Federal Reserve Bank of Minneapolis
Research Department Working Paper

TOWARD A STATISTICAL MACRODYNAMICS: AN ALTERNATIVE MEANS OF INCORPORATING MICRO FOUNDATIONS

Michael J. Stutzer

Working Paper 242
PACS File 3300

Revised November 1983

Support provided while visiting at the Department of Economics, Brown University, is gratefully acknowledged.

Introduction

Economic systems studied in Macroeconomics, General Equilibrium Theory, and International Trade are large, complex systems composed of numerous interacting subsystems. Macroeconomics attempts to determine government policy impacts on the behavior of aggregates resulting from the interaction of large numbers of consumers and firms. General Equilibrium Theory studies the more detailed behavior of consumers and firms via the concept of interacting markets. International Trade studies a world economy composed of numerous countries interacting through world markets.

In their attempt to understand the behavior of such complex systems, economists have followed the reductionist lead of physical scientists. They have done so by attempting to compose their theories from detailed theories of individual subsystem behavior. The "New Classical Macroeconomics" attempts to exploit dynamic optimization-based microeconomic theories of firm and consumer behavior in constructing theories explaining the behavior of macroeconomic aggregates.[1/] The "New Industrial Organization" hopes to explain oligopolistic industry behavior by formulating optimization-based subtheories which predict a firm's production, entry, and exit decisions.[2/] Many positive and normative models of economic growth have concentrated on understanding the case of a single, isolated economy's growth as a prelude to understanding the growth of economies in interaction with the world economy.

Unfortunately, interactions among large numbers of heterogeneous subsystems are often less well understood and harder to analyze than are the isolated subsystems themselves. This situation has led modellers to adopt many simplifying assumptions. For example, to derive testable hypotheses, many macroeconomic theories resort to the assumption that all firms are identical, or alternatively, that aggregate output behaves as if it were produced by a single firm.[3/] To derive testable hypotheses via comparative statics, general equilibrium theorists resort to assuming gross substitutes and other strong assumptions about the nature of market interaction.[4/] Simon and Ando's (1961) assumption of "near decomposability" is another simplifying assumption made in the face of such complexity. In all of these cases, such assumptions are not made solely to ensure computational tractability. They are also necessitated because of the lack of detailed data about the interactions of interest (e.g., cross elasticities).

The difficulty of deriving testable hypotheses in large, dynamic systems whose interactions are poorly understood has also plagued physical scientists. But physicists and mathematicians have developed general methods of analyzing such systems. Perhaps the most successful method applies to so-called "large, weakly interacting" systems, to be defined below. The method is called Equilibrium Statistical Mechanics, or the Gibbs Formalism. It is the thesis of this paper that this method, which has been successfully applied to complex dynamic problems in such diverse fields as Physics (Reif, 1965), Population Biology (Kerner, 1972), and Neurology (Cowan, 1968), can also be fruitfully applied to generate testable hypotheses in analogous complex, dynamic economic systems. To develop this thesis, a general measure theoretic version of the Gibbs Formalism, derived by R. M. Lewis (1960) and exposited by Truesdell (1960) is briefly presented. The reader is strongly urged to see these two papers for a rigorous detailed exposition. An alternative, information-theoretic justification for the Gibbs Formalism is also presented.

To illustrate the concepts and their application throughout the paper, the Gibbs Formalism is applied to an extremely simplified dynamic competitive model of the economy. Even though the model is oversimplified, we will see that it predicts many empirical findings.

First, though, it will later prove instructive to consider the physical problem which first illustrated the need for and usefulness of the Gibbs Formalism: the ideal gas problem. Formal analogies between the physics of the gas and the economics of our competitive model will also be discussed throughout the paper.

Dynamic Optimization in a Gas

Consider the problem of predicting the distribution of atoms' speeds in a low density (more precisely, ideal) monatomic gas heated to some temperature T within a container of volume V, like a balloon. The classical atomic theory of matter posits that the speeds are affected by the weak gravitational attraction the atoms have for each others, as well as by their banging into the walls of the container. Classical physicists of the 19th century believed that they had a good theory predicting the motion of an individual gas atom i (i.e., the ith subsystem) of mass M in isolation, i.e., when uninfluenced by other atoms and when moving in unenclosed space. Denote the ith atom's position at time t by $\left(q_1^i(t), q_2^i(t), q_3^i(t)\right) \triangleq q^i(t)$ and its time derivative (i.e., velocity vector) at t by $\left(\dot{q}_1^i(t), \dot{q}_2^i(t), \dot{q}_3^i(t)\right) \triangleq \dot{q}_i(t)$. Then, according to the Principle of Least Action, the ith gas atom in isolation moves as if it were an Euler path, although not a global minimizer, of the following dynamic minimization problem:

$$(1) \qquad \min_{q_i, \dot{q}_i} \int_0^T L^i(q^i, \dot{q}^i) = \int_0^T \sum_{j=1}^{3} \frac{M\dot{q}_j^{i2}}{2}\, dt$$

whose Euler paths are found by computing

$$(2) \qquad d/dt\ \partial L^i/\partial \dot{q}_j^i - \partial L^i/\partial q_j^i = d/dt\ M\dot{q}_j^i = M\ddot{q}_j^i = 0,\ j = 1,\ 2,\ 3.$$

The latter equality in (2) is a special case of Newton's Law F = MA, where F = 0. For any initial condition $(q^i(0),\dot{q}^i(0))$, the solution of (2) is $q_j^i(t) = \dot{q}_j^i(0)t + q_j^i(0)$. In isolation, the ith atom travels in a straight line with a constant velocity vector $\dot{q}^i(0)$. An alternative characterization of this is to transform problem (1) into a Hamiltonian formulation by introducing additional, so-called conjugate coordinates,

$$(3) \qquad p_j^i = \partial L^i/\partial \dot{q}_j^i = M\dot{q}_j^i,$$

the Hamiltonian H^i via the Legendre transformation,

$$(4) \qquad H^i(p^i,q^i) = \sum_{j=1}^{3} p_j^i \dot{q}_j^i - L^i(q^i,\dot{q}^i) = \sum_{j=1}^{3} {p_j^i}^2/2M$$

and defining the Hamiltonian differential equations

$$(5) \qquad \dot{q}_j^i = \partial H^i/\partial p_j^i \qquad j = 1, 2, 3 \qquad \dot{p}_j^i = -\,\partial H^i/\partial q_j^i$$

whose solution also yields the Euler paths of (1).[5] The Hamiltonian (4) is constant, or <u>conserved</u>, along trajectories $(p^i(t),q^i(t))$ in the ith atoms' six-dimensional state space, and is termed a <u>conservation law</u>. While either (1) and (2) or (4) and (5) provides a complete characterization of the behavior of the ith gas atom in isolation, the Hamiltonian formulation will prove most useful in what is to follow.

When a large number N of these gas atoms are placed in a container of volume V, they cannot travel in straight lines for very long. They are no longer isolated, for each of them must bounce off the walls of the vessel, and they are weakly mutually attracting. As a result, (4) is no longer a conservation law for an individual atom. This makes the prediction of the behavior of the system of N atoms extremely complicated. How can this behavior be predicted?

With accurate knowledge about the detailed nature of the atoms' interactions with the container and each other, it would be theoretically possible to predict the position and velocity of all atoms for all future times. This could be accomplished by formulating and solving a differential equation system incorporating these interactions. Then, the distribution of speeds and other distributions dependent on the motion of the atoms could be tabulated from the solution, given the initial position and velocity of each atom.

Unfortunately, this method is infeasible for two reasons. First, accurate knowledge about the interactions is difficult to obtain. But even if it weren't, the huge dimensionality of the system ($N \approx 10^{23}$ for a mole of gas) precludes one from measuring the initial conditions and from integrating the resulting gigantic system of complicated differential equations forward.

Fortunately in 1902, with the brilliant work of Maxwell and Boltzmann to guide him, Josiah Willard Gibbs published a method for solving this and other related problems. Because of the aforementioned analytical problems, Gibbs' method does not attempt to precisely predict the future values of system variables. Rather, it attempts to predict the distributions of system variables, and to link these distributions to system parameters. More precisely, for the ideal gas, assume that the dynamic behavior of its state vector $\gamma = (p^1,q^1,p^2,q^2\ldots p^N,q^N)$ is a trajectory from an ergodic dynamical system, and that interactions between atoms are "weak," in a sense to be defined later. Then, Gibbs' method predicts that its invariant probability measure has the so-called canonical density $g(\gamma)$:

$$\text{(C)} \qquad g(\gamma) = \prod_{i=1}^{N} \frac{e^{-\frac{1}{kT} H^i(p^i,q^i)}}{\iiint\limits_D \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{kT} H^i(p^i,q^i)}}\, dp_1^i dp_2^i dp_3^i dq_1^i dq_2^i dq_3^i$$

where D is the region the gas' container occupies in space, T is the gas temperature, K is Boltzmann's constant, and H^i is given by (4). Ergodicity implies that a typical realization of γ behaves as if it was generated by sampling from the canonical density (C). Thus, for a particular gas, suppose one wanted to compute the distribution that $\gamma(t)$ displays over an infinite time span. Gibbs' method predicts that its density is the canonical density (C). Identical replicate gases, differing only with respect to the initial conditions of their atoms, behave as different trajectories from the system with (C) density.

For example, suppose one wanted to compute the distribution of speeds among all N identical atoms with mass M. The ith atom's speed $s_i(t)$ is the norm of its velocity vector, which, when expressed in terms of conjugate coordinates, is

$$(6) \qquad s_i(t) = \sqrt{\sum_{j=1}^{3} {P_j^i}^2}/M.$$

The invariant density of (6) can be shown, via the standard change of variables in (C), to equal

$$(6a) \qquad g(s_i) = 4\pi(M/2\pi KT)^{3/2} s_i^2 e^{-Ms_i^2/2KT}; \text{ with mean}$$

$$(6b) \qquad E(s_i) = (8KT/\pi M)^{1/2} .$$

and is dubbed a Maxwellian density. The invariant density (6a) for s_i yields the speed distribution a particular atom of mass M traces out over an infinite time span. It also yields the distribution of all identical mass M atomic speeds in the gas held at temperature T, because each atoms' behavior is a different trajectory from the system with density (6a). Because the expected speed (6b) is inversely proportional to the square root of M, the average speed over time of some atom would be twice as fast as that of another atom four times its mass. Other comparative dynamics results like this one allow the researcher to test the theory on actual time series data.

There is a striking analogy between the gas problem and typical problems of dynamic microeconomic-based system models. These problems also involve large numbers of subsystems (i.e., agents, markets, individual economies, etc.) whose individual behavior is described either by dynamic optimization problems analogous to (1), or by differential or difference equations derived by other methods like stochastic optimization[6/] or by extending static theories into dynamic contexts.[7/] The task of aggregating the behavior of those heterogeneous subsystems is also plagued by lack of data and computational intractability. The underlying parameters of the individual economic subsystems and the exogenous parameters of external forces affecting the system play the role that M, T, and V do in the gas problem. Given values of such exogenous economic parameters, one can apply Gibbs' method to predict both cross-section and time-series distributions of endogenous economic variables, and to generate testable hypotheses through comparative dynamics relationships analogous (6b). Confirmation of such hypotheses corroborates the maintained assumptions about the isolated dynamic subsystem models, and about the "weak" nature of the system's interactions.

Furthermore, even if the values of some exogenous parameters are unknown, it is often possible to estimate parameters using data on observable endogenous series, and tests of the theory can be based on the estimated model.

The Gibbs Formalism According to Lewis

The following is intended to be a simplified, application-oriented exposition of the aforementioned paper by Lewis[8/] and the interpretations given it by Truesdell[9/]. Lewis' notation is normally used throughout. To simplify the exposition, some measure theoretic details as well as proofs are ignored herein. The reader is urged to see Lewis' paper for these details.

The first key concept in the Gibbs Formalism is a system theoretic description of the ith subsystem in isolation. Ergodic theory is used in the description.[10/]

The ith subsystem in isolation, or complete space, is a four-tuple $(\Gamma_i, A_i, T_t^i, m_i, y^i)$, where:

(a) Γ_i is an A_i-measurable space, σ-finite with respect to the countably additive measure m_i. The state of the subsystem at time t, denoted $\gamma^i(t)\varepsilon\Gamma_i$, is the minimum amount of information the analyst needs to forecast the future behavior of the isolated subsystem.

(b) The family of transformations, or state transition functions T_t^i: $\Gamma_i \to \Gamma_i$, $t\varepsilon[0,\infty)$, represent the ith subsystem's dynamics. Given an initial state at $\gamma^i(\tau)\varepsilon\Gamma_i$ at some time τ, $T_t^i\ (\gamma^i(\tau)) = \gamma^i(\tau+t)$, the state which will prevail at t units of time later. Lewis also requires the semigroup composition property characteristic of dynamical systems, i.e., $T_t^i \circ T_s^i = T_{t+s}^i, t,s\varepsilon[0,\infty)$. Denoting the Lesbesgue measurable subsets of $[0,\infty)$ by L, it is also required that the family T_t^i is a measurable transformation from $([0,\infty) \times \Gamma_i, L \times A_i)$ into (Γ_i, A_i).

(c) The invariant measure m_i is a countably additive measure on A_i which is preserved by the state transition functions T_t^i. The measure m_i is said to be preserved by T_t^i if $m_i\left({T_t^i}^{-1}(A)\right) = m_i(A)$ for all $t\varepsilon[0,\ \infty)$ and $A\varepsilon A_i$, i.e., if the measure of the set of all initial states $\gamma^i(\tau)$ which are transformed into states $\gamma^i(\tau+t)\varepsilon A$ after t time units have passed, is equal to the measure of A itself.

(d) A k-vector of real-valued, Borel measurable functions $y^i = (y^i_1, \ldots, y^i_k)$, where each y^i_j: $\Gamma_i \to R$, is termed a complete invariant vector, or is termed a complete vector of constants of the motion, or is termed a complete vector of conservation laws, which has the following properties:

(i) $y^i_j(T^i_t(\gamma)) = y^i_j(\gamma^i)$ for every t in $[0,\infty)$ and almost every γ^i, for each $j = 1, \ldots, k$. Each y^i_j is termed an invariant function, or a constant of the motion, or a conservation law because the value y^i_j is constant, or conserved, along a particular system trajectory starting at γ^i.

(ii) Any other conservation law z: $\Gamma_i \to R$ can be written as $z(\gamma^i) = \Phi^i(y^i(\gamma^i))$ for almost every γ^i, where Φ^i is Borel measurable on the real Borel space R^k. This means that y^i is the largest vector of "functionally independent" conservation laws.

An example of a subsystem in isolation will now be introduced. It is used throughout the rest of the paper to illustrate the concepts introduced there.

Example: A Dynamic, Competitive Industry

Consider the following simplified, partial equilibrium dynamic model of a competitive industry, denoted by i. Constant returns to scale in production are assumed for its identical firms. Except for the possibility of different endowments, households are also assumed to be identical. Without a loss of generality, this permits us to analyze the situation in terms of a single representative firm and household.

A representative household is assumed to own some nonnegative capital stock $q^i(t)$, and to rent it to a representative firm. It uses the rental income from capital and firm profits to purchase firm output, consuming some and investing the rest. Each household acts as if it maximizes the same instantaneous utility $U^i(c^i(t))$ subject to the budget constraint:

$$(7) \qquad \max_{c^i, q^i, \dot{q}^i} \int_0^\infty U^i(c^i)dt$$
$$\text{s.t.} \quad c^i + \dot{q}^i = w^i(t)q^i + \pi^i; \; q^i(0) \text{ given}$$

where the dot denotes the time derivative, where π^i is economic profit from the representative firm, and where $w^i(t)$ is the real rental of capital relative to the price of output.

The representative firm simultaneously chooses $q^i(t)$ to maximize profit at each point in time:

$$(8) \qquad \max_{q_i} \pi^i(t) = \beta_2^i q^i(t) - w^i(t)q^i(t) \; .$$

Equilibrium then requires $w^i(t) \equiv \beta_2^i$, the assumed constant, positive productivity of capital in this industry. Because of this, economic profits are zero in equilibrium, and the $q^i(t)$ stream chosen by households via (7) trivially solves (8) as well. Substituting (8) and the budget constraint into (7) and multiplying by -1 then yields the household's problem:

$$(9) \qquad \min_{q^i, \dot{q}^i} \int_0^\infty - U^i(\beta_2^i q^i - \dot{q}^i).$$

We restrict attention to utilities U^i in which the Euler equation and transversality condition for (9) yield its solution. One such environment, which will be used throughout the paper, is

$$(10) \qquad U^i = -\beta_1^i(c^i - \beta_3^i)^2 \; ,$$

where β_1^i and β_3^i are positive constants. That (10) is such a utility is shown by Hadley and Kemp.[11] Introduce the conjugate coordinate p^i as in (3) or via Pontryagin's maximum principle[12]

(11) $$p^i = \partial L^i/\partial \dot{q}^i = \partial U^i/\partial c^i = -2\beta_1^i(c^i - \beta_3^i)$$

invert (11) to solve $c^i = \beta_3^i - p^i/2\beta_1^i$, then define a Hamiltonian H^i via the Legendre transformation as in (4),

(12) $$H^i(p^i, q^i) = p^i\dot{q}^i + U^i = p^{i^2}/4\beta_1^i + (\beta_2^i q^i - \beta_3^i)p^i$$

then solve the Hamiltonian differential equations (5), and require that $(p^i(0), q^i(0))$ satisfy the transversality condition

(13) $$H^i(p^i, q^i) = 0.$$

H^i can be thought of imputed "income," where U^i is the value of consumption, and $p^i\dot{q}^i$ is the value of investment valued at its marginal opportunity cost, i.e., the marginal utility of consumption. Assuming $0 \leq q^i(0) \leq \beta_3^i/\beta_2^i$, the solution is:

(14) $$q^i(t) = \beta_3^i/\beta_2^i - \left(\beta_3^i/\beta_2^i - q^i(0)\right)e^{-\beta_2^i t} .$$

(15) $$p^i(t) = 4\ B_1^i(\beta_3^i - \beta_2^i q^i(t))$$

For the utility (10), the ith subsystem in isolation is thus specified as: $\Gamma_i = \{(p^i, q^i) \mid p^i \varepsilon [0, 2\beta_1^i\beta_3^i], q^i \varepsilon [0, \beta_3^i/\beta_2^i]\}$. A_i is the class of Borel measurable sets for the topology on Γ_i inherited from the Euclidean space R^2. T_t^i is the C^∞ flow derived from the Hamiltonian differential equations for (12). The invariant measure m_i is the Lesbesgue measure, i.e., area on (Γ_i, A_i). This follows from Liouville's Theorem, which implies that the Lesbesgue measure is preserved by any flow induced by a C^∞ a vector field (i.e., autonomous differential equation) on a compact, closed, and orientable C^∞ manifold whose divergence vanishes.[13/] Finally, $y^i = H^i$ defined by (12), as time differentiation verifies that any time independent Hamiltonian is a conservation law for the flow it induces.

The second key concept needed is the concept of weakly interacting subsystems. Consider a subsystem i interacting with other subsystems. Then, the state transition functions T_t^i and conservation law(s) it possessed in isolation may no longer be valid. In the gas problem, the ith gas atom in a gas will not travel in a straight line forever, as its T_t^i requires, due to collisions with the container and forces exerted on it by other atoms. As a consequence, its energy (4) is no longer conserved. In the economy, intermediate inputs, nonseparable utilities, and numerous other interactions with other markets preclude (7), (14) and (15) from holding for the ith industry. If the T_t^i and its associated complete vector of conservation laws no longer govern the behavior of the ith subsystem, then just what laws do govern its behavior? And how can one discover such laws without detailed information about the nature of the interactions?

The assumption of weak interaction is the key idea needed to answer these questions. Consider the system composed of N subsystems. It is described by a system state space with states $\gamma = (\gamma^1, \ldots, \gamma^N) \in \Gamma = \prod_{i=1}^{N} \Gamma_i$, the Cartesian product of the individual subsystems' state spaces. In the presence of ill-specified interactions among its subsystems, the system's state transition functions T_t are unknown. The assumption of weak interaction is that the unknown T_t have the following two properties:

Property (i): a complete vector of conservation laws for T_t on Γ is given by $y_j(\gamma) = \sum_{i=1}^{N} y_j^i(\gamma^i)$, $j = 1, \ldots, k$, and

<u>Property (ii)</u>: T_t preserves the invariant measure $m = \prod_{i=1}^{N} m_i$, on $(\prod_{i=1}^{N} \Gamma_i, \prod_{i=1}^{N} A_i)$, i.e., the <u>product measure</u> on Γ. As an additional regularity condition, Lewis assumes that $m(y^{-1}(I))$ is finite for every finite rectangle I in R^k.

The intuitive meaning of property (i) is that although the interaction makes each y_j^i vary over time, the interaction is weak enough so that the total variations cancel in the aggregate $y_j = \sum_{i=1}^{N} y_j^i$. In the gas problem, property (i) is the assumption that while the energy H^i of the ith atom given by (4) changes over time, the total energy of the gas $\sum_{i=1}^{N} H^i$ <u>is</u> conserved. That is, the atoms exchange energy with each other and their container in such a way as to conserve the total energy. In an economy of N interacting competitive industries, property (i) is the assumption that while the imputed income H^i, given by (12), of each individual industry is no longer conserved (i.e., no longer zero), inputed national income $\sum_{i=1}^{N} H^i$ <u>is</u> conserved. The industries exchange imputed income among themselves in such a way that the total national imputed income is conserved. But other than the possibility that the economy may behave <u>as if</u> this were true, is there any plausible economic theoretical basis for this assumption?

To examine the economic theoretical plausibility of Property (i) in our example, suppose there were no "interactions" across the N industries, in the following sense: In choosing among the N consumption goods c^i; i = 1, ..., N, each household rents capital $q^i(t)$ and invests in a representative firm i in each of N industries, i = 1, ..., N. Constant returns to scale prevails in each industry, and no intermediate inputs are needed. Then, households would act as if they solved.

$$(16) \qquad \max_{\substack{c^i, q^i, \dot{q}^i \\ i=1,\dots,N}} \int_0^\infty \sum_{i=1}^N U^i(c^i)dt.$$

$$\text{s.t.} \sum_{i=1}^N (c^i + \dot{q}^i) = \sum_{i=1}^N (w^i(t)q^i + \pi^i)$$

Due to constant returns to scale, $\beta_2^i \equiv w^i(t)$ and $\pi^i = 0$ for all i. Then, the additive separability in (16) permits us to show that each industry will follow its isolated path, i.e., for each i = 1, ..., N, H^i given by (12) is conserved, from which follows that $\sum_{i=1}^N H^i$ is also conserved.

However, in the presence of structural interactions causing deviations from (16), each H^i may not be conserved. For example, suppose the utility function is $\sum_{i=1}^N U^i(c^i) + I(c^1,\dots,c^N)$, where I is a nonlinear term representing the possibility that a good's consumption affects the marginal utilities of other goods' consumptions. Or, suppose that labor or intermediate goods are necessary for production of some goods, so that production functions for some industries do not depend solely on the capital allocated to them. Then, each H^i from (12) will not be conserved. Proposition (i) is perhaps plausible from the viewpoint of this economic theory if, when N is large, the interactions mentioned above are "weak" in the sense that while <u>each</u> H^i is not conserved, $\sum_{i=1}^N H^i$ will be. Real world evidence that cross elasticities of demand are insignificant for a large fraction of pairs of goods might help support the as yet ill-specified claim that the effects of the nonlinear I are weak.

Clearly, in no way do these speculations constitute an economic theoretical basis for the "realism" of Property (i). Such a theoretical basis must await research proving that the solution to a well-specified model compatible with existing economic theory and containing such interactions does indeed satisfy Property (i).

The intuitive meaning of Property (ii) can be grasped by thinking of $m_i(\gamma^i)$ as measuring the nonnormalized probability density of observing the ith subsystem in state γ^i, conditional on the information in Property (i) about the unknown T_t. Property (ii) is then interpreted as the assumption that the probability density of observing the system in state $\gamma = (\gamma^1,\ldots,\gamma^N)$, i.e., the probability density of the random vector γ, is the product of its individual component's probabilities densities. That is, Property (ii) is the assumption that, conditional on (i), the component subsystems behave in a probabilistically independent fashion. Once again, we do not attempt to produce a detailed structural model compatible with this assumption.

The assumption of weak interaction is impossible to verify directly, precisely because we <u>don't</u> specify observable interactions. For the same reason, we do not "justify" it through detailed economic structural theories of interaction. Rather, it should be viewed as a maintained "as-if" assumption to help predict system behavior. As such, it is a "reasonable" assumption only if it leads to useful predictions and insights. Its "truth" or "falsity" is irrelevant for this purpose. This view will be strengthened by the information theoretic route to the Gibbs Formalism presented later. In closing, it is interesting to note that progress in providing an analogous structural physical theory basis for the Gibbs Formalism applied to the gas problem has come only in the last decade, despite its widespread success and acclaim over the past 75 years.[14/]

<u>The third and final concept of the Gibbs Formalism is the assumption that each subsystem is small compared to the rest of the system, or, equivalently, that the system is large relative to each of its subsystems.</u> The precise, measure theoretic definition of a large system is given in Lewis, and is too complicated to state here. <u>Roughly</u>, the system is <u>large</u> relative to

each subsystem i if its conservation laws $y_j = \sum_{i=1}^{N} y_j^i$, $j = 1, \ldots, k$, would not "vary much" with ceteris paribus variations in any single y_j^i, $i=1, \ldots, N$. In the gas problem, the assumption of a large system is satisfied if there are so many gas atoms that ceteris paribus fluctuations in the energy of any one atom would hardly influence the gas' total energy. In the economy, this assumption implies that there are enough goods so that the contribution of any industry's imputed income is small compared to national imputed income. As before, no detailed structural "justification" for this assumption is given.

It is important to note that this third assumption can be weakened, if the analyst is willing to forego the possibility of predicting the long-run behavior of individual large subsystems. That is, if some subsystem i's y_j^i do significantly contribute to the total y_j, it is not possible to calculate the long-run behavior of γ^i, although it is still possible to compute the long-run behavior of the small subsystems' states and of variables depending solely on a small subsystem's state. Thus, the existence of large subsystems (i.e., atoms, industries, etc.) in this sense does not totally invalidate the Gibbs Formalism. It merely restricts the type of predictions that can be reliably made. Finally, the regularity condition in Property (ii) is satisfied in our example, due to the finite measure of the product space Γ.

Long-Run Behavior of Subsystems in Large, Weakly Interacting Systems

Lewis, in his Theorem 2, shows how the long-run time average f_i^* of any real valued, m_i-integrable function $f_i(\gamma^i(t))$ of the ith subsystem's state vector γ^i can be computed in large, weakly interacting systems.15/ Let $T_t(\gamma^i(\tau)) \triangleq \gamma^i(\tau+t)$. Then, the long-run time average f_i^* of some function $f_i(\gamma^i(t))$ starting from initial condition $\gamma^i(0)$, exists for almost every (m_i) $\gamma^i(0)$, and is computed by

(17) $$f_i^* \stackrel{\Delta}{=} \lim_{T\to\infty} \frac{1}{T}\int_0^T f_i(\gamma^i(t))dt = \frac{1}{z_i}\int_{\Gamma_i} f_i(\gamma^i)e^{-\alpha\cdot y^i(\gamma^i)}dm_i \stackrel{\Delta}{=} \bar{f}_i$$

where

(18) $$z_i(\alpha) = \int_{\Gamma_i} e^{-\alpha\cdot y^i(\gamma^i)}dm_i$$

The function $z_i(\alpha)$ is called the partition function for the ith subsystem, and is assumed to exist in some open neighborhood of k-vectors α. The correct value of the vector α to use in computing $\bar{f}_i$ in (17) depends on the value assumed by the subsystem's complete vector of conservation laws y^i, as described in Lewis' corollary to his Theorem 2. Fortunately, the value of α is independent of the particular subsystem one is interested in. For most applications envisioned here, one would not need to know the value of α, so one would not need to know the (average) value of y^i. Rather, one could treat α as a vector of free parameters, and use econometric methods to estimate α from data.

Equations (17) and (18) have a nice probabilistic interpretation. Equation (17) shows how the long-run time average of any m_i-integrable function of the ith subsystem's state γ^i can be computed by a so-called phase average $\bar{f}_i$ over its state space Γ_i. The multivariate function $e^{-\alpha\cdot y^i(\gamma^i)}$ can be thought of as determining a "probability density" for the ith subsystem's state γ^i, with $z_i(\alpha)$ thought of as its normalization constant. To see this, let f_i be the indicator function for any measurable set E_i. The long-run time average of this indicator function yields the long-run fraction of time the state $\gamma^i(t)$ spends in E_i, and is dubbed the mean sojourn time of the subsystem in E_i. The mean sojourn time is a frequentist way to define the probability $p_i(E_i)$ that the subsystem state is in some subset E. Computing (17) for this indicator f_i, one obtains:

$$(19) \qquad p_i'(E_i) = \frac{1}{z_i(\alpha)} \int_{E_i} e^{-\alpha \cdot y^i(\gamma^i)} dm_i.$$

When the subsystem has only $k = 1$ independent conservation law, the probability distribution in (19) is the famous canonical distribution of Gibbs. For the general case of $k > 1$, Truesdell has dubbed it the polycanonical distribution.

The polycanonical distribution is also of use in calculating statistics of some time series f_i other than its long-run average. For example, to compute the long-run variance of some square integrable f_i

$$(20) \qquad \text{var } f_i \triangleq \lim_{T\to\infty} \frac{1}{T} \int_0^T \left(f_i(\gamma^i(t)) - f_i^*\right)^2 dt$$

one simply expands (20) and applies (17) to obtain

$$(21) \qquad \text{var } f_i = \overline{f_i^2} - \overline{f}_i^2 ,$$

which is the familiar mean (i.e., phase average) squared minus the square of the mean. Other long-run moments are similarly computed.

The polycanonical density can be similarly used to compute long-run covariances between two time series $f_i(\gamma^i(t))$ and $g_i(\gamma^i(t))$. Defining the long-run covariance of square integrable f_i and g_i as

$$(22) \qquad \text{cov } (f_i,g_i) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \left(f_i(\gamma^i(t))-f_i^*\right) \left(g_i(\gamma^i(t)-g_i^*\right)dt$$

one expands (22) and again applies (17) to obtain the familiar representation

$$(23) \qquad \text{cov}(f_i,g_i) = \overline{f_i g_i} - \overline{f}_i \overline{g}_i.$$

To make matters concrete, consider our ith industry (10), with conservation law (12). The invariant measure m_i is the Lesbesgue measure μ. Using the canonical distribution (19), the mean sojourn time of $\gamma^i = (p^i,q^i)$ in some measurable subset E_i of Γ_i is given by

$$(24) \qquad p_i'(E_i) = z_i(\alpha)^{-1} \int_{E_i} \exp(-\alpha({p^i}^2/4\beta_1^i + (\beta_2^i q^i - \beta_3^i)p^i))d\mu$$

where

$$(25) \qquad z_i(\alpha) = \int_0^{\beta_3^i/\beta_2^i} \int_0^{2\beta_1^i\beta_3^i} \exp\,(-\alpha({p^i}^2/4\beta_1^i + (\beta_2^i q^i - \beta_3^i)p^i))dp^i dq^i.$$

Unfortunately, the integrals in (24) and (25) cannot be evaluated in terms of elementary functions. In the empirical application below, they will be numerically integrated by computer.

An Application of the Subsytem's Canonical Density

The marginal density $g_i(q^i)$ of (24), defined by

$$(26) \qquad g_i(q^i) = z_i(\alpha)^{-1} \int_0^{2\beta_1^i\beta_3^i} \exp(-\alpha({p^i}^2/4\beta_1^i + (\beta_2^i q^i - \beta_3^i)p^i))dp^i$$

yields the density of a representative firm's capital stock in the ith industry. Note that it only depends on the parameters $(\beta_1^i, \beta_2^i, \beta_3^i)$ characterizing the ith industry, rather than on their values for other industries. Because of this, it is possible to fit this density to industry data without losing too many degrees of freedom. There is a paucity of time series data on actual individual firms' input utilization, though. However, because our model assumes that all firms in an industry are identical except for their initial capital stocks, each firm's capital stock time series can be viewed as a trajectory generated by the dynamical system whose invariant density is (26). Then, (26) should also yield the stationary density of capital stock <u>across</u> firms in the ith industry.

Differentiating (26) under the integral sign twice, we find that $dg_i/dq^i < 0$ and $d^2g_i/d{q^i}^2 > 0$, a result which holds up for any utility U^i yielding an interior solution to (9).[16] Under the interpretation above, this means that the density of firm capital q^i in the ith industry is a decreasing, convex function of firm capital q^i. Because of the assumed constant returns to scale, the density of firm output $\beta_2^i q^i$ is proportional to that of q^i. Thus, choosing either firm output or capital as a measure of firm size, the density of firm size should be convex downward. Furthermore, by additionally assuming that the capital/labor ratio is a constant across firms in the ith industry, easily obtainable histograms of the industry's employment/firm should also decline throughout the range of firm employment.

Industry data seems to support this theoretical prediction. For example, Lawrence Klein[17/] states

> "There are so many firms with only a few employees that there appear to be continuously falling frequency distributions. They do not, as in the case of income, rise to a modal peak then decline. They begin with a modal peak and then decline throughout."

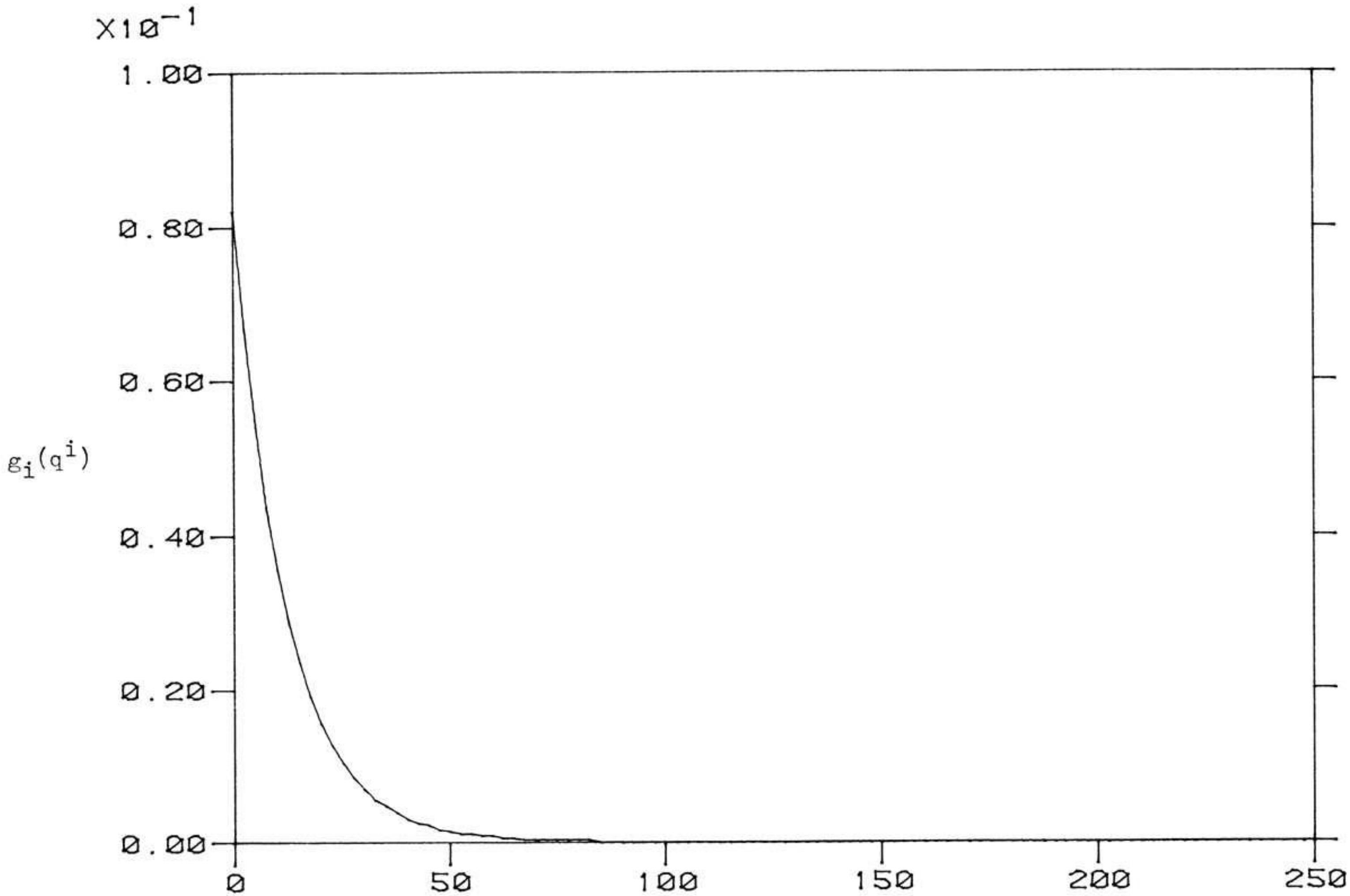
Corroboration of this finding comes from the U.S. Census Bureau's Census of Manufacturers. Every five years, their census reports the distribution of establishment employment size in the 20, two-digit SIC industries. A typical distribution from the latest (1977) census is that for SIC 29, Petroleum and Coal Products, shown in Table 1. By experimenting with different values for α and the parameter values β_1^i, β_2^i, β_3^i in (26), we have been able to produce a distribution similar to this in a hypothetical industry with no firms larger than 250 employees, by assuming that the capital/labor ratio is one. Its density is graphed in Figure 1. It is difficult to accurately compute the requisite numerical integrals when the domain of integration is much larger than this. Hopefully, proper scaling of data will help avoid this problem in the future. To summarize, we have shown that a weakly interacting economy with dynamic competitive industries, whose firms produce under constant returns to scale, qualitatively yields the observed distribution of establishment size.

SIC 29 Petroleum and Coal Products			Canonical Density (26)[18]
Employees	# of Establishments	% of Total	% of Total
1-4	710	32.2	33.4
5-9	422	19.1	22.0
10-19	300	13.6	24.3
20-49	348	15.8	18.0
50-99	148	06.7	2.1
100-249	150	06.8	0.1
250-499	63	02.9	--
500-999	45	02.0	--
1000-2499	18	00.8	--
72500	2	0.1	--
TOTAL	2206		

TABLE 1: An actual industry vs. a canonical density (26)

$\alpha = 12.5 \quad \beta_1^i = .25 \quad \beta_2^i = .008 \quad \beta_3^i = 2$

Figure 1 Marginal Canonical Density for $\alpha = 12.5$, $\beta_1^i = .25$, $\beta_2^i = .008$, $\beta_3^i = 2$



In the future, we hope to find minimum chi-square estimates for α, β_1^i, β_2^i, and β_3^i for numerous industries, under the theoretical restriction that α is constant across industries. Such a procedure would also permit us to test the hypothesis that α is constant across industries.

Comparative Subsystem Dynamics in Large, Weakly Interacting Systems

The statistics of real valued functions of the ith subsystem's state, $f_i(\gamma^i(t))$, depend on the vector of parameters β^i through the poly-canonical distribution (19). Assuming that $y^i(\gamma^i;\beta^i)$ is differentiable in β^i and that β^i is a known or estimated parameter vector, one can differentiate relations like (17) or (21) with respect to β^i to determine changes in the long-run statistics with respect to changes in β^i. This exercise is dubbed comparative subsystem dynamics. Assume that these parameter changes occur "slowly enough" so that long-run time averages are good approximations to the actual partial time means $\frac{1}{T}\int_0^T f_i(\gamma_i(t))dt$ that occur between parameter changes. Then, comparative subsystem dynamics give a good indication of the actual subsystem behavior over time. Such "slowly varying" parameter changes are called quasi-static in physics. The assumption of quasi-static change is, roughly, the usual assumption used by economic analysts in justifying the relevance of comparative statics exercises.

For the density tabulated in Table 1 and graphed in Figure 1, we compute the change in mean firm capital stock and in mean household consumption resulting from a ceteris paribus increase of 1 percent in each parameter. The results are summarized in Table 2.

	Original Parameters	1% increase in: α	β_1^i	β_2^i	β_3^i
mean firm capital $\bar{q}^i$	12.652	13.939	12.502	12.527	12.478
mean houshold consumption $\bar{c}^i = \beta_3^i - \bar{p}^i/2\beta_1^i$	.362	.172	.360	.363	.362

TABLE 2: Comparative Subsystem Dynamics of Density (26)

$\alpha = 12.5, \quad \beta_1^i = .25, \quad \beta_2^i = .008, \quad \beta_3^i = 2.0$

From Table 2, we see that the mean capital stock per firm falls and mean household consumption rises when the productivity (rental) of capital β_2^i, increases, in accord with intuition. The quantitative increase in mean consumption is small, though. Also, note that changes in the systemwide parameter α have a far more dramatic effect on these variables than do the other parameters, which only affect this particular industry.

Long-Run Behavior of a System in Weak Interaction with Its Environment

Many variables of interest depend on the states of numerous subsystems. For example, a macroeconomist interested in total national output, consumption, or investment in our economy needs to compute statistics of functions $f(\gamma^1(t),\ldots,\gamma^N(t))$. To do so, we assume that the system itself is in weak interaction with some ill-understood environment. Redenoting the system $\left(\prod_{i=1}^N \Gamma_i, \prod_{i=1}^N A_i, \sum_{i=1}^N y^i, \prod_{i=1}^N m_i, T_t\right)$ by (Γ',A',y',m',T_t'), the environment is specified as another subsystem, with state space Γ'', measurable sets A'', state transition functions T_t'', invariant measure m'' and a complete vector of conservation laws y''. The assumption of weak interaction between the system and its environment is that $y = y' + y''$ is a complete vector of conservation laws for the coupled system/environment $\Gamma = \Gamma' \times \Gamma''$, the product measure $m = m' \times m''$ is an invariant measure for the (unknown) system/environment state transition functions T_t on Γ, and that $m(y(I)) < \infty$, for every finite rectangle I in R^k. It is also essential to assume, as before, that the combined system/environment is large relative to the system. That is, it is assumed that y "does not vary too much" with ceteris paribus fluctuations of y', the precise definition being given in Lewis, p. 360. In the economic example used here, the environment would include exogenous elements which impinge on the nationwide economy and that are not explicit in the model.

Given these additional assumptions, one can again apply Theorem 2 of Lewis[19/] to compute time averages of system, dubbed macro, variables $f(\gamma^1(t),\ldots,\gamma^N(t))$, via the use of a polycanonical density. Remembering that $\gamma' = (\gamma^1,\ldots,\gamma^N)$

$$(27) \qquad f^*(\gamma'(0)) = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(\gamma'(t))dt = \frac{1}{z'(\alpha)}\int_{\Gamma'} f(\gamma')e^{\alpha\cdot y'}dm' \triangleq \bar{f}$$

where the system's partition function $z'(\alpha)$ is:

$$(28) \qquad z'(\alpha) = \int_{\Gamma'} e^{-\alpha\cdot y'}dm' = \int_{\Gamma_1}\int_{\Gamma_2}\cdots\int_{\Gamma_N} e^{-\alpha\cdot y^1}e^{-\alpha\cdot y^2}\cdots e^{-\alpha\cdot y^N}dm_N dm_{N-1}\cdots dm_1$$

or, rewriting (27) for the indicator function f on a set $E = \prod_{i=1}^{N} E_i$, $E_i \in A_i$, and using (19), find:

$$(29) \qquad p'(E) = \prod_{i=1}^{N} p_i'(E_i) \; .$$

Furthermore, the long-run variance of a square integrable (w.r.t. (29)) macro variable f can be computed by (20) with $f = f(\gamma'(t))$ and will yield a formula similar to (21);

$$(30) \qquad \text{var } f = \overline{f^2} - \bar{f}^2 \; .$$

Finally, the long-run covariance between two square integrable macro variables $f(\gamma'(t))$ and $g(\gamma'(t))$ can be computed by (22) yielding:

(31) $$\mathrm{cov}\ (f,g) = \overline{fg} - \bar{f}\ \bar{g}.$$

An Application of Macrodynamics

Suppose we are interested in computing moments of aggregate output, or the equivalent aggregate income. Letting n denote the number of households, aggregate income y is:

(32) $$y = n \sum_{i=1}^{N} w^i q^i = n \sum_{i=1}^{N} \beta_2^i q^i$$

The joint distribution (29) factors into the product of its marginals, so the component random variables in the sum (32) are independent. The absolute values of the components are also uniformly bounded by the constant $M = \max_{i=1,\ldots N} n\beta_3^i$; because $q^i \varepsilon [0,\beta_3^i/\beta_2^i]$. Therefore, if we additionally assume that β_1^i, β_2^i, β_3^i and α are such that the series $(\mathrm{var}\ (y_m))^{1/2}$ diverges, where $y_m = n \sum_{i=1}^{m} \beta_2^i q^i$, it follows from the Lindeberg Central Limit Theorem[20/] that the standardized y_m converge in distribution to the standard normal. One economy obviously satisfying this assumption is that where all industries are identical.

Given the above assumption, one might expect to find that detrended national output will be normally distributed about its trend over an infinite time span. To test this prediction, we first detrended the postwar, quarterly time series of U.S. real gross domestic product, starting in 1947. To do so, we regressed this series on a fourth order polynomial time trend. The 146 estimated residuals formed the detrended output series. As predicted, it does indeed appear that this series is normally distributed, for its mean is zero to six decimal places and its median is only .63, which is quite close to its mean when measured relative to its standard deviation of \$25.8 billion. Thus, the distribution appears symmetric. More importantly, the distribution passed the Kolmogorov-Smirnov test for normality, as described in Kendall and Stuart[21/], at the .05 level of significance.

When Parameters Are Constant But Unknown: Gibbsian Econometric Modelling of Time Series Data

To test the theory within regimes of constant $\beta = (\beta^1,\ldots,\beta^N)$, or to determine any unknown parameters in the α and β vectors, one could simultaneously estimate the vectors α and β of the polycanonical density $z'(\alpha,\beta)^{-1}e^{-\alpha\cdot y'(\gamma';\beta)}$ from time series data on observable system time series data $f(\gamma'(t))$. To do so, note that (27) yields an exact relationship between the long-run mean f^* of an observable f and its phase average $\overline{f}$ with respect to the polycanonical density. Using E for the expectation operator, the effects of measurement, specification and/or other modelling errors might be summarized by the more general hypothesis:

$$(32) \qquad E\left(f^* - \int_{\Gamma'} \frac{fe^{-\alpha\cdot y'(\gamma';\beta)}dm'}{z'(\alpha;\beta)}\right) \triangleq E\big(F(\gamma';\alpha,\beta)\big) = 0.$$

In the jargon of time series econometrics, (32) is a "population orthogonality condition" implied by this theory. The theory delivers as many such orthogonality conditions as there are observable series f. Given a finite series of past observations $f(\gamma'(0)),\ldots,f(\gamma'(T))$ on at least $k + r$ (i.e., the dimensionality of (α,β)) observable f's, one should be able to apply the Generalized Method of Moments (GMM) estimator of Hansen (1982) to obtain consistent and asymptotically normal estimators $\hat{\alpha}$ and $\hat{\beta}$. The number of observable series $f(\gamma'(t))$ is, in principle, infinite. Thus, there will be far more than $k + r$ orthogonality restrictions at our disposal. Hansen has also concocted a test of the "over-identifying restrictions", i.e., the number of restrictions we have in excess of $k + r$, which utilizes the chi-squared distribution. Such a test provides a formal means of testing the theory within regimes of constant β.

If the state γ' itself is observable, one could also utilize minimum chi-square estimation to fit the system's polycanonical density, as envisioned earlier. Furthermore, even if γ' is not observable, the density of some observable $f(\gamma')$ can be derived from the polycanonical density and the functional form of f. It can then be fit to data.

If these econometric tests are successful, the resulting estimated polycanonical density $z'(\hat{\alpha},\hat{\beta})^{-1}e^{\alpha\cdot y'(\gamma';\hat{\beta})}$ provides what I dub a <u>Gibbsian Econometric Model</u> (GEM) of the system. The GEM could be of use for unconditional forecasting of time averages and other moments of observables $f(\gamma(t))$. It may also be of use for conditional forecasting of policy interventions, which can be modelled by predictable, quasi-static changes of α and β.

An Alternative Derivation of The Polycanonical Distribution: The Maximum Entropy Formalism

Another argument supporting the plausibility of the polycanonical distribution is grounded in information theory. This argument was initially advanced and promoted by E. T. Jaynes in a series of papers[22/] and has been dubbed the Maximum Entropy Formalism (MEF). The MEF has also been widely used in regional and urban economics for modelling transportation problems and interregional commodity flows.[23] The following presentation of this formalism follows results from the book by Guiasu.[24/]

As earlier, suppose (Γ',A') is a measurable space, with a probability measure m'. The measure m' is to be thought of as a prior probability measure, summarizing the researcher's subjective probability of events in A'. Now, suppose on the basis of some other information gleaned about the subsystem, the researcher revises these probabilities, replacing m' with a measure p', absolutely continuous with respect to m'. Both of these measures contain information about the subjective likelihood of various events in A'. But how can we quantify the amount of information, or equivalently, the lack of uncertainty, expressed by a particular measure? And how can we quantify the variation of information that occurs when passing m' to a p' which is absolutely continuous with respect to m'? By the Radon-Nikodym Theorem, there exists a nonnegative real valued function on Γ' such that $p'(E) = \int_E \phi dm'$. Then, with motivation to be provided shortly, define the variation of information $I(p'|m')$ in passing from m' to p' to be:

(33) $$I(p'|m') \triangleq \int_{\Gamma'} \phi(\gamma')\ln\phi(\gamma')dm'$$

Definition (33) has many intuitively appealing properties capturing the flavor of the change in information in moving from m' to p'. For example, suppose that $p' = m'$. Then there has obviously been no change in information, and definition (33) accordingly yields $I(p'|m') = 0$. A simple argument in Guiasu[25] shows that I is nonnegative, finite in value when ϕ is m'-square integrable, and zero only when $p' = m'$. Thus, an actual change of distribution always conveys some information.

For the moment, assume that Γ' is a discrete set $\{\gamma_1',\ldots,\gamma_n'\}$, with $p'(\{\gamma_i'\}) = p_i'$. Then, if m' is the uniform probability,

(34) $$I(p'|\text{uniform}) = \sum_{i=1}^{n} np_i'\ln(np_i') \cdot \frac{1}{n} = \ln(n) + \sum_{i=1}^{n} p_i'\ln p_i' \;.$$

Defining the entropy of the measure p' by $E(p') = -\sum_{i=1}^{n} p_i'\ln p_i'$

(35) $$E(p') = \ln(n) - I(p'|\text{uniform}) = -\sum_{i=1}^{n} p_i'\ln p_i' \;.$$

Because I is nonnegative, we see that $E(p')$ attains its maximum of $\ln n$ when p' is the uniform distribution. Interpreting the entropy E as the "amount of uncertainty" in the measure p', we would then conclude that the uniform measure has the most uncertainty. This is also in accord with intuition.

One could possibly construct alternatives to the entropy (35) as ways to quantify the amount of uncertainty in a measure in a discrete space Γ'. Khinchin[26] postulated four eminently reasonable axioms as desiderata for a reasonable quantification of the amount of uncertainty in p'. Khinchin then showed that the only possible functions satisfying these axioms are proportional to the entropy $-\sum_{i=1}^{n} p_i' \log p_i'$, where the logarithm can have any base. Thus, for discrete state spaces Γ', we can appeal to these reasonable axioms to support the claim that our entropy $E(p') = -\sum_{i=1}^{n} p_i' \ln p_i'$ reasonably captures the concept of the amount of uncertainty.

Suppose that the <u>only</u> information about the discrete distribution that one has can be expressed in terms of the set of expected value constraints:

$$(36) \qquad \overline{y_j'} = \sum_{i=1}^{n} p_i' y_j'(\gamma_i'), \quad \text{for } j = 1, \ldots, k;\ k < n.$$

where each y_j' is a real valued, Borel-measurable function.

Suppose that, knowing (36), one wanted to revise the prior distribution m' that one held prior to knowing <u>any</u> information (36). It seems reasonable to choose p' to be the revised distribution containing the most uncertainty $E(p')$ subject to the information constraints (36). That is, one would choose nonnegative probabilities $p_1', \ldots, p_n'$ summing to one and maximizing the strictly concave function (35) subject to the linear equality constraints (36). The approach of choosing p' to maximize entropy subject to whatever informational constraints the researcher possesses is termed the Maximum

Entropy Formalism (MEF). Note from (35) that MEF is equivalent to minimizing the variation of information $I(p' | \text{uniform})$ obtained when passing from a complete lack of information, i.e., from the uniform prior distribution m', to the information embodied in (36).

A trivial modification of Theorem 16.1 in Guiasu shows that the solution maximizing (35) subject to (36) and the normalization constraint is:

(37) $$p_i' = z'(\alpha)^{-1} e^{-\alpha \cdot y'(\gamma_i')}$$

(38) $$z'(\alpha) = \sum_{i=1}^{n} e^{-\alpha \cdot y'(\gamma_i')}$$

where the dot denotes the inner product of the k-vector y' and the vector α of Lagrange multipliers for (36). The measure p' is a discrete version of the polycanonical distribution. Such a distribution is "maximally noncommittal," in that it conveys no more information than that contained in the explicitly recognized constraints (36). Any other distribution implicitly contains more information than that explicitly recognized, and is thus ad hoc.

When the state space Γ' contains a continuum of elements, the axiomatic result of Khinchin alluded to earlier fails to go through. Therefore, the axiomatic basis for choosing the polycanonical distribution as the continuous distribution containing the most uncertainty subject to information constraints is not present.

However, the desirable properties of the variation of information (33) mentioned above do not require a discrete sample space. So, it still seems reasonable to choose the distribution p' causing the least variation of information in moving from a prior m' to p', subject to any information constraints. Assume that the only information present is represented by the expected value constraints:

(39) $\int_{\Gamma'} \phi(\gamma')y_j'(\gamma')dm' = \overline{y_j'}$ $j = 1, \ldots, k$

where, on each Γ', y_j' is a real valued, Borel-measurable function. Then, minimizing (33) subject to (39), $\phi \geq 0$, and $\int_{\Gamma'} \phi dm' = 1$, is equivalent to maximizing $-I(p'|m')$, dubbed generalized entropy, subject to these constraints. Replacing the discrete entropy E by -I, and trivially modifying Theorem 16.1 in Guiasu:

(40) $\phi(\gamma') = z'(\alpha)^{-1}e^{-\alpha \cdot y'}$

(41) $z'(\alpha) = \int_{\Gamma'} e^{-a \cdot y'}dm'$

(42) $p'(E) = z'(\alpha)^{-1}\int_{E'} e^{-\alpha \cdot y'}dm'$

Thus, independent of theory, MEF has delivered the polycanonical distribution. Among distributions absolutely continuous with respect to m'[27/] and compatible with the constraints (39), the polycanonical distribution is the one which causes the minimum variation of information from m'. Any other distribution implicitly implies a bigger change in information than that warranted solely by the explicitly recognized constraints (39), and is thus ad hoc.

Towards a Statistical Macrodynamics: Epilog

It has been argued herein that the Gibbs Formalism provides a useful, tractable means of formulating testable hypotheses, at both the micro and macro level, in complex, dynamic economic systems. These testable hypotheses are formulated from a model of the economic system that forces the researcher to specify only:

(1) A reasonable dynamic model of each subsystem's isolated behavior. Such a model is conveniently generated by either deter-

ministic or stochastic dynamic optimization based behavioral hypotheses, but need not be.

(2) A measure-theoretic hypothesis that the system is large relative to its subsystems, which are in weak interaction with one another. This hypothesis is only necessary when one wishes to formulate testable hypotheses about an individual subsystem's behavior, i.e., about endogenous micro variables.

(3) A measure-theoretic hypothesis that the system's exogenous environment is large relative to the system, and is in weak interaction with the system. This hypothesis is only necessary when one wants to formulate testable hypotheses about systemic behavior, i.e., about endogenous macro variables.

<u>or</u>, from the Maximum Entropy Formalism for statistical inference,

(4) one must specify <u>only</u> that the mean values of the conservation laws are known.

A rejection of testable hypotheses based on the Gibbs Formalism might be remedied by changing the specification of (1), i.e., by proposing alternative micro models of subsystem behavior. Acceptance of these testable hypotheses corroborates the correctness of the specification of (1).

Footnotes

[1] See Sargent, Chapter 16.

[2] See Baumol.

[3] See Sargent, Chapter 16.

[4] See Quirk and Saposnik, Chapter 6.

[5] See Samuelson (1972) for a summary of the relations between dynamic optimization problems and their Hamiltonian formulations.

[6] See Lucas and Prescott for an example.

[7] This is sometimes done by time differentiation of static equilibrium conditions. See, for example, Sargent, Chapter 1.

[8] See Lewis.

[9] See Truesdell.

[10] See the excellent lecture notes of Petersen for an introduction to Ergodic Theory.

[11] See Hadley and Kemp, pp. 66-67.

[12] See Hadley and Kemp, Chapter 4.

[13] See Cornfeld, Fomin and Sinai, pp. 47-48.

[14] See Lanford (1974) for details of their progress.

[15] See Lewis, p. 362. In accord with Lewis' ideas we have in mind the interpretation (II) on p. 368, i.e., that the ith subsystem is in weak interaction with the N-1 others. The concordance with Lewis' notation is as follows:

Denote $(\Gamma_i, A_i, T_t^i, m_i, y_i) = (\Gamma', A', T_t', m', y')$.

Define $(\Gamma'', A'', T_t'', m'', y'') = \left(\prod_{j\neq i} \Gamma_j, \prod_{j\neq i} A_j, \prod_{j\neq i} T_t^j, \prod_{j\neq i} m_j, \sum_{j\neq i} y^j \right)$.

Assume that m" is preserved by T''_t and that y" is a complete set of conservation laws for $\prod_{j \neq i} \Gamma_j$. Then, Γ' and Γ'' are said to be in weak interaction if the product measure m' x m" on $\Gamma' \times \Gamma''$ is preserved by T_t, and if y' + y" is a complete set of conservation laws for T_t on $\Gamma = \Gamma' \times \Gamma''$. The definition of Γ'' being larger than Γ' is given on p. 360.

16/First, we must find the Hamiltonian for the general case. Because U^i is assumed strictly concave, one can invert $p^i = dU^i/dc^i$ to obtain $c^i = c^i(p^i)$. Then, $\dot{q}^i = \beta_2^i q^i - c^i(p^i)$ and $H^i = U^i + p^i \dot{q}^i = U^i(c^i(p^i)) + p^i \beta_2^i q^i - p^i c^i(p^i)$. Differentiating $g_i(q^i)$ twice for this general H^i shows that it is downward sloping and strictly convex.

17/See Klein, p. 150.

18/These were computed by integrating (26) over the rounded ranges [0,5], [5,10], [10,20], etc., by a Gaussian integration routine.

19/See Lewis, p. 368.

20/See Ash, pp. 336-337.

21/See Kendall and Stuart, vol. 2, pp. 477-488.

22/See the referenced papers of Jaynes.

23/See Wilson.

24/See Guiasu, Chapters 1, 2 and 16.

25/See Guiasu, Chapter 2, Theorem 2.2.

26/See Khinchin, Uniqueness Theorem, p. 9. Also, see Guiasu, Chapter 2.

27/The invariant measure m' of our theory was not normalized. If $m'(\Gamma') < \infty$, as in our example, then interpretation of m' as a prior probability is correct if we reinterpret m' to be $m'/m'(\Gamma')$. If $m'(\Gamma') = \infty$, though, then we can only interpret $I(p'|m')$ to be the variation of information obtained when passing from the invariant measure m' to the probability measure

p'. As long as ϕ is m'-square integrable, as assumed, the theorems cited in the paper are still valid.

References

Ash, R., Real Analysis and Probability, Academic Press, 1972.

Baumol, W., "Contestable Markets: An Uprising in the Theory of Industry Structures," American Economic Review, V. 72, #1 (March 1982).

Cornfeld, I. P., Fomin, S. V., Sinai, Ya. G., Ergodic Theory, Springer-Verlag, 1982.

Cowan, J. D., "Statistical Mechanics of Neutral Nets," in E.R. Caniello, ed., Neutral Networks, Springer-Verlag, New York 1968.

Guiasu, S., Information Theory with Applications, McGraw-Hill, 1977.

Hansen, L. P. "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica, V. 50, No. 4, July 1982.

Hadley, G. and M. C. Kemp, Variational Methods in Economics, North-Holland, 1971.

Jaynes, E., "Information Theory and Statistical Mechanics," Physical Review, vol. 106, 1957, p. 620.

Jaynes, E., "Where do we Stand on Maximum Entropy?" in Levine and Tribus, eds. The Maximum Entropy Formalism, MIT Press, 1979.

Kendall, M. and A. Stuart, The Advanced Theory of Statistics, vol. 2, 3rd ed., Griffin, 1973.

Kerner, E. H., Gibbs Ensemble: Biological Ensemble, Gordon and Breach, New York, 1972.

Kerner, E. H., "Dynamical Aspects of Kinetics," Bulletin of Mathematical Biophysics, V. 26, 1964. (Also reprinted in Kerner (1972).)

Khinchin, Mathematical Foundations of Information Theory, Dover Publications, 1957.

Klein, L. R., An Introduction to Econometrics, Prentice-Hall, 1962.

Lanford, O. E., "Time Evolution of Large Classical Systems" in J. Moser, ed. Dynamical Systems Theory and Applications, Springer-Verlag Lecture Notes in Physics, No. 38, 1974.

Lewis, R. M., "Measure-Theoretic Foundations of Statistical Mechanics," Archives of Rational Mechanics and Analysis, vol. 5, 1960.

Lucas, R. and E. Prescott, "Investment Under Uncertainty," Econometrica, V. 39, #5, 1971.

Petersen, K., Introductory Ergodic Theory, Department of Mathematics, University of North Carolina, Chapel Hill, N.C., 1971.

Reif, F., Fundamentals of Statistical and Thermal Physics, McGraw-Hill, 1965.

Samuelson, P., "The General Saddlepoint Property of Optimal Control Motions," Journal of Economic Theory, V. 5, 1972.

Sargent, T., Macroeconomic Theory, Academic Press, 1979.

Sato, R., Theory of Technical Change and Economic Invariance, Academic Press, 1981.

Simon, H. and A. Ando, "Aggregation of Variables in Dynamic Systems," Econometrica, V. 29, April 1961.

Truesdell, C., "Ergodic Theory in Classical Statistical Mechanics," in P. Caldirola, ed., Ergodic Theories, Proc. of International School of Physics (Enrico Fermi), Academic Press, 1960.

Wilson, A. G., Entropy in Urban and Regional Modelling, Pion Limited, 1970.