ECONOMIC RESEARCH REPORTS

INFERENCE IN MODELS WITH MULTIPLE EQUILIBRIA

by

Boyan Jovanovic

R.R. #86-17 September 1986

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

INFERENCE IN MODELS
WITH
MULTIPLE EQUILIBRIA

Boyan Jovanovic*

New York University

June 1986

* Original draft November 1985. Technical assistance from the C.V. Starr Center of Applied Economics is gratefully acknowledged as is computing assistance by Francesco Goletti.

Introduction

Models with multiple equilibria have become common in economics, and are likely to remain so, at least for the forseeable future. Consequently there is a need for a general framework within which statistical inference can be drawn in the context of such models, and this paper is an attempt in this direction.

Existing theory of structural inference in economics is based on the post-war work of Koopmans and his co-workers. This work does not form an adequate basis for inference in models with multiple equilibria, since it assumes that when the exogenous data (both 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.

An early application by Wald (1950) of Neyman-Pearson hypothesis testing to incomplete equation systems appears to have been the only serious attempt to deal with non-uniqueness. Wald considers the case where the number of endogenous variables exceeds the number of equations and concentrates on how hypotheses might be tested in that situation. His treatment is brief, however, and its generalization to other context is unclear. Moreover, the source of non-uniqueness in models in which agents optimize is never that the number of restrictions falls short of the number of endogenous variables: rather, it is usually found in the nonlinearity of the equation system, or in a particular failure of equilibria to be regular.[1]

---

[1] Of course linear systems too can produce non-uniqueness even when the number of equations is the same as the number of endogenous variables. In the standard linear simultaneous equations model $Ay + Bx = u$ (where y, x and u are vectors of endogenous, exogenous and latent variables), if the square matrix A is singular, infintely may solutions for y exist.

Plan and Summary. Section two introduces the terminology -- basically the definition of "structure" of Koopmans and Reiersol (1950), but then drops the assumption that the endogenous variables can be uniquely recovered from the exogenous variables. Consequently the joint distribution of the observables is not uniquely determined by the structure. In this sense, the predictive content of a structure is reduced.

The extent to which non-uniqueness reduces the refutability of a structure is examined in section 3, which contains two rather negative results: the set of distributions consistent with a given structure must be (a) convex, and (b) closed under the formation of a variety of non-linear operations on the endogenous random variables. Consequently, if a structure gives rise to two different distributions of observables, then it also gives rise to a whole spectrum of distributions in between, and indeed to others that are not convex combinations. These results are damaging for the refutability of models with multiple equilibria, and we seek to remedy this later on in section 7.

Following a definition of identification in section 4, sections 5 and 6 present two examples, the second of which shows that multiple equilibrium is not always a curse and that a model's parameters can sometimes be identified in spite of the presence of great many equilibria, although this is likely to be an exception.

Section 7 introduces a fairly general class of models in which the problems that multiplicity causes for inference are not as severe. These are models of competition between negligible, anonymous agents. They generally have fewer equilibria and the strategic interactions among agents are less complicated; examples of such models follow in the subsequent two sections.

The best known example is of course that of Walras ("competitive") equilibrium.

Section 10 and beyond deal specifically with estimation and other types of inference. Section 10 presents consistency results for two nonparametric estimators for models in the anonymous-negligible class (of section 7). One estimator is applicable in cases where the equilibrium can be treated as a fixed effect, the other in cases when a random effect treatment is warranted (as with panel data -- see Chamberlin, 1984 -- it is usually clear whether a fixed or random effect specification ought to be used).

Section 11 briefly discusses maximum likehood methods, both the fixed and random effect formulation. No consistency theorems are given since the standard arguments of Wald (1949) or Kiefer and Wolfowitz (1956) apply once the parameter space is correctly specified.

Section 12 discusses the Neyman-Pearson theory of testing as it applies in general to models with multiplicity of equilibria. The main twist here is that the testing of structures with multiple equilibria is not always equivalent to the testing of composite hypotheses. When this is not so, a simple extension is proposed.

Section 13 contains some remarks on Bayesian reference which requires that prior probabilities be put not only over structures but over equilibria as well. Some criteria for choosing the prior are briefly mentioned.

The final section deals with some specific issues relevant to the application of ML and other estimators that essentially minimize the distance between theory and data. Such estimators are biased (sometimes even asymptotically so) towards vacuous structures simply because such structures can fit any configuration of facts. The larger the number and diversity of equilibria, the more vacuous a structure is. We demonstrate a case of asymptotic bias of ML by example. We then look at the finite sample bias for a different example in which ML is consistent, and find that under reasonable conditions the bias disappears fairly quickly.

2. General formulation.

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 the vector space U.

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

The vector u is exogenous, the vector y endogenous. Observable exogenous variables are ignored in this paper; their presence is easily dealt with by trivial additions to the notation.

Structure. Koopmans and Reiersol (1950, p. 168) define a structure as a pair $s = (\nu, \phi)$, where

$$\nu \in M_U \tag{2.1}$$

and where ϕ is a set of simultaneously valid relationships

$$\phi(y, u) = 0. \tag{2.2}$$

Koopmans and Reiersol then assume that the equation system (2.2) admits a unique solution for y in terms of u, except possibly for a set of u's of

ν - measure zero. Each structure s then implies a unique probability distribution of y.

Two remarks may be made here. First, Koopmans and Reiersol introduced (2.2) as an illustration of the familiar case in which the set of restrictions imposed by an economic structure consists entirely of a system of equalities (e.g. a system of accounting identities and necessary conditions arising when differentiable functions are maximized at interior points of feasible sets). In general, the system (2.2) would have to be replaced by a system of inequalities (e.g., in discrete choice models, or models in which agents' objective functions are not differentiable with respect to their decision variables). We shall continue, however, to refer to (2.2) as <u>the</u> entire set of restrictions implied by the structure with the understanding that these restrictions may be of a more general nature.

Second, ϕ is assumed to be restricted to the set Φ of sets of relations that admit at least one solution for y for each u in U, that is, at least one equilibrium exists for each u; when this is not true, one can simply restrict the support of ν to that subset of U on which at least one equilibrium exists.

<u>Multiple Equilibria</u>. 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

$$\Psi_{\phi}(u) = \{y \in Y \mid \phi(y,u) = 0\}. \tag{2.3}$$

This is the reduced form correspondence, or the equilibrium correspondence. If equilibrium is always unique, $\Psi_{\phi}(u)$ is a singleton for each u, and a reduced form (function) exists.

3. Restrictions on observables.

Let

$$A_\phi = \{\text{measurable } \psi\colon U \to Y \mid \psi(u) \in \overline{\Psi}_\phi(u),\ \forall\ u \in U\}$$

be the set of measurable selections from the correspondence Ψ_ϕ. This is the set of all measurable functions ψ satisfying $\phi[\psi(u),u] = 0$ everywhere.

Let $S = (M \times \Phi)$ be the set of admissable structures (where $M \subseteq M_U$ and where Φ is a set of structural relations). Its generic element is $s = (\nu,\phi)$. Let Γ be the correspondence of S into M_Y defined by

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

The set of $\Gamma(s)$ is the collection of all distributions of the observables that are consistent with the structure s.

Some remarks are now given concerning this definition..

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

2. The definition of Γ would be vacuous if A_ϕ were empty, but this is generally not the case: Ψ is usually closed-valued and u.h.c. (see e.g., Hildenbrand 1974 p. 30 and Green 1984, thm. 3) in which case the non-emptiness of A_ϕ follows from a lemma in Hildenbrand (1974, p. 55).

3. The definition of Γ excludes distributions that might be somehow derived from non-measurable ψ solving (2.2). When ϕ is not continuous, such non-measurable reduced forms may exists. But since we are not able to assign a μ in M_Y to such a ψ (because for at least some Borel sets in Y, their

inverse image is not a Borel set), we are forced to ignore non-measurable reduced forms. There is a fundamental sense, however, in which the restriction (3.1) is too stringent. A non-measurable ψ can arise when agents' payoffs are not continuous because then their decision correspondences are not u.h.c. and may not have measurable selections. One could of course increase the number of measurable solutions by coarsening the topology on Y, (and hence the Borel sigma algebra) but this would entail the cost of reducing the number of probability statements that can be made. In the coarsest case, with the sigma algebra consisting of Y and the empty set, every function whose range is in Y is measurable but no interesting probability statements are possible.

4. One might argue the converse: the restriction (3.1) is too loose, because we should only admit continuous ψ's -- a much smaller class than the set of measurable solutions to (2.2). The problem with this is that a u.h.c. correspondence (Ψ) does not in general have a continuous selection, globally on U. (See Allen (1985) for a recent discussion in the context of Walras equilibria). And when continuous selections do exist, it is hard to think of good reasons for arbitrarily excluding all other selections. On the other hand, the exclusion of non-measurable ψ's is here made on grounds of necessity alone.

Two theorems characterizing Γ. Since $\Psi_\phi(u)$ is not in general a convex set (for fixed u), the set A_ϕ is not convex in general. First we show that if A_ϕ has more than one element, then it is quite large. Let $B(U)$ be the Borel subsets of U. For $B \in B(U)$, let

$$I_B(u) = \begin{cases} 1 & \text{if } u \in B, \\ 0 & \text{otherwise} \end{cases}$$

be the indicator function of the set B. Then if $\psi_1, \psi_2 \in A_\phi$, let

$$\psi_B(u) = I_B(u)\psi_1(u) + [1-I_B(u)]\psi_2(u). \tag{3.2}$$

If $B \in B(U)$, ψ_B is measurable, and hence $\psi_B \in A_\phi$. Thus for any two reduced forms ψ_1 and ψ_2, we obtain in addition an entire family $(\psi_B)_{B\in B(U)}$ of such functions, each of which is also in A_ϕ. Thus A_ϕ may include a large number of random variables (i.e., measurable ψ's), but one still needs to determine how different their distributions are, and this is a question about Γ. The answer is that although A_ϕ is not in general a convex set, Γ is convex-valued for each s:

<u>Theorem 3.1</u>: Let $U \subseteq R^m$, and let ν be absolutely continuous w.r.t. Lebesgue measure on R^m. Let $\mu_1, \mu_2 \in \Gamma(s)$. Let $\mu_\alpha = \alpha\mu_1 + (1-\alpha)\mu_2$. Then $\mu_\alpha \in \Gamma(s)$ for all $\alpha \in [0,1]$.

<u>Proof</u>: Let ψ_i: $U \to Y$ be such that $\mu_i = \nu \cdot \psi_i^{-1}$, $i=1,2$. The existence of measurable ψ_i is immediate from (3.1). Let h: $U \to \{0,1\}$ be measurable, and for each u, set $\psi_h(u) = h(u)\psi_1(u) + [1-h(u)]\psi_2(u)$. For any Borel subset B of y,

$$\psi_h^{-1}(B) = [h^{-1}(1) \cap \psi_1^{-1}(B)] \cup [h^{-1}(0) \cap \psi_2^{-1}(B)]$$

which is a Borel subset of U since it is obtained by a finite number of operations on Borel sets. Hence ψ_h is measurable. Since ψ_1 and ψ_2 are reduced forms, so is ψ_h. Since $h^{-1}(1) \cap h^{-1}(0) = \emptyset$,

$$\nu(\psi_h^{-1}(B)) = \nu(\{h^{-1}(1) \cap \psi_1^{-1}(B)\}) + \nu(\{h^{-1}(0) \cap \psi_2^{-1}(B)\}).$$

Since ν is absolutely continuous w.r.t. Lebesque measure, one can construct on the probability space $(U, B(U), \nu)$ a r.v. $h_\alpha: U \to \{0,1\}$ such that (a) $\nu(\{h^{-1}(1)\}) = \alpha$, where $\alpha \in [0,1]$, and (b) h_α is independent of ψ_1 and ψ_2. Then

$$\nu(\psi_{h_\alpha}^{-1}(B)) = \alpha\nu(\psi_1^{-1}(B)) + (1-\alpha)\nu(\psi_2^{-1}(B))$$
$$= \alpha\mu_1(B) + (1-\alpha)\mu_2(B).$$

Since B and α are arbitrary, setting $\mu_\alpha = \nu \cdot \psi_{h_\alpha}^{-1}$ yields the assertion of the theorem. Q.E.D.

This result is damaging to the refutability of structures yielding multiple equilibria. If μ_1 and μ_2 are different, then these two implications of s are augumented by all convex combinations of μ_1 and μ_2. The next result is somewhat surprising; it says that if μ_1 and μ_2 are in $\Gamma(s)$, there will, in addition to all convex combinations of μ_1 and μ_2, exist other distributions in $\Gamma(s)$.

Theorem 3.2: Let $\mu_i = \nu \cdot \psi_i^{-1} \in \Gamma(s)$, $i=1,2$, where ψ_i are such that $\nu(\{u|\ \psi_1(u) > \psi_2(u)\}) > 0$ and $\nu(\{u|\psi_1(u) < \psi_2(u)\}) > 0$. Then there exists μ and μ' in $\Gamma(s)$ such that neither μ nor μ' is a convex combination of μ_1 and μ_2.

Proof: Let $\psi_+ = \max(\psi_1, \psi_2)$ and $\psi_- = \min(\psi_1, \psi_2)$, and let $\mu = \mu \cdot \psi_+^{-1}$ and $\mu' = \nu \cdot \psi_-^{-1}$. Then $\mu, \mu' \in \Gamma(s)$. Moreover, $\int y d\mu > \max(\int y d\mu_1, \int y d\mu_2)$ so that the first moment of μ cannot be a convex combination of the first moments of μ_1 and μ_2, and so the assertion of the theorem is true for μ. Similarly $\int y d\mu' < \min(\int y d\mu_1, \int y d\mu_2)$. Q.E.D.

Both theorems are of practical relevance. The construction of say ψ_+ in the proof of theorem (3.2) involves a nonlinear operation on the random variables ψ_1 and ψ_2, namely $\max(\psi_1,\psi_2)$. This construction is a special case of (3.2) with $B = \{u \in U | \psi_1(u) > (u)\}$. Other non-linear operations can be effected with other B's. As a result, a narrow parametric family (such as the normal) cannot represent all $\mu \in \Gamma(s)$ when $\Gamma(s)$ is not a singleton, because such a family is not closed with respect to the variety of nonlinear operations admitted by (3.2).

The linear and nonlinear operations on random variables based on (3.2) are similar to certain operations taken in the regression-switching literature. This is no accident since the ψ_i in (3.2) can be thought of as regimes and $I_B(\cdot)$ as the selection-rule over regimes. Theorem 3.2 then states that the class of mixtures over regimes (see Quandt 1972) is not in general large enough to capture all distributions implied by the structure. The reason is that systematic selections also are possible such as ψ_+ or ψ_-, the latter being routine 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. The disequilibrium models, however, typically involve unique equilibria in the sense that once price is specified exogenously the system of equations admits a unique solution.

4. Identification.

Identification holds if the structure can be recovered uniquely from the distribution of the observables, μ, and if this true for all each μ in the range of Γ. Let ρ be a metric on S. A necessary and sufficient condition for identification is

$$(4.1) \qquad \Gamma(s) \cap \Gamma(s') = \emptyset,$$

for all s and s' in S for which $\rho(s,s') > 0$. If this condition is met, $\Gamma^{-1}(\mu)$ is a function whose domain is $\Gamma(S)$, and whose range is S. The identification condition is appropriate even when $\Gamma(s)$ is a singleton for each s in S; i.e. even where there is no structure in S leading to non-uniqueness.

There is no necessary relation between nonuniqueness and lack of identifiability - neither is implied by the other. Nevertheless, non-uniqueness makes it "easier" for (4.1) to fail. As an extreme illustration, suppose that there is some s_0 in S for which $\Gamma(s_0) = M_Y$ -- if s_0 is true, then no conceivable set of observations can refute it because every outcome is an equilibrium. Then $\Gamma(s_0) \cap \Gamma(s) \neq \emptyset$ for every s in S.

When (4.1) fails, partial identification is often possible -- this will be shown by example.

5. Example 1: A two-person game.

In a two-person game, the payoff functions are

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

and

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

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 $U = [0,1]^2$, ν is the uniform distribution, and the remaining structural characteristic is just ϕ; the analyst knows a priori that $\phi \in (0,1] = \Phi$.

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 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

$$(y=)\ \psi_1(u) = 0 \qquad \text{all } u \in U$$

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

(Since there are no equilibria in which the two players do not always mimic one another, we write $y \equiv x_1 = x_2$). We then have $Y = \{0, 1\}$, and M_Y is a subset of R^2, the first coordinate measuring the probability that $y = 0$, and the second the probability that $y = 1$. Then

$$\mu_1 \equiv \nu \cdot \psi_1^{-1} = (1,0).$$

and

$$\mu_2 \equiv \nu \cdot \psi_2^{-1} = (1-\phi^2, \phi^2).$$

These two points and their convex combination comprise $\Gamma(s)$. See figure 1. Since the uniform ν is the only distribution of u admitted a priori, Γ is effectively a function only of ϕ. Moreover, the two components of μ must add up to one, so it is sufficient to specify only the second component of μ (the probability that $y=1$), which we denote by $\underline{\mu}$. Then (by a slight abuse of notation) we take Γ to be correspondence of Φ into $[0,1]$, getting

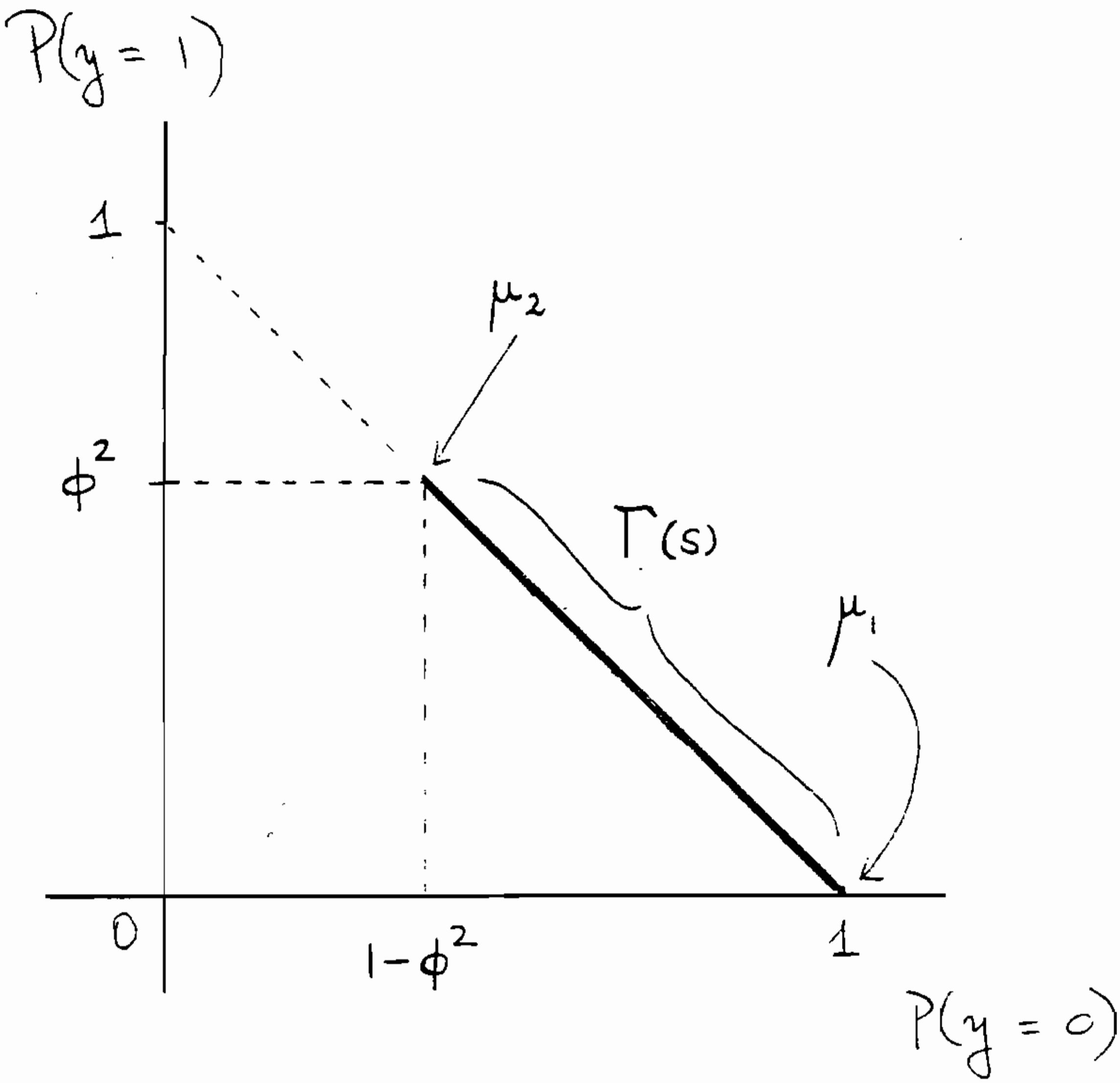

FIGURE I

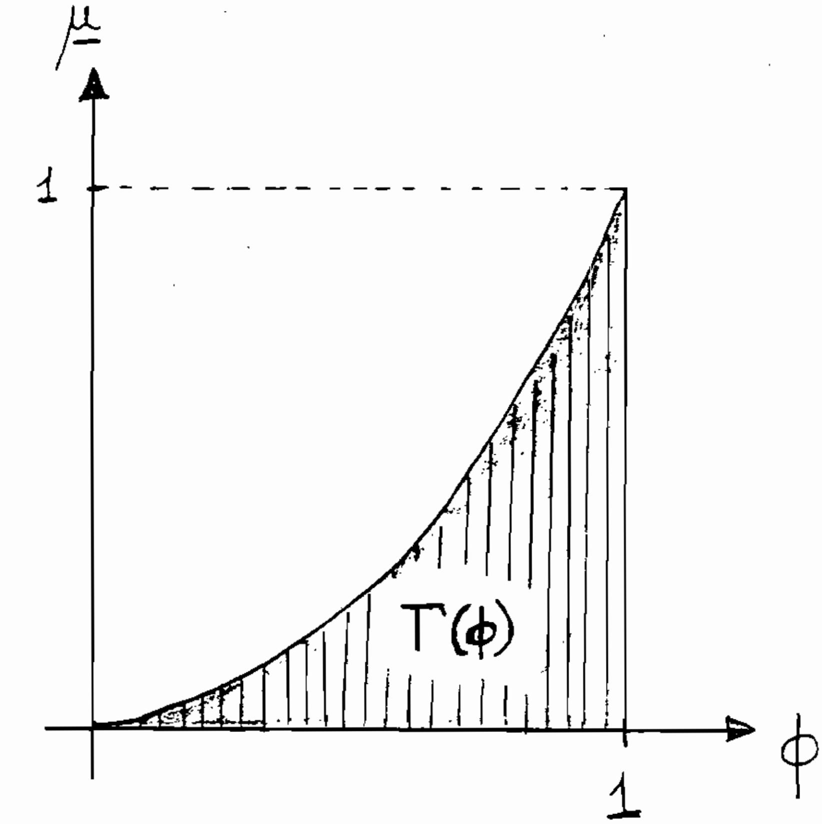

Figure 2A

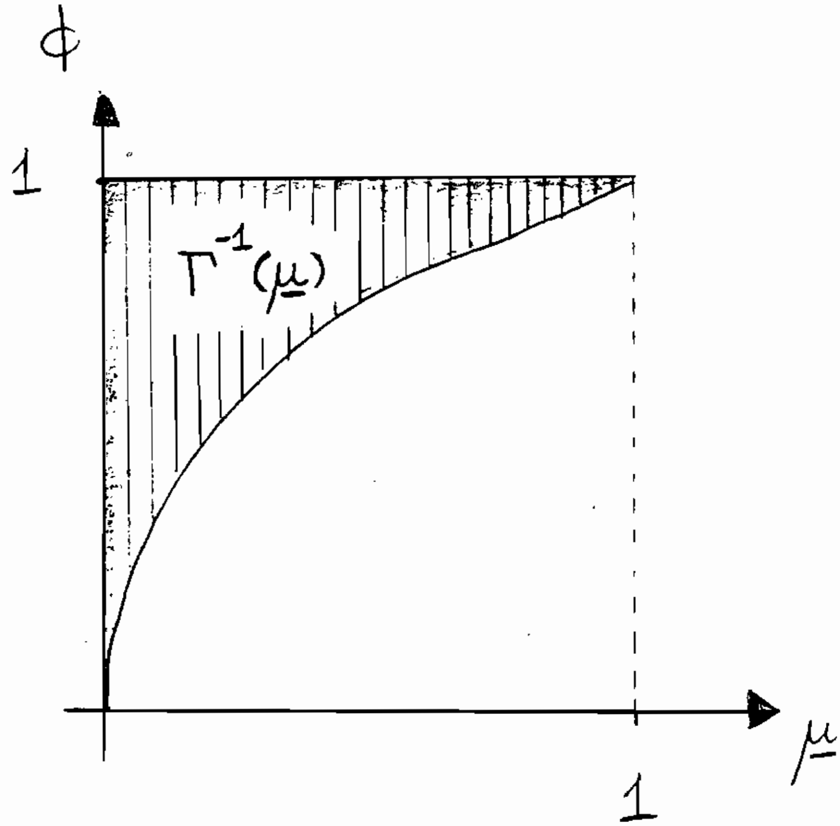

Figure 2B

$$\Gamma(\phi) = [0,\phi^2] \qquad \phi \in \Phi \quad ,$$

(5.1)

$$\Gamma^{-1}(\underline{\mu}) = [\sqrt{\underline{\mu}},1] \qquad \underline{\mu} \in [0,1].$$

Figures 2A and 2B illustrate Γ and Γ^{-1}.

Turning to identification, condition (4.1) fails: for any $\phi,\phi' \in \Phi$ with $\phi < \phi'$,

(5.2) $$\Gamma(\phi) \cap \Gamma(\phi') = [\phi^2,(\phi'^2)] \neq \emptyset.$$

In spite of the failure of (4.1), certain subsets of parameter space Φ can be identified. If $\underline{\mu}$ can be consistently estimated then from (5.1) we can infer that $\phi \in [\sqrt{\mu},1]$.

6. Example 2: Speculative bubbles.

An asset pays out a constant dividend $\gamma \geq 0$ at the end of each period. Let y_t be the price of the asset at t, and $E_t(\cdot)$ the expectation operator based on information available up to and including t. Under risk-neutrality the arbitrage condition is

(6.1) $$y_t = \beta(\gamma + E_t y_{t+1}), \qquad t \in T,$$

where $\beta < 1$ is the discount factor and T the non-negative integers. The unique stationary solution to (6.1) is

(6.2) $$y_t = \alpha = \beta\gamma/(1-\beta) \ , \qquad t \in T.$$

The non-stationary solutions to (6.2) are of the form

(6.3) $$y_t = \alpha + \epsilon_t, \qquad t \in T,$$

where $(\epsilon_t)_{t\in T}$ is a sequence of (possibly degenerate) random variables satisfying

(6.4) $\qquad E_t(\epsilon_{t+1}) = \beta^{-1}\epsilon_t$, and $\epsilon_t \geq -\alpha$; $\qquad t \in T$.

Thus (6.1) has a continuum of deterministic and of random solutions.

Using the language of section 2, we let $y = (y_t)_{t\in T}$, and $Y = R^{\infty}_{+}$, the latter endowed with the product topology. The ϵ_t can be thought of as extrinsic (sunspot) random variables; however, they should be thought of as endogenous in the sense that the joint distribution of the ϵ's is not determined by the structure, but is instead one of many possible equilibrium outcomes.

Let U be the unit interval, and ν Lebesgue measure on this interval. The space of realizations of the (ϵ_t) sequence will be denoted by R^{∞}, again endowed with the product topology. For each Borel measure η on R^{∞} satisfying (6.4), there is a family of measurable functions $(h^{\eta}_{t,\alpha})_{t\in T, \alpha\in[0,\infty)}$ (where $h^{\eta}_{t,\alpha}$: $U \rightarrow [-\alpha,\infty)$) such that if $h^{\eta}_{\alpha} = (h^{\eta}_{1,\alpha}, h^{\eta}_{2,\alpha}, \ldots)$, (so that h^{η}_{α}: $U \rightarrow [-\alpha.\infty)^{\infty}$),

(6.5) $$\eta = \nu\cdot(h^{\eta}_{\phi})^{-1}.$$

Let N_{ϕ} be the set of distributions η that satisfy (6.4). Then the set A_{ϕ} of section 3 is now given by

(6.6) $$A_{\alpha} = \cup_{\eta\in N_{\alpha}} (\psi^{\eta}_{\alpha})$$

where ψ^{η}_{α}: $U \rightarrow Y$ is defined by

(6.7) $$\psi^{\eta}_{\alpha} = \alpha + h^{\eta}_{\alpha}.$$

The definition of $\Gamma(s)$ in (3.1) is still valid, and the hypotheses of theorem (3.1) are met here.

Identification. Since any ν on U other than Lebesque measure is ruled out a priori, the only structural parameters of relevance are β and γ. Since there is never any change in market fundamentals (γ is constant), all price movements are attributable to bubbles. Interestingly, even if γ is not observable, both β and γ can be consistently estimated, unless the equilibrium is stationary in which case only α is estimable. [(6.3) and (6.4) imply that $E_t(y_{t+1}-\alpha) = \beta^{-1}(y_t-\alpha)$, so that $y_{t+1} = \alpha(1-\beta^{-1}) + \beta^{-1}y_t + v_t$ where v_t is uncorrelated with y_t. Thus if y_t exhibits some variation, both α and β are consistently estimable. Paradoxically, the "bad" equilibria are more informative about the structural parameters; the example is similar in spirit to others (e.g., Sargent and Wallace (1985), Woodford (1986 especially sec. 4)) in which identification can be secured in spite of severe multiplicity of equilibria.

7. Negligible agents in anonymous situations.

The assumptions of negligibility and anonymity deserve special attention because of the tractability they bring to the mathematical formulation the statistical analysis of equilibrium outcomes. We first sketch the theory of anonymous games between negligible agents (Mas-Colell (1985), Green (1984), Jovanovic and Rosenthal (1986)), and then compare matters to the formulation in section 2.

Let Z be the set of actions, z, that each player can take, and M_Z the set of probability measures on Z. A player is characterized by a payoff function $v: Z \times M_Z \to R.$ This specification embodies the anonymity assumption: in addition to his own action, the player cares only about the distributions of the actions of others. Let V be the set of admissable payoff functions. A game is then characterized by a measure η on V. Given a game η, a measure τ on $V \times Z$ is a <u>Nash equilibrium distribution</u> if (denoting by τ_v and τ_z the marginal of τ on V and Z respectively),

$$\text{(7.1)} \qquad \begin{aligned} &\text{(i)} && \tau_v = \eta \text{ and} \\ &\text{(ii)} && \tau(\{(v,z) \mid v(z,\tau_z) \geq v(Z,\tau_z)\}) = 1 \end{aligned}$$

The negligibility assumption is embodied in (ii): in deviating from z to, say, z′ in Z, the player does not alter τ_z. Informally, (i) and (ii) state that if almost all players behave optimally in response to τ_z, then τ_z will indeed be the resulting distribution of actions.

Now we shall compare this formulation with that of section 2. If $I = [0,1]$ is the set of agents,

$$\text{(7.2)} \qquad U = V^I \quad \text{and} \quad Y = Z^I.$$

Now one can compare (7.1) to (2.2). In (2.2), ϕ relates one sequence of (or a continuum of) random variables (u) to another (y) while (7.1) relates one <u>distribution</u> (τ_v) of the elements of one sequence of (or of a continuum of) random variables, to the <u>distribution</u>, τ_z, of another. The latter approach is simpler, and enables us to prove some general results in section 10.

The new space of latent variables, η, is M_v; and the space of endogenous data is M_z. The equilibrium correspondence of M_v into M_z is

$$\text{(7.3)} \qquad \Delta(\eta) = \{m \in M_z \mid m = \tau_z \text{ and } \tau \text{ satisfies (7.1)}\}.$$

This is the analogue of Γ.

In contrast to Γ, Δ is not a convex-valued correspondence; versions of theorems (3.1) and (3.2) do not hold in this context, and that is a virtue of the approach. The failure Δ to in general be convex-valued is shown in the context of example 4 below.

Identification. Let $\bar{\rho}$ be a metric on M_V generating the topology of weak convergence. The condition for the identification is that

(7.4) $$\Delta(\eta) \cap \Delta(\eta') = \emptyset$$

for all η and η' in M_V for which $\bar{\rho}(y,y') > 0$.

8. Example 3:

The first illustration of the approach in section 7 has $Z = R$, and hence M_Z consists of probability distributions on the line. For $z \in R$ and $m \in M_R$, the payoff function is

(8.1) $$v(z,m) \equiv \theta\bar{m}z - z^2/2 - \epsilon z,$$

where $\bar{m} \equiv \int z dm(z)$ is the average action of all players, and where $\epsilon \in R$ is a player-specific parameter independent over players, having density $g(\epsilon)$ with mean zero. The analyst wishes to estimate θ.

The reaction function for player ϵ is

(8.2) $$z = \theta\bar{m} - \epsilon \equiv R(\theta,\bar{m},\epsilon).$$

Nash equilibrium requires that in addition to (8.2), $\bar{m} = \int R(\theta,\bar{m},\epsilon)g(\epsilon)d\epsilon$, i.e.,

(8.3) $$\bar{m} = \theta\bar{m}.$$

If equations (8.2) and (8.3) hold, then (7.1) holds. The distribution η over payoff functions V is determined uniquely by θ and g. Given θ, g and $\bar{m}$, (8.2) shows that the density of y, denoted by f(y), satisfies

$$(8.4) \qquad f(y) = g(y-\theta\bar{m}) = g(y-\bar{m}).$$

Thus once g and θ are given, $\bar{m}$ is a sufficient statistic for the equilibrium that obtains. Since θ is the only unknown structural parameter (it is assumed for simplicity that the analyst knows g), and since $\bar{m}$ is a sufficient statistic for equilibrium m, we may specialize (7.3) to a correspondence $\bar{\Delta}$ relating θ and $\bar{m}$ only:

$$(8.5) \qquad \bar{\Delta}(\theta) = \begin{cases} \emptyset & \theta < 0 \\ 0 & \theta \in [0,1) \cup (1,\infty) \\ (-\infty,\infty) & \theta = 1 \end{cases}$$

The graph of $\bar{\Delta}$ is a subset of R^2 and is depicted in figure 3. No equilibrium exists for negative θ's; any $\bar{m}$ is an equilibrium when $\theta = 1$, and for the remaining positive θ's, equilibrium is unique at zero.

Although $\bar{\Delta}$ is here convex-valued, the same is not generally true of Δ. This will be shown in the next example. Note too that identification (7.4) fails.

9. Example 4: A continuum version of example 1.

This example is in a sense a continuum version of example 1, although its structure is more general. It has elements in common with the Diamond-Mortensen formulation of complementarities in search. There are two actions $z \in \{0,1\} = Z$. The payoff function is

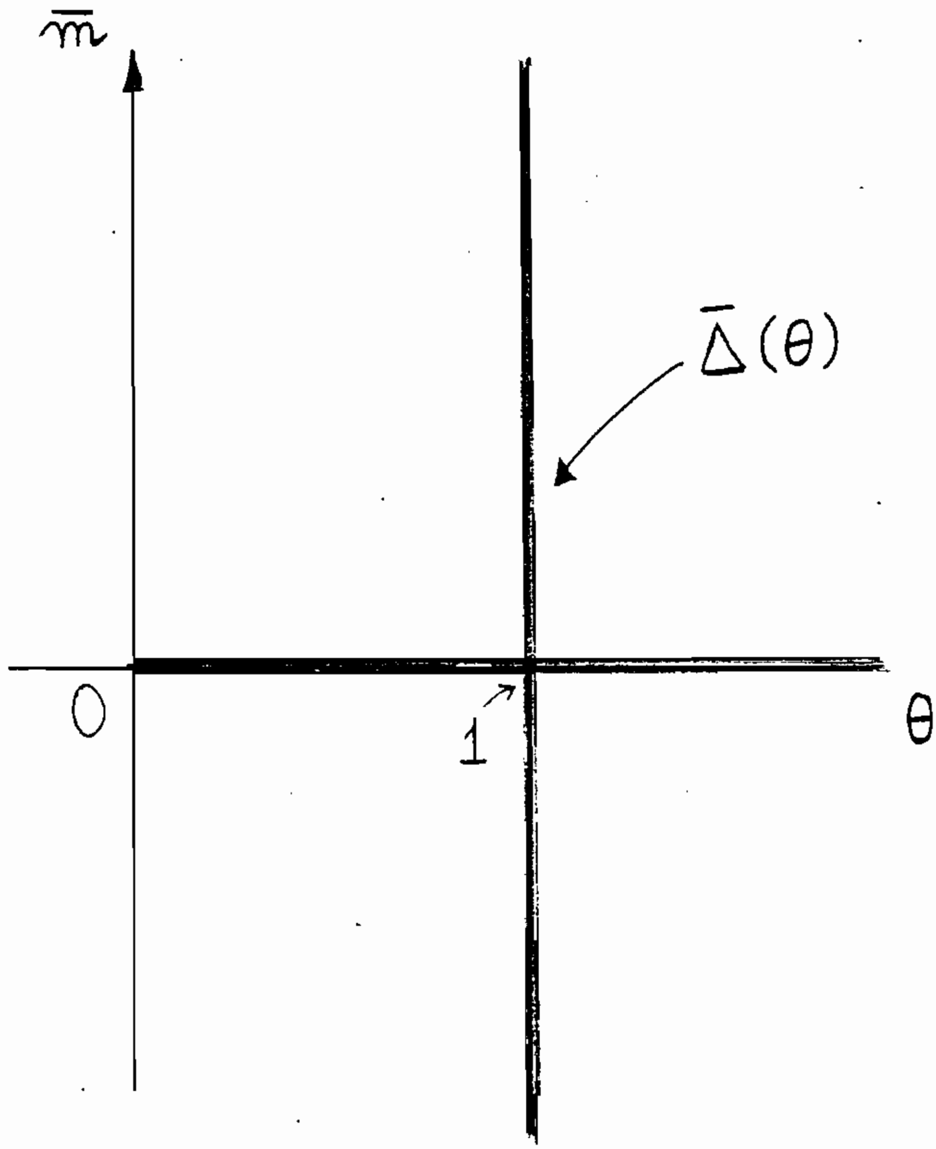

Figure 3

$$(9.1)\qquad v(z,m) = \begin{cases} h(\bar{m},\phi) - \epsilon & \text{if } z=1, \\ 0 & \text{if } z=0 \end{cases}$$

where $\bar{m} = \int z dm(z)$ is the average action in the population of all players, ϕ is a parameter vector, h is a continuous function, and ϵ a player-specific parameter with C.D.F. $G(\epsilon)$. When

$$(9.2)\qquad h(m,\phi) = \phi\bar{m} \quad \text{and} \quad G = \epsilon; \quad (\epsilon \in [0,1]),$$

we get the continuum version of example 1. Player ϵ chooses $z=1$ if and only if $\epsilon < h(\bar{m},\phi)$; therefore Nash equilibrium requires that

$$(9.3)\qquad \bar{m} = G[h(\bar{m},\phi)].$$

As in example 3, since G and h are given and only ϕ is not known to the analyst, and since $\bar{m}$ is then a sufficient statistic for m, the equilibrium correspondence is

$$(9.4)\qquad \bar{\Delta}(\phi) = \{\bar{m} \in [0,1] \,|\, \bar{m} = G[h(\bar{m},\phi)]\}.$$

To see that $\bar{\Delta}$ is not generally convex valued, just choose G and h such that for some ϕ, (9.3) has exactly two solutions for $\bar{m}$.

To compare matters at hand with the first example, assume that (9.2) holds. Our task is to compare the size of $\bar{\Delta}(\phi)$ with the size of $\Gamma(\phi)$ (see eq. (5.1)). Since any $\bar{m} \in \bar{\Delta}(\phi)$ solves

$$(9.5)\qquad \bar{m} = \phi\bar{m} \qquad (\bar{m} \in [0,1]),$$

we find that

$$(9.6)\qquad \bar{\Delta}(\phi) = \begin{cases} 0 & \phi \in [0,1) \\ [0,1] & \phi = 1 \end{cases}$$

and

$$\bar{\Delta}^{-1}(\bar{m}) = \begin{cases} 1 & \text{if } \bar{m} \in (0,1] \\ [0,1] & \text{if } \bar{m} = 0 \end{cases}$$

These relations are illustrated in figures 4A and 4B which should be compared with figures 2A and 2B. The predictive content of the continuum model is larger -- the Lebesgue measure of the graphs of $\bar{\Delta}$ and $\bar{\Delta}^{-1}$ is zero, whereas the Lebesgue measure of the graph of Γ and Γ^{-1} in figure 2 is $\int_0^1 \phi^2 d\phi = 1/3$. Moreover, the identifiability properties are more satisfactory as well, although full identifiability in the sense of (4.1) or (7.4) does not hold since $\bar{\Delta}(\phi) \cap \bar{\Delta}(\phi') = \{0\} \neq \emptyset\}$.

10. Consistent nonparametric estimation

When it is not known what equilibrium the economy will be in, the equilibrium becomes an additional parameter, and it must be an admissable equilibrium for the structure under consideration. Formally, the enlarged parameter space is the graph of the equilibrium correspondence, this graph is a subset of $S \times M_Y$:

(10.1) $$\Omega = \{(\mu,s) \in M_Y \times S \,|\, \mu \in \Gamma(s), \quad \Gamma(s) \neq \emptyset\}.$$

The theorems of this section, however, deal only with the negligible-agents case of section 7, where the parameter space is

(10.2) $$\Omega' = \{(m,\eta) \in M_Z \times M_V \,|\, m \in \Delta(\eta), \quad \Delta(n) \neq \emptyset\}.$$

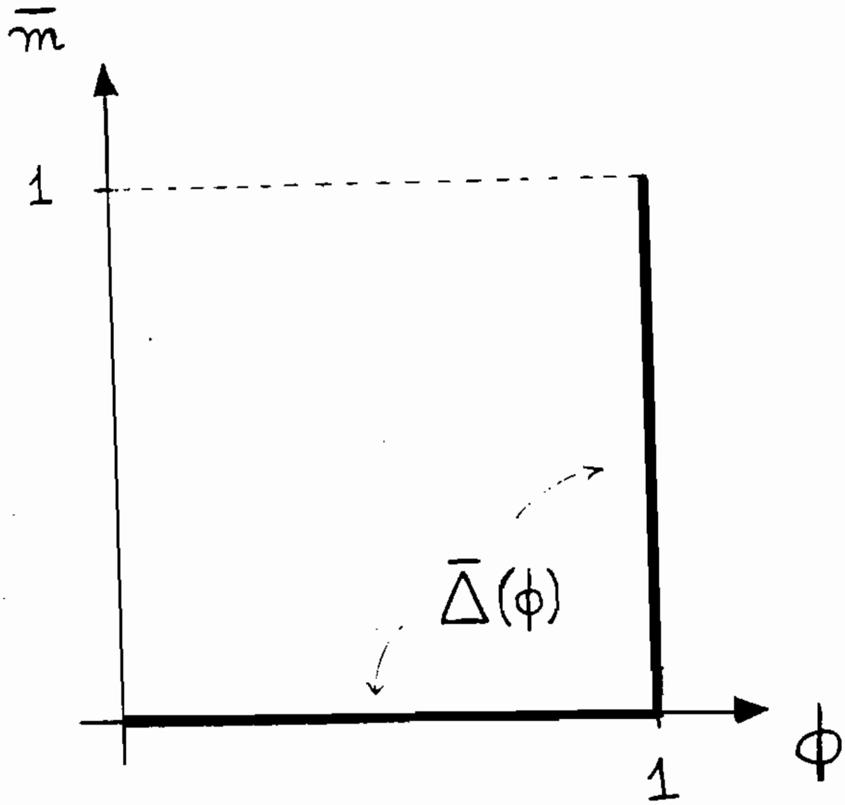

Figure 4A

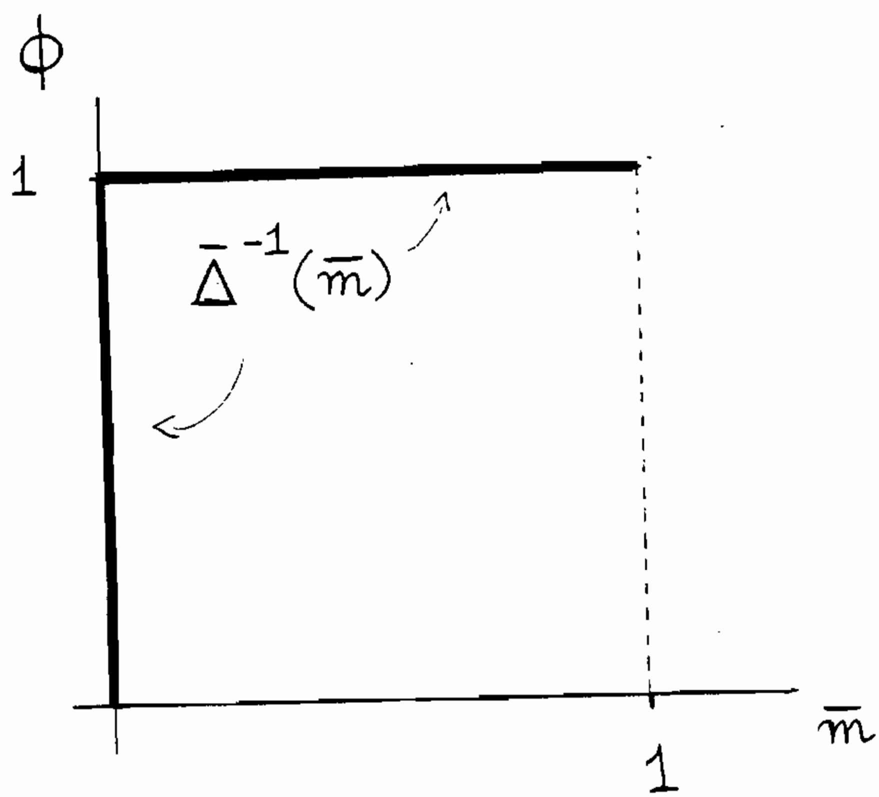

Figure 4B

When (10.2) is relevant and when certain other conditions prevail, a consistent estimator is available, and this section is concerned with two such estimators.

Let $\hat{z} = (z_1, \ldots z_n)$ be be the vector consisting of actions taken by a subset of size n of players from the game defined in section 7. The true structure and distribution of actions will be denoted by (η, m). The sample distribution of action is

$$(10.3) \qquad \hat{m}_n(A) \equiv n^{-1} \sum_{i=1}^{n} I_A(z_i),$$

where $I_A(\cdot)$ is the indicator of A, a Borel subset of Z. Assume that a priori, η confined to a compact subset of M_V, denoted by N. Now define the estimator t_n by

$$(10.4) \qquad t_n(\hat{z}) = [\tilde{m}_n, \Delta^{-1}(\tilde{m}_n)] \subset \Omega',$$

where

$$\tilde{m}_n = \underset{m' \in \Delta(N)}{\operatorname{argmin}} \ \bar{\rho}(m', \hat{m}_n),$$

and where $\bar{\rho}$ is a metric on M_Z that generates the topology of weak convergence. In words, the estimator t_n does the following: the estimated distribution of actions, $\tilde{m}_n$, is that member of equilibrium set, $\Delta(N)$, which is the closest to the sample distribution, $\hat{m}_n$. (When more than one closest member exists, any selection rule will do -- see the proof of lemma 1.) The structural estimate is then obtained by taking $\Delta^{-1}(\tilde{m}_n)$ which is non-empty since by construction $\tilde{m}_n \in \Delta(N)$. (One can not use $\hat{m}_n$ itself as the estimate because $\Delta^{-1}(\hat{m}_n)$ may be empty).

<u>Lemma 1</u>. A sequence $\tilde{m}_n$ exists, and if Z is separable, $P(\tilde{m}_n \to m) = 1$.

Proof: Since $\bar{\rho}$ is continuous, the existence of $\hat{m}_n$ follows if $\Delta(N)$ is compact. From Green (1984, thm. 3), Δ is u.h.c.; moreover, $\Delta(\eta)$ is closed-valued. Since N itself is compact, $\Delta(N)$ is itself compact (Hildenbrand. 1974, p. 24, proposition 3).

Turning to convergence, we note that although $\tilde{m}_n$ is not uniquely defined (because there may be more than one minimizer), since $m \in \Delta(N)$, $\bar{\rho}(\tilde{m}_n,\hat{m}_n) \le \bar{\rho}(m,\hat{m}_n)$. But since Z is separable, the Glivenko-Cantelli theorem (Hildenbrand 1974, p. 52) implies that $P[\lim_{n\to\infty} \bar{\rho}(m,\hat{m}_n) = 0] = 1$ (Hildenbrand 1974, p. 52). Then from the triangle inequality,

$$\bar{\rho}(\tilde{m}_n,m) \le \bar{\rho}(\tilde{m}_n,\hat{m}_n) + \bar{\rho}(\hat{m}_n,m) \le 2\bar{\rho}(m,\hat{m}_n),$$

and the second assertion of the lemma follows. Q.E.D.

Note that while $\Delta^{-1}(\tilde{m}_n)$ is never empty, it is not always a singleton. When $\Delta^{-1}(\tilde{m}_n)$ is not a singleton, the estimator t_n does not single out any one particular structure within that subset. This estimator is the subject of the next theorem:

Theorem 10.1: If Z is a separable metric space and if

(i) the payoff functions v are continuous,

(ii) eq. (7.4) holds (i.e., full identification),

(iii) the z_i $(i=1,\ldots,n)$, are i.i.d. draws from m,

the estimator t_n is strongly consistent as $n \to \infty$. That is,

$$P[\lim_{n\to\infty} t_n = (m,\eta)\,|\,m,\eta] = 1.$$

Proof: From lemma 1,

$$P(\lim_{n\to\infty} \tilde{m}_n = m) = 1. \tag{10.5}$$

From condition (ii) of the theorem, Δ^{-1} is not a correspondence; instead it is a function mapping $\Delta(N)$ into M_v. From (i) and from Green (1984, theorem 3), Δ is a u.h.c. correspondence. Hence Δ^{-1} maps closed sets into closed sets (Hildenbrand 1974, p. 21). And since Δ^{-1} is a function, it must be continuous. Therefore (10.5) implies that

$$(10.6) \qquad P[\lim_{n\to\infty} \Delta^{-1}(\bar{m}_n) = \Delta^{-1}(m)] = 1.$$

Then condition (ii), along with (10.5) and (10.6) implies the assertions of the theorem. Q.E.D.

Theorem 10.1 presents a positive result, but its conditions may not hold in practice, especially the second and third conditions. We shall now proceed to analyze the consequences of a failure of assumption (iii). As it stands, the assumption is that all the power, as n becomes large, comes from observing the actions of agents from a single economy and that the latter finds itself in a particular equilibrium. We may instead assume the reverse -- that we have just one observation per game, but that the number of games we can observe gets large. Such an assumption may be warranted when one has cross-country data, or more generally, when one has repeated observations of the same experiment that each time involves a different collection of players.

If each experiment is indeed the same (i.e. η is the same) but the equilibrium is perhaps different in each experiment, we can regard equilibrium as an incidental parameter, with some distribution λ over $\Delta(\eta)$. In this case the sample distribution $\hat{m}_n$ defined in (10.3) converges to m^* which is defined as follows: for any Borel subset A of Z,

$$(10.7) \qquad m^*(A) = \int_{\Delta(\eta)} m(A)\, d\lambda(m)$$

where $\lambda \in M_{\Delta(\eta)}$, the latter being the set of probability measures (i.e., "mixtures") over distributions in $\Delta(\eta)$.

Although this is not a discussion of maximum likelihood estimation, we are following the logic of Kiefer and Wolfowitz's (1956) treatment of incidental parameters. Thus the new parameter space is

$$(10.8) \qquad \Omega^* = \{(\lambda,\eta) \in M_{M_Z} \times M_V | \eta \in N, \text{ and } \lambda \in M_{\Delta(\eta)}\},$$

which should be compared with Ω' of (10.2).

The next step is to define an estimator. In doing so, it is helpful to note that in (10.7), m^* depends on η only in so far as the latter restricts λ to lie in $M_{\Delta(\eta)}$. The set of all possible λ's is $\bigcup_{\eta \in N} M_{\Delta(\eta)} \equiv M^*$. Hence (10.7) can be re-written as

$$(10.9) \qquad m^* = \int m d\lambda(m) \equiv \xi(\lambda),$$

so that $\qquad \xi: M^* \to M_Z$.

<u>D.1</u>. The family of mixtures $\xi(M^*)$ is <u>identifiable</u> if the function $\xi^{-1}:\xi(M^*) \to M^*$ exists.

This definition is analogous to the standard one employed in the parametric case (e.g., Teicher 1961, 1963): Knowledge of the mixture allows one to recover the mixing distribution. Next define the correspondence Δ_0: N $\to M_Z$ by $\Delta_0(\eta) \equiv M_{\Delta(\eta)}$. Thus $\Delta_0(\eta)$ is the set of mixtures that are admissable when η is true. When (D.1) holds, we define the estimator T_n by

$$(10.10) \qquad T_n(\hat{z}) = [\xi^{-1}(\tilde{m}_n), \Delta_0^{-1}(\xi^{-1}(\tilde{m}_n))] \subset \Omega^*.$$

The true parameter point in Ω^* will be denoted by (λ,η).

Theorem 10.2: Let the assumptions of theorem 3 hold, with the exception that (iii) is replaced by

(iii)′ the z_i (i=1,...,n) are i.i.d. draws from m^*, and with the added assumption that

(iv) (D.1) holds.

Then the estimator T_n is strongly consistent as $n\to\infty$. That is,

$$P[\lim_{n\to\infty} T_n = (\lambda,\eta) | \lambda,\eta] = 1.$$

Proof: Reasoning as in the proof of lemma 1, we have (a) $P(\tilde{m}_n \to m^*) = 1$. By construction, $m^* = \xi(\lambda)$, and since ξ is invertable, (b) $\lambda = \xi^{-1}(m^*)$.

Since (7.4) holds, the supports of the various mixing distributions admissible under different η's do not intersect. That is, $\bar{\rho}(\eta,\eta') > 0$ implies $\Delta_0(\eta) \cap \Delta_0(\eta') = \emptyset$. But then the function Δ_0^{-1} exists, and (c) $\Delta_0^{-1}(\lambda) = \eta$.

In view of (a), (b) and (c), the proof will be complete if we can show that ξ^{-1} and Δ_0^{-1} are both continuous.

To show continuity of ξ^{-1}, we observe that ξ is continuous (under the topology of weak convergence on its domain and range). Moreover ξ is a linear functional. But by (iv), ξ^{-1} exists, hence it must be continuous (and linear).

To show continuity of Δ_0^{-1}, we note that the u.h.c. property of Δ implies that Δ_0 is also u.h.c. Then one can repeat the argument of theorem 3 (relating to the continuity of Δ^{-1}) to conclude that Δ_0^{-1} is continuous. Q.E.D.

11. Maximum Likelihood Estimation

Let $\hat{y}$ be the observed subset of Y, and let $\hat{Y}$ be the subspace of Y to which $\hat{y}$ belongs; $\hat{Y}$ is always finite dimensional. The likehood of $\hat{y}$ will be denoted by

(11.1) $$L(\hat{y};\mu)$$

If the marginal of μ on $\hat{Y}$ has a density, $L(\cdot)$ is simply the joint density of $\hat{y}$. Maximizing the likehood amounts to finding that structure s and equilibrium μ that solve:

(11.2) $$\max_{s\in S}\left\{\max_{\mu\in\Gamma(s)} L(\hat{y};\mu)\right\}$$

This problem decomposes into two sub-problems. First, the ML estimate of μ, is gotten by

(11.3) $$\hat{\mu} = \underset{\mu\in\Gamma(S)}{\operatorname{argmax}}\ L(\hat{y},\mu),$$

where $\Gamma(S) = U_{s\in S}\,\Gamma(s)$. Assuming that this defines $\hat{\mu}$ uniquely, the ML estimate of s, denoted by $\hat{s}$, is

(11.4) $$\hat{s} = \Gamma^{-1}(\hat{\mu}),$$

and if (4.1) holds (complete identification), $\hat{s}$ is other also uniquely defined. [The unlikely case in which there are multiple maxima in (11.3), obvious modification is required].

The above is a "fixed-effect" treatment of equilibria, because all the data are assumed to come from the same structure <u>and</u> the same equilibrium distribution μ. Instead, however, the data may be drawn from different games or economies which have the same s but possibly different μ's. Suppose that $\hat{y}_j$: is the sample taken from economy j $(j \in J)$, and that its likelihood is $L(\hat{y}_j;\ \mu_j)$. Following Kiefer and Wolfowitz's (1956) treatment of the incidental parameters problem, one may solve

(11.5) $$\max_{s,G}\ \prod_{j\in J}\left\{\int_{\Gamma(s)} L(\hat{y};\mu)\,dG(\mu)\right\}$$

for the ML estimate $\hat{s}$, $\hat{G}$, where the G's are mixtures over distributions μ. This amounts to a "random effects" treatment of equilibria.

For the negligible agent case of section 7, the development parallels the above. The likelihood of a sample $\hat{z} = (z, \ldots, z_n)$ of agent's will be written as

$$(11.1)' \qquad L'(\hat{z};m),$$

the assumption being that all the agents are sampled from the same economy with equilibrium distribution m. If the sampling is independent,

$$(11.6) \qquad P(z_1 \in A_1, \ldots, z_n \in A_n) = \prod_{j=1}^{n} m(A_j)$$

for given sets $A_j \subset Z$ [One of the benefits of the negligibility assumption is that in contrast to the general case, (11.6) is generally valid]. Maximizing the likelihood then amounts to solving

$$(11.2)' \qquad \max_{\eta \in N} \left\{ \max_{m \in \Delta(\eta)} L'(\hat{z};m) \right\}.$$

This too decomposes into two sub-problems. First, the ML estimate $\hat{m}$ is

$$(11.3)' \qquad \hat{m} = \underset{m \in \Delta(N)}{\operatorname{argmax}} \; L'(\hat{z};m).$$

Assuming that this defines $\hat{m}$ uniquely, the ML estimate $\hat{\eta}$ is given by

$$(11.4)' \qquad \hat{\eta} = \Delta^{-1}(\hat{m})$$

and if (7.4) holds, $\hat{\eta}$ is uniquely defined.

In view of (11.6), we can be more specific about the random effects treatment of equilibria in the negligible-agent case. Suppose that for economy j $(j \in J)$ we have a sample of size n_j of independent draws of agents' actions $\hat{z}_j = (z_1, \ldots, z_{n_j})$. If m has density $g_m(z)$ at $z \in Z$,

$$L'(\hat{z}_j;m) = \prod_{i=1}^{n_j} g_m(z_j),$$

and the analog of (11.5) is

$$(11.5)' \qquad \max_{\eta,\lambda} \prod_{j \in J} \{\int_{\Delta(\eta)} L'(\hat{z}_j;m)d\lambda(m)\}$$

where the λ's are mixtures over distributions m.

12. Testing

This section and the next deal only with the general case. The extensions to the negligible agent case are straightfoward. Suppose that one wishes to test the hypothesis that $s = s_0$. When equilibrium under s_0 is unique and $\Gamma(s_0)$ is a singleton, this is a "simple" hypothesis in the sense that under s_0 the distribution of the observables is determined uniquely. When $\Gamma(s_0)$ contains more than one point, the hypothesis is "composite".

The following statement is always true:

$$(12.1) \qquad s = s_0 \Rightarrow \mu \in \Gamma(s_0);$$

while, if (4.1) holds (complete identification), the following statement is also true:

$$(12.2) \qquad s = s_0 \Leftrightarrow \mu \in \Gamma(s_0).$$

Therefore, under complete identification, testing for the truth of _any_ structural hypothesis s_0 in S is equivalent to testing for the _composite_ hypothesis that $\mu \in \Gamma(s)$. Once a particular s_0 is specified, however, the equivalence in (12.2) will remain true so long as

$$\Gamma(s_0) \cap \Gamma(s) = \emptyset \tag{12.3}$$

for all s in S, $(s \neq s_0)$. This condition is weaker than (4.1).

The remainder of this section is devoted to showing that even when (12.3) (and hence (4.1)) fails, the Neyman-Pearson theory of hypothesis testing (Kendall and Stuart 1961, chapters 22, 23) readily extends to structures with multiple equilibria. Suppose that we wish to test the hypothesis that $s = s_0$ against the alternative that $s \in A \subsetneq (S - s_0)$. The optimal test is to look for a rejection region $w \subseteq \hat{Y}$ that solves the problem:

$$\max_{w} \left\{ \min_{\mu' \in \Gamma(A)} \int_{w} L(\bar{y};\mu')d\bar{y} \right\} = 1 - \beta, \tag{12.4}$$

subject to

$$\max_{\mu \in \Gamma(s_0)} \int_{w} L(\tilde{y};\mu)d\bar{y} \leq \alpha. \tag{12.5}$$

The pre-assigned quantity α is the size of the test. The constraint (12.5) states that the probability of a type I error may not exceed α no matter what μ in $\Gamma(s_0)$ is true. In (12.4), the objective is to maximize the power of the test against the least favorable alternative in $\Gamma(A)$. Thus, no matter what $\mu' \in \Gamma(A)$ is considered, the probability of a type II error is no greater than β. [When $\Gamma(s)$ and $\Gamma(A)$ are both singletons, the problem becomes the familiar one of finding a best critical region for testing a simple null against a simple alternative (Kendall and Stuart 1961, p. 174)].

13. Bayesian Procedures.

In this section we briefly mention some points of relevance to the Bayesian approach.

1. <u>The support of the prior</u>. The prior measure π should have Ω (or Ω') as its support. Then

$$\int_{\Gamma(s)} \pi(d\mu|s) = 1,$$

where $\pi(\cdot|s)$ is the conditional measure on μ given s. One difficulty with this requirement is that sometimes the size of $\Gamma(s)$ or its exact nature is unknown. For instance, for certain structures s we may be able to prove the existence of an equilibrium (i.e., non-emptiness of $\Gamma(s)$), but we may have little by way of characterization. This is a general problem, not specific to the Bayesian approach.

2. <u>Nonparametric Bayes methods</u>. When the general approach of section 2 is used, theorems (3.1) and (3.2) imply that in general no known parametric family can include every distribution in $\Gamma(s)$ when the latter is not a singleton (in specific examples $\Gamma(s)$ may be included as part of a parametric family - e.g., example 1 and figure 1). Nonparametric Bayes methods may be needed; recent discussion of these is in Diaconis and Freedman (1986).

3. <u>Choice of prior</u>. The measure π is a joint measure over structures and their equilibria. Since $\Gamma(s)$ will often include an infinite number of μ's, (thus 3.1 and 3.2) we immediately come up against the familiar problem that there are many ways to assign a prior in such a way as to assign zero weight to each μ in $\Gamma(s)$. But even when a structure has a finite number of equilibria (it is often the case that $\Delta(\eta)$, for fixed η, has finitely many elements), what criteria are there for assigning prior weight to equilibria? The following is a list of the sorts of equilibria that, according to some writers, are especially deserving:

(a) Symmetric equilibria, when they exist.

(b) "Naively" stable equilibria, e.g., Walras equilibria that are stable in the face of fictitious tatonnement. Sophisticated extensions appear in the game theory (Harsanyi and Selten, 1984) and in the rational expectations literature (Lucas 1985).

(c) Simple equilibria, in which strategies depend upon the fewest number of variables. (This encompasses (a) as a special case).

(d) Equlibria that are focal points.

14. Some applied problems.

This section has three subsections. The first illustrates the use of the ML and nonparametric estimators. The second and third illustrate the bias (first asymptotic and then finite-sample) that these estimators exhibit towards models that have less predictive content.

(A) Illustration of the non-parametric and ML estimators. This subsection briefly deals with the estimator t_n of eq.(10.4) and the fixed-effect ML estimator of eq.(11.3), as applied to example 4 (section 9), with the additional assumption of eq. (9.2) added on. The analyst has observations $(z_1,\ldots z_n)$ -- actions taken by a random sample of n players. The likelihood of the sample is proportional to $\bar{m}^k(1-\bar{m})^{n-k}$. The non-parametric estimate and the ML estimate of $\bar{m}$ are both equal to k/n. The estimate is strongly consistent by the SLLN. But by (9.7), the estimate of ϕ is in both cases given by

$$\hat{\phi} = \begin{cases} 1 & \text{if } k > 0 \\ [0,1] & \text{if } k = 0 \end{cases}$$

Then $\hat{\phi} \xrightarrow{a.s.} \phi$ if and only if $\phi = 1$ and $\bar{m} > 0$. This poor performance of both estimators is due to the failure of the identification condition (7.4) to hold, so that the hypothesis of theorem 10.1 are not met.

(B) Asymptotic bias of the ML and nonparametric estimators where identification fails. When the number of parameters changes as the structure varies, one can expect (a) finite sample bias even when identication holds, and (b) when identification fails (so that theorem 10.1 does not apply) even asymptotic bias towards structures that have less predictive power. The subsection uses example 3 to show that (b) can be true, and we shall do so for the ML estimator only -- the argument for the nonparametric estimator is virtually the same.

Once again we assume that a random sample $(z_1,\ldots,z_n)$ of agents' actions is available. The density of each observation is given in (8.4). According to (11.3)' one can first compute the ML estimate of $\bar{m}$, which solves:

$$\max_{\bar{m}\in R} \sum_{i=1}^{n} \ln g(z_i - \bar{m})$$

(note that $\Delta(N) = R$ so that $\bar{m}$ is unrestricted). Let $\hat{m}_n(z)$ denote the (set of) solution(s) to (14.1). If lng is strictly concave (a mild restriction met by many densities including the normal -- see Pratt, 1981), $\hat{m}_n$ will be a function from $R^n \to R$; moreover, $\hat{m}$ will be a strictly increasing function. Then the set of points $\hat{m}_n^{-1}(0)$, (a subset of R^n) is of measure zero under the true distribution $\prod_i g(\cdot - \bar{m})$, since g has no mass points. Therefore, $\hat{m}_n \neq 0$ with probability one, and this is true for each n. Then from (11.4) and (8.5), the ML estimate of θ, namely

$\bar{\Delta}^{-1}(\hat{m}_n)$, is equal to one with probability one, and this too is true for all n. Thus for all θ except $\theta = 1$, the ML estimator is inconsistent.

The intuition for all this result is straightfoward. Suppose indeed that $\theta \neq 1$, so that $\bar{\Delta}(\theta) = 0$. Then $\hat{m}_n \xrightarrow{a.s.} 0$. Nevertheless, $P(\hat{m}_n = 0) = 0$ for each fixed n. Hence for each n, no matter how large, $\theta = 1$ yields the highest likelihood, with probability one.

(C) Finite sample bias. Even if the identification conditions are fullfilled so that the nonparametric estimator (and probably the ML estimator) is consistent, one expects finite-sample bias towards structures with less predictive power. The following modification of example 4 (section 9) is designed to give us some idea about the magnitude of the bias, and about the rate at which it goes away assymptotically.

Assume that (9.1) holds and that $\phi \in \{\phi_2, \phi_2\}$; i.e., there are only two structures admissable a priori. Assume also that G and h are chosen in such a way that

$$\text{(14.2)} \qquad \bar{\Delta}(\emptyset) = \begin{cases} \frac{1}{2} & \phi = \phi_1 \\ \{\frac{1}{2} - 2\gamma, \frac{1}{2} + 2\gamma\} & \text{if} \quad \phi = \phi_2. \end{cases}$$

That is, structure ϕ_1 has a unique equilibrium, structure ϕ_2 has two equilibria. Suppose as in (A) above that a sample $(z_1, \ldots z_k)$ is given, and let k once more be the number of players choosing z=1. The likelihood of k is

$$\text{(14.3)} \qquad \binom{n}{k} \bar{m}^k (1 - \bar{m})^{n-k}.$$

The ratio k/n is, for large n, roughly normal with mean $\bar{m}$ and variance $\bar{m}(1-\bar{m})/n$. The ML estimate of $\bar{m}$ maximizes (14.3) subject to the constraint that $\bar{m} \in \bar{\Delta}(\phi)$. Our interest is in the question of how frequently the structure ϕ_1 is picked when it is true, compared to the frequency with which ϕ_2 is chosen where it is true. To obtain the exact finite-sample distribution of the ML estimate is a difficult task, so we shall use the normal approximation. The sole purpose of the discussion of example 4 is to provide an economic justification for what follows.

Let $\bar{x}_n = k/n$. For fixed n, the distribution of $\bar{x}_n$ is assumed to be $N(\frac{1}{2}, n^{-1}\sigma^2)$ when ϕ_1 is true, and either $N(\frac{1}{2} - 2\gamma, n^{-1}\sigma^2)$ or $N(\frac{1}{2} + 2\gamma, n^{-1}\sigma^2)$ where ϕ_2 is true. This situation is depicted in figure 5. [The model of example 4 implies that when ϕ_1 is true we ought to have $\sigma^2 = \frac{1}{4}$, whereas when ϕ_2 is true, we should have $\sigma^2 = (\frac{1}{2} + 2\gamma)(\frac{1}{2} - 2\gamma) = \frac{1}{4} + 2\gamma$. We shall ignore this difference in the interest of tractability. This can be justified either by assuming that γ is small, or by imagining a somewhat different model that will imply identical σ's) As n increases all three distributions become tighter. Let $\hat{\phi}_n$ be the ML estimate of ϕ. Then

$$\hat{\phi}_n = \begin{cases} \phi_1 & \text{if } \bar{x}_n \in [\frac{1}{2} - \gamma, \frac{1}{2} + \gamma] \\ \phi_2 & \text{otherwise} \end{cases} \qquad (14.4)$$

Elementary calculation leads to

$$P(\hat{\phi}_n = \phi_1 | \phi = \phi_1) = \xi(\gamma n/\sigma) - \xi(-\gamma n/\sigma) = P_1 , \qquad (14.5)$$

where ξ is the cumulative standard normal integral, and to

$$P(\hat{\phi}_n = \phi_2 | \phi = \phi_2) = 1 - [\xi(3\gamma n/\sigma) - \xi(\gamma n/\sigma)] = P_2. \qquad (14.6)$$

Because of the symmetry, it makes no difference in the calculation of P_2 whether $\bar{m} = \frac{1}{2} + 2\gamma$ or $\bar{m} = \frac{1}{2} - 2\gamma$. We expect $P_2 > P_1$ since structure ϕ_2 offers two distributions while structure ϕ_1 offers only one. Indeed $P_2 - P_1$ may be thought of as the bias in favor of the structure with two equilibria.

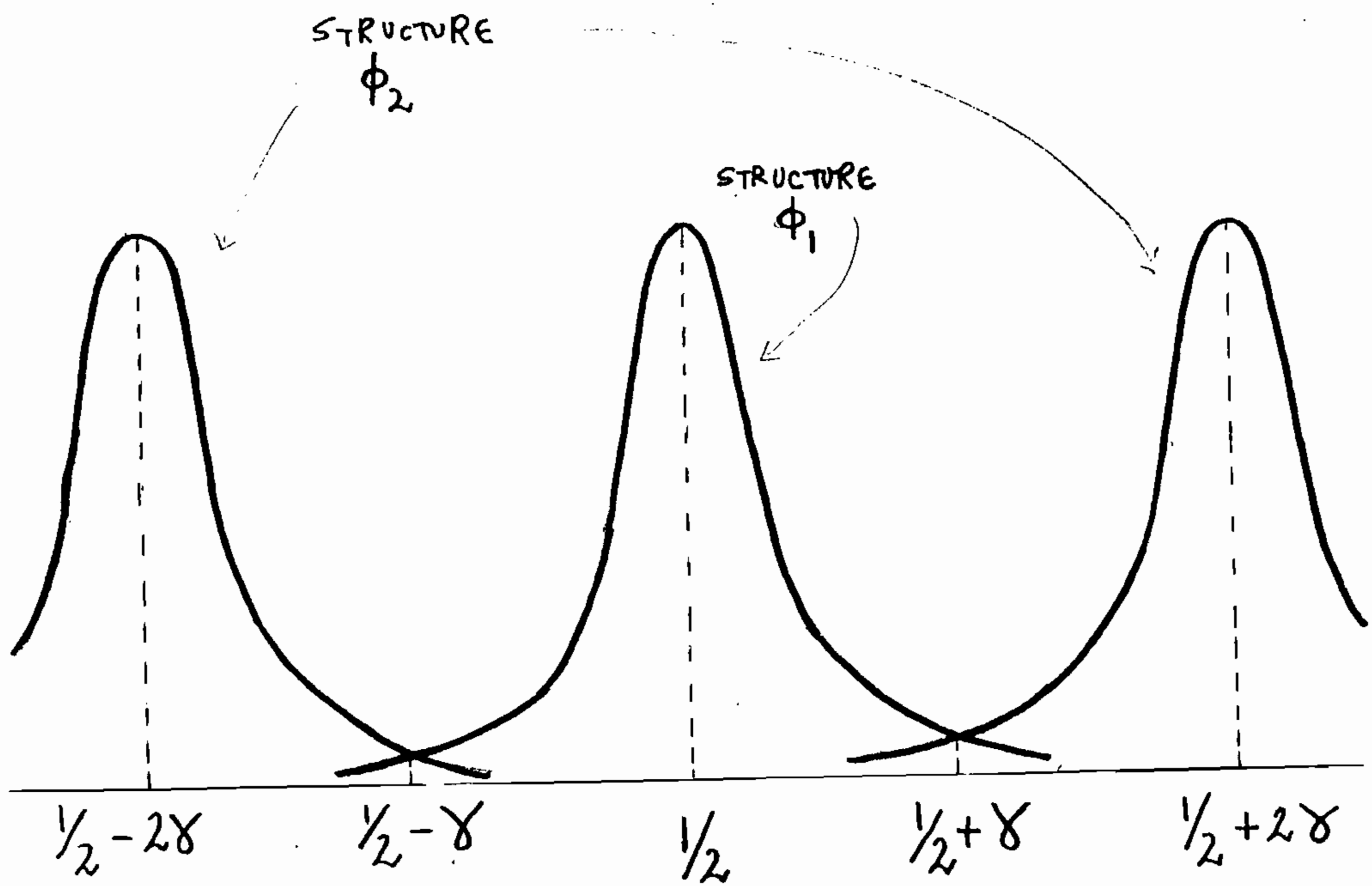

Figure 5

Table 1 plots the values of P_1, P_2, and P_2-P_1 as a function of $\gamma n/\sigma$. The table shows that if γ and σ are of the same order of magnitude, the bias P_2 - P_1 is not a problem since it goes rapidly to zero as soon as four or more observations are available. The general point that if the predictions of the two structures differ significantly relative to the noise in the data, the bias is not large, and it quickly goes away.

TABLE 1

γ_n/σ	P_1	P_2	$P_2 - P_1$
0.00	0.0000	1.0000	1.0000
.01	.0080	.9920	.9840
.02	.0160	.9841	.9681
.03	.0239	.9761	.9522
.04	.0319	.9682	.9363
.05	.0399	.9603	.9204
.06	.0478	.9525	.9047
.07	.0558	.9447	.8889
.08	.0638	.9371	.8733
.09	.0717	.9294	.8577
.1	.0800	.9219	.8423
.2	.1585	.8535	.6950
.3	.2358	.8020	.5662
.4	.3108	.7705	.4597
.5	.3829	.7583	.3753
.6	.4515	.7617	.3102
.7	.5161	.7759	.2598
.8	.5763	.7963	.2201
.9	.6319	.8194	.1875
1.0	.6827	.8427	.1600
1.5	.8664	.9332	.0668
2.0	.9545	.9773	.0228
2.5	.9876	.9938	.0062
3.0	.9973	.9987	.0013
3.5	.9995	.9998	.0002
4.0	.9999	.9999	.0001

Note: Differences inexact due to rounding error.

References

Allen, B. "Continuous Random Selections from the Equilibrium Correspondence," CARESS working paper # 85-25, University of Pennsylvania, April 1985.

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

Diaconis, P. and D. Freedman. "On the Consistency of Bayes Estimates," Annals of Statistics 14, no. 1 (1986): 1-26.

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

Green, E. "Continuum and Finite-Player Noncooperative Models of Competition," Econometrica 52, no. 4 (July 1984): 975-995.

Harsanyi, J. and R. Selten. "The Tracing Procedure," International Journal of Game Theory, 1984.

Hildenbrand, W. Core and Equilibria of a Large Economy. Princeton University Press, Princeton, N.J. 1984.

Jovanovic, B and R. Rosenthal. "Anonymous Sequential Games," Starr Center working paper # 86-12 New York University, March 1986.

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.

Lucas, R.E. Jr. "Adaptive Behavior and Economic Theory," unpublished manuscript, University of Chicago, 1985.

Mas-Colell, A. "On a Theorem of Schmeidler," Journal of Mathematical Economics 1985.

Pratt, J. "Concavity of the Log likelihood," Journal of the American Statistical Association March (1979): 103-106.

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

Sargent, T. J. and N. Wallace. "Identification and Estimation of a Model of Hyperinflation with a Continuum of 'Sunspot' Equilibria." Discussion paper # 220, University of Minnesota, September 1985.

Teicher, H. "Identifiability of Mixtures, "Annals of Mathematical Statistics 32 (1961): 244-248.

Teicher, H. "Identifiability of Finite Mixtures, "Annals of Mathematical Statistics 34 (1963): 1265-69.

Wald, A. "Note on the Consistency of the Maximum Likelihood Estimate," Annals of Mathematical Statistics 20 (1949): 595-60.

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

Woodford, M. "Expectations, Finance and Aggregate Instability," unpublished, New York University, March 1986.