1 The enigma of simplicity

Hugo A. Keuzenkamp, Michael McAleer and Arnold Zellner

Introduction

Many scientists believe that simplicity is a crucial element in their quest for knowledge. The order, which is found in chaos, it is thought, facilitates understanding, prediction and intervention. This seems particularly proper in statistical science. Simplicity seems to be related to inductive power. Some even hold simplicity as a sign of beauty and assign it intrinsic value (see Derkse, 1992). Simplicity has many faces.

Still, exactly how simplicity is used in scientific reasoning remains enigmatic. Until recently, simplicity remained part of the realm of metaphysics. Recent arguments in the philosophy of science, statistics and econometrics generated wide agreement that the role of simplicity in scientific reasoning is in need of clarification. A multidisciplinary conference was organized and held at Tilburg University in January 1997 to help clarify the definition and role of simplicity in scientific reasoning.[1] The fruits of the conference are presented in this volume of papers.

The idea that simplicity matters in science is as old as science itself. Consider Occam's razor, 'entia non sunt multiplicanda praeter necessitatem': entities are not to be multiplied beyond necessity.[2] A problem with Occam's razor is that nearly everybody seems to accept it, but few are able to define its exact meaning and to make it operational in a non-arbitrary way. Hence, the theme for this volume might well be summarized by the general question of how to clarify the enigma of simplicity in scientific inference. This leads to several specific questions, namely:

[1] Conference on Simplicity and Scientific Inference, Tilburg, The Netherlands, 9–11 January 1997, sponsored by the Center for Economic Research. Eighteen participants from nine countries attended the conference, with the following disciplinary backgrounds (with numbers in parentheses): econometric theory (6), applied econometrics (3), philosophy of science (3), information theory (2), and one from each of statistical theory, probability theory, economic game theory and history of economic thought.

[2] See Keuzenkamp (2000, ch. 5) for further discussion.

1. What is meant by simplicity?
2. How is simplicity measured?
3. Is there an optimum trade-off between simplicity and goodness-of-fit?
4. What is the relation between simplicity and empirical modelling?
5. What is the relation between simplicity and prediction?
6. What is the connection between simplicity and convenience?

In this introduction, we will briefly discuss these questions and note how different contributions to this volume relate to them. The volume has two parts. Part I, 'The Importance of Simplicity', consists of six chapters where the epistemological issues are the main focus. Part II, 'Simplicity in Theory and Practice', consists of ten chapters in which statistical and econometric dimensions of simplicity are the main topic. Part II is concluded with contributions on communication, application and, finally, some reflections by Nobel laureates in economics.

1 What is meant by simplicity?

That simplicity matters in science is a fact which is revealed in scientific practice, but also more explicitly by the answers to a questionnaire which was sent to Nobel laureates in economics. McAleer reports on this in chapter 17. An interesting finding is that the preference for simplicity is almost (not completely – Lawrence Klein being an exception) universal. Equally interesting is that scientists admit to having little analytical understanding of simplicity. Kenneth Arrow made the point succinctly in writing 'I have not thought deeply about it, although it certainly plays a great role in my thinking and that of my colleagues. I know certainly that the term has a number of different meanings.'

Simplicity may mean 'paucity of parameters' (Simon and many others in this volume), 'plausibility' (e.g. Sober), 'strong falsifiability', 'communicability' and many other things. The contributions to this volume illustrate that different meanings do indeed prevail, but also that there is much common ground between scientists of totally different backgrounds.

2 How is simplicity measured?

Sober notes in his contribution that philosophers have little advice on the measurement of simplicity (or plausibility). Moreover, they fail to justify simplicity as a guide to hypothesis choice and have a hard time weighing the desideratum 'simplicity' against other desiderata of scientific models. For Sober, the conclusion is that simplicity may have no global justification, in which case its justification might be problem-specific. Hamminga sympathizes with that conclusion. In 'What explains complexity?', he argues that it is difficult to find anyone opposing simplicity as a desirable feature of a theory or model. However, facing some choice, the recommendation to maximize simplicity is in danger of being empty unless the problem is specified more precisely.

Despite the deep philosophical problems with an operational definition of simplicity, many attempts to make simplicity measurable have been made. A simple measure of complexity seems to be the number of parameters of a statistical model. Such a measure lies behind nearly all more sophisticated measures of simplicity. It is not without problems. Parametric simplicity implies a formal (mathematical) structure – which to some will be quite complex and may, therefore, violate another meaning of simplicity: communicability. Still, as the language of economics is by and large mathematical, this criticism does not seem to be fatal. Indeed, most authors in this volume suggest that simplicity is related in some way to paucity of parameters. But even then, the meaning may remain unclear.

The most prominent early contributor to the statistical literature on simplicity is Harold Jeffreys. His contribution is an integral part of his Bayesian theory of probability and philosophy of science. He summarized his view regarding simplicity pointedly as 'the simpler laws have the greater prior probability' (1961, p. 47).[3] Jeffreys has pioneered the theory of simplicity by providing an operational measure of simplicity (or its counterpart, complexity). The simplicity postulate of Jeffreys in its operational form attaches a prior probability of 2^{-c} to the disjunction of laws of complexity c, where c = order + degree + absolute values of the coefficients of scientific laws, expressed as differential equations.

Jeffreys, who regarded his proposal as a first step in the direction of measuring complexity, did not propose a formal justification for the simplicity postulate and its operational form. Therefore, some critics claim that the measure is arbitrary and sensitive to the choice of the operational 'language' and measurement system which are used in scientific models. Others regard it as a serious contribution to the understanding of the role of simplicity in scientific inference. Various chapters in this volume help to clarify and elaborate Jeffreys' intuitive definition and contribute to appraise its importance in scientific methodology.

The 'language problem' is well known in the philosophy of science, for example, as the so-called riddle of induction, due to Goodman (see the contribution of Sober). In this 'riddle', two propositions are compared:

[3] See also Wrinch and Jeffreys (1921).

(1) all emeralds are green; and (2) all emeralds are green until the year 2050, after that, they are blue. If one defines 'grue' as green until 2050 and blue afterwards, and 'bleen' as blue until 2050 and green afterwards, then the propositions can be rephrased as (1′) all emeralds are grue until the year 2050, after that, they are bleen; and (2′) all emeralds are grue. 'Why', asks Sober, 'should we adopt one language, rather than another, as the representational system within which simplicity is measured?' Philosophers have no answer to this question. Happily, it turns out that the intuitive measure of Jeffreys is an approximation of a formally justifiable measure of complexity, which has been derived in information theory (see below). It can be shown that, asymptotically, the measurement language is not relevant for this information-theoretic measure of simplicity.

In this volume, Zellner analyses and elaborates Jeffreys' approach in detail, and provides illustrative examples. Zellner has long advocated that economists should follow the advice of natural scientists and others to keep their models sophisticatedly simple, especially as simple models seem to work well in practice. Zellner reviews Jeffreys' measure of the complexity of differential equations. Subsequently, he extends it to the context of difference equations and other models that are frequently encountered in statistics and econometrics, among others, allowing for measurement equations as well as state equations, and introducing measures of complexity for various distributional assumptions for error terms.

As noted above, Jeffreys' measure can be related to the information-theoretical measure, due to Kolmogorov. Kolmogorov complexity theory is explained in the chapter by Vitányi and Li (chapter 8). The confrontation between philosophy and information theorists is a rare occasion for improving our understanding of simplicity. The benefit is even larger thanks to Rissanen's contribution on description length (see chapter 9). Those two chapters provide an exceptionally lucid introduction to a highly relevant theme. Ploberger and Phillips (chapter 10) expand this theme to the econometrics context.

Vitányi and Li define Kolmogorov complexity as 'the quantity of absolute information in an individual object'. Rissanen argues that the most appealing concrete complexity measure of a data set is the length of the shortest binary program for the data in a universal computer language. That language, however, is too large to provide a useful complexity yardstick for statistical applications because no algorithm exists to find it, so that finding such a model is a non-computable problem. It is possible to modify the idea by selecting a family of probability distributions as the language (as such, a non-trivial modelling problem), which provides codes as 'programs' for the data. The shortest code length is then taken as the complexity of the data set relative to the chosen family, and the associated distribution is its 'best' description for inductive purposes. Stochastic complexity separates the model from the noise and enables the complexity of the model to be defined exactly. In his chapter, Rissanen provides an operational statistical device to implement those insights, which is the so-called 'minimum description length principle'. This principle also provides an answer to the question, is there an optimum trade-off between simplicity and goodness-of-fit? We turn to this question now.

3 Is there an optimum trade-off between simplicity and goodness-of-fit?

The idea that there is a trade-off between goodness-of-fit and simplicity is common knowledge in statistics. But just how the trade-off should be made remains unclear. Surprisingly, the theory of statistics does not provide a unique answer. Fisher's theory of reduction is based on 'sufficient statistics', where loss of information is measured against an arbitrary (or, should we say, conventional) significance level of mostly 5 per cent. In econometrics, parameter deletion tests are normally based on similar criteria, or on correlation indices, which are corrected for degrees of freedom (like Theil's corrected R^2). Alternatively, statistical information criteria are used to trade-off simplicity and goodness-of-fit. The latter approach is closest to most contributions in this volume.[4]

In chapter 5, Forster argues that simplicity has appeared in statistics with a form and precision that it never attained in the methodology of science. Forster lays a foundation for all forms of model selection, from hypothesis testing and cross-validation to Akaike's information criterion (AIC) and Schwarz's Bayesian information criterion (BIC). These methods are evaluated with respect to a common goal of maximizing predictive accuracy. The general theme is that, along with descriptive accuracy, simplicity is an important ingredient for predictive performance of statistical models.

'Occam's bonus' by Anthony Edwards (chapter 7) notes that the year of the Tilburg Simplicity Conference, 1997, coincided with the twenty-fifth anniversary of three classic contributions to the literature on the

[4] A perspective that is highly relevant to the issue at stake, but not elaborated on in this volume, is the method of posterior odds in Bayesian inference, due to Jeffreys. Posterior odds for alternative models (say a simple model versus a complicated model) reflect relative goodness-of-fit, the extent to which prior information is in agreement with sample information, prior odds and a penalty for relative complexity. See Zellner (1984, repr. 1987), chapters 3.6 (which relates the posterior odds ratio to the Akaike Information Criterion) and 3.7.

trade-off between simplicity and goodness-of-fit. They are Akaike's development of his AIC, Edwards' monograph on likelihood, and Nelder and Wedderburn's generalized linear models (which referred to the trade-off as the 'deviance'). Edwards examined the rate of exchange (trade-off) between support (log-likelihood) and simplicity. He discussed what increase in support is required to justify an increase in complexity of the model, such as the addition of a new parameter, with Occam's bonus referring to such a trade-off. There is a trade-off in allowing for the complexity of a probability model: too little complexity and the model will be so unrealistic as to make prediction unreliable; too much complexity and the model will be so specific to the particular data as to make prediction unreliable.

4 What is the relation between simplicity and empirical modelling?

There has been a long discussion in the econometric literature on the relation between empirical modelling and the complexity of models. One theme, discussed in Keuzenkamp and McAleer (1995), is whether 'general-to-specific' modelling strategies have specific merits (the answer was negative, which is consistent with earlier assessments of Jeffreys and Zellner). There are more issues that relate empirical modelling and simplicity.

Ploberger and Phillips, for example, explain why econometricians are interested in a particular aspect of Rissanen's theoretical results. In a typical empirical modelling context, the data-generating process (DGP) of a time series is assumed to be known up to a finite-dimensional parameter. In such cases, Rissanen's theorem provides a lower bound for the empirically achievable distance between all possible data-based models and the 'true' DGP.[5] This distance depends only on the dimension of the parameter space. Ploberger and Phillips examine the empirical relevance of this notion to econometric time series. Next, they discuss a new version of the theorem that allows for non-stationary DGPs. Non-stationarity is relevant in many economic applications and it is shown that the form of non-stationarity affects, and indeed increases, the empirically achievable distance to the true DGP.

In chapter 11 of this volume, on 'Parametric versus non-parametric inference: statistical models and simplicity', Aris Spanos compares and contrasts two approaches to statistical inference, parametric and non-parametric. The relation between simplicity and goodness-of-fit (or, in

[5] The notion of a 'true' DGP (with probability 1) is controversial, both in probability theory and in the philosophy of science.

his words, statistical adequacy) can best be analysed in a parametric setting. Spanos argues that simplicity is a pragmatic, as opposed to an epistemic, attribute, such as empirical adequacy, which constitutes a necessary condition for choosing among theories using simplicity as a criterion. Spanos relates the notion of simplicity to relevant themes in statistical modelling, such as misspecification testing, data mining and the pre-test bias problem.

Antonio Aznar, Isabel Ayuda and Carmen García-Olaverri examine 'The role of simplicity in an econometric model selection process' (chapter 12). The authors argue that the first condition to be met by any model to be used in empirical science is that it provides a good fit to the available data. However, in order to achieve a good model, this fit is a necessary, though not sufficient, condition. Agreement with the facts can take many forms, but not all of them are equally acceptable. Therefore, apart from goodness-of-fit, models should also satisfy a second condition, associated with the concept of simplicity. The authors justify and analyse a bipolar approach to the evaluation of econometric models, based on a trade-off between goodness-of-fit and simplicity, which move in opposite directions. Alternative forms in which different model selection criteria combine the two elements are also analysed.

Moving from a parametric to a non-parametric theme, Dirk Tempelaar examines 'Simplicity in a behavioural, non-parametric context' (chapter 13). As the concept of simplicity is inextricably related to models, the definition of simplicity depends critically on the modelling paradigm chosen – the theme which is discussed by Sober and Hamminga as well. Tempelaar argues that the parametric versions of the concept of simplicity constitute, in some sense, a paradox. Simplicity, or its mirror image, complexity, is particularly useful in modelling exercises in which theory plays a rather modest role, and the data are more important in deciding upon the model structure. Consequently, there is a need for principles such as Occam's razor. However, with weak theory and strong data, deep structural parameters are scarce, or even absent, thereby questioning the legitimacy of parametric implementations of Occam's razor. An alternative non-parametric definition of the concepts of simplicity and complexity, based on 'behavioural modelling', is sketched.

Uwe Jensen examines 'The simplicity of an earnings frontier' (chapter 16) and provides estimates of an extended human capital model with imperfect information on the employee's side as a stochastic earnings frontier. The costs of the information imperfection are measured by the inefficiency terms of the frontier model. This individual inefficiency in finding suitable jobs is shown to be considerable. The frontier function approach leads to a sensible interpretation of the deviations of empirical income from estimated maximum possible income. Jensen applies the concept of simplicity and descriptive accuracy to examine whether the standard least squares estimation approach or the frontier approach are to be preferred. Comparisons of the models are conducted verbally as there is no computable simplicity criterion in this context.

Herbert Simon contributed chapter 3, entitled 'Science seeks parsimony, not simplicity: searching for pattern in phenomena'. He argues that science aims at parsimony, the ratio of the number of symbols in raw data to the number that will describe their pattern, and thereby describe and explain the facts parsimoniously, enhancing the falsifiability of theories, and bringing simplicity in its wake. By synthesizing simple hypotheses before complex ones, the most parsimonious hypothesis that will account for the phenomena will be obtained. Theories built on structural relations allow predicting the behaviour of systems under partial change in structure. The approximately hierarchical structure of many phenomena admits relatively simple theories, each describing system behaviour at a different level of the hierarchy. Simon relates the roles that parsimony and simplicity have played in his own research, especially in economics and statistics.

Continuing in this vein, Marcel Boumans explores an approach in which complex systems are considered as hierarchic in 'A macroeconomic approach to complexity' (chapter 4). A system is hierarchic when it is composed of interrelated subsystems, each of them being, in turn, hierarchic until some lower level of subsystems that are treated as black boxes is reached. According to Jeffreys, the internal structure of a black box can be taken to be as simple as possible. If the complexity of the system arises from the complicated interaction of the parts, computer simulations are possible. However, if the complexity arises from the multitude of the parts, Simon suggests approximating the system by a nearly decomposable system to simplify the analysis. Boumans explores the development of Simon's approach and its methodology in macroeconomics.

5 What is the relation between simplicity and prediction?

One of the most important justifications for aiming at (some degree of) simplicity is that it supports the predictive performance of models. This justification has been given by Jeffreys and, explicitly or implicitly, many statisticians (see also Forster and Sober, 1994). Based on the principle of Occam's razor, it is widely believed that the better a theory compresses the data, the better the theory generalizes and predicts unknown data. This belief is vindicated in practice but, Vitányi and Li note, has apparently not been rigorously justified in a general setting. For that purpose, they propose the use of the Kolmogorov complexity. They examine the relation between data compression and learning, and show that compression is almost always the best strategy, both in hypothesis identification by using the minimum description length principle and in prediction methods.

6 What is the connection between simplicity and convenience?

In chapter 15, 'Communication, complexity and coordination in games', Mattias Ganslandt investigates how the transmission of information determines collective behaviour in coordination games. Pre-play communication should help players to avoid coordination failures. Furthermore, transmission of information should help players to optimize their collective behaviour. The complexity of an equilibrium is defined as the length of the description the players attach to the equilibrium strategy profile. Ganslandt shows that simplicity can be used to choose from multiple strict Nash equilibria, and that choosing an equilibrium is a trade-off between efficiency and ease of describability (the mirror image of complexity). It is also shown that simple patterns of behaviour occur if talk is costly.

Hamminga focuses on why an increase in complexity is such an almost universally observed empirical phenomenon in dynamic processes (chapter 6). From a philosophical point of view, econometrics is claimed to be a special case of this general process of the evolution of nature towards complexity. Simplicity discussions should be viewed from the perspective of individuals striving to maintain simplicity while their surroundings become more complex. Perhaps the comments by Jeffreys on Schrödinger's wave equation, discussed in Zellner's contribution (chapter 14), clarify some of the issues raised here. Jeffreys suggests that this apparently highly complex differential equation may be the gradual result of successively better approximations: it was likely to be the simplest among competing partial differential equations, not the simplest in an absolute sense.

7 Conclusion

The chapters in this volume clarify the enigma of simplicity, without providing the final answers to all questions raised above. The growing interest in the analysis of simplicity and scientific inference merits further research, where it is likely that interactive contributions from different disciplines will yield much synergy. We hope that this volume stimulates such research and, paraphrasing Occam, would like to thank our contributors for their enormous effort in making their chapters not complex beyond necessity.

REFERENCES

Derkse, W. (1992). *On Simplicity and Elegance*. Delft: Eburon.

Forster, M. R. and E. Sober (1994). How to tell when simpler, more unified, or less *ad hoc* theories will provide more accurate predictions. *British Journal for the Philosophy of Science* 45: 1–35.

Jeffreys, H. (1961). *Theory of Probability*. Oxford: Clarendon Press.

Keuzenkamp, H. A. (2000). *Probability, Econometrics and Truth*. Cambridge: Cambridge University Press.

Keuzenkamp, H. A. and M. McAleer (1995). Simplicity, scientific inference and econometric modelling. *Economic Journal* 105: 1–21.

Wrinch, D. and H. Jeffreys (1921). On certain fundamental principles of scientific inquiry. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science* 42: 369–90.

Zellner, A. (1984, repr. 1987). *Basic Issues in Econometrics*. Chicago: University of Chicago Press.