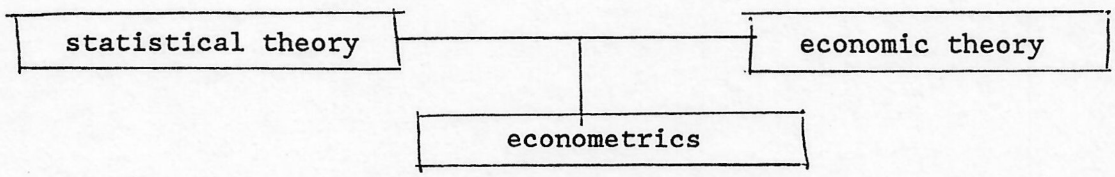
The Influence of Economic Theory on Econometrics

Arnold Faden

November 30, 1993

Introduction

When the Econometric Society was founded in 1932, the declared goal was to build econometrics on the twin foundations of statistical theory and economic theory (and perhaps economic data as the third input).



This program has been adhered to faithfully, but not always with happy results. Econometrics has inherited bad genes as well as good, and defects in its parents have passed down as defects in the offspring. I will concentrate on the inheritance from economic theory. Specifically, I will argue that the incorrect preoccupation of economic theory with equilibrium and simultaneous equations led to a decade of wrong-headed development by the Cowles Commission, followed by a decade of first-generation economy-wide econometric models equally wrong-headed.

Another topic to be covered is the resolution of the conflict between "subjective" and "objective" probability. Without this, the very meaning of stochastic econometric models is problematic.

From Economic Theory to Econometrics

Where do econometric relations come from? There are two sources. The first is the "epsilon-method": Take a relation postulated by economic theory, and add an epsilon (a "random shock") to it. Thus $F(x_1, \ldots, x_n) = 0$ becomes $F'(x_1, \ldots, x_n, \epsilon) = 0$.

The second source relates to forecasting. To forecast x_t, one thinks of plausible past variables (including lagged x values), and a plausible functional form, and puts them together, again including epsilons: $x_t = F(x_{t-1}, \ldots, y_{t-1}, \ldots, \epsilon_t \ldots)$.

In general, the first source yields "structural equations", while the second source yields various time series approaches, such as ARMA models, distributed lags, and combinations of these. In general, the former are thought of as deeper and more justified theoretically than the latter.

Time and Uncertainty in Economic Theory

Most of the relations postulated in economic theory are timeless and deterministic - that is, they do not in themselves have any reference to time or probability. (Think of utility functions, demand functions, material balances, input-output relations, etc.). Econometrics exists as a separate discipline precisely because these extra-theoretic considerations must be brought in to establish contact between economic theory and the real world.

It behooves us to examine economic theory very carefully to see if anything essential is left out by the omission of time and uncertainty - and the closely related concepts of equilibrium and causation.

Take uncertainty first. The justification of determinism goes something like this: In reality there are no exact relations among variables (except for definitive identities). But there is a central tendency that arises from the interaction of rational agents, and the actual world consists of this tendency buffeted by random disturbances. Thus the analogy is to a distribution and, say, its mean: The best that can be done is to predict what the mean is.

However, this analogy breaks down for relations. Suppose theory postulates $F(x, y) = 0$ - say a demand relation, where x is quantity and y is price. In general one can solve for either variable in terms of the other: $y = g(x)$ or $x = h(y)$ - these both say the same thing. Now introduce uncertainty. One's first thought is to introduce a bivariate distribution $p(x,y)$. But this just gives a single point (Ex, Ey) as central tendency, not a schedule between pairs of x,y values. To get a schedule, postulate a family of distributions over y, indexed by x: $p(y|x)$, with $E(y|x) = g(x)$, the regression line. Reversing the roles of x and y, however, with $p(x|y)$ and $E(x|y) = h(y)$ gives a system which is incompatible with the first ($g(x)$ and $h(y)$ are no longer inverses of each other - the problem of two regressions). The attempt to resolve these problems gives rise to new concepts apparently not present in economic theory - the distinction between exogenous and endogenous variables above all.

These are pervasive problems: They arise for any deterministic relation among variables. With 3 or more variables in a relation the problems get worse, as they do with multiple relations. (How to keep these relations from getting mixed up - the identification problem).

Similar problems arise with time. Most relations postulated in economic theory are equilibrium relation - say $F(x_t, y_t) = 0$. One realizes that these relations do not hold exactly out of equilibrium, but again the equilibrium relations are the central tendency to which the system tends to return when buffeted by dynamic disturbances. And again, the attempt to dynamize economic relations can be done in many ways: Theory radically underdetermines the possible dynamizations, just as it underdetermines the possible stochastizations.

Things might seem to get worse when one considers both time and uncertainty simultaneously. But perhaps the problems introduced can be resolved only in tandem. For one thing, time yields a before-and-after asymmetry among variables, and asymmetry is what is needed to resolve the problem of the two regressions. For another thing, what would uncertainty be without a future to be uncertain about. As someone said, "Time is what prevents everything from happening at once."

Dynamics and Equilibrium

We now bring these considerations to an abstract setting. A (deterministic, discrete time) dynamical system consists of a set of possible states S, and a dynamical law $F:S \to S$, the interpretation being that $s_{t+1} = F(s_t)$ gives the transition from the state at time t to $t+1$.

An equilibrium is a fixed point of F: $F(s) = s$. A dynamical system may have many, one, or no equilibria at all. And even if equilibria exist, the system need not be approaching any.

An invariant set $S' \subseteq S$ is a subset such that $F(S') \subseteq S'$ (once the system is in S' it stays there.)

More generally, a dynamical law may be of the form $F:S\times\ldots\times S \to S$, in which the preceding n states determine the next. However, we may think of $S\times\ldots\times S$ as itself the state space, which reduces things to the preceding. Next, the state space may factor, say $S = S_1\times\ldots\times S_k$, so that a state is represented by a "vector" $(x_1, \ldots, x_k)$. The dynamical law then splits into k components $F_1, \ldots, F_k$, namely, $x_{j,t+1} = F_j(x_{1t}, \ldots, x_{kt})$. This introduces the notion of causality among the components: write $x_i \to x_j$ ("component i has a direct causal influence on j") if F_j depends nontrivially on x_i. For example, given state space (x_1, x_2), suppose $x_{1,t+1} = F_1(x_{1t})$. Then the factor space S_1

develops autonomously, and can be investigated without considering x_2. More generally, given dynamical system (S, F), a generalized factor is an onto mapping g: $S \rightarrow Q$. This factor develops autonomously if, for all s,s' such that $g(s) = g(s')$, it happens that $g(F(s)) = g(F(s'))$. This determines a dynamical law on Q, which can be investigated independently of the full system S.

For example the factor may be in equilibrium while the full system is not.

A parametric system is one with a family of laws F_ω indexed by a parameter space Ω; which can be written as $F:\Omega \times S \rightarrow S$. However this introduces nothing new: Think of the overall state space as being $\Omega \times S$ with components Ω, S developing by $s_{t+1} = F(\omega_t, s_t)$, and $\omega_{t+1} = G(\omega_t, s_t) \equiv \omega_t$, so that the parameter space is an autonomous component that is, in fact, always in equilibrium. (Parameters should always be thought of in this manner. It would save much confusion in statistics and econometrics.)

Consider a factored system $S_1 \times S_2$. Factor x_1 is in temporary equilibrium if there is a subset $S_2' \subseteq S_2$ such that $F_1(x_1, x_2) = x_1$ for all x_2 in S_2', and the current x_2 component is in S_2'. Thus x_1 will stay where it is as long as x_2 remains in S_2', which may not be forever.

Finally, a non-autonomous system is one with a changing law of motion: F is indexed by time t. Trivially this may be autonomized by thinking of the state as $T \times S$, where θ, the "clock" variable, develops by the rule $\theta_{t+1} = G(\theta_t, s_t) \equiv \theta_t + 1$.

Economic Theory and Equilibrium

We now apply some of these ideas to economic theory. Note first that it concerns itself overwhelmingly with equilibrium: consumer equilibrium, equilibrium of the firm, supply-demand equilibrium, the classical stationary state, underemployment equilibrium, etc., up to "general" equilibrium. (Even so-called "disequilibrium" models--Barro-Grossman, et al--actually deal with temporary equilibrium).

And of genuinely dynamic models, one would guess that most of them are concerned with approaching an equilibrium.

Now, if one thinks of economics as approaching a very complicated reality by a series of successive approximations, then this preoccupation with equilibrium is understandable: Equilibrium states are easy to grasp and easy to find and solve for, compared with finding, grasping and solving for the full dynamic models in which they are embedded. The trouble arises when one forgets how poor and incomplete one's approximation is, and "takes the limits of one's vision for the limits of the world." This impedes the path to further progress, and distorts practice. My thesis is that preoccupation of economic theory with equilibrium has distorted econometric practice.

How is equilibrium represented in economic theory? Take the most common case, in which the state space is (formally) R^n, Euclidean n-space, so that an equilibrium point is a vector $(x_1, \ldots, x_n)$, the components referring to stocks, flows, prices, wages, incomes, etc. The equilibrium is usually represented implicitly, by a system of <u>simultaneous algebraic equations</u> $F_i(x_1, \ldots, x_n) = 0$, $i = 1, \ldots, n$ (typically, first-order optimality conditions and market clearing conditions, and perhaps institutional constraints).

Are these equations just a fancy way of representing the equilibrium, or do they have intrinsic significance? Suppose there are k additional "exogenous" variables or parameters $y_1, \ldots, y_k$ among these equations, so that the equations are more correctly represented as $F_i(x_1, \ldots, x_n, y_1, \ldots, y_k) = 0$. Then we are dealing with a family of equilibria indexed by y, which have solutions ("reduced forms") $x = G(y)$. Evidently this schedule says more than just specifying the single solution x^o corresponding to a particular y^o. Do the "structural equations" $F_i(x, y) = 0$ have any "surplus meaning" over the reduced forms $x = G(y)$? Only in the vague sense that they give a clue as to the dynamic system from which the equilibrium arose.

<u>Subjective and Objective Probability</u>

Now to bring in probability. A random variable Q may be thought of as a question, the possible answers to which are $A_1, \ldots, A_n$ (take n finite for simplicity). The distribution of this random variable is given by the probabilities $P_i = \text{Prob}(A_i)$, satisfying $P_i \geq 0$, $\Sigma P_i = 1$, and refers to the cognitive attitude of some mind at some time toward Q. Thus probabilities may vary from one mind to another, and in the same mind over time, and in that sense are "subjective."

If several minds are "similar", and they are all exposed to the same information (perhaps pooled by communication), then they may tend to have similar cognitive attitudes, so that there is rough intersubjective agreement. This is not quite the same as "objective" probability.

Now the world we live in is a dynamical system, but we are not sure which one. The corresponding cognitive attitude is represented by a probability distribution over alternative dynamical systems. This has two aspects: uncertainty about the law of development, and uncertainty about the state the system is in. But the former can be reduced to the latter as follows: Let Ω be a space of parameters indexing possible laws. The augmented state space is then $\Omega \times S$ rather than S alone. The law for this space is known: $s_{t+1} = F(\omega_t, s_t)$, and $\omega_{t+1} = \omega_t$, where $F(\omega_t, \cdot)$ is the law indexed by ω_t. Thus all uncertainty resides in the state of the system.

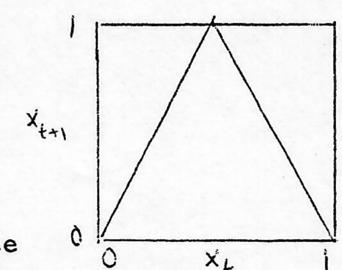
The foregoing gives the "subjective" (or "Bayesian") concept of probability, as reflecting uncertainty. There is, however, another concept which I will (but probably shouldn't) call "objective" probability. (Invariant measure is a better term).

Consider a dynamical system with state space S and law $F: S \to S$. A probability distribution P on S is invariant if it stays the same under the transformation F. Specifically, $P(F^{-1}(A)) = P(A)$ for all (measurable) sets A.

Here are some examples.

(i) Uniform motion around a circle: the uniform distribution is invariant.

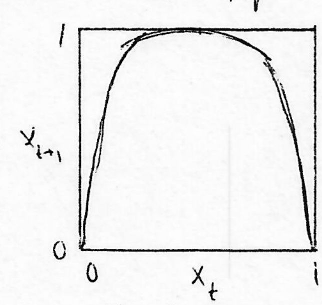
(ii) Non-uniform motion around a circle (these examples are in continuous time). Let $S(\theta)$ be the speed at angle θ. The distribution with density function proportional to $1/S(\theta)$ is invariant. (To see this, note that the probability mass passing any point per unit time = speed $\times$ density, is constant). P(A) = fraction of time spent in region A.



tent map

(iii) The tent map $x \to \min\,[2x, 2-2x]$ on [0,1] The uniform distribution is invariant.

(iv) The logistic $x \to 4x - 4x^2$; the "arc sin" distribution is invariant: $P(x) = (x - x^2)^{-\frac{1}{2}}/\pi$



logistic

For any dynamical system there may be none, one, or many invariant probabilities. (The set of invariant probabilities always is convex, since if P, Q are invariant so is $\lambda P + (1 - \lambda)Q$.) For example, uniform motion on the line has no invariant probabilities. (The uniform distribution on the line _is_ invariant, but this is improper, since infinite). On the other hand, for the trivial system $F(s) \equiv s$ (no motion), _all_ probability distributions are invariant.

Let S, $F: S \to S$ be a dynamical system, and P a probability distribution on S, not necessarily invariant. The dynamical law induces a transformation on P to say P', namely, $P'(A) = P(F^{-1}(A))$, all measurable $A \subseteq S$. The invariant probabilities are simply the equilibrium points of this transformation. One may want to exclude "unstable" equilibria suitably defined, e.g., mass one concentrated at s_0, where s_0 is an ordinary unstable equilibrium point.

To call these invariants "objective" probabilities has some justification: They do _seem_ to have something to do with "frequentist" probability, and with the probabilities that appear in scientific theories and that seem to be "out there" and not inside our heads. Take the ideal case when a dynamical system has a unique proper stable distribution P. Then P has a physical interpretation: P(A) is the fraction of time the system spends in state A.

It is not true, however, that subjective probabilities are approximations of the objective ones toward which they should aim. On the contrary, the "objective" probability P of a system represents a state of maximal ignorance about the system. Any partial observation of a system gives a clue as to the state it is actually in now, and should modify P via Bayes theorem. Objective P is more appropriate as the _prior_ distribution of Bayesian inference.

Independent Identically Distributed Random Variables

Consider a system exhibiting "sensitive dependence on initial conditions" such as the tent map, and which has an additional property which we may call ergodicity (which means, roughly, that any initial distribution which is at all "spread out" transforms to the unique invariant distribution (the uniform distribution for the tent map). The _relaxation time_ is how long it takes for virtually complete "memory loss" to occur (this depends on the precision of the initial distribution).

Suppose one makes repeated "imprecise" observations on this system at intervals exceeding the relaxation time. This provides an "objective" way of generating iid random variables in a deterministic world! And, further, one on which all observers will agree provided their ability to observe precisely is uniformly limited. Example: the tent map, where in state x, the observer can determine only whether $x < \frac{1}{2}$ or $x > \frac{1}{2}$.

Markov Chains and Stationary Processes

Once a source of iid random variables exists, the generalization to Markov chains is easy. Let the state of a dynamical system be given by $s = (x,y)$, where $x_{t+1} = F(x_t, y_t)$ and the y's are iid with y_t not influenced by $x_{t-1}, x_{t-2}, \ldots$. Then the x's are a Markov chain - that is, $P(x_{t+1}|x_t, x_{t-1}, \ldots) = P(x_{t+1}|x_t)$.

This property may be described as path-independence: The conditional distribution of x_t given all preceding x's depends only on the last, so x_{t-1} summarizes all information from the past. Now path independence also characterizes dynamical systems in general: If $s_{t+1} = F(s_t)$, then knowledge of s_{t-1}, s_{t-2} doesn't help, since s_{t+1} is already determined completely by s_t.

The Markov processes stochasticizes dynamical systems in the following sense: In place of the deterministic law F one has a conditional probability distribution over the state space S. A dynamical system is the special case where this distribution is degenerate for all s. Conversely, if P_t is the probability distribution of the variable x_t, then the P_t's themselves form a (deterministic!) dynamical system, the dynamical law $P_{t+1} = F(P_t)$ given by

(*) $$P_{t+1}(A) = P(x_{t+1} \in A) = \int_S P(x_{t+1} \in A|x_t)\, P_t(dx_t)$$

But more than this: The entire preceding discussion on dynamical systems can be "stochasticized" concept by concept. In particular, consider equilibrium. Point s is an equilibrium point if $F(s) = s$, and then s, s, s, ... is a possible history compatible with the dynamical law. For Markov processes, the corresponding concept is an equilibria distribution P, one satisfying $F(P) = P$, where F is given by (*). Then each x_t has the same distribution P. But even more is true: The joint distribution of x_t, x_{t+1} is the same for all t, and so on for $x_t, x_{t+1}, \ldots, x_{t+n}$ for any n. That is, the joint distribution is shift invariant, or stationary. Stationarity is the stochasticization of the equilibrium concept. (Note that a deterministic stationary process must be a constant).

(Technical note: must all stationary processes arise from Markov processes in this way? Well, just as there are two-stage dynamical systems with laws $F: S \times S \to S$, there are two-stage Markov processes where

$$P(x_t | x_0, \ldots, x_{t-1}) = P(x_t | x_{t-2}, x_{t-1})$$

In both cases, the state space may be redefined as $S \times S$. Similarly, there are n-stage processes for any n. For example, an ARIMA (p,d,q) process is Markov with $n = \max(p+d, q)$. All of these may give rise to stationary processes. But the set of all these processes is in a sense "dense" in the space of all processes, and their equilibria are "dense" in the space of all equilibria.)

Econometrics Once Again

Having made this lengthy detour into fundamentals, we return to economic theory and econometrics. The aim of both is to get close to the causal structure of the economic system, theory by developing plausible hypothetical models, econometrics by identifying real world situations with these models at least approximately. Because of our limited abilities to observe, probability should be incorporated organically into these models, even in the realm of pure theory.

Econometric theory stresses the import of "structural relations" over "reduced forms". But what is "structural"? Should it not mean causal structure? This is, after all, the deepest level one can hope to attain, and the one that is relevant for policy purposes. (Another, more subtle, virtue of causal models is their "modularity", their ability to fit together to make more complete causal models, a virtue not shared by simultaneous equation models).

However, looking at what econometric theory calls structural shows a curious reversal. Causal influence goes from past to future. Simultaneous equations cannot do this, almost by definition. Reduced forms, on the contrary, do have this form (which, by the way, doesn't necessarily mean they are capturing causal structure at all well).

What is the source of this strange preoccupation with simultaneous equations? I believe it derives from economists' preoccupation with equilibrium. (Recall general equilibrium theory, usually considered to be the core of economic theory. But even Marshallian theory boils down mostly to the equilibrium of supply and demand. The only real exception is capital theory).

Now equilibrium is not something that should be assumed. As noted above, dynamical systems need not have equilibrium points, and even if they do, the system need not be at or even approaching it. There is one further drawback to equilibrium from the point of view identifying causal structure: It gives no clue as to how the equilibrium arose. Only by disturbing the equilibrium can one obtain this information.

Consider a two-dimensional state space $s = (x,y)$ (say price and quantity) and suppose for simplicity that the dynamics is represented as a simple Markov chain: s_{t+1} has a distribution dependent on s_t. (A common specification is to have the conditional expectation $E(s_{t+1}|s_t) = G(s_t)$, with dispersing independent of s_t). Suppose, for example, that

(†)
$$x_{t+1} = a + bx_t + cy_t + u_t$$
$$y_{t+1} = d + ex_t + fy_t + v_t,$$

where u_t, v_t are jointly iid and independent of s_t, s_{t-1}, ... These are the "true" structural equations. Suppose now, believing that the system is "stationary", one tries to represent this structure. One says, á la Cowles Commission, that it is not legitimate to regress y on x, say, because the variables are "jointly dependent." One represents this as a pair of simultaneous equations. Which pair? To judge from (†), the best pair might be

(††)
$$x_t = a + bx_t + cy_t + u_t$$
$$y_t = d + ex_t + fy_t + v_t$$

This disturbs the situation, but at least the true dynamics can be recovered. Will (††) be the estimated equations? Well, no. For one thing, they are not identified (b,c,e,f are in general all $\neq 0$.) Then one must seek further "identifying restrictions", find instrumental variables for estimation, etc. It is all very confusing.

Note what the assumption of stationarity means here. (The stationary distribution is the invariant measure for the Markov chain, hopefully unique and stable). It represents a condition of <u>maximal ignorance</u> concerning the true state, and is destroyed by any partial observation of that state.

One standard argument for simultaneous equations is that the impact of a change in external parameter can be gauged with them, but not with the reduced forms. Add a parameter θ to the preceding system, so that it reads, say, $E(s_{t+1}|s_t, \theta) = G(s_t, \theta)$. Some or all of a,b,c,d,e,f may depend on θ. The trouble is, apart from getting the equations right, that θ may enter both equations.

What happens if we deal with variables that are not close to the causal structure (as almost always happens, since the variables we observe are aggregated to a greater or longer extent)? It would appear there is even less justification for a "simultaneous equations" treatment here, since the "true" structure is not attainable in any case. What one should do is postulate a causal structure that makes theoretical sense, even if the variables are not directly observable.