also follow directly from Proposition 1. Maximality is a consequence of the central postulate of statistical physics (cf. Lieb and Yngvason 1999). □

Monotonicity is the usual understanding of preference representation through an index function. It is included as a special case in the more general understanding of preference

representation of Proposition 1. Additivity states that the utility derived from a compound act is the sum of utilities of corresponding simple acts. The additivity property must be seen in light of our definition of compounding as playing the simple acts consecutively (Definition 2). Thus, following the logic of consequentialism, there is no loss or gain in terms of utility to be made from the process of compounding itself.[9] Formally, the extensitivity property is thus a special case of additivity within our framework. Both properties, additivity and extensitivity, reflect the Cartesian product structure of the state space of compound systems. Lastly, maximality implies that the constant act is always the most preferred act on any set $\mathcal{G}(\overline{y})$.

In statistical physics and information theory, a function with these properties is called *additive entropy*. Essentially, the uncertainty utility function $H^n_{\overline{y}}$ is, thus, an additive entropy function. As entropy is a statistical measure of the homogeneity of a distribution, the interpretation of our representation of uncertainty preferences is quite simple and intuitive: of two acts with equal total payoff volume over all potential outcomes, the one with the more homogenous distribution (i.e. higher entropy) of relative payoff shares is preferred.

3.3 Definition of uncertainty aversion

We now formalize the notion of aversion to Knightian uncertainty and clarify how to compare two individuals in terms of their degree of uncertainty aversion.

Definition 3 (uncertainty aversion)**.** *A decision maker is said to be uncertainty averse for given $\overline{y} > 0$ if and only if $f^c \succ f$ for all $f, f^c \in \mathcal{G}(\overline{y})$ with $f \neq f^c$.*

Definition 3 defines an uncertainty-averse individual with respect to a specific set $\mathcal{G}(\overline{y})$ as someone who always prefers certainty to even the slightest uncertainty as expressed by the non-uniform distribution of the total payoff volume $\overline{y}$ over possible states of the world. For example, if $n = 2$ and $\overline{y} = 10\$$, an uncertainty-averse individual would always prefer the constant act $f^c = (5\$,5\$)$ to any other element from $\mathcal{G}(\overline{y})$, such as $(3\$,7\$)$ or

[9] Luce et al. (2008) have used an entropy-based modelling approach to account for the utility drawn from the process of compounding itself (which they refer to as gambling), but their approach takes place within expected utility.

(1$,9$). The cases of uncertainty neutrality and an uncertainty-loving attitude can then be defined in the obvious way.

In risk theory, the maximum willingness to pay for certainty in the context of a given act is captured by the risk premium of that act, which is the difference between expected payoff from the act and the individual's certainty equivalent. Here, we can define the concept of an uncertainty premium in a similar manner using the idea of a *uniformly-distributed equivalent* developed by Atkinson (1970) in the context of social welfare theory.[10]

Definition 4 (uncertainty premium)**.** *Assume that for any $f \in \mathcal{F}$ there exists exactly one uniformly-distributed equivalent act $f^{\mathrm{UDE}} = \eta^f \cdot \underline{1} \in \mathcal{G}(\overline{y}^{f^{\mathrm{UDE}}})$ with $\eta^f := \overline{y}^{f^{\mathrm{UDE}}}/n$ such that $f^{\mathrm{UDE}} \sim f$. Then the uncertainty premium Π^f of act f is*

$$\Pi^f := \frac{1}{n}\left(\overline{y}^f - \overline{y}^{f^{\mathrm{UDE}}}\right) . \tag{3}$$

The crucial assumption here is that such a uniformly-distributed equivalent act f^{UDE} exists for any $f \in \mathcal{F}$, i.e. there always exists a positive amount of money that leaves the decision maker indifferent between playing f and receiving η^f for sure. Thus, f^{UDE} is a special constant act, which also fulfills this indifference condition. A person is uncertainty averse (neutral, loving) with regard to act f if and only if $\Pi^f > 0$ $(= 0, < 0)$.[11] These observations naturally allow us to compare uncertainty attitudes of different persons.

Definition 5 (Interpersonal comparison of uncertainty aversion)**.** *For given $\overline{y} > 0$ and two individuals A and B with uncertainty preference relations $\succeq_A$ and $\succeq_B$, A is said to be more uncertainty averse than B if and only if $\Pi^f_A > \Pi^f_B$ for all $f \in \mathcal{G}(\overline{y})$.*

This means that for A to be more uncertainty averse than B for given $\overline{y}$, A's uncertainty premium needs to be greater than B's for every act from $\mathcal{G}(\overline{y})$. It is of course

[10] Atkinson (1970) speaks of *egalitarian equivalence*. When transferring his formal idea from the context of income distribution within a society to the context of decision-making under uncertainty and, hence, distribution of payoffs over different potential outcomes, the notion of "egalitarian" does not fit any more. Hence, we have replaced it by "uniform distribution".

[11] While uncertainty-neutral and uncertainty-loving preferences are perfectly well conceivable, we cannot model these within our formal framework, because each entropy function becomes maximal for the constant act from any given $\mathcal{G}(\overline{y})$ (cf. Corollary 1). That means, entropic uncertainty utility functions as given by Axioms 1–7 always represent uncertainty aversion. See Sections 3.4 and 4 for details.

possible that for some $\overline{y}$, A is more uncertainty averse than B, while for some other $\overline{y}'$, it is the other way round. Uncertainty aversion and particularly its magnitude are thus local characteristics of a person (i.e. depending on the level of payoff) rather than global ones.

3.4 Special case: $n = 2$

As an instructive special case, we consider a two-state world ($n = 2$). Instead of analyzing $H^2_{\overline{y}}(s_1,s_2)$ subject to the constraint $s_1 + s_2 = 1$, we may as well equivalently analyze the function $\tilde{H}^2_{\overline{y}}(s) \equiv H^2_{\overline{y}}(s,1-s)$, where $s_1 = s$ and $s_2 = 1 - s$ so that the explicit constraint is implicitly contained. As any function $H^n_{\overline{y}}$ is only defined where $\sum_i s_i = 1$ holds, the two functions $H^2_{\overline{y}}$ and $\tilde{H}^2_{\overline{y}}$ are fully equivalent and have the same characteristics. For ease of notation, we will refrain from indexing utility functions in the remaining part of this section. For $n = 2$, it is possible to transfer some concepts of risk aversion from risk theory, and to conveniently illustrate them with two-dimensional graphs. For example, we can suggest an inverse measure of absolute uncertainty aversion similar to the Arrow-Pratt measure of absolute risk aversion (Pratt 1964, Arrow 1965):

Definition 6 (inverse measure of absolute uncertainty aversion)**.** *For $n = 2$ and any uncertainty utility function $H(s_1,s_2)$ in the sense of Proposition 1, we employ the equivalent utility function $\tilde{H}(s) \equiv H(s,1-s)$ with $s_1 = s$ and $s_2 = 1 - s$, which we assume to be strictly increasing on [0,0.5], to define on [0,0.5]*

$$\Lambda(s) := -\frac{\tilde{H}''(s)}{\tilde{H}'(s)} \tag{4}$$

as an inverse measure of absolute uncertainty aversion for $\tilde{H}(s)$.

As shown in Pratt (1964), this measure is invariant under linear-affine transformations, a feature which makes it attractive for our theory as well. However, two remarks seem appropriate: First, while the measure Λ of uncertainty aversion can be nicely illustrated and interpreted for $n = 2$, it is unclear whether a similar measure exists for $n > 2$. Formal generalizations to higher dimensions have been proposed and discussed (e.g. Duncan 1977, Levy and Levy 1991), but so far only for certain special cases. Second, just like the utility

function $H_{\overline{y}}(s)$, the measure $\Lambda(s)$ is a function of relative payoff shares s and not absolute payoffs y.

Consider the following thought experiment. For given $\overline{y} > 0$, two uncertainty-averse persons A and B are offered a sequence of choices. In each round, A and B can choose between playing an act from the set $\mathcal{G}(\overline{y})$ or to accept a payment of η with $0 \leq \eta < \overline{y}/2$ for sure. In the first round ('base scenario'), the act offered is the constant act $(\overline{y}/2, \overline{y}/2)$, which corresponds to certainty. In each subsequent round, the uncertain act offered becomes continuously more uncertain by featuring a more unequal payoff distribution over the two potential states, that is, an increased possible win over the base scenario in case of the good outcome and an accordingly increased loss in the bad outcome. The certain payment η remains the same in all rounds. As both persons are uncertainty averse and η is strictly smaller than the payoff from the constant act in the base scenario, $\overline{y}/2$, in the base scenario both persons will choose playing the act rather than accepting η for sure. They will continue choosing the uncertain act, depending on the exact level of the certain payment η, while the uncertainty is still small. But as the offered act becomes continuously more uncertain from round to round while the certain payment η remains constant, there will eventually be a level of uncertainty, where one person, say A (without loss of generality), is just indifferent between playing the uncertain act and accepting the certain payment η. For all subsequent rounds, this person will then choose the certain payment η over playing the uncertain act.

If the switching-level of uncertainty at which a person switches from playing the uncertain act to accepting the certain payment η is different between persons A and B, say A switches already at a smaller level of uncertainty than B, and this holds for every admissible level of η, then this can be used to label person A as more uncertainty averse (in the sense of Definition 5) than B. The following lemma states that the measure Λ (from Definition 6) provides such an interpersonal comparison.

Lemma 2 (Interpersonal comparison of uncertainty aversion for $n = 2$)**.** *Let A and B be two uncertainty-averse individuals that are offered the choice of either playing uncertain acts from $\mathcal{G}(\overline{y})$ for given $\overline{y} > 0$, or to accept a certain payment of η with $0 \leq \eta < \overline{y}/2$*

instead. Denote by y_H and y_L the two outcomes of uncertain acts offered, where $y_L + y_H = \overline{y}$ and $y_H > y_L$. Assuming the correspondence $\Pi \propto y_H - y_L$, it holds that

$$\Lambda_B(s) \geq \Lambda_A(s) \quad \Longleftrightarrow \quad \Pi_B(s) \leq \Pi_A(s) \quad \forall s \in [0{,}0.5]\,.$$

Proof. See Appendix C. □

Lemma 2 thus establishes Λ as an (inverse) measure of the degree of uncertainty aversion of a decision maker for the case $n = 2$: the larger Λ, the smaller the degree of uncertainty aversion. Because of its invariance under linear-affine transformations, it can be used for interpersonal comparison of uncertainty aversion.

We illustrate Lemma 2 and the underlying thought experiment in Figure 1. At first sight, the figure looks much like the well-known graph from risk theory, but there are some important differences. First and foremost, the horizontal axis displays s and thus, each point on it fully represents an act, rather than money values or wealth levels. Second, because both individuals are uncertainty-averse, both utilities become maximal for $s^c = 1/2$, which of course represents the constant act. For drawing the graph in a handy manner, we exploit invariance of the utility function $H(s)$ up to linear-affine transformations, so that $\tilde{H}_A(0) = \tilde{H}_B(0) = 0$ and $\tilde{H}_A(s^c) = \tilde{H}_B(s^c)$ without loss of generality. Third, a change of sign in Λ does *not* mean a change from uncertainty-averse to uncertainty-loving preferences, which is an important difference to the case of risk. Instead, a change in Λ only means a change in the degree of uncertainty aversion. Generally, smaller values of Λ reflect more uncertainty-averse preferences, and negative values of Λ mean an even higher uncertainty-aversion than positive values.

Furthermore, if we denote by s_{i*} the act at which person i is just indifferent between uncertain act and certain payment η, then we know that $\tilde{H}_A(s_{A*})$ must be equal to the utility drawn from the certain payment η.[12] At this point s_{A*} however, and at the level of the certain payment η offered, person B still prefers the uncertain act to the certain payment η. For given η, it is only at a higher level of uncertainty, $s_{B*} < s_{A*}$, that B is just indifferent between the uncertain act represented by s_{B*} and the certain payment η. What's more, in order for person B to be just indifferent between the uncertain act

[12] We may assert this because Definition 5 warrants such a statement even though it is naturally true that the certain payment η belongs to $\mathcal{G}(2\eta) \neq \mathcal{G}(\overline{y})$.

represented by s_{A*} and some certain payment, it would take a certain payment larger than just η, say $\eta + \xi$. The extra amount of money ξ is equivalent to the difference in utility levels $H_B(s_{A*}) - H_B(s_{B*})$. Hence, B's uncertainty premium at the act represented by s_{A*} is smaller than A's. Because person B's utility curve lies above person A's utility curve for all $s \in [0,s^c]$ and therefore, for all acts from the specific set $\mathcal{G}(\overline{y})$, B's uncertainty premium will be smaller for all acts from $\mathcal{G}(\overline{y})$. Thus, according to Definition 5, B can be said to be less uncertainty averse than A.

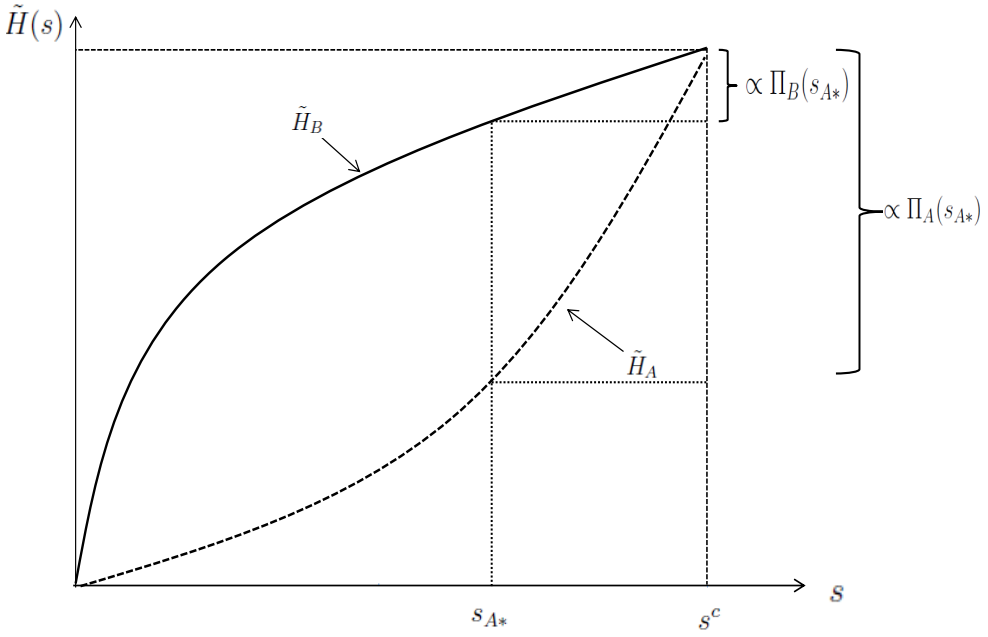


Figure 1: Uncertainty utility curves for two uncertainty-averse individuals, A and B, with differing degrees of uncertainty aversion. Every point on the s axis represents a different act from one particular set $\mathcal{G}(\overline{y})$. s_{A*} is the point at which A is just indifferent between the uncertain act represented by s_{A*} and the certain payment η. Individual i's uncertainty premium Π_i for any act on the s axis cannot be directly drawn into the graph, but it is proportional to the difference in utility from $H(s^c)$, as indicated for the point s_{A*} in the figure.

4 Example: A one-parameter function

We propose a generalized one-parameter entropy measure known from statistical physics and information theory – Rényi's (1961) generalized entropy – as one possible functional representation of preferences satisfying Axioms 1–7 (Section 4.1). The positive, real-valued parameter represents the individual's degree of uncertainty aversion. We subsequently illustrate the behavior of the proposed preference index with a stylized simple decision problem between simple acts and compare it to other decision criteria such as the maximin, maximax, minimum regret rules, the Hurwicz criterion and the Laplace Principle in Section 4.3. In this section, we are concerned with decisions between simple acts. That means, we have n states of the world and a given value for $\overline{y}$. We therefore omit the subscript $\overline{y}$ in the following.

4.1 Properties of Rényi's generalized entropy

We start by technically introducing Rényi's (1961) generalized entropy function, before we elaborate on its interpretation in the context of modeling preferences and decision making under Knightian uncertainty.

Definition 7 (Rényi entropy)**.** *For $n \in \mathbb{N}$, $s \in S^n$ and $\alpha > 0$, the Rényi entropy of order α is the function $H_\alpha^n : S^n \to \mathbb{R}$ with*

$$H_\alpha^n(s) = \begin{cases} \frac{1}{1-\alpha} \ln\left(\sum_{i=1}^n s_i^\alpha\right) & : \alpha > 0, \quad \alpha \neq 1 \\ -\sum_{i=1}^n s_i \ln s_i & : \alpha = 1 \end{cases}. \tag{5}$$

The expression for $H_1^n(s)$ is the continuous extension of the general expression of H_α^n for the limit $\alpha \to 1$. It has been proposed independently by Shannon (1948) and Wiener (1948). It is sometimes referred to as Shannon-Weaver-entropy because it has been popularized by Shannon and Weaver (1949).[13] Other notable special cases are $H_0^n(s) = \ln n$, the Hartley entropy (Hartley 1928), and $H_\infty^n(s) = \min_i\{-\ln s_i\} = -\ln\left(\max_i\{s_i\}\right)$, which

[13]The base of the logarithm used to calculate the entropies can be arbitrarily chosen. Rényi (1961) introduced his generalized entropy using the ld function, i.e. $\log_2$. Naturally, the choice of a particular base does not affect any result as long as the same base is used consistently.

is also known as min-entropy. The following lemma connects the decision framework from Section 3 with Rényi's generalized entropy.

Lemma 3 (Numerical representation of Axioms 1–7)**.** *Rényi's generalized entropy is one possible representation of uncertainty preferences satisfying Axioms 1–7.*

Proof. Using Definitions 1 and 2 as well as Lemma 4, this is an insertion exercise. □

Thus, because Rényi's generalized entropy fulfills Axioms 1–7, we can interpret it as one possible representation of an individual's preferences under Knightian uncertainty. Being one particular such function, it has all the properties that additive entropy functions have in general: continuity, monotonicity, additivity and extensivity (Proposition 1 and Corollary 1). The next lemma states some specific properties of Rényi's generalized entropy.

Lemma 4 (Properties of Rényi's generalized entropy)**.** *In addition to the properties stated in Proposition 1 and Corollary 1, Rényi's generalized entropy (Equation 5) has the following properties for all $\alpha > 0$, $n \in \mathbb{N}$, $s \in S^n$ and for every permutation matrix P:*

1. *Symmetry:* $H_\alpha^n(s) = H_\alpha^n(Ps)$.
2. *Maximality:* $H_\alpha^n(\frac{1}{n}\underline{1}) = \ln n > H_\alpha^n(s)$ *for every* $s \in S^n \setminus \left\{\frac{1}{n}\underline{1}\right\}$.
3. *Minimality:* $H_\alpha^n(Ps) = 0$ *for* $s = (1,0,\ldots,0)$.
4. *Dependence on* α: $\frac{\mathrm{d}}{\mathrm{d}\alpha} H_\alpha^n(s) < 0$.

Proof. Symmetry simply carries over from the underlying functions, ln and the summation. Minimality can be verified by insertion, and the maximality statement directly follows from solving the associated optimization problem. The proof of the dependence-on-α property can be found in Beck and Schlögl (1993). □

The formal properties of Rényi's generalized entropy stated in Lemma 4 can be interpreted in terms of uncertainty preferences. The symmetry property states that the sequence of the payoff shares that result from an act does not affect the value of H_α^n so that it does not matter in what sequence these shares are numbered. With regard to de-

cision theory, this is a central assumption with regard to the decision maker's preferences, which will be discussed in greater detail in Section 5. For now, we just note that it implies that the Rényi individual is probabilistically sophisticated (Machina and Schmeidler 1992) with uniform subjective beliefs, i.e. the decision maker implicitly applies the Laplacian Principle of Insufficient Reason (Laplace 1820). The maximality property tells us that H_α^n reaches its unique maximum for a completely uniform distribution. This maximum value is equal to $\ln n$ and hence independent of α. As discussed in the context of Corollary 1 in Section 3.2, maximality represents a strict preference of certainty to uncertainty on any set $\mathcal{G}(\overline{y})$. Conversely, acts where the payoff volume is concentrated in just one state are always least preferred (minimality). The last property, dependence on α, is directly relevant for modelling uncertainty aversion. It means that, for any given act $f \in \mathcal{G}(\overline{y})$ of arbitrary dimension n, it holds that $H_0^n(s^f) > H_1^n(s^f) > \ldots > H_\infty^n(s^f)$. That is, for given act and, thus, uncertainty, the greater the parameter α in the utility function, the smaller the resulting value of the utility function. This leads to the following statement.

Proposition 2 (Uncertainty-aversion parameterization in Rényi's generalized entropy)**.** *The parameter α in Rényi's generalized entropy (Equation 5) measures uncertainty aversion: the greater the parameter α in the utility function, the smaller – ceteris paribus – the resulting value of the utility function.*

Proof. Assume that two individuals, A and B, who accept Axioms 1–7 have Rényi's generalized entropy as representation of their uncertainty preferences. Lemma 4 then holds, especially Property 4, dependence-on-α. Then, for all $n \in \mathbb{N}$, $\overline{y} > 0$, and all $f \in \mathcal{G}(\overline{y})$, $H_{\alpha_A}^n(s^f) < H_{\alpha_B}^n(s^f)$ is equivalent to $\alpha_A > \alpha_B$. Hence, for all $s \in S^n$ the graph of utility function of individual A is below that of individual B if and only if $\alpha_A > \alpha_B$. Arguing along the same line as in the thought experiment from Section 3.4 (illustrated graphically in Figure 1), based on a suitable generalization of "more uncertain" from $n = 2$ to any n, this means that A is more uncertainty averse than B. □

4.2 Special case: $n = 2$

As previously, we employ the special case of a two-state world ($n = 2$) to provide a neat illustration. Hence, we plot in Figure 2 the fully-equivalent function

$$\tilde{H}^2_\alpha(s) \equiv H^2_\alpha(s,1-s) = \frac{1}{1-\alpha}\ln\left[s^\alpha + (1-s)^\alpha\right]; \quad \alpha > 0 \neq 1. \tag{6}$$

It can easily be verified that $\tilde{H}^2_\alpha(s)$ has all characteristics stated in Lemma 4. It is thus still an entropy function, but with effectively reduced number of arguments.

Figure 2 illustrates that the numerical utility that a decision maker attaches to a Knightian act critically depends on the parameter α. Observe that utility curves are symmetric to $s = 0.5$ by construction (cf. Lemma 4, Property 1). For $0 < \alpha < 2$, we observe that $\tilde{H}'_\alpha(s) > 0$ and $\tilde{H}''_\alpha(s) < 0$ for every $s \in [0, 0.5]$, while for $\alpha > 2$, it holds that $\tilde{H}'_\alpha(s) > 0$ and $\tilde{H}''_\alpha(s) > 0$ for s smaller than some threshold value s_T, for which $\tilde{H}''_\alpha(s_T) = 0$ holds, and $\tilde{H}''_\alpha(s) > 0$ for $s > s_T$. Our measure of absolute uncertainty aversion from Definition 6 then becomes

$$\Lambda(s) = \frac{\frac{s^\alpha\alpha}{s} - \frac{(1-s)^\alpha\alpha}{1-s}}{s^\alpha + (1-s)^\alpha} - \frac{\frac{s^\alpha\alpha^2}{s^2} - \frac{s^\alpha\alpha}{s^2} + \frac{(1-s)^\alpha\alpha^2}{(1-s)^2} - \frac{(1-s)^\alpha\alpha}{(1-s)^2}}{\frac{s^\alpha\alpha}{s} - \frac{(1-s)^\alpha\alpha}{1-s}}, \quad \alpha > 0 \neq 1. \tag{7}$$

We plot $\Lambda(s)$ in Figure 3 for different values of the parameter α. We observe that the value of Λ at any point s is determined by α and s and that, for all s, smaller values of α imply greater values of Λ (cf. Lemma 4, Property 4). This boils down to the following statement.

Lemma 5 (α as measure of uncertainty aversion)**.** *For any two individuals A and B with Rényi preference representations, it holds that $\Lambda_B(s) > \Lambda_A(s)$ if and only if $\alpha_B < \alpha_A$. That is, the greater α, the greater the degree of uncertainty aversion.*

Proof. Follows directly from Lemma 2 and Proposition 2. □

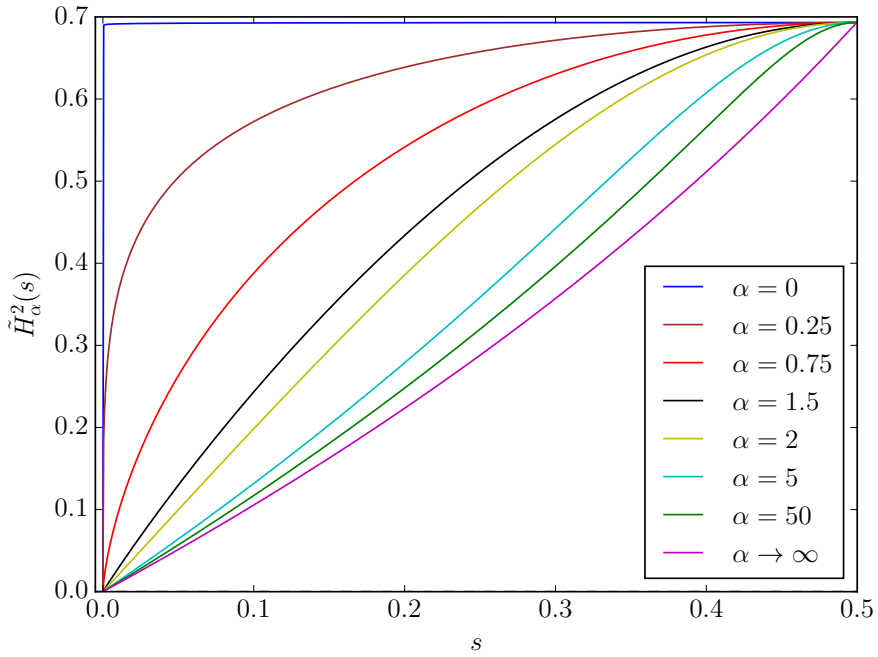


Figure 2: Graph of uncertainty averse preferences represented by Rényi's generalized entropy (Equation 5) for various parameter values $\alpha > 0$ and only two possible outcomes, $n = 2$, such that $s_1 = s$ and $s_2 = 1 - s$. Note that, by construction, the plot is symmetric to the vertical axis at $s = 0.5$.

Lemma 5 is thus a confirmation of the general Proposition 2 for the special case $n = 2$. As before, the sign of $\Lambda(s)$ may change without implying a switch from uncertainty averse to uncertainty-loving preferences (cf. Section 3.4).

4.3 Comparison of decision rules

In this section, we review well-established decision criteria under uncertainty from the literature and compare these to our framework using a stylized and static sample decision problem. The criteria from the literature include the maximin rule (Wald 1949) and its optimistic counterpart, the maximax rule, Laplace's principle of insufficient reason (Laplace 1820), the rule of minimum regret (Niehans 1948, Savage 1951) and the Hurwicz criterion (Arrow and Hurwicz 1977) which is a linear combination of maximin and maximax which weighs possible maximum and minimum payoffs in each state according to the decision maker's optimism. The latter rule is sometimes also called α-maximin. Concise overviews can be found in Luce and Raiffa (1989) and Heal and Millner (2013).

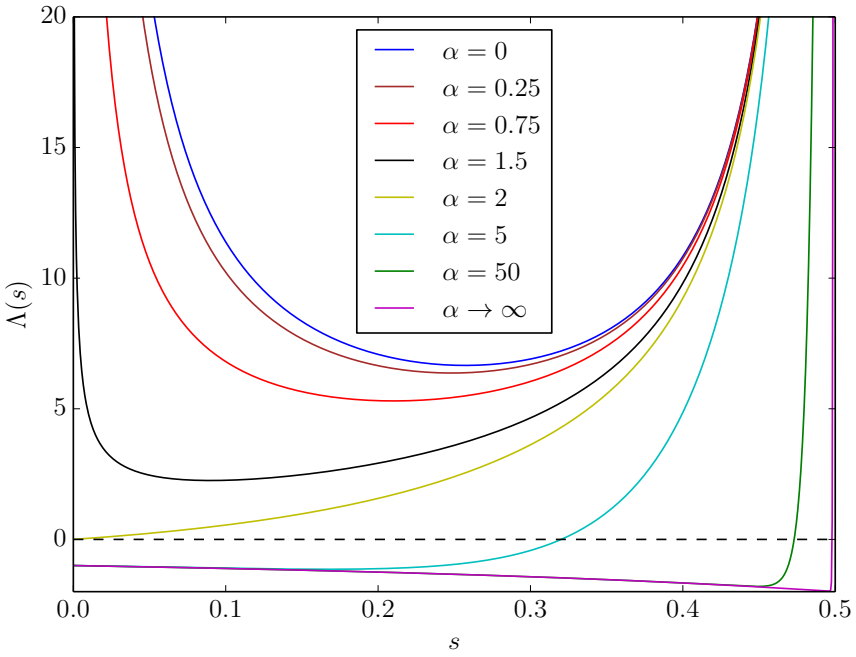


Figure 3: Graph of the coefficient of absolute uncertainty aversion $\Lambda(s)$ (cf. Equation 7) for various values of $\alpha > 0$.

In the following, we take a closer look at these rules. While terms such as maximin ('maximize the minimum over all possible acts') and maximax ('maximize the maximum over all possible acts') are self-explanatory, this is less true for the other three decision rules mentioned. Pierre-Simon Laplace's 1820 principle of insufficient reason[14] states that there is no reason to assume that one specific state of the world is more probable than another one when probabilities are unknown. Hence, they should all be given equal probability weight. Strictly speaking, Laplace's principle is thus a rule for assigning probabilities to outcomes and not a decision rule in itself. However, the wording 'Laplace principle' is often used synonymously with 'Laplacian expected utility', which refers to an expected utility maximizer applying the Laplace principle to calculate expected utility. The rule of minimum regret is based on the idea to minimize the maximum possible 'regret': for each possible state of the world, the act that leads to the highest payoff is set as reference point relative to which the 'regret' is calculated as possible payoff that would potentially be foregone if the respective state of the world materialized. So, in the optimal case regret is zero. The alternative that minimizes possible regret over all states of the world is considered the best choice in this decision framework. Thus, 'regret' is quite similar to

[14]The principle was renamed 'the principle of indifference' by Keynes (1921).

the concept of opportunity cost, but unlike its well-known sibling, it can attain a value of zero. Moreover, focusing on minimizing a negative quantity rather than maximizing a positive one, the rule expresses a very cautious, if not pessimistic attitude of the decision maker towards uncertainty. Mathematically, the rule of minimum regret – sometimes also referred to as 'Savage-Niehans rule' – is to choose the act k from $\mathcal{G}(\overline{y})$ which minimizes possible 'regret', i.e. which minimizes the expression

$$\sum_{i=1}^{n} \left[(\max_{k} u(y_i^k)) - u(y_i^k) \right] . \tag{8}$$

Eventually, the Hurwicz rule generalizes the maximin and maximax criteria: for each alternative k, the function

$$\Phi(y^k) = \lambda \max_i \{u(y_i^k)\} + (1-\lambda) \min_i \{u(y_i^k)\}, \quad 0 \leq \lambda \leq 1 \tag{9}$$

is evaluated and compared to the function values of the alternatives. The associated decision rule is $\max_k \Phi(y^k)$. λ thus reflects the individual's optimism as a greater λ gives more weight to the maximum payoff of one particular act k and hence less weight to the minimum. Similarly, Heal and Millner (2013) give an interpretation of $1-\lambda$ as representing 'aversion to a lack of knowledge'. Hence, with $\lambda = 1$, we recover the maximax rule while $\lambda = 0$ leads again to the maximin criterion.

As exemplary decision problem, we look at the following textbook example: an individual has to take a decision between three acts, f, g and h. The acts are known to generate the following payoffs (in monetary units):

$$\begin{aligned} y^f &= (300, 150, 250, 300)\,, \\ y^g &= (60, 60, 60, 820)\,, \\ y^h &= (15, 280, 340, 365)\,. \end{aligned}$$

Act f is very even but does neither have an especially large maximum payoff nor a particularly low minimum possible payoff. In fact, it guarantees the maximal minimum payoff out of the three alternatives. Hence, the maximin criterion would select act f. Act g offers the potentially highest possible win out of all three uncertain prospects but it only does so in one out of four possible states of the world whereas in the other three

states, we end up having only 60 monetary units. Obviously, the maximax criterion would rate g highest and f lowest. Act h features the smallest minimum but otherwise it offers three potentially large payoffs as compared to f and g. A risk-neutral Laplace individual would be indifferent between the three acts, while a risk-averse one would prefer f. The rule of minimum regret would lead to the choice of f while h and g would be tied. The advice that the Hurwicz criterion gives us critically depends on the choice of λ. A rather pessimistic individual ($\lambda = 0.1$) would choose act f while for any $\lambda \geq \frac{9}{61}$, act g would be preferred. This choice is made by any sufficiently optimistic individual – i.e. $\lambda \geq \frac{9}{61}$ – and as λ is further increased, we can observe a change of the second most preferred act from g to h. The complete rankings of acts are given in Table 1.

Table 1: Orderings over the Knightian acts f, g and h that result from different decision rules from the literature.

decision criterion	choice ordering
maximin	$f \succ g \succ h$
maximax	$g \succ h \succ f$
risk-neutral Laplace EU	$f \sim g \sim h$
risk-averse Laplace EU	$f \succ h \succ g$
minimum regret	$f \succ g \sim h$
Hurwicz, $\lambda = 0.1$	$f \succ g \succ h$
$\lambda = 0.2$	$g \succ f \succ h$
$\lambda = 0.8$	$g \succ h \succ f$

In Table 2, we illustrate our proposed index $H^4_\alpha(s^k)$ for the three Knightian acts f, g and h for different parameter values of uncertainty aversion α. In our framework, a comparison of differences in the uncertainty utility index is meaningful for the same individual due to Proposition 1 (uniqueness up to linear-affine transformations). We see that, although the preference over the acts always remains $f \succ h \succ g$, the overall level of well-being attached to a single act drastically depends on the degree of uncertainty aversion α. For example, an individual with a very low level of uncertainty aversion ($\alpha = 0.1$), the respective index values of H provided by the three uncertain prospects are within a range of 0.077 from act f (best) to act g (worst), whereas at high levels of uncertainty aversion (e.g. $\alpha = 50$), the difference is 1.016. Thus, an individual relatively uncaring towards Knightian uncertainty would gain relatively little in terms of preference

satisfaction when swapping from prospect g to prospect f. On the other hand, the very same swap would mean an over six times higher level of preference satisfaction to a highly uncertainty-averse decision maker.

Table 2: H^4_α scores of the three Knightian acts f, g and h for different degrees of uncertainty aversion α. The resulting preference ordering is $f \succ h \succ g$.

uncertainty aversion	uncertainty utility		
α	$H^4_\alpha(s^f)$	$H^4_\alpha(s^g)$	$H^4_\alpha(s^h)$
0.1	1.383	1.306	1.338
0.5	1.369	0.983	1.215
1	1.354	0.660	1.151
3	1.309	0.291	1.103
5	1.283	0.243	1.090
10	1.252	0.216	1.070
20	1.230	0.204	1.048
50	1.214	0.198	1.025

In summary, combining the results from Tables 1 and 2, we find that the overall ranking of acts from our method is different from the other criteria. However, the most preferred option is the same as with the maximin rule and a pessimistic Hurwicz individual. That being said, a risk-averse Laplacian individual would arrive at the very same ranking as our Rényi individual, but this comes with the above disclaimer.

5 Discussion and conclusion

We discuss some key points of the paper and provide a conclusion and some outlook in the following.

Summary. Based on a set of seven axioms on a preference relation over Knightian acts, we have provided a proof that there exists a real-valued, additive and extensive representation of these preferences on any set $\mathcal{G}(\overline{y})$, which contains all Knightian acts with positive total payoff volume $\overline{y}$. Moreover, we have shown that this representation is monotonous on $\mathcal{G}(\overline{y})$ and unique up to linear-affine transformations. Within our framework, we have furthermore provided a new and very parsimonious definition of the notion

of uncertainty aversion, based on which we also formalize the concept of an uncertainty premium, which makes interpersonal comparison of uncertainty attitudes possible.

We have illustrated our approach with a suitable function from information theory fulfilling our axioms – Rényi's generalized entropy (Rényi 1961) – which we have demonstrated to contain a parameter that captures the degree of an individual's aversion to Knightian uncertainty. Finally, we have compared our approach in a static sample decision problem to other methods from the literature. We have found that our preference index produces a ranking different from the other decision rules considered. However, the most preferred act coincides with the one preferred by an individual with maximin preferences and with the choice a very pessimistic Hurwicz individual would make.

(In-)Completeness of preferences. The arguably strongest assumption required to bring our framework to life is Axiom 3. It assumes completeness only on the subset $\mathcal{G}(\overline{y})$ of all simple acts $\mathcal{F}$. Thus, for two arbitrary Knightian acts from $\mathcal{F}$, we do not globally assume that the individual will be able to state her preferences, i.e. they may be incomparable. Much rather, we require her to have complete preferences only locally, that is, on the subset of Knightian acts with payoff volume $\overline{y}$, $\mathcal{G}(\overline{y})$. Indeed, we think that this kind of locally complete preferences is in fact normatively much more compelling than assuming completeness on $\mathcal{F}$, even more so in the case of Knightian uncertainty. In fact, the implications of the completeness assumption for economic theory have been vividly discussed from the outset. von Neumann and Morgenstern (1944) themselves considered it 'very dubious whether the idealization which treats this postulate as a valid one, is appropriate or even convenient' (ibid.: p. 630). Others like Luce and Raiffa (1957) criticized the possibility of intransitivities if individuals exacted decisions between alternatives that might be 'inherently incomparable'. In the same vein, R.J. Aumann (1962, 1964) doubted the normative appeal of an an a-priori exclusion of the possibility of an individual to be unwilling or unable to arrive at preference statements for at least some acts. In our view, Aumann's point was that the inability to state one's preferences regarding a decision might be the result of rational thinking and judgment, so there is no reason to make completeness of preferences a standard of any rational choice theory, an argument ultimately very similar to Putnam (1986). Despite all this, the only other contribution we know of that discusses and uses incompleteness of preferences in the context of Knightian uncertainty is Bewley (2002). He replaces completeness with an

inertia assumption, which states that an alternative is only accepted if preferred to the status quo. Moreover, an individual may assert that two alternatives are incomparable. The main difference to our approach however is that Bewley works within the Anscombe-Aumann framework, which relies on subjective and objective probabilities, a concept we have deliberately avoided here for reasons laid out earlier in this paper.

The role of $\overline{y}$ and its intuitive appeal. The sum of payoffs $\overline{y}$ over all possible states of the world plays a central role in our theory. One might wonder about desirability and intuitive appeal of this feature. First and foremost, the issue is closely related to the above question of whether or not to assume incomplete preferences. We argue in favor of incompleteness here, but it is clear that it naturally comes at a price. From a technical point of view, the problem with incomplete preferences is that one cannot have a complete representation either, so one might run into issues with dominance. On the other hand, for each $\overline{y} > 0$, $\mathcal{G}(\overline{y})$ is the largest possible set of acts which cannot possibly dominate each other, so it seems natural to build a theory around this. As to intuition, consider the possibility that you are promised a slice of your favorite pie tomorrow, but the size of the slice will depend, for some reason, on (1) what you do today, and (2) what state of the world materializes tomorrow. Even before you worry about your options to act and single outcomes, the most natural question seems 'How large is that pie anyway?'. Only as a second step will you probably think about what is best to do given the possible – fundamentally uncertain – states of the world.

Probabilistic sophistication. A central feature of the Rényi preference representation from Section 4 is the symmetry of $H(s)$ (cf. Lemma 4), which implies that the sequence of payoffs in any given payoff distribution does not matter to the individual. She only cares about payoffs, not about the state of the world which causes the payoff. This property has been termed 'event exchangeability' (Chew and Sagi 2006: 771), and it interestingly implies the individual to be *probabilistically sophisticated* (Machina and Schmeidler 1992) with uniform subjective beliefs. That is, because any two outcomes of any act are exchangeable for a Rényi individual, the concrete choice of the Rényi function implies that the decision maker follows Laplace's principle of insufficient reason in a non-expected utility framework. In that respect, our result might be one step towards a result that parallels the contribution of Gravel, Marchant and Sen (2012), who provided a complete axiomatic foundation of Laplace's principle in the expected utility setting.

Connection to Laplacian expected utility. In the following, we compare the standard expected utility (EU) framework to our approach. Starting from expected utility with concrete utility function $u(y) = \ln y$, and assuming that we have an individual that applies Laplace's principle, we can generally say that the ranking of acts will coincide with a ranking done by our Rényi individual. We make this explicit in Appendix D. In that particular case, the connection between Laplacian EU and our framework can be established, because a strictly increasing monotonous transformation from the EU functional to the Rényi functional can always be constructed. The situation is however not clear for general utility functions $u(y)$ or representations of H other than the one presented here, so a general correspondence between these two frameworks cannot be established here, and its existence seems questionable to us. However, the positive message to be learned from this observation is that the theory developed here is not completely detached from other theories, but much rather shares a boundary point with them.

(Non-) Additive entropies in economics. We have seen that our axiomatic foundation of preferences allows for the existence of an additive preference representation, and that such functions are called additive entropies in physics and information theory. For completeness, we should say that entropies have been used before in economic theory. Luce, Ng and Marley (2008) have proposed to use entropies to model the utility which individuals derive from the process of gambling itself. This approach seems interesting, because it departs from the typical consequentialist setting in economics, in that it matters to the decision maker how an outcome is obtained rather than just caring about what outcome is obtained. However, this is quite a different use of the concept of entropy, and unrelated to our concept presented here. Second, there are also non-additive entropies, such as the Tsallis entropy (Tsallis 1988). Non-additivity is interesting from the point of view of ranking and valuing compound acts. With additivity, any compound act is just as good as the sum of preference satisfaction obtained from its constituents (cf. Corollary 1). Non-additivity would enable us to model situations where compound acts are better (worse) than just the sum of its constituents. While such preferences would be an interesting object of study, we do not know of a formal axiomatization that could be transferred in a way similar to the one presented here.

Practicability. One might think that it is highly unlikely that in any real-world decision problem, every option to act will live in the same subspace $\mathcal{G}$ of $\mathcal{F}$. But then, one

might as well wonder whether the theory presented here was practically of any relevance. After all, we cannot hope that nature will give us problems taylor-made to the tight frame of our theories. But we can assure the critical reader that there is some hope. since it is almost always possible to approximate elements of $\mathcal{F}$ through elements of some subspace $\mathcal{G}$. An example should clarify the matter: Suppose we want to compare the acts f_1, f_2, g_3, g_4 where, unfortunately, f_1 and f_2 live in $\mathcal{F}$ whereas $g_3, g_4 \in \mathcal{G}$. We cannot directly compare these options, but we can find the elements closest to f_1 and f_2 in the subspace $\mathcal{G}$. Linear algebra teaches us that this is possible for any act space $\mathcal{F}$ with *finite* dimension (cf. Schönhage 1971), which should easily be the case for any real-world problem. The question then is if the price that we pay in terms of approximation errors is reasonably low or too high. This necessarily depends on the problem at hand and should be judged on a case-by-case basis.

Conclusion and outlook. In a nutshell, we have shown how a parsimonious set of axioms leads to a special preference representation under Knightian uncertainty, which is additive and extensive. We have illustrated how our setup could be used to model and inter-personally compare different attitudes towards such uncertainty. From here, several research fields seem open: First, there is the problem field of measuring uncertainty aversion, theoretically as well as experimentally. Theoretically, one could think of a transfer of concepts from general relativity, where measuring curvature that is invariant under certain coordinate transformations in high dimensional spaces is formalized. Experimentally, it would be interesting to assess uncertainty attitudes in different settings and contexts. Moreover, our ideas might find applications in measuring product and technological diversity, in institutional design, or in generalizing the concepts of (self-)insurance and protection (Ehrlich and Becker 1972) to situations of Knightian uncertainty.

Acknowledgments

The authors would like to thank the following persons: Dave Abson, Simon Grant, Nikolai Hoberg, Mark Machina, Jack Stecher and Martin F. Quaas for critical discussion of ideas and many helpful comments on earlier versions of this manuscript. Financial support

under grant number 01LL0911E from the German Federal Ministry of Education and Research is gratefully acknowledged.

Appendix

The calculation rules for the compounding of acts are basically given by the rules that underlie the compounding operation, i.e. standard addition and multiplication (cf. Definition 2).

A Proof of Lemma 1

Assume $f \oplus h \succeq g \oplus h$ (see premise of Lemma 1). Let furthermore $\epsilon = \frac{1}{n}$ where $n \in \mathbb{N}$. Then, we have by Axioms 1 and 6

$$f \oplus \epsilon h \sim (1-\epsilon)f \oplus \epsilon f \oplus \epsilon h \tag{10}$$

From Axiom 1, we know that

$$(1-\epsilon)f \sim (1-\epsilon)f \tag{11}$$

From our premise and Axiom 5, we moreover know

$$\epsilon f \oplus \epsilon h \succeq \epsilon g \oplus \epsilon h \tag{12}$$

Equations 11 and 12 can be combined via Axiom 4 to yield

$$(1-\epsilon)f \oplus \epsilon f \oplus \epsilon h \succeq (1-\epsilon)f \oplus \epsilon g \oplus \epsilon h \tag{13}$$

which we may combine with Equation 10 so as to read

$$f \oplus \epsilon h \succeq (1-\epsilon)f \oplus \epsilon g \oplus \epsilon h \sim (1-2\epsilon)f \oplus \epsilon f \oplus \epsilon g \oplus \epsilon h \tag{14}$$

Using again Equation 12 together with Axioms 1 and 4, we may write

$$f \oplus \epsilon h \succeq (1-2\epsilon)f \oplus 2\epsilon g \oplus \epsilon h \tag{15}$$

Repeating this succession of steps n times yields

$$f \oplus \epsilon h \succeq (1 - n\epsilon) f \oplus n\epsilon g \oplus \epsilon h \tag{16}$$

which can be rewritten as

$$f \oplus \epsilon h \succeq g \oplus \epsilon h \tag{17}$$

And from here, we may conclude by Axiom 7 that $f \succeq g$ indeed.

□

B Proof of Proposition 1

While this proof follows Lieb and Yngvason (1999) in terms of argumentation and layout, we transfer it from thermodynamics to our setting of preferences over Knightian acts. It consists of three main steps, which we sketch briefly in the following: First, we show in Lemma 6 that for any act $f \in \mathcal{G}(\overline{y})$ there exists at least one real number, which serves as 'utility scale', and which can be defined as mixture parameter of two other arbitrary acts from $\mathcal{G}(\overline{y})$. We define the supremum of all possible such mixture parameters as the 'canonical utility' on $\mathcal{G}(\overline{y})$. Here, 'canonical' is to be understood in the sense of a standardization, which is achieved by the supremum operation. Economically, this can be seen as calibration of the utility scale. In Lemma 7, we demonstrate the equivalence of $\preceq$ on $\mathcal{G}(\overline{y})$ and $\leq$ on $\mathbb{R}$. Third, we show in Lemma 8 that $H_{\overline{y}}(s^f)$ is unique on $\mathcal{G}(\overline{y})$. These ingredients allow for a brief formulation of the proof, which we subsequently provide. Finally, Lemma 9 establishes uniqueness up to linear-affine (cardinal) transformations.

Lemma 6. *Suppose that f_0 and f_1 are two acts in $\mathcal{G}(\overline{y})$ with $f_0 \prec f_1$. Define for $\lambda \in \mathbb{R}$*

$$H_\lambda = \{f \in \mathcal{G}(\overline{y}) : (1 - \lambda) f_0 \oplus \lambda f_1 \preceq f\}$$

Then

1. *$\forall f \in \mathcal{G}(\overline{y})$, there is a $\lambda \in \mathbb{R}$ such that $f \in H_\lambda$*
2. *$\forall f \in \mathcal{G}(\overline{y})$, $\sup\{\lambda : f \in H_\lambda\} < \infty$*

In words, (1) for every Knightian act, or equivalently act f, there exists a real number such that $f \in H_\lambda$ and (2) this real number is bounded above.

Proof. 1. If $f_0 \preceq f \Rightarrow f \in H_0$ by Axiom 2. For general f, we claim for some $\alpha \geq 0$

$$(1+\alpha)f_0 \preceq \alpha f_1 \oplus f \tag{18}$$

and hence

$$(1-\lambda)f_0 \oplus \lambda f_1 \preceq f \quad \text{with} \quad \alpha = -\lambda$$

If Equation 18 were not true, then $\alpha f_1 \oplus f \preceq (1+\alpha)f_0 \quad \forall \alpha > 0$ and so, by Axioms 5 and 6

$$f_1 \oplus \frac{1}{\alpha} f \preceq f_0 \oplus \frac{1}{\alpha} f_0$$

By Axiom 7, this would imply $f_0 \preceq f_1$, in contradiction to the assumption

2. This is essentially the same argument, i.e. proof by contradiction: If $\sup\{\lambda : f \in S_\lambda\} = \infty$, then we would have for some sequence of λ's tending to ∞

$$(1-\lambda)f_0 \oplus \lambda f_1 \preceq f$$

which would imply by Axioms 4 and 6 that

$$f_0 \oplus \lambda f_1 \preceq f \oplus \lambda f_0$$

and by Axiom 5

$$\frac{1}{\lambda} f \oplus f_1 \preceq \frac{1}{\lambda} f \oplus f_0$$

which would imply by the continuity axiom (Axiom 7) that $f_1 \preceq f_0$.

□

In the next step, we assume Knightian acts $f_0, f_1 \in \mathcal{G}(\overline{y})$ with $f_0 \preceq f_1$ and define for arbitrary $f \in \mathcal{G}(\overline{y})$

$$H_{\overline{y}}(s^f) := \sup\{\lambda : (1-\lambda)f_0 \oplus \lambda f_1 \preceq f\} \tag{19}$$

the canonical utility on $\mathcal{G}(\overline{y})$ with reference points f_0 and f_1 in the space $\mathcal{G}(\overline{y})$. Then Lemma 6 guarantees that $H_{(\overline{y})}(s^f)$ is well-defined and bounded above.

Lemma 7 (Equivalence of $\preceq$ and $\leq$)**.** *Assume $f_0 \preceq f_1$ as before and $a_0, a_1, a_0', a_1' \in \mathbb{R}$ with $a_0 + a_1 = a_0' + a_1'$. Then the following are equivalent*

1. $a_0 f_0 \oplus a_1 f_1 \preceq a_0' f_0 \oplus a_1' f_1$
2. $a_1 \leq a_1'$ *(and hence $a_0 \geq a_0'$)*

Furthermore, $\sim$ holds in 1. if and only if $a_1 = a_1'$ and $a_0 = a_0'$.

Proof. Assume that $a_0 + a_1 = a_0' + a_1' = 1$ and that all a's are strictly positive.
1. $\Rightarrow$ 2.: We write $\lambda = a_1$ and $\lambda' = a_1'$. We deliberately assume that $\lambda > \lambda'$ which violates 2. above to show that this assumption leads to a contradiction. If indeed $\lambda > \lambda'$, then we have

$$(1-\lambda) f_0 \oplus \lambda f_1 \preceq (1-\lambda') f_0 \oplus \lambda' f_1$$

and by Axioms 4 and 6, we get

$$(1-\lambda) f_0 \oplus \lambda' f_1 \oplus (\lambda - \lambda') f_1 \preceq (1-\lambda) f_0 \oplus (\lambda - \lambda') f_0 \oplus \lambda' f_1$$

From this, applying Axioms 4 and 6 again, we arrive at $(\lambda - \lambda') f_1 \preceq (\lambda - \lambda') f_0$ which yields by Axiom 5 that $f_1 \preceq f_0$ which is the contradiction we were looking for.
2. $\Rightarrow$ 1.:

$$\begin{aligned} (1-\lambda) f_0 \oplus \lambda f_1 &\overset{\text{A4/A6}}{\sim} (1-\lambda') f_0 \oplus (\lambda' - \lambda) f_0 \oplus \lambda f_1 \\ &\overset{\text{A4/A5}}{\preceq} (1-\lambda') f_0 \oplus (\lambda' - \lambda) f_1 \oplus \lambda f_1 \\ &\overset{\text{A4/A6}}{\sim} (1-\lambda') f_0 \oplus \lambda' f_1 \end{aligned} \tag{20}$$

which only holds for $\lambda' > \lambda$. □

Lemma 8. *(Uniqueness of canonical utility $H_{\overline{y}}$) Let $H_{\overline{y}}$ denote the canonical utility from Equation 19 on $\mathcal{G}(\overline{y})$ with respect to the reference acts $f_0 \preceq f_1$. If $f \in \mathcal{G}(\overline{y})$, then*

$$\lambda = H_{\overline{y}}(s^f)$$

is equivalent to

$$f \sim (1-\lambda)f_0 \oplus \lambda f_1$$

Proof. First, if $\lambda = H_{\overline{y}}(s^f)$, then by definition of the supremum, there is a sequence $\varepsilon_1 \geq \varepsilon_2 \geq \ldots \geq 0$ converging to zero such that

$$(1-(\lambda-\varepsilon_n))f_0 \oplus (\lambda-\varepsilon_n)f_1 \preceq f \quad \forall n$$

By Axiom 6

$$\begin{aligned}(1-\lambda)f_0 \oplus \lambda f_1 \oplus \varepsilon_n f_0 &\sim (1-\lambda+\varepsilon_n)f_0 \oplus (\lambda-\varepsilon_n)f_1 \oplus \varepsilon_n f_1 \\ &\preceq f \oplus \varepsilon_n f_1 \end{aligned} \tag{21}$$

By Axiom 7, we get

$$(1-\lambda)f_0 \oplus \lambda f_1 \preceq f$$

On the other hand, since λ is the supremum we have by Axiom 3

$$f \preceq (1-(\lambda+\varepsilon))f_0 \oplus (\lambda+\varepsilon)f_1 \quad \varepsilon > 0$$

which means

$$f \oplus \varepsilon f_0 \preceq (1-\lambda)f_0 \oplus \lambda f_1 \oplus \varepsilon f_1$$

and so, by Axiom 7 again

$$f \preceq (1-\lambda)f_0 \oplus \lambda f_1 \quad \Longrightarrow f \sim (1-\lambda)f_0 \oplus \lambda f_1 \quad \text{when} \quad \lambda = H_{\mathcal{G}(\overline{y})}(f)$$

If, conversely, $\lambda' \in [0,1]$ is such that

$$f \sim (1-\lambda')f_0 \oplus \lambda' f_1$$

then by Axiom 2

$$(1-\lambda)f_0 \oplus \lambda f_1 \sim (1-\lambda')f_0 \oplus \lambda' f_1$$

and thus $\lambda = \lambda'$ by Lemma 7. □

Hence, for every $f \in \mathcal{G}(\overline{y})$, there is a unique $\lambda \in \mathbb{R}$, namely $\lambda = H_{\overline{y}}(f)$, such that

$$f \sim (1-\lambda)f \oplus \lambda f \sim (1-\lambda)f_0 \oplus \lambda f_1 \tag{22}$$

Put differently, any Knightian act $f \in \mathcal{G}(\overline{y})$ is always representable by a linear mixture of two arbitrary, non-identical acts from $\mathcal{G}(\overline{y})$ with mixture parameter λ.

The proof of Proposition 1 can now be provided.

Proof of Proposition 1:

Proof. 1. $\Rightarrow$ 2.: Let $\lambda_i = H_{\overline{y}}(s^{g_i})$ and $\lambda_i' = H_{\overline{y})}(s^{g_i'})$. By Lemma 8, we know that $g_i \sim (1-\lambda_i)f_0 \oplus \lambda_i f_1$ and $g_i' \sim (1-\lambda_i')f_0 \oplus \lambda_i' f_1$. By the consistency axiom (Axiom 4) and Axiom 6, we know – if $t_1 + \ldots + t_N = t_1' + \ldots + t_M'$ (for all $N \geq 1$ and $M \geq 1$) – that

$$t_1 g_1 \oplus \ldots \oplus t_N g_N \sim \sum_{i=1} t_i(1-\lambda_i)f_0 \oplus \sum_i t_i \lambda_i f_1$$

and

$$t_1' g_1' \oplus \ldots \oplus t_N' g_N' \sim \sum_{i=1} t_i'(1-\lambda_i')f_0 \oplus \sum_i t_i' \lambda_i' f_1$$

Statement 2 now follows from Lemma 7. The implication 2. $\Rightarrow$ 1. is obvious. □

In the following last Lemma, we demonstrate that $H_{\mathcal{G}(\overline{y})}$ is unique up to cardinal transformations.

Lemma 9 (Cardinality of $H_{\overline{y}}$)**.** *If $H_{\overline{y}}^*$ is a function on $\mathcal{G}(\overline{y})$ satisfying*

$$(1-\lambda)f \oplus \lambda g \preceq (1-\lambda)f' \oplus \lambda g'$$

if and only if

$$(1-\lambda)H_{\overline{y}}^*(s^f) + \lambda H_{\overline{y}}^*(s^g) \leq (1-\lambda)H_{\overline{y}}^*(s^{f'}) + \lambda H_{\overline{y}}^*(s^{g'}) \quad \forall f,g,f',g' \in \mathcal{G}(\overline{y})$$

then $H_{\overline{y}}^*(s^f) = aH_{\overline{y}}(s^f) + b$ with $a = H_{\overline{y}}^*(s^{f_1}) - H_{\overline{y}}^*(s^{f_0}) > 0$ and $b = H_{\overline{y}}^*(s^{f_0})$. $H_{\overline{y}}$ is the canonical utility on $\mathcal{G}(\overline{y})$ with reference acts f_0 and f_1.

Proof. From Equation 22, we have by hypothesis on $H^*_{\overline{y}}$ and $\lambda = H_{\overline{y}}$

$$\begin{aligned} H^*_{\overline{y}}(s^f) &= (1-\lambda)H^*_{\overline{y}}(s^{f_0}) + \lambda H^*_{\overline{y}}(s^{f_1}) \\ &= \left[1 - H_{\overline{y}}(s^f)\right] H^*_{\overline{y}}(s^{f_0}) + H_{\overline{y}}(s^f) H^*_{\overline{y}}(s^{f_1}) \\ &= \left[H^*_{\overline{y}}(s^{f_1}) - H^*_{\overline{y}}(s^{f_0})\right] H_{\overline{y}}(s^f) + H^*_{\overline{y}}(s^{f_0}) \end{aligned} \tag{23}$$

The last line implies that $a = H^*_{\overline{y}}(s^{f_1}) - H^*_{\overline{y}}(s^{f_0}) > 0$ since $f_0 \preceq f_1$ by assumption. This establishes cardinality of the Knightian utility index $H_{\overline{y}}$ on $\mathcal{G}(\overline{y})$ and thus completes the proof. □

C Proof of Lemma 2

Let $\overline{y} > 0$ be given. Consider acts from the set $\mathcal{G}(\overline{y}) \subset \mathcal{F}$ and two individuals, A and B, accepting Axioms 1–7. By Proposition 1, we can attach $\tilde{H}$-values properly to all acts from $\mathcal{G}(\overline{y})$ for both individuals. Without loss of generality, we may use uniqueness of $\tilde{H}$ up to linear-affine transformations and normalize $\tilde{H}_A(0) = \tilde{H}_B(0) = 0$ and $\tilde{H}_A(s^c) = \tilde{H}_B(s^c) = 1$. Recall from Definition 5 that if A is more uncertainty averse than B for some act from $\mathcal{G}(\overline{y})$, then this will also be the case for any other act from $\mathcal{G}(\overline{y})$. We start by proving '⇒' before we prove '⇐'.

'⇒': Assume $\Lambda_B(s) \geq \Lambda_A(s)$ for all $s \in I = [0, 0.5]$, i.e. $\tilde{H}_B(s)$ is curved more steeply towards the s axis than $\tilde{H}_A(s)$ on the entire interval I. $\tilde{H}_i(s)$ are continuous (Proposition 1) and strictly monotonically increasing on I (Definition 6), so that the only possibility for the constant act $s^c = 0.5$ from $\mathcal{G}(\overline{y})$ to be the utility maximum for both individuals is that $\tilde{H}_B(s)$ lies above $\tilde{H}_A(s)$ for all $s \in I$. By assumption, it holds for both individuals that $\Pi \propto y_H - y_L$, which directly translates to $\Pi \propto s^c - s$. We may then conclude that the uncertainty premium of person i depends on the corresponding difference in utilities, so that $\Pi_i(s) \propto \tilde{H}_i(s^c) - \tilde{H}_i(s)$, where of course $\tilde{H}_A(s^c) = \tilde{H}_B(s^c) = 1$. From there, utilizing $\tilde{H}_B(s) > \tilde{H}_A(s)$ for all $s \in I$, it follows that $\Pi_B(s) < \Pi_A(s)$ for all $s \in I$.

'⇐': Assume $\Pi_B(s) < \Pi_A(s)$ for all $s \in I = [0, 0.5]$, i.e. the uncertainty premium of B is smaller than that of A for every act $s \in I$. Because of the normalization of H_i and because we know – by the same argument as above – that $\Pi_i(s) \propto 1 - \tilde{H}_i(s)$,

it follows that $H_B(s) > H_A(s)$ for all $s \in I$. But then, because of $\tilde{H}_i(0) = 0$ and $\tilde{H}_i(0.5) = 1$, it is then necessarily true for every $s \in I$ that $\tilde{H}_B(s)$ is curved more steeply towards the s axis than $\tilde{H}_A(s)$, and therefore $\Lambda_B(s) \geq \Lambda_A(s)$ for every $s \in I$.

□

D Connection to expected utility

If $u(y) = \ln y$, then there is a direct correspondence between a Laplace-EU individual and a Rényi individual. To see this, consider a two-state act with one good outcome, y_H and one bad outcome y_L. The Rényi functional then reads

$$H = \frac{1}{1-\alpha} \ln \left[\left(\frac{y_L}{y_L + y_H} \right)^{\alpha} + \left(\frac{y_H}{y_L + y_H} \right)^{\alpha} \right] \tag{24}$$

while the Laplacian expected utility reads

$$\mathbb{E}U = \frac{1}{n}(\ln y_L + \ln y_H) = \frac{1}{n} \ln(y_L \cdot y_H) \tag{25}$$

For the relative ranking of any two acts, the ln functions matter. The question is thus, whether there exists a strictly monotonous, i.e. order preserving, transformation

$$T : y_L \cdot y_H \mapsto \left(\frac{y_L}{y_L + y_H} \right)^{\alpha} + \left(\frac{y_H}{y_L + y_H} \right)^{\alpha} \tag{26}$$

At least one such transformation exists:

$$T : x \mapsto \frac{(\frac{x}{y_H})^{\alpha} + (\frac{x}{y_L})^{\alpha}}{(\frac{x}{y_L} + \frac{x}{y_H})^{\alpha}} \tag{27}$$

It holds that $T'(x) > 0$ for all $\alpha > 0$ and such a transformation can be constructed for all $n > 2$ as can be seen by complete induction. It is however unclear whether such a transformation can be found for any $u(y)$ and any H.

References

Andersen, S., G.W. Harrison M.I. Lau and E.E. Rutström (2006), Elicitation using multipleprice list formats, *Experimental Economics* 9, 383–405

Al-Najjar, N.I. and J. Weinstein (2009), The ambiguity aversion literature: a critical assessment, *Economics and Philosophy* 25, 249–284

Allais, M. (1953), Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine, *Econometrica* 21(4), 503–546

Alpert, M. and H. Raiffa (1982), A progress report on the training of probability assessors, In: Kahnemann, D., P. Slovic and A. Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, UK, pp. 294–305

Anscombe, F.J. and R.J. Aumann (1963), A definition of subjective probability, *The Annals of Mathematical Statistics* 34, 199–205

Arrow, K.J. (1964), The role of securities in optimal allocation of risk-bearing, *Review of Economic Studies* 31, 91–96

Arrow, K.J. (1965), *Aspects of the Theory of Risk-Bearing*, Yrjö Jahnssonin Säätio, Helsinki

Arrow, K.J. and L. Hurwicz (1977), An optimality criterion for decision-making under ignorance, In: Arrow, K.J. and L. Hurwicz (ed.), *Studies in Resource Allocation Processes*, Cambridge University Press, Cambridge, pp. 463–75

Aumann, R.J. (1962), Utility theory without the completeness axiom, *Econometrica* 30(3), 445–462

Aumann, R.J. (1964), Utility theory without the completeness axiom: a correction, *Econometrica* 32(1), 210–212

Beck, C. and T. Schlögl (1993), *Thermodynamics of Chaotic Systems: An Introduction*, Cambridge University Press, Cambridge, UK.

Bewley, T.F. (2002), Knightian decision theory. Part I, *Decisions in Economics and Finance* 25, 79–110

Chew, S.H. and J.S. Sagi (2006), Event exchangeability: probabilistic sophistication without continuity or monotonicity, *Econometrica* 74(3), 771–786

de Finetti (2008)[1979], ed. by A. Mura, transl. by H. Hosny, with an introduction by M.C. Galavotti, *Philosophical Lectures on Probability*, Springer, Dordrecht

Debreu, G. (1959), *Theory of Value: An Axiomatic Analysis of General Equilibrium*, Yale University Press, New Haven, CT

Duncan, G.T. (1977), A matrix measure of multivariate risk aversion, *Econometrica* 45(4), 895–903

Ehrlich, I. and G.S. Becker (1972), Market insurance, self–insurance and self–protection, *Journal of Political Economy* 80(4): 623–648.

Ellsberg, D. (1961), Risk, ambiguity, and the savage axioms, *Quarterly Journal of Economics* 75, 643–669

Feduzi, A. and J. Runde (2011), The uncertain foundations of the welfare state, *Journal of Economic Behavior & Organization* 80, 613–627

Fyfe, J.C., N.P. Gillett and F.W. Zwiers (2013), Overestimated global warming over the past 20 years, *Nature Climate Change* 3, 767–769

Gilboa, I. (2010), Questions in decision theory, *Annual Review of Economics* 2, 1–19

Gilboa, I. and D. Schmeidler (1989), Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics* 18, 141–153

Gravel, N., T. Marchant and A. Sen (2012), Uniform expected utility criteria for decision making under ignorance or objective ambiguity, *Journal of Mathematical Psychology* 56(5), 297–315

Halevy, Y. (2007), Ellsberg revisited: an experimental study, *Econometrica* 75(2), 503–536

Hartley, R.V.L. (1928), Transmission of information, *Bell System Technical Journal* 7(3), 535–563

Heal, G. and A. Millner (2013), Uncertainty and decision in climate change economics, *Centre for Climate Change Economics and Policy Working Paper* No. 128

Hirshleifer, J. (1965), Investment decision under uncertainty, *Quarterly Journal of Economics* 79, 509–536

Hirshleifer, J. (1966), Investment decision under uncertainty, *Quarterly Journal of Economics* 80, 252–277

Keynes, J.M. (1921), *A Treatise on Probability*, Macmillan, London

Klibanoff, P., M. Marinacci and S. Mukerji (2005), A smooth model of decision making under ambiguity, *Econometrica* 73(6), 1849–1892

Knight, F. (1921), *Risk, Uncertainty, and Profit*, 1st edition. Houghton Mifflin Company, The Riverside Press, Cambridge, MA

Laplace, P.S. (1820), *Théorie Analytique des Probabilités*, Mme Ve Courcier, Paris

Levy, H., and A. Levy (1991), Arrow-Pratt measures of risk aversion: the multivariate case, *International Economic Review* 32(4), 891–898

Lieb, E. and J. Yngvason (1999), The physics and mathematics of the second law of thermodynamics, *Physics Reports* 310, 1–96

Luce, R.D. (1972), Conditional expected, extensive utility, *Theory and Decision* 3, 101–106

Luce, R.D., C.T. Ng and A.A.J. Marley (2008), Utility of gambling I: entropy modified linear weighted utility, *Economic Theory* 36, 1–33

Luce, R.D. and H. Raiffa (1957), *Games and Decisions: Introduction and Critical Survey*, Wiley, New York, NY

Maccheroni, F., M. Marinacci and A. Rusticchini (2006), Ambiguity aversion, robustness, and the variational representation of preferences, *Econometrica* 74(6), 1447–1498

Machina, M. and M. Rothschild (2008), Risk, In: Durlauf, S.N. and L.E. Blume (eds.), *The New Palgrave Dictionary of Economics*, Macmillan, London

Machina, M. and D. Schmeidler (1992), A more robust definition of subjective probability, *Econometrica* 60, 745–780

Mehta, V., G. Meehl L. Goddard, J. Knight, A. Kumar, M. Latif, T. Lee, A. Rosati and D. Stammer (2009), Decadal climate predictability and prediction: where are we? *The 8th Workshop on Decadal Climate Variability: Summary*

Millner, A., R. Calel, D.A. Stainforth and G. MacKerron (2013), Do probabilistic expert elicitations capture scientists' uncertainty about climate change, *Climate Change* 116(2), 427–36

Millner, A., S. Dietz and G. Heal (2010), Ambiguity and climate policy, *NBER Working Paper* No. 16050, June 2010

Nau, R.F. (2006), Uncertainty aversion with second order utilities and probabilities, *Management Science* 52, 136–145

Niehans, J. (1948), Zur Preisbildung bei ungewissen Erwartungen, *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 84, 433–456

Polasky, S., S.R. Carpenter, C. Folke and B. Keeler (2011), Decision-making under great uncertainty: environmental management in an era of global change, *Trends in Ecology and Evolution* 26(8), 398–404

Pratt, J.W. (1964), Risk aversion in the small and in the large, *Econometrica* 32, 122–136

Putnam, H. (1986), Rationality in decision theory and in ethics, *Critica: Revista Hispanoamericana de Filosofia* 18(54), 3–16

Rényi, A. (1961), On measures of entropy and information, in J. Neyman (ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley, CA, pp. 547–561

Rubinstein, A. (2006), *Lecture Notes in Microeconomic Theory: The Economic Agent*, 1st edition. Princeton University Press, Princeton, NJ

Savage, L.J. (1951), The theory of statistical decision, *Journal of the American Statistical Association* 46, 55–67

Savage, L.J. (1954), *The Foundations of Statistics*, Wiley and Sons, New York, NY, 2nd revised and enlarged edition (1972), Dover Publications, New York, NY

Schmeidler, D. (1989), Subjective probability and expected utility without additivity, *Econometrica* 57(3), 571–587

Schönhage, A. (1971), *Approximationstheorie*, de Gruyter, Berlin

Shannon, C.E. (1948), A mathematical theory of communication, *Bell System Technical Journal* 27, 379–423 and 623–656

Shannon, C.E. and W. Weaver (1949), *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL

Sornette, D. (2003), *Critical Phenomena in Natural Sciences: Chaos, Fractals, Self Organization and Disorder: Concepts and Tools*, 2nd edition, Springer, Berlin

Traeger, C. (2011), Subjective risk, confidence and ambiguity, *CUDARE Working Paper* No. 1103

Tsallis, C. (1988), Possible generalization of Boltzmann-Gibbs statistics, *Journal of Statistical Physics* 52, 479–487

von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, 3rd edition. Princeton University Press, Princeton, NJ

Wald, A. (1949), Statistical decision functions, *The Annals of Mathematics and Statistics* 20(2), 165–202

Wiener, N. (1948), *Cybernetics or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge, MA

Yaari, M. (1969), Some remarks on measures of risk aversion and on their uses, *Journal of Economic Theory* 1, 315–329

Zweifel, P. and R. Eisen (2012), *Insurance Economics*, Springer, Berlin and Heidelberg