The probability approach in economic methodology

Keuzenkamp, H.A.

Publication date:
1990

CentER
for
Economic Research

Discussion paper

[redacted]

TILBURG UNIVERSITY

No. 9044

THE PROBABILITY APPROACH IN ECONOMIC METHODOLOGY: ON THE RELATION BETWEEN HAAVELMO'S LEGACY AND THE METHODOLOGY OF ECONOMICS

by Hugo A. Keuzenmkamp

August 1990

ISSN 0924-7815

THE PROBABILITY APPROACH IN ECONOMIC METHODOLOGY

ON THE RELATION BETWEEN HAAVELMO'S LEGACY AND THE METHODOLOGY OF ECONOMICS

Hugo A. Keuzenkamp[1]

CentER for Economic Research
Tilburg University
Box 90153
5000 LE Tilburg, The Netherlands

December 1989

Revised: June 1990

[1] I would like to thank Ton Barten, Philippe Deschamps, Arie Kapteyn, Michael McAleer and Mark Steel for helpful comments. The usual disclaimer applies.

Abstract

In his masterful 1944 manifesto, Haavelmo tried to convince econometricians of the idea that methodology and appraisal of (economic) theories are a fundamentally probabilistic matter. He helped to provide foundations for identifying, estimating and testing simultaneous economic relations. Haavelmo and his colleagues at the Cowles Commission were very successful, but *one* issue turned out to be overwhelmingly more difficult to solve than expected: the choice of the "best" economic theory or econometric model. This is of interest for econometricians as well as for those involved in the methodology of economics. This paper is primarily addressed to the latter group. The statistical and econometric literature on testing non-nested hypotheses proves to be of interest for economic methodology. Hence, the *probability approach in economic methodology* is a natural successor to Haavelmo's *probability approach in econometrics*. Obvious as this may sound to those familiar with probabilistic inference, this is a fairly new message for economic methodologists. For bad reasons, philosophy of science and economic methodology treated statistical inference as not their cup of tea. This led to an artificial distinction between probabilistic and non-probabilistic inference.

Contents

1. Introduction

Despite his aversion for publicity and fame, the 1989 Nobel Memorial prize in Economics was awarded to the econometrician Trygve Haavelmo. This honours his contributions to econometrics, especially his supplement to the 1944 edition of Econometrica, *The Probability Approach in Econometrics*. Haavelmo provided the first thorough analysis of how probability methods could and should be used in economics. His work extended Tinbergen's efforts to specify, estimate and test simultaneous models in economics.[2] However, we will not evaluate here his work on the statistics of systems of equations, but instead focus on another issue: the use of probability theory in economic inference.

Haavelmo emphasized the need for proper probability theory in econometrics. At the time, the most popular and most developed probability theory was the frequentist methodology. The methods of Neyman, Pearson and Fisher, the leading statisticians of the day, were extended to the econometrics domain, and many successful results emerged. Hence, this resulted in a legacy that emphasized frequentist probability techniques for empirical research.

Model identification was among the most important issues addressed by Haavelmo (other contributors were Working, Frisch, Tinbergen and Marschak, to mention just a few). The problem was how empirical observations could be used to trace structural equations, such as a demand function, where demand depends on prices. As supply depends on prices as well, and only equilibrium points (where demand equals supply) are observed, it was not immediately clear how empirical data related to economic theory, i.e. how data on prices and quantities could identify theoretical relations such as the demand function. The well known rank and order conditions provided a solution to this problem. Prior

---

[2] Of course, Haavelmo was not alone in his efforts. The members of the Cowles Commission in particular were actively involved in this *research programme*. Their activities were comparable with a political struggle, Haavelmo provided the party manifesto. The developments that led to Haavelmo's manifesto are discussed in Morgan (1990).

restrictions were needed for identification. Initially, this was not really a *probabilistic* issue, as it was assumed that these restrictions were given by Economic Theory, with absolute certainty. As soon as this assumption became suspect the probability approach in econometrics lost some of its appeal. Haavelmo's foundations became questionable.

The early hopes of the famous Cowles Commission (of which Haavelmo was a member from 1943 to 1947) were very optimistic, but quickly undermined. The immediate goal was to come with definite models of the economy that could be used to guide policy. Without much hesitation, the director of the Cowles Commission could write his research agenda for the near future[3]:

1945-6: work on method(ology) to be completed in the main;
1946-8: final application of method to business cycle hypotheses and to (detailed) single market problems;
1948-9: discussion of policy. Extension to international economics;

(J. Marschak, 1944 memorandum to Social Science Research Committee, quoted in Epstein 1987, p. 62). A similar agenda would work for the development of the atomic bomb, but statistical evaluation of economic theories turned out to be not that easy, not to speak of social engineering[4].

Controversies on the appropriate inferential method, the statistical model and economic theory continued. The fundamental problem of theory appraisal was never really solved. It became evident that there was more uncertainty then initially expected: the problem was not just to estimate a given model, but also to choose the model to start with. Haavelmo delegated this problem to the economic theorists (1944, p. 71), and invoked what Leamer (1978) dubbed the "axiom of correct specification" to

---

[3] This research agenda corresponds to what Lakatos (1970) calls the "positive heuristic" of a research programme. See below.

[4] It is fair to say that the amount of money spent on the making of the atomic bomb is incomparable to the money spent on the development of econometrics: Rhodes (1986) provides an astonishing account of the magnitude and economic costs of this piece of physical engineering.

justify further probabilistic inference[5].

Despite such shortcomings, there is little doubt that Haavelmo and his Cowles colleagues convinced most of the econometric profession with the claim that proper probability foundations were a necessary condition for good inference. On the other hand, those who studied methodology of economics (from a philosophy of science perspective) were not overly impressed and neglected most of econometrics and statistics. One reason might be that the original research agenda had little to say about model appraisal, the problem of methodology. Still, this is a weak excuse, certainly today but also during the early years of econometrics.

Probabilistic inference is so obviously a candidate for appraisal of theories that one needs good reasons to neglect it. It is doubtful that such reasons exist. And *if* reasons can be found, they should be given[6]. Furthermore, one would like to see alternatives. In a recent discussion on such alternatives for dealing with uncertainty (i.e. fuzzy logic and artificial intelligence), Lindley (1987) made the same assertion as Haavelmo did: the only satisfactory description of uncertainty is provided by probability theory, as no better methods are available.[7] Hence, if there is uncertainty about theories, it is best expressed in probabilistic terms. The specification and appraisal of economic models can be interpreted in this light.

This paper evaluates the scope and limits of probability methods in economic methodology. What may someone who is interested in economic

---

[5] This "axiom" was first put forward by R.A. Fisher; Koopmans referred to it as "Fisher's axiom of correct specification" in 1937 (see Spanos (1989, p. 411)).

[6] Of course, theories are not *only* evaluated from the point of view of empirical evidence. Conceptual issues are important as well. Still, a theory that has no empirical support whatsoever is not a very attractive one.

[7] However, Shafer (1987) disagrees by giving more credit to the artificial intelligence program. As far as I am aware, there are no applications of fuzzy logic, and hardly any applications of artificial intelligence, to empirical economics. See also the paper by Swamy et.al. (1985), who try to reconcile fuzzy logic with Bayesian probability theory.

methodology learn from the econometric literature on testing economic theories? More than seems to be suggested by most of the methodological literature. Furthermore, the problems of econometric inference did already pop up in the early days of the subject, hence it is interesting to see how they were treated then, and now. This part of Haavelmo's legacy is worth investigating in detail.

The paper is organized as follows. Section 2 of this paper starts with a brief overview of some ideas in philosophy of science. Section 3 starts with a discussion of probabilistic foundations of economic inference. In section 4 we discuss the problem of observational equivalence of rival models. Section 5 deals with statistical inference, and the difference between the frequentist and the Bayesian approach in probability theory. Section 6 deals with testing, comparing and appraising rival hypotheses. Section 7 concludes the paper.

2. A Short Detour on Philosophy of Science

Science involves formulating and appraising hypotheses, theories (sets of hypotheses), or research programmes ("theories in development", see Lakatos 1970 and below). Ideally, one would like to obtain a definite solution for scientific problems, but usually there are several competing theoretical explanations. In that case, we must try to find the best or most useful explanation available. Therefore, we need methods to evaluate theories.

Philosophy of science, the discipline that takes a closer look at such appraisal, was dominated for some time by the ideas of Popper (1959). He emphasized the notion of *crucial tests* (tests leading to complete rejection) and *logical falsifiability* in evaluating theories. Popper's philosophy is problematic for many reasons (see e.g. Hausman 1989), but it is evident that crucial tests in economics are rare or do not exist at all. Furthermore, the essence of probabilistic theories is that they regard events as *improbable*, not *impossible* (events with probability zero). Logical falsifiability relates to the latter case but is not of much interest for the former. Hence probabilistic inference is better suited for scientific appraisal than non-probabilistic inference.

Popper's student, Lakatos, tried to overcome some of the problems of Popperian methodology. He dropped the crucial test and instead emphasized the *dynamics* of theory development. Hence, Lakatos (1970) substituted *research programmes* for Popper's theories. A theory, $\mathcal{T}$, is just one instance of a research programme, $\mathcal{RP}$, at a given point in time. A falsification of $\mathcal{T}$ is not automatically a rejection of $\mathcal{RP}$. Falsifying a theory was replaced by measuring the degree of "progressiveness" of an $\mathcal{RP}$. A programme is theoretically progressive if succeeding theories predict novel facts (facts unforeseen by earlier theories). The programme is empirically progressive if these predictions are realized.

But Lakatos's suggestion of appraising theories by comparing their rate of progressiveness or degeneration is not easily applied to economics. A research programme is a rather vague notion. Scientists may disagree about what belongs to a specific $\mathcal{RP}$ and what does not (see Feyerabend, 1975). For example, if we replace the theory of diminishing marginal utility of money by a successor, which combines this theory with the general theory of relativity, we seem to make a "progressive" step. Of course, this is not what Lakatos had in mind: in order to avoid such nonsense-progression he introduced the *hard core*. The hard core defines the general principles which are indisputable, at least among the adherents of a particular research programme. Lakatos's hard core is of some help, but is insufficient to define what exactly belongs to the theory. An alternative to avoid the above-mentioned nonsensical conjunction of two theories is to introduce the notion of *irrelevant conjunction* (for a discussion, with a Bayesian solution to the problem, see Rosenkrantz (1983)).

It is difficult to define the "hard core" of a programme without arbitrariness. Competing programmes may apply to partly non-overlapping areas of interest. This leads to well known problems as *incommensurability* (Kuhn's (1962) notion that new theories give new interpretations of events, and even to the language describing these events), and to the *Duhem-Quine* problem (one is never sure just what is being tested, the hypothesis of interest or an auxiliary hypothesis). Furthermore, whereas Popper still made an (unsuccessful) attempt to contribute to the theory of probabilistic inference (cf. his *propensity theory of probability*), Lakatos has a rather condescending attitude towards this subject. Both of them reject inductive reasoning.

The philosopher Nancy Cartwright (1983) has made clear that the relation between pure theory, a model and measured variables is not at all as well defined as one would like to see. She emphasizes that explanatory theories (*fundamental laws*) in physics are always false, strictly spoken (compare Lakatos's dictum: "all theories are born refuted"). Instead, *phenomenological laws* are empirically informative. There is a trade-off between explanatory power and empirical validity (Cartwright 1983, p. 3):

> We have detailed expertise for testing the claim of physics about what happens in concrete situations. When we look to the real implications of our fundamental laws, they do not meet these ordinary standards. (...) When it comes to the test, fundamental laws are far worse off than the phenomenological laws they are supposed to explain."

Speaking of *laws* in economics may be misleading, therefore we call them *models*, $\mathcal{M}$.[8] It is clear that $\mathcal{M}$ relates to $\mathcal{RP}$, but this relation is not one of logical, strictly deductive, entailment.

Take, for example, Ian Hacking's short history of the discovery and interpretation of Faraday's magneto-optical effect. In the 19th century, G.B. Airy:

> "showed how to represent it [the Faraday effect] analytically within the wave theory of light. The equations for light had contained some second derivatives of displacement with respect to time. Airy assessed some *ad hoc* further terms, either first or third derivatives. *This is a standard move in physics. In order to make the equations fit the phenomena, you pull from the shelf some fairly standard extra terms for the equations, without knowing why one rather than another will do the trick.*" (Hacking 1983 p. 211, emphasis added)[9].

---

8 Cartwright (1983) distinguishes pure theories, models and phenomenological laws because her main interest is to criticize scientific realism. Her point is that "phenomenological laws are indeed true of the objects in reality -or might be; but the fundamental laws are true only of objects in the model." (op.cit. p.4). We are less interested in the problem of scientific realism and more in the problem of testing or comparing economic theories. These comparisons take place at the level of empirical models which are related to developing theories. Lakatos's term *research programme* serves this goal better than fundamental laws do, and we will use the word *model* in the sense of Cartwright's phenomenological law. The dynamics of statistical inference falls under the heading of *modelling*.

9 Airy could not *explain* the Faraday effect, he only could give a good

Hence, theories in physics are often enriched with additional elements, not contained in the pure theory. That sounds familiar to the economist. But if pure theories are wrong anyway, if we are *a priori* certain they are, then does it make sense to use probability methods to evaluate them? The answer is yes. Even if we *know* that the pure theory is false, strictly spoken, we would like to know how well the theory represents reality[10].

The Popper-Lakatos tradition (still alive at LSE in the person of John Watkins, and in economics defended by Mark Blaug and Joop Klant) has lost ground to alternative philosophies of science. Some of them are hardly worth the label "philosophy" (e.g. Caldwell 1982, who does not pretend to provide a true alternative philosophy but may be interpreted as a warning against following dogmatically a one-sided philosophy such as Popper's), but others are true attempts to solve the deep methodological problems. Particularly since the seventies, probabilistic reasoning is in the philosophical picture. Howson and Urbach (1989) provide a comprehensive treatment of a probabilistic philosophy of science (from a Bayesian perspective). In the following sections, we will discuss the validity of probabilistic inference in economics.

3. Economic Inference and Probability

Haavelmo's argument to defend the use of probabilistic inference closely resembles Hacking's point or Cartwright's emphasis on the importance of phenomenological laws. The following quotation (Haavelmo 1944, p. iv) makes this clear.

> "In fact, if we consider actual economic research -even that carried out by people who oppose the use of probability schemes- we find that

---

description of what happened. It was Lorentz who provided a theoretical explanation in terms of electron theory. Cartwright (1983, p.2) calls Airy's analysis a phenomenological law, and Lorentz's *explanation* a theoretical law.

[10] Compare Box (1979): "All models are wrong but some are useful".

> it rests, ultimately, upon some, perhaps very vague, notion of probability and random variables. For whenever we apply a theory to facts we do not -and we do not expect to- obtain exact agreement. Certain discrepancies are classified as "admissible", others as "practically impossible" under the assumptions of the theory. And the *principle* of such classification is itself a theoretical scheme, namely one in which the vague expressions "practically impossible" or "almost certain" are replaced by "the probability is near to zero", or "the probability is near to one".
> This is nothing but a convenient way of expressing opinions about real phenomena. But the probability concept has the advantage that it is "analytic", we can derive new statements from it by the rules of logic. Thus, starting from a purely formal probability model involving certain probabilities which themselves may not have any counterparts in real life, we may derive such statements as "The probability of A is almost equal to 1". Substituting some real phenomena for A, and transforming the statement "a probability near to 1" into "we are almost sure that A will occur", we have a statement about a real phenomenon, the truth of which can be tested."

Let us now see exactly how probability enters economics. First of all, human behaviour, the underlying determinant of economic data, is to a degree erratic. This is the unexplained part of behaviour which can be modelled in probabilistic terms. This randomness may be due to human nature or, for example, to the fact that behaviour of rational individuals depends partly on their expectations of future states of the world.

Secondly, the variables that are part of the theoretical model have to be matched with observational counterparts. This match may be imperfect, with "noise" entering due to the fact that only approximate variables can be observed. Measurement error was Haavelmo's main reason for invoking probability theory. A related issue is the seperation of object and subject which hampers, for example, empirical research to the permanent income hypothesis. The econometrician has to identify permanent shocks in individual incomes from observed data, whereas individuals often know already before the actual shock occurs that it is imminent. The result may be that the econometrician falsely rejects the theory (see Campbell 1987 for an excellent treatment of this problem).

Thirdly, the theoretical form of the model may leave room for many interpretations. In theoretical models, functional forms are often left unspecified, and constraints such as concavity and stability conditions may be all that are available. The empirical model will always be mis-specified, to some degree. For example "economic nature" may be a highly nonlinear system of dynamic equations, whereas the model is a linear simplification of this system. Haavelmo (1944, p. 71)) left this specification problem to economic theory: "Let [*f*] be our economic theory to be tested, the random variables having certain prescribed distribution properties. The principal task of economic theory is to make a fruitful choice of the *forms f*." Modern developments in econometrics brought this problem within the realm of econometrics (two different approaches are mis-specification testing (e.g. Hausman (1978)) or specification search (Leamer 1978))[11].

Fourthly, the theoretical model may simply be wrong. Not only, as noted before, because theoretical models, even in physics, are always wrong. Also, from a purely logical point of view, if there are two rival explanations for some empirical occurrences, which are partly or fully exclusive to each other, then not both of them can be "true" at the same time.

Hence we have four sources of uncertainty: a probabilistic nature, measurement error, specification error and theory uncertainty. Given these sources of uncertainty, probabilistic methods may prove invaluable to give the theories empirical relevance, to quantify their match with the data. In section 4 below, two different probabilistic methodologies are discussed: the *frequentist* and the *Bayesian* one. The underlying philosophies are quite different. Haavelmo adopted the frequentist approach to probabilistic inference, but we will see that there are some

---

[11] Spanos (1989) argues that Haavelmo is a forerunner of the "Spanos-approach to econometrics", in which the empirical economic model and statistical model are distinct. This leads to the mystical claim that "At the statistical-model level, testing the underlying assumptions constitutes proper misspecification testing, but at the empirical-model level, the same tests can only be described as "diagnostic checking"." (Spanos 1989, p. 423). This kind of schizophrenia is absent in Haavelmo (1944). Haavelmo's probability foundation does not relate to testing the validity of the statistical model, but relates primarily to the probabilistic analysis of a simultaneous model, i.e. the joint p.d.f.. It also is prepostorous to see in Haavelmo an early LSE-style econometrician, i.e. one in the tradition of Sargan and Hendry. Two key-stones of this tradition are extensive testing, and specifying sound short- and long term dynamics of a model. There is nothing whatsoever in Haavelmo's work to validate Spanos's suggestion that the Nobel-laureate may be regarded as a honourable member of this school in econometrics.

implicitly Bayesian ideas in his work as well[12]. Of course, this doesn't make him a Bayesian hero.

It is interesting to see how Haavelmo formulated his programme (1944, pp. 8-9):

> "That is, the model will have an economic meaning only when associated with a design of actual experiments that describes -and indicates how to measure- a system of "true" variables (or objects) x_1, x_2, ..., x_n that are to be identified with the corresponding variables in the theory. (...) The model thereby becomes *an a priori hypothesis* about real phenomena, stating that every system of values that we might observe of the "true" variables will be one that belongs to the set of value-systems that is admissible within the model. (...) Hypotheses in the above sense are thus the joint implications -and the only testable implications, as far as *observations* are concerned- of a theory *and* a design of experiments. It is then natural to adopt the convention that a theory is called true or false according as the hypotheses implied are true or false, when tested against the data chosen as the "true" variables. Then we may speak, interchangeably, about testing hypotheses or testing theories."

Haavelmo's methodology seems thus a fairly straightforward implementation of the *modus tollens* (if the model is rejected, then the underlying theory is rejected):

$$\mathcal{M}(\mathcal{T},\mathcal{D}):\ \mathcal{M} = \{\mathcal{H}\}\ ,\ \sim\mathcal{H} \xrightarrow{p} \sim(\mathcal{T},\mathcal{D}) \qquad \textbf{(3.1)}$$

where $(\mathcal{T},\mathcal{D})$, the theory and the experimental design, defines the model, and $\{\mathcal{H}\}$ denotes the set of a priori hypotheses implied by the model. The $\sim$ sign is the logical operator "not true". The sign $\xrightarrow{p}$ is used to denote a relation for probabilistic inference to distinguish it from a purely logical (deductive) relation (of course it does not indicate convergence in probability).

Above it was argued that the relation between theory and empirical model is more diffuse than 3.1 suggests[13]. It is better to replace $\mathcal{T}$, the

---

[12] Not unlike Fisher's interpretation of probability. Fisher accepted probability as a measure of evidence, whereas Neyman rejected such an interpretation.

[13] This point is not entirely neglected by Haavelmo, note *The Meaning of*

theory in 3.1, by a research programme $\mathcal{RP}$.[14] The statistical rejection of a hypothesis may inspire an adjustment in the $\mathcal{RP}$ and not to the rejection of a theory (see Barten and Keuzenkamp, forthcoming). Statistical inference is a process of learning, not a once and for all decision. At a given point in time, this inference results in a favourite model $\mathcal{M}(t)$, which is the one that survives the feedback process of selection and variation. Of course, there may be more of such models. If two competing models emerge, then the first issue that arises is to see if data can discriminate between the models. This point is related to the identification problem, extensively discussed in Haavelmo (1944).

This identification problem, at least as it was discussed initially, points to a very specific characteristic of the occurrence of probability in economics: simultaneity. In economic relations, everything depends on everything. A demand equation has prices as right hand side variables, but these prices depend again on the quantities demanded and supplied. This basic problem, inherent to economics, makes probabilistic inference in economics more difficult than straightforward philosophical discussions of probabilistic inference would suggest. Hence we will deal with this problem in some more detail in the next section.

4. Identification and Observational Equivalence

The identification problem belongs to the most interesting topics in econometrics. Haavelmo made significant contributions to the solution of the problem: he discussed the order and rank condition in a very general way. Closely related to identification is observational equivalence. For example, if a demand function and a supply function are not identified, then estimation of one of them is observationally equivalent to

---

the Phrase 'To Formulate Theories by Looking at the Data', section 17 of Haavelmo (1944). See also below.

14 The term research programme remains a vague one. If an $\mathcal{RP}$ has any empirical meaning at all, than we may interpret it as the *model space*. Below, we make an effort to provide a Bayesian account of an $\mathcal{RP}$.

estimating the other. The same can occur when testing rival models, say $\mathcal{M}_1$ and $\mathcal{M}_2$. Either model provides a set of variables and functional relations in the context of a stochastic scheme to explain the dependent variables, $\mathbf{y}$. Thus, we have two densities, $p(\mathbf{y}|\mathcal{M}_1)$ and $p(\mathbf{y}|\mathcal{M}_2)$. We define observational equivalence as:

$$p(\mathbf{y}|\mathcal{M}_1) = p(\mathbf{y}|\mathcal{M}_2) \tag{3.1}$$

which says that the data look identical, whether the first or the second "window" is used. The data cannot discriminate between the models. One reason may be the fact that the *structure* of the models differs, but the models as they are to be estimated are the same. The reduced forms of the models are then equivalent. The problem was recognized in the early days of econometrics, with Haavelmo discussing it at length in his (1944) essay. He presented the order and rank conditions for identification, but his interest was not in comparing models but in showing under what conditions one would be allowed to give structural interpretations to parameters. For example, if prices and quantities are correlated, it is not necessary that this relation represents a *demand* function: if identification conditions are not fulfilled, then the estimated function could as well be a *supply* function. This problem was discussed by Working in the 1920s.

An interesting example of observational equivalence of rival models is presented in Sargent (1976), who shows an implication of Lucas's (1976) critique on econometric policy evaluation. The idea is this. There may exist two competing $\mathcal{RP}$s, $\mathcal{RP}_1$ and $\mathcal{RP}_2$, which have very different characteristics but yet are indistinguishable if they are translated to empirical models. The estimated parameters do not reveal whether they identify the first or second theory. In Sargent's example, one theory has Keynesian features (discretionary monetary feedback rules can be used to reduce unemployment), and the other has monetarist features (money cannot have long term real effects). Assume that, according to model one, changes in last year's money supply changes current unemployment, whereas the other model argues that only *unanticipated* changes matter. The underlying theories are very different but, in the absence of other information, a model that estimates unemployment conditional on past money does not say anything at all about the validity of one or the other model. In the extreme case that individuals have perfect foresight, that is, they know any change in the policy rule, it is impossible to infer with statistical methods which model is "true". In other cases, additional prior information, such as the dynamics of the economy or the distribution of unanticipated shocks, is needed in order to identify the parameters of a model. The use of such prior information, and anyway the question "Where does the model come from", has led to two different approaches in the theory of probability: the frequentist and the Bayesian. These, and their relevance for a probability approach in the methodology of economics, are discussed in the next section.

5. Statistical Inference: the Frequentist and Bayesian Approach in Probability Theory

In section 3 we saw that Haavelmo delegated the choice of fruitful models to test to economic theory. Hence, according to his methodology, the model is given to the statistician. This is also the starting point for the frequentist approach to statistical inference, at least this was the situation around 1940.

The frequentist theory relates a model directly to the outside world (i.e. the data) via an "objective" mechanism. Furthermore, this theory is founded upon the occurrence of repetitive events, a class to which theory appraisal does not belong. Von Mises, one of the founders of frequency theory, claims that "...if one talks of the probability that the two poems known as the *Iliad* and the *Odyssey* have the same author, no references to a prolonged sequence of cases is possible and it hardly makes sense to assign a *numerical* value to such a conjecture." (cited in Barnett 1973, p. 8). Similarly, theory appraisal has often been excluded by adherents of the frequentist approach. A theory is right or it is false, if it is false then it is not a valid base for probabilistic inference.

Frequentist theory assumes that there is no uncertainty with respect to the parameters of a model: the "true" parameters are fixed quantities, the only problem is that we cannot observe them directly.[15] To make inferential statements, the frequentist needs to assert beforehand that one model is "true" (Leamer's axiom of correct specification). That is a remarkable assumption if two competing theories, which are *a priori* equally valuable, are evaluated. Either, the frequentist who wants to evaluate such theories has to take the foundations of frequentist probability theory with a grain of salt, or inference is not necessary as it is already known that one or the other model is true.[16]

The Bayesian perspective is different. Repetitive events and truth are not the starting points for inference, but rather degrees of belief or betting ratios. For a Bayesian, the model is a useful window from which to view the world, and probability is the measure of uncertainty. If new data become available, they are used to revise the uncertainty. Hence, the Bayesian perspective emphasizes consistent sequential learning. The Bayesian is not interested in the properties of a hypothetical infinite number of random drawings from some population, but only in the data-information as it is available. Prior information (or degrees of belief) are combined with new information to obtain posterior beliefs. All statements are conditional on this (prior and data) information.

Bayesian methods have often been criticized, particularly because of their dependence on prior probabilities. In the world of frequentist probability, such prior information should have the form of known frequencies. In that case, there is no reason to distinguish the "prior" from sample information. In other cases, a frequentist will reject any

---

[15] Mises (1957, p. 158) disagrees: "I do not understand the many beautiful words used by Fisher and his followers in support of the likelihood theory. The main argument, namely, that p is not a variable but an "unknown constant", does not mean anything to me." Mises does not provide an alternative inspiration for economists in their practical statistical work insofar they work with small samples: Mises vehemently criticizes small sample theory. And, as we just saw, using probability for epistemological inference is also unwarranted in the eyes of Mises.

Furthermore, it must be said that speaking of *the* frequentist approach can be dangerous: notable differences between frequentist interpretations (such as Neyman's, Fisher's or Mises's) exist (see Jeffreys 1961).

[16] In section 6 below, the issue of non-nested testing is discussed in detail. Some recent frequentist approaches to this problem drop the assumption that one of the models should be regarded as "true", but at significant costs (loss of power).

use of prior information. For example, probabilistic prior ignorance is a contradiction in terms, if the terms are dictated by the frequentist. A Bayesian, who thinks in terms of beliefs not frequencies, does not have such hesitation, particularly after the formalization provided by Jeffreys (1961). It is true, though, that Bayesian "ignorance" has to obey some very strict rules. The Bayesian rejoinder to the frequentist school is that you cannot do without these prior probabilities if you want to answer basic inferential questions.

Most of modern econometrics stems from the frequentist tradition. The statistical theory of point estimation, estimation of confidence intervals and "hypothesis testing", developed among others by Fisher, Neyman and Pearson, were introduced into econometrics by Haavelmo in 1944. Of course, some of these statistical methods had already been applied by other econometricians, but Haavelmo provided probabilistic foundations for their use in interdependent systems of equations. First, he had to justify the use of a mathematical theory of repetitive events, for it was not immediately clear how unique economic data could be interpreted as such. Haavelmo (1944, p. 51) invented a clever way out:

> "There is no logical difficulty involved in considering the "whole population as a sample", for the class of populations we are dealing with does *not* consists of an infinity of different individuals, it consists of an infinity of possible *decisions* which might be taken with respect to the value of *y*. And all the decisions taken by all the individuals who were present during one year, say, may be considered as one sample, all the decisions taken by, perhaps, the *same* individuals during another year may be considered as *another* sample, and so forth. From this point of view we may consider the total number of possible observations (the total number of decisions to consume *A* by all individuals) as result of a sampling procedure, which *Nature* is carrying out, and which we merely watch as passive observers."

Haavelmo needed this metaphysical justification because at the time there were no real alternative methods of inference. The Bayesian approach was highly suspect. The leading statisticians of the prewar era rejected the Bayesian methodology (although Neyman became more positive towards the Bayesian approach some twenty years later, and Fisher developed a *quasi* Bayesian method called *fiducial inference* (see Barnett 1973)). The first edition of Jeffreys's Bayesian *Theory of Probability* appeared in 1939 and it took a considerable amount of time before this book was recognized as a significant contribution to the theory of probability. Haavelmo hardly could and certainly did not make use of Jeffreys's work. Still, he gives support to what may be regarded as an implicitly Bayesian interpretation of statistics (1944, p. 48)[17]:

> "we considered "frequency of occurrence" as a practical counterpart to probability. But in many cases such an interpretation would seem rather artificial, e.g., for economic time series where a repetition of the "experiment", in the usual sense, is not possible or feasible. Here we might then, alternatively, interpret "probability" simply as a measure of our *a priori confidence* in the occurrence of a certain event."

In the present paper we want to see how this relates to "confidence in competing hypotheses or models", an issue not directly addressed by Haavelmo but a logical complement to his work.

The question for the next section is how two rival models, M_1 and M_2, can be compared. First, we have to find out how to obtain them. This is the domain of statistical modelling, an issue large and difficult enough for a separate PhD-thesis[18].

A problem closely related to statistical modelling is the validity of statistical tests of hypotheses if a theory is false anyway. In the background of hypothesis testing lingers a *maintained hypothesis*: that part of the theory that remains unquestioned. For example, this may be assumption of valid conditioning, the axiom of correct specification. Haavelmo relies on such an assumption (1944, p. 74, see also p. 71): "This means that we are sure -or that we accept without test- that the theory is all right so far as the forms of the functions f are concerned." It is the task of economic theory to make fruitful choices of the forms f.

The statistical methods of modelling and hypothesis testing lead to a problem that occurs if there is only one sample from which to infer both the form of the model and the estimates of the parameters. This is known

---

17 This comes fairly close to Fisher's interpretation of probability: a measure of evidence. There is a strong difference between Fisher on the one hand and Neyman and Pearson on the other hand with regard to this interpretation of probability. The Neyman-Pearson approach deals with the risk of taking decisions, not with truly inferential problems.

18 Work in progress.

as the data mining problem, but it was already known to the early econometricians. As noted before, Haavelmo spent a section on it. A quote may be illuminating:

> "In general, whenever we can establish that certain data satisfy certain relationships, we add something to our knowledge, namely a restriction of the class of a priori admissible hypotheses. *The real difficulty* lies in deciding whether or not a given relation is actually compatible with the data: and the important thing to be analyzed is the reliability of the test by which the decision is made, since we have to deal with stochastic relations and random variables, not exact relations.
> From this point of view there is, therefore, no justified objection against trying out various theories to find one which "fits the data". But objections may be made against certain *methods of testing the fit*." (1944, p. 83)

Haavelmo refers to the case where the set of *a priori* admissible hypotheses, in his terminology Ω^0, is not fixed but changes during the specification search: "What is *not* permissible is to let Ω^0 be a *function of the sample point.*" (op.cit.). Note that this is similar to the Bayesian idea of *coherence*, due to De Finetti (see Hill 1974). It also supports Hendry's preference for testing from general to specific, where Ω^0 is the most general hypothesis one can think of. A problem for strictly applying frequentist methods remains: the sampling distributions of the test statistics are usually not known if sequential tests are performed.

Data mining violates strict Bayesian principles as well. Coherent behaviour assumes the existence of a complete prior over all possible models, whereas in reality some models suggest themselves only after inspection of the data. Hence, a Bayesian researcher presents results conditional on the model(s) under scrutiny, and has to entertain non-dogmatic priors if prior omniscience is unavailable (see also Hill 1986). There exists no theory of scientific creativity: both Bayesian and Frequentist theories fail in this respect, as does any other philosophy.

6. Testing, Comparing and Appraising Rival Models

Whether Bayesian or frequentist tools were used in the inferential stage, the researcher will try to end up with a favoured hypothesis, which can be compared with a rival hypothesis. We can use probability methods to evaluate these hypotheses. But such evaluation is remote from questions like "Is monetarism superior to Keynesianism?". Statistical tools can be used to compare well defined statistical models, further inference to less well defined research programmes is only possible if researchers agree on the terms that define the relations between an $\mathcal{RP}$ and the empirical model. Empirical support is not the only criterion for appraising theories, but it is an important one. If both competing parties would agree that comparing $\mathcal{M}_1$ and $\mathcal{M}_2$ will give a definite answer to the question posed above, we may obtain a direct test. Not yet a crucial one: this would only be a limiting case for the event that all probability mass would go to just one of the two hypotheses. But the test depends on the agreement between individuals about the relevance of $\mathcal{M}_1$ for $\mathcal{RP}_1$, this agreement is a matter of taste and judgment. Hence, theory appraisal is a subjective matter, but agreement on the terms for appraisal is possible. It primarily depends on agreement about the validity of statistical models.

The Friedman-Meiselman effort to test monetarism versus Keynesianism was not accepted by Keynesians, because they disagreed that the Keynesian theory was represented by a sound Keynesian hypothesis. Note, by the way, how Friedman and Meiselman characterized advanced statistical methods for comparing theories:

> "The evidence is so one-sided that its import is clear without the nice balancing of conflicting bits of evidence, the sophisticated examination of statistical tests of significance, and the introduction of supplementary information that the economic statistician repeatedly finds necessary in trying to decide questionable points, and that is indeed a major source of pride and pleasure in his craft" (Friedman and Meiselman 1963, p. 186).

Not everyone agreed that the evidence was so one-sided. And it was the craft of economists using sophisticated statistical tools that showed Friedman and Meiselman's approach was founded on a misspecified model. The discussion turned into a debate on statistical testing of competing models.

As will be clear from the previous section, there are two general approaches to the problem comparing or testing rival models: the frequentist and the Bayesian. It is not the purpose of this paper to present a rigourous analysis of these two approaches (see e.g. Barnett 1973). Instead we concentrate on the basic ideas in order to be able to make clear what this statistical literature may have to do with the methodology of economics.

In the frequentist approach there are basically three different strategies for testing rival models: the embedding or comprehensive model strategy, the generalized likelihood ratio strategy and the symmetric (or equivalence) strategy. The comprehensive model is a nesting of the competing models in a more general, embedding model. More specifically, let us assume that, after some experimentation and hypotheses testing, two rival models resulted: $\mathcal{M}_1$ and $\mathcal{M}_2$. Assume both are linear models given by:

$$\mathcal{M}_1: \quad Y = X\beta_1 + \varepsilon_1 , \qquad \varepsilon_1 \sim N(0,\sigma_1^2 I) \tag{5.1}$$

$$\mathcal{M}_2: \quad Y = Z\beta_2 + \varepsilon_2 , \qquad \varepsilon_2 \sim N(0,\sigma_2^2 I) \tag{5.2}$$

where **X** and **Z** may or may not have some (not complete) overlap. Testing these non-nested hypotheses can be done by regarding them as two restricted forms of a more general model, the embedding model:

$$\mathcal{M}^*: \quad Y = X\beta_1^* + Z\beta_2^* + \varepsilon^* \tag{5.3}$$

The test now is to see if either $\mathbf{b}_1^*$ or $\mathbf{b}_2^*$ are significantly different from zero (using an *F*-test, for example). The test leads to some difficulties, however. First, we may find that either $\mathbf{b}_1^*$ or $\mathbf{b}_2^*$ is significantly different from zero which seems to be an informative result. It may happen, however, that a *joint* test on both $\mathbf{b}_1^*$ and $\mathbf{b}_2^*$ leads to a contradictory result. Secondly, it is not clear what should be done if both or none of the restrictions are rejected (if none of the restrictions is rejected, then we are left with an embedding model in which we are not *per se* interested). Thirdly, collinearity between (subsets of) the rival regressors may decrease the power of the test (Pesaran (1982) discusses different aspects of the comprehensive testing approach).

An alternative to the comprehensive model approach is due to Cox (1961, 1962), who also inspired Atkinson and MacKinnon to develop extensions of the previous approach. Cox defines a test statistic:

$$T_1 = L_{1,2} - \mathbb{E}_{\mathbf{b}_1}\{L_{1,2}\} \tag{5.4}$$

where $L_{1,2}$ is the log-likelihood ratio, $\mathbb{E}_{\mathbf{b}_1}\{L_{1,2}\}$ is the expected value (if $\mathcal{M}_1$ were true) of the difference between the sample log-likelihood ratio given $\mathcal{M}_1$ and the sample log-likelihood ratio given $\mathcal{M}_2$ (the likelihoods are evaluated at the maximum likelihood estimates). Under the null, T_1 is asymptotically normal distributed. Note that the procedure is asymmetric: in the case given above, the distribution of the parameters of $\mathcal{M}_2$ is evaluated conditionally on the truth of $\mathcal{M}_1$ ($\mathcal{M}_1$ is called the *reference hypothesis*). Hence, the inferential results may depend on whether $\mathcal{M}_1$ or $\mathcal{M}_2$ is made the reference hypothesis. The motivation for Cox's test was partly a dissatisfaction with the use of improper priors in a Bayesian framework for model comparison (see below for a discussion).

The Cox-test showed up in econometrics around 1970 (see e.g. Gaver and Geisel 1974, who refer to applications by Dhrymes *et.al.* and Pesaran). Hendry, Mizon and Richard use it for their theory of *encompassing* (see e.g. Mizon and Richard 1986). Encompassing is both used to compare rival models, and to see whether a particular model is an adequate representation of the data. We will concentrate on its application to evaluating rival models. The method starts with the *data generating process*. If the *DGP* were known, it would be possible to predict how specific models would behave. Hence, if one model can predict the behaviour of parameters of the other one, then a feature of the "true" *DGP* is captured. Cox's test (or alternative non-nested hypotheses tests), on which the encompassing philosophy relies, is regarded as a statistical tool that may be used to approximate the *DGP* closer and closer. The adherents of encompassing relate this to Lakatos's notion of progressiveness (Mizon and Richard 1986). Whether this is really what Lakatos had in mind may be questioned[19], furthermore, there are some

[19] Empirical progress according to Lakatos is the result of novel theoretical predictions which turn out to be correct. Mizon and Richard are writing about new *interpretations* of given data.

problems with regard to the idea of such scientific progression (see Feyerabend (1975, Ch. 15) for a strong critique on Lakatos's view on the development of physics). Other criticisms on the encompassing methodology are related to the evaluation of rival theories at their maximum likelihood estimates (the Bayesian would compare the complete densities), and the problem of the choice of appropriate significance levels and the level of power. The fact that two rival theories may have different prior probabilities should be part of an inferential strategy as well.

Encompassing of $\mathcal{M}_2$ by $\mathcal{M}_1$ implies to interpret the parameters of $\mathcal{M}_2$ within the context of $\mathcal{M}_1$. The parameters of $\mathcal{M}_2$ are interpreted *as if* $\mathcal{M}_1$ is true (this gives *pseudo true values* for the parameters in $\mathcal{M}_2$, see White 1982). If no information is lost by interpreting the parameters in $\mathcal{M}_2$ as the pseudo true ones under $\mathcal{M}_1$, then $\mathcal{M}_1$ is said to encompass $\mathcal{M}_2$. Florens, Hendry and Richard (1989) extend the methodology of Mizon and Richard (1986) to a Bayesian setting in which pseudo true values are not needed.

A different frequentist approach to the problem of model choice is presented in Vuong (1989) and for a more simple case in Lien and Vuong (1987), who explicitly drop the assumption that either one or both models should be "true" in the statistical sense (like Mizon and Richard, Vuong relies on pseudo-true values). Vuong's is the third strategy in the frequentist domain, the *symmetric* or *equivalence* approach (which dates back to work of Hotelling in 1940)[20]. In contrast to the other two approaches, there is no implicit assumption that either one of the rival models is true, or that the embedding model is true. Instead, Vuong provides a very general framework in which nested, non-nested and overlapping models can be compared, with the possibility that one or both models may be mis-specified. An information criterion is used (the Kullback Leibler Information Criterion, KLIC), which measures the distance between a given distribution and the "true" distribution. The

---

[20] Theil presented another symmetric "test", by showing that maximizing the R^2 corrected for degrees of freedom is equivalent to minimizing the expected unexplained variance of the regression, which would lead to obtaining the correct model "on the average". For a discussion, see Gaver and Geisel 1974, p. 53.

idea is "to define the "best" model among a collection of competing models to be the model that is closest to the true distribution" (Vuong 1989, p. 309). The advantage of generality is countered by a loss of power to discriminate between the models, which together with the computational complications is a drawback of this approach.

The second methodology is based on the Bayesian perspective and considers the problem as a decision-theoretic one, to which Jeffreys's (1961) *posterior odds ratio's* can be applied. Zellner (1971) and Gaver and Geisel (1974) provide discussions of this application to comparing models. The basic idea is very simple. Instead of applying Bayes's theorem to infer about parameters, we now apply it to inference of models. More explicitly: assume as before we have two models to compare: $\mathcal{M}_1$ and $\mathcal{M}_2$, with parameter vectors β_1 and β_2 respectively. The prior probability density functions for these vectors are $p(\beta_i|\mathcal{M}_i)$, $i=1,2$. The prior probabilities of the models are $p(\mathcal{M}_i)$. These probabilities are revised in the light of the data to posterior probabilities of the models using Bayes's Theorem:

$$p(\mathcal{M}_i|\mathbf{y}) = \frac{p(\mathcal{M}_i)p(\mathbf{y}|\mathcal{M}_i)}{p(\mathbf{y})} \tag{5.5}$$

The posterior odds ratio for comparing the two models is:

$$\mathbf{POR} = \frac{p(\mathcal{M}_1|\mathbf{y})}{p(\mathcal{M}_2|\mathbf{y})} = \frac{p(\mathcal{M}_1)}{p(\mathcal{M}_2)} \cdot \frac{\int p(\beta_1|\mathcal{M}_1)p(\mathbf{y}|\beta_1,\mathcal{M}_1)d\beta_1}{\int p(\beta_2|\mathcal{M}_2)p(\mathbf{y}|\beta_2,\mathcal{M}_2)d\beta_2} \tag{5.6}$$

(see Zellner 1971, p. 298). The posterior odds equals the *prior odds* times the ratio of the weighted or *averaged likelihoods* of the two models. The weighting is the result of the uncertainty about the parameters β_i which underlies the basic philosophy of Bayesian inference. The frequentist approach evaluates the likelihood ratio at the (maximum likelihood) point estimates for β_i, the Bayesian evaluates the whole posterior density for β_i. Hence, **5.6** remains different from Vuong's tests, even though Vuong drops the claim that any of his models may be regarded as if they represent the truth.

The prior odds ratio is simply explained as one's betting ratio of model 1 versus model 2 before considering the evidence as provided by the (new) data. If there is no *a priori* reason to prefer one over the other model, then the prior odds ($p(\mathcal{M}_1)/p(\mathcal{M}_2)$) may be set to one.

The second part of formula 5.6, called the Bayes Factor, is the ratio of the marginal densities of **y** in the light of model 1 (with parameters β_1) and model 2 (with parameters β_2) respectively. Models 1 and 2 are the competing models that are used as "windows" through which the data can be seen. The Bayes factor reveals something of the quality of the windows. If the prior confidence is very much in favour of $\mathcal{RP}_1$ and not $\mathcal{RP}_2$, and you think the window provided by $\mathcal{M}_1$ is *a priori* satisfactory, your prior odds will be high. But the data may force you to adjust your opinion: if the Bayes factor favours $\mathcal{M}_2$ this has to lead to a revision of the prior odds. A rational researcher has either to revise the confidence in $\mathcal{RP}_1$, or to adjust the research programme such that it might be able to regain confidence. The comparison between the rival models is not a real test in the sense that a decision must be made about dropping one or the other model. Instead, it is a way to obtain information on the relative quality of a model in a certain setting, this information may be used for further inference. The Bayesian view deals with learning from data, this view is particularly useful in the context of appraising rival theories.

An outright rejection of a research programme is not demanded from a rational investigator. In this sense, the Bayesian approach is in agreement with Lakatos's. The difference is that Bayesianism allows to quantify, to a certain degree at least, the otherwise vague notion of "progress" and "degeneration". Let us see how this is done.

Say we have two competing theories (e.g. the permanent income hypothesis versus the Keynesian consumption theory, for a treatment of the former see e.g. the paper by Campbell 1987, quoted before), $\mathcal{T}_1$ and $\mathcal{T}_2$. They are not static but develop over time, i.e. they are part of competing research programmes. At a given point of time, $\mathcal{T}_1$ may have a higher probability than $\mathcal{T}_2$, measured in posterior odds, but it is possible that more progress is made around $\mathcal{T}_2$ than in the research programme of its competitor. Let us assume that the theories and experimental designs have resulted in two rival models, for example linear regression models. For example, $\mathcal{M}_1$ implies that temporary fluctuations in current income do not matter for consumption behaviour, whereas $\mathcal{M}_2$ implies a strong positive correlation. We can calculate the posterior odds for these models. They can be formulated according to **5.6** or more simply as:

$$\mathbf{POR} = \frac{p(\mathcal{M}_1)\cdot p(\mathrm{Y}|\mathcal{M}_1)}{p(\mathcal{M}_2)\cdot p(\mathrm{Y}|\mathcal{M}_2)}$$

Assume this ratio favours $\mathcal{M}_1$, the permanent income hypothesis, **POR** is bigger than one. The adherent of $\mathcal{M}_2$ does not give up his theory but changes the theory by adding some additional arguments, $\mathcal{A}$. For example, he may introduce some dynamics in the Keynesian consumption function in the form of an error correction specification. If these additions fit in very well with $\mathcal{M}_2$ (i.e. have high prior probability given $\mathcal{M}_2$, which amount to say that the theoretical innovation is not *ad hoc* in Lakatos's first sense), and if they, together with $\mathcal{M}_2$, lead to new empirical predictions, the result will be a new posterior odds ratio, which is equal to:

$$\mathbf{POR}^* = \frac{p(\mathcal{M}_1)\cdot p(\mathrm{Y}|\mathcal{M}_1)}{p(\mathcal{M}_2)\cdot p(\mathrm{Y}|\mathcal{M}_2,\mathcal{A})\cdot p(\mathcal{A}|\mathcal{M}_2)}$$

In other words, the odds change favourably to the second theory if the additional or augmentive hypothesis $\mathcal{A}$ has high probability given $\mathcal{M}_2$ (i.e. given the environment of its research programme), and if $\mathcal{M}_2$ and $\mathcal{A}$ jointly predict the occurance of data better than $\mathcal{M}_2$ itself did. Another way of viewing this is to define $\mathcal{M}_2$ and $\mathcal{A}$ together as a new model, $\mathcal{M}_3$, and calculate the posterior odds of the old versus the new instance of this research programme. If the odds are bigger than one, the innovation is "progressive". It will also lead to an improvement of the odds versus the competitor, as straightforward calculation shows.

If the data are so outspokenly at odds with theory $\mathcal{M}_2$ that the prior odds are blown away by the data information, and if no additional hypothesis can be thought of that is both likely in the context of this theory and in the context of the data, than the opinions of researcher of rival research programmes may converge towards $\mathcal{M}_1$. Subjectivity (in the sense of differences of private opinions) vanishes in the limit, where one theory is true and the sample-size grows to infinity. In this case, the method of posterior odds and frequentist non-nested testing will lead to the same decision.

The Bayesian methodology (whether the posterior odds approach, or the encompassing approach as presented in Florens *et.al.* (1989)) has the advantage that it deals explicitly with uncertainty about parameters and models. This is at the cost of analytical tractability: calculation of 5.6 can be cumbersome if informative priors are used, and may lead to nonsensical results if non-informative priors are used (see Gaver and Geisel 1974). In fact, the latter implies that it is impossible to evaluate rival models without any *a priori* information. Judging theories only on ground of maximum likelihood evaluation is as misleading as calculating posterior odds based upon non-informative priors.

The foregoing discussion of frequentist and Bayesian strategies to evaluate rival models would be of little interest if applications were absent. Most econometric research fits in the frequentist tradition, hence most applications of testing rival models should be found here. Indeed, quite a few applications appear in the literature. Tests of Keynesian versus New-Classical explanations of unemployment are presented in McAleer *et. al.* (1990), who also provide references to non-nested tests of investment models, income versus wealth in money demand equations, and the choice of the appropriate interest rate in a money demand function.

Applications of the Bayesian posterior odds are rare. Neftçi (1984) uses POR's to answer the question "Are economic time series asymmetric?". Geweke (1987) analyzes the cyclical behaviour of real GDP and uses posterior odds ratios to compare models with damped, oscillatory or explosive behaviour. I am not aware of applications to compare models that belong to rival research programmes, even though in such an application the Bayesian philosophy comes to its full attractiveness.

7. Conclusion

Haavelmo proposed to make better use of probabilistic underpinnings in empirical economics. His major interest was in the simultaneous character of economic models, but he also gave a number of interesting general methodological statements. In this paper, we tried to see if there is any link between Haavelmo's claim that statistical work in economics needed sound probabilistic foundations, and the problem of appraising economic theories.

The first point of this paper is to emphasize the importance of probability methods in economic inference. It is sad that this should be emphasized, but methodologists have not shown much interest in this area so far. Secondly, we showed how data might be informative on rival economic theories. Furthermore, we analyzed some difficulties in interpreting the results of statistical inference if "the" model is not given, and compared frequentist and Bayesian alternatives.

Thirdly, it was argued that use of probabilistic inference in economics involves subjectivity. This subjectivity is inevitable. Haavelmo (1944, p. 3) already acknowledged the limitations of probabilistic inference:

"(...) it is not to be forgotten that they [our explanations] are all our own artificial inventions in a search for an understanding of real life; they are not hidden truths to be "discovered"."

An artificial invention is man-made, by definition. It involves taste, judgment and creativity. Decision criteria, such as the 95% significance level, are matters of taste as well (the existing literature on optimal choice of significance levels has a Bayesian spirit, like applications of the principle of minimizing the maximum regret. The choice of the regret function remains arbitrary (for a discussion see Amemiya 1985, p. 52-55). Rules of Bayesian Inference set bounds to the arbitrariness of model selection or scientific progress.

The demand of (neo-) Popperians for more testing in economics is misguided if a proper foundation for probabilistic inference is not provided in the mean time. The Bayesian philosophy provides both a sound theory of inference and guidelines for the use and interpretation of probabilistic "tests" in econometrics. This conclusion has brought us far from the initial work of Haavelmo, but provides solutions to problems that were raised already by Haavelmo and his contemporaries, when they defended the probabilistic methodology in econometrics.

Literature

Amemiya, Takeshi (1985), *Advanced Econometrics*, Harvard U.P., Cambridge

Barnett, Vic (1973), *Comparative Statistical Inference*, J. Wiley, London

Barten, Anton and Hugo A. Keuzenkamp (forthcoming), Rejection without Falsification: the story of testing homogeneity in consumer demand theory, mimeo

Bruce Caldwell (1982), *Beyond Positivism, Economic Methodology in the Twentieth Century*, George Allen and Unwin, London

Campbell, John (1987), Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis, *Econometrica* **55** no. 6, p. 1249-1273

Cartwright, Nancy (1983), *How the Laws of Physics Lie*, Clarendon Press, Oxford

Cox, D.R. (1961), Tests of Separate Families of Hypotheses, *Proc. Berkeley Symp. 4th.* **1**, p. 105-123

Cox, D.R. (1962), Further Results on Tests of Separate Families of Hypotheses, *J.Roy.Stat.Soc. Ser. B* **24**, 406-424

Epstein, Roy (1987), *A History of Econometrics*, Elseviers Science Publishers, Amsterdam

Feyerabend, Paul (1975), *Against Method*, Verso, London, 1978

Florens, Jean-Pierre, David F. Hendry and Jean François Richard (1989), Encompassing and Specificity, *mimeo*

Friedman, Milton and D. Meiselman (1963), The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897-1958, in *Stabilization Policies: Commission on Money and Credit*, Prentice Hall, Englewood Cliffs

Geweke, John (1988), The Secular and Cyclical Behavior of Real GDP in Nineteen OECD Countries, 1957-1983, *J. of Business & Economics Statistics* **6**, no.4, Oct. 1988 p.479-486

Haavelmo, Trygve (1944), *The Probability Approach in Econometrics*, supplement to *Econometrica* **12**, July

Hacking, Ian (1983), *Representing and Intervening*, CUP, Cambridge

Hausman, Jerry A. (1978), Specification Tests in Econometrics, *Econometrica* **46** no. 6, November, p. 1251-1271

Hausman, Daniel (1989), Economic Methodology in a Nutshell, *Journal of Economic Perspectives* Spring 1989, **3** no.2 p. 115-127

Hill, Bruce M. (1974), On Coherence, Inadmissibility and Inference About Many Parameters in the Theory of Least Squares, in: *Studies in Bayesian Econometrics and Statistics* (S. Fienberg and A. Zellner, Eds.), North Holland, Amsterdam

_____ (1986), Some Subjective Bayesian Considerations in the Selection of Models, *Econometric Reviews* **4**(2), p. 191-246

Howson, Colin and Peter Urbach (1989), *Scientific Reasoning: The Bayesian Approach*, Open Court, La Salle, Illinois

Jeffreys, Harold (1961), *Theory of Probability* (3rd ed.), Clarendon, Oxford

Kuhn, Thomas S. (1962), *The Structure of Scientific Revolutions*, U. of Chicago Press, Chicago

Lakatos, Imre (1970), The Methodology of Scientific Research Programmes, in: Lakatos and Musgrave (1970), *Criticism and the Growth of Knowledge*, CUP, Cambridge

Leamer, Edward E. (1978), *Specification Searches, ad hoc inference with nonexperimental data*, John Wiley, New York

Lien, Donald and Quang H. Vuong (1987), Selecting the Best Linear

Regression Model (A Classical Approach), *Journal of Econometrics* **35**, suppl., p. 3-23

Lindley, Dennis V. (1987), The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence and Expert Systems, *Statistical Science* **2**, p. 17-24

Lucas, Robert E. Jr. (1976), Econometric Policy Evaluation, A Critique, in: K.Brunner and A.H. Meltzer (eds.), *The Phillips Curve and Labour Markets, Suppl. to Journal of Monetary Economy*

Mises, Richard von (1957), *Probability, Statistics and Truth*, 2nd revised ed., first ed. (German) 1928, (English) 1939, Dover 1981

Mizon, Grayham E. and Jean François Richard (1986), The Encompassing Principle and its Application to Testing Non-Nested Hypotheses, *Econometrica* **54**, p. 657-678

Morgan, Mary (1990), *The History of Econometric Ideas*, CUP, Cambridge

Neftçi, Salih N. (1984), Are Economic Time Series Asymmetric Over the Business Cycle?, *J.Political Economy* **92** no.2, p.307-328

Pesaran, M.H. (1982), On The Comprehensive Method of Testing Non-Nested Regression Models, *J. of Econometrics* **18** no. 2, p. 263-274

Popper, Karl R. (1959), *The Logic of Scientific Discovery*

Rhodes, Richard (1986), *The Making of the Atomic Bomb*, Penguin Books, London

Rosenkrantz, Roger (1983), Why Glymour is a Bayesian, in: John Earman (ed.), *Minnesota Studies in the Philosophy of Science, Volume X: Testing Scientific Theories*, U. of Minnesota Press, Minneapolis

Sargent, Thomas J. (1976), The Observational Equivalence of Natural and Unnatural Rate Theories of Macroeconomics, *Journal of Political Economy* **84**, no. 3

Shafer, Glenn (1987), Probability Judgment in Artificial Intelligence and Expert Systems, *Statistical Science* **2**, p. 3-16

Spanos, Aris (1989), On Rereading Haavelmo: A Retrospective View of Econometric Modeling, *Econometric Theory* **5**, p. 405-429

Swamy, P.A.V.B., R.K. Conway, and P. von zur Muehlen (1985), The Foundations of Econometrics, Are There Any?, *Econometric Reviews* **4** no. 1, p. 1-61

Vuong, Quang H. (1989), Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica* **57**, p. 307-333

White, Hal (1982), Maximum Likelihood Estimation of Misspecified Models, *Econometrica* **50**, p. 1-26

Zellner, Arnold (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York

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

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
8932	E. van Damme, R. Selten and E. Winter	Alternating Bid Bargaining with a Smallest Money Unit
8933	H. Carlsson and E. van Damme	Global Payoff Uncertainty and Risk Dominance
8934	H. Huizinga	National Tax Policies towards Product-Innovating Multinational Enterprises
8935	C. Dang and D. Talman	A New Triangulation of the Unit Simplex for Computing Economic Equilibria
8936	Th. Nijman and M. Verbeek	The Nonresponse Bias in the Analysis of the Determinants of Total Annual Expenditures of Households Based on Panel Data
8937	A.P. Barten	The Estimation of Mixed Demand Systems
8938	G. Marini	Monetary Shocks and the Nominal Interest Rate
8939	W. Güth and E. van Damme	Equilibrium Selection in the Spence Signaling Game
8940	G. Marini and P. Scaramozzino	Monopolistic Competition, Expected Inflation and Contract Length
8941	J.K. Dagsvik	The Generalized Extreme Value Random Utility Model for Continuous Choice
8942	M.F.J. Steel	Weak Exogenity in Misspecified Sequential Models
8943	A. Roell	Dual Capacity Trading and the Quality of the Market
8944	C. Hsiao	Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data
8945	R.P. Gilles	Equilibrium in a Pure Exchange Economy with an Arbitrary Communication Structure
8946	W.B. MacLeod and J.M. Malcomson	Efficient Specific Investments, Incomplete Contracts, and the Role of Market Alternatives
8947	A. van Soest and A. Kapteyn	The Impact of Minimum Wage Regulations on Employment and the Wage Rate Distribution
8948	P. Kooreman and B. Melenberg	Maximum Score Estimation in the Ordered Response Model

No.	Author(s)	Title
8949	C. Dang	The D_3-Triangulation for Simplicial Deformation Algorithms for Computing Solutions of Nonlinear Equations
8950	M. Cripps	Dealer Behaviour and Price Volatility in Asset Markets
8951	T. Wansbeek and A. Kapteyn	Simple Estimators for Dynamic Panel Data Models with Errors in Variables
8952	Y. Dai, G. van der Laan, D. Talman and Y. Yamamoto	A Simplicial Algorithm for the Nonlinear Stationary Point Problem on an Unbounded Polyhedron
8953	F. van der Ploeg	Risk Aversion, Intertemporal Substitution and Consumption: The CARA-LQ Problem
8954	A. Kapteyn, S. van de Geer, H. van de Stadt and T. Wansbeek	Interdependent Preferences: An Econometric Analysis
8955	L. Zou	Ownership Structure and Efficiency: An Incentive Mechanism Approach
8956	P.Kooreman and A. Kapteyn	On the Empirical Implementation of Some Game Theoretic Models of Household Labor Supply
8957	E. van Damme	Signaling and Forward Induction in a Market Entry Context
9001	A. van Soest, P. Kooreman and A. Kapteyn	Coherency and Regularity of Demand Systems with Equality and Inequality Constraints
9002	J.R. Magnus and B. Pesaran	Forecasting, Misspecification and Unit Roots: The Case of AR(1) Versus ARMA(1,1)
9003	J. Driffill and C. Schultz	Wage Setting and Stabilization Policy in a Game with Renegotiation
9004	M. McAleer, M.H. Pesaran and A. Bera	Alternative Approaches to Testing Non-Nested Models with Autocorrelated Disturbances: An Application to Models of U.S. Unemployment
9005	Th. ten Raa and M.F.J. Steel	A Stochastic Analysis of an Input-Output Model: Comment
9006	M. McAleer and C.R. McKenzie	Keynesian and New Classical Models of Unemployment Revisited
9007	J. Osiewalski and M.F.J. Steel	Semi-Conjugate Prior Densities in Multi-variate t Regression Models

No.	Author(s)	Title
9007	J. Osiewalski and M.F.J. Steel	Semi-Conjugate Prior Densities in Multivariate t Regression Models
9008	G.W. Imbens	Duration Models with Time-Varying Coefficients
9009	G.W. Imbens	An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling
9010	P. Deschamps	Expectations and Intertemporal Separability in an Empirical Model of Consumption and Investment under Uncertainty
9011	W. Güth and E. van Damme	Gorby Games - A Game Theoretic Analysis of Disarmament Campaigns and the Defense Efficiency-Hypothesis
9012	A. Horsley and A. Wrobel	The Existence of an Equilibrium Density for Marginal Cost Prices, and the Solution to the Shifting-Peak Problem
9013	A. Horsley and A. Wrobel	The Closedness of the Free-Disposal Hull of a Production Set
9014	A. Horsley and A. Wrobel	The Continuity of the Equilibrium Price Density: The Case of Symmetric Joint Costs, and a Solution to the Shifting-Pattern Problem
9015	A. van den Elzen, G. van der Laan and D. Talman	An Adjustment Process for an Exchange Economy with Linear Production Technologies
9016	P. Deschamps	On Fractional Demand Systems and Budget Share Positivity
9017	B.J. Christensen and N.M. Kiefer	The Exact Likelihood Function for an Empirical Job Search Model
9018	M. Verbeek and Th. Nijman	Testing for Selectivity Bias in Panel Data Models
9019	J.R. Magnus and B. Pesaran	Evaluation of Moments of Ratios of Quadratic Forms in Normal Variables and Related Statistics
9020	A. Robson	Status, the Distribution of Wealth, Social and Private Attitudes to Risk
9021	J.R. Magnus and B. Pesaran	Evaluation of Moments of Quadratic Forms in Normal Variables

No.	Author(s)	Title
9022	K. Kamiya and D. Talman	Linear Stationary Point Problems
9023	W. Emons	Good Times, Bad Times, and Vertical Upstream Integration
9024	C. Dang	The D_2-Triangulation for Simplicial Homotopy Algorithms for Computing Solutions of Nonlinear Equations
9025	K. Kamiya and D. Talman	Variable Dimension Simplicial Algorithm for Balanced Games
9026	P. Skott	Efficiency Wages, Mark-Up Pricing and Effective Demand
9027	C. Dang and D. Talman	The D_1-Triangulation in Simplicial Variable Dimension Algorithms for Computing Solutions of Nonlinear Equations
9028	J. Bai, A.J. Jakeman and M. McAleer	Discrimination Between Nested Two- and Three-Parameter Distributions: An Application to Models of Air Pollution
9029	Th. van de Klundert	Crowding out and the Wealth of Nations
9030	Th. van de Klundert and R. Gradus	Optimal Government Debt under Distortionary Taxation
9031	A. Weber	The Credibility of Monetary Target Announcements: An Empirical Evaluation
9032	J. Osiewalski and M. Steel	Robust Bayesian Inference in Elliptical Regression Models
9033	C. R. Wichers	The Linear-Algebraic Structure of Least Squares
9034	C. de Vries	On the Relation between GARCH and Stable Processes
9035	M.R. Baye, D.W. Jansen and Q. Li	Aggregation and the "Random Objective" Justification for Disturbances in Complete Demand Systems
9036	J. Driffill	The Term Structure of Interest Rates: Structural Stability and Macroeconomic Policy Changes in the UK
9037	F. van der Ploeg	Budgetary Aspects of Economic and Monetary Integration in Europe

No.	Author(s)	Title
9038	A. Robson	Existence of Nash Equilibrium in Mixed Strategies for Games where Payoffs Need not Be Continuous in Pure Strategies
9039	A. Robson	An "Informationally Robust Equilibrium" for Two-Person Nonzero-Sum Games
9040	M.R. Baye, G. Tian and J. Zhou	The Existence of Pure-Strategy Nash Equilibrium in Games with Payoffs that are not Quasiconcave
9041	M. Burnovsky and I. Zang	"Costless" Indirect Regulation of Monopolies with Substantial Entry Cost
9042	P.J. Deschamps	Joint Tests for Regularity and Autocorrelation in Allocation Systems
9043	S. Chib, J. Osiewalski and M. Steel	Posterior Inference on the Degrees of Freedom Parameter in Multivariate-t Regression Models
9044	H.A. Keuzenkamp	The Probability Approach in Economic Methodology: On the Relation between Haavelmo's Legacy and the Methodology of Economics

P.O. BOX 90153, 5000 LE TILBURG, THE NETHERLAND