
Probability, Econometrics and Truth

The methodology of econometrics

Hugo A. Keuzenkamp



PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge CB2 2RU, UK www.cup.cam.ac.uk

40 West 20th Street, New York, NY 10011-4211, USA www.cup.org

10 Stamford Road, Oakleigh, Melbourne 3166, Australia

Ruiz de Alarcón 13, 28014 Madrid, Spain

© Cambridge University Press 2000

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2000

Printed in the United Kingdom at the University Press, Cambridge

Typeface 10/12 Times *System* 3B2 [kw]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Keuzenkamp, Hugo.

Probability, econometrics and truth: the methodology of econometrics/Hugo Keuzenkamp.

p. cm.

Includes indexes.

ISBN 0 521 55559 8 (hardback)

1. Econometrics. 2. Probabilities. I. Title

HB139.K48 2000

330'.01'5195-dc21

99-058732

ISBN 0 521 553 59 8 hardback

Contents

	<i>page</i>
<i>Introduction</i>	vii
1 The philosophy of induction	1
2 Probability and indifference	21
<i>Intermezzo: a formal scheme of reference</i>	28
3 Reactive frequency and induction	33
4 Probability and belief	67
5 The theory of simplicity	98
6 From probability to econometrics	119
7 Econometric modelling	142
8 In search of homogeneity	180
9 Positivism and the aims of econometrics	213
10 Probability, econometrics and truth	262
<i>Personalia</i>	276
<i>References</i>	281
<i>Name Index</i>	299
<i>Subject Index</i>	307

1 The philosophy of induction

[S]ome other scientists are liable to say that a hypothesis is definitely proved by observation, which is certainly a logical fallacy; most statisticians appear to regard observations as a basis for possibly rejecting hypotheses, but in no case for supporting them. The latter attitude, if adopted consistently, would reduce all inductive inference to guesswork.

Harold Jeffreys ([1939] 1961, p. ix)

1 Introduction

Occasionally, the aspirations of econometrics are frustrated by technical difficulties which lead to increasing technical sophistication. More often, however, deeper problems hamper econometrics. These are the problems of scientific inference – the logical, cognitive and empirical limitations to induction. There is an escapist tendency in econometrics, which is to seek salvation in higher technical sophistication and to avoid deeper philosophical problems. This is reflected by the erosion of an early foothold of empirical econometrics, *Econometrica*. The share of empirical papers has declined from a third in the first (1933) volume to a fifth in recent volumes. This is not because most empirical values for economic variables or parameters have been settled. Despite the ‘econometric revolution’, there is no well established numerical value for the price elasticity of bananas. If *Econometrica* were to publish an issue with well established econometric facts, it might be very thin indeed. The factual knowledge of the economy remains far from perfect, as are the ability to predict its performance, and the understanding of its underlying processes. Basic economic phenomena, such as the consumption and saving patterns of agents, remain enigmatic. After many years of econometric investigation, there is no agreement on whether money causes output or not. Rival theories flourish. Hence, one may wonder what the added-value of econometrics is. Can we learn from experience in economics, and, if so, does econometrics itself serve this purpose? Or, were the aspirations too high after all, and does the sceptical attitude of Keynes half a century ago remain justified today?

2 Humean scepticism

An important issue in the philosophy of science is how (empirical) knowledge can be obtained.¹ This issue has a long history, dating back (at least) to the days of the Greek Academy, in particular to the philosopher Pyrrho of Elis (c. 365–275 BC), the first and most radical sceptic. Academic scepticism, represented for example by Cicero (106–43 BC), is more moderate than Pyrrho's. The ideas of Pyrrho (who did not write books, 'wisely' as Russell, 1946, p. 256, remarks) are known via his pupil Timon of Phlius (c. 320–230 BC) and his follower Sextus Empiricus (second century AD), whose work was translated into Latin in 1569. A few earlier translations are known but they have probably only been read by their translators. The 1569 translation was widely studied in the sixteenth and seventeenth centuries. All major philosophers of this period referred to scepticism. René Descartes, for example, claimed to be the first philosopher to refute scepticism.

One of the themes of the early sceptics is that only deductive inference is valid (by which they mean: logically acceptable) for a demonstrative proof, while induction is invalid as a means for obtaining knowledge. Perception does not lead to general knowledge. According to Russell (1946, p. 257),

Scepticism naturally made an appeal to many unphilosophic minds. People observed the diversity of schools and the acerbity of their disputes, and decided that all alike were pretending to knowledge which was in fact unattainable. Scepticism was a lazy man's consolation, since it showed the ignorant to be as wise as the reputed men of learning.

Still, there was much interest in scepticism since the publication of the translation of Sextus Empiricus' work, not only by 'unphilosophic minds'. Scepticism has been hard to refute. Hume contributed to the sceptical doctrine (although he did not end up as a Pyrrhonian, *i.e.* radical sceptic). The result, 'Humean scepticism', is so powerful, that many philosophers still consider it to be a death blow to induction, the 'scandal of philosophy'.²

Hume ([1739] 1962) argues that the empirical sciences cannot deliver causal knowledge. There are no rational grounds for understanding the causes of events. One may observe a sequence of events and call them cause and effect, but the connection between the two remains hidden. Generalizations deserve scepticism. Hume (Book I, Part III, section 12, p. 189) summarizes this in two principles:

that there is nothing in any object, considered in itself, which can afford us a reason for drawing a conclusion beyond it; and, that even after the observation of the

frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience.

The ‘scandal of philosophy’ is fundamental to empirical scientific inference (not just econometrics). It has wider implications (as Hume indicates) than denying causal inference. For example, does past experience justify the expectation of a sunrise tomorrow? The question was raised in discussing the merits of Pierre Simon de Laplace’s ‘rule of succession’, a statistical device for induction (see chapter 2).³ Another example, popular in philosophy, deals with extrapolation to a population instead of the future: if only white swans have been observed, may we infer that all swans are white? (This is the classic example of an affirmative universal statement.)

The sceptical answer to these questions is negative. The rules of deductive logic prohibit drawing a general conclusion if this conclusion is not entailed by its propositions. There is no *logical* reason why the next swan should be white. Of course, swans can be defined to be white (like statisticians who define a fair die to be unbiased), making black swans a contradiction in terms. An alternative strategy is to conclude that all *known* swans are white. The conclusion is conditional on the observed sample. Hence, the choice is between formulating definitions or making conditional enumerations. But most empirical scientists want to make generalizations. This is impossible if the induction problem proves insurmountable. Therefore, an understanding of induction is essential.

The logical form of the induction problem is that all *observed* X are Φ does not entail that all X are Φ . The next three chapters, dealing with probabilistic inference, consider a more delicate, probabilistic form of the induction problem: given that *most* observed X are Φ , what can be said about X in general? The source of Humean scepticism follows from the conjunction of three propositions (Watkins, 1984, p. 3):

- (i) there are no synthetic *a priori* truths about the external world;
- (ii) any genuine knowledge we have of the external world must ultimately be derived from perceptual experience;
- (iii) only deductive derivations are valid.

The conjunction of (i), (ii) and (iii) does not allow for inferring knowledge beyond the initial premises. In this sense, inductive inference is impossible.

John Watkins (p. 12) argues that a philosophical, or ‘rational’ answer to scepticism is needed, because otherwise it is likely to encourage irrationality. Watkins holds that Hume himself regarded philosophical scepticism as an academic joke. Indeed, Hume uses the expression *jeux d’esprit* (in *A letter From a Gentleman to his Friend in Edinburgh*, included

as an appendix to Hume [1748] 1977, p. 116). Describing the person who is afflicted by Pyrrhonism, Hume (p. 111) concludes:

When he awakes from his dream, he will be the first to join in the laugh against himself, and to confess, that all his objections are mere amusement.

Amusement, Watkins (1984, p. 12) argues, does not qualify as a rational answer to scepticism. In fact, Hume's response is more elaborate than the quotation suggests. It relies on conventionalism (see below). I agree with Watkins that, formally, conventionalism is not very appealing (although conventions have much practical merit). Fortunately, there are alternatives. Once the source of Hume's problem (the threefold conjunction just mentioned) is clarified, the merits of those alternative responses to scepticism can be appraised.

Watkins (pp. 4–5) discusses a number of strategies as responses to Hume's problem. The most interesting ones are:

- the naturalist (ignoring the conjunction of propositions (i)–(iii));
- the apriorist (denying proposition (i));
- the conjecturalist (amending proposition (ii)); and
- the probabilist strategy (which takes odds with proposition (iii)).

A more detailed discussion of the probabilist strategy will be given in the next three chapters, while the remainder of this book considers how well this strategy may work in econometrics.

3 Naturalism and pragmatism

Descartes argued that one should distrust sensations. Insight in causal relations results from mere reasoning. Hume, criticizing Cartesian 'dogmatic rationalism', argues that such plain reasoning does not suffice to obtain unique answers to scientific questions. Cartesian doubt, 'were it ever possible to be attained by any human creature (as it plainly is not) would be entirely incurable' (Hume [1748] 1977, p. 103). It would not yield true knowledge either: 'reasoning *a priori*, any thing might appear able to produce anything' (*Letter From a Gentleman*, p. 119). Cartesian doubt is unacceptable to Hume ([1739] 1962, p. 318). It gave him a headache:

The *intense* view of these manifold contradictions and imperfections in human reason has so wrought upon me, and heated my brain, that I am ready to reject all belief and reasoning, and can look upon no opinion even as more probable or likely than another.

But this does not make Hume a Pyrrhonian or radical sceptic. He is rescued from this philosophical 'melancholy and delirium' by nature.

His naturalist strategy is to concede that there is no epistemological answer to scepticism, but to deny its importance. It is human nature to make generalizing inferences, the fact that inference is not warranted from a logical point of view has no practical implications. Hume (*An Abstract of a Book Lately Published, Entitled, A Treatise of Human Nature, Etc.*, in Hume [1739] 1962, p. 348) concludes,

that we assent to our faculties, and employ our reason only because we cannot help it. Philosophy would render us entirely *Pyrrhonian*, were not nature too strong for it.

The great subverter of Pyrrhonism, Hume ([1748] 1977, p. 109) writes, is 'action, and employment, and the occupations of common life'. Not reasoning, but custom and habit, based on the awareness of constant conjunctions of objects, make human beings draw inferences (p. 28). This response is known as conventionalism. According to Hume (p. 29), custom is the 'great guide of human life', and without custom or habit, those who are guided only by Pyrrhonian doubt will 'remain in a total lethargy, till the necessities of nature, unsatisfied, put an end to their miserable existence' (p. 110). Reason is the slave of our passions.

A pinch of Pyrrhonian doubt remains useful, because it makes investigators aware of their fallibility (p. 112). The fact that one cannot obtain absolute certainty by human reasoning does not imply universal doubt, but only suggests that researchers should be modest (*Letter From a Gentleman*, p. 116). But many scientists will feel embarrassed by the conclusion that custom is the ultimate foundation of scientific inference. Watkins, for example, rejects it. However, conventionalism may be rationally justified. This has been attempted by some adherents of the probabilistic approach. Other strategies related to Hume's conventionalism are instrumentalism (developed by John Dewey) and pragmatism, or pragmaticism, as Charles Peirce christened it. These hold that hypotheses may be accepted and rejected on rational grounds, on the basis of utility or effectiveness. The pragmatic approach can be combined with the probabilistic strategy. But it is not free of problems. Most importantly, it is an invitation to scientific obscurantism (should a theory be useful to the learned – who qualifies? – or to the mighty?). A problem with conventionalism is to give an answer to the question 'where do these conventions come from?' and to provide a rational justification for the conventions (evolutionary game theory has been directed to this question). Lawrence Boland (1982; also 1989, p. 33) argues that neoclassical economists deal with the induction problem by adopting a conventionalist strategy. Econometricians base much of their work on another convention concerning the size of a test: the well known 5% significance level. This

convention has its roots in a quarrel between Karl Pearson and R. A. Fisher, two founders of modern statistics (see chapter 3, section 3.2).

4 Apriorism

The *apriorist strategy* to the problem of scepticism denies proposition (i), concerning the absence of synthetic *a priori* truth. Immanuel Kant invented this notion of *a priori* synthetic truth, true knowledge that is both empirical and based on reasoning. It is neither analytic nor synthetic.⁴ The canonical example of an *a priori* synthetic truth is Kant's Principle of Universal Causation, which is his response to Humean scepticism. Kant argued that everything must have a cause: 'Everything that happens presupposes something upon which it follows in accordance with a rule' (translated from *Kritik der reinen Vernunft*, Kant's most important work, published in 1781; in Krüger, 1987, p. 72). This doctrine is also known as causal determinism, or simply as causalism (Bunge [1959] 1979, p. 4).

Unlike Hume, John Stuart Mill endorsed Kant's principle: for Mill, induction is the search for causes. Mill distinguishes four 'canons of induction', given in his *Logic*, III (viii); Mill [1843] 1952):

- the method of agreement;
- the method of difference;
- the method of residues;
- the method of concomitant variations.

These methods are based on the 'principle of uniformity of nature', which holds that the future will resemble the past: the same events will happen again if the conditions are sufficiently similar. The method of difference starts from the premise that all events have a cause. The next step is to give an exhaustive list of possible causes, and select the one(s) which always occurs in common with the event, and does not occur if the event does not occur. A problem is to select this exhaustive list of possible causes.

Keynes ([1921] CW VIII, p. 252) refers to the principle of uniformity of nature in his discussion of reasoning by analogy, and suggests that differences in position in time and space should be irrelevant for the validity of inductions. If this principle forms the basis for induction, it cannot itself be founded upon inductive arguments. Furthermore, it is doubtful that experience validates such a strong principle. Nature seems much more erratic and surprising than the principle of uniformity of nature suggests. Still, the late philosopher Karl Popper ([1935] 1968, p. 252) explicitly argues that 'scientific method presupposes *the immutability of natural processes*, or the "principle of the uniformity of nature"'.

Likewise, Bernt Stigum (1990, p. 542) argues that this principle is a necessary postulate of epistemology. Some probability theorists advocate a statistical version of this synthetic *a priori* truth: the stability of mass phenomena (see in particular the discussion of von Mises in chapter 3, section 2).

In the social sciences, it is not the uniformity of nature which is of interest, but the relative stability of human behaviour. A more apt terminology for the principle would then be the 'principle of stable behaviour'. Consider the axioms of consumer behaviour. If one assumes that preferences are stable (Hahn, 1985, argues this is all the axioms really say), then accepting these axioms as *a priori* truths warrants inductive generalizations. This principle solves, or rather, sidesteps, the Humean problem. If it is accepted, generalizations from human experience are admissible. But again this postulate is doubtful. Too frequently, humans behave erratically, and on a deeper level, reflexivity (self-fulfilling prophecies) may undermine uniform regularities in the social sciences. It suffers from the same problems as the principle of uniformity of nature: either it is false, or its justification involves infinite regress. But a weaker principle of stable behaviour may be accepted, by giving a probabilistic interpretation to the generalization. There should be an appreciable (non-zero) probability that stable behaviour may be expected. This is the basis for rational behaviour. A fair amount of stability is also necessary (not sufficient) for scientific inference: otherwise, it is impossible to 'discover' laws, or regularities.

It is hard to imagine interesting *a priori* synthetic truths specific to economics. The axioms of consumer behaviour are not generally accepted as true. An investigation of their validity cannot start by casting them beyond doubt (chapter 8 provides a case history of 'testing' consumer demand theory). Bruce Caldwell (1982, p. 121) discusses praxeological axioms of Austrian economists as an example of Kant's *a priori* synthetic propositions. The Austrian Friedrich von Wieser argued that a cumbersome sequence of induction is not needed to establish laws in economics. He claimed (cited in Hutchison, 1981, p. 206) that we can 'hear the law pronounced by an unmistakable inner voice'. Ludwig von Mises made apriorism the cornerstone of his methodology. The problem of this line of thought is that inner voices may conflict. If so, how are we to decide which voice to listen to?

5 Conjecturalism

The *conjecturalist strategy* denies Watkins' proposition (ii) and instead holds that scientific knowledge is only negatively controlled by experi-

ence: through falsification. Popper provided the basic insights of the conjecturalist philosophy (also known as methodological falsificationism) in his *Logik der Forschung* in 1934 (translated as Popper, [1935] 1968). This nearly coincides with one of the first efforts to test economic theory with econometric means (Tinbergen, 1939b). Followers of Popper are, among others, Imre Lakatos and Watkins. I will first discuss Popper's views on inference, then Lakatos' modified conjecturalism.

5.1 Popper's conjecturalism

Popper's impact on economic methodology has been strong. Two pronounced Popperians in economics are Mark Blaug (1980) and Terence Hutchison (1981). Moreover, statisticians and econometricians frequently make favourable references to Popper (Box, 1980, p. 383, n.; Hendry, 1980; Spanos, 1986) or believe that Popper's is 'the widely accepted methodological philosophy as to the nature of scientific progress' (Bowden, 1989, p. 3). Critics claim that the real impact of Popperian thought on economic inference is more limited (see also De Marchi, 1988; Caldwell, 1991).

5.1.1 Falsification and verification

Scientific statements are those which can be refuted by empirical observation. Scientists should make bold conjectures and try to falsify them. This is the conjecturalist view in a nutshell. More precisely, theories are thought of as mere guesses, conjectures, which have to be falsifiable in order to earn the predicate scientific. The *modus tollens* (if p , then q . But not- q . Therefore, not- p) applies to scientific inference – if a prediction which can be deduced from a generalization (theory) is falsified, then that generalization itself is false. The rules of deductive logic provide a basis for scientific rationality and, therefore, make it possible to overcome the problems of Humean scepticism. Falsifiability distinguishes science from non-science (the demarcation criterion). The growth of knowledge follows from an enduring sequence of conjectures and refutations. Theories are replaced by better, but still fallible, theories. Scientists should remain critical of their work.

So far, there seems not much controversial about the conjecturalist approach. The tentative nature of science is a commonplace. Popper went beyond the commonplace by constructing a philosophy of science on it, methodological falsificationism. A source of controversy is Popper's critique of logical positivism, the philosophy associated with the *Wiener Kreis*.⁵ A related source is his obnoxious rejection of induction.

Logical positivism holds that the possibility of empirical *verification*, rather than falsification, makes an empirical statement ‘meaningful’ (the meaning lies in its method of verification). There are many problems with this view, but Popper aimed his fire at an elementary one: affirmative universal statements, like ‘all swans are white’, are not verifiable. In response to Popper’s critique, Carnap dropped the verifiability criterion and started to work on a theory of confirmation (see also chapter 4, section 3.1). Again, this theory was criticized by Popper.

The logical difference between verification and falsification is straightforward. The observation of a white swan does not imply the truth of the claim ‘all swans are white’. On the other hand, observing a black swan makes a judgement about the truth of the claim possible. In other words, there is a logical asymmetry between verification and falsification. This asymmetry is central to Popper’s ideas: ‘It is of great importance to current discussion to notice that falsifiability in the sense of my demarcation principle is *a purely logical affair*’ (Popper, 1983, p. xx; emphasis added). This logical affair is not helpful in guiding the work of applied scientists, like econometricians. It should have real-world implications. For this purpose, Popper suggests the crucial test, a test that leads to the unequivocal rejection of a theory. Such a test is hard to find in economics.

According to Popper, it is much easier to find confirmations than falsifications. In the example of swans this may be true, but for economic theories things seem to be rather different. It is not easy to construct an interesting economic theory which cannot be rejected out of hand. But if verification does not make science, Popper needs another argument for understanding the growth of knowledge. Popper ([1935] 1968, p. 39) bases this argument on severe testing:

there is a great number – presumably an infinite number – of ‘logically possible worlds’. Yet the system called ‘empirical science’ is intended to represent only *one* world: the ‘real world’ or ‘world of our experience’ . . . But how is the system that represents our world of experience to be distinguished? The answer is: by the fact that it has been submitted to tests, and has stood up to tests.

Experience is the sieve for the abundance of logically possible worlds. The difference with induction results from a linkage of experience with falsifications: experience performs a *negative* function in inference, not the positive one of induction.

Popper’s idea that the truth of a theory cannot be proven on the basis of (affirming) observations, is not revolutionary – indeed, it basically rephrases Hume’s argument. Obviously, it was known to the logical positivist. And it had already been a common-sense notion in the statistical literature for ages (in fact, Francis Bacon had already made the

argument, as shown by Turner, 1986, p. 10). One can find this, explicitly, in the writings of Karl Pearson, Ronald Aylmer Fisher, Harold Jeffreys (see the epigraph to this chapter), Jerzy Neyman and Egon Pearson,⁶ Frederick Mills, Jan Tinbergen,⁷ Tjalling Koopmans (1937) and probably many others. They did not need philosophical consultation to gain this insight, neither did they render it a philosophical dogma according to which falsification becomes the highest virtue of a scientist.⁸ Econometricians are, in this respect, just like other scientists: they rarely aim at falsifying, but try to construct satisfactory empirical models (see Keuzenkamp and Barten, 1995). Of course, 'satisfactory' needs to be defined, and this is difficult.

Jeffreys' epigraph to this chapter can be supplemented by a remark made by the theoretical physicist Richard Feynman (1965, p. 160): 'guessing is a dumb man's job'. A machine fabricating random guesses may be constructed, consequences can be computed and compared with observations. Real science is very different: guesses are informed, sometimes resulting from theoretical paradoxes, sometimes from experience and experiment. Jeffreys argues that one may agree with Popper's insight that confirmation is not the same as proof, without having to conclude that confirmation (or verification) is useless for theory appraisal, and induction impossible.

5.1.2 The crucial test

An important example to illustrate Popper's ([1935] 1968) methodological falsificationism is Einstein's general theory of relativity, which predicts a red shift in the spectra of stars.⁹ This is the typical example of a prediction of a novel fact which can be tested. Indeed, a test was performed with a favourable result. But Paul Feyerabend (1975, p. 57, n. 9) shows that Einstein would not have changed his mind if the test had been negative. In fact, many of Popper's examples of crucial tests in physics turn out to be far more complicated when studied in detail (see Feyerabend, 1975; Lakatos, 1970; Hacking, 1983, chapter 15, agrees with Lakatos' critique on crucial tests, but criticizes Lakatos for not giving proper credit to empirical work).

For several reasons, few tests are crucial. First, there is the famous 'Duhem–Quine problem'. Second, in many cases rejection by a 'crucial test' leaves the researcher empty handed. It is, therefore, unclear what the implication of such a test should be – if any. Third, most empirical tests are probabilistic. This makes it hard to obtain decisive inferences (this will be discussed below).

The Duhem–Quine problem is that a falsification can be a falsification of anything. A theory is an interconnected web of propositions. Quine ([1953] 1961, p. 43) argues,

Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. . . . Conversely, by the same token, no statement is immune to revision. Revision even of the logical law of the excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle?

For example, rejection of homogeneity in consumer demand (see chapter 8) may cast doubt on the homogeneity proposition, but also point to problems due to aggregation, dynamic adjustment, the quality of the data and so on. In the example ‘all swans are white’ the observation of a green swan may be a falsification, but also evidence of hallucination or proof that the observer wears sunglasses. The theory of simplicity provides useful additional insights in the Duhem–Quine thesis (see chapter 5, section 3.2).

Second, what might a falsification imply? Should the theory be abandoned? Or, if two conflicting theories are tested, how should falsifications be weighted if both theories have defects? This is of particular interest in economics, as no economic theory is without anomalies. If induction is impossible, is the support for a theory irrelevant? Popper ([1963] 1989, chapter 10) tries to formulate an answer to these questions by means of the notion of verisimilitude or ‘truthlikeness’. In order to have empirical content, verisimilitude should be measurable. But this brings in induction through the back door (see chapter 3, section 5.3).

In a review of Popper’s methodology, Hausman (1988, p. 17) discusses the first and second problem. He concludes that Popper’s philosophy of science is ‘a mess, and that Popper is a very poor authority for economists interested in the philosophy of science to look to’. Hausman shows that a Popperian either has to insist on logical falsifiability, in which case there will be no science (everything will be rejected), or has to consider entire ‘test systems’ (in Popper’s vocabulary, ‘scientific systems’), in which case severe testing has little impact on the hypothesis of interest. The reason for the latter is that such a test system combines a number of basic statements and auxiliary hypotheses. If, as Popper claims, confirmation is impossible, one is unable to rely on supporting evidence from which one may infer that the auxiliary hypotheses are valid. A falsification, therefore, can be the falsification of anything in the test system: the Duhem–Quine problem strikes with full force. One should be able to rely on some rational reason for tentative acceptance of ‘background

knowledge'. Crucial tests and logical falsification are of little interest in economic inference. The complications of empirical testing, or what Popper also calls 'conventional falsification', are much more interesting, but Popper is of little help in this regard. Lakatos (1978, pp. 165–6) reaches a similar conclusion as Hausman:

By refusing to accept a 'thin' metaphysical principle of induction Popper fails to separate rationalism from irrationalism, weak light from total darkness. Without this principle Popper's 'corroborations' or 'refutations' and my 'progress' or 'degeneration' would remain mere honorific titles awarded in a pure game . . . only a positive solution of the problem of induction can save Popperian rationalism from Feyerabend's epistemological anarchism.

Lakatos' own contribution is evaluated in section 5.2.

The third problem of crucial tests, the probabilistic nature of empirical science, is of particular interest in econometrics. Popper ([1935] 1968, p. 191) notes that probability propositions ('probability estimates' in his words) are not falsifiable. Indeed, Popper (p. 146) is aware that this is an

almost insuperable objection to my methodological views. For although probability statements play such a vitally important role in empirical science, they turn out to be in principle *impervious to strict falsification*. Yet this stumbling block will become a touchstone upon which to test my theory, in order to find out what it is worth.

Popper's probability theory is discussed in chapter 3, section 5. There, I argue that it is unsatisfactory, hence the stumbling block is a painful one. One objection can already be given. In order to save methodological falsificationism, Popper proposes a methodological rule or convention for practical falsification: to regard highly improbable events as ruled out or prohibited (p. 191; see Watkins, 1984, p. 244 for support, and Howson and Urbach, 1989, p. 122 for a critique). This is known as *Cournot's rule*, after the mathematician and economist Antoine-Augustin Cournot, who considered highly improbable events as physically impossible. Cournot's rule has been defended by probability theorists such as Emile Borel and Harald Cramér (1955, p. 156), without providing a deeper justification. The rule has even been used to support the hypothesis of divine providence: life on earth is highly improbable and must be ruled out, if not for the existence of the hand of God.

The problem with the rule is where to draw the line: when does improbable turn into impossible? Popper ([1935] 1968, p. 204) argues that a methodological rule might decree that only reasonably fair samples are permitted, and that predictable or reproducible (*i.e.* systematic) devia-

tions must be ruled out. But the concept of a fair sample begs the question. If an experiment can be repeated easily (in economics, controlled experiments are rare and reproducible controlled experiments even more so), this may be a relatively minor problem, but otherwise it can lead to insoluble debates. Is a test result, with an outcome that is improbable given the hypothesis under consideration, a fluke or a straightforward rejection? If ten coins are tossed 2,048 times, each particular sequence is extremely improbable, and, by Cournot's principle, should be considered impossible.¹⁰ Cournot's rule does not provide sound guidance for the following question: What will be a falsification of the statement that most swans are white, or 'Giffen goods are rare'? This type of proposition will be discussed in the treatment of the probabilistic strategy (their empirical content, not their falsifiability, is of practical interest).

5.1.3 Critical rationalism

Caldwell (1991, p. 28) argues that, despite its drawbacks, Popper's falsificationism partly captures the spirit of economic inference. The fact, however, that economists occasionally test theories and sometimes conclude that they reject something, does not imply that methodological falsificationism is an apt description of this part of economic inference. Methodological falsificationism deals with a logical criterion to be applied to logical caricatures of scientific theories (in this sense, Popper operates in the tradition of the *Wiener Kreis*). There are neither crucial tests in economics, nor is there a strong desire for falsifications.

Caldwell continues that falsificationism may be abandoned, but that Popper's true contribution to scientific method is to be found in critical rationalism. This is a much weakened version of methodological falsificationism and merely implies that scientists should be (self-) critical. It reduces Popper's philosophy to a platitude (some authors indeed claim that Popper trivializes science, e.g. Feyerabend, 1987). Unlike methodological falsificationism, critical rationalism purports to be a historically accurate description of science. By showing that, in a number of historical cases, good science accords to critical rationalism, Popper suggests that the stronger programme of methodological falsificationism is supported. But those case studies provide weak evidence (Hacking, 1983), or even are 'myths, distortions, slanders and historical fairy tales' (Feyerabend, 1987, p. 185). This does not mean that critical rationalism is invalid; on the contrary, it is a principle which has been advocated by a wide range of scientists, before and after the publication of Popper's views on methodology.

For example, the motto of Karl Pearson ([1892] 1911) is a statement due to Victor Cousin: '*La critique est la vie de la science*' (i.e. criticism is

the life of science; see also Pearson, p. 31, among many other places where he emphasizes the importance of criticism). More or less simultaneously with Pearson, Peirce introduced the theory of fallibilism, of which falsificationism is a special case (see e.g. Popper, [1963] 1989, p. 228). According to fallibilism, research is stimulated by a state of unease concerning current knowledge. Research intends to remove this state of unease by finding answers to scientific puzzles. In Humean vein, Peirce (1955, p. 356) argues, ‘our knowledge is never absolute but always swims, as it were, in a continuum of uncertainty and of indeterminacy’. Methodological falsificationism is different from Peirce’s philosophy, by actually *longing* for a state of unease rather than attempting to remove it.

5.1.4 *Historicism*

Although Popper’s philosophy of science is in many respects problematic, in particular if applied to the social sciences, he has contributed an important insight to the social sciences that is usually neglected in discussions of (economic) methodology. This is his critique of ‘historicism’. Historicism, Popper ([1957] 1961) argues, starts from the idea that there is such a thing as a historical necessity of events, or social predestination. He considers Marx and Hegel as examples of historicists (although they had a very different interpretation of historicism and Popper’s interpretation of their work is controversial).¹¹ Disregarding the historicist merits of Marx and Hegel, this may be Popper’s most interesting book for economists.

Popper’s critique of historicism is related to his idea that knowledge is conjectural and universal theories cannot be proven. This does not preclude growth of knowledge, but this growth results from trial and error. It cannot be predicted. The ability to predict the future course of society is limited as it depends on the growth of knowledge.

One of the most interesting parts in the *Poverty of Historicism* is his discussion of the so-called *Oedipus Effect*. Popper (p. 13; see also Popper, 1948) defines it as:

the influence of the prediction upon the predicted event (or, more generally, ... the influence of an item of information upon the situation to which the information refers), whether this influence tends to bring about the predicted event, or whether it tends to prevent it.

This problem of reflexivity, as it is also known, undermines methodological falsificationism. Popper’s views on historicism and the special characteristics of the social sciences suggest that falsificationism cannot play quite the same role in the social sciences as in the natural sciences. Popper

([1957] 1961, pp. 130–43) is ambiguous on this point. He remains convinced of the general importance of his methodology, the unity of method. Meanwhile, he recognizes the difference between physics and economics:

In physics, for example the parameters of our equations can, in principle, be reduced to a small number of natural constants – a reduction that has been carried out in many important cases. This is not so in economics; here the parameters are themselves in the most important cases quickly changing variables. This clearly reduces the significance, interpretability, and testability of our measurements. (p. 143)

Popper swings between the strong imperatives of his own methodology, and the more reserved opinions of his former colleagues at the London School of Economics (LSE), Lionel Robbins and Friedrich Hayek. Popper's compromise consists of restricting the domain of the social sciences to an inquiry for conditional trends or even singular events, given an *a priori* axiom of full rationality.¹² Popper calls this the zero method. Methodological falsificationism may be applied to this construct, but the rationality postulate is exempted from critical scrutiny. It is an *a priori* synthetic truth (see also Caldwell, 1991, pp. 19–21).

The initial conditions may change and this may invalidate the continuation of the trend. Hence, social scientists should be particularly interested in an analysis of initial conditions (or situational logic). The difference between prognosis and prophecy is that prophecies are unconditional, as opposed to conditional scientific predictions (Popper, [1957] 1961, p. 128). In the social sciences, conditions may change due to the unintended consequence of human behaviour (this is a cornerstone of Austrian economic thought). Mechanic induction like extrapolation of trends is, therefore, not a very reliable way of making forecasts. The econometric implications of the Oedipus Effect and the lack of natural constants in economics deserve more attention. These implications are of greater interest to the economist or econometrician than the methodology of falsificationism or Popper's ideas on probability.

5.2 Lakatos and conjecturalism

What does Lakatos offer to rescue the conjecturalist strategy? Because of the problems involved with methodological falsificationism, he proposes to give falsifications less impact. He rejects the crucial test or 'instant falsification', and instead emphasizes the dynamics of theory development.

5.2.1 Research programmes

This dynamics can be evaluated by considering a theory as a part of an ongoing *research programme*. A theory is just one instance of a research programme, RP, at a given point in time. How do you decide whether a succeeding theory still belongs to an RP? This question should be settled by defining the essential characteristics of an RP, by its *hard core* and the guidelines for research, the *heuristic*. The hard core consists of the indisputable elements of a research programme. The positive heuristic provides the guidelines along which research should proceed. The negative heuristic of a research programme forbids directing the *modus tollens* at the hard core (Lakatos, 1978, p. 48).

According to Lakatos (p. 48), the hard core of an RP is irrefutable by the methodological decision of its proponents. A falsification of the theory is not automatically a rejection of an RP. Falsifying a theory is replaced by measuring the degree of progressiveness of an RP. Lakatos (p. 33) distinguishes three kinds of progress. A research programme is

- *theoretically* progressive if 'each new theory has excess empirical content over its predecessor, that is, if it predicts some novel, hitherto unexpected fact'
- *empirically* progressive if some of these predictions are confirmed
- *heuristically* progressive if it avoids auxiliary hypotheses that are not in the spirit of the heuristic of a research programme (Lakatos would call such hypotheses *ad hoc*₃, where *ad hoc*₁ and *ad hoc*₂ denote lack of theoretical and empirical progressiveness, respectively).

It is not easy in economics to apply Lakatos' suggestion of appraising theories by comparing their rate of progressiveness or degeneration. A research programme is a vague notion. Scientists may disagree about what belongs to a specific RP and what does not (see Feyerabend, 1975). The problem culminates in the so-called tacking paradox (see Lakatos, 1978, p. 46). If the theory of diminishing marginal utility of money is replaced by a successor, which combines this theory with the general theory of relativity, an apparently progressive step is being made. Of course, this is not what Lakatos has in mind: this is why he emphasizes the consistency with the positive heuristic of an RP. An alternative for avoiding nonsensical combination of two theories is to introduce the notion of *irrelevant conjunction* (see Rosenkrantz, 1983, for a discussion and a Bayesian solution to the problem).

Lakatos' suggestions are of some help for understanding ('rationally reconstructing') economic inference, but in many cases they are too vague and insufficiently operational. Chapter 8 provides a case study showing how difficult it is to apply them to an important episode in the history of applied econometrics: testing homogeneity of consumer demand.

Competing research programmes may apply to partly non-overlapping areas of interest. This leads to problems such as the already mentioned Duhem–Quine problem, and *incommensurability*, Thomas Kuhn’s (1962) notion that new theories yield new interpretations of events, and even of the language describing these events. Furthermore, whereas Popper still made an (unsuccessful) attempt to contribute to the theory of probabilistic inference (cf. his propensity theory of probability, and the notion of verisimilitude), Lakatos has a critical attitude towards this subject.

5.2.2 *Growth and garbage*

Both Popper and Lakatos attack inductivism, but Lakatos is more radical in his critique of empirical tests. These tests are a key element in Popper’s epistemology, but not so in Lakatos’. This is clear from the *Methodology of Scientific Research Programmes*, in which falsifications are less crucial than in Popper’s work. Instead of falsifiability and crucial tests, Lakatos advocates a requirement of growth. Can statistical methods help to obtain knowledge about the degree of empirical progress of economic theories or research programmes? Lakatos’ scant remarks on this issue provide little hope (all quoted from Lakatos, 1970, p. 176). To start, the requirement of continuous growth

hits patched-up, unimaginative series of pedestrian ‘empirical’ adjustments which are so frequent, for instance, in modern social psychology. Such adjustments may, with the help of so-called ‘statistical techniques’, make some ‘novel’ predictions and may even conjure up some irrelevant grains of truth in them. But this theorizing has no unifying idea, no heuristic power, no continuity.

These uncharitable statements are followed by such terms as *worthless*, *phony corroborations*, and, finally, *pseudo-intellectual garbage*. Lakatos concludes:

Thus the methodology of research programmes might help us in devising laws for stemming this intellectual pollution which may destroy our cultural environment even earlier than industrial and traffic pollution destroys our physical environment.

Whoever is looking for a Lakatosian theory of testing will have a hard job. Lakatos’ approach is nearly anti-empirical. In his case studies of physics, experimenters are repeatedly ‘taught lessons’ by theoreticians; bold conjectures are made despite seemingly conflicting empirical evidence, and so on (see Hacking, 1983, chapter 15, for a devastating critique of Lakatos’ account of many of these experiments).

Lakatos’ approach is of little help for economists and, as I will argue later on (chapter 8), it does not provide a basis for a useful econometric

methodology. If testing and falsifying of all propositions are approached rigorously, not much of economics (or any other science) will be left. On the other hand, if induction is rejected and severe testing as well, one ends up with ‘anything goes’ or ‘anarchism in disguise’ (Feyerabend, 1975, p. 200). Watkins (1984, p. 159), the saviour of methodological falsificationism, agrees with this characterization of Feyerabend and makes the following *reductio ad absurdum*:

If you could tell, which you normally cannot, that Research Program 2 is doing better than Research Program 1, then you may reject Research Program 1, or, if you prefer, continue to accept Research Program 1.

What is missing in the conjecturalist strategy is a clear view on the utility of theories, the economy of scientific research (emphasized by Peirce) and the positive role of measurement and evidence. A theory of inductive reasoning remains indispensable for understanding science.

6 Probabilism

The probabilistic strategy may offer garbage, if applied badly (as Lakatos suggests) but it may also yield insights in scientific inference: in its foundations and in appraising econometric applications.

This brings us to the probabilist strategy, which claims that Hume wants too much if he requires a proof for the truth of inductive inferences, and amends proposition (iii) of section 2 above. A logic of ‘partial entailment’ is proposed: probability logic. This strategy has been investigated by John Maynard Keynes, Harold Jeffreys, Hans Reichenbach, Rudolf Carnap and many others. Alternatively, the problem of Humean scepticism may be resolved by providing a probabilistic underpinning of the principle of the uniformity of nature, which has been investigated with the Law of Large Numbers. This approach has been taken by Richard von Mises.

The following three chapters discuss these and other versions of a probabilistic strategy for scientific inference. The probabilistic strategy deserves special attention because it can serve as a foundation for econometric inference, the topic of this book. Furthermore, uncertainty is particularly important in economics. If falsifying economic theories is feasible at all, then it must be probabilistic falsification. The issue of probabilistic *testing* can only be well understood if a good account of probabilism is presented (note that testing is clearly neither the only nor the ultimate aim of probabilism). If probabilistic falsification is impossible, then other methods are needed for appraising economic theories given the uncertainty that is inextricably bound up with economics.

Notes

1. There are some excellent introductions to the philosophy of science, written for and by economists. Blaug (1980) and Caldwell (1982) are particularly strong on methodology, but both have little to say about econometrics. Darnell and Evans (1990) combine methodology and econometrics, but their treatment is brief and sometimes ambiguous. More recent developments in methodology can be found in Backhouse (1994).
2. The phrase is due to C. D. Broad (see Ramsey, 1926, p. 99; and Hacking, 1975, p. 31). Frank Ramsey (1926, pp. 98–9) denies that there is no answer to Hume's problem, but 'Hume showed that it [i.e. inductive inference] could not be reduced to deductive inference or justified by formal logic. So far as it goes his demonstration seems to be final; and the suggestion of Mr Keynes that it can be got round by regarding induction as a form of probable inference cannot in my view be maintained. But to suppose that the situation which results from this is a scandal to philosophy is, I think, a mistake.'
3. The problem is not trivial. Keynes ([1921] CW VIII, p. 418) notes how Laplace calculated that, 'account be taken of the experience of the human race, the probability of the sun's rising tomorrow is 1,826,214 to 1, this large number may seem in a kind of way to represent our state of mind of the matter. But an ingenious German, Professor Bobek, has pushed the argument a degree further, and proves by means of these same principles that the probability of the sun's rising every day for the next 4000 years, is not more, approximately, than two-thirds, — a result less dear to our natural prejudices.' See also Pearson ([1892] 1911, p. 141) for a discussion of Laplace and the probability of sunrise.
4. A proposition is analytic if the predicate is included in the subject (e.g. all econometricians are human). A synthetic proposition is not analytic (usually thought to be based on matters of fact; e.g. all econometricians are wise). Quine ([1953] 1961) contains a classic critique of the distinction.
5. The *Weiner Kreis* was the influential group of scientists who met regularly in Vienna during the 1920s and 1930s. In 1922 the group was formed by Moritz Schlick. Some prominent members were Rudolph Carnap, Otto Neurath and Hans Hahn. Other regular participants were Herbert Feigl, Philipp Frank, Kurt Gödel, Friedrich Waisman and Richard von Mises. The group built on the positivist doctrines of Henri Poincaré, and in particular the Austrian physicist and philosopher Ernst Mach. In addition, they used advances in logic due to Gottlob Frege, to Russell and Whitehead, whence logical positivism. A closely related view is logical empiricism, associated with Hans Reichenbach and, again, Carnap. Caldwell (1982) defines logical empiricism as the mature version of logical positivism. The different branches of twentieth-century positivism are often grouped as neo-positivism. Popper is occasionally associated with neo-positivism, but Hacking (1983, p. 43) argues convincingly that Popper does not qualify as a positivist.
6. Although with a twist in their case, as they are interested in behaviour rather than inference.

7. Tinbergen (1939a, p. 12) argues that econometrics cannot prove a theory right, but it may show that some theories are not supported by the data. He never refers to Popper, or any other philosopher (Popper, [1957] 1961, on the other hand, contains a reference to Tinbergen who notes that constructing a model is a matter of trial and error). Koopmans makes a remark similar to Tinbergen's. If researchers still speak of verification of theories, then this should not be taken literally: very few would deny Hume's argument.
8. According to Mark Blaug, an important difference between Popper and those statisticians is Popper's banning of immunizing stratagems. Whether an absolute ban would benefit science is doubtful. See Keuzenkamp and McAleer (1995) for a discussion of 'ad hocness' and inference, in particular the references to Jeffreys.
9. Popper regards Einstein as the best example of a scientist who took falsification seriously, because of his bold predictions. Einstein has the highest number of entries in the name index of Popper ([1935] 1968) (thirty-six times). Carnap (with a score of thirty-five) is a close second, but does not share in Popper's admiration.
10. The example is not imaginary. W. Stanley Jevons performed this experiment to 'test' Bernoulli's law of large numbers. This and more laborious though equally meaningless experiments are reported in Keynes ([1921] CW VIII, pp. 394–9).
11. Hegel, for example, believed in the historical relativity of truth, a tenable position from the point of view presented in this book.
12. '[A]nd perhaps also on the assumption of the possession of complete information' (Popper, [1957] 1961, p. 141) – a remarkably un-Austrian assumption!