Discussion Papers
Department of Economics
University of Copenhagen

No. 12-16

Experiments, Passive Observation and Scenario Analysis: Trygve Haavelmo and the Cointegrated Vector Autoregression

Kevin Hoover and Katarina Juselius

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Experiments, Passive Observation and Scenario Analysis: Trygve Haavelmo and the Cointegrated Vector Autoregression

Kevin Hoover, Department of Economics and Department of Philosophy, Duke University, and
Katarina Juselius, Department of Economics, University of Copenhagen

Abstract

The paper provides a careful, analytical account of Trygve Haavelmo's unsystematic, but important, use of the analogy between controlled experiments common in the natural sciences and econometric techniques. The experimental analogy forms the linchpin of the methodology for passive observation that he develops in his famous monograph, The Probability Approach in Econometrics (1944). We show how, once the details of the analogy are systematically understood, the experimental analogy can be used to shed light on theory-consistent cointegrated vector autoregression (CVAR) scenario analysis. CVAR scenario analysis can be seen as a clear example of Haavelmo's "experimental" approach; and, in turn, it can be shown to extend and develop Haavelmo's methodology and to address issues that Haavelmo regarded as unresolved.

Keywords: Trygve Haavelmo, experiments, passive observation, CVAR, scenario analysis, probability approach, econometrics

JEL Codes: C30, C31, B41, B31, C50

1 The Problem of Passive Observation

Perhaps the two most vital contributions of Trygve Haavelmo's path-breaking, magisterial monograph, *The Probability Approach in Econometrics* (1944), was to recast the application of the theory of probability to economic data in order to bring economics models into the scope of formal statistical analysis (see Juselius 2012) and to develop a methodology of passive observation (including the contextualization of techniques such as multivariate regression and the identification of simultaneous systems of equations) as an analogue to controlled experiments. In Haavelmo's formulation, Nature is cast in the role of the experimenter, and the econometrician uses a model to define a perspective or "point of view" in which the formulation of the model and the statistical techniques applied to it play the role of experimental controls (Haavelmo 1944, pp. 1, 6, 14-15, 51).[1]

It is hard to miss the weight that Haavelmo places on the notion of experimentation: variants on the root "experiment" occurs fifty-four times throughout the monograph. In the first forty or so pages, most of the references are in variants on the construction "design of experiment." Yet for all this attention, Haavelmo's discussion is unsystematic, and a careful account of the role that the simile of experimentation plays in *The Probability Approach* should be illuminating.[2] We argue that, when systematically laid out, Haavelmo's understanding of passive observation as analogous to experimentation is rich and nuanced and that it provides a deeper view of his methodological vision than available hitherto, as well as a framework for understanding how some modern practices neglect issues that he found vital. It also allows us to cast other modern practices – especially scenario analysis in the context of the cointegrated vector autoregression (CVAR) as closely related to Haavelmo's experimental methodology (Juselius 2006, 2012). In taking this view, we reject the idea that Haavelmo is a strong apriorist who provides no account of how economists learn from theories as opposed to merely testing them (cf. Heckman 1992, 2000, and Eichenbaum 1995).

[1]Morgan 1990, ch. 8, provides a now standard account of these developments in the history of econometrics; see also Hendry, Spanos, and Ericsson (1989) for a précis of the main achievements of Haavelmo's *Probability Approach*.

[2]Boumans (2005, 2012) has previously treated related aspects of Haavelmo's methodology.

2 Ideal Experiments

Controlled experiments are a well-known method of isolating and measuring causal processes. The logic of the controlled experiment is described by John Stuart Mill's *method of difference:* "If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon" (Mill 1843[1973], p. 391). While Mill leaves the causal direction undetermined, when the experimenter can select the *controls* (the circumstances in common) and the *intervention* (i.e., the one difference) directly, rather than passively observe them, even the causal direction is determined. Controlled experiments can be extended in some cases to determine not only the existence of a cause but its quantitative measure by applying the method of difference jointly with Mill's *method of concomitant variation*: "Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation" (Mill 1843[1973], p. 401).

Ideally, control in the controlled experiment is comprehensive. Every relevant factor is set to determinate levels and interventions are independent of one another and of the outcome variable. Some of the virtues of the randomized controlled trial – often called the "gold standard" of statistical investigation – arise from the fact that randomization helps to secure both sorts of independence.

Haavelmo (1954, p. 2) refers to structural relations as "any economic relation associated with, and valid for, a specified economic structure that could conceivably be reproduced experimentally." Thus, the purposes of engaging in experimentation are closely related to our interest in structural relations:

> The study of structural relations may serve at least these three purposes:
>
> 1. To satisfy scientific curiosity.
>
> 2. To study the functioning of alternative structures that could have practical interest from the point of view of economic *reform.*

3. To explain current events in the actual economic structure under which an economy is at present operating. [Haavelmo 1954, p. 2]

The first is related to the notion that the purpose of an experiment is causal articulation – to learn which causes are operative, to map their interconnections, and to quantify their strengths. Haavelmo (1954, p. 3) goes on to point out that our understanding of more complicated economic structures derives from "piecing together relations derived from the consideration of relatively simple partial experimental designs, or structures." The second is related to the use of economic models in conducting counterfactual experiments, which frequently form the basis of policy advice. And the third is related to prediction and *ex post* analysis of the actual paths of the economy.

Experiments require an experimenter. In economics experimentation is sometimes feasible, but very often not, so that inference must rely on passive observation (Haavelmo 1944, p. 7). For Haavelmo, a key idea is that Nature may be thought of as an experimenter: "Nature is steadily turning out [experiments] from her own enormous laboratory, . . . which we merely watch as passive observers" (Haavelmo 1944, p. 14; see also pp. 9, 16).

The distinction between truly controlled experimentation and passive observation is this: In a truly controlled trial, we may think of the outcome y as a function of the intervention variable x *ceteris paribus:* $y = f(x)$, with other relevant factors set a fixed levels, In the case of passive observation, we may think of the other relevant factors as observed and brought into the function, so that they too may vary: $y = f(x, a, b, c, ...)$. In the case of passive observation, "control" is not so much as matter of setting particular values or literal *ceteris paribus* as it is accounting for the variation in relevant confounding factors. Such accounting is the natural domain of statistical analysis in which literal control can often be replaced by conditioning on other variables.

Although Haavelmo typically refers to Neyman and Pearson as the source of his statistical approach, the application of probability and statistics to controlled experiments was already well developed in the 1930s, particularly in the work of R.A. Fisher, including *The Design of Experiments* (1935). Haavelmo's achievement is, in part, an adaptation of techniques to suit the special problems that economics raises for imposing statistical controls. How can economic models be formulated so that statistical methods would be adequate substitutes for literal experimental control? : "we try to take care of the *ceteris paribus* conditions ourselves, by statistical devices of clearing the data from influences not taken account of in the theory (e.g., by multiple correlation anslysis)" (Haavelmo 1944, p. 17; see also pp. 17-19).

The range of natural variation of the variables imposes an important limitation on passive observation in comparison to active experimentation. The difficulty is that a variable that may be causally important may nonetheless display no actual variation. That is no problem when a model is used to simulate the actual behavior of variables *ceteris paribus* the constant variables but it poses a much starker problem when trying to recover the causal structure of the model from observable data or when the model is used for counterfactual analysis that involves interventions on hitherto constant variables.[3]

Comprehensive control is an exacting standard – even for true controlled experiments – and Haavelmo (1944, pp. 49-50) proposes that some of the causally relevant factors may be unobservable and/or indeterminable which provides the rationale for stochastic models. Haavelmo (1944, p.48) does not make any strong claims for any particular interpretation of probability. Its basis may be ontic (i.e., probability is true feature of reality) or epistemic (i.e., it is a reflection of our ignorance), but Haavelmo (1944, pp. 51-52) observes that if it has a variety of independent sources of variation, a random variable will conform to well-defined distributions, such as the normal, and can be treated by the tools of probability theory and statistics. In introducing stochastic models, it may appear that we have left the domain of *ideal* experiments. But this is not the case. We do not give up on a complete characterization of the relevant variables; rather we characterize their behavior under broader but still comprehensive categories (Haavelmo 1944, p. 2). The move to stochastic models does not depart from the ideal unless a more specific, nonstochastic characterization is actually available.

Stochastic models connect interventions and controls with outcomes and rely on the analogy to controlled experiments in just the same way as nonstochastic models, except that some variables must be generated as random draws from a probability distribution. If such draws are repeated – e.g., in the manner of Monte Carlo experiments – then a probability distribution for all the variables of a model may be built up.

[3]Frisch refers to problem of the recovery of the underlying structure – a problem related to, but not identical with, the identification problem, as the *inversion problem* (Frisch 1926, p. 86 cited by Louçã 2007, p. 11; see also Frisch 1939 and Louçã 2007, p. 95, *passim*). The pitfalls of such counterfactual analysis were addressed directly in Marschak (1953) and have become a mainstay of macroeconometric thinking since acquiring the name "noninvariance" or "Lucas critique" (Lucas 1976).

For a stochastic model that involves time-dated variables, Haavelmo (1944, pp. 48-49) famously characterizes the realized time series as a single draw from n-dimensional distribution rather than n-draws from a one-dimensional distribution. In many cases, the partition of the model into deterministic and stochastic elements can be seen equivalently as n-draws from a *conditional* probability distribution of the random terms in which the conditioning variables are updated for each draw. (The details of Haavelmo's effort to bring probability to bear on economic time-series is addressed more fully in Juselius 2012).

3 The Design of Measurement Experiments

We are used to thinking of ("crucial") experiments as tests of the truth of a theory or as a means of identifying causes, but many experiments presuppose the broad truth and causal articulation of the underlying theory and seek only to quantify an unknown value (Haavelmo 1944, p. 14). Haavelmo's common phrase "the design of experiment" is deployed most often in just this latter context of quantification. The distinction between variables and parameters is, for Haavelmo (1944, p. 3) a relative one. Variables typically refer to the objects of investigation; parameters are introduced by the analyst; both can be quantified using experimental methods. Thus, Haavelmo (1944, p. 1) sees the most basic act of observation as an experiment. For example, to convert a "formal mathematical scheme" such as the analysis of choice as the interaction of indifference surfaces with budget constraints into economics, we must design the

> experiments that would indicate, first, what real phenomena are to be identified with the theoretical prices, quantities, and income; second, what is to be meant by an "individual"; and, third, how we should arrange to observe the individual actually making his choice . . . In the verbal description of his model in "economic terms," the economist usually suggests, explicitly or implicitly, some type of experiments or controlled measurements designed to obtain the real variables for which he thinks that his model would hold. [Haavelmo 1944, pp. 6-7]

Later Haavelmo (1954, p. 2) glosses the notion of experimental design in the context of obtaining values for variables as "all the things one would have to write in an instruction to even the most intelligent assistant observer in order to communicate . . . a desired procedure of collecting appropriate data."

Experiments used to obtain values for variables can be seen as measuring devices in which an *a priori* experimental design is applied to the world (cf. Hoover 1994 on the notion of "econometrics as measurement"). *A priori* in this case does not mean non-empirical nor unrelated to the acquisition of knowledge in the past, but only that the theoretical model, the design of the measuring experiment, when applied to the real world, provides a maintained *perspective* or "point of view" – a "classification of real phenomena[,] . . . viewing reality through the framework of some scheme" (Haavelmo 1944, p. 1; see also p. 11):

> 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. [Haavelmo, p. 9]

Of course, a central point about perspectives is that they can be different, that one may view reality from alternative, but no less correct, points of view.[4]

Haavelmo does not provide a really good concrete example of multiple, equally acceptable perspectives; yet he does give an important general characterization in the context of parameter estimation. We need not, he argues, work with a probability distribution (i.e., a perspective) in which a desired parameter, say, α appears directly: "In general, any kind of data following a probability distribution which depends in a known way upon α, may serve as a means of estimating α, provided that the method of estimation is based on the appropriate stochastic specification of the data . . ." (Haavelmo 1954, p. 5). While these are not Haavelmo's own examples, we can easily illustrate the point by considering the equivalence of models transformed from levels to levels and differences or from real terms to nominal terms (see section 6 below).

[4] Hoover (2012a) develops the idea of a *perspectival realism* such as Haavelmo suggests here.

4 Measurement, Observation, and Testing

In an ideal case, Haavelmo conceives of the testing of a theoretical model as simply conducting a measurement. If the relationships and values of the theoretical model are completely matched to the relationships and values obtained by using the same design of an experiment but applying it to the real world, then one would judge the model to be true: "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" (Haavelmo 1944, p. 9). Haavelmo would regard the simulation of a theoretical model that perfectly matched the observed behavior of the variables in the world as sufficient evidence for its truth. The issue is only slightly more complicated in this case that there are multiple valid modeling perspectives on the data (see section 3 above). In that case, so long as we can derive the probability distribution of our theoretical model from the one that actually characterizes the observed relationships, we may judge the theoretical model to be true.

Theory may be incomplete or have only qualitative implications. In one type of case, theory may determine the values of variables or parameters only loosely or within a range. Theory may not, for example, assign a parameter α a particular value, but perhaps only a range – say, $\alpha > 0$. A successful test, then, measures the analogue to α in the world to be within that range. In a second type of case, theory may imply only a broad property such as a homogeneity restriction (e.g., the neutrality of money) or the selected characteristic of a probability distribution. Then, a class of theoretical models, rather than a unique theoretical model, would be supported by their consistency with the observed or measured relationships (Haavelmo 1944, pp. 82-83).

In all of these cases, the logic that relates theory to the world through the design of an experiment expressed in a theoretical model is the same as that for observation or the collection of data discussed in section 3. The theoretical model implies a set of procedures that applied in ideal circumstances would elicit information from the world. If the empirical model displays verisimilitude with respect to the world observed according to the experimental design implied by the theoretical model, then the theoretical model is supported. Both observation of particular variables and testing of the model follow Haavelmo's template for measurement, although the one seeks to find a value, the other a relationship: "The essential feature of . . . a rule of measurement is that it does not a priori impose [the theoretical restriction at stake] upon the variables to be measured" (Haavelmo 1944, p. 13). A genuine measurement or a genuine test might impose a particular theoretical perspective on the data, but it must leave open alternative possibilities within that perspective.[5]

The distinction between testing and observation is thus one of degree, not of kind. When using an experimental design to measure the value of a variable or parameter (or even a relationship), there may be no degree of freedom with respect to the maintained theoretical assumptions, even though different values are permissible. The measurement would not be checkable against any other standard than its conforming to the *a priori* design of an observational experiment. Such a measurement cannot constitute a test of maintained assumptions, since it builds them in from the beginning. Only by embedding those assumptions in a broader maintained or *a priori* framework that is consistent with their either holding or not holding could one construct a genuine test.

What is a stake may be illustrated by an example that is not Haavelmo's. The United States Congressional Budget Office (1995) publishes estimates of the natural rate of unemployment. These estimates are backed out of a particular specification of an expectations-augmented Phillips curve of a quite specific specification. They, thus, presuppose the experimental design and do not provide a test of it. The Phillips curve is used as a measuring instrument; and, while it may supply data that are used to test other hypotheses, it is not directly involved in a test of what it estimates.[6]

Both tests and observations are kinds of measurements in Haavelmo's scheme. An observed variable or relationship is defined by the application of an experimental design to the real world. In contrast, the test is defined by the application of the experimental design of the theoretical model to the data generated by observation. A genuine test is possible only if the design of the test experiment is independent of the design of the observational experiment – that is, a genuine test requires that there is nothing in the design of the observational experiment that guarantees its concordance with the theoretical model.

---

[5] Haavelmo (1944, p. 66; cf. p. 10) sees the relationship between measurement and testing as so close that he refers to estimation as "a particular form of testing hypotheses."

[6] See Boumans (2005, ch. 5, appendix) for a useful discussion and another example. Duarte and Hoover (2012) provide another relevant example in the use of economic models to measure shocks.

A further consequence of Haavelmo's view is that observations are relative to a point of view – that is, at one level or another observations build in a theoretical perspective and none is a completely raw or free-floating fact. The real world constrains us, to be sure, but what we see in the real world depends, in part, on how we look at it.

5 Design of Observational Experiments

Testing for Haavelmo is conceived of as a matter of matching theoretical models to observations, which are themselves also the product of experimental design. In each case, our object is to provide a model of an experiment:

> The idea behind this is, one could say, that Nature has a way of selecting joint value-systems of the "true" variables such that these systems are as if the selection had been made by the rule defining our theoretical model. [Haavelmo 1944, p. 9]

But the observations must not be constructed in such a way that they necessarily confirm the theoretical model. Stochastic models undermine the strategy to the degree that any apparent mismatch between theory and observation may be explained as a rare event rather than as a failure of the model under test (Haavemo 1944, p. 2). Although, the details are beyond our current scope, Haavelmo's (1944, ch. 4) discussion of Neyman and Pearson's framework for statistical hypothesis testing aims to supply practicable standards for drawing a pragmatic distinction between events that support a match between theoretical model and observation and events that do not support such a match.

Haavelmo's account of the ideal test provides too "clean" a view of econometric methodology. It would be easy to read it as, *a priori* theory proposes and observation disposes, which in turn supports a caricature: the economic theorist, working in isolation, passes his hypotheses to the econometrician who "[a]rmed with an array of tools ... goes about his grim task – testing and rejecting models," reporting "yes" or "no" but not otherwise interacting with the theorist (Eichenbaum 1995, p. 1619). The vision of the econometrician as the as grim executioner of models that fail to fit reality, gives him a role as a gatekeeper, but not as a constructive contributor to the accumulation of empirical knowledge (cf. Heckman 1992, pp. 883-884; 2000, pp. 86-87). The caricature misses Haavelmo's actual vision of econometric practice in two related respects.

First, as already noted in section 3, *a priori* for Haavelmo does not imply considerations independent of all empirical considerations, experience, and so forth:

> It is almost impossible, it seems, to describe exactly how a scientist goes about constructing a model. It is a creative process, an art, operating with rationalized notions of some real phenomena and of the mechanism by which they are produced. The whole idea of such models rests upon a belief, already backed by a vast amount of experience in many fields, in the existence of certain elements of invariance in a relation between real phenomena, provided we succeed in bringing together the right ones. [Haavelmo 1944, p. 10]

Rather *a priori* refers to a maintained perspective or point of view that is a) independent of the experimental design for measuring the observations on which it will be tested and b) allows no feedback from the current observations to the current experimental design. Feedback from observational results to new experimental designs, however, is an essential element in the growth of knowledge. Haavelmo sees the requirement of an *a priori* experimental design as a condition of interpretability. So, for example, it is only within a framework in which the space of admissible hypotheses is set out in advance and held constant that the notions of size and power (or type I and type II error) have precise, quantifiable counterparts. Apriorism of this sort is part of his general view that knowledge is perspectival: we can understand – or even properly observe – empirical reality only through the theoretical framework of a well-defined experimental design (cf. Hoover 2012a). Such apriorism does not rule out learning from the data: "It is clearly irrelevant how we happen to choose the hypothesis to be tested . . . In particular, the hypothesis might be one that suggests itself *by inspection of the data*" (Haavelmo 1944, p. 83).

The second way in which the caricature of Haavelmo's econometrician as the grim executioner of models misses his actual methodology builds on the last point. Haavelmo sees economic knowledge as advancing through the interplay of theory and observation. He quotes approvingly Bertrand Russell (1927, p. 194; see Haavelmo 1944, p. 14): "The actual procedure of science consists of an alternation of observation, hypothesis, experiment, and theory." He himself writes: "In scientific research – in the field of economics as well as in other fields – our search for 'explanations' consists of digging down to more fundamental relations than those that appear before us when we merely 'stand and look' " [Haavelmo 1944, p. 38]. Even in the context of formal Neyman-Pearson statistical tests, Haavelmo observes that the class of admissible hypotheses might be incomplete and that examining the power of tests against hypotheses *outside* the admissible class might be enlightening. Haavelmo's vision is one in which the outcomes of one experiment shapes new questions, new perspectives, and new experimental designs.

The upshot is that rather than an overly clean view of the nature of empirical investigation, Haavelmo takes a messy view – even in the case of real-world controlled experiments, but especially for the case of passive observation. With experimentation as the governing simile, much of *The Probability Approach* is devoted to understanding the nature of the empirical mess and to proposing workable strategies for managing it.

Haavelmo distinguishes three types of variables – where "variable" also comprises parameters and relationships – each defined in relationship to the experiments in which it participates. In effect, Haavelmo sees two worlds – the world of theory and the world of reality. The world of theory is the home of precisely defined concepts and relationships, a world undisturbed by unknown or unaccounted for factors, a world of ideal experimentation. This world is populated by *theoretical variables*. An experiment in this world amounts to evaluating a counterfactual claim: "the most interesting [structural relations] are those for which the associated design of experiment consists in fixing a set of datum-parameters or 'independent variables,' the 'outcome' of the experiment being the choice of a particular value of some dependent variable" (Haavelmo 1954, p. 3). In the context of a formal theoretical model, such outcomes might be deduced mathematically or simulated – in the latter case, the experimentation is literal though it takes place in an artificial world. Ideal testing, as we saw in section 3, amounts to establishing the perfect (or, at least, adequate match between) the theoretical model and the measured behavior of variables in the world.

The world of reality lacks the crisp characterization of the artificial world of theory. In Haavelmo's account, it is populated by two types of variables:

"true" and "observational" variables (Haavelmo 1944, p. 5).[7] The *true variables* are those that correspond perfectly to the experimental design for the *measurement* of a variable.

> But the theoretical variables are not defined as identical with some "true" variables. For the process of correct measurement is, essentially, applied to each variable separately. To impose some functional relationship upon the variables means going much further. We may express the difference by saying that the "true" variables (or time functions) represent our ideal as to accurate measurements of reality "as it is in fact," while the variables defined in a theory are the true measurements that we should make if reality were actually in accordance with our theoretical model. [Haavelmo 1944, p. 5]

Haavelmo's point is this: ideally, the values of true variables would correspond conceptually to the theoretical variables, but the relationships in which they stand one with another are not built into the experimental design through which they are measured. Rather whether true variables correspond to theoretical variables is the central question of testing, and it must remain an open question in the design of an empirical model to be settled only with regard to the data; otherwise, no genuine test is possible.

The *observational variables* are the variables that are actually collected. Haavelmo's distinction is vital; for he proposes that the testing relationship is a relationship between theoretical and true variables. If observational variables are to be meaningful, they must correspond more or less closely to true variables. The failure of such correspondence poses a serious methodological challenge. The researcher must either assure such correspondence or find a way to compensate for it absence.

At least three questions can be raised with respect to any variable: First, *what is it*? The question has two senses. What is its value? And what is it conceptually? This is the central question in the comparison for real-world variables. Does the observational variable capture the concept aimed at for the true variable? For various practical reasons, it may not.

[7]Haavelmo is fond of scare quotes. Partly, this is a matter of personal style, but the fact that they are most consistently used with reference to epistemological concepts, such as truth and reality possibly reflects a fear of committing himself in fraught philosophical debates (see Hoover 2012a). Not sharing his diffidence, we shall not adopt his practice.

Second, *how does it behave*? Theoretical variables behave according to the rules embedded in theoretical models. True and observational variables behave according to the rules governing reality. But many features of reality may serve to obscure or undercut the relationship between true and observational variables.

Third, *what does it mean*? For Haavelmo (1944, p. 6) meaning is a matter of tested theory. Theoretical models taken as formal systems have no meaning. The models and their variables gain their meaning through being applied to the world, through the relationships among the theoretical variables corresponding to those among the true variables. But this is not a one-way street. Haavelmo (1944, p. 12) suggests that the natural history of most sciences begins with ill-formulated metaphysical theories and are made increasingly complex to deal with the challenges posed by observational data. At some point, "clearing work is needed, and the key to such clearing is found in *a priori* reasoning, leading to the introduction of some very general – and often very simple – principles and relationships, from which whole classes of apparently very different things may be deduced." Theory thus helps in the conceptual understanding of true variables and, therefore, of the relationships among them, and consequently also in the design of appropriate experiments to make practical observations.

An example of our own may help to clarify Haavelmo's distinctions. Keynesian theory classifies a person as involuntarily unemployed when he is not employed and the real wage exceeds his marginal disutility of labor – i.e., he would like to work at the going wage but is not offered a job (Keynes 1936, ch.2). The theoretical concept suggests an experimental design to elicit unemployment data: People should first be divided into those working and those not working. Those not working should be surveyed and asked whether they would be willing to work for the actual wage rate being paid to workers in jobs that they would be qualified to do.

The data needed to construct the true unemployment rate would have to survey every person in the economy. In practice, of course, there are practical difficulties. Surveys will always be more limited, so that questions of representativeness arise. People may not answer the survey honestly. And in a world of heterogeneous labor, determining which of the variety of wage rates is relevant is daunting. Thus, the observed unemployment rate may fail to conform in various ways to the true rate. But many of the problems are remediable, so that with sufficient will and resources, greater conformity between the two variables might be obtained.

Some of the practical problems, however, may bleed into conceptual problems. In practice, the U.S. Bureau of Labor Statistics actually conducts its surveys somewhat differently (see Hoover 2012b, ch. 12, section 12.2). They ask two questions: Are you working? And if not, have you actively sought work in the past two weeks? The first question establishes the division of persons into workers and nonworkers, and the second is supposed to establish the division between the involuntary and voluntarily unemployed. The conceptual basis for the second division is not the same as that proposed in Keynesian theory, so that reported unemployment data diverge from the true data needed to test any Keynesian theory involving involuntary unemployment. In practice, the mismatch may prove inconsequential if, in fact, the time-series behavior of the unemployment data approximated that of a conceptual purer involuntary unemployment data sufficiently well (Haavelmo 1944, p. 7). But there are in many other cases, a range of deeper problems, not so easily solved by an appeal to approximation.

The distinction between theory and model is not drawn explicitly in *The Probability Approach*, but it is implicit in Haavelmo's term "theoretical model." Roughly, Haavelmo treats theory as abstracted from real phenomena and, possibly, conceptually incomplete. For example, economic theory may tell us that demand and price are related inversely, but may be indifferent among further possible concretizations – for example, among particular values for the price elasticity or among other variables that are taken to influence demand. There exist, then, a variety of theoretical models compatible with the theory. Many of these models may be mutually incompatible. Others may form families of compatible models, the members of which are consistent, as long as restricted to *specific* domains defined by *ceteris paribus* assumptions. For example, a demand relationship $q = f(p)$ may be compatible with $q = g(p, y)$, so long as they share, say, a related functional form, and the appropriate *ceteris paribus* condition holds: $y = \bar{y}$. This is one sense in which different compatible perspectives on the same reality are possible.

Theoretical models posit stable relationships. A deep problem in economics is that there is no guarantee that correspondingly stable relations exist in reality. Haavelmo (1944, p. 13) calls relationships that *in fact* are stable under experimentation *autonomous* (see Aldrich 1989). Autonomy is not a property of the abstract theoretical model, but a property of observable reality. *No* experiment – either controlled experiments of laboratory science or the hypothetical experiments invoked in passive observation – are *completely* controlled (Haavelmo 1944, p. 18). In setting up an experiment, we frequently ignore potentially influential factors because they are known to be stable or they are inaccessible or simply because we are ignorant. Should these factors change, the observed relationship will fail to be autonomous.

In the case that we know that influential factors are constant, autonomy in observational models is analogous to *ceteris paribus* conditions in theoretical models. A shift in the value of an influential factor shifts the domain of applicability of the observational model. While we may account for the effects of shifts in known factors in the observational model, whether we are able to incorporate them into the theoretical model depends on whether they are comprehensible given the conceptual resources of the theory.

Haavelmo also recognizes that autonomy can be threatened by factors that are inaccessible to us or of which we are ignorant. Observable relationships typically hold only in an "environment" or "milieu" that cannot be fully specified in advance and so open up the possibility of unpredictable collapses of autonomy (Haavelmo 1944, section 8; 1954, p. 2). These may to some degree be treated as exceptional cases and captured through statistical methods *ex post*: "The construction of systems of autonomous relations is, therefore, a matter of intuition and factual knowledge; it is an art" (Haavelmo 1944, p. 29). But too many exceptions, too many failures of autonomy, sap the value of the observational model (Haavelmo 1944, p. 25). Luck matters.

As we have seen already, testing for Haavelmo amounts to checking the match between the theoretical model and the true model of the observations. An autonomous stochastic model must, according to Haavelmo, account for the stochastic behavior of the observable variables as it is in reality, not as it is posited by some theoretical model. In the case of passive observation

> we can only try to adjust our theories to reality as it appears before us... We try to choose a theory and a design of experiments to go with it, in such a way that the resulting data would be those which we get by passive observation of reality. [Haavelmo 1944,p. 14]

He argues that stochastic variables are interpretable only in a well-defined stochastic model and "it is ... important ... not to force certain data into an alien model" (Haavelmo 1954, p. 6; cf. p. 5 and 1944, p. iv). A stochastic model adequate to the observed variables can be used to test a theoretical model in Haavelmo's view as long as the key features of the theory are preserved in a way that conforms both to the theoretical model and to the model of the true variables. The central difficulty of empirical research is establishing the necessary conformity to justify tests and to make their conclusions meaningful.

Haavelmo sees empirical research as an iterative process. When the model of the observable data do not sufficiently match the theoretical data to permit a useful test, then there are two choices, bring the data closer to the theory or bring the theory closer to the data. Haavelmo provides an example. Consider a theory that the quantity of a good (y) depends on its price (p).[8] The theory is realized in a theoretical stochastic model:

(A) $y = \alpha p + u$

(B) p can be deliberately fixed for experimental purposes.

(C) For every fixed value of p, u is an unobservable random variable with a known distribution which does not depend on the value of p. The u's are independent in repeated trials.

(D) $E(y) = \alpha p+$ constant

(E) α is an unknown parameter. [Haavelmo 1954, appendix]

Haavelmo treats the model as an experimental design in which the behavior of the data can be determined through repeated realizations of the random term u.

Haavelmo then asks what happens if the data are not, in fact, generated in the manner that the theoretical models supposes, but instead from a time series process:

(a) $y(t) = \alpha p(t) + w_1(t) + hw(t)$

(b) $p(t) = \beta p(t-1) + w_2(t) + kw(t)$

(c) w, w_1, w_2 are mutually and serially independent (unobservable) random variables with known distributions.

(d) α is the same unknown parameter as in (A). β, h, and k are unknown constants. [Haavelmo 1954, appendix]

[8] Haavelmo actually writes x rather than y in initially defining the quantity demanded, but he writes y in the actual demand functions and in all but one other case in working out the example. It appears to be a slip, and we write y consistently.

The second, observed model is incompatible with the first theoretical model. In particular, as Haavelmo points out, except for special choices of the parameters, $E(y(t)) \neq \alpha p +$ constant. To insist on using the theoretical model in the face of the fact that the time-series model describes the actual behavior of the data is precisely to force "certain data into an alien model."

Several issues arise with respect to this example. What is the status of the second model? Is it a new theoretical model or is it an observational model that must be interpreted through the first theoretical model? Haavelmo is noncommittal: one could "if one wants" offer an interpretation of the new elements of the structure. He gives a little story; but the story draws on very different conceptual resources than those that we usually associate with economic theory; and, in any case, it seems to be optional. So, another interpretation would be to regard it as a purely observational construct. One way that we could have obtained the model is through an iterative process in which statistical tests would have revealed that key stochastic assumptions of the first model, such as (B) the serial independence of the u's would have failed. This might suggest some sort of dynamic process.[9] The second model would, then, be a good guess, informed by the test of the first model.

Were we to change Haavelmo's example a little and to treat w_1,w_2 not as independent and unobservable random variables, but as other observable variables governed by processes similar to (b), then changes in these variables would have been revealed as failures in the implicit assumption of autonomy of the demand curve (A). Extra-statistical knowledge (and, again, good guessing) might have suggested that these variables be introduced as additional controls.

We can, therefore, think either of adapting theory to observation (i.e., replacing the first model with the second as the theoretical model) or as providing a better experimental design that takes account of features of the data not addressed in the theory in such a way that we can accurately measure the true variables – in this case, particularly, the value of α. Either way, the first model will never be accepted on any well formulated test. Under the second way of interpreting the example, Haavelmo's condition (d) that asserts that the same α appears in both models can be rationalized in the following way. The first model is rejected against the more general second model. But the features that produce the rejection (i.e., the failure to capture the dynamics) do not touch the interpretive core (i.e., the meaning of α); hence it is accept-

[9] See Hoover, Johansen, and Juselius (2008) for a similar analysis of another example.

able to use the simpler first model to represent a class of models that embed the same linear law of demand and to use if for conceptual analysis.

A final issue is that Haavelmo draws special attention to the role of w, the common random shock to y and p. We shall come back to this issue in the next section.

The key lessons from Haavelmo's illustration are these: In order to isolate the relevant true variables and true relationships (to control the experiment), a process of adapting either the observational design or the theoretical model is essential. While the adaptations can take place on either pole, only actual examination of the data will point out when adaptations are required, so the process is not purely theoretical, but rather fundamentally empirical. The relationships of stochastic variables can be interpreted legitimately only within a stochastic model consonant with the observable data and more general than any relationships to be tested. Some aspects of an appropriate stochastic model may have clear theoretical interpretations, others may be at best empirically warranted, *ex post* adjustments that deliver an interpretable stochastic model.

We address the way in which these features play out in the cointegrated vector autoregression in the next section.

6 Experiments and CVAR Scenarios

The power of Haavelmo's use of the experiment as a simile for understanding passive observation is well illustrated in the light that it casts upon "scenario analysis" in the context of the cointegrated vector autoregression (CVAR). To take a concrete case, we refer to the test of monetarism on Danish data discussed by Juselius (2006, ch. 2) and Juselius and Johansen (2006).

Monetarism comprises, among other features, the quantity theory of money, which implies that money is neutral in the long run though it affects real variables in the short run (Friedman 1956, 1969). Friedman (1956) argues that the quantity theory depends fundamentally on the notion of a stable demand-for-money function that comprises a transactions demand dependent on nominal income and, therefore, in aggregate on real GDP and the price level, and an asset demand for which there are two relevant margins: between money and (i) real goods with an opportunity cost measured by the rate of inflation and (ii) other financial assets with an opportunity cost measured as the yield differential between an alternative asset and money.

Monetarism is compatible with a variety of relationships among real and nominal variables in the short run (Friedman 1974). As we saw in section 4, Haavelmo does not argue that a theory tested through an experimental design must be complete. Rather the theoretical model may exemplify a class of models with specific generic features leading to a test of a class of models. We note that, in addition, to the quantity theory, most versions of monetarism also subscribe to two doctrines that are related to the quantity theory in various ways: first, Irving Fisher's hypothesis that a nominal interest rate is the sum of an *independent* real rate of interest and an inflation rate; and, second, the expectations theory of the term structure of interest rates. Our focus is on long-run versions of these doctrines as well as of the quantity theory of money. In addition, many macroeconomists treat the quantity theory as requiring a minimum of two independent types of shock – demand and supply shocks. We also restrict our attention to versions of monetarism in which the rational-expectations hypothesis holds.

Haavelmo's notion of the design of experiments already constrains the formulation of a model of the true variables. On the one hand, the true variables must be adequate counterparts of the theoretical variables; and, on the other, the observational variables must adequately represent the true variables. Furthermore, the model must be general enough to allow a generic test of the key theoretical relationships. The design strategy thus operates along two margins governing the relationships (i) between the theoretical and the true variables and (ii) between the true and the observed variables. The observed Danish counterparts to the theoretical variables in Juselius (2006) are: the logarithm of the M3 monetary aggregate (m), the logarithm of real gross national expenditure (y^r), the logarithm of the consumer price index (p), the own yield on M3 (R_m), and the yield on a long-term government bond (R_b). Of course, the inflation rate is then Δp; the real stock of money, $m-p$; the opportunity cost against real goods (the real short rate of interest), $R_m - \Delta p$; and the opportunity cost against financial assets, $R_b - R_m$. With respect to the first margin, the conceptual match between the theoretical variables (broad money, the general price level, level of activity and opportunity costs) is reasonably good. However, the underlying theory is not completely specific and, we could, for instance, make other choices of variables for, say, the long-term rate of interest or the price level with equal claim to being true variables. With respect to the second margin, even taking these particular choices as given, we cannot rule out some deviation between the actual measurements of these variables and the ideally correct measurements.

A test of monetarism requires a model of the true variables in which particular monetarist claims may or may not hold. A VAR representation of these data provides a very high level of generality in which to frame such a test. It is a structured characterization of the information in the data. The VAR model can be written as

$$x_t = \mu_0 + \Pi_1 x_{t-1} + \ \ldots + \Pi_k x_{t-k} + \Phi D_t + \varepsilon_t, \quad t = 1,\ 2, \ldots,\ T, \qquad (1)$$

where x_t is a vector of variables, μ_0 is a vector of constants; D_t is a vector of deterministic terms (such as, trends, and dummy variables) ε_t is a vector of identically distributed normal random variables; and Π and Φ are matrices of parameters; all of conformable dimensions; and there exists a sufficient set of initial conditions for the lagged $x's$. The VAR is similar to Haavelmo's (1954) dynamic demand model (see Section 5 above), although both models are more general than the ones that Haavelmo (1944, 1954) typically contemplates, since each is compatible with all the variables, except the deterministic terms, being endogenously determined.[10]

To provide an adequate account of the stochastic mechanism, we must work both margins: on the one hand, providing an adequate empirical characterization of the data; on the other, transforming it into a perspective in which the theoretical and true relations among the data can be compared. Such a framework in which theory can be brought to bear on the appropriate data is essential; for as Haavelmo observes:

> A sample of observations is just a set of cold, uninteresting numbers unless we have a theory concerning the stochastic mechanism that has produced them. [Haavelmo, 1950: p. 265].

The design of the experiment in Haavelmo's view involves establishing a perspective on the data in several stages. The VAR model in (1) provides a conceptual framework. The first stage uses misspecification tests to determine whether the model is an adequate description of the data. If not, it must be turned into one through further specification for which the tests may provide some useful guidance. But we note that (1) is not perfectly general, as we have supposed linearity in parameters. Linearity is a perspective imposed on the data that *could* correspond to a true feature of the world,

[10] Haavelmo typically refers to endogenous variables as *dependent* and exogenous variables as *independent* or *datum-variables.*

but which may also work when a first-order Taylor expansion is an adequate approximation to an inherently nonlinear economic model. Also, in practice, we contemplate a restricted set of deterministic variables, such as constant, trend, dummies. Furthermore, formulating an adequate model of the true variables is a process that is informed by the actual properties of the observed data:

> The economist ... is presented with some results which, so to speak, Nature has produced in all their complexity, his task being to build models that explain what has been observed. [Haavelmo 1944, p. 7].

To secure a stable well-formulated stochastic structure, such as Haavelmo requires for interpretability, it be may necessary to control for important events or institutional change (such as reforms, interventions, wars) through additional variables (particularly, deterministic variables, such as step or impulse dummies) or additional lags. In some cases, the precise variables, lags, or functional forms may be anticipated on the basis of background knowledge or shrewd theoretical insights, but in others, it will be a matter of *ex post* adjustments to the actual behavior of the observed data informed by either general characteristics of the data or by formal specification tests. (See Juselius (2006, 2012) for a more detailed account of the formulation of an adequate VAR representation.)

At this stage, it is essential to follow Haavelmo's advice to use the conceptual resources of economic theory to design an experiment that is statistically adequate and, at the same time, formulated in a manner that the key propositions of the theoretical model can be adequately represented and tested. We refer to this formulation as a scenario – that is, a concrete specialization or set of restrictions on a more general stochastically and observationally adequate representation of the observed data that corresponds to the distinguishing features of the theoretical model. The general representation is, thus, the controlled framework within which the experiment is run, and the scenario is the experiment to be run within that framework.

To turn to the specific case, the long-run relationships of monetarism are naturally expressed as cointegrating relationships, so it is natural to reformulate equation (1) as a CVAR:

$$\triangle x_t = \mu_0 + \alpha\beta' x_{t-1} + \Gamma_1 \triangle x_{t-1} + \Phi D_t + \varepsilon_t, \tag{2}$$

where for p variables and r cointegrating relationships, $\Pi = \alpha\beta^{'}$ and α and β are $p \times r$ matrices, $r < p$, and $\beta' x_t$ defines the stationary combination of nonstationary variables. Here both for ease of exposition and because a VAR in levels with only two lags characterizes the Danish data well, the CVAR is formulated with only one lagged difference.

It is useful to also represent (2) in its inverted form in which the variables are expressed as a function of shocks and deterministic components:

$$x_t = C\sum_{i=1}^{t}\varepsilon_i + C\mu_0 t + C\sum_{i=1}^{t}\Phi D_i + C^*(L)(\varepsilon_t + \Phi D_t) + \tilde{X}_0, \tag{3}$$

where $C = \beta_\perp(\alpha_\perp(I - \Gamma_1)\beta_\perp)^{-1}\alpha_\perp$, measures the long-run impact of a shock to the system; $C^*(L)(\varepsilon_t + \Phi D_t)$ is an infinite lag polynomial describing the impulse response function of the stochastic shocks, ε_t,and deterministic shocks, D_t; and $\tilde{X}_0$ contains the initial values, $x_{0,}x_{-1}$, of the process and the initial value of the short-run dynamics $C^*(L)\varepsilon_0$. The representation (3) describes a decomposition of the vector process, x_t, into stochastic trends, $C\sum_{t=1}^{t}\varepsilon_t$, deterministic trends, $C\mu_0 t$, cycles, $C^*(L)\varepsilon_t$, and irregular components, ε_t and D_t.

As a first step to bring the general CVAR perspective to bear on monetarism, we let $x_t = [m_t, p_t, y_t^r, R_{m,t}, R_{b,t}]'$. We note that statistical tests suggest the variables are nonstationary: that p and m are I(2) and the remaing variables are I(1). This leads to a tentative decomposition of the data vector into two stochastic trends, one deterministic time trend, and a stationary cycle component:

$$\begin{bmatrix} m_t \\ p_t \\ y_t^r \\ R_{m,t} \\ R_{b,t} \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ 0 \\ 0 \\ 0 \end{bmatrix} \left[\sum\sum u_{1i}\right] + \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \\ d_{31} & d_{32} \\ d_{41} & d_{42} \\ d_{51} & d_{52} \end{bmatrix} \begin{bmatrix} \sum u_{1,i} \\ \sum u_{2,i} \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ 0 \\ 0 \end{bmatrix} [t] + X_0 \tag{4}$$

where X_0 is a catch-all for stationary components and initial values. The specification is constrained both by facts about the data, particularly the integration properties, and background assumptions to the monetarist scenario, particularly the assumption of two autonomous shocks – an aggregate demand or nominal shock ($u_{1,t}$) and an aggregate supply (or real) shock ($u_{2,t}$).

Under the testable assumption that m_t and p_t are I(2), the nominal shock, $u_{1,t}$, cumulates twice to describe a second-order stochastic trend in money and prices and once to describe a first-order stochastic inflation trend. The real shock, $u_{2,t}$,cumulates once, consistent with the real variables being I(1). The linear deterministic trends for m_t, p_t, and y_t^r reflected in assumption that $g_1, g_2, g_3 \neq 0$ corresponds to the fact that the average growth rates of m_t, p_t, and y_t^r are significantly different from zero. For the Danish data, it is also necessary to include one step-dummy in the cointegration relations to account for the change in the equilibrium mean associated with the deregulation of capital movements in 1983 which caused (i) a strong reallocation of money holdings; (ii) a change in the inflation rate; and (iii) a change in the risk premium of the interest rate spread.

A formulation such as CVAR model (4) addresses one of Haavelmo's (1954, appendix) concerns (discussed in section 5 above), that models would very commonly be driven by common shocks. He worried that this would be a confounding feature that would stand in the way of recovering the underlying structure from the observations. While the concern is a live one for stationary data, for nonstationary data, nearly the opposite is the case: cointegrating relations among variables with genuinely common stochastic and deterministic trends are easier to observe and cointegrating relations among variables with unrelated stochastic trends are easier to reject than for long-term relationships among stationary variables. Nonstationarity amounts to Nature performing very dramatic experiments that reveal, rather than cloak, structure. They also help to resolve another of Haavelmo's characteristic concerns – the idea that a *potentially* important casual variable might show too little variance either to be an important actual cause or to support a precise parameter estimate. Again, because the variance of nonstationary series increases over time, we generally do not have to wait very long for potential factors to be revealed. This is because parameter estimates among nonstationary variables are superconsistent, converging at a rate of $1/T$ rather than $1/\sqrt{T}$ as with stationary variables.[11]

CVAR model (4) turns out to be an observationally consistent description of the Danish data. It is, at once, constrained enough to characterize the data and flexible enough to provide a key part of the experimental design – the controlled framework in which characteristic monetarist propositions

[11] The superconsistency of estimates also eliminates simultaneity bias in cointegrating relationships, which was another of Haavelmo's (1943) central concerns (see Juselius 2012).

can be tested. The other key part is the monetarist scenario itself: the long-run neutrality of money implies that money and prices are homogeneous of degree one; a stable long-run demand for money implies that velocity is stationary; the expectations-theory of the term structure implies that the differential between short and long rates of interest is stationary; and the Fisher hypothesis implies that the real rate of interest is stationary. More explicitly:

5.A m_t and v_t are homogeneous of degree 1: the neutrality of money

5.B $m - p - y^r \sim I(0)$: the velocity of money

5.C $R_m - R_b \sim I(0)$: the term structure

5.D $R_m - \Delta p \sim I(0)$: the Fisher hypothesis

5.E expectations are rational

The assumption of homogeneity can be formulated as the restriction in CVAR model (4). In practice, the neutrality between money and prices (5.A) is accepted for the Danish data. Homogeneity allows us to impose a useful simplification on CVAR model (4) that amounts to a more restrictive change of perspective. Under this assumption, (Kongsted, 2005) shows that CVAR model (4) can be reformulated in an (almost) equivalent form:

$$\begin{bmatrix} m_t - p_t \\ \Delta p_t \\ y_t^r \\ R_{m,t} \\ R_{b,t} \end{bmatrix} = \begin{bmatrix} d_{11} - d_{21} & d_{12} - d_{22} \\ c_{21} & 0 \\ d_{31} & d_{32} \\ d_{41} & d_{42} \\ d_{51} & d_{52} \end{bmatrix} \begin{bmatrix} \sum u_{1,i} \\ \sum u_{2,i} \end{bmatrix} + \begin{bmatrix} g_1 - g_2 \\ 0 \\ g_3 \\ 0 \\ 0 \end{bmatrix} [t] + X_0 \tag{5}$$

where all variables are now at most I(1).

Incorporating these assumptions as restrictions on CVAR model (5) results in a model consistent with the monetarist scenario:

$$\begin{bmatrix} m_t - p_t \\ \Delta p_t \\ y_t^r \\ R_{m,t} \\ R_{b,t} \end{bmatrix} = \begin{bmatrix} 0 & d_{12} \\ c_{21} & 0 \\ 0 & d_{12} \\ c_{21} & 0 \\ c_{21} & 0 \end{bmatrix} \begin{bmatrix} \sum u_{1,i} \\ \sum u_{2,i} \end{bmatrix} + \begin{bmatrix} g \\ 0 \\ g \\ 0 \\ 0 \end{bmatrix} [t] + X_0 \tag{6}$$

where real income and real money stock share the real stochastic trend, $\sum u_{2,i}$ and the deterministic time trend; and inflation and the two nominal interest rates share the nominal trend, $\sum u_{1,i}$.

We have, so far, ignored the stationary components of the data in CVAR models (4) - (6). An important feature of the CVAR formulation is the dichotomy between the long-term, nonstationary component and the short-term, stationary component. The superconsistency of the parameter estimates among nonstationary variables allows a substantially independent analysis of the nonstationary component. The error-correction of deviations from cointegrating relationships is one of the forces that affects short-term adjustment behavior, as, in effect, these deviations are another element of the stationary component. Because of the superconsistency of the parameter estimates of the cointegrating relationships, the stationary component can be analyzed as a second stage of a complete analysis conditional on having adequately identified the cointegrating relationships.

The long-term monetarist scenario defines a class of models compatible with a variety of short-term specifications. For example, an extended monetarist scenario, including the hypotheses that money causes prices in the long run and that money causes nominal income in the short-run would be evaluated in a model that specified both the contemporary and lagged dynamics of the variables. We do not, however, pursue these extensions here.[12]

The actual empirical investigation of Danish data shows that each of the key monetarist relationships in 5.B-5.C are too persistent to be stationary (Juselius 2006, p. 188). Although this is an important negative result, at least for a particular class of monetarist theories, such results may more generally yield positive information. For example, the fact that some cointegrating relationships are supported and others contradicted for a particular scenario, could suggest the empirically relevant direction in which it would be helpful to reformulate the theoretical model. More concretely, in a version of monetarism with rational expectations, just the fact that an *ex post* step dummy variable is necessary to formulate an adequate stochastic specification (as in the Danish data) helps to direct further investigation. The usual approaches to rational expectations (5.E) assume stationary probabil-

[12] Giese (2008) provides an illustration, using a similar scenario analysis to investigate a common specialization of the expectations theory of the term-structure model in which, in addition to their cointegrating relationship, the short rate of interest *causes* the long rate. She rejects the relevant zero restriction and, indeed, finds evidence for the weak exogeneity of long rates.

ity distributions in order to apply the law of iterative expectations (Sargent 1987, ch. 3, Sargent and Lundqvist 2004, ch. 2). A location shift in the stochastic trend violates this assumption (Hendry and Mizon, 2010). There are several approaches that might be taken to address this issue within a rational-expectations framework: for example, restricting the scope of rational expectations to the periods between location shifts as in Lucas (1976) or modeling learning as in Sargent (1994) or Evans and Honkopohja (2001). Another alternative, the imperfect knowledge economics (IKE) of Frydman and Goldberg(2007, 2011), is compatible not only with the structural breaks in 1983, but also with the positive finding of persistence in the real interest rate, the term spread, and the velocity of money. These further approaches call for new experimental designs and the evaluation of new scenarios.

7 Experiments and the Growth of Economic Knowledge

The scope for actual experimentation in economics is limited. Nonetheless, Haavelmo found the simile of the controlled experiment to be a fruitful way to formulate a positive methodology of passive observation. Unfortunately, the constructive aspects of Haavelmo's methodology have frequently been neglected to the point that it has been caricatured the almost wholly negative view of the economist as the grim executioner of theories. Haavelmo has been thought to advocate relying on a nonempirical, *a priori* economic theory to propose hypotheses that are then accepted or rejected mechanically on the basis of statistical tests. Such a characterization is wide of the mark. Haavelmo in fact provided a subtle account of the constructive interplay between theory and observation. In his account, experiments are designed using theoretical perspectives and revised and elaborated in light of the observational results of those experiments. Passive observation is understood through analogy to experiment in which statistical techniques serve a similar function to experimental controls, while the interplay between theory and observation is fundamentally similar to the case with actual experiments. Haavelmo's ideas illuminate the practice of a theory-consistent CVAR scenario, which in turn pushes the experimental analogy beyond the point that Haavelmo left it six decades ago. A theory consistent CVAR scenario is often informative about how to modify the theory model when the correspondence between the theoretical and observed structure is weak. In Haavelmo's words:

> we can only try to adjust our theories to reality as it appears before us. And what is the meaning of a design of experiment in this case. It is this: We try to choose a theory and a design of experiments to go with it, in such a way that the resulting data would be those which we get by passive observation of reality. And to the extent that we succeed in doing so, we become masters of reality – by passive agreement. [Haavelmo, 1944, p.14]

8 References

Aldrich, J. (1989) Autonomy,*Oxford Economic Papers* N.S. 41(1), 15-34.

Boumans, M. (2005) *How Economists Model the World into Numbers*. London: Routlege.

Boumans, M. (2012) Haavelmo's Methodology of Empirical Science Outside the Laboratory, unpublished typescript, University of Amsterdam.

Congressional Budget Office (1995) CBO's Method for Estimating Potential Output, CBO Memorandum, October 1995, at ftp://ftp.cbo.gov/3xx/doc312/potout.pdf

Duarte, P. G. and K. D. Hoover (2012) Observing Shocks. In H. Maas and M. Morgan, (eds.) *Histories of Observation in Economics*, supplement to *History of Political Economy* 44(5), forthcoming.

Evans, G. W. and S. Honkapohja (2001) *Learning and Expectations in Macroeconomics*. Princeton: Princeton University Press.

Eichenbaum, M. (1995) Some Comments on the Role of Econometrics in Economic Theory, *Economic Journal* 105(433), 1609-1621.

Fisher, R.A. (1935) *The Design of Experiments*. Edinburgh: Oliver and Boyd.

Friedman, M. (1956) The Quantity Theory of Money: A Restatement. In Milton Friedman (ed.) *Studies in the Quantity Theory of Money*. Chicago: University of Chicago Press.

Friedman, M. (1969) The Optimal Quantity of Money. In *The Optimal Quantity of Money and Other Essays*. London: Macmillan.

Friedman, M. (1974) A Theoretical Framework for Monetary Analysis. In Robert J. Gordon (ed.) *Milton Friedman's Monetary Framework*. Chicago: University of Chicago Press.

Frisch, R. (1926) *Sur les Semi-invariants et Moments Employés dans l'Etude des Distributions Statistiques.* Oslo: Jacob Dybwad for Akademi i Oslo.

Frisch, R. (1939) A Note on Errors in Time Series. *Quarterly Journal of Economics* 53(4), pp. 639-640.

Frydman, R. and M. Goldberg (2007) *Imperfect Knowledge Economics: Exchange Rates and Risk.* Princeton: Princeton University Press.

Frydman, R. and M. Goldberg (2011) *Beyond Mechanical Markets: Risk and the Role of Asset Price Swings.* Princeton: Princeton University Press.

Giese, J. V. (2008) Level, Slope, Curvature: Characterizing the Yield Curve in a Cointegrated VAR Model. *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 2, 28.

Haavelmo, T. (1943) The Statistical Implications of a System of Simultaneous Equations. *Econometrica* 11(1), pp. 1-12.

Haavelmo, T. (1944) The Probability Approach to Econometrics. *Econometrica*, 12 (Supplement), iii-118.

Haavelmo, T. (1950) Remarks on Frisch's Confluence Analysis and its Use in Econometrics, in ed. T. C. Koopmans, editor. *Statistical Inference in Dynamic Models, Cowles Commission Monograph*, Vol. 10. New York: Wiley, pp. 258-265.

Haavelmo, T. (1954) Structural Models and Econometrics, unpublished typescript, University of Oslo.

Heckman, J. J. (1992) Haavelmo and the Birth of Modern Econometrics: A Review of the History of Econometric Ideas by M. Morgan. *Journal of Economic Literature* 30(2), pp. 876-886.

Heckman, J. J. (2000) Causal Parameters and Policy Analysis in Economics: A Twentieth Century Perspective. *Quarterly Journal of Economics* 115(1), 45-97.

Hendry, D. F. and G. E. Mizon (2010). On the Mathematical Basis of Inter-temporal Optimization,University of Oxford, Economics Department, Discussion Paper, 497.

Hendry, D. F., A. Spanos, and N. R. Ericssson (1989) The Contributions to Econometrics in Trygve Haavelmo, *Sosialøkonomen* no. 11, pp. 12-17.

Hoover, K. D. (1994) Econometrics as Observation: The Lucas Critique, Causality and the Nature of Econometric Inference. *Journal of Economic Methodology* 1(1), 65-80.

Hoover, K. D. (2012a) Pragmatism, Perspectival Realism, and Econometrics. In A. Lehtinen, J. Kuorikoski and P. Ylikoski (eds.) *Economics for*

Real: Uskali Mäki and the Place of Truth in Economics. London: Routledge, pp. 223-240.

Hoover, K. D. (2012b) *Applied Intermediate Macroeconomics.* Cambridge: Cambridge University Press.

Hoover, K. D., S. Johansen, and K. Juselius. (2008) Allowing the Data to Speak Freely: The Macroeconometrics of the Cointegrated Vector Autoregression. *American Economic Review* 98(2), 251-255.

Juselius, K. (2006) *The Cointegrated VAR Model. Methodology and Applications* Oxford: Oxford University Press, 2006.

Juselius, K. (2012) Haavelmo's Probability Approach and the Cointegrated VAR, unpublished typescript, University of Copenhagen.

Keynes J. M. (1936) *The General Theory of Employment Interest and Money.* London: Macmillan.

Kongsted, H. C. (2005) Testing the Nominal-to- Real Transformation, *Journal of Econometrics* 124(2), 202-225.

Louçã, F. (2007) *The Years of High Econometrics: A Short History of the Generation that Reinvented Economics.* London: Routledge.

Lucas, R. E. Jr. (1976) Econometric Policy Evaluation: A Critique, reprinted in Lucas (1981), pp. 104-130.

Marschak, J. (1953) Economic Measurements for Policy and Prediction. In W.C. Hood, and T. Koopmans (eds.) *Studies in Econometric Method.* Cowles Commission Monograph vol. 14. New York: Wiley, pp. 1-26.

Mill, J. S. (1843[1973]) A System of Logic, vol. 1. In J.M Robson (ed.) *The Collected Works of John Stuart Mill*, vol. 7. Toronto: University of Toronto Press, 1973.

Morgan, M. S. (1990) *The History of Econometric Ideas.* Cambridge: Cambridge University Press.

Russell, B. (1927) *The Analysis of Matter.* New York: Harcourt, Brace.

Sargent, T. J. (1987) *Dynamic Macroeconomics.* Cambridge, MA: Harvard University Press.

Sargent, T. J. (1994) *Bounded Rationality in Macroeconomics: The Arne Ryde Memorial Lectures.* Oxford: Oxford University Press.

Sargent, T. J. and L. Lundqvist (2004) *Recursive Macroeconomic Theory*, 2^{nd} edition. Cambridge, MA: MIT Press.