

Tilburg University

Simplicity, scientific inference and econometric modelling

Keuzenkamp, H.A.; McAleer, M.

Publication date:
1994

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Keuzenkamp, H. A., & McAleer, M. (1994). *Simplicity, scientific inference and econometric modelling*. (CentER Discussion Paper; Vol. 1994-56). Unknown Publisher.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

CBM

841 R

8414

1994

NR.

56

entER

for

Economic Research

Discussion paper



Center
for
Economic Research

8414
1994
56

No. 9456

**SIMPLICITY, SCIENTIFIC INFERENCE
AND ECONOMETRIC MODELLING**

by Hugo A. Keuzenkamp *R18*
and Michael McAleer

August 1994

*JEL B0
JEL B4
JEL C5*

*Econometric Models
Economic School*

! Economic Methodology

ISSN 0924-7815



**SIMPLICITY, SCIENTIFIC INFERENCE
AND ECONOMETRIC MODELLING***

Hugo A. Keuzenkamp

Department of Economics
Tilburg University

Michael McAleer

Department of Economics
University of Western Australia

Revised: June 1994

* The authors wish to thank Dhammika Dharmapala, Jan Magnus and seminar participants at La Trobe University, the London School of Economics, University of Western Australia, University of Limburg, and participants at a workshop on econometric methodology at CentER, Tilburg University, for helpful comments and discussions. The second author wishes to acknowledge the financial support of the Australian Research Council and the Royal Netherlands Academy of Arts and Sciences. Much of this paper was written while the second author was visiting CentER at Tilburg University and the Tinbergen Institute at Erasmus University Rotterdam.

Abstract

Although simplicity (parsimony) seems to be important in scientific inference, it is difficult to find a satisfactory operational definition that can be used to implement an optimal level of simplicity. Many statisticians and econometricians agree that one of their goals is the reduction of data (i.e. formulating models that contain the relevant information in the data, for example, by means of sufficient statistics). However, it is not generally agreed how this goal should be achieved if the specification of the model is not given *a priori*. Two issues are discussed in this paper. The first is whether a formal definition and justification of simplicity in scientific inference can be found, and whether an optimal level of simplicity is obtainable. A definition of simplicity is possible, as are the optimum conditions for the desired degree of simplicity. The model of inference used here relates Bayesian inference to algorithmic information theory. Simplicity is examined in the light of induction, the Duhem-Quine thesis, and bounded rationality. The second issue relates to the role that simplicity might play in econometric modelling. This is elucidated with some remarks on the 'general to specific' approach to modelling and discussions on the purpose of a model.

Keywords: simplicity, Occam's Razor, information criteria, general to specific.

JEL-codes: B0, B4, C5.

"is it not natural to begin any attempt at analysing the economic mechanism by making the simplest assumption compatible with general theory?"

Jan Tinbergen (1940).

0. Preamble

'The march of science is towards unity and simplicity.' (Poincaré, 1902/1952, p. 173). In econometrics, this march threatens to become lost in a maze of specification uncertainty. The familiar problem of data mining is that there are many different ways to deal with specification uncertainty. Reductionism (Hendry, 1993), sensitivity analysis (Leamer, 1978) and profligate Vector Autoregressive (VAR) modelling (Sims, 1980) share, as their starting point, a general, high-dimensional model. The reductionist approach attacks specification uncertainty by means of a general to specific 'simplification search' (in the terminology of Leamer, 1978) of conditioning and marginalizing, in which a sequence of (possibly) asymptotically independent significance tests is used. However, the notion of 'simplicity' is not explicitly formalized, but instead, a conventionally chosen significance level is used. Sims' VAR approach uses information criteria in order to obtain models with 'optimal' lag lengths, but no effort is made to use such criteria for imposing additional zero restrictions on the parameters of the model. As a result, VAR models are often criticized for being 'over-parameterized,' or insufficiently parsimonious (i.e. simple).

In this paper, an effort is made to formalize simplicity (parsimony) and to show how it relates to scientific inference and econometric modelling. In particular, use will be made of insights in algorithmic information theory, the theory of inductive reasoning, and Kolmogorov complexity theory. It will be argued that simplicity is a vital element in a theory of scientific inference. A definition of simplicity is possible, as are the optimum conditions for the desired degree of simplicity. Understanding simplicity is seen to be important for specifying hypotheses and selecting models.

The paper is organized as follows. Three main sections discuss the background and formalization of simplicity with regard to *a priori* probability and information theory (Section 1), its meaning for scientific inference in general and its relation to induction, the Duhem-Quine thesis and bounded rationality (Section 2), and its meaning for econometric modelling in particular (Section 3). Concluding comments are given in Section 4.

1. Formalizing simplicity

1.1 Background: Occam's Razor and the Law of Parsimony

A famous maxim of both philosophy, in general, and scientific inference, in particular, is *Occam's Razor*: 'Entia non sunt multiplicanda praeter necessitatem' (entities are not to be multiplied beyond necessity). This maxim is usually attributed to the Oxford Franciscan scholasticist William of Ockham (±1285-1349), although it cannot be found in his writings.¹ Perhaps for this reason, different versions of the principle circulate,² and even the spelling of Ockham's name varies—the razor has been used to shave it to Occam. Occam's razor is supposed to be one of the canons of scholasticism (see, e.g. Pearson, 1911, p. 393), although Thorburn (1918) argues that, apart from its commonsense as a sound rule of methodology, the razor has little to do with mediaeval scholasticism. Related canons, cited by Pearson (1911, p. 393), are *Principia non sunt cumulanda*, and a statement that can be found in the writings of Ockham as well as those of his teacher Duns Scotus (±1266-1308), *Frustra fit per plura, quod potest fieri per pauciora* (it is in vain to do by many what can be done by fewer).³ This is known as the *Law of Parsimony* (or *Parcimony*).

Whatever the antecedents of Occam's Razor, as a methodological principle it is useful but vague. Other things equal, a more elaborate model cannot fit the data worse than a specific (restricted) version of it. Occam's razor suggests deleting those extensions of a model that are irrelevant to the aim of the model, examples of aims being description and prediction. It is a recurring theme in the writings of early positivists, such as Mach, Poincaré, and Pearson (see e.g. Pearson, 1911). If two hypotheses H_i and H_j describe facts D equally well, i.e. $P(D|H_i) = P(D|H_j)$, the principle says that the simpler of the two should be preferred.

But what is the motivation? Some efforts have been made to base simplicity on metaphysical grounds. An example is the view of Sir Isaac Newton (1642-1725) that '*Natura*

¹ A detailed historical account of Occam's Razor is given in Thorburn (1918), who argues that it is a modern myth. The formula was probably first used by John Ponce of Cork in 1639, and attributed to Ockham by the historian Tennemann in 1812. The English title 'Occam's Razor' was invented by the Scottish philosopher, William Hamilton (1788-1856; not to be confused with the mathematician) in 1852.

² For example, Quine (1987, p. 12) writes: "*Entia*," William of Ockham had intoned, "*non multiplicanda sunt praeter necessitatem*." No reference is given. Rissanen (1983, p. 421) represents 'Ockham's razor' as 'plurality is not to be assumed without necessity.' Again, references are missing.

³ We traced the following versions of Occam's Razor in the economics literature. Friedman (1953, p. 13n) discusses 'Occam's razor' as an 'arbitrary principle', without citing the Latin formula. Klant (1979, p. 49) presents 'Occam's Razor' as '*Entia explicantia non sunt amplificantia praeter necessitatem*.'

enim simplex est et rerum causis superfluis non luxurial' (*Principia Mathematica*, 1687, cited in Pearson, 1911, p. 92).⁴ This amounts to the that nature is simple, and one ought not seek superfluous causes. Such a view has remained popular among physicists. For example, it can be found in a paper on the principles of scientific inquiry by Wrinch and Jeffreys (1921, p. 380): 'The existence of simple laws is, then, apparently, to be regarded as a quality of nature.' We will return to their writings below. Another modern version of Newton's view is given by the physicist Richard Feynman, who asks, How is it possible that we can guess nature's laws? His answer is: 'I think it is because nature has a simplicity and therefore a great beauty' (Feynman, 1965, p. 173). One might object that this view of the physicist is metaphysical speculation, not based on facts. A diametrical speculation is expressed by Leamer and Hendry, who both hold that 'Nature is complex and Man is simple' (Hendry *et al.*, 1990, p. 185).

Occam's razor does not imply that nature is either simple or complex, but only suggests that simplicity is a sound device for inference. It serves as a rule of methodology, not as a metaphysical dogma (Thorburn, 1918, p. 352). Such a view can be found in the writings of positivists such as Mach and Pearson, and also in Peirce's publications on pragmatic philosophy:

'Hypothesis in the sense of the more facile and natural, the one that instinct suggests, that must be preferred; for the reason that, unless man have [*sic*] a natural bent in accordance with nature's, he has no chance of understanding nature at all.' (Peirce, *Abduction and Induction*, in Peirce, 1955, p. 156)

However, there is still the objection that Occam's razor is not operational. When, for example, are new elements in a theory 'redundant'? To answer this question, it is necessary to face the trade-off between simplicity and (descriptive) accuracy. This trade-off has been studied by some of the founders of probability theory. To the best of our knowledge, this attempt was made by Gottfried Wilhelm Leibniz (1646-1716).⁵ He argues that an hypothesis is more probable than another in proportion to its simplicity (economy of assumptions) and its power (number of phenomena that can be explained by the hypothesis) (see Keynes, 1973, p. 303; Cohen, 1989, p. 27). Keynes does not say whether Leibniz elaborates on the possibility of formalizing a trade-off, but it is unlikely that Leibniz did. Keynes himself did not. He needed a principle of parsimony for his own (logical) probability theory, which is known as his *principle of limited independent variety*. This is a principle, invoked in order to ascertain non-zero probabilities a

⁴ Thorburn (1918, p. 349) cites the third edition of the *Principia Mathematica* (1726, p. 387), where the formula '*Natura nihil agit frustra, et frustra fit per plura quod fieri potest per pauciora*' appears, with no mention of Ockham.

⁵ Leibniz used the formula '*Entia non esse multiplicanda praeter necessitatem*' in his inaugural dissertation (see Thorburn, 1918, p. 346), without specific reference to Ockham.

priori if one attempts to make probabilistic inferences (see Keynes, 1973, and Section 2 below). It is not a methodological rule to trade-off simplicity and goodness of fit. Keynes' younger colleague, Frank Ramsey, mentioned the trade-off and suggested making use of maximum likelihood:

'In choosing a system we have to compromise between two principles: subject always to the proviso that the system must not contradict any facts we know, we choose (other things being equal) the simplest system, and (other things being equal) we choose the system which gives the highest chance to the facts we have observed. This last is Fisher's "Principle of Maximum Likelihood", and gives the only method of verifying a system of chances.' (Cited in Edwards, 1992, p. 248)

Kemeny (1953) discusses the same issue. He defines compatibility of a theory with the observations by the condition that the observations must lie within a 99 percent confidence interval (Kemeny, 1953, Definition 1, p. 398). Kemeny suggests adopting a rule of inference, which is to select the simplest hypothesis compatible with the observed data. This is his 'Rule 3' (Kemeny, 1953, p. 397), which has been named 'Kemeny's Rule' by Li and Vitányi (1992). Although intuitively appealing, this is rather *ad hoc*. On the suggestion of philosopher Nelson Goodman, Kemeny (1953, p. 408) remarks that it would be of interest to find a criterion combining an optimum of simplicity and compatibility based on first principles. The optimum trade-off is not presented. Hempel (1966, p. 40) is equally unsuccessful in providing an optimal trade-off.

It is, therefore, often argued that Occam's razor is arbitrary, even as a methodological rule. For example, Friedman (1953, p. 13) argues that predictive adequacy is the first criterion to judge a theory, simplicity the second, while realism of assumptions is of little interest. A well known problem of Friedman's remarks on methodology is his vagueness on terminology: prediction, realism and simplicity are not defined. Moreover, some Bayesian philosophers of science do not support the view that prediction is paramount in the sense that it provides greater support for a theory than does, say, the testing of auxiliary hypotheses (for further details, see Dharmapala and McAleer, 1994). Principles of simplicity that are used in econometrics (such as Theil's R^2 , or the information criteria of Akaike and Schwarz) are often discussed as being useful but *ad hoc* (see e.g. Judge *et al.*, 1985, p. 888; Amemiya, 1985, p. 55).

1.2 Simplicity and *a priori* probability

Wrinch and Jeffreys (1921) provide an explicit discussion on the relevance of simplicity in scientific inference. They note that 'it will never be possible to attach appreciable probability to an inference if it is assumed that all laws of an infinite class, such as all relations involving only analytic functions, are equally probable *a*

priori.' (Wrinch and Jeffreys, 1921, p. 389). In order to enable inference, they invoke two premises: 'Every law of physics is expressible as a differential equation of finite order and degree, with rational coefficients' and 'the simpler the law, the greater is its prior probability' (1921, p. 386). The first premise is like Russell's claim that scientific laws of physics can only be expressed as differential equations (see Russell, 1927, p. 122). Once laws are expressed in differential form, they can be ordered. This is where the two premises meet, as 'The most natural way of well-ordering these are such that those of low order and degree, and involving no numerical constants other than small integers and fractions with small numerators and denominators, come earliest in the sequence.' (1921, p. 390). Although highly informal, this is the first attempt to operationalize the notion of simplicity. Once this is done, the ordering is used in a probability ranking, using the second premise. In practice, Wrinch and Jeffreys (1921, p. 386) argue, simplicity 'is a quality easily recognizable when present'.

In his later writings, Jeffreys further developed his views on simplicity. Jeffreys (1961) argues that there is good reason to give a simple hypothesis a higher prior probability than a complex hypothesis, in particular because simple hypotheses tend to yield *better predictions* than complex hypotheses (i.e. have superior inductive qualities). The argument does not rest on a view that nature is simple. Simple hypotheses are favoured *a priori* on grounds of degrees of rational belief (Jeffreys, 1961, pp. 4-5). As a result, hypotheses of interest can be analyzed using strictly positive prior probabilities. The argument is known as the Jeffreys-Wrinch Simplicity Postulate, henceforth denoted as SP:

Simplicity Postulate (SP). *'The set of all possible forms of scientific laws is finite or enumerable, and their initial probabilities form the terms of a convergent series of sum 1'* (Jeffreys, 1957, p. 36)

Jeffreys (1961, p. 47) summarizes his view as 'the simpler laws have the greater prior probability'. The SP in the crude operational form given in Jeffreys (1961, p. 47) is to attach prior probability 2^{-c} to the disjunction of laws of complexity c , where c is measured by the sum of the order, the degree, and the absolute values of the coefficients of scientific laws, expressed as differential equations. Jeffreys proposes assigning a uniform prior probability distribution to laws with an equal degree of complexity. A serious problem is that his measure of complexity is arbitrary, although more explicit than the one presented in the paper with Wrinch.⁶ Furthermore, as he acknowledges, it

⁶ As an example, Jeffreys (1961, p. 46) considers the hypothesis:

$$s = a + ut + (1/2)gt^2 + a_3t^3 + \dots + a_n t^n.$$

If the first three terms on the right hand side can account for most variation in the data, it would be 'preposterous' to increase the number of adjustable parameters to or even beyond the number of observations: 'including too many terms will *lose* accuracy in

rules out measures of complexity of hypotheses which cannot be formulated as differential equations. Jeffreys provides neither a formal justification for the SP nor for the 2nd rule, and acknowledges that the prior probabilities assigned to scientific laws may not be sufficiently precise (Jeffreys, 1961, pp. 48-49). The SP is an heuristic principle, with one feature that has to be retained in any event:

'If a high probability is ever to be attached to a general law, without needing a change in the form of the law for every new observation, some principle that arranges laws in an order of decreasing initial probability with increasing number of adjustable parameters is essential' (Jeffreys, 1961, p. 48).

A justification for a modified SP, and suggestions for obtaining sufficiently precise prior probabilities needed to implement the modified SP, are given below.

1.3. Simplicity and information theory

A justification for the trade-off between descriptive accuracy and parsimony can be found in the literature on algorithmic information theory. Think of an hypothesis as a string of binary digits (bits), denoted by $\{0,1\}^*$. Similarly, the observations can be represented by another string of bits. It is possible to show that the simplicity of an hypothesis is related to the length of the string describing the hypothesis, and this length in turn is related to a probability measure. A problem is that this length depends on the language in which the hypothesis is expressed. It can be shown, however, that a minimum code length for an hypothesis does exist (Rissanen, 1983), but this code length is not computable. The problem then is to find a good approximation to this minimal code length. Hence, a formal justification of Occam's razor is feasible, although the implementation of the resulting principle is arbitrary. Heuristic arguments will be presented to substantiate this claim.⁷

According to Bayes' Theorem, the posterior probability for an hypothesis H_i given data D is proportional to the prior times the likelihood:

$$P(H_i|D) \propto P(H_i)P(D|H_i). \quad (1)$$

prediction instead of gaining it.'
 Jeffreys (1961, pp. 47-49) provides four examples for evaluating the complexity for differential equations, two of them derived from special cases of the equation above. Complexity is given by the sum of the order, degree and the absolute values of the coefficients. For example, if $s=a$, $ds/dt=0$, which results in a complexity of $1+1+1=3$. For the case $s=a+ut+(1/2)gt^2$, he obtains $d^2s/dt^2=0$ (sic!), which yields a complexity of $2+1+1=4$. A major problem with interpreting these calculations is that Jeffreys does not define what is meant by order, degree and the absolute value of the coefficients. As such, his measure is unclear.

⁷ Proofs can be found in Li and Vitányi (1990, 1992).

It is assumed that the hypotheses under consideration are at most countably infinite.⁸ The goal is to maximize the posterior probability $P(H_i|D)$: the more elaborate is H_i , the better will the hypothesis fit the data and the larger is the likelihood, $P(D|H_i)$. Simultaneously, however, according to the SP, the prior probability, $P(H_i)$, declines with the increasing complexity of the hypothesis.

The more elaborate is an hypothesis, the greater is the number of bits required for its binary representation. Rewriting (1) in negative logarithms (where $\log(\cdot)$ denotes logarithms with base 2, and natural logarithms are indicated by $\ln(\cdot)$), yields:

$$-\log P(H_i|D) = -\log P(D|H_i) - \log P(H_i). \quad (2)$$

Maximizing the posterior probability (selecting the hypothesis that has the highest support of the data) is equivalent to minimizing the expression given in (2).

It can be shown that $\log P(H_i)$ is related to the *descriptive complexity* K of the hypothesis H_i and $\log P(D|H_i)$ is related to the 'self-information' of the data given model H_i (see e.g. Li and Vitányi, 1992). Roughly speaking, K measures the minimum number of bits required to encode a proposition. The relation between complexity and a probability distribution has been established by Solomonoff (1964a) and Levin (see Rissanen, 1983). Solomonoff discusses a general problem of induction: extrapolating a sequence of symbols drawn from some finite alphabet. Given such a sequence, denoted by S , what is the probability that it will be followed by a sequence a ? Hence, the problem is to calculate:

$$P(Sa|S) = \frac{P(S|Sa)P(Sa)}{P(S)}, \quad (3)$$

One of the examples, presented in Solomonoff (1964b), is the extrapolation of a Bernoulli sequence. The analytic approach starts from (2). The prior probability in (2) is obtained by examining how the strings of symbols might be produced by a 'Universal Turing Machine' (UTM).⁹ The crucial insight is that strings with short and/or numerous descriptions (program codes that yield the relevant string as output) are assigned high *a priori*

⁸ Such an assumption is made in many rival approaches to inference. Wrinch and Jeffreys (1921) make it in the context of (necessarist) Bayesian induction. Popper (1959) invokes it for his methodological falsificationism. Good (1968, p. 125), who is a subjective Bayesian, argues that parameter values can, in practice, take only computable values in the sense of Turing, provided that they are specified in a finite number of words and symbols. Ramsey (1974, p. 27), who gives a frequentist approach to model discrimination, takes the number of models as countable for convenience.

⁹ A Turing Machine is basically a computer with infinitely expandable memory. A UTM is a Turing Machine which can simulate any other Turing Machine. If a program code (a finite string of 0's and 1's) c which is fed to the UTM yields S as output, then this is represented by $UTM(c) = S$. There may be many such strings c , but it may be the case that the UTM is unable to produce output. This is known as the halting problem.

probability. Solomonoff gives intuitive as well as quasi-formal motivations for this assignment (a purely formal motivation can be found in Li and Vitányi, 1992). The intuitive motivation is based on two arguments. First, the resulting approach of inference is explicitly related to Occam's razor, 'one interpretation of which is that the more "simple" or "economical" of several hypotheses is the more likely.' (Solomonoff, 1964a, p. 3). Second, strings with many possible descriptions yield high *a priori* probability because of the 'feeling that if an occurrence has many possible causes, then it is more likely.' (Solomonoff, 1964a, p. 8). The quasi-formal motivation for Solomonoff's argument to relate the prior probability to simplicity comes from coding theory, from which a measure is derived about the likelihood that a will follow S . This measure, characterized by the crucial insight presented above, obeys 'most (if not all) of the qualities desired in an explication of Carnap's probability.' (Solomonoff, 1964a, p. 8). Later developments of algorithmic learning theory have improved upon Solomonoff's pioneering work, without losing the basic insights about the relation between simplicity and probability. The result is known today as the Solomonoff-Levin distribution with respect to a UTM, which will be denoted by $P_{UTM}(H_i)$. In case of (3), it can be shown that $2^{-K(Sa)} \leq P_{UTM}(Sa)$ (e.g. Li and Vitányi, 1992). It turns out that Jeffreys' prior probability distribution, based on the Jeffreys-Wrinch Simplicity Postulate, is an approximation to the ideal (or 'universal', as Rissanen, 1983 calls it) prior probability distribution.¹⁰ The Solomonoff-Levin distribution is not based on subjective Bayesian betting, but on 'objective' features of the problem of interest. Therefore, it may be regarded as an element in a 'necessarist' theory of Bayesian inference.

An apparent weakness of the arguments given so far is that the resulting probability distribution is not effectively computable in the sense of Turing. This is due to the halting problem. There are limitations on the ability of the Turing machine to foresee whether a particular program code c will provide any output or will run forever, entering infinite loops. Therefore, a computable approximation to $P_{UTM}(Sa)$ in (3) must be obtained. Solomonoff (1964b) contains a number of examples in which his insights are applied. Rissanen (1983) and Li and Vitányi (1992) show how a number of approaches to inductive inference can be derived from Solomonoff's approach. These approaches include Maximum Likelihood (ML), Maximum Entropy (ME) and Minimum Description Length (MDL, also known as global maximum likelihood), and other approaches that are not used in economics, such as Gold's inductive inference and Valiant learning. Among the most interesting

¹⁰ Jeffreys' distribution has some undesirable properties, most importantly, it is 'improper' (it does not sum to 1). An additional advantage of the Solomonoff-Levin distribution is of course that in this case, a formally satisfactory definition of complexity is given. The formalization has been advanced independently by other authors than Solomonoff as well, in particular, by Kolmogorov - whence the current terminology of 'Kolmogorov-complexity'. See Li and Vitányi (1990) for a useful survey.

computable approximations is Rissanen's (1983) MDL principle.

1.4 Minimum Description Length and Maximum Likelihood

Rissanen (1983) shows the links between the MDL principle and ML and ME, respectively. The rather obvious link to the SP will also be presented. First consider MDL. Recall the code length of a model, parameterized by θ , and the data, represented by x . The length $l(x, \theta)$ consists of the self information plus the model code, and is given by

$$l(x, \theta) = -\log P(x|\theta) + l(\theta). \quad (4)$$

The formula is related to Bayes' theorem, written in logarithmic form. If $2^{-l(\theta)}$ is written as $Q(\theta)$ (which can be shown to satisfy the conditions for a probability distribution, see Rissanen, 1983), then

$$2^{-l(x, \theta)} = P(x|\theta)Q(\theta). \quad (5)$$

The parameter for the code length, θ , is an integer and must have a prior probability distribution on all its possible values, i.e. on all integers. To encode an integer scalar θ , approximately $\log_2 n(\theta)$ bits are required (where the integer $n(\theta)$ is the binary representation of θ divided by its precision). The prior probability $Q(n) = 2^{-\log(n)} = 1/n$ seems a natural prior probability distribution (namely, the uniform distribution proposed by Jeffreys), but this is not a proper one as it does not sum to 1. Another (related) problem is that the computer must be able to separate different strings, without dropping the binary character of encoding. The prefix property is needed to solve this difficulty, where a prefix code is a program or the binary representation of a model. This implies that a number greater than $\log_2 n$ bits is required to encode the integer n . The exact number of minimally required bits of the prefix code is not computable (see Chapter 4, subsection 4.5.3, of Keuzenkamp, 1994). The required code length can be approximated by $\log^*(n) \equiv \log_2(n) + \log_2 \log_2(n) + \dots$ (the finite sum involves only the non-negative terms; see Rissanen, 1983, p. 424).¹¹ The proper prior probability that results is:

$$Q(n) = 1/n \times (1/\log_2 n) \times \dots \times (1/\log_2 \dots \log_2 n) \times (1/c), \quad (6)$$

¹¹ Rissanen (1983) shows that functions other than the \log^* function also have the desired properties, but \log^* is the most efficient to represent the integers. These functions share the property that the first term ($\log x$) is dominant.

where $c = 2.865\dots$. The first (and dominant) fraction corresponds to Jeffreys' non-informative prior.

The universal prior probability distribution is used to construct an 'objective' method of inference, which is sometimes called global maximum likelihood, or the principle of minimum description length (for derivations, see Rissanen, 1983, 1987). Let $P(x|\theta)$ be the likelihood of the N observations of data x , parameterized by the k -dimensional parameter vector θ . $I(\theta) = M(\theta)/N$, where $M(\theta)$ is the Hessian of $-\log P(x|\theta)$ (the information matrix) evaluated at the maximum likelihood estimate. Then the minimal description length MDL of the data encoded with help of the theory, and the theory itself, is (up to a term of order $(\log k)/N$):

$$MDL = \min_{\theta, k} \left\{ -\log P(x|\theta) + \frac{k}{2} \log \left(\frac{2\pi e N}{k} \right) + k \log \|\theta\| \right\}. \quad (7)$$

The third term containing the norm of θ , evaluated at the optimum, makes the criterion invariant to all non-singular linear transformations. As noted in Rissanen (1987, p. 94), the first two terms in the minimized expression correspond to Schwarz's (1978) Bayesian information criterion, which is:

$$SIC = \min_{\theta, k} \left\{ -\ln P(x|\theta) + \frac{k}{2} \ln N \right\}. \quad (8)$$

A related information criterion which is slightly simpler and was derived earlier, is the Akaike Information Criterion, AIC :

$$AIC = \min_{\theta, k} \left\{ -\ln P(x|\theta) + k \right\}. \quad (9)$$

The AIC is derived as an asymptotic approximation to the Kullback-Leibler distance (or entropy). Minimizing AIC has also been called the *Principle of Parsimony* (Sawa, 1978). There are at least three problems related to the AIC . First, it does not define a proper density (Rissanen, 1987, p. 92). Second, it fails to give a consistent estimate of k (see also Schwarz, 1978). Finally, the dependence on some 'true' distribution is problematic for its own sake, as it relies on an assumption that the pseudo-true model is nearly true (Sawa, 1978, p. 1277; see Zellner, 1978, for a comparison of AIC with Bayesian posterior odds; see also Judge *et al.*, 1985; Amemiya, 1985).

Rissanen (1983, p. 428) remarks that applying the MDL principle, while retaining only the first two terms of (7), has been used successfully for the analysis of autoregressive (AR) and autoregressive moving average (ARMA) models. Furthermore,

'In contingency tables, the criterion, measuring the total amount of information, offers a perhaps speculative but nonetheless intriguing possibility to discover automatically inherent links as "laws of nature" in experimentally collected

data. In the usual analyses such links had to be first proposed by humans for a statistical verification or rejection.' (Rissanen, 1983, p. 428)

In other words, Rissanen advocates the automation of science. However, his own methods confirm the view that the smaller is the number of observations, the more arbitrary will be the computable approximation to the 'universal' prior probability distribution.

For completeness, compare Rissanen's prior (6) to Jeffreys' inference based on an improper uninformative prior and his original simplicity postulate. This starts from encoding an hypothesis H to an integer belonging to $\mathbb{N}^+ = \{1, 2, \dots\}$. Examples of efficient coding can be found in Solomonoff (1964b). Assign prior probability $1/i$ to integer i .

Although this results in an improper distribution, as $\sum_{i=1}^{\infty} 1/i$ diverges from 1, the procedure may work well if only a small number of hypotheses is being considered. The prior is an (improper) approximation to the universal prior. To see this, note that $1/i = 2^{-\log(i)}$, whereas Rissanen's (proper) approximation to the universal prior is $2^{-\log^* i}$. The relation between *MDL* and the maximum entropy principle of Jaynes (1989) is established in Rissanen (1983, p. 429) and Li and Vitányi (1992, pp. 375-377).

Obviously, the method of ML is another special case of global maximum likelihood. Given an hypothesis H that defines the parameters θ of a model, and given the data x , the goal is to maximize $\ln P(x|\theta)$. Hence, if the total code length is given by $l(x, \theta) = -\log P(x|\theta) + l(\theta)$, then the ML approach amounts to minimizing the code length for a fixed parameterization θ , as only $\log P(x|\theta)$ is evaluated under the 'axiom of correct specification'. In econometric practice, pure maximum likelihood (i.e. inference without evaluating different specifications) is rarely applied. The attractive feature of global maximum likelihood is that it provides a well-founded criterion for the trade-off between goodness of fit and parsimony. The *modified simplicity postulate* is to minimize the total code length needed to describe the data with respect to a particular model. Unlike Jeffreys' original simplicity postulate, the modified postulate is well defined. An unavoidable arbitrariness in implementing this postulate remains, but the amount of arbitrariness can be made precise and is easy to understand.

2. Simplicity and scientific inference

2.1 Simplicity and induction

Popper (1959, p. 385) argues that simple theories should have low (or zero) prior probability: 'simplicity, or paucity of parameters, is linked with, and tends to increase with, improbability rather than probability.' He concludes that all theories of interest

(in particular, theories that are simple and easy to falsify, i.e. theories with 'high empirical content') are simple to such an extent that they even have zero prior probability. If Popper is correct, then the Bayesian (inductive) approach to inference cannot be sustained, as all posterior probabilities of non-tautological hypotheses would be zero. This is why Popper (1959) argues in favour of falsificationism: if a degree of support for scientific propositions cannot be established, at least false propositions can be eliminated. Science thereby progresses by a sequence of conjectures and refutations.

Such arguments have been criticized by the Bayesian philosopher Howson (1988), who sees no role for simplicity in epistemology. Howson rejects the simplicity postulate, arguing that it is as arbitrary as Popper's opposing argument (see also Howson and Urbach, 1989, p. 292). Howson and Urbach are not the only Bayesians to reject the importance of simplicity. For example, Leamer (1978, p. 203) claims to be agnostic. Leamer (in Hendry *et al.*, 1990, p. 184) argues that simplicity may be helpful for the purpose of communication, but not for inference. However, it will be argued that simplicity is essential for a theory of inference.

Two scholars of induction, Keynes and Jeffreys, emphasized that, in order to use probabilistic reasoning for scientific inference, one has to map probabilities *a priori* to probabilities *a posteriori*. Those probabilities refer to propositions (scientific hypotheses). The probability *a priori* of an hypothesis should not be negligible, otherwise the posterior probability will be zero. If there is no prior information on the merits of alternative hypotheses, it is necessary to make an 'uninformative' prior probability assignment. The uniform distribution has been used for this purpose. If the number of alternative hypotheses is unbounded, each hypothesis will have a prior probability of zero. Hence, the posterior probability of scientific hypotheses will be zero as well, so that inductive inference is not possible. Keynes and Jeffreys suggested different approaches to solve this problem of induction, with Keynes invoking the 'principle of limited independent variety' and Jeffreys the 'Simplicity Postulate'.

Keynes (1973, p. 277) argues that, if every separate configuration of the universe were subject to its own governing law, prediction would become impossible and the inductive method useless. The inductive hypothesis, as he calls his 'principle of limited independent variety', states that, as the number of independent constituents of a system together with the laws of necessary connection become more numerous, inductive arguments become less applicable (pp. 279-280). He does not give a formal definition of these terms, but the principle can be understood as stating that, for inductive inference, the propositions that constitute the premises of an inductive argument must have a high degree of homogeneity. In other words, an object of inductive inference should not be infinitely complex or be determined by an infinite number of generators (p. 286-7). The

reason for this fundamental requirement is that strictly positive prior probabilities are needed for inductive inference. In Keynes' view, these prior probabilities are assessed by reliance on *analogy*. If every fact would have its own cause or generator, then this method of reasoning by means of analogy breaks down, and induction becomes impossible.

Keynes does not argue that this principle is always satisfied (the debate with Tinbergen, for example, was inspired by Keynes' view that in the case of inference with regard to investment, it is *not* satisfied). Rather, Keynes holds that, given *a priori* uncertainty, such a principle is needed if appreciable *a posteriori* probabilities regarding propositions of interest are to be obtained. How to assess the validity of using the principle in specific problems of inference remains unclear. Keynes' colleague at Cambridge, Jeffreys, also tried to solve the problem of obtaining non-arbitrary *a priori* probabilities in the absence of information, but from a different perspective. Even in cases of *a priori* uncertainty, and in case of an unlimited number of possible explanations ('independent variety', in Keynes' theory), strictly positive prior probabilities may be obtained by ranking the possible explanations in order of their complexity.

Jeffreys (1961, p. 342) interprets the statement 'entities are not to be multiplied without necessity' as follows:

'Variation is random until the contrary is shown; and new parameters in laws, when they are suggested, must be tested one at a time unless there is specific reason to the contrary.' (italics in original)

In the preface to the third edition of his *Theory of Probability*, Jeffreys (1961, p. viii) notes that the implications of this principle are contrary to 'the nature of induction as understood by philosophers' (presumably, Popper is implicated). The important point, which can already be found in Pearson (1911), is that one starts with the hypothesis that variation in the data is random, and then gradually elaborates upon a model to describe the data in order to improve upon the approximation. Furthermore,

'Any clearly stated law has a positive prior probability, and therefore an appreciable posterior probability until there is definite evidence against it' (Jeffreys, 1961, p. 129).

The simplest hypothesis is that variation is random until the contrary is shown, the onus of the proof resting on the advocate of the more complicated hypothesis (Jeffreys, 1961, pp. 342-3).

It is now possible to compare Jeffreys' approach with Popper's. Although it is acknowledged that Jeffreys' SP is made operational in an arbitrary way, the underlying argument is not arbitrary. More recent advances, in particular due to Rissanen (1983), provide a non-arbitrary operational form of the SP. On the other hand, Popper's critique

of assigning positive probability *a priori* to simple propositions is self-defeating. The argument is based on two points: first, the idea that the number of possible propositions is infinite; second, that the empirical content of simple propositions is high, so that the chances of falsification are also high. On the first point, if the number of conjectures is indeed infinite, the chance that a sequence of conjectures and refutations will come to an end by hitting the truth is zero. A methodology based on gradual approximation does not suffer from such a weakness. The second point underlying Popper's argument is due to confusing $P(H)$ and $P(D)$ (see Section 1.3 above and, for further details, Keuzenkamp, 1994, p. 74), a topic that is beyond the scope of this paper.

2.2 Simplicity and the Duhem-Quine thesis

The Duhem-Quine thesis of testing theories suggests that a negative test result cannot disconfirm a theory: the rejection affects one (or more) element(s) of a whole test system, but does not indicate which element is invalidated by the test. If a prediction of a scientific theory is shown to be wrong, one can adjust any part of the test system (sometimes called 'web' or 'cluster'), and not just the hypothesis of interest. The simplicity postulate reduces the impact of the Duhem-Quine thesis, although its logical status is not altered.

According to the modified simplicity postulate, simple models have stronger *a priori* predictive power (in the sense of minimizing a mean squared prediction error) than complex models. Adding *ad hoc* auxiliary hypotheses to a model, and using them for modified predictions based on the revised test system, will reduce the *a priori* predictive power of the model. The reason is that *ad hoc* is unrelated to the other theoretical notions that are used to establish a model. Hence, *ad hoc* alterations of theory need a larger additional code (bits) and, therefore, increase the total descriptive length. Negative test results should lead to efforts to adapt the model as coherently as possible, minimizing *ad hoc* (hence, additional code length). This view receives support from Quine's (1987, p. 142) suggestion to introduce the 'maxim of minimum mutilation' in responding to a test result: 'disturb science as little as possible, other things equal'.

This is important for discussions of data mining. The data miner has absolute freedom to make any changes in the specification of a model. As the best (perfect) fit that can be obtained will be the one resulting from a model with as many parameters as there are observations, the data miner tends to make an implicit trade-off between fit and simplicity. For example, the SAS statistical package provides statistical routines where, given a set of data for particular variables, the linear model yielding a maximum R^2 corrected for degrees of freedom is selected. The stepwise regression algorithm is a

simplified version of this automated modelling procedure. A major problem with such routines is the *ad hoc* nature of the resulting models. Simplicity and fit are considered and, if description of a particular set of data is the only purpose of interest, then the resulting model may be satisfactory. Similarly, if a Box-Jenkins model is chosen by optimizing a particular information criterion, a satisfactory model for describing the given data may be obtained. However, such models are idiosyncratic and do not share general results with other models. Hence, investigators with broader aims (such as scientific inference, induction) are unlikely to opt for mechanical modelling using, e.g., stepwise algorithms. The role of economic theory is to highlight general characteristics of different sets of data, in which case economic theory is a simplifying device.¹² This can be formulated in terms of the modified simplicity postulate. If an investigator analyzes consumption in both the USA and the UK, applying Box-Jenkins techniques is likely to yield two statistically adequate models with very different specifications and, hence, a larger number of bits will be required to describe them. However, a theoretical straitjacket is likely to yield models that are slightly worse in terms of goodness of fit, but with greater simplicity in terms of the number of bits due to the similar specifications involved. This analysis of the Duhem-Quine problem avoids the problems that result from a Lakatosian interpretation of scientific research programmes (advocated by Cross, 1982).¹³

2.3 Bounded Rationality

The theory of bounded rationality, due to Simon (see Simon, 1986, for references), is currently experiencing a revival in game theory, as well as in the theory of learning in rational expectations models. Situations of boundedly rational choice emerge for basically the same reasons as those that prevent use of the Solomonoff-Levin distribution directly. Some problems are not computable. In microeconomic phenomena, this bears on situations of strategic behaviour: in complex situations, one may opt for simple strategies. In macroeconomics, disputes on the validity of some economic theories originate from different views on where exactly to locate bounded rationality: for example, consumers who suffer from money illusion, or entrepreneurs who are unable to discriminate between relative and absolute price changes. These implications for economic

¹² Recall that the information criteria, presented above, are derived by starting from prior ignorance about rival models. Once an investigator has been able to extract general information from specific data, the assumption of prior ignorance is no longer valid and informative probability distributions may be used.

¹³ This interpretation, due to Lakatos (1970), depends on the notions of a hard core, protective belt and positive heuristic, which are hard to define. The criteria to decide whether to continue supporting a research programme, or to drop it, are arbitrary and undefined.

theory extend beyond the scope of this paper, but there is a relation with (econometric) inference.

Keynes' (1973) famous example of the beauty contest is related to this argument. According to Keynes, the probability that one model will be chosen as the newspaper readers' beauty queen cannot be calculated. It involves infinite regress on behalf of the readers of the newspaper, who not only have to state their own preferences, but also to guess the preferences and guesses of other readers (the newspaper game was not to choose the beauty queen, but to guess who would be chosen). Keynes' example is an intuitive version of the halting problem. His conclusion is that, in a broad class of applications, it is not possible to formulate numerically precise probability judgments. However, it may still be feasible to assign probability orderings.¹⁴

If the problem of deriving explicit and non-arbitrary prior probability distributions did not exist, then most philosophers (perhaps even statisticians) would agree that Bayesian inference is optimal (namely, it results in coherent probability judgments and satisfies the likelihood principle). Bayesian prediction is an instance of Solomonoff's predictor which is considered ideal for the purpose of predictive inference (Li and Vitányi, 1992). However, these optimal qualities may be too good to be true. The Bayesian inference machine (conditioning, marginalizing and mapping) is similar to Laplace's demon.¹⁵ The demon is empirically omniscient and knows the deterministic laws of physics (such as gravitation). A *Bayesian* demon, to paraphrase the metaphor, is logically omniscient by having an exhaustive set of hypotheses with their ideal universal prior probabilities. The Bayesian demon knows in advance how to respond to every new piece of information: there can be no surprises (this is similar to the complete markets hypothesis of an Arrow-Debreu economy, where it is possible to make perfect contingency plans).

Kolmogorov complexity theory suggests why this fiction of the Bayesian demon must be defective. Kemeny expresses his doubts as follows:

'Few, if any modern philosophers still expect fool-proof rules for making inductive inferences. Indeed, with the help of such rules we could acquire infallible knowledge of the future, contrary to all our empiricist beliefs.' (Kemeny, 1953, p. 711).

Laplace's demon did not survive the quantum revolution in physics. Similarly, the Bayesian demon is set back by Gödel's incompleteness theorem or, which amounts to the same thing, the non-computability of the universal prior (Li and Vitányi, 1990, pp. 208-

¹⁴ Good (1965) and Leamer (1978) have continued this line of thought.

¹⁵ Laplace's deterministic view on the character of the universe can be found in his *Théorie analytique des probabilités*, not in his *Traité de Mécanique Céleste*; see Suppes (1970, p. 32). Laplace uses probability theory for problems where ignorance of complex causes prevails.

214, establish the link between complexity and Gödel's theorem). Even the most extreme 'objectivist' or 'necessarist' version of Bayesianism, based on the notion of a universal prior, must allow for a pinch of subjectivism.

The possibility of indeterminate decision problems, due to the non-computability problem, has recently been discussed by game theorists (e.g. Binmore, 1991; Rubinstein, 1990). The problem is relevant for situations of strategic decision making. Examples are tax wars, interest rate and exchange rate policy on the macro level, or currency and stock market speculation at the micro level. Complexity considerations affect the choice of strategies (see Rubinstein, 1990, p. 17). Although outcomes of decisions need not be completely chaotic (namely random without following a probability law), they may well be.

3. Simplicity and econometric modelling

3.1 General to specific modelling

An implication of the argument on bounded rationality given in the preceding section is that the search for 'the' Data Generating Process, by those who hold the DGP as more than a convenient fiction, is in vain. Such a process does not exist: it is a metaphysical fiction, so that the DGP has the same fate as Laplace's long-deceased demon. Bounded rationality implies that some decisions are made on arbitrary grounds, yielding arbitrary consequences. Of course, it is also possible that 'rules of thumb' emerge, perhaps conventions that have proved themselves in previous instances. Such rules may generate relative stability until they eventually collapse due to some shock. The characteristics of the successive process are not determined, and a regime shift may result. Hence, a stable DGP is either a fiction or a meaningless tautology. As a formal tool for modelling, the less suggestive terminology 'general model' is to be preferred.

If the data are represented by a statistical model, it is desired that the error terms and residuals of the model are stable, and obey certain probability laws. However, unlike the claims of reductionists, nothing forces the (metaphysical) DGP to being stable. Although constant parameters may be desirable, this is not because human nature or society is stable. Models with time varying parameters may be less useful in physics, but can be helpful in the social sciences.

The frequentist interpretation of probability, which presumes the existence of a stable universe (consider for example Von Mises' conditions of convergence and randomness, discussed in Keuzenkamp, 1994), does not apply to the social domain. This does not make inference impossible, but one should explicitly acknowledge the cognitive limitations in the theory of inference that is used. Bayesian inference with a 'human face' can deal with those limitations.

Now consider the reductionist approach to econometrics. This starts from the presumption of the DGP and views econometrics as the reduction of this DGP to a parsimonious model. It is argued that one has to start from a general model and test downward (i.e. test successive restrictions on the general model). The general model is, in fact, a set of models containing all possible simplifications as elements. Hence, the larger is the set of models, the higher is the joint prior probability of all enclosed models. In Poirier's (1988) terminology, the 'window' is wide. However, there is no particular reason why the most general model considered deserves special credence. The argument that test statistics are not valid, in the Neyman-Pearson sense, if they are applied within the context of invalid reductions, is problematic, since even the most general conceivable model is likely to be 'wrong'. This is so even if the reductionist econometrician obeys Haavelmo's (1944) imperative of not choosing the general model by observing the data. Testing downward is sensible if one favours parsimony, but the theory of reduction does not offer satisfactory principles of simplicity. This critique is amplified by the fact that Neyman-Pearson tests treat the null and the alternative hypotheses in a quite arbitrary asymmetric way. In contrast, inference based on the modified simplicity postulate does not suffer from this bias (see Keuzenkamp and Magnus, forthcoming, for discussion of different approaches to testing in econometrics).

In empirical investigations, simple models can inspire knowledge, while complex models rarely do. According to the modified simplicity postulate, simple models should come first to mind (or, what is not quite the same, have lower descriptive complexity, hence higher prior probability). It may be the case, of course, that the investigator chooses a 'microcosm' (in the words of Savage, 1954) that is too small. The investigator may become aware of this if, for example, the chosen set of models does not perform satisfactorily. There is presently no formal theory that provides an objective criterion to suggest when the 'microcosm' should be enlarged, but misspecification tests seem to be relevant for this purpose. As soon as a wider set of models is considered, however, the MDL principle can be applied to evaluating these new models.

The idea that the only valid way of inference proceeds by testing downward, starting from maximal attainable complexity, rests on a misconception of scientific inference. Jeffreys (1961, p. vii) remarks that

'scientific method depends on considering at the outset the hypothesis that variation of the data is completely random, and modifying it step by step as the data are found to support alternatives.'

Furthermore, Jeffreys (1957, p. 78) argues that the starting point for inference is to regard all variation in data as random, and

'then successive significance tests warrant the treatment of more and more of it as predictable, and we explicitly regard the method as one of successive

approximation.⁷

Jeffreys (1957, p. 76) also argues that if the method of successive approximation is not used, one is committed to explaining in detail every separate residual. Jaynes (1989) has a similar view on scientific progress, which is based on the Maximum Entropy Principle. Rosenkrantz (1983, p. 78) defends such a strategy, and calls it 'structured focusing'. His approach starts with constructing a good first-order approximation (representation, in the terminology of Hacking). Anomalies are observed and accommodated by more refined approximations. The reverse approach, from general to simple, may be very cumbersome, and may even lead the researcher astray (see the views of DNA researcher Watson, cited in Rosenkrantz, 1983). In econometrics, Zellner is a rare exception in support of the simplicity postulate; Zellner (1982) opposes a 'top-down' (general to specific) approach.

There exists widespread disagreement on the role of simplicity in inductive (econometric) inference. The basic insight presented here is that simplicity, or its mirror image complexity, can be measured by the minimum number of bits needed in a computer program that describes a theory and its observations. Note that Lucas (1980) states as an explicit goal of economics to write such a computer program. At the optimum level, the joint complexity of data, as described by the model, and the model description itself, are minimized.

This way of formalizing simplicity is instructive for various reasons. It clarifies a limitation to induction, due to the non-computability of an ideal universal prior. The skillful (but subjective) hand of the econometrician is indispensable. Econometric inference is boundedly rational. The general to specific approach suffers from the weakness that it is not possible in empirical applications to estimate a general model incorporating all conceivable variables that may affect the variables of interest. Obviously, one cannot estimate the DGP, even if it were to exist. A general empirical model is, literally, false. What is of interest is how well the model is able to represent the data, predict new data, or provide insights to economic policy. Whether the model mimics the DGP is a problem of metaphysics. If there is reason to think that models of different complexity might be equal candidates for some given purpose, all of them should be considered and their quality evaluated using a modified simplicity postulate.

3.2 The purpose of the model

Has simplicity a virtue independent of the goal of inference? Rissanen's earlier cited remark, wherein he reveals a hope for the automation of science, suggests that simplicity is an objective characteristic independent of subjective judgments of an investigator or independent of the context of modelling. Rissanen (1987, p. 96) argues that the introduction of subjective judgments in inferential problems make the resulting

inferences 'strictly speaking unscientific'. Keuzenkamp (1994, Chapter 4) deals with problems of objectivity, and argues that such objectivity is misleading (this is particularly true for economic inference, where data information is limited and value judgments are inevitable). The modified simplicity postulate may help, however, to yield greater agreement on how to trade off parsimony and descriptive accuracy. Algorithmic information theory shows why certain constraints to objective knowledge may exist, but it does not lead to the conclusion that subjective judgments are unscientific.

Subjective judgments matter, and so does the context of inference. Prediction is different from regulation and control, for example. In general, more complex models are needed if policy intervention is the purpose of modelling (this conviction at least stimulated large scale macroeconomic modelling in the '60s and '70s, and subsequently their complex successors with more elaborate microfoundations). Nevertheless, this does not violate the (modified) SP. Given the goal of inference, the rule should be to select the set of models that is able to meet the goal, so that the SP is used in order to choose the appropriate model.

It might be objected that the MDL principle does not pay sufficient attention to the purpose of the modeller. If, from the MDL perspective, Einstein's relativity theory is superior to classical mechanics in describing red shifts in the spectra emitted by stars, should it be used to predict the position at time t of a football that has been set in motion at $t-1$? The answer is negative: if the input data are macroscopic, then it is likely that inductive inference related to the position of the football will yield superior results using classical mechanics. The MDL principle would suggest that the extra bits to describe the relativity part of the more complex model do not improve the predictions and should, therefore, be disregarded. No physicist would be likely to consider starting from the relativity model and test downwards, in this case.

If the goal of inference is to make money, for example, by predicting stock prices, the criterion is not to maximize simplicity but to maximize dollars. The example in Rissanen (1987) is interesting in this respect: applying the MDL principle leads to the random walk model of stock prices. A model with trend fits the data better, but the extra storage code needed to describe the trend parameter does not weigh against the improvement in the fit of the model. Rissanen (1987, p. 97) concludes, wrongly, that this proves that 'any successful stock advice must either utilize inside information, which is not in the data, or it is entirely the result of luck'. While the quoted remark may be true, this cannot be inferred from the fact that the trend in the data has only limited descriptive power. A money-maker might still prefer the model *with* trend (but even then, transaction costs are probably higher than the expected returns).

4. Concluding comments

In this paper, the concept of simplicity and its effect on different approaches to the problem of specification uncertainty were evaluated. Currently, statistical reductionism is in vogue. It argues that proper inference starts with a general model, or ideally, even the Data Generating Process, and descends step by step to the more specific and simple model, until a critical line on loss of information is crossed. However, general to simple is an inversion of the history of scientific inference. Scientific progress in physics illustrates a sequence of increasing complexity: consider planetary motion, from circles to ellipses, to even more complicated motions based on relativity; from molecules to atoms, and to even smaller particles. Keplers' laws were misspecifications, but were they invalid? The same question can be raised for cases in empirical econometrics, such as the investigation of consumer demand. The first econometric studies were based on very simple models, with few observations. Current investigations are not infrequently based on more than 10,000 observations, and are much more general in being less aggregated with more flexible functional forms. Were the pioneers wrong and misguided by their simple misspecifications? Although reductionism does not pretend to be a theory of scientific development, such questions point to the defect it has as a modelling strategy. All models are wrong: Kepler's, Tinbergen's, and Hendry's alike. Some, however, are useful. Kepler's clearly were, as were Tinbergen's. Whether the same applies to the models (and hence, the modelling *strategies*) of reductionist econometricians still has to be shown.

Given the limited amount of data, a general to specific approach (i.e. starting from a very large dimensional parameter space) is not feasible: one typically commences inference with relatively simple models. However, this may not be the most important problem with reductionism. Following the arguments of this paper, a relatively complex model should receive low prior probability if inductive inference is the goal.¹⁶ Therefore, the only justification for making models more complex is upon empirical observation that the simple model is defective. This does not imply that practical researchers have to follow a simple to general approach instead. Box (1980, p. 425) suggests that scientific progress involves an iterative inductive-deductive course. Although this point of view has merit, it would be useful to add another iterative process, namely between approximation (inspired by the simple to general approach) and reduction. Applied econometricians usually start with relatively simple models, which become gradually more complicated if additional data become available or if the simple models are unable to cope with the specific demands of the investigator. Adding variables

¹⁶ For *deductive inference*, e.g. refining general equilibrium theory, complexity does not have to receive a penalty, as there is no trade-off such as the one that inspires the modified simplicity postulate.

(or lags) may also result in the belief that some others may be redundant, resulting in a simplifying iteration (such as the 'simplification search' of Leamer, 1978). The iteration is also between Savage's (1954) grand world and the search for a small world microcosm. Specification freedom is a nuisance to purists, but is an indispensable aid to practical econometricians. The key to coping with specification uncertainty is to obtain a better understanding of the importance of simplicity.

References

- Amemiya, Takeshi (1985), *Advanced Econometrics*, Harvard University Press, Cambridge.
- Binmore, Ken (1991), Modeling rational players II, *Economics and Philosophy* 4, 9-55.
- Box, G.E.P. (1980), Sampling and Bayes inference in scientific modelling and robustness, *Journal of the Royal Statistical Society A*, 143, 383-430 (with discussion).
- Cohen, L. Jonathan (1989), *An Introduction to the Philosophy of Induction and Probability*, Clarendon Press, Oxford.
- Cross, Rod (1982), The Duhem-Quine thesis, Lakatos and the appraisal of theories in macroeconomics, *Economic Journal* 92, 320-340.
- Dharmapala, Dhammika and Michael McAleer (1994), Prediction and accommodation in econometric modelling, *Environmetrics* 5, to appear.
- Edwards, A.W.F. (1992), *Likelihood*, expanded edition (first ed. 1972), Johns Hopkins University Press, Baltimore.
- Feynman, Richard P. (1965), *The Character of Physical Law*, MIT Press, Cambridge.
- Friedman, Milton (1953), *Essays in Positive Economics*, Chicago University Press, Chicago.
- Good, Irving J. (1965), *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*, MIT Press, Cambridge.
- ____ (1968), Corroboration, explanation, evolving probability, simplicity and a sharpened razor, *British Journal for the Philosophy of Science* 19, 123-143.
- Haavelmo, Trygve (1944), *The Probability Approach in Econometrics*, supplement to *Econometrica* 12, July.
- Hempel, Carl G. (1966), *Philosophy of Natural Science*, Prentice Hall, Englewood Cliffs.
- Hendry, David F. (1993), *Econometrics: Alchemy or Science, Essays in Econometric Methodology*, Blackwell, Oxford.
- ____, Edward E. Leamer and Dale J. Poirier (1990), The ET dialogue: a conversation on econometric methodology, *Econometric Theory* 6, 171-261.
- Howson, Colin (1988), On the consistency of Jeffreys' simplicity postulate, and its role in Bayesian inference, *The Philosophical Quarterly* 38, 68-83.
- ____ and Peter Urbach (1989), *Scientific Reasoning: The Bayesian Approach*, Open Court, La Salle, Illinois.
- Jaynes, Edwin T. (1989), *Papers on Probability, Statistics and Statistical Physics*, edited by Roger D. Rosenkrantz, Kluwer, Dordrecht.
- Jeffreys, Harold (1957), *Scientific Inference* (second edition; first edition 1931), Cambridge University Press, Cambridge.
- ____ (1961), *Theory of Probability* (third edition; first edition 1939), Clarendon, Oxford.

- Judge, George G., W.E. Griffiths, R. Carter Hill, Helmut Lütkepohl and Tsoung-Chao Lee (1985), *The Theory and Practice of Econometrics*, 2nd ed., New York, Wiley.
- Kemeny, John G. (1953), The use of simplicity in induction, *The Philosophical Review* **62**, 391-408.
- Keuzenkamp, Hugo A. (1994), *Probability, Econometrics and Truth*, unpublished Ph.D. thesis, Tilburg University.
- and Jan Magnus (forthcoming), On tests and significance in econometrics, *Journal of Econometrics*.
- Keynes, John Maynard (1973), *A Treatise on Probability, The Collected Writings of John Maynard Keynes*, Volume VIII, edited by Donald Moggridge, St. Martin's Press, New York.
- Klant, J.J. (1979), *Spelregels voor Economen*, (translation: *The Rules of the Game*, Cambridge University Press), Stenferd Kroese, Leiden.
- Lakatos, Imre (1970), Falsification and the methodology of scientific research programmes, in: Lakatos and Musgrave (eds.), *Criticism and the growth of knowledge*, Cambridge University Press, Cambridge, 91-196.
- Leamer, Edward E. (1978), *Specification Searches*, Wiley, New York.
- Li, Ming and Paul M.B. Vitányi (1990), Kolmogorov complexity and its applications, in J. van Leeuwen (ed.), *Handbook of Theoretical Computer Science*, Elsevier, Amsterdam.
- (1992), Inductive reasoning and Kolmogorov complexity, *Journal of Computer and System Sciences* **44**, 343-384.
- Lucas, Robert E. (1980), Methods and problems in business cycle theory, reprinted in Lucas (1981), *Studies in Business-Cycle Theory*, The MIT Press, Cambridge, 271-296 from *Journal of Money, Credit and Banking*, **12**.
- Pearson, Karl (1911), *The Grammar of Science*, third edition (first edition 1892), Adam & Charles Black, London.
- Peirce, Charles S. (1955), *Philosophical Writings of Peirce*, selected and edited by Justus Buchler, Dover Publications, New York.
- Poincaré, H. (1902), *Science and Hypothesis*, 1952 Dover edition, New York.
- Poirier, Dale (1988), Frequentist and subjectivist perspectives on the problem of model building in economics (with discussion), *Journal of Economic Perspectives* **2**, 121-170.
- Popper, Karl R. (1959), *The Logic of Scientific Discovery*, 3rd revised edition, Harper and Row, New York.
- Quine, Willard van Orman (1987), *Quiddities*, Harvard University Press, Cambridge.
- Ramsey, James B. (1974), Classical model selection through specification error tests, in Paul Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 13-47.
- Rissanen, Jorma (1983), A universal prior for integers and estimation by minimum description length, *Annals of Statistics* **11**, 416-431.

- ____ (1987), Stochastic complexity and the MDL principle, *Econometric Reviews* **6**, 85-102.
- Rosenkrantz, Roger (1983), Why Glymour is a Bayesian, in John Earman (ed.), *Testing Scientific Theories, Minnesota Studies in the Philosophy of Science X*, University of Minnesota Press, Minneapolis, 69-97.
- Rubinstein, Ariel (1990), New directions in economic theory - bounded rationality, *Working Paper* no. 11-90, Foerder Institute for Economic Research, Tel Aviv.
- Russell, Bertrand (1927), *An Outline of Philosophy*, George Allen & Unwin, London.
- Savage, Leonard J. (1954), *The Foundations of Statistics*, second revised edition, Dover Publications, New York 1972.
- Sawa, Takamitsu (1978), Information criteria for discriminating among alternative regression models, *Econometrica* **46**, 1273-1291.
- Schwarz, G. (1978), Estimating the dimension of a model, *Annals of Statistics* **6**, 461-464.
- Simon, Herbert A. (1986), Rationality in psychology and economics, in Robin M. Hogarth and Melvin W. Reder (eds.), *Rational Choice*, University of Chicago Press, Chicago, 25-40.
- Sims, Christopher (1980), Macroeconomics and reality, *Econometrica* **48**, 1-48.
- Solomonoff, Ray (1964a), A formal theory of inductive inference I, *Information and Control* **7**, 1-22.
- ____ (1964b), A formal theory of inductive inference II, *Information and Control* **7**, 224-254.
- Suppes, Patrick (1970), *A Probabilistic Theory of Causality*, North Holland, Amsterdam.
- Thorburn, W.M. (1918), The myth of Occam's razor, *Mind* **27**, 345-53.
- Tinbergen, Jan (1940), On a method of statistical business-cycle research. A reply, *Economic Journal* **50**, 141-154.
- Wrinch, Dorothy and Harold Jeffreys (1921), On certain fundamental principles of scientific inquiry, *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science* **42**, 369-390.
- Zellner, Arnold (1978), Jeffreys-Bayes posterior odds ratio and the Akaike information criterion for discriminating between models, *Economic Letters* **1**, 337-342.
- ____ (1982), Basic issues in econometrics, past and present, *American Economist* **26**, 5-10.
- 42**, 278-284 (with discussion).

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9355	Z. Yang	A Simplicial Algorithm for Computing Robust Stationary Points of a Continuous Function on the Unit Simplex
9356	E. van Damme and S. Hurkens	Commitment Robust Equilibria and Endogenous Timing
9357	W. Güth and B. Peleg	On Ring Formation In Auctions
9358	V. Bhaskar	Neutral Stability In Asymmetric Evolutionary Games
9359	F. Vella and M. Verbeek	Estimating and Testing Simultaneous Equation Panel Data Models with Censored Endogenous Variables
9360	W.B. van den Hout and J.P.C. Blanc	The Power-Series Algorithm Extended to the <i>BMAP/PH/1</i> Queue
9361	R. Heuts and J. de Klein	An (s,q) Inventory Model with Stochastic and Interrelated Lead Times
9362	K.-E. Wärneryd	A Closer Look at Economic Psychology
9363	P.J.-J. Herings	On the Connectedness of the Set of Constrained Equilibria
9364	P.J.-J. Herings	A Note on "Macroeconomic Policy in a Two-Party System as a Repeated Game"
9365	F. van der Ploeg and A. L. Bovenberg	Direct Crowding Out, Optimal Taxation and Pollution Abatement
9366	M. Pradhan	Sector Participation in Labour Supply Models: Preferences or Rationing?
9367	H.G. Bloemen and A. Kapteyn	The Estimation of Utility Consistent Labor Supply Models by Means of Simulated Scores
9368	M.R. Baye, D. Kovenock and C.G. de Vries	The Solution to the Tullock Rent-Seeking Game When $R > 2$: Mixed-Strategy Equilibria and Mean Dissipation Rates
9369	T. van de Klundert and S. Smulders	The Welfare Consequences of Different Regimes of Oligopolistic Competition in a Growing Economy with Firm-Specific Knowledge
9370	G. van der Laan and D. Talman	Intersection Theorems on the Simplotope
9371	S. Muto	Alternating-Move Preplays and $vN - M$ Stable Sets in Two Person Strategic Form Games
9372	S. Muto	Voters' Power in Indirect Voting Systems with Political Parties: the Square Root Effect

No.	Author(s)	Title
9373	S. Smulders and R. Gradus	Pollution Abatement and Long-term Growth
9374	C. Fernandez, J. Osiewalski and M.F.J. Steel	Marginal Equivalence in v -Spherical Models
9375	E. van Damme	Evolutionary Game Theory
9376	P.M. Kort	Pollution Control and the Dynamics of the Firm: the Effects of Market Based Instruments on Optimal Firm Investments
9377	A. L. Bovenberg and F. van der Ploeg	Optimal Taxation, Public Goods and Environmental Policy with Involuntary Unemployment
9378	F. Thuijsman, B. Peleg, M. Amitai & A. Shmida	Automata, Matching and Foraging Behavior of Bees
9379	A. Lejour and H. Verbon	Capital Mobility and Social Insurance in an Integrated Market
9380	C. Fernandez, J. Osiewalski and M. Steel	The Continuous Multivariate Location-Scale Model Revisited: A Tale of Robustness
9381	F. de Jong	Specification, Solution and Estimation of a Discrete Time Target Zone Model of EMS Exchange Rates
9401	J.P.C. Kleijnen and R.Y. Rubinstein	Monte Carlo Sampling and Variance Reduction Techniques
9402	F.C. Drost and B.J.M. Werker	Closing the Garch Gap: Continuous Time Garch Modeling
9403	A. Kapteyn	The Measurement of Household Cost Functions: Revealed Preference Versus Subjective Measures
9404	H.G. Bloemen	Job Search, Search Intensity and Labour Market Transitions: An Empirical Exercise
9405	P.W.J. De Bijl	Moral Hazard and Noisy Information Disclosure
9406	A. De Waegenaere	Redistribution of Risk Through Incomplete Markets with Trading Constraints
9407	A. van den Nouweland, P. Borm, W. van Golstein Brouwers, R. Groot Bruinderink, and S. Tijs	A Game Theoretic Approach to Problems in Telecommunication
9408	A.L. Bovenberg and F. van der Ploeg	Consequences of Environmental Tax Reform for Involuntary Unemployment and Welfare

No.	Author(s)	Title
9409	P. Smit	Arnoldi Type Methods for Eigenvalue Calculation: Theory and Experiments
9410	J. Eichberger and D. Kelsey	Non-additive Beliefs and Game Theory
9411	N. Dagan, R. Serrano and O. Volij	A Non-cooperative View of Consistent Bankruptcy Rules
9412	H. Bester and E. Petrakis	Coupons and Oligopolistic Price Discrimination
9413	G. Koop, J. Osiewalski and M.F.J. Steel	Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function
9414	C. Kilby	World Bank-Borrower Relations and Project Supervision
9415	H. Bester	A Bargaining Model of Financial Intermediation
9416	J.J.G. Lemmen and S.C.W. Eijffinger	The Price Approach to Financial Integration: Decomposing European Money Market Interest Rate Differentials
9417	J. de la Horra and C. Fernandez	Sensitivity to Prior Independence via Farlie-Gumbel-Morgenstern Model
9418	D. Talman and Z. Yang	A Simplicial Algorithm for Computing Proper Nash Equilibria of Finite Games
9419	H.J. Bierens	Nonparametric Cointegration Tests
9420	G. van der Laan, D. Talman and Z. Yang	Intersection Theorems on Polytopes
9421	R. van den Brink and R.P. Gilles	Ranking the Nodes in Directed and Weighted Directed Graphs
9422	A. van Soest	Youth Minimum Wage Rates: The Dutch Experience
9423	N. Dagan and O. Volij	Bilateral Comparisons and Consistent Fair Division Rules in the Context of Bankruptcy Problems
9424	R. van den Brink and P. Borm	Digraph Competitions and Cooperative Games
9425	P.H.M. Ruys and R.P. Gilles	The Interdependence between Production and Allocation
9426	T. Callan and A. van Soest	Family Labour Supply and Taxes in Ireland
9427	R.M.W.J. Beetsma and F. van der Ploeg	Macroeconomic Stabilisation and Intervention Policy under an Exchange Rate Band

No.	Author(s)	Title
9428	J.P.C. Kleijnen and W. van Groenendaal	Two-stage versus Sequential Sample-size Determination in Regression Analysis of Simulation Experiments
9429	M. Pradhan and A. van Soest	Household Labour Supply in Urban Areas of a Developing Country
9430	P.J.J. Herings	Endogenously Determined Price Rigidities
9431	H.A. Keuzenkamp and J.R. Magnus	On Tests and Significance in Econometrics
9432	C. Dang, D. Talman and Z. Wang	A Homotopy Approach to the Computation of Economic Equilibria on the Unit Simplex
9433	R. van den Brink	An Axiomatization of the Disjunctive Permission Value for Games with a Permission Structure
9434	C. Veld	Warrant Pricing: A Review of Empirical Research
9435	V. Feltkamp, S. Tijs and S. Muto	Bird's Tree Allocations Revisited
9436	G.-J. Otten, P. Borm, B. Peleg and S. Tijs	The MC-value for Monotonic NTU-Games
9437	S. Hurkens	Learning by Forgetful Players: From Primitive Formations to Persistent Retracts
9438	J.-J. Herings, D. Talman, and Z. Yang	The Computation of a Continuum of Constrained Equilibria
9439	E. Schaling and D. Smyth	The Effects of Inflation on Growth and Fluctuations in Dynamic Macroeconomic Models
9440	J. Arin and V. Feltkamp	The Nucleolus and Kernel of Veto-rich Transferable Utility Games
9441	P.-J. Jost	On the Role of Commitment in a Class of Signalling Problems
9442	J. Bendor, D. Mookherjee, and D. Ray	Aspirations, Adaptive Learning and Cooperation in Repeated Games
9443	G. van der Laan, D. Talman and Zaifu Yang	Modelling Cooperative Games in Permutational Structure
9444	G.J. Almekinders and S.C.W. Eijffinger	Accounting for Daily Bundesbank and Federal Reserve Intervention: A Friction Model with a GARCH Application
9445	A. De Waegenaere	Equilibria in Incomplete Financial Markets with Portfolio Constraints and Transaction Costs
9446	E. Schaling and D. Smyth	The Effects of Inflation on Growth and Fluctuations in Dynamic Macroeconomic Models

No.	Author(s)	Title
9447	G. Koop, J. Osiewalski and M.F.J. Steel	Hospital Efficiency Analysis Through Individual Effects: A Bayesian Approach
9448	H. Hamers, J. Suijs, S. Tijs and P. Borm	The Split Core for Sequencing Games
9449	G.-J. Otten, H. Peters, and O. Volij	Two Characterizations of the Uniform Rule for Division Problems with Single-Peaked Preferences
9450	A.L. Bovenberg and S.A. Smulders	Transitional Impacts of Environmental Policy in an Endogenous Growth Model
9451	F. Verboven	International Price Discrimination in the European Car Market: An Econometric Model of Oligopoly Behavior with Product Differentiation
9452	P.J.-J. Herings	A Globally and Universally Stable Price Adjustment Process
9453	D. Diamantaras, R.P. Gilles and S. Scotchmer	A Note on the Decentralization of Pareto Optima in Economies with Public Projects and Nonessential Private Goods
9454	F. de Jong, T. Nijman and A. Röell	Price Effects of Trading and Components of the Bid-ask Spread on the Paris Bourse
9455	F. Vella and M. Verbeek	Two-Step Estimation of Simultaneous Equation Panel Data Models with Censored Endogenous Variables
9456	H.A. Keuzenkamp and M. McAleer	Simplicity, Scientific Inference and Econometric Modelling

PO BOX 00150 5000 IRLANDS

Bibliotheek K. U. Brabant



17 000 01191536 1

68
66
14