STYLIZED FACTS, THE PURPORT OF AN ECONOMIC THEORY, AND SCIENTIFIC EXPLANATION IN ECONOMICS AND ECONOMETRICS[1]

by

Bernt P. Stigum

November 8, 2004

http://folk.uio.no

I. Introduction

Economic science is systemized knowledge about the nature of social reality that pertains to economic matters. No doubt, economists believe that their science is not fictional. They develop theories and collect data to determine the empirical relevance of their theories. They provide scientific explanations for regularities that they observe in their data. And they use their theories and data to make predictions about the future of relevant economic phenomena and events. However, in doing all this they face a problem that I believe most of them do not take seriously enough.

1.1 *A Serious Problem*

In the first chapter of his classic treatise, **The Probability Approach in Econometrics** (Haavelmo, 1944) Trygve Haavelmo makes two observations that I can use to explicate the nature of the problem that economic scientists are facing.

(1) 'When [asked] about the actual *meaning* of [a theoretical variable, the] answer we might give consists, at best, of a tentative description involving words

which we have learned to associate, more or less vaguely, with certain real phenomena' (Haavelmo 1944, 4).

(2) 'It is never possible – strictly speaking – to avoid ambiguities in classifications and measurements of real phenomena. [In] most cases we are not even able to give an unambiguous description of the *method* of measurement to be used, nor are we able to give precise rules for the choice of *things to be measured* in connection with a certain theory' (Haavelmo 1944, 4).

These observations have severe implications for the kind of empirical analysis that Haavelmo and his many followers today are advocating. In any empirical analysis in which the relevance of an economic theory matters; i.e., in a theory-data confrontation, the researcher in charge (RIC) is to distinguish between true, theoretical, and observational variables. The value of a true variable (or time function) is taken to be an accurate measurement of reality "as it is in fact." The value of a theoretical variable is the true measurement that RIC should make if reality actually were in accordance with the theoretical model (Haavelmo 1944, 5). Finally, the value of an observational variable is the value of the variable that RIC has observed in a pertinent field survey or in relevant government publications as the case may be. **If one takes the quotes from Haavelmo at face value, the concept of a true variable and the true value of a theoretical variable have no definite meaning. Also, the references of RIC's observational variables are ill defined.**

Strictly speaking, the theory in an economic theory-data confrontation is a theory about undefined terms; e.g., about toys in a toy economy. The names of the terms indicate the kind of situation in social reality about which the originator of the theory was theorizing. The data in a theory-data confrontation are data that RIC has constructed from observations that he has obtained from field surveys or government publications. The references of these data belong for the most part in a socially constructed world of ideas. The toy economies and economists' world of ideas have little in common with social reality – a fact which renders successful searches for knowledge about economic aspects of social reality problematic.

I believe that the given problem can be handled in a reasonable way. A proper understanding of the purport of an economic theory and meaningful bridge principles that link up theoretical variables and relevant data can enable economists to learn about characteristic features of social reality.

1.2 *The Purport of an Economic Theory*

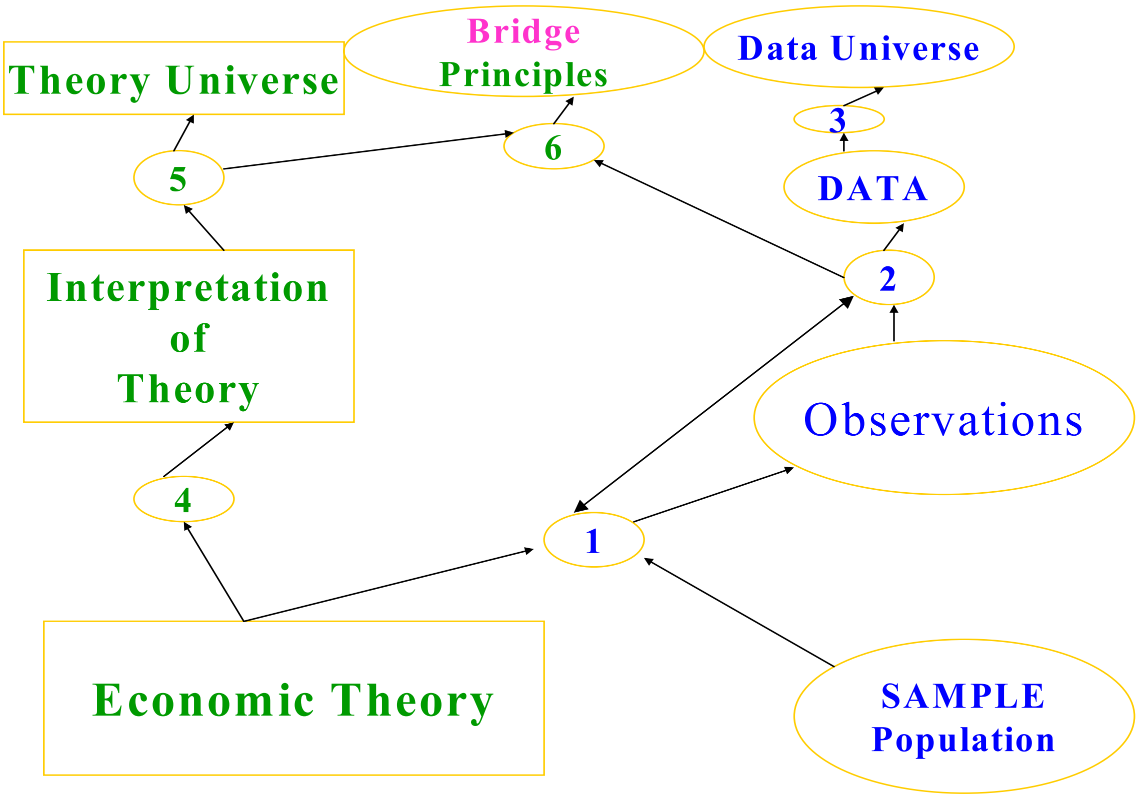
To me the intended interpretation of an economic theory of choice or development delineates the positive analogies that the originator of the theory considered sufficient to describe the kind of situation that he had in mind.[2] For example, in an economic theory of the firm one learns that the manager of a firm chooses his input-output strategy so as to maximize the firm's profits. This characterization provides a succinct description of an important characteristic feature of firm behavior in the theory's reference group of firms in social reality. Similarly, in an economic theory of a certain kind of financial markets the theory may insist that the family of equilibrium yields on the pertinent instruments are co-integrated ARIMA processes. This characterization describes a characteristic feature of the probability distributions that govern the behavior over time of equilibrium yields in such markets. Whether the positive analogies in question have empirical relevance is a query that can be answered only by confronting the given theories with appropriate data.

In reading the preceding observations on the purport of an economic theory it is important to keep the following points in mind. (1) To say that profit maximization is a positive analogy of firm behavior is very different from saying that the firm behaves as if it were maximizing profits. The managers of the firms in the theory's reference group do, by hypothesis, choose their input-output vectors so as to maximize their firms' profits. (2) Even though an accurate description of the behavior of a given firm would exhibit many negative analogies of firm behavior, the positive analogies that the theory identifies must not be taken to provide an approximate description of firm behavior. (3) The understanding of the theory and the data one possesses determine what kind of questions about social reality one can answer in a theory-data confrontation. With the first and third point I rule out of court the instrumentalistic view of economic theories expounded by Friedman (1953). With the second and the third point I want to distance my view from the idea that the theorems of an economic theory are tendency laws in the sense that Mill (1836) gave to this term.[3]

1.3 *Bridge Principles in Theory-Data Confrontations*

Haavelmo saw no need for the use of bridge principles in the kind of empirical analysis he was advocating**. If the references of the true variables differed from the references of the observational variables, RIC could correct his data or adjust his theory so as to make 'the facts [that he considered] to be the [values of the] 'true' variables relevant to the theory'** (Haavelmo 1944: 7). However, when the concept of a true variable has no definite meaning and the references of RIC's data belong in a world of ideas, Haavelmo's cavalier attitude to empirical analysis cannot be justified.

Fig. 1 **A Theory-Data Confrontation**



To me the theory and the data in a theory-data confrontation are given and RIC cannot escape the problem of figuring a way to relate his theoretical variables to his data. A look at the prototype of a theory-data confrontation pictured in Fig. 1 shows what I have in mind. On the left side of the figure are boxes that contain information pertaining to the relevant theory; i.e., the theory itself, models of the theory, and the part of the theory that is at stake in the empirical analysis. The last part comprises the ingredients by which the theory universe is constructed. On the right hand side of the figure are boxes that contain information concerning the data generating process; i.e., the sample population on whose characteristics observations are based, the observations, data that the researcher has constructed, and the data universe in which all the pertinent data variables reside. The two universes are disjoint and connected by a bridge. The bridge consists of assertions, called bridge principles, that describe how variables in the theory universe are related to variables in the data universe. No meaningful empirical analysis can take place before RIC has formulated all the relevant bridge principles.

Formally, the Theory Universe is a pair, (Ω_T, Γ_t), where Ω_T is a subset of a vector space and Γ_t is a finite set of axioms that the members of Ω_T must satisfy. The Data Universe is a pair (Ω_P, Γ_p), where Ω_P is a subset of a vector space and Γ_p is a finite set of axioms that the members of Ω_P must satisfy. The bridge principles, $\Gamma_{t,p}$, comprise a finite set of assertions concerning vectors in a subset of $\Omega_T \times \Omega_P$ that I denote by Ω and call the *sample space*. More specifically, let ω_T and ω_P, respectively, be vectors in Ω_T and Ω_P, and let $\omega = (\omega_T, \omega_P)$ be a vector in $\Omega_T \times \Omega_P$. For all $\omega \in \Omega$, the members of $\Gamma_{t,p}$ describe how the components of ω_T are related to the components of ω_P.

E 1.1 Consider an empirical analysis of the market for a given commodity that Aris Spanos describes in (Spanos 1989, 409-410). In the associated formal theory-data confrontation, the theory universe satisfies the conditions:

$\omega_T \in \Omega_T$ only if $\omega_T = (Q^d, Q^s, P, W^d, W^s, u, v)$, where $(Q^d, Q^s) \in R_+^{2N}$, $P \in R_{++}^N$, $(W^d, W^s, u, v) \in R^{4N}$, and $N = \{0,1,\ldots\}$. Also, there exist triples of constants, $(\alpha_0, \alpha_1, \alpha_2)$ and $(\beta_0, \beta_1, \beta_2)$, with $\alpha_0 >> 0$, $\alpha_1 < 0$, $\beta_0 < 0$, and $\beta_1 > 0$, such that for all $\omega_T \in \Omega_T$ and all $t \in N$

$$Q^d_t = \alpha_0 + \alpha_1 P_t + \alpha_2 W^d_t \,;\; Q^s_t = \beta_0 + \beta_1 P_t + \beta_2 W^s_t \text{, and } Q^d_t = Q^s_t$$

where Q^d_t, Q^s_t, P_t, W^d_t, and W^s_t denote the t^{th} component of the respective variables.

As to the data universe,

$\omega_P \in \Omega_P$ only if $\omega_P = (q^*,p^*,z^{*d},z^{*s})$, where $q^* \in R_+^N$, $p^* \in R_{++}^N$, $z^{*d} \in R^N$, and $z^{*s} \in R^N$.

Finally, the bridge principles:

$\Omega \subset \Omega_T \times \Omega_P$, and for all $\omega \in \Omega$ and all $t \in N$,

$q^*_t = Q^d_t + u_t$, $p^*_t = P_t + v_t$, $z^{*d}_t = W^d_t$, and $z^{*s}_t = W^s_t$.

In this formal market the Q^d_t (and Q^s_t) denotes so many units of the commodity in question that the agents in the market plan to buy (sell) at the price P_t; the W^d_t and W^s_t are variables that account for factors, other than the Q's and the P, that are operative in the market; the q^*_t and p^*_t record actual sales and prises in the various periods; the z^{*d}_t and z^{*s}_t are, respectively, the observed values of W^d_t and W^s_t; and u_t and v_t are error terms.[4] Except for the error terms and my insistence that the market in the theory universe is in equilibrium in each period, my model of the given market differs from Spanos' model (cf. his equations 1a, 1b, 2a, 2b) in that the equations that constitute the bridge between the two universes pertain only to the values of the respective variables in the sample space. The latter is a significant difference between the two models.

It is important to keep in mind that the bridge principles in a theory-data confrontation reflect RIC's beliefs as to how his theoretical and data variables are related to one another. Chances are that RIC has little empirical evidence on which to base his beliefs. However, depending upon the subject matter of his theory, he may obtain good help from relevant economic theorems. For example, in a study of consumer choice RIC may relate the price of a commodity and the commodity itself to a price index and a commodity index that he has constructed. The Hicks-Leontief aggregation theorem tells him when such bridge principles are justified. In a study of firm behaviour the theory may insist that the firm's production function is a function of labor and capital and RIC must construct aggregates that he can relate to these theoretical variables. Theorems in Stigum (1967 and 1968) describe aggregates that he can use and delineate conditions under which such bridge principles are justified.

1.4 *Stylized Facts and Scientific Explanations*

Economic theory-data confrontations occur in many different situations. In some of these confrontations economists try to establish the empirical relevance of a theory. In others econometricians search for theoretical explanations of observed regularities in their data. In still others economic researchers produce evaluations of performance and forecasts for business executives and government policy makers. In this paper I shall discuss theory-data confrontations in which the RIC searches for a scientific explanation of regularities in his data that economists like to identify as *stylized facts*; e.g., characteristic differences in cross-section and time-series estimates of the consumption function, (Friedman, 1957) and (Modigliani and Brumberg, 1955), or cointegration among yields on U.S. Treasury Bills (Hall *et al*, 1992).

An explanation is an answer to a why question. It makes clear and intelligible something that is not known or understood by the person asking the question. A scientific explanation is an explanation in which the ideas of a scientific theory play an essential role. In economics this scientific theory is an economic theory, and the ideas of the theory are used to provide scientific explanations of regularities in RIC's data.

The form in which the causes of events and the reasons for observed phenomena are listed and used in scientific explanations will differ among scientists, even within the same discipline. There is, therefore, a need for formal criteria by which we can distinguish good from bad scientific explanations. These criteria must list the necessary ingredients of a scientific explanation and explicate the ideas of a logically adequate and an empirically adequate scientific explanation. I provide formal criteria for scientific explanations in economics and econometrics and exemplify their use in a laboratory test and in an analysis of the Treasury Bill market.

1.4.1 *Hempel's Deductive Nomological Scheme for Scientific Explanations*

The most influential formal account of scientific explanation is Carl G. Hempel's deductive-nomological scheme (DNS) for scientific explanations (Hempel, 1965, pp. 245-251). According to Hempel, a scientific explanation of an event or a phenomenon must have four ingredients: (1) A sentence, E, that describes the event or phenomenon in question; (2) a list of sentences, $C_1,\ldots,C_n$, that describes relevant antecedent conditions; (3) a list of general laws, $L_1,\ldots,L_k$; and (4) arguments that demonstrate that E is a logical consequence of the Cs and the Ls. Such an explanation is adequate if it is both logically and empirically adequate. An explanation is logically adequate if at least one L plays an essential part in the demonstration that the *explanandum* E is a logical consequence of the *explanans*, the family of C's and L's. Moreover the explanation is empirically adequate only if (1) it is possible, at least in principle, to establish by experiment or observation whether the C's that are used in condition 4 are satisfied, and (2) the L's that are used in condition 4 have been subjected to extensive tests and have passed them all.[5]

E 1.2: I am to explain why Per, a human being, cannot live forever. To do that I let P denote Per and let E(x) insist that x is mortal. Then E becomes E(P). Next, I let C(x) assert that x is a human being and observe that C(P). Finally, I let L denote a law of biology which insists that $(\forall x)[C(x) \supset E(x)]$. Now, in first-order logic L implies that $[C(P) \supset E(P)]$, and C(P) and $[C(P) \supset E(P)]$ imply that E(P). Consequently, E(P) is a logical consequence of C(P) and L as required in condition 4 of Hempel's scheme.

The scientific explanation of E(P) in E 1.2 is logically adequate since L plays an essential role in the proof that E(P) is a logical consequence of C(P) and L. The explanation is also empirically adequate since observation alone suffices to determine whether P satisfies C, and since history accounts for numerous tests of L all of which have failed to falsify L.

Hempel's DNS has attractive features, and Hempel and others have used it to give interesting scientific explanations of events and phenomena in different sciences (cf. for example, Hempel, 1965, pp. 335-338 and Nagel,1961, pp. 30-32). Even so, influential philosophers of science have criticized it for many failings and questioned the possibility of applying it in the social sciences because of the paucity of general laws in these sciences.

A law is a general lawlike statement which insists that individuals or objects of a given kind must have certain properties or that individuals or objects of different given kinds must be related in a certain way. Also a general lawlike statement is a law only if it is a logical consequence of an accepted scientific theory and its truth value is independent of place and time. Examples are: "Copper expands when heated" (CEH) and "When electrones and antielectrones meet head on, they annihilate each other" (EAN). CEH is a theorem in thermodynamics and statistical mechanics, and EAN is a theorem in quantum mechanics. Both theorems are believed to be valid irrespective of place and time.

There are many economic theories and all sorts of interesting economic theorems. Therefore, the claim that there are no laws in economics amounts to insisting that there are no theorems in economics whose empirical relevance is independent of time and place. This claim cannot be true. To wit: Two of Nassau Senior's fundamental postulates of economics, the first and the last, Gresham's law, and Paul Samuelson's Fundamental Theorem of consumer choice (cf. Stigum, 1990, pp. 4, 554, and 186) are theorems of economics that economists believe are valid irrespective of place and time. How prevalent such laws are in economics is, however, uncertain.

In the notion of a scientific explanation that I shall develop in this paper I will allow any number of economic theories to play the role of a law in Hempel's scheme. Consequently, the demands for logical and empirical adequacy of a scientific explanation on which I shall insist will appear much weaker than the DNS adequacy requirements. Still my criteria for the empirical adequacy of a scientific explanation cannot differ that much from the criteria on which philosophers of science, today, insist. Here are two observations in support of my contention: (i) A law of nature is not an assertion that has a truth value. It is a statement that comes with a list of situations in which it has proven possible to apply it (Toulmin, 1953, pp. 86-87).
(ii) Scientists apply theories selectively in scientific explanations. For example, they say that Newton's theory can be used to explain the tides even though they know that Newton's laws do not satisfy Hempel's critera for empirical adequacy. Whether a theory explains some fact or other, is independent of whether the real world, as a whole, fits the theory (Van Frassen, 1980, p. 98).

1.4.2 *Empirical Relevance and Hempel's Symmetry Thesis*

A scientific explanation is empirically adequate only if the theory that is used in the explanation is relevant in the empirical context in which the explanation is carried out. In this section I explicate the two notions, empirical context and empirical relevance, in theory-data confrontations in which the data generating process (DGP) is deterministic; that is, in the kind of confrontations for which Hempel developed his DNS. In Section 3 I shall discuss the meaning of the same two notions in situations where the data generating process is random; that is, in the kind of situations that econometricians usually face. .

When the data generating process is deterministic, the *empirical context* in which the theory-data confrontation takes place is a pair, an accurate description of a sampling scheme and a family of data admissible models of the data universe, (Ω_P, Γ_p). A model of (Ω_P, Γ_p) is *data admissible* if the actual data in the pertinent theory-data confrontation satisfy the strictures on which the model insists.

Next, let (Ω_T, Γ_t) and $\Gamma_{t,p}$ denote the theory universe and the bridge principles in the present theory-data confrontation. Also, let IM_p denote the intended family of data admissible models of the data universe and let $\Im_p$ denote the family of collections of sentences about individuals in the data universe that are logical consequences of Γ_t, Γ_p, and $\Gamma_{t,p}$. There is one collection of such sentences for each member of the intended interpretation of $(\Gamma_t, \Gamma_p, \Gamma_{t,p})$. *The pertinent theory, is relevant in the empirical context that IM_p delineates if and only if there is at least one member of the intended interpretation of Γ_t, Γ_p, $\Gamma_{t,p}$ such that the corresponding member of $\Im_p$ has a model that is a member of IM_p .*

An interpretation of the theory universe comprises a family of models of Γ_t. The requirement that there be a member of the intended interpretation of Γ_t, Γ_p, $\Gamma_{t,p}$ such that the corresponding members of $\Im_P$ *are valid in some member of IM_P* singles out one model of (Ω_T, Γ_t) that is empirically relevant in the given theory-data confrontation. I believe that the RIC ought to search for a as large as possible family of models of the theory universe that satisfies the required condition.

In theory-data confrontations in which $\Gamma_{t,p}$ has many models it is often the case that the RIC ends up checking whether there is a model of $\Gamma_{t,p}$ that will allow him to conclude that his theory has empirical relevance. This may sound strange, but is in accord with the condition of empirical relevance that I delineated above. The next example shows what I have in mind.

E 1.3 I believe that individuals rank random prospects according to their expected utility and that individual utility functions are linear. I also believe that people tend to overvalue low probabilities and undervalue high probabilities. To test these hypotheses I face a student of mine with a large number, N, of simple prospects, ask for his certainty equivalents of these prospects, and construct his utility function.

The axioms of my test concern one undefined term, *the Sample Space*, Ω, that must satisfy the following set of axioms. The definitional axiom is as follows:

U 1: There are sets, Ω_T and Ω_P of ordered (2N+1)-tuples, ω_T and ω_P, such that $\Omega \subset \Omega_T \times \Omega_P$.

The Γ_t axioms are:

U 2: $\omega_T \in \Omega_T$ only if $\omega_T = (p,x,U)$, where $(p,x) \in ([0,1]\times[0,1000])^N$ and $U(\cdot):[0,1000]\rightarrow[0,1]$.

U 3: For all $\omega_T \in \Omega_T$, $U(x_i)=p_i$, $i=1,...,N$, where (p_i,x_i) is the i^{th} component of (p,x).

U 4: For all $\omega_T \in \Omega_T$, $x_i=1000p_i$, $i=1,...,N$.

The Γ_p axioms are:

U 5: $\omega_P \in \Omega_P$ only if $\omega_P=(q,z,W)$, where $(q,z) \in ([0,1]\times[0,1000])^N$ and $W(\cdot):[0,1000]\rightarrow[0,1]$.

U 6: For all $\omega_P \in \Omega_P$, $W(z_i)=q_i$, $i=1,...,N$, where (q_i,z_i) is the i^{th} component of (q,z).

The $\Gamma_{t,p}$ axioms are:

U 7: For all $(\omega_T,\omega_P) \in \Omega$, $x_i = z_i$. $i=1,...,N$.

U 8: There is an $\alpha\in(0,0.5]$ such that, for all $(\omega_T,\omega_P) \in \Omega$ and $i=1,...,N$,

$$p_i = \begin{cases} \alpha + \alpha^{-2}(q_i-\alpha)^3 \text{ if } 0\leq q_i\leq\alpha \\ \alpha + (1-\alpha)^{-2}(q_i-\alpha)^3 \text{ if } \alpha<q_i\leq 1 \end{cases}$$

In the interpretation of the Γ_t axioms that I intend, U(·) is the utility function of a student, p_i is the student's perceived probability, and x_i is his certainty equivalent of a random prospect which promises 1000 with (perceived) probability p_i and 0 with (perceived) probability $(1-p_i)$. With this interpretation of U(·), p, and x, it follows from U 3 and U 4 that, for all $\omega_T \in \Omega_T$ and i=1,..,N, $U(x_i) = x_i/1000$.

In the intended interpretation of the Γ_P axioms, q_i is a probability with which I describe random prospects to my student, that is, 1000 with probability q_i and 0 with probability $(1-q_i)$. Also, for each q_i, z_i is the certainty equivalent of a random prospect that the student records. Finally, W(·) is the utility function that I will construct on the basis of the student's responses to my qs.

In the interpretation of the $\Gamma_{t,p}$ axioms that I intend, U 8 describes how a student's perceived probabilities vary with quoted probabilities. The axiom delineates one way in which the student might overvalue low probabilities and undervalue high probabilities. Finally U 7 insists that I will obtain accurate observations on x.

From the preceding axioms I derive T 1 to test the empirical relevance of U 2, U 3, and U 4. This theorem is a proposition concerning individuals in Ω_P.

T 1: Let $\alpha\in(0,0.5]$ be the α of U 8. Then, for all $(\omega_T,\omega_P) \in \Omega$ and i=1,...,N.

$$W(z_i) = \begin{cases} \alpha(1+([z_i-1000\alpha]/1000\alpha)^{1/3}) \text{ if } 0\leq z_i\leq 1000\alpha \\ \alpha(1+((1-\alpha)/\alpha)([z_i-1000\alpha]/1000(1-\alpha))^{1/3} \text{ if } 1000\alpha\leq z_i\leq 1000. \end{cases}$$

This theorem is true in some model of U 5 and U 6 but not in all models of these axioms. Also T 1 is true in all models of U 1-U 8. Hence, I can use T 1 to test the empirical relevance of U 2-U 4 by checking whether T 1 is true in one of the data admissible models of Γ_p that the intended interpretation of (Ω_P,Γ_p) contains. The intended interpretation of (Ω_P,Γ_p) in the present case is determined by my student's responses to my queries. A member of this interpretation is a model of Ω_P that contains the N values of q_i that I quoted, the corresponding certainty equivalents, z_i, that my student recorded, and a function, $W(\cdot):[0,1000]\rightarrow[0,1]$, that satisfies the conditions, $W(z_i) = q_i$, i = 1,…,N.

If there is no model of $\Gamma_{t,p}$; i.e., if there is no $\alpha\in(0,0.5]$, such that T 1 is valid in a member of the intended interpretation of (Ω_P,Γ_p), I must conclude that Γ_t, $\Gamma_{t,p}$ or both lack relevance in the given empirical context

The empirical relevance of the pertinent theory has interesting implications for my account of scientific explanation. According to Hempel, scientific explanation and prediction are two sides of the same coin. Specifically, in an adequate DNS in which the C's are valid statements the logical arguments that are used to establish the explanandum E constitute a potential predictor of the occurrence of E. In my account, the logical arguments in a logically adequate explanation of E demonstrate that E is valid in any world in which the bridge principles are valid. They do not show that E is valid in the Real World.[6] Hence, they cannot be used as potential predictors of E. Hempel's symmetry thesis concerning explanation and prediction holds in my account only if the explanation is both logically and empirically adequate; i.e., only if the theoretical arguments in the explanation are relevant in the empirical context in which the explanation is carried out.

2. **Scientific Explanation in Economics**

In this section of the paper I shall present a formal characterization of scientific explanations in economics for situations in which there is no sample population, S, and just one model in the intended interpretation of Γ_p. The situations I have in mind are analogues of the kind of situation for which Hempel developed his DNS; i.e. a situation in which the explanation can be based on physical or economic theories of deterministic character (cf. Hempel, 1965, p. 351). It is, therefore, interesting that my characterization of scientific explanations, with the proper translation, can be made to fit Hempel's scheme. I explain how in Section 2.3,

2.1 *The Characterization*

I have observed certain stylized facts in a set of data and search for a scientific explanation of these regularities. Formally, my search can be described as follows:

SE 1: Let (Ω_P,Γ_p) be some given data universe. Assume that the components of the vectors in Ω_P denote so many units of objects that have been observed and that Γ_p delineates the salient properties of these data. Also, let H be a finite family of assertions concerning Ω_P, and suppose that there is a data admissible model of (Ω_P,Γ_p), M, in which all the assertions of H are true. M is taken to be the intended model of the data universe and it is assumed that H delineates the characteristics of the data that are to be explained. Then, *to give a scientific explanation of H, means to find a theory universe, (Ω_T,Γ_t), a sample space, $\Omega \subset \Omega_T \times \Omega_P$,and a collection of bridge principles, $\Gamma_{t,p}$, that in Ω link Ω_T with Ω_P such that H becomes a logical consequence of Γ_t, Γ_p, and $\Gamma_{t,p}$.*

Such an explanation is *logically adequate*, if H is not a logical consequence of Γ_p alone. It is *empirically adequate* if M is a model of all the logical consequences of Γ_t, $\Gamma_{t,p}$, and Γ_p that concern components of ω_P.

In SE 1 there are several things to notice and to keep in mind for later discussion. Firstly, in SE 1, as in Hempel's DNS, the *explanandum* is a logical consequence of the *explanans*, and the explanation is logically adequate if at least one of the components of Γ_t plays an essential role in the proof of H. As in the case of Hempel's DNS, the logical adequacy of a scientific explanation alone does not in my scheme entail the symmetry thesis. The reason why is that I cannot assert $\Gamma_{t,p}$ and claim that the principles it comprises are valid. At best, I can insist that $\sim\Box\sim\Gamma_{t,p}$; i.e., that the given bridge principles are valid in some world, and hope that all the logical consequences of Γ_t , $\Gamma_{t,p}$, and Γ_p are valid in M.

Secondly, the empirical adequacy of an SE 1 scientific explanation hinges on the predictive powers of Γ_t. Specifically, the scientific explanation of H is empirically adequate if Γ_t is relevant in the empirical context that M determines. It is, therefore, not strange that the required empirical relevance of Γ_t entails that a logically and empirically adequate scientific explanation of H could have been used to predict the happening of H. Hempel' symmetry thesis does apply to logically and empirically adequate SE 1 explanations.

Thirdly, both the empirical and the logical adequacy of an SE 1 scientific explanation hinge on the validity of the bridge principles that are adopted. This is an important aspect of SE 1 the meaning of which I explicate in discussing the modal-logical rendition of SE 1 in section 22.3. in (Stigum, 2003).

2.2 *An Example*[7]

My explication of scientific explanation might sound unfamiliar. So here is an example to fix ideas. Suppose that I, for some student, have constructed the function, $W(\cdot):[0,1000] \rightarrow [0,1]$, in the way I described the construction of $W(\cdot)$ in E 1.3, and suppose that this new $W(\cdot)$, at each observed x, satisfies equation (1).

$$W(x) = 0.5(1+((x-500)/500)^{1/3}),\ x\ \varepsilon\ [0,1000]. \qquad (1)$$

Suppose also that I, for the same student, have constructed a function, $V(\cdot)$: $[0,1000] \rightarrow [0,1]$, in the following way: I assigned the values 0 and 1, respectively, to $V(0)$ and $V(1000)$. Also, whenever $V(\cdot)$ had been defined at x and y, I let $V(C(x,y)) = (1/2)V(x) + (1/2)V(y)$, where $C(x,y)$ equaled the student's certainty equivalent of the option, x with probability 1/2 and y with probability 1/2. Finally, suppose that, for all $x\ \varepsilon\ [0,1000]$ at which I have defined $V(\cdot)$, I have found that it satisfies the relation:

$$V(x) = x/1000,\ \ x\ \varepsilon\ [0,1000]. \qquad (2)$$

Now, I am asked to give a scientific explanation of equations (1) and (2) and the inequalities in equation (3).

$$V(x) \text{ is } \begin{cases} > W(x) \text{ if } x\ \varepsilon\ (0, 500), \text{ and} \\ < W(x) \text{ if } x\ \varepsilon\ (500,1000). \end{cases} \qquad (3)$$

To do that I must describe the data universe, (Ω_P,Γ_p), formulate the assertion to be explained, H, find the required theory universe, (Ω_T,Γ_t), and delineate the bridge principles that are to connect Ω_T with Ω_P.

The Data Universe and the Explanandum, H

I begin by letting Ω_P be a set of seventuples, ω_p, that satisfy the following axioms:

EB 1: $\omega_P \varepsilon \Omega_P$ only if $\omega_P = (q,x,W,y,z,C,V)$, where $q \varepsilon [0,1]$, $x,y,z \varepsilon [0,1000]$, $W(\cdot):[0,1000] \rightarrow [0,1]$, $y \leq z$, $V(\cdot): [0,1000] \rightarrow [0,1]$, $V(0) = 0$, $V(1000) = 1$, and $C(\cdot):[0,1000]^2 \rightarrow [0,1000]$.

EB 2: For all $\omega_P \varepsilon \Omega_P$, $W(x) = q$, and $V(C(y,z)) = (1/2)V(y) + (1/2)V(z)$.

EB 3. If $(q,x,W,y,z,C,V) \varepsilon \Omega_P$, then $(q,x,W,0,1000,C,V) \varepsilon \Omega_P$. Also, there exist two uniquely determined pairs of numbers in [0,1000], (v_y,w_y) and (v_z,w_z), that are independent of q and x and satisfy the relations, $y = C(v_y,w_y)$, $z = C(v_z,w_z)$, $(q,x,W,v_y,w_y,C,V) \varepsilon \Omega_P$, and $(q,x,W,v_z,w_z,C,V) \varepsilon \Omega_P$. If $y = 0$, $(v_y,w_y) = (0,0)$, and if $z = 1000$, $(v_z,w_z) = (1000,1000)$. Otherwise, $v_y < w_y$ and $v_z < w_z$.

When reading these axioms, observe that the uniqueness of the pairs, (v_y,w_y) and (v_z,w_z), on which I insist in EB 3, is a characteristic of the sampling scheme that I used in the construction of $V(\cdot)$. It is not a property of certainty equivalents as such. With that in mind, I can state the explanandum, H, as follows :

H: For all $\omega_P \varepsilon \Omega_P$, (i) the value of $W(\cdot)$ at x satisfies equation (1) with q instead of p; (ii) the values of $V(\cdot)$ at y, z, and C(y,z) satisfy equation (2); and (iii) the value of $W(\cdot)$ at x and the value of $V(\cdot)$ at C(y,z) satisfy the inequalities in equation (3) whenever $x = C(y,z)$.

This assertion is true in some model of (Ω_P,Γ_p), but it is not a logical consequence of EB 1, EB 2, and EB 3.

The Theory Universe .

Next I must search for a useful theory universe, (Ω_T,Γ_t). Here is one possibility. Let Ω_T be a set of sextuples, ω_T, that satisfy the following axioms:

EB 4: $\omega_T \varepsilon \Omega_T$ only if $\omega_T = (p,r,U,s,t,\hat{C})$, where $p \varepsilon [0,1]$, $r,s,t \varepsilon [0,1000]$, $U(\cdot): [0,1000] \rightarrow [0,1]$, $s \leq t$, and $\hat{C}(\cdot):[0,1000]^2 \rightarrow [0,1000]$.

EB 5: For all $\omega_T \varepsilon \Omega_T$, $U(r) = p$ and $U(\hat{C}(s,t)) = (1/2)U(s) + (1/2)U(t)$.

EB 6: For all $\omega_T \varepsilon \Omega_T$, $r = 1000p$, and $\hat{C}(s,t) = (1/2)s + (1/2)t$.

In these axioms the triple (p,r,U) plays the same role as the triple, (p,x,U), played in section 17.3.1 (cf. axioms U 2-U 4). Also, $\hat{C}(s,t)$ is to be interpreted as the certainty equivalent of the random option, s with perceived probability 1/2 and t with perceived probability 1/2. In its intended interpretation, (Ω_T, EB 4-EB 6) is the universe of a theory in which the decision maker orders prospects according to their perceived expected value.

The Bridge Principles

Finally, I must describe how, in the sample space, the individuals in Ω_T are related to the individuals in Ω_P. I shall insist that I have accurate observations on r,s,t, and $\hat{C}$. Also, I shall assume that the given student's perceived probabilities**,** in an appropriate way, overvalue low probabilities and undervalue high probabilities.

EB 7: The sample space, Ω, is a subset of Ω_T x Ω_P.

EB 8: For all $(\omega_T,\omega_P) \varepsilon \Omega$, $r = x$, $s = y$, $t = z$, and $\hat{C}(s,t) = C(y,z)$.

EB 9: For all $(\omega_T,\omega_P) \varepsilon \Omega$, $p = 0.5 + 4(q-0.5)^3$.

EB 10: If $(\omega_T,\omega_P) \varepsilon \Omega$ and $(\omega_T,\omega_P) = (p,r,U,s,t,\hat{C},q,x,W,y,z,C,V)$, then $(p,r,U,0,1000,\hat{C},q,x,W,0,1000,C,V) \varepsilon \Omega$, $(p,r,U,v_y,w_y,\hat{C},\mathbf{q},x,W,v_y,w_y,C,V) \varepsilon \Omega$, and $(p,r,U,v_z,w_z,\hat{\mathbf{C}},q,x,W,v_z,w_z,C,V) \varepsilon \Omega$, where the pairs, (v_y,w_y) and (v_z,w_z), are as described in EB 3.

If I pick (Ω_T,Γ_t) and $\Gamma_{t,p}$ as described above, I can show by simple algebra that, for all $(\omega_T,\omega_P) \varepsilon \Omega$, the value of $W(\cdot)$ at x must satisfy equation (1). Moreover, I can show, first by simple algebra, that $C(y,z) = (1/2)y + (1/2)z$, and then, by an obvious inductive argument, that the value of $V(\cdot)$ at y,z, and C(y,z) must satisfy equation (2). But if that is so, the value of $W(\cdot)$ at x and the value of $V(\cdot)$ at C(y,z) must satisfy the inequalities in equation (3) whenever x = C(y,z). Hence, (Ω_T,Γ_t) and $\Gamma_{t,p}$ provide the required scientific explanation of H.

The explanation of H that I have delineated is obviously logically adequate. Since I have no relevant data, I cannot determine whether the explanation is empirically adequate.

2.3 *Sundry Remarks*

There are several striking features of the preceding example of a scientific explanation that I should point out. According to expected utility theory (EUT), the relation, V(x) = W(x), $x \varepsilon [0,1000]$, must hold if the student's perception of probabilities are veridical. This is true, moreover, regardless of whether W(·) is linear. Also, the relationship between V(·) and W(·) depicted in equation (3) seems to be a general feature; i.e., a stylized fact, of experimental results that Maurice Allais and his followers have recorded during the last fifty years (cf. for example Allais 1979, pp. 649-654). Note, therefore, that I use EUT to provide a scientific explanation of equations (1)- (3). In contrast, Allais and his followers have used the necessity of the equality of V(·) and W(·) and equation (3) to discredit the descriptive power of EUT in risky situations. My analysis shows that they had good reasons for their disbelief only if their subjects' perception of probabilities was veridical.

A logically and empirically adequate SE 1 explanation differs in interesting ways from an adequate DNS explanation. Hempel's Cs, Ls, and E concern individuals in one and the same universe. This universe is in SE 1 my data universe. There the members of Γ_p play the roles of Hempel's Cs, the translated versions of the members of Γ_t play the roles of Hempel's Ls, and H has taken the place of Hempel's E. Hence, with the proper translation, an SE 1 explanation can be made to look like a DNS explanation. Still, there is a fundamental difference. Hempel's criteria for a DNS explanation to be empirically adequate insist that his Ls must have been subjected to extensive tests and have passed them all. I only insist that my Ls, the members of Γ_t, be relevant in the given(!) empirical context in which the explanation is carried ouit.

There is an interesting second way in which an SE 1 explanation differs from a DNS explanation. This difference enables a logically and empirically adequate SE 1 explanation to reason away one of the most serious problems that Hempel's DNS has faced (cf. my remarks at the end of section 22.1). The problem arises when the same question calls for different answers depending on the circumstances in which it is asked. Then an adequate DNS explanation may be good in one situation and meaningless in another. In logically and empirically adequate SE 1 explanations this problem does not arise. The reason why is that the meaning of H in SE 1 varies with the empirical context, M. H may have the same meaning in two different Ms. Also, the same M may be a model of two different Γ_ps. In SE 1 H and M are given in advance. Different Ms and/or different Γ_ps might call for different Γ_ts and $\Gamma_{t,p}$s in the explanation of H.

2.4 *A Potential Criticism*

Judging from my reading of Bas Van Frassen's wonderful book on **The Scientific Image** (Van Frassen, 1980), Van Frassen would not be satisfied with SE 1 and its demands for logical and empirical adequacy. He would agree that it is correct to require that the scientific theory on which the explanation is based be empirically relevant in the context in which the explanation is carried out. However, in order that this requirement be meaningful, the description of the "context" must be adequate. Van Frassen would fault my description on two accounts. It fails to provide criteria by which a person can judge the contextual relevance of the theory. It also fails to specify a contrast class for H. These qualms about SE 1 are interesting. So I shall take time to explain what Van Frassen has in mind.

First, the contextual relevance of the theory in SE 1: Van Frassen believes that a person who is asked to explain the occurrence of an event or an observed phenomenon will start by looking for salient features of the cause of the event or salient reasons for the existence of the phenomenon. What appears salient to a given person depends on his orientation, his interests, and various contextual factors. For example, the cause of a youngster's death might be 'multiple hemorrhage' to a physician, 'negligence on the part of the driver' to a barrister, and 'a defect in the construction of the car brakes' to a car mechanic (Van Frassen, 1980, p. 125). Therefore, the reasons why one theory is chosen instead of another ought to appear in the description of the context in which the given person carries out his scientific explanation.

Then the required contrast class for H. A *contrast class* for the *explanandum* is a finite family of assertions, {H, A, B, …, K}, that concern individuals in the data universe and have the property that H and the negations of all the other assertions in the class are logical consequences of the pertinent *explanans.* Van Frassen believes that the singling out of salient causal factors of an event or salient reasons of a phenomenon depends on the range of contrasting alternatives to the *explanandum*. For example, it makes a difference in explaining why a given person has paresis whether the contrast class is his brother or the members of his country club all of whom have a history of untreated syphilis (Van Frassen, 1980, p. 128). By specifying a contrast class, the scientist can communicate to interested parties what question he is out to explain.

The given ideas of Van Frassen about the pragmatics of scientific explanation have important bearings on SE 1. I shall mention two of them. Firstly, the axioms of the theory in SE 1 need not constitute more than a small part of the axioms of a complete theory. For example, in 2.1 the axioms of the relevant complete theory are the axioms of von Neumann and Morgenstern's expected utility theory together with the additional assumption that the utility function is linear. The axioms of the theory universe in 2.1 postulated only the latter assumption. By delineating an appropriate contrast class for H, I can introduce the complete theory, so to say, via the back door. In 2.1, for example, I can insist that the contrast class of H is{H, [W(x) = x/1000], [W(x) = V(x)]}. Besides H, this contrast class contains two assertions on which the expected utility theory insists.

Secondly, in my theory of scientific explanations I have been content to come up with one possible explanation of the *explanandum*. It is often the case that different scientific theories can be used to provide an explanation of one and the same *explanandum*. If two different theories can explain a given H, then some of the predictions of one are likely to constitute a contrast class for the other and *vice versa*. The two contrast classes might give us reasons to prefer one of the explanations over the other. Such uses of contrast classes appear in my discussion of testing one theory against another in chapter 18 of (Stigum, 2003).

3. Scientific Explanation in Econometrics

The situation envisaged in SE 1 is similar to the experimental tests of physical theories which Pierre Duhem described in his book, **The Aim and Structure of Physical Theory** (cf. Duhem, 1954, pp.144-147). It differs, however, from the situations econometricians usually face when they search for the empirical relevance of economic theories. In SE 1 H is a family of sentences each one of which has a truth value in every model of (Ω_P,Γ_p), and all of which are true in some model of (Ω_P,Γ_p). In contrast, in enconometrics H is often a family of statistical relations. One H might insist that "on the average, families with high incomes save a greater proportion of their incomes than families with low incomes." Another H might claim that "the prices of soybean oil and cottonseed oil vary over time as if they were two cointegrated ARIMA processes." These assertions are about properties of the data generating process. They need not have a truth value in a model of (Ω_P,Γ_p).

When H is a family of statistical relations, a scientific explanation of H must be based on statistical arguments. Before I can describe the characteristics of scientific explanation in such cases I must, first, sketch the formal structure of theory-data confrontations in which the data generating process is random and discuss what it means for the relevant theory to be empirically relevant in such a case.

3.1 *Theory-data Confrontations with random data generating Processes*

.

The structure of a theory-data confrontation in which the data generating process is stochastic has four parts. The first three are the two universes and the bridge between them. However, in this case the data universe is not taken to be an independent entity. Instead, it forms one part of a triple, $((\Omega_P,\Gamma_p), \Im p, Pp(\cdot))$, where $\Im_p$ is a family of subsets of Ω_P and $P_p(\cdot)$: $\Im_p \rightarrow [0,1]$ is a probability measure. Subject to the conditions on which Γ_p insists, $P_p(\cdot)$ induces a joint probability distribution of the components of ω_P that I denote by FP. In a formalized theory-data confrontation with a random data generating process, FP plays the role of the *true* probability distribution of the components of ω_p.

The fourth part of the formal structure comprises three elements: a sample population, S, a function, $\Psi(\cdot)$:S$\rightarrow\Omega$, and a probability measure, $P(\cdot)$:$\aleph\rightarrow$ [0,1] on subsets of Ω. I imagine that the pertinent econometrician is sampling individuals in S, for example, firms or historical data. S is usually a finite or a countably infinite set. With each point $s \in S$ I associate a pair of vectors, $(\omega_{Ts},\omega_{Ps}) \in \Omega$. I denote this pair by $\Psi(s)$ and insist that $\Psi(s)$ is the value at s of the function, $\Psi(\cdot)$. I also imagine that there is a probability measure, $Q(\cdot)$, on a σ-field of subsets of S, $\aleph_S$, that determines the probability of observing an s in the various members of $\aleph_S$. The properties of $Q(\cdot)$ are determined by $\Psi(\cdot)$ and the probability measure, $P(\cdot)$, that assigns probabilities to subsets of Ω in $\aleph$. Specifically, $\aleph_S$ is the inverse image of $\aleph$ under $\Psi(\cdot)$ and, for all $B \in \aleph$,

$$Q(\Psi^{-1}(B)) = P(B \cap \text{range}(\Psi))/P(\text{range}(\Psi)),$$

where $\text{range}(\Psi) = \{\omega \in \Omega : \omega = \Psi(s) \text{ for some } s \in S\}$. Finally, I imagine that the econometrician has obtained a sample of N observations from S, $s_1,\ldots, s_N$, in accordance with a sampling scheme, ξ. $P(\cdot)$ and ξ determine a probability distribution on subsets of Ω^N. The marginal distribution of the components of ω_P determined by $P(\cdot)$ and $\Gamma_{t,p}$ I denote by MPD. The marginal distribution of the sequence of ω_{Ps}s determined by $P(\cdot)$, $\Gamma_{t,p}$ and ξ, I refer to as the pseudo data generating process and denote it by PDGP. The *true* data generating process I identify with the joint distribution of the ω_{Ps}s that is induced by FP and ξ, and I denote it by DGP.[8]

Strictly speaking, in a formalized economic theory-data confrontation in which the data generating process is taken to be stochastic, the *empirical context* in which the theory-data confrontation takes place is a triple, an accurate description of a sampling scheme, the intended interpretation of the data universe, (Ω_P, Γ_p), and the *true* probability distribution of the components of ω_P, FP. I shall refer to this empirical context by the name, Real World. The empirical context in which the theory-data confrontation *actually* takes place is a different triple, a not-necessarily accurate description of a sampling scheme, the intended interpretation of (Ω_P, Γ_p), and an interpretation of MPD, the marginal probability distribution of the components of ω_P. I like to think of this empirical context as one of the worlds in which all the bridge principles are valid.

As to the relevance of the theory. Let the core structure of the theory-data confrontation be as described above, and let IM_p and IM_{MPD} , respectively, denote the intended family of data admissible models of the data universe and the MPD. For simplicity I shall assume that the members of IM_{MPD} are parametric. Next, let $\Im_p$ denote the family of collections of sentences about individuals in the data universe that are logical consequences of Γ_t, Γ_p, and $\Gamma_{t,p}$. There is one collection of such sentences for each member of the intended interpretation of (Γ_t, Γ_p , $\Gamma_{t,p}$). Finally, let $\Im_{MPD}$ denote the family of collections of sentences about the parameters of the MPD that are logical consequences of Γ_t, Γ_p, $\Gamma_{t,p}$ and the axioms that determine the characteristics of P(·). Each member of $\Im_{MPD}$ is a collection of sentences about the parameters of the MPD that is determined by a member of the intended interpretation of Γ_t, Γ_p, $\Gamma_{t,p}$ and the axioms that determine the characteristics of P(·). Now, *the theory is relevant in the empirical context that IM_p and IM_{MPD} delineate if and only if* there is at least one member of the intended interpretation of Γ_t, Γ_p, $\Gamma_{t,p}$ and the axioms that determine the characteristics of P(·) such that *(1) the corresponding member of $\Im_p$ has a model that is a member of IM_p , and (2) the corresponding member of $\Im_{MPD}$ is valid in some member of IM_{MPD}.*

3.2 A Characterization of Scientific Explanation for Econometrics

With that much said about the formal structure of a theory-data confrontation in which the data generating process is random I can provide a succinct characterization of scientific explanations in such cases. To wit SE 2.

SE 2: Let (Ω_P,Γ_p) be some given data universe; let $\aleph_P$ be a σ-field of subsets of Ω_P; let $P_p(\cdot)$: $\aleph_P \rightarrow [0,1]$ be a probability measure; and let FP denote the joint probability distribution of the components of the vectors in Ω_P which, subject to the conditions on which Γ_p insists, is determined by $P_p(\cdot)$. Also, let H_1 and H_2 , respectively, be a finite family of assertions concerning the characteristics of the vectors in Ω_P and the FP, and let $\Im_1$ and $\Im_2$, respectively, be a family of data admissible models of H_1 and (Ω_P,Γ_p) and of H_2 and the FP. Finally, suppose that $\Im_1$ is the intended interpretation of H_1 and (Ω_P,Γ_p) and that $\Im_2$ is the intended interpretation of H_2 and the FP. Then*, to give a scientific explanation of the pair, (H_1 , H_2), means to find a theory universe, (Ω_T,Γ_t), a sample space, $\Omega \subset \Omega_T \times \Omega_P$, a finite set of bridge principles, $\Gamma_{t,p}$, that in Ω relates members of Ω_T to members of Ω_P, a probability* *measure, P(·), on subsets of Ω, and a probability distribution, MPD, of the components of the vectors in* Ω_P *that is determined by* Γ_t *,* Γ_p*,* $\Gamma_{t,p}$ *, and the axioms of P(·) such that* H_1 *becomes a logical consequence of* Γ_t *,* Γ_p*, and* $\Gamma_{t,p}$ *, and such that the given MPD has the characteristics of FP on which* H_2 *insists.*

Such an explanation is *logically adequate* if the pair, (H_1 , H_2), is not a logical consequence of Γ_p and the axioms of $P_p(\cdot)$. The explanation is *empirically adequate* if there is a model of Γ_t, $\Gamma_{t,p}$, Γ_p, and the axioms of P(·) and an associated model of the MPD that have the following properties: (1) The logical consequences of Γ_t, $\Gamma_{t,p}$, and Γ_p that concern characteristics of the vectors in Ω_P has a model that is a member of $\Im_1$; and (2) there is a member of $\Im_2$ whose associated FP shares with the model of MPD the characteristics on which the logical consequences of Γ_t, $\Gamma_{t,p}$, Γ_p, and the axioms of P(·) insist.

It might seem strange that the sample population, S, plays no role in SE 2. The reason is simple. In SE 2 I imagine that I have data of one of three kinds: (i) Cross-section data; i.e., a finite sequence of vectors of observations on different individuals that pertain to some given point in time; (ii) time series of vectors of observations on some individual or some aggregate of individuals; and (iii) panel data; i.e., a time series of vectors of observations on more than one individual. For the purposes of the explanation in SE 2 the data are given, and so are H_1 and H_2. Hence, in an SE 2 explanation I can work directly with P(·) and $P_p(\cdot)$ without involving Q(·), $\aleph_S$, and S.

I will need Q(·), $\aleph_S$, and S when I investigate the empirical adequacy of an SE 2 explanation. Since I do not know $\Im_1$ and $\Im_2$, I cannot establish that a given scientific explanation is empirically adequate. The best I can hope for is that I have reasons to believe that the likelihood that the explanation is empirically adequate is high. To do that I use data to construct a data admissible model of the MPD and to delineate the contours of a 95% confidence band around the parameters of this MPD. A large enough family of data admissible models of the MPD whose parameters belong to this confidence band is likely to contain an MPD whose characteristics some model of the FP in $\Im_2$ shares.

3.3 *An Example*

It is a difficult task to give logically and empirically adequate SE 2 explanations of regularities in the kind of data I have in mind for SE 2. So a detailed example is called for. In this section I shall recount an SE 2 explanation in (Stigum, 2003, chapter 23) that Heather Anderson, Geir Storvik and I (ASS) gave of the following stylized facts that A. Hall, H. Anderson, and C.L.M. Granger (HAG) have discovered (cf. Hall, Anderson, and Granger, 1992).[9]

(1) "Yields to maturity of U.S. Treasury bills are cointegrated I(1) processes;"

(2) "during periods when the Federal Reserve specifically target short-term interest rates, the spreads between yields of different maturity define the cointegrating vectors."

The Treasury bill market is part of the Money market in the U.S.A.. The yields in the Money market are interrelated. Consequently, when one delineates the relevant positive analogies for the functioning of the Treasury bill market, one must take into account how the functioning of the remainder of the Money market influences the determination of yields in the bill market.

There are many different Money market instruments even when one distinguishes them just by the name of the issuer and the kind of issue; e.g., Treasury versus General Electric, and three months bills versus six months bills. There are fabulously many more when one distinguishes instruments by maturity as well. For the purposes of ASS's analysis it seemed unnecessary to take into account this multiplicity of Money market instruments. So, to keep arguments clear and simple, ASS argued as if there were just two bills and just one other Money market instrument. In due course it will appear that their arguments' gain in clarity did not come at the expense of a reader's loss in insight.

The ASS Data Universe

The individuals in HAG's own data universe were series of daily bid and asked quotes on eleven Treasury bills; one series for bills with one month to maturity, another for bills with two months to maturity, and so on to a series with eleven months to maturity. From these series they obtained eleven nominal yield-to-maturity series by taking the average bid and asked quotes of the day as the prices of the respective bills and by insisting that the length of a month be 30.4 days.

HAG's sample consisted of 228 observations on each yield series, dating from January 1970 until December 1988. ASS's sample consisted of 150 observations on the same yields dating from January 1983 until June 1995. Each observed yield pertained to the last trading day of the month and was taken from the Fama Twelve Month Treasury Bill Term Structure File of the Center for Research in Securities Prices at the University of Chicago.

For the sake of brevity ASS took the monthly yield series, instead of the bid and asked quotes, to be basic elements in their data universe, (Ω_P, Γ_p). Also, for the sake of clarity and simplicity, ASS assumed that they had observations only on bills of two different maturities, Ḱ(1) and Ḱ(2); e.g., three and six months bills, or bills that will mature in one and two months. Since ASS had monthly observations, they let Ḱ(1) and Ḱ(2) denote series of yields of bills that, respectively, mature in one and two months.

In addition to their observations on the two Treasury bill yields ASS had a corresponding series of monthly observations on the overnight Federal Funds rate. Each observation recorded the weekly effective (annualized) rate on overnight Federal Funds. Data (from 1984 onwards) was obtained from the Federal Reserve, and earlier data was obtained from the Federal Reserve Bulletin. Also, the weekly data was converted to monthly observations by using the last observation for each month. I shall denote the series of observations on the Federal Funds rate by ff.

With these assumptions in mind, I need only three axioms to characterize the individuals in ASS's data universe. In the intended interpretation of the axioms the $\hat{y}_j$, j = 1,2,3 are auxiliary series that are needed to delineate the properties of ASS's data. Also, the numbers in N are taken to denote consecutive 'months' beginning at some arbitrary point in time.

D 1: $\omega_P \in \Omega_P$ only if $\omega_P = (Ḱ(1), Ḱ(2), ff, \hat{y}_1, \hat{y}_2, \hat{y}_3)$, where $Ḱ(j) \in (R_+)^N$, j = 1,2, $ff \in (R_+)^N$, and $\hat{y}_j \in (R)^N$, j = 1, 2, 3; and N = {0,1,2,...}.

D 2: For each $\omega_P \in \Omega_P$ and all t ε N,

$\acute{K}(1)_t = \max(\hat{y}_{1t},0)$, $\acute{K}(2)_t = \max(\hat{y}_{2t},0)$, and $ff_t = \max(\hat{y}_{3t},0)$, (4)

where $\acute{K}(1)_t$, $\hat{y}_{1t}$, $\acute{K}(2)_t$, $\hat{y}_{2t}$, ff_t, $\hat{y}_{3t}$ denote the t^{th} component of $\acute{K}(1)$, $\hat{y}_1$, $\acute{K}(2)$, $\hat{y}_2$, ff, and $\hat{y}_3$.

Instead of thinking of ω_P as a sixtuple of series, one can think of ω_P in Ω_P as a vector valued function, $(\acute{K}(\cdot,\omega_P), ff(\cdot,\omega_p), \hat{y}(\cdot,\omega_P))$: $N \to R_+^2 \times R_+^1 \times R^3$, defined by

$$(\acute{K}(t,\omega_P), ff(t, \omega_P), \hat{y}(t,\omega_P)) = \omega_{Pt}\ , t \in N, \text{ and } \omega_P \in \Omega_P, \qquad (5)$$

where ω_{Pt} is the t^{th} component of ω_P. If one thinks of the elements of Ω_P in that way, I can state the third axiom concerning the individuals in Ω_P as follows:

D 3: Let Ω_P be as above, and let the vector-valued function, $(\acute{K}, ff, \hat{y})(\cdot)$: $N \times \Omega_P \to R_+^2 \times R_+^1 \times R^3$, be as described in equation 5. Also, let $\aleph_P$ be the standard Borel field of subsets of Ω_P. Then there exists a probability measure, $P_p(\cdot)$: $\aleph_P \to [0,1]$, such that, relative to $P_p(\cdot)$, the probability distributions of the family of random vectors, $\{(\acute{K}, ff, \hat{y})(t,\omega_P); t \in N\}$, equal the corresponding family of probability distributions of the process that generates the individuals in Ω_P.

With axioms D 1-D 3 in hand I proceed to formulate the assertion of which ASS intended to give a scientific explanation.

H. *Let $P_P(\cdot)$: $\aleph_P \to [0,1]$ be the probability measure on whose existence I insisted in D 3. Also, let the vector-valued function, $(\acute{K},ff,\hat{y})(\cdot)$: $N \times \Omega_P \to R_+^2 \times R_+^1 \times R^3$, be as described in equation 5. Then, relative to $P_p(\cdot)$, the family of random vectors, $\{(\acute{K},ff,\hat{y})(t,\omega_P); t \in N\}$, satisfies the following conditions:*

(i) for all t ε N, $(\acute{K},ff)(t,\omega_P) = \max(\hat{y}(t,\omega_P),0)$ a.e.;

(ii) $\{\hat{y}(t,\omega_P); t\ \varepsilon\ N\}$ is a vector-valued ARIMA process with one unit root; and

(iii) the two first components of $\{\hat{y}(t,\omega_P); t\ \varepsilon\ N\}$ are cointegrated with cointegrating vector (-1,1,0); i.e., $\{\hat{y}_2(t,\omega_P) - \hat{y}_1(t,\omega_P); t\ \varepsilon\ N\}$ is a wide- sense stationary process.

This H does not sound quite like HAG's dictum. However, if one interprets HAG's assertions with their footnote 5 in mind, one must end up with H as stated. In footnote 5 HAG observe that yields to maturity cannot be integrated processes in the strict sense because nominal yields are bounded below at zero while integrated processes are unbounded.

The ASS Theory Universe

Next I shall describe the theory universe in ASS's scientific explanation of H. For that purpose I need six axioms. In the statement of the axioms, with j = 1,2, the K(j), G(j), $\lambda(j)$, and Λ are to be interpreted as the theory universe's series versions of variables that at each $t \varepsilon N$ denote, respectively, the equilibrium yield at date t of a j-month pure discount bond ($K(j)_t$), the equilibrium rate of return from contracting at day t to buy a 1-month pure discount bond (j-1) months from day t ($G(j)_t$), the expected value at date t of $K(1)_{t+1}$ ($\lambda(1)_t$) and FF_{t+1} ($\lambda(2)_t$), and an error term. Also, η_j and u_j are series of error terms, and FF is the series of equilibrium yields to maturity of the one and only non-Treasury-bill security in the money market. One can think of the latter as the "Federal Funds Rate" in the ASS theory universe. Finally, N = {0,1,2,...}, and y_1, y_2, and y_3 are auxiliary series whose meanings are determined by the axioms. The members of N are taken to denote consecutive 'months' beginning at some arbitrary point in time.

The Axioms

B 1: $\omega_T \varepsilon \Omega_T$ only if $\omega_T \varepsilon (R^{16})^N$. and

$$\omega_T = (K(1), K(2), G(1), G(2), FF, \lambda(1), \lambda(2), \Lambda, \eta_1, \eta_2, y_1, y_2, y_3, u_1, u_2, u_3).$$

B 2: For each $\omega_T \varepsilon \Omega_T$ and $t \varepsilon N$,

$$K(1)_t = \max(y_{1t},0) \text{ and } G(1)_t = K(1)_t, \quad (6)$$

$$y_{2t} = (1/2)[y_{1t} + G(2)_t], \text{ and} \quad (7)$$

$$K(2)_t = \max(y_{2t}, G(2)_t, 0). \quad (8)$$

where $K(j)_t$, $G(j)_t$, y_{jt}, respectively, denote the t^{th} component of K(j), G(j), and y_j, j = 1,2.

B 3: For each $\omega_T \varepsilon \Omega_T$, and $t \varepsilon N$,

$$G(2)_0 = \lambda(1)_0, \Lambda_0 = 0, \text{ and}$$

$$G(2)_t = \lambda(1)_t + \Lambda_t, \text{ for } t \geq 1. \quad (9)$$

where $G(2)_t$, $\lambda(1)_t$, Λ_t, respectively, denote the t^{th} component of $G(2)$, $\lambda(1)$, and Λ.

B 4: For each $\omega_T \varepsilon \Omega_T$ and $t \varepsilon N$,

$$y_{1t+1} = \lambda(1)_t + \eta_{1t+1}, \text{ and} \quad (10)$$

$$y_{3t+1} = \lambda(2)_t + \eta_{2t+1}. \quad (11)$$

Also, $\eta_{10}=0$, $\eta_{20}=0$; and there exists a positive pair, $(\acute{y}_1,\acute{y}_3)$, such that $y_{10}=\acute{y}_1$, and $y_{30}=\acute{y}_3$.

B 5: For each $\omega_T \varepsilon \Omega_T$ and all $t \varepsilon N$,

$$FF_t = \max (y_{3t},0) \quad (12)$$

B 6: Let $\lambda = (\lambda(1),\lambda(2))$. There exists a 2×2 matrix, $\varphi = (\varphi_{ij})$, with strictly positive diagonal elements and with largest absolute eigenvalue less than one, such that, for each $\omega_T \varepsilon \Omega_T$,

$$\lambda_0' = \varphi(y_{10},y_{30})', \text{ and}$$

$$\lambda_t' - \lambda_{t-1}' = \varphi((y_{1t},y_{3t})' - \lambda_{t-1}'), \ t \geq 1. \quad (13)$$

The Dynamics of the Market

In ASS's scientific explanation of H the preceding axioms are taken to delineate important positive analogies of the behavior over time of equilibrium yields in the U.S. money market. Some of them, e.g., B 2, describe relationships between different yields that must hold because of the possibilities for arbitrage in the market. Others, e.g., B 4 and B 6, describe essential features of the dynamics of the money market. The following discussion will attest to that.

The axioms B 2-B 6 have logical consequences that concern the possible validity of HAG's two stylized facts in ASS's theory universe. To see why, observe first that.the λ_t of B 6 is to represent the theoretical money market's prediction of the most likely value of $(K(1)_{t+1},FF_{t+1})$ conditional upon the observed values of $(K(1)_s,FF_s)$, $s = 0,1,...,t$. That such an interpretation of λ_t is a possibility can be inferred from B 2, B 4, B 5, and theorem T 1. The latter is a simple logical consequence of B 6.

T 1 Let I_2 be the 2×2 identity matrix, and let φ be the 2×2 matrix in B 6. Then, for each $\omega_T \varepsilon \Omega_T$ and all $t \varepsilon N$,

$$\lambda'_t \;=\; \Sigma_{\,0\le s\le t}\,(I_2 - \varphi)^s\varphi(y_{1t-s},y_{3t-s})' \qquad (14)$$

HAG claimed that the behavior over time of the yield to maturity on a Treasury bill resembles the behavior of an I(1) process. The possibility of such an interpretation of the pair (K(1),FF) and K(2) can be gathered from B 2 and the following three theorems.

T 2 Let φ be the 2×2 matrix of B 6, and let $\eta=(\eta_1, \eta_2)'$. For each $\omega_T \varepsilon \Omega_T$,

$$(y_{11},y_{31})' = \varphi(y_{10},y_{30})' + \eta_1, \text{ and}$$

$$(y_{1t+1}, y_{3t+1})' = (y_{1t}, y_{3t})'+\eta_{t+1} - (I-\varphi)\eta_t,\ t \ge 1 \qquad (15)$$

T 3 Let φ be the 2×2 matrix of B 6, and let $\xi\ \varepsilon$ R be defined by (16).

$$\xi_0=0,\ \xi_1=(1/2)[(1+\varphi_{11})\eta_{11}-(1-\varphi_{11})y_{10}+\varphi_{12}y_{30}+\varphi_{12}\eta_{21}+\Lambda_1], \text{ and}$$

$$\xi_t = (1/2)[(\eta_{1t} - \eta_{1t-1}) + \varphi_{11}(\eta_{1t} + \eta_{1t-1})+\varphi_{12}(\eta_{2t} + \eta_{2t-1}) + \Lambda_t - \Lambda_{t-1}],\ t\ge 2 \quad (16)$$

Then, for each $\omega_T\ \varepsilon\ \Omega_T$,

$$y_{20} = (1/2)[(1+\varphi_{11})y_{10} + \varphi_{12}y_{30}], \text{ and}$$

$$y_{2t} = y_{2t-1} + \xi_t\ ,\ t \ge 1. \qquad (17)$$

T 4 For each $\omega_T\ \varepsilon\ \Omega_T$ and all $t\ \varepsilon$ N,

$$K(2)_t = \max\ [(1/2)(G(1)_t + G(2)_t\ ,0] \qquad (18)$$

HAG also insisted that the yields to maturity of Treasury bills are cointegrated and that the spreads between yields of different maturity define the cointegrating vectors. The next theorem establishes the possibility of HAG's dictum being correct in ASS's theory universe. There yields to maturity on "Treasury bills" might be realizations of cointegrated I(1) processes, and the spread between them might determine their cointegrating relationship.

T 5 Let φ be the 2×2 matrix of B 6. For each $\omega_T\ \varepsilon\ \Omega_T$,

$$y_{20}-y_{10} = (1/2)[\Lambda_0+y_{11}-y_{10})-\eta_{11}], \text{ and}$$

$$y_{2t}-y_{1t} = (1/2)(\Lambda_t-(1-\varphi_{11})\eta_{1t}+\varphi_{12}\eta_{2t}),\ t \ge 1 \qquad (19)$$

Cointegrated processes have common trends. Looking at the common trend of y(1) and y(2) will provide us with new insight into the dynamics of the ASS theory universe's money market. For that purpose, let

$$CTR_t = (\varphi_{11},\varphi_{12})[(\acute{y}_1,\acute{y}_3)' + \Sigma_{0\leq s\leq(t-1)} \eta_{t-1-s}]$$

Then it follows from equations 15, 10, and 9 that

$$y(1)_t = CTR_t + \eta_{1t}, \text{ and}$$

$$G(2)_t = CTR_t + \Lambda_t + (\varphi_{11},\varphi_{12})\eta_t.$$

But, if that is so, then the equilibrium condition in equation 8 implies that

$$y(2)_t = CTR_t + (1/2)[\Lambda_t + \eta_{1t} + (\varphi_{11},\varphi_{12})\eta_t].$$

The equations for y_{1t} and y_{2t} justify referring to CTR_t as the common trend of these variables. In the complete axiom system CTR_t behaves as a generalized random walk.

It is interesting to note here that CTR_t is a function of η_{2t} as well as η_{1t}. Hence "the other part" of the money market plays an essential role in the construction of the common trend of Treasury Bill yields.

It is also interesting to note that the y_1 and y_3 series need not be cointegrated. In the complete axiom system they will be cointegrated ARIMA processes if and only if there is a pair, α_1 and α_2, such that $(\alpha_1,\alpha_2)\varphi = 0$. The reason why is explicated in T 6.

T 6 Let φ be the 2×2 matrix of B 6. For each $\omega_T \in \Omega_T$,

$$(y_{1t+1}, y_{3t+1})' = \varphi(\acute{y}_1,\acute{y}_3)' + \varphi \Sigma_{0\leq s\leq(t-1)} \eta_{t-s} + \eta_{t+1} \text{ , and}$$

$$y_{2t} = (\varphi_{11},\varphi_{12})[(\acute{y}_1,\acute{y}_3)' + \Sigma_{0\leq s\leq(t-1)} \eta_{t-1-s}] + (1/2)[\Lambda_t + \eta_{1t} + (\varphi_{11},\varphi_{12})\eta_t].$$

Since the first component in the equations for y_{2t} is CTR_t, we need only prove the expression for (y_{1t+1}, y_{3t+1}). The proof of that fact goes as follows.

$$\begin{aligned}(y_{1t+1}, y_{3t+1})' &= \lambda_t' + \eta_{t+1} \\ &= \lambda_{t-1}' + \varphi\eta_t + \eta_{t+1} \\ &= \lambda_0' + \varphi \Sigma_{0\leq s\leq(t-1)} \eta_{t-s} + \eta_{t+1} \\ &= \varphi(\acute{y}_1,\acute{y}_3)' + \varphi \Sigma_{0\leq s\leq(t-1)} \eta_{t-s} + \eta_{t+1} .\end{aligned}$$

If a pair, (α_1, α_2), exists such that $(\alpha_1,\alpha_2)\varphi = 0$, CTR_t or some constant multiple of CTR_t becomes the common trend of y_1 and y_3.

The ASS Bridge Principles

So much about the ASS (Ω_T, Γ_t) for now. Next I shall write down the bridge principles in ASS's scientific explanation of H, and describe several of the properties of Treasury-bill yields in the data universe that one can derive from them and the B-axioms. I begin with the bridge principles.

G 1: The sample space, Ω, is a subset of $\Omega_T \times \Omega_P$; i.e., $\Omega \subset \Omega_T \times \Omega_P$.

G 2: There exists a 3×3 matrix, $\psi = (\psi_{ij})$, with strictly positive diagonal elements and with largest absolute eigenvalue less than one, such that, for each $\omega \; \varepsilon \; \Omega$,

$$\hat{y}_0 = \acute{y}_0, \; u_0 = 0, \text{ and}$$

$$\hat{y}_t - \hat{y}_{t-1} = \psi(y_t - \hat{y}_{t-1}) + u_t, \; t \geq 1, \qquad (20)$$

where $\hat{y} = (\hat{y}_1,\hat{y}_2,\hat{y}_3)'$, $y = (y_1,y_2,y_3)'$ and $u = (u_1,u_2,u_3)'$

Characteristic Behavior in the Data Universe

From the characterization of ASS's theory universe and from G 1 and G 2 one can derive the following two interesting theorems.

T 7 Let U denote the forward shift operator, let ξ be as described in equation 16, and let $\zeta = (\zeta_1,\zeta_2,\zeta_3) \; \varepsilon \; (R^3)^N$ be defined for arbitrary t by equations 21-23.

$$\zeta_{1t}=u_{1t}-u_{1t-1}+\psi_{11}[\eta_{1t}-(1-\varphi_{11})\eta_{1t-1}+\varphi_{12}\,\eta_{2t-1}]+\psi_{13}[\eta_{2t}-(1-\varphi_{22})\eta_{2t-1}+\varphi_{21}\eta_{1t-1}]+\psi_{12}\,\xi_t \; (21)$$

$$\zeta_{2t}=u_{2t}-u_{2t-1}+\psi_{21}[\eta_{1t}-(1-\varphi_{11})\eta_{1t-1} + \varphi_{12}\eta_{2t-1}]+\psi_{23}[\eta_{2t}-(1-\varphi_{22})\eta_{2t-1}+\varphi_{21}\eta_{1t-1}]+\psi_{22}\xi_t \; (22)$$

$$\zeta_{2t}=u_{2t}-u_{2t-1}+\psi_{31}[\eta_{1t}-(1-\varphi_{11})\eta_{1t-1}+\varphi_{12}\eta_{2t-1}]+\psi_{33}[\eta_{2t}-(1-\varphi_{22})\eta_{2t-1}+\varphi_{21}\eta_{1t-1}] +\psi_{32}\xi_t \; (23)$$

For each $\omega \; \varepsilon \; \Omega$,

$$\hat{y}_0 = \acute{y}_0,$$

$$(I-(I-\Psi)U^{-1})\hat{y}_1 \; = \; \Psi \begin{matrix} \varphi_{11}y_{10}+\varphi_{12}y_{30}+\eta_{11} \\ \\ y_{20}+\Lambda_1 \end{matrix} +u_1, \text{ and}$$

$$(I - (I - \psi)U^{-1})(I-IU^{-1})\hat{y}_t = \zeta_t \, , \; t \geq 2, \qquad (24)$$

T 8 Let ε_t be defined by equation 25 for all t ε N-{0,1}:

$$\varepsilon_t \; = \; [I - (I - \Psi)U^{-1}]^{-1} \zeta_t \, , \qquad (25)$$

where ζ is as described in equations 21-23. Then, for each $\omega \; \varepsilon \; \Omega$ and all t ε N-{0,1},

$$\hat{y}_{2t-1}-\hat{y}_{1t-1} = y_{2t}-y_{1t} + (-1,1,0)\psi^{-1} (u_t-\varepsilon_t);$$

$$= (1/2)(\Lambda_t-(1-\varphi_{11})\eta_{1t}+\varphi_{12}\eta_{2t}) + (-1,1,0) \, \psi^{-1} (u_t-\varepsilon_t) \qquad (26)$$

These two theorems and theorems T 2, T 3, and T 5 suggest that the ŷ and the y series might each be realizations of two cointegrated I(1) processes with the same cointegrating vector (-1,1,0).

The ASS P-Axioms

To establish that the ŷ and the y series in fact are realizations of cointegrated I(1) processes I must specify the stochastic properties of the components of ω. These properties can be deduced from ASS's three axioms, P 1, P 2, and P 3, and from the conditions I delineated in axioms B 2 - B 6, D 2, and G 1 - G 2.

P 1: Let the vector valued function, (K,G,FF,λ,Λ,η,y,u,Ḱ,ff,ŷ)(·): $N \times \Omega \rightarrow R^{22}$ be defined for all t ε N by equation 27:

$$(K,G,FF,\lambda,\Lambda,\eta,y,u,\acute{K},ff,\hat{y})(t,\omega) = \omega_t,\ t \in N, \text{ and } \omega \in \Omega. \quad (27)$$

Let ℵ be a σ-field of subsets of Ω, and suppose that the functions (K,...,ŷ)(t,·): $\Omega \rightarrow R^{22}$, t ε N, are measurable with respect to ℵ. There exists a probability measure, P(·): ℵ → [0,1], relative to which the family of functions,{(Λ,η,u)(t,ω);t ≥ 1,}is a vector-valued wide-sense stationary process.

P 2 Let P(·): ℵ → [0,1] be as described in P 1. Relative to P(·), the families of functions, {η(t,ω);t ≥ 1} and {u(t,ω);t ≥ 1}, constitute purely random processes with means zero and covariance matrices, Σ_η and Σ_u.

P 3: Let P(·): ℵ → [0,1] be as described in P 1. Also, let the function, (K,G,FF,λ,Λ,η,y,u,Ḱ,ff,ŷ)(·): $N \times \Omega \rightarrow R^{22}$, be as described in equation (27). Then, relative to P(·),for each t ∈ N-{0}, the Λ, η, u, and y components of this family of functions satisfy the following conditions:

$$E\{\eta(t,\omega) \mid (y_1,y_3)(0),\ldots,(y_1,y_3)(t-1)\} = 0 \text{ a.e.}$$

$$E\{\Lambda(t,\omega) \mid (y_1,y_3)(0),\ldots,(y_1,y_3)(t)\} = \Lambda(t,\omega) \text{ a.e., and}$$

$$E\{u(t,\omega) \mid (y_1 y_3)(0),\ldots,(y_1,y_3)(t)\} = 0 \text{ a.e.}$$

In reading these axioms there are several things to notice: (1) ASS do not insist that Σ_η is diagonal. (2) The family of functions, {Λ(t,ω);t ≥ 0}, need not constitute a purely random process. (3) ASS have only delineated conditions on the Λ, η, u, and y components of (K,G,FF,λ,Λ,η,y,u,Ḱ,ff,ŷ)(·). However, from P 1-P 3, the axioms in the ASS theory universe, and the ASS bridge principles one can derive all the conditions on the components of $(K,G,FF,\lambda,\acute{K},ff,\hat{y})(\cdot)$ that ASS needed in their scientific explanation of HAG's two stylized facts. (4) The characteristics of the probability distributions of the family of random variables, $\{(\acute{K},ff,\hat{y})(t,\omega);\ t \geq 0\}$, that one derives from P 1-P 3, B 1-B6, and G 1-G 2 are characteristics of the MPD.

The ASS Scientific Explanation

With P 1- P 3 in hand, theorem T 14 below becomes an immediate consequence of B 4, G 2, P 1- P 3, and theorems T 6- T8 and T 10, T 12, and T 13. Therefore, there is no need to spell out the details of a proof here. In reading the theorem, note that the random process, $\{\hat{y}(t,\omega);\ t \in N\}$, is defined on $(\Omega,\aleph)$ and not on $(\Omega_P,\aleph_P)$. Hence, H is not an immediate consequence of T 14 and D 2.

T 14 Let $P(\cdot): \aleph \rightarrow [0,1]$ be the probability measure in P 1-P 3. Relative to $P(\cdot)$, the λ, y, and $\hat{y}$ components of the family of function, $\{(K,G,FF,\lambda,\Lambda,\eta,y,u,\acute{K},ff,\hat{y})(t,\omega);t\geq 0\}$, satisfy the following conditions:

(i) For all $t \geq 1$, with $P(\cdot)$-probability one,

$$\lambda(t,\omega) = E((y_1,y_3)(t+1)|(y_1,y_3)(0),\ldots,(y_1,y_3)(t)) \tag{28}$$

(ii) For all $t\ \varepsilon\ N$, $(\acute{K},ff)(t,\omega) = \max\ (\hat{y}(t,\omega),0)$ a.e.

(iii) The family of functions $\{y(t,\omega);t\ \varepsilon\ N\}$ is an I(1) vector-valued ARIMA process the first two components of which are cointegrated with cointegrating vector, (-1,1,0).

(iv) The family of functions, $\{\hat{y}(t,\omega);t\ \varepsilon\ N\}$ is an I(1) vector-valued ARIMA process the first two components of which are cointegrated with cointegrating vector, (-1,1,0).

T 14(ii) and T 14(iv) imply that there is an MPD, i.e., a probability distribution of the family of random vectors on $(\Omega, \aleph)$, $\{(\acute{K},ff,\hat{y})(t,\omega);t\geq 0\}$, that Γ_t, $\Gamma_{t,p}$, and the axioms of $P(\cdot)$ determine and that satisfies the strictures on which H insists. But if that is so, then standard arguments (cf. Stigum, 1990, pp. 344 -347) suffice to establish the existence of a probability measure, $Þ(\cdot): \aleph_P \rightarrow [0,1]$, relative to which the family of functions, $\{(\acute{K},ff,\hat{y})(t,\omega_P);t \in N\}$, which was defined in equation 5, satisfies the conditions of assertion H. To wit: The given MPD consists of a family of finite probability distributions of the random vectors in $\{(\acute{K}_t, ff_t, \hat{y}_t); t \in N\}$. These probability distributions determine a probability measure, Þ(·): $\aleph_p \rightarrow [0,1]$, relative to which the family of functions, $\{(\acute{K}, ff, \hat{y})(t, \omega_P); t \in N\}$, satisfies the conditions on which H insists.

The probability measure Þ(·) need not be the same as the probability measure $P_p(\cdot)$ on whose existence I insisted in D 3. However, the probability distributions of the family $\{(\acute{K}, ff, \hat{y})(t, \omega_P): t \in N\}$ relative to Þ (·) and $P_p(\cdot)$ share the conditions on which H insists. From this and the preceding paragraph it follows that H is true in all models of D 1-D 3, B 1-B 6, G 1-G 2, and P 1-P 3. Since H is true in some, but not all models of D 1- D 3, I conclude that B 1-B 6, G 1-G 2, and P 1-P 3 provide a logically adequate scientific explanation of H.

Taking stock of ASS's formal scientific explanation H so far, calls for several remarks. Note first that one can think of H as a pair, (H_1, H_2), where H_1 repeats what is said in D 2 and H_2 comprises the assertions in H. Next, note that the ASS theory universe, the ASS bridge principles, and the ASS axioms of P(·) have many models. The models of the theory universe vary with the values of the pair, $(\acute{y}_1, \acute{y}_3)$, in B 4 and the matrix, φ, in B 6. The models of the bridge principles vary with the values of the vector, $\acute{y}_0$, and the matrix, Ψ, in G 2. And the models of the P(·)- axioms vary with the values of the components of the matrices, Σ_η and Σ_u, in P 2. In developing ASS's scientific explanation of H I established its logical adequacy without referring to any of these models. Hence the MPD in ASS's explanation comprises a large family of MPDs that the models of B 1-B 6, G 1- G2, and the axioms of P(·) determine. When ASS set out to establish the empirical adequacy of their explanation, they applied numerical analysis to estimate one model of the quadruple, $(\Gamma_t, \Gamma_p, \Gamma_{t,p}, P(\cdot))$, derived the associated model of the MPD, and checked whether the explanation of H that the two models provided was empirically adequate. The explanation was to be deemed empirically adequate if there existed a family of models of the MPD that with probability 0.95 contained a model whose salient characteristics were shared by some model of the FP in $\Im_2$.

The Empirical Adequacy

ASS's scientific explanation of H is logically adequate. Whether it is also empirically adequate remains to be seen. To be reasonably certain of the empirical adequacy of B 1- B 6, G 1-G 2, and P 1-P 3 one must search for a family of models of the MPD, $\Im$, that with probability 0.95 contains a model whose salient characteristics are shared by FP. Finding an appropriate $\Im$ is a bit of a fishing expedition that I describe next.

A State Space Formulation of Equation (24)

P 1-P 3 do not put stringent conditions on the probability distributions of η, Λ, and u. So ASS began their search for an appropriate $\Im$ by assuming that η_1 , η_2 , Λ, and the components of u constitute six independently distributed purely random processes. For the sake of argument they also, initially, assumed that all these processes are Gaussian.

Next they reformulated the relations in (24) so that they could be expressed by a state space model. For that purpose they let $v(t)=(u_1(t),u_2(t),u_3(t),\eta_1(t),\eta_2(t),\Lambda(t))'$ and observed that

$$\zeta(t) = C_0v(t) - C_1v(t-1),\ t = 1,2,\dots , \qquad (29)$$

with

$$C_0 = \begin{array}{cccccc} 1 & 0 & 0 & \Psi_{11} + 1/2\Psi_{12}(1+\varphi_{11}) & \Psi_{13} + 1/2\Psi_{12}\varphi_{12} & 1/2\Psi_{12} \\ 0 & 1 & 0 & \Psi_{21} + 1/2\Psi_{22}(1+\varphi_{11}) & \Psi_{23} + 1/2\Psi_{22}\varphi_{12} & 1/2\Psi_{22} \\ 0 & 1 & 0 & \Psi_{31} + 1/2\Psi_{32}(1+\varphi_{11}) & \Psi_{33} + 1/2\Psi_{32}\varphi_{12} & 1/2\Psi_{32} \end{array} \text{, and}$$

$$C_1 = \begin{array}{cccccc} 1 & 0 & 0 & (\Psi_{11}+1/2\Psi_{12})(1-\varphi_{11})-\Psi_{13}\varphi_{21} & -(\Psi_{11}+1/2\Psi_{12})\varphi_{12}+\Psi_{13}(1-\varphi_{22}) & 1/2\Psi_{12} \\ 0 & 1 & 0 & (\Psi_{21}+1/2\Psi_{22})(1-\varphi_{11})-\Psi_{23}\varphi_{21} & -(\Psi_{21}+1/2\Psi_{22})\varphi_{12}+\Psi_{23}(1-\varphi_{22}) & 1/2\Psi_{22} \\ 0 & 0 & 1 & (\Psi_{31}+1/2\Psi_{32})(1-\varphi_{11})-\Psi_{33}\varphi_{21}) & -(\Psi_{31}+1/2\Psi_{32})\varphi_{12}+\Psi_{33}(1-\varphi_{22}) & 1/2\Psi_{32} \end{array}$$

Also, they let $A = I_3 - \Psi$, and wrote equation (24) as

$$\hat{y}(t) - \hat{y}(t-1) = A(\hat{y}(t-1) - \hat{y}(t-2)) + C_0v(t) - C_1v(t-1) \qquad (30)$$

Then with $x(t) = (\hat{y}(t) - \hat{y}(t-1), v(t), v(t+1))$ the searched for state space form of (24) became equations (31) and (32).

$$x(t) = \begin{matrix} A & -C_1 & C_0 \\ 0 & 0 & I_6 \\ 0 & 0 & 0 \end{matrix} \; x(t-1) + \begin{matrix} 0 \\ 0 \\ v(t+1) \end{matrix} \qquad (31)$$

and

$$\hat{y}(t) = \hat{y}(t-1) + \{I_3 \;\; 0 \;\; 0\}\; x(t) \qquad (32)$$

Equation (31) describes the dynamics of the state vector, while equation (32) insists that $\hat{y}(t)$ is observed at each time point and that there are no observation errors.

Estimation of Parameters

There are nineteen parameters to be estimated, the nine components of Ψ, the four components of φ, and the six variances of the components of v. Unfortunately, the likelihoods of the parameter sets,

$$\{\sigma_{u1}^2, \sigma_{u2}^2, \sigma_{u3}^2, \sigma_{\Lambda}^2\} \text{ and}$$

$$\{\sigma_{u1}^2 + 1/4\Psi_{12}^2\sigma_{\Lambda}^2, \sigma_{u2}^2 + 1/4\Psi_{22}^2\sigma_{\Lambda}^2, \sigma_{u3}^2 + 1/4\Psi_{32}^2\sigma_{\Lambda}^2, 0\},$$

are indistinguishable, which goes to show that one cannot obtain identifiable estimates of all the four variances of u and Λ. Since ASS could not be certain of the validity of the bridge principles anyway, they decided to assign the value zero to the u-terms in equations (20) and (24). That left them with sixteen unknown parameters in a new state-space model in which $v(t) = (\eta_1(t),\eta_2(t),\Lambda(t))'$ and in which the first three columns of the C_0 and C_1 matrixes are deleted.

To estimate the values of the remaining sixteen parameters ASS proceeded as follows. They calculated the likelihood of their sample observations with the help of a Kalman filter (Harvey, 1989) and used the resulting likelihood and Powell's method of numerical optimization (Press et al, 1992, pp. 412) to search for the parameter values at which the likelihood would attain its maximum height. Powell's method is not fast, but it seemed to give reliable optima.

To obtain uncertainty measures and confidence intervals for their maximum likelihood parameter estimates, ASS used the method of bootstrapping (cf. (Efron, 1982) and (Davison and Hinkley, 1997)). The main ideas behind the bootstrapping procedure are threefold: (1) Generate a series of artificial data sets by simulation following (almost) the same model as the one that generated the original data set; (2) estimate the values of the model parameters anew with each one of the simulated data set; and (3) use the series of parameter estimates obtained from a large number of data sets to determine meaningful measures of the uncertainty that is attached to the original parameter estimates. In the present case ASS applied the ideas of bootstrapping in the following way. First, they used the original parameter estimates and the state-space model to obtain estimates of the components of the error term, v(t), for all pertinent t. Next, they generated new series of error terms by resampling (drawing with replacement) from the estimated time series of v(t). Then they used each new time series of v(t) and the state-space model with the original parameter estimates to construct time series of values of ŷ(t), ŷ(t-1), and ŷ(t-2). At last they

	Estimate	Bias	St. error	95% Conf. Interval	
ση,1	0.356	-0.179	0.022	0.493	0.577
ση,2	0.632	-0.446	0.027	1.022	1.128
σΛ	10.54	-0.779	1.553	8.046	14.119
φ11	0.787	0.003	0.056	0.678	0.899
φ12	0.216	-0.106	0.059	0.208	0.435
φ21	0.68	-0.024	0.076	0.558	0.849
φ22	0.632	-0.049	0.081	0.526	0.837
ψ11	0.604	-0.006	0.016	0.577	0.638
ψ12	0.012	-0.004	0.002	0.012	0.019
ψ13	0.15	0.026	0.01	0.103	0.144
ψ21	0.562	0.005	0.007	0.544	0.57
ψ22	0.064	-0.003	0.002	0.065	0.071
ψ23	0.088	-0.008	0.005	0.087	0.109
ψ31	-0.01	-0.002	0.015	-0.039	0.02
ψ32	0.031	-0.018	0.004	0.042	0.057
ψ33	0.284	-0.016	0.007	0.287	0.313
λ1φ	0.682	-0.085	0.094	0.607	0.976
λ2φ	0.101	-0.03	0.054	0.002	0.199
λ1ψ	0.825	-0.019	0.02	0.603	0.681
λ2ψ	0.265	-0.004	0.009	0.253	0.287
λ3ψ	0.062	-0.001	0.002	0.059	0.068

Table 1 *shows the parameter estimates, estimates of bias and standard errors, and 95% confidence intervals. The confidence intervals were obtained by using the simplest percentile method.*

used each series of the latter variables to estimate the model parameters anew. In that way they obtained a series of parameter values that they could use to construct the uncertainty measures for which they searched. ASS carried out a total of one thousand bootstrap simulations.

Table 1 and the Empirical Adequacy of the Explanation of HAG's Two Stylized Facts

Suppose, for now, that ASS's assumptions concerning the probability distributions of η, Λ, and u are correct. This assumption, the estimates of φ, Ψ, Σ_η, and σ_Λ^2 that one finds in Table 1, and arbitrary values of $(\acute{y}_1,\acute{y}_3)$ and $\acute{y}_0$ determine a model of the quadruple, $(\Gamma_t, \Gamma_p, \Gamma_{t,p}, P(\cdot))$. From this model I can derive a model of the MPD. . Here is how: From the estimates in Table 1 and equation (24) I deduce that

$$\hat{y}(t)-3.048\hat{y}(t-1)+3.3614\hat{y}(t-2)-1.5689\hat{y}(t-3)+0.2555\hat{y}(t-4)=B(U)\zeta(t),\ t\geq 4 \qquad (33)$$

$$\Delta\hat{y}(t) - 2.048\Delta\hat{y}(t-1) + 1.3134\Delta\hat{y}(t-2) - 0.2555\Delta\hat{y}(t-3) = B(U)\,)\zeta(t),\ t\geq 4, \text{ and} \qquad (34)$$

$$\hat{y}_2(t-1) - \hat{y}_1(t-1) = (1/2)[\Lambda_t - 0.213\eta(1)_t + 0.216\eta(2)_t] - (-1,1,0)\Psi^{-1} \in_t , \qquad (35)$$

where

$$B(U) = \begin{matrix} 1 -1.652U^{-1} + 0.6675U^{-2} & -0.562U^{-1} + 0.4024U^{-2} & 0.0174U^{-2} \\ -0.012U^{-1} + 0.0133U^{-2} & 1-1.112U^{-1} + 0.2835U^{-2} & -0.031U^{-1} + 0.0123U^{-2} \\ -0.15U^{-1} + 0.1415U^{-2} & -0.088U^{-1} + 0.1191U^{-2} & 1-1.332U^{-1}+0.364U^{-2} \end{matrix}$$

$$\Psi^{-1} = \begin{matrix} 1.527382 & 0.1227996 & -0.8447666 \\ -15.867849 & 17.1085015 & 3.0796806 \\ 1.785835 & -1.8631533 & 3.1552191 \end{matrix}$$

and, with $D(U) = Det(I-AU^{-1}) = 1 - 2.048U^{-1} + 1.3134U^{-2} - 0.2555U^{-3}$,

$$\varepsilon(t) = D(U)^{-1}B(U)\zeta(t).$$

From the estimates in Table 1 and equations (24) and (29) I can also deduce that

$$\Delta\hat{y}(t) = A^* \Delta\hat{y}(t-1) + C_0^*(\eta,\Lambda)(t)' - C_1^*(\eta,\Lambda)(t-1)' , \qquad (35)$$

where

$$A^* = \begin{matrix} 0.396 & -0.012 & -0.13 \\ -0.562 & 0.936 & -0.088 \\ 0.01 & -0.031 & 0.716 \end{matrix} ; \qquad C_0^* = \begin{matrix} 0.6147220 & 0.151296 & 0.0060 \\ 0.6191840 & 0.094912 & 0.0320 \\ 0.0176985 & 0.287348 & 0.0155 \end{matrix}$$

$$C_1^* = \begin{matrix} 0.5032780 & 0.021504 & 0.0320 \\ 0.4668160 & 0.006464 & 0.0320 \\ -0.1998185 & 0.039820 & 0.0155 \end{matrix}$$

Finally, from equations (16), (21)-(23) and ASS's assumptions concerning η and Λ, and from the estimates in Table 1 I can deduce that the covariance structure of ζ is given by equations (36) and (37).

$$Var[\zeta(t)] = C_0\Sigma C_0' + C_1\Sigma C_1' = \begin{matrix} 0.154 & 0.154 & 0.026 \\ 0.154 & 0.172 & 0.021 \\ 0.026 & 0.021 & 0.077 \end{matrix} \qquad (36)$$

$$Cov[\zeta(t), \zeta(t-1)] = -C_0\Sigma C_1' = \begin{matrix} 0.001 & -0.007 & 0.031 \\ -0.003 & -0.020 & 0.031, \\ 0.0130 & 0.012 & -0.02 \end{matrix} \qquad (37)$$

where $\Sigma = Var(v(t)) = diag(0.127, 0.399, 111.092)$.

In looking back at the description of the MPD that equations (36)-(37) provide, notice that the characteristic polynomials, $P_{\hat{y}}(z)$ and $P_{\Delta\hat{y}}(z)$, of equations (33) and (34) are given by (38) and (39), respectively.

$$P_{\hat{y}}(z) = (1-z)\,(z - 0.366508)(z - 0.742084)(z - 0.939408) \qquad (38)$$

$$P_{\Delta\hat{y}}(z) = (z - 0.366508)(z - 0.742084)(z - 0.939408) \qquad (39)$$

Hence, $\Delta\hat{y}(t)$ satisfies the equation of a wide-sense stationary autoregressive moving average process, i.e., of an I(0) ARIMA process. Also, $\hat{y}(t)$ satisfies the equation of an I(1) ARIMA process. This and equation (26) imply that the ŷ(t)-process has all the characteristics on which H(ii) and H(iii**)** insist. Since the components of Ḱ(1), Ḱ(2), and ff_t are non-negative by construction, the MDP of the vector, (Ḱ(t,·), ff(t, ·), ŷ(t,·)), has the characteristics on which H, and hence H_2, insists.

The equations (33)-(39) delineate one member of ℑ, the sought-for family of models of the MDP. I obtain other models by varying, within the pertinent confidence bands, the values of the components of φ and Ψ and the values of the variances of η and Λ. There are no restrictions on the values that the variances can assume. However, the φs and Ψs that I pick must be so that the absolute values of their characteristic roots are less than 1. In describing the contour of ℑ ASS let the limits on the pertinent variances equal the estimated 0.95 confidence bands. I shall use the estimated confidence bands for the φs and Ψs as well believing that the associated confidence bands for the characteristic roots that they determine are not too different from the confidence bands that ASS have estimated. The reason for my belief is that each bootstrap estimate of the characteristic roots is based on the φ and Ψ values of the pertinent sample. Thus the uncertainty measures and the confidence bands for the characteristic roots in Table 1 take the relation between the φ and Ψ values and the values of the characteristic roots into account.

The probability distributions of η and Λ

If the MPD that ASS have estimated is data admissible, I obtain the sought for family of models of the MPD by varying the parameters within their 0.95% confidence bands. The MPD is data admissible only if the probability distribution of η and Λ satisfy the conditions that ASS have imposed on them. I shall describe the results of various tests of these conditions next.

24.4.4.1 ***Checking for Independence***

The Kalman filter provides estimates of the components of v(t) that one can use for model checking. Figure 3.1 shows scatter plots of all combinations of the components of η. The independence assumption seems to be realistic for these variables

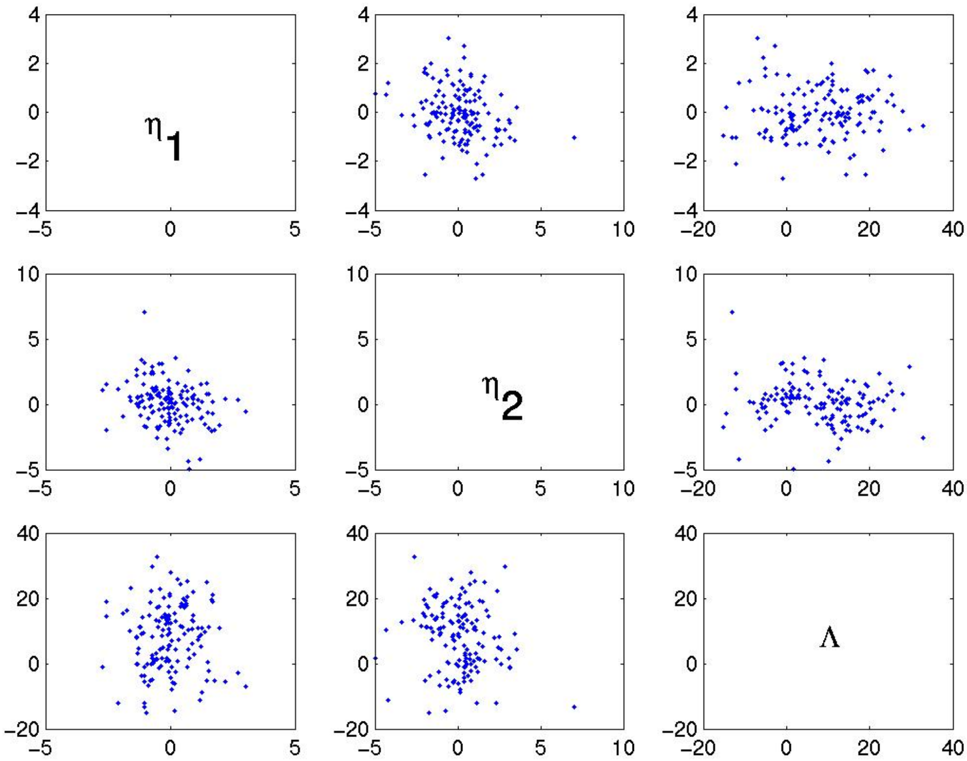


Figure 3.1: *Scatterplots of each pair of random terms calculated from the estimated model by the Kalman filter. On both axes the order of the random terms are η_1, η_2, Λ. η and Λ.*

To provide a second check on the independence assumption ASS also calculated the autocorrelations and the partial autocorrelations of the components of η and Λ. The results are displayed in Figures 3.2 and 3.3. The estimated values of the autocorrelations indicate that independence in time seems reasonable for η_1 and η_2. The small temporal dependence for the latter variable can be blamed on the estimation procedure. Temporal independence for Λ is more suspect. As Figure 3.4 suggests, an autoregressive model of Λ of order two might be a better model for Λ. This possibility is interesting in as much as Λ_t, in contradistinction to the components of η_t, is measurable with respect to the σ-field that the pairs, $(y_1,y_3)(0),\ldots,(y_1,y_3)(t)$, generate. However, ASS did not follow it up with new estimates here.

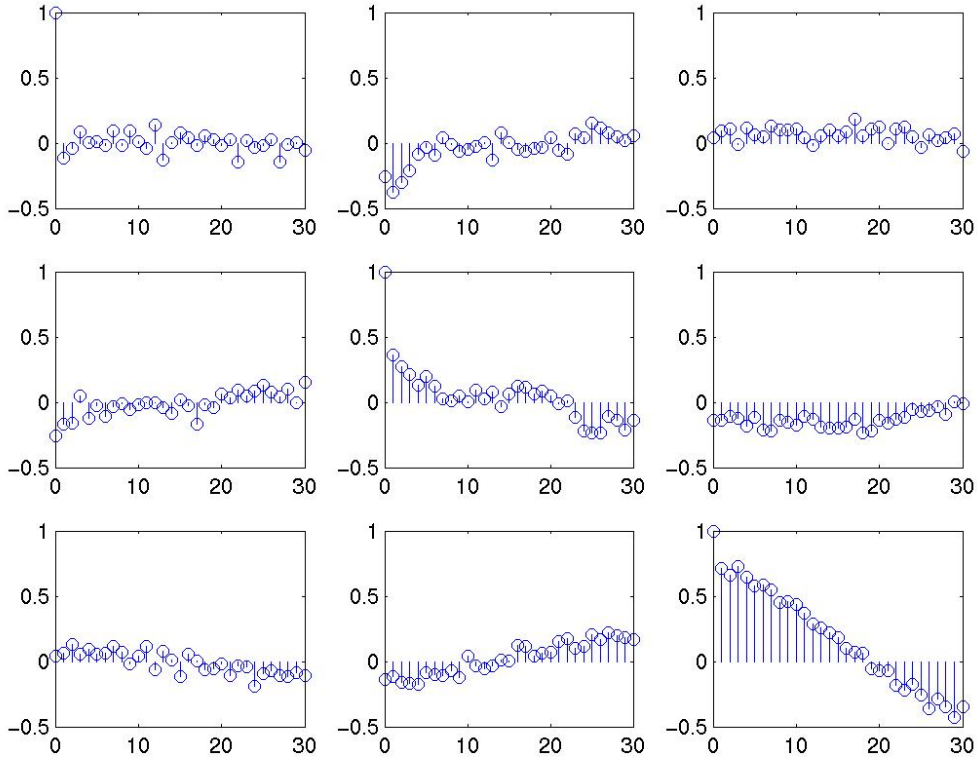


Figure 3.2: *Estimated autocorrelation functions for* (η_1, η_2, Λ).

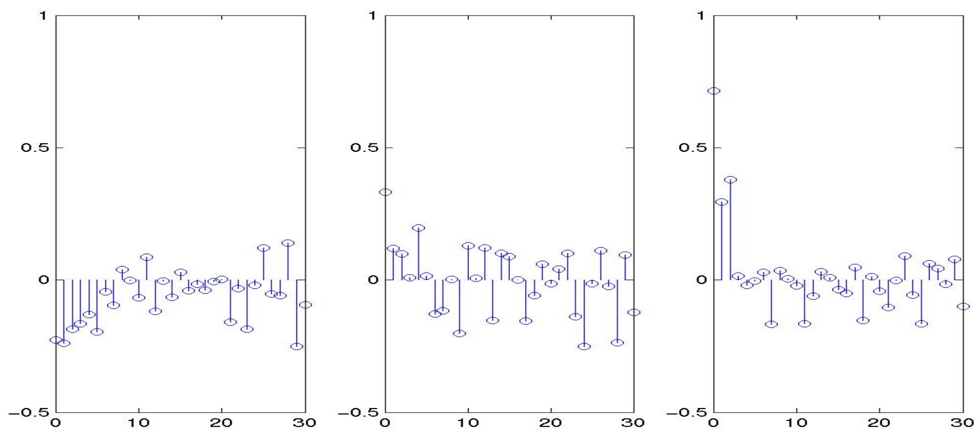


Figure 3.3: *Estimated partial autocorrelation functions for* η_1, η_2 *and* Λ.

Checking for Normality

Figure 3.4 displays QQ.-plots of the (standardized) values of η and Λ. The plots indicate that a Gaussian distribution is reasonable for η_1 and Λ, while there are a couple of outliers in η_2 .

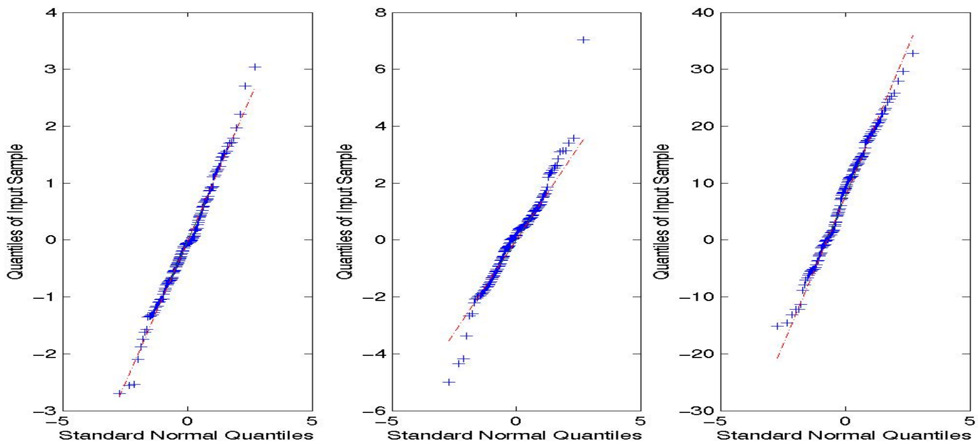


Figure 3.4: *QQ-plot of ordered residuals against quantiles in the Gaussian distribution. The order of the variables are η_1, η_2 and Λ.*

Summing Up

I noted above that the search for a family of models of the data universe within which ASS's scientific explanation of HAG's two stylized facts is empirically relevant is like a fishing expedition.. In ASS's statistical calculations they assumed that the two η-processes and the Λ-process were independently distributed, purely random Gaussian processes. Their statistical results do not ensure that the Λ- process is purely random and that the η_2 - process is Gaussian. **Since they did not insist on**

these characteristics in the P(·) axioms, they decided to ignore this uncertainty and considered this part of their statistical analysis completed.

A Last Remark Concerning the Generality of ASS's Scientific Explanation

If the H I formulated is a correct rendition of HAG's two stylized facts, the generality of ASS's scientific explanation of H rests and falls on whether one loses insight in the workings of the U.S. Money market by considering just two bills and one other Money market instrument. Now, it is significant here that axioms B 2, B 4, and G 2 easily can be generalized to a market with many more securities. From this it follows that one's gain in clarity from ASS's simplifying assumption about the number of bills and other Money market instruments has not been at the expense of a loss in generality.

FOOTNOTES

1. This is meant to be a discussion paper for a lecture on "Stylized Facts, the Purport of an Economic Theory, and Scientific Explanation in Economics and Econometrics. Most of the material is taken from Chapters 22 and 23 in (Stigum, 2003).
2. A positive analogy for a group of individuals (or a family of events) is a characteristic that the members of the group (family) share. A negative analogy is a characteristic that only some of the members of the group (family) share.

3. There are interesting observations on *ceteris paribus* clauses and tendency laws in Mark Blaug's discussion of Mill (Blaug, 1990, pp. 59-69) and in Daniel M. Hausmann's account of inexactness in economic theory (Hausmann,1992, ch. 8). Also, Lawrence Summer's stories of successful pragmatic empirical work (Summers, 1991, pp. 140-141) provide insight into the way economic theorists learn about the positive analogies that their theories identify.
4. There are many ways to interpret the error terms; e.g., as measurement errors. Judging from (Spanos, 1995) and private communications, Spanos now favors a dynamic relationship between planned sales and purchases and actual sales and purchases. Be that as it may. The important idea that I am trying to convey in the example is that the domain of definition of the bridge principles is the sample space and that the sample space is a proper subset of $\Omega_T \times \Omega_P$.
5. Originally, Hempel insisted that the "sentences constituting the explanans must be true" (Hempel, 1965, p.248).
6. Here I am using the vernacular of modal logic. If an assertion, A, is true necessarily, A is true in all possible worlds. If there is a possibility that A be true, there are worlds in which A is true. In symbols, $\Box A$ insists that A is true necessarily and $\sim\Box\sim A$ claims that it is possible that A is true.
7. I have made use of versions of this example in several articles. The present version is identical with the one I presented in Chapter 22 in (Stigum, 2003).
8. To make my description of the PDGP and the DGP as simple as possible I have assumed implicitly that the sampling scheme is such that the ω_{PS} are identically and independently distributed.

9. Roughly speaking, an ARIMA process is a stochastic difference equation whose characteristic polynomial has one or more roots of absolute value 1. It is said to be integrated of order one, and denoted by I(1), if it has just one such root. An ARMA process is a stochastic difference equation whose characteristic polynomial has no roots of modulus 1. Such a process is said to be integrated of order 0, and denoted by I(0). A multivariate I(1) ARIMA process is cointegrated if there is a linear combination of the variables that possesses all the characteristics of a wide-sense stationary process. Relevant details concerning characteristics of ARIMA processes the reader can find in (Stigum, 1975) and in (Stigum, 2003, chapter 23).

REFERENCES

Allais, M., 1979, "The So-called Allais Paradox and Rational Decisions Under Uncertainty," in:*Expected Utility Hypothesis and the Allais Paradox***,** M. Allais and O. Hagen (eds.), Boston: Reidel.

Blaug, M., 1980, *The Methodology of Economics*, Cambridge: Cambridge University Press.

Duhem, P., 1954, *The Aim and Structure of Physical Theory***,** P. P. Wiener (trans.), Princeton: Princeton University Press.

Efron, B., 1982, *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia: Society for Industrial and Applied Mathematics.Fisher, R. A., 1922, "On the Mathematical Foundation of Theoretical Statiostics," *Philosophical Transactions of the Royal Society of London*, Series A, 222, 309-

368

Friedman, M., 1953, “The Methodology of Positive Economics,” in: *Essays in Positive Economics by M. Friedman, Chicago: Chicago University Press*

Friedman, M., 1957. *A Theory of the Consumption Function*, Princeton: Princeton University Press,

Hall, T., H. Anderson, and C. W. J. Granger, 1992, “A Cointegration Analysis of Treasury Bill Yields,” *The Review of Economics and Statistics* 74, 116–126.

Hempel, C. G., 1965, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science,* New York:The Free Press.

Haavelmo, T., 1944, *The Probability Approach in Econometrics*, Econometrica, 12, 1-118, Supplement

Mill, J. S., 1836, “On the Definition of Political Economy; and on the Method of Investigation Proper to It” reprinted in: *Collected Works, Essays on Economy and Society*, *Vol. 4*, J. M. Robson (ed.), Toronto: University of Toronto Press, 1967.

Modigliani, F., and R. Brumberg, 1955, "Utility Analysis and the Consumption Function: An Interpretation of Cross-section Data," in *Post-Keynesian Economics*, K. K. Kurihara, (ed.), London: Allen and Unwin.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannert, 1992, *Numerical Recipes in*, Cambridge: Cambridge University Press.

Spanos, A. 1989, “On Rereading Haavelmo: A Retrospective View of Econometric Modeling,” **Econometric Theory**, vol. 5, pp. 405-429

Spanos, A. 1995, “ On Theory Testing in Econometrics: Modeling with Nonexperimental Data,“ **Journal of Econometrics**, Vol. 67, pp 189-226

Stigum, B. P., 1975, “Asymptotic Properties of Autoregressive Integrated Moving Average Processes,” *Stochastic Processes and Their Applications*, Vol. III, pp. 315-344.

Stigum, B. P., 1990, *Toward a Formal Science of Economics*, Cambridge: MIT Press.

Stigum, B. P., 1995, ”Theory-Data Confrontations in Economics,” in: *Dialogue: Canadian Philosophical Review* 34, 581–604

Stigum, B. P., 1998, “Scientific Explanation in Econometrics,” in: *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, S. Stroem (ed.), Cambridge: Cambridge University Press.

Summers, L. H., 1991, "The Scientific Illusion in Empirical Macroeconomics," *Scandinavian Journal of Economics*, 93, 128-148.

Toulmin, S., 1953, *The Philosophy of Science: An Introduction*, New York: Harper & Row.

Van Frassen, B. C., 1980, *The Scientific Image*, Oxford: Oxford University Press.