ECONOMIC RESEARCH REPORTS

OBSERVABLE IMPLICATIONS OF MODELS WITH MULTIPLE EQUILIBRIA

by

Boyan Jovanovic

R.R. #88-20

June 1988

C. V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

OBSERVABLE IMPLICATIONS OF MODELS

WITH

MULTIPLE EQUILIBRIA

Boyan Jovanovic*

New York University

Revised June 1988

* I would like to thank the C.V. Starr Center for Applied Economics for technical assistance.

Observable Implications of Models

With

Multiple Equilibria

ABSTRACT

Existing theory on structural inference developed by Koopmans and others is inadequate as a basis for inference in models with multiple equilibria, since it assumes that systems of structural equations have unique solutions. This paper presents a general framework in which issues of estimation, identification and so forth can be posed, and it proves three propositions of a general nature concerning the issue of observable implications of such models. The key assumption maintained throughout the paper is that the set of observable implications derives from the set of measurable selections from the equilibrium correspondence. The general conclusion is that unless some special assumptions are made about the range of values that the observable and unobservable variables can take on, models with multiple equilibria tend quickly to become vacuous.

1. Introduction

Models with multiple equilibria are now so common in economics that we need a general framework for statistical inference in such models. Existing theory on structural inference developed by Koopmans and others is inadequate in this regard, since it assumes that once the exogenous data (observable and unobservable) are specified, the endogenous variables can be uniquely determined. It is precisely this assumption that fails when a model has multiple equilibria: A complete specification of the environment does not lead to a unique solution for the endogenous variables. The only serious attempt to deal with non-uniqueness seems to be Wald (1950), who discusses hypothesis-testing for the case where the number of endogenous variables exceeds the number of equations.

2. General formulation.

Let U and Y be vector spaces, each endowed with a metric. The variables under consideration are divided into two groups:

(a) latent variables, $u \in U$. The vector u is not observed by the analyst, but perhaps some or even all of its components are observed by the economic actors. The symbol M_U denotes the set of all Borel probability measures on U.

(b) observed variables, $y \in Y$. The analyst can observe y, or at least some of its components. The symbol M_Y denotes the set of all Borel probability measures on Y.

The vector u is exogenous, the vector y endogenous. Observable exogenous variables are ignored; their presence is easily dealt with by trivial additions to the notation (see remark B of section 5). Throughout, the topology on M_Y and M_U is that of weak convergence.

Structure. In the spirit of Koopmans and Reiersol (1950, p. 168), define a structure as a pair $s = (\nu, \phi)$, where $\nu \in M_U$ and where ϕ is a relation (i.e., a subset of $Y \times U$) such that $(y,u) \in \phi$. The following three examples illustrate how ϕ may arise.

1. Optimization problems: For a given $f: Y \times U \to R$, let $(y,u) \in \phi$ if $y = \underset{y}{\operatorname{argmax}} f(\tilde{y},u)$. Here the problem is not multiple equilibria, but multiple maximizers. Such multiple maximizers are unlikely, however. For instance, f might be a polynomial (of finite order) in $\tilde{y} \in R^1$, with the coefficients being strictly monotone functions of u. If ν is atomless, multiple maximizers occur on a set of u's of ν-measure zero.

2. Linear simultaneous equation-systems. Let $Y = R^m$, $U = R^n$, and let A be an $(n \times m)$ matrix. Then let $(y,u) \in \phi$ if $Ay = u$. If $m > n$, infinitely may solutions for y exist, for each u. Even if $m = n$, A may be singular, although this is generically not the case.

3. Nonlinear simultaneous equation-systems. Let $Y = R^m$, $U = R^n$, and let $f: R^{m+n} \to R^m$ be given. Then let $(y,u) \in \phi$ if $f(y,u) = 0$. The Arrow-Debreu model with m+1 commodities results in a system of this sort (where u may represent (uninsurable) endowments, f the vector of excess demands, and y the vector of relative prices). Alternatively, f may be the vector of first-order conditions in an m-player game, with y a vector

of actions, u a vector of shocks to payoff functions, with u being common knowledge amongst the players but not observed by the analyst. Competitive equilibria and Nash-equilibria are often not unique.

Let Φ be the set of all relations on (say the collection of all Borel subsets of) $Y \times U$ and let $S \subset \Phi \times M_U$ be the set of a priori admissible structures. Multiple Equilibria arise when a complete specification of the environment (i.e., of the exogenous data) does not lead to a unique determination of the endogenous variables. Let

$$\Psi_\phi(u) = \{y \in Y \mid (y,u) \in \phi\}$$

be the reduced-form correspondence, or the equilibrium correspondence. If equilibrium is always unique, $\Psi_\phi(\cdot)$ is a function. It is assumed that $\Psi_\phi(u)$ is non-empty on U.

3. Restrictions on observables:

What does structure s imply about the distribution of observables? Matters are simple if we allow, for each fixed u, that y be chosen from the set $\Psi_\phi(u)$ with the (possible) aid of an extrinsic random selection-device. Let $M_{Y\times U}$ be the set of Borel measures on $Y \times U$. For a $\tau \in M_{Y\times U}$, let τ_Y and τ_U denote the marginals on Y and U respectively. Let

$$T_0(s) = \{\tau \in M_{Y\times U} \mid \tau(\phi) = 1, \text{ and } \tau_U = \nu\}.$$

This is the set of measures that assign unit-measure to ϕ and whose marginal on U is ν. That <u>every</u> τ in $T_0(s)$ is the result of using <u>some</u> (possibly random) equilibrium-selection-device conditional on u, follows because probability measures on Y, indexed by u, can be constructed in this case (Billingsley 1979, p. 381). In this sense, $T_0(s)$ is the collection of all joint distributions of y and u that are consistent with s. The set of all <u>marginals</u> on Y is

$$T(s) = \{\mu \in M_Y | \mu = \tau_Y \text{ and } \tau \in T_0(s)\}.$$

Since it is the y's that we see and not the u's, it is T that we are primarily interested in. Both T(s) and $T_0(s)$ are <u>convex</u>.

The purist might, however, wish to disallow extrinsic randomness in the choice of equilibria. In that case, he would be left with only those distributions that can be generated by the measurable selections from Ψ_ϕ. That is, if $A_\phi = \{\text{measurable } \psi: U \to Y | \psi(u) \in \Psi_\phi(u), \forall u \in U\}$, the set of distributions of y generated by elements of A_ϕ is[1]

$$\Gamma(s) = \{\mu \in M_Y | \mu = \nu \cdot \psi^{-1}, \ \psi \in A_\phi\}.$$

Clearly, $\Gamma(s) \subseteq T(s)$ for each s. But how much larger is T ? The first proposition states that when ν is atomless, Γ and T are, for all practical purposes, identical:

[1] The meaning of $\nu \cdot \psi^{-1}$ is that for any Borel subset B of Y, $\mu(B) = \nu(\{\psi^{-1}(B)\})$.

Proposition 1: Let Y be compact, and ν atomless. Then $T(s)$ is the closure of $\Gamma(s)$.

Proof: See the appendix. Some remarks on this result are:

1. While T and Γ are virtually indistinguishable, T is easier to characterize. Moreover, when Y is compact, so is T,[2] and so it is preferable to Γ on grounds of expediency: open set problems in the choice of a "best" distributions do not arise if T is used as the menu. For instance, the problem of maximizing a likelihood function by choosing from among distributions contained in $T(s)$, and then by choosing s, will, by Weierstrass's theorem, have a solution (for its first stage, at least) when the likelihood function is continuous on $T(s)$.

2. Atomlessness of ν is sufficient, not necessary. What is needed is that one can "distribute" the u's among the solutions y and for this, enough u's are needed. The proposition would fail if, for instance, there were 3 possible equilibrium y's and only two u's. Then $\Gamma(s)$ could not contain a distribution that assigns positive measure to each of the y's, and yet such a distribution would be in $T(s)$. In econometric applications, however, ν is usually assumed to be atomless.

3. Even when ν is atomless, generally, $\Gamma \subset T$, because no measurable function can oscillate fast enough to attain certain distributions in T. The discussion of eq. (1) in Feldman and Gilles (1985) is relevant here.

4. If equilibrium is unique for all u so that Ψ_ϕ is a function, then $T(s)$ and $\Gamma(s)$ are singletons and $T = \Gamma$ for all s.

---

[2] One virtue of the topology of weak convergence in this context is that the set of probability measures over a compact set is itself compact.

5. If $\psi_i \in A_\phi$ $(i=1,2)$, and if I_B is the indicator of the set B, the regime $\psi_B(u) = I_B(u)\psi_1(u) + [1-I_B(u)]\psi_2(u)$ is also in A_ϕ, as long as B is a Borel set in U. Thus A_ϕ must be closed with respect to operations of this sort.[3]

We now return to the question of the size of T(s). Clearly, multiplicity is not a problem if it occurs on a set of u's of ν-measure zero, since T(s) is then single-valued. More generally, if $U^*(s)$ is the subset of U on which more than one equilibrium exists, and if $\beta(Y)$ is the collection of Borel subsets of Y, we have

<u>Proposition 2</u>: $$\sup_{\mu,\mu'\in T(s)} \{ \sup_{B\in\beta(Y)} |\mu(B) - \mu'(B)| \} \le \nu(U^*(s)).$$

<u>Proof</u>: Fix B, and write U^* for $U^*(s)$. From proposition 1, there are ψ_1 and ψ_2 in A_ϕ such that $|\mu(B) - \mu'(B)| < |\nu(\psi_1^{-1}(B)) - \nu(\psi_2^{-1}(B))| + \epsilon$, for any $\epsilon > 0$. But $\nu(\psi_i^{-1}(B)) = \nu(\psi_i^{-1} \cap U^*) + (\psi_i^{-1}(B)\cap(U^*)^c)$. Since ψ_1 and ψ_2 coincide on $(U^*)^c$, $\nu(\psi_1^{-1}(B) \cap (U^*)^c) = \nu(\psi_2^{-1}(B) \cap (U^*)^c)$ so that $|\nu(\psi_1^{-1}(B)) - \nu(\psi_2^{-1}(B))| \le |\nu(\psi_1^{-1}(B) \cap U^*) - \nu(\psi_2^{-1}(B) \cap U^*)| \le \nu(U^* \cap \Psi_\phi^{-1}(B))$.

---

3 These operations resemble those in the regression-switching literature. This is no accident since the ψ_i can be thought of as regimes, and one can contemplate various selection-rules over regimes (see Quandt 1972). Proposition 1 states that the class of such selection rules is not in general large enough to capture all distributions implied by the structure. Systematic selections are also used in the disequilibrium literature (e.g. Goldfeld and Quandt 1975) where the choice of the demand or supply regime is based on whichever is smaller. Disequilibrium models, however, typically involve unique equilibria in the sense that once price is specified exogenously, the system of equations admits a unique solution.

Since the second sentence of the proof is true for all positive ϵ, the argument is complete. Q.E.D.

Therefore, if $\nu(U^*(s))$ is small, multiplicity can be safely be ignored, at least in the sense that an arbitrary selection of a μ from $T(s)$ cannot be far off the mark. Let F_μ and $F_{\mu'}$ be the CDF's implied by μ and μ' on the vector space Y. Then proposition 2 implies that

$$\sup_{\mu,\mu'\in T(s)} \{ \sup_{y\in Y} |F_\mu(y) - F_{\mu'}(y)| \} \le \nu(U^*(s) \cap \Psi_\phi^{-1}(-\infty,y])) \le \nu(U^*(s)).$$

This is a useful inequality since it is stated in terms of CDF's, just as the law of large numbers is often expressed in a similar way, as for instance in its Glivenko-Cantelli version (Billingsley 1979, p. 232).

<u>The dimensionality of T and Γ</u>. Can T, or at least Γ be represented as a finite-dimensional family of distributions? Generally, the answer will be no, although there are two important exceptions. First, if Y is finite, then T and Γ have dimensions at most one-less-than the number of points in Y (the example in the next section illustrate this point). Second, when U is finite, A_ϕ is simply a set of vectors, each vector having as many components as the number of points in U, and as a result Γ(s) (but not necessarily T(s)!) is finite-dimensional.

Under fairly weak conditions, however, T and Γ will be infinite dimensional:

Proposition 3: Let B be an open set in $U \subseteq R^n$, and let ψ and ψ' be two selections from A_ϕ. Suppose that

(a) u has strictly positive density on B;

(b) $\psi(B) \cap \psi'(B) = \emptyset$;

(c) ψ is continuous and strictly monotone on B.

Then $\Gamma(s)$ and $T(s)$ are infinite-dimensional.

Proof: B contains a collections of sets $(B_j)_{j=1}^{\infty}$ for which $\bigcap_{j=1}^{\infty} \psi(B_j) = \emptyset$ (by (c)), for which $\nu(B_j) > 0$ all j (by (b)), and for which $\psi(B_j)$ is open in Y, all j (by (c)). Now let $(C_{j\theta})_{\theta\in(0,1)}$ be a family of subsets of B_j such that $\theta' > \theta \Rightarrow C_{j\theta} \subset C_{j\theta'}$ and such that $\nu(C_{j\theta}) < \nu(C_{j\theta'})$. Then in the spirit of remark 5 following proposition 1, let

$$\psi_j(u,\theta) \equiv I_{C_{j\theta}}(u)\psi(u) + [1 - I_{C_{j\theta}}(u)]\psi'(u).$$

As θ increases from zero to one, ψ_j shifts continuously more and more mass onto $\psi(B_j)$ and away from $\psi'(B_j)$. This yields a one-parameter family (indexed by $\theta_j \in (0,1)$). Since they assign different measures to an open set, each of these distributions is different under the weak-convergence topology. But j is arbitrary. All of the distributions obtained by varying θ_j, j=1,2,... are different because $\cap\psi(B_j) = \emptyset$ and because $\psi(B_j) \cap \psi'(B_j) = \emptyset$ by (b). Thus Γ has dimensionality at least $(0,1)^\infty$, and since $\Gamma \subseteq T$, the claim follows. Q.E.D.

Identification. Identification holds if the structure can be recovered uniquely from the distribution of the observables, μ, and if this true for all μ in the range of T. A necessary and sufficient condition for identification (with or without nonuniqueness) is

$$(1) \qquad T(s) \cap T(s') = \emptyset,$$

for all s and s' in S for which $s \neq s'$. If (1) holds, T^{-1} is a function. There is no necessary relation between nonuniqueness and lack of identifiability -- neither implies the other. Nevertheless, non-uniqueness makes it "easier" for (1) to fail, because it makes $T(s)$ larger.

4. Example: A two-person game. Let the payoff functions be

$$\Pi_1(x_1,x_2,u_1) = \begin{cases} \phi x_2 - u_1 & \text{if } x_1 = 1 \\ 0 & \text{if } x_1 = 0 \end{cases} \quad \text{and} \quad \Pi_2(x_1,x_2,u_2) = \begin{cases} \phi x_1 - u_2 & \text{if } x_2 = 1 \\ 0 & \text{if } x_2 = 0 \end{cases}.$$

Here $x_i \in \{0,1\}$ is player i's action, and $u_i \in [0,1]$ is his type. The players' types are common knowledge among the players; the analyst, however, knows only that they are independently and uniformly distributed on $[0,1]^2$. Thus ν is the uniform distribution. The remaining structural characteristic is just ϕ; the analyst knows a priori that $\phi \in (0,1]$.

One pure-strategy Nash equilibrium is $x_1 = x_2 = 0$ for all $(u_1, u_2) \in U$. A second is $x_1 = x_2 = 1$ for $(u_1,u_2) \in [0,\phi]^2$, and zero otherwise. There are no mixed strategy equilibria except on a subset of U of ν-measure zero, nor are there equilibria where $x_i \neq x_j$ for any $u \in U$. The reduced forms implied by the two equilibria discussed above are

$$\psi_1(u) = 0 \quad \text{all } u \in U, \qquad \psi_2(u) = \begin{cases} 1 & \text{for all } u \in [o,\phi]^2 \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mu_1 = \nu \cdot \psi_1^{-1} = (1,0)$, and $\mu_2 = \nu \cdot \psi_2^{-1} = (1-\phi^2, \phi^2)$. The first coordinate of each of the μ_i measures the probability that $y = x_1 = x_2 = 1$, the second that $y = 0$. Then $T(s) = \{(p_1,p_2) \in R^2 | p_1 \in [1-\phi^2, 1]$, and $p_1+p_2=1\}$, which is convex. It is immediate that $T(s) = \Gamma(s)$ so that Proposition 1 holds trivially. The distributions in $T(s)$ can be represented by a one-parameter family -- the strict monotonicity condition of Proposition 3 fails here. Proposition 2 becomes relevant as $\phi \to 0$. The identification condition (1) fails here: If s' and s differ in that $\phi' > \phi$ (while ν remains the uniform distribution on the unit-square), then $T(s) \subset T(s')$. Nevertheless, certain subsets of the parameter space can be identified: If μ can be consistently estimated and if $\hat{p}_1$ is the estimate of its first coordinate, then we can infer that $\phi \in [\sqrt{\hat{p}_1}, 1]$.

5. Extensions

A. Structures with multiple equilibria are composite hypotheses. Let $\Omega = \{(\mu,s) \in M_Y \times S | \mu \in T(s)\}$ be the graph of T. If, in addition to the structural parameters, we also think of equilibria as parameters, then Ω is the appropriate parameter-space. A particular hypothesis $H_0 : s = s_0$ is "composite"[4] in the sense that the "nuisance" parameter μ is left unspecified by the hypothesis tested except for the constraint that $\mu \in T(s_0)$. Whether a "fixed effect" or a "random effect" treatment of these nuisance parameters is called for will be transparent from the way that data is gathered, just as it is in other contexts (see Chamberlain 1984). Bayesian methods in the present context call for prior beliefs over Ω, and not just over S as is true when uniqueness holds). Note also that the more vacuous the structure s_0 (i.e., the larger is $T(s_0)$), the more likely it is to contain a μ that fits the facts.

B. Observable exogenous variables. Formally, shifts in such variables will affect the shape of ϕ. But they may also affect the way equilibria are selected. To briefly indicate some of the possibilities, consider the example of section 4, and assume that $\phi = 2/3$, and that u_1 and u_2 are now observable by the analyst. Moreover, let u_1 and u_2 be perfectly correlated (this makes no difference to the calculation of the equilibria). Write $u = u_1 = u_2$. Thus when $u > 2/3$, $y = 0$. Then assume that $y = 1$ for $u \in [1/2,2/3]$ and $y = 0$ otherwise. Thus we have rigged up a selection mechanism in such a way that the correlation between y and u is positive, even though economic intuition says that it should be negative since u is the (marginal) cost incurred as x increases from zero to one. Of course, there are other selection rules which lead to a negative correlation between y and u. Note that this negative conclusion was reached in an example in which the endogenous variable can take on just two values, and that the situation can only get worse in general.

---

[4] See Kendall and Stuart (1961), ch. 23.

C. <u>Local uniqueness does not imply even locally-continuous selections</u>. Consider the following two statements: (A) If equilibrium is locally unique, when the exogenous data change slightly, we can stay in the neighborhood of the old solution, rather than "jump" to another solution far away, even when other such distant solutions exist. (B) Selections from Ψ_ϕ should be continuous. When a continuous selection does not exist, we should choose ones that are the "most" continuous in some sense. In short, avoid selections that "jump around."

Statement (A) is based on the following logic: (i) "If u <u>changes</u> from u_0 to $u_0+\epsilon$, then y should not <u>change</u> by much". But to endorse (B) is to assert that (ii) "If, rather than at u_0, u had been at $u_0+\epsilon$ <u>instead</u>, then rather than at y_0, y would have been at $y_0+\delta$ <u>instead</u>". But (A) does not imply (B). Whatever notions of stickiness, inertia, focal points, etc. are invoked to lend creedence to (A), they cannot justify (B). That is why the present paper opts only for measurable selections.

Appendix: Proof of proposition 2

Since the range of Ψ_0 lies in a compact set, it has an ϵ-cover with a finite number of elements. Let $y_1, \ldots, y_n$ be the centers of the $n(\epsilon)$ neighborhoods of the cover. For each u in U let $\Psi_n(u) = \Psi(u) \cap \{y_1 \ldots, y_n\}$, (where we drop the subscript ϕ from Ψ for the remainder of the proof). The range of Ψ_n is a finite set. The Hausdorff distance between Ψ_n and Ψ is, for each μ, at most ϵ. It follows that, as $\epsilon \to 0$ (and hence $n(\epsilon) \to \infty$), any measure on the graph of Ψ can be weakly approached arbitrarily closely by a measure on the graph of $\Psi_{n(\epsilon)}$. Therefore the proof can be completed by showing that the set of measures on the graph of Ψ_n with a marginal on U given by ν actually coincides, for each n, with the set of measures one can obtain by using the measurable selections from Ψ_n, because the latter are by construction, for each n, contained in A_ϕ.

For $k = 1, \ldots, 2^{n(\epsilon)}$, let (B_k) be the subsets of $\{y_1, \ldots, y_n\}$. Let $C_k = \{u \in U | \Psi_n(u) = B_k\}$. Then the (C_k) form a partition of U. Let $h(y|u)$ be the given probability that y is selected (from amongst the $\{y_1, \ldots, y_n\}$) conditional on u, so that $\mu(y_j) = \int_U h(y_j|u)d\nu(u)$ is given. We shall now construct a measurable $\psi: U \to Y$ that induces this measure. Partition C_k into (C_{kj}) such that $\nu(C_{kj}) = \int_{C_k} h(y_j|u)d\nu(u)$, (all j such that $y_j \in B_k$). This can be done because ν is atomless. (The subsets C_{kj} are not uniquely defined, but this is immaterial). Now (C_{kj}) is also a partition of U. Then define $Y: U \to Y$ by $\psi(u) = y_j$ for $u \in C_{kj}$. Then $\nu(\{\psi^{-1}(y_j)\}) = \int_U h(y_j|u)d\nu(u)$. Therefore ψ is the measurable selection from $\Psi_{n(\epsilon)}$ for which $u = \nu \cdot \psi^{-1}$. Q.E.D.

References

Billingsley, P. Probability and Measure, New York: John Wiley and Sons, 1979.

Chamberlain, G. "Panel Data," Handbook of Econometrics II, ed. by Z. Griliches and M. Intrilligator, North Holland: New York 1984.

Feldman, M., and C. Gilles. "An Expository Note on Individual Risk without Aggregate Uncertainty," Journal of Economic theory, 35 (1985): 26-32.

Goldfeld, S.M., and R.E. Quandt. "Estimation in a Disequilibrium Model and the Value of Information, Journal of Econometrics 3(1975): 325-48.

Kendall, M.G., and A. Stuart. The Advanced Theory of Statistics, 2; Griffin, London, 1961.

Kiefer, J., and J. Wolfowitz. "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters." Annals of Mathematical Statistics 27 (1956): 887-906.

Koopmans, T.C., and O. Reiersol. "The Identification of Structural Characteristics", Annals of Mathematical Statistics 21 (1950): 165-181.

Quandt, R.E. "A New Approach to Estimating Switching Regressions," Journal of the American Statistical Association 67, no. 338 (June 1972): 306-310.

Wald, A. "Remarks on the Estimation of Unknown Parameters in Incomplete Systems of Equations," in Statistical Inference in Dynamic Economic Models, T.C. Koopmans, ed., John Wiley. 1950.