Uncertainty and Disagreement in Equilibrium Models*

Nabil I. Al-Najjar[†] Eran Shmaya[‡]

October 2012;
This version: February 2014

Abstract

Leading equilibrium concepts require agents' beliefs to coincide with the model's true probabilities and thus be free of systematic errors. This implicitly assumes a criterion that tests beliefs against the observed outcomes generated by the model. We formalize this requirement in stationary environments. We show that there is a tension between the requirements that beliefs can be tested against systematic errors, on the one hand, and allowing agents to disagree or be uncertain about the long-run fundamentals. We discuss the implications of our analysis in the contexts of asset pricing and dynamic games.

*We thank Luciano De Castro for his comments, and Luciano Pomatto and Jaber Zarezadeh for their research assistance.

[†] Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston IL 60208. Link to Research Page.

[‡] Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston IL 60208.

Contents

1 Introduction

Leading equilibrium concepts such as Bayesian-Nash equilibrium and rational expectations equilibrium incorporate the idea that agents' beliefs about future outcomes coincide with the model's true probabilities.A natural motivation for this requirement is that the competitive pressures in economic and strategic interactions create strong incentives to avoid systematic forecast errors.

Requiring beliefs to be free of systematic errors implicitly assumes that there is a criterion or test to verify that they are consistent with observed outcomes. This paper formalizes this requirement and examines its implications for the sort of equilibria one might expect to arise. Roughly, an equilibrium is *testable* if every non-equilibrium belief can be rejected with positive probability by a test that compares that belief with observed outcomes. Testability is motivated by, and is a stylized representation of, statistical tests that outside observers, such as an econometrician, might use to take theoretical models to the data. Our focus, however, is on the economic and strategic implications of testability rather than the empirics of testing per se.

We first take the perspective of an outside observer who models an infinite-horizon economy as a stochastic process P. This process describes the assumed evolution of exogenous variables, as well as the optimizing behavior of agents and the resulting equilibrium outcomes. Leading equilibrium concepts require agents' beliefs to agree with P. We start with the case that the outside observer directly assumes that the system is in equilibrium, without explicitly modeling the agents' optimization. Our central assumption is that P is stationary. Stationarity is a natural assumption, widely used in dynamic stochastic models for its tractability and conceptual appeal. It is also necessary for the application of many standard empirical methods. We illustrate the role and limitations of stationarity in a number of important contexts, such as consumption-based asset pricing, Markov perfect equilibria, and Bayesian-Nash equilibrium in Markovian environments.

Our main result states that requiring an equilibrium to be testable is equivalent to any one of the following three properties. First, an equilibrium P is testable if and only if it precludes disagreement: any alternative belief Q

that is consistent with observed outcomes with probability one must be equal to P. Second, an equilibrium is testable if and only if agents' beliefs preclude structural uncertainty about long-run fundamentals, in the sense that no realization of past observations can change their opinions about the model's long-run properties. Intuitively, the absence of structural uncertainty means that agents have nothing further to learn from data about structural parameters. Learning may still occur, but only for predicting short-run outcomes. Third, testability is related to the properties of moment conditions used in the econometrics of non-linear dynamic stochastic models. We show that an equilibrium is testable if and only if the empirical moment conditions identify the model's true parameters asymptotically.[1]

We interpret these findings as highlighting a connection (and potential tension) between compelling desiderata of dynamic stochastic models. Testing agents' equilibrium beliefs against actual observations reflects the view that the concept of equilibrium should have observable implications, and not be just an article of faith. A non-testable equilibrium is one for which there exists alternative beliefs that cannot be rejected with positive probability, regardless of the amount of data. Testability is also naturally related to empirical estimation of equilibrium models, a link formalized by its equivalence to the consistency of empirical moment conditions. Given the wide use of moment conditions in empirical work, we view this consistency as another desirable property a model should have.

Our main theorem says that testability precludes disagreement and structural uncertainty about long-run fundamentals, two properties that are increasingly important in modeling economic and financial phenomena. For example, the absence of disagreement, in the form of common or concordant priors, implies no trade theorems that are widely viewed as inconsistent with observed investors' behavior and volume of trade. A natural response to this conflict between empirical evidence and theoretical predictions is to weaken the common prior assumption so agents can disagree on how to interpret information.[2] Relatedly, structural uncertainty about long-run fundamentals

[1] We focus on the generalized method of moments (GMM) introduced by Hansen (1982). See Section 3.3 for a detailed formal discussion.

[2] There is a large and growing literature on the role of disagreement in explaining the

and its gradual resolution through learning are increasingly viewed as relevant to understanding a number of empirical puzzles. In their survey of the subject, Pástor and Veronesi (2009) suggest that "[m]any facts that appear baffling at first sight seem less puzzling once we recognize that parameters are uncertain and subject to learning."[3] Our theorem indicates a potential tension between disagreement and structural-uncertainty, on the one hand, and the usual conceptual justification for equilibrium and its empirical estimation, on the other.

A stochastic process describing a dynamic economic system in equilibrium must take into account that agents' decisions are optimal given their beliefs. Section 5 focuses on how learning and equilibrium relate to testability and the other properties discussed above. In the case of passive learning, Proposition 5.1 shows, under mild additional assumptions, that structural uncertainty about exogenous variables tends to lead to non-stationary endogenous responses. Intuitively, uncertainty about long-run fundamentals implies that an agent's beliefs will be non-stationary as uncertainty is resolved through learning. This result is in the spirit of Nachbar (2005)'s theorem on the tension between learning and optimization in repeated games. We discuss Nachbar's work in Section 6.

Although leading equilibrium concepts require agents' beliefs to agree with the model's true probabilities, models differ in how this requirement is implemented. In a Bayesian-Nash equilibrium (with a common prior), agents are only required to have a subjective belief about how Nature selects the true parameters of the model. A more demanding approach is rational expectations where beliefs are objective and coincide with the model's observed empirical frequencies. Beliefs in a rational expectations equilibrium are testable, while beliefs in a Bayesian-Nash equilibrium can accommodate

trade volume and other puzzle in financial markets. See Hong and Stein (2007) for an excellent survey. Ross (1989), for instance, notes that: "It seems clear that the only way to explain the volume of trade is with a model that is at one and the same time appealingly rational and yet permits divergent and changing opinions in a fashion that is other than ad hoc."

[3] Lewellen and Shanken (2002), Brav and Heaton (2002), and Weitzman (2007) are early examples of works that introduce learning about fundamentals to explain equity premia, risk-free rates, excess volatility, and predictability of returns. See Pástor and Veronesi's survey of the literature.

structural uncertainty and disagreement. The ideal would be to "have it both ways," *i.e.,* to maintain the flexibility of the Bayesian-Nash equilibrium without sacrificing testability or identification via moment conditions. Our results point out to difficulties in reaching such compromise.

The message of this paper is not that disagreement and structural uncertainty, as we define them, have no empirical content. Rather, the difficulty is in reconciling them with the idea of a "true model" shared by agents and outside observers.[4] As noted in Section 5.3, subjective disagreement and structural uncertainty have observable implications on behavior, even though there may be no "true model" to objectively test beliefs against, or that can be recovered from empirical moment conditions.

The paper proceeds as follows. In Section 2 we introduce the central concept of testability. We discuss models of asset pricing and game theory to motivate the main concepts. Section 3 introduces the properties of disagreement, structural uncertainty, and the consistency of empirical moment conditions. Section 4 states our main theorem, while Section 5 connects the theorem to learning and equilibrium. Finally, Section 6 provides additional connections to the literature.

2 Testing Beliefs

2.1 Preliminaries

We consider an infinite horizon model, with time periods denoted $n = 0, 1, \ldots$. In period n, an outcome s_n in a finite set S is realized.[5] Let $H = S \times S \times \cdots$ denote the set of infinite histories. The set of histories of length n is denoted H_n, and the (finite) algebra of events generated by these histories is $\mathcal{H}_n$. Stochastic processes induce distributions on the set of infinite histories H, with the σ-algebra $\mathcal{H}$ generated by the finite-horizon events $\cup_{n=1}^{\infty} \mathcal{H}_n$.

[4] See Sargent (2008): "a rational expectations equilibrium asserts that the same model is shared by (1) all of the agents within the model, (2) the econometrician estimating the model, and (3) nature, also known as the data generating mechanism."

[5] The finiteness of S is convenient to avoid inessential technical complications.

A probability distribution P on H will stand for either the true (exogenous or equilibrium) process, or for an agent's belief about that process. Which of these interpretations is being referred to will be clear from the context. Our focus is on *stationary distributions*. Recall that a probability distribution P is stationary if, for every k, its marginal distribution on the algebra $\mathcal{H}_l^{l+k}$ generated by coordinates $l, \ldots, l+k$ does not depend on l. See, for example, Gray (2009). Let $\mathcal{P}$ denote the set of stationary distributions over H.

2.2 Tests

Equilibrium concepts, such as Nash equilibrium and rational expectation equilibrium, require agents' beliefs to coincide with the true equilibrium probabilities. Consider an outside observer (a modeler or an econometrician) who assumes that agents hold equilibrium beliefs and observes a realization of the process. Verifying that agents' beliefs coincide with the model's probabilities requires an objective criterion, or mechanism that compares beliefs with observed outcomes. Statistical testing is a natural language to formalize this requirement.

Fix a stationary stochastic process P which we interpret as the "equilibrium" process. This process describes the evolution of the model's exogenous variables (*e.g.,* income and technology shocks) as well as endogenous variables (*e.g.,* strategies, consumption levels, or asset pricing). A statistical *test* is any function

$$T : \mathcal{P} \times H \to \{0, 1\}$$

that takes as input a distribution P and a history h and returns a yes/no answer. Interpret an outcome $T(P, h) = 1$ to mean "history h is consistent with the process P;" otherwise, h is inconsistent with P. The set:

$$T_P \equiv \{h : T(P, h) = 1\}$$

is the set of all observations viewed as consistent with P under this test. We may interpret T_P as the empirical predictions of P relative to T: if P is correct, then observations must be in T_P.

Tests in this paper depend on the entire infinite realization of the process, while real-world statistical tests use finite observations.[6] On the other hand, asymptotic testing makes testability an easier hurdle to overcome: if an equilibrium theory is not testable in our sense, then it is certainly not testable using finite tests. This has the implication that failure of testability in our model is not a short-term phenomenon caused by scarcity of data.

The *Type I error* of a test T at a process P is the number $1 - P(T_P)$ representing the probability that the test rejects P when it is true. Statistical models usually require tests with low Type I error. In our asymptotic testing setting, it is natural to consider tests that are *free of Type I error*, *i.e.,* ones for which $P(T_P) = 1$ for every P. This makes for sharper statements and interpretation of the results. For tests based on finite observations, small but positive Type I errors are more natural. See the discussion in Section 6.4.

Definition 1. *A stationary distribution P is* testable *if for every stationary $Q \neq P$ there is a test T such that*

1. *T is free of Type I error;*
2. *$Q\big(T_P\big) < 1$.*

We can interpret the definition in two ways. First, think of P as the theory held by an outside observer, such as a theorist or an econometrician, of the data generating process implied by an economic model. The theory P is testable if, when P is incorrect and the correct alternative Q is presented, T can be proven wrong with some probability. In statistical language, testability requires that for any such Q, there is a test T for P against Q that has *some* power. That is, the Type II error of this test is not 100%.

We may alternatively interpret P as the belief of an agent within the model. Equilibrium requires agents to have unquestioning faith in their equilibrium beliefs. Such agents have no use for tests that check whether their beliefs are right or wrong (since they are convinced they know the truth). On the other hand, agents less certain about their knowledge may view the testability of their beliefs as a desirable criterion.

[6] Formally, a finite test must be measurable with respect $\mathcal{H}_n$ for some finite n. The set of finite tests is obviously a subset of the set of all tests. Section 6.4 discusses finite tests.

Our notion of testability is meant to shed light on the implications of equilibrium rather than as a guide for designing practical testing procedures. Actual empirical tests are finite and must therefore tolerate some Type I error. Requiring asymptotic tests to be free of such errors, as we do, is conceptually simpler and, in some cases, results in stronger theorems. In summary, testability retains some essential features of real world empirical tests, but abstracts from others in order to focus on theoretical questions about equilibrium beliefs.

Before turning to testability in economic models, we provide a stylized mathematical example:

Example 1. *For $0 \leq p \leq 1$, let μ_p denote the i.i.d. distribution on coin tosses with probability p of success. We will prove later that $P = \mu_p$ is testable. In fact, the test can be chosen independently of the alternative Q: It checks that every block of k consecutive outcomes appear in the realized sequence at frequency $1/2^k$.*

As an example of a non-testable belief let $P = \frac{1}{2}\mu_{2/3} + \frac{1}{2}\mu_{1/3}$. Under this distribution the observations are i.i.d. but with unknown mean; the agent believes the coin is $\mu_{2/3}$ or $\mu_{1/3}$ with equal probability. This belief is non-testable: any alternative belief $Q = \alpha\mu_{2/3} + (1-\alpha)\mu_{1/3}$ for some $\alpha \in (0,1)$ cannot be separated from P in the sense of Definition 1.

2.3 Example: Asset Pricing

As a first illustration, consider the canonical consumption-based asset pricing model. The primitive is a stochastic process $\{c_n\}$, interpreted as the consumption of a representative agent with a time separable utility $\sum_{n=1}^{\infty} \delta^n u(c_n)$. Here, u is a (differentiable) period utility and $\delta \in [0,1)$ is the discount factor. Assume a finite outcome space rich enough to model the consumption process and any collection of asset returns or other variables of interest.[7]

The marginal rate of substitution between consumption in periods $n+1$

[7] We assumed, for technical convenience, that the outcome space is finite. In asset pricing theory, it is more natural to consider continuous outcome spaces. We can either extend the model to continuous spaces, or assume S to be some appropriately fine grid.

and n is the random variable

$$m_n \equiv \delta \, \frac{u'(c_{n+1})}{u'(c_n)}, \tag{1}$$

also known as the stochastic discount factor.

We assume that m_n is stationary with distribution P.[8] In a consumption-based asset pricing model, the process $\{m_n\}$ determines the equilibrium rates of returns of *all* assets. Specifically, an asset is represented by a stochastic process $\{R_n\}$ (not necessarily stationary) giving the gross rate of return at time n of a dollar invested in that asset at time $n-1$. Equilibrium requires $\{R_n\}$ to satisfy the Euler equation $E_P\big[m_{n+1}\, R_{n+1} \,|\, h^n\big] = 1$, for all positive probability histories h^n.

The usual practice, as in Lucas (1978)-style endowment economies, is to take the consumption process as exogenous, presumably being the outcome of an unmodeled intertemporal optimization.[9] Once consumption and utility are given, equilibrium requires that assets be priced according to the Euler equation.

The model postulates P as the true process governing how $\{m_n\}$ evolves. Our formal notion of testability attempts to give an operational meaning to the statement "P is the true process." Suppose that the modeler, or the representative agent, is confronted by an arbitrageur with an alternative theory $Q \neq P$ of the stock market. Testability captures the intuition that the modeler's theory P must have some observable implications that can be used to "prove P wrong" in the presence of an alternative Q. Since even a wrong theory can give, by accident, more accurate predictions than the true theory, the statement "proving P wrong" must be qualified.[10] The notion of testability accomplishes this by requiring the modeler (or the agent) to

[8] The stationarity of the discount factor can be derived from primitive assumptions about the consumption process and the utility function. Here we assume stationary directly.

[9] See Cochrane (2005) for a textbook exposition. What is considered endogenous and exogenous is, obviously, model-dependent. We follow the common practice in asset pricing theory by assuming that consumption is exogenous, although a richer model would treat it as endogenously determined from more primitive processes.

[10] A theory that predicts a fair coin to land *Heads* in a given toss is more accurate than the true theory (that the coin is fair) about half of the time!

produce, for any competing theory $Q \neq P$, an objective criterion that does not fail P, but that has some power against Q.

2.4 Example: Markov Perfect Equilibria (MPE)

Our next example is a dynamic game with a Markovian state variable. These models are extensively used in industrial organization, political economy, among other fields.

There is a finite set of states, X, and players, I. The set of action profiles is $A = A^1 \times \cdots \times A^I$, where A^i is player i's set of actions. In accordance with our notation, we write $S = X \times A$ to denote the outcome of the game in any given round.

In each period $n = 0, 1, \ldots$, players choose action profile $a_n \in A$ then a state x_n is realized. The distribution of x_n given period $n-1$'s state x_{n-1} and period n's action profile a_n is determined by a known time-invariant transition function $\pi : X \times A \to \Delta(X)$. Note that the timing in our formulation is not standard: we assume that players choose actions before state is realized. This timing matters for our passive learning example in Section 5.1. We assume for simplicity that $\pi(x, a)$ has full support for every outcome (x, a). The initial outcome (x_0, a_0) is chosen according to some distribution p on S. A strategy for player i specifies his mixed action at every stage as a function of past states and action profiles. A strategy is Markovian if players' actions depend only on the current state. Thus, a Markovian strategy for player i is of the form $\sigma^i : X \to \Delta(A)$.

A Markovian strategy profile σ and the transition function π induces a unique stationary Markovian distribution P over $\Delta(H)$ that describes the steady state of the play. The steady state distribution does not depend on the initial choice of (x_0, a_0).

Agent i maximizes discounted expected utility, with period utility $u^i(x, a)$ and discount factor δ. A profile σ is a Markov perfect equilibrium (MPE) if for each player i, σ^i is optimal against the profile of strategies of his opponents, σ^{-i}.[11] The MPE model is widely used in applied work to model, among other things, industry dynamics following the seminal work of Ericson

[11] As usual, all deviations are allowed, and not just Markovian ones.

and Pakes (1995). An attractive feature of this model is that it lends itself naturally to empirical implementation, an issue we shall return to below.[12]

We will show later that an MPE distribution P on S^∞ is testable: given an alternative stationary (nor necessarily Markovian) theory $Q \neq P$ of how the game evolves, we can construct a Type I error free statistical test for P against Q with some power, i.e., such that P is rejected with positive probability under Q. We cannot in general conclude that P will necessarily be rejected with *high* probability. For example, if Q assigns high probability $1-\epsilon$ to P and probability ϵ to some $P' \neq P$, then P is accepted most of the time.

A variant of the above model is one where we assume that an analyst observes the players' actions and some function $f : X \to X'$ of the state.[13] For example, in empirical models it is common to assume that players condition their actions on disturbances that are unobservable by the analyst. Every MPE distribution induces a belief of the analyst over $(A \times X')^{\mathbb{N}}$. The analyst' belief is typically not a Markov process but is still stationary. We will show later that it is also testable.

3 Disagreement and Structural Uncertainty

We introduce three properties a dynamic economic model may have: the possibility of disagreement, absence of structural uncertainty, and the feasibility of empirical estimation using moment conditions. Our main theorem will show that all three properties are equivalent to testability.

3.1 Disagreement

A standard assumption in economic models is that agents share a common prior (Aumann (1976 and 1978)) and, as a result, a common interpretation of information. This assumption has been questioned on a number of grounds. First, that agents may disagree about how to interpret information seems

[12]In empirical models, strategies usually depend also on unobservable disturbances.

[13] In particular, if f is constant then the analyst only see the players actions but not the state of nature.

both intuitive and consistent with the basic axioms of rationality. Second, the absence of disagreement leads to paradoxical theoretical conclusions, such as the no-trade theorems (Milgrom and Stokey (1982)), that are difficult to reconcile with reality. Third, as noted in the introduction, there is large empirical evidence that seems difficult to reconcile with common interpretation of information.[14]

A natural way to introduce disagreement is to assume that agents have different prior beliefs about the underlying uncertainty. Heterogenous prior models are used to generate realistic trade volumes and to account for other asset pricing anomalies. See Hong and Stein (2007) and Pástor and Veronesi (2009) for surveys.

An agent's uncertainty about the true process P is formally represented by a distribution $\mathcal{Q} \in \Delta(\Delta(H))$ on the set of stochastic processes. To any belief $\mathcal{Q}$ corresponds a process $Q \in \Delta(H)$ with identical distribution on sample paths.[15] Since there is no difference between the belief $\mathcal{Q}$ and the process Q on observable phenomena, we use the latter to describe beliefs.

Definition 2. *Two beliefs Q_1, Q_2 are* compatible *if for every event B, $Q_1(B) = 1$ if and only if $Q_2(B) = 1$.*

Compatibility of beliefs is the property known as mutual absolute continuity, a condition that appears in the seminal paper by Blackwell and Dubins (1962) and introduced to the study of learning in repeated games by Kalai and Lehrer (1993). Blackwell and Dubins showed that compatible beliefs "merge," in the sense of generating the same predictions in the limit as data increases.[16]

Compatibility is a common requirement in heterogenous belief models. It has the interpretation that, while beliefs may initially disagree, their dis-

[14] There are motives for trading under common priors (unanticipated liquidity and rebalancing needs). Hong and Stein (2007) suggest that these motives are far too small to explain the 51 trillion dollar volume of trade in equity in 2005, say.

[15] Define $Q(A) \equiv \int_{\Delta(H)} P(A)\mathcal{Q}(\mathrm{d}P)$. Note that since the set of stationary distributions is closed and convex, if a belief $\mathcal{Q}$ is concentrated on stationary distributions, then its reduction Q is also stationary.

[16] See Blackwell and Dubins (1962) or Kalai and Lehrer (1993) for a formal statement of this well-known result.

agreement must eventually vanish. Incompatible beliefs must continue to disagree even in the limit, with infinite data.

Definition 3. *A stationary belief P* precludes disagreement *if for every stationary belief Q compatible with P, we have $Q = P$.*

As the name suggests, such process P leaves no room for disagreement with a compatible belief. The following stylized example illustrate the mathematical structure of the definition:

Example 2. *Consider again the belief $P = \mu_{1/2}$ in Example 1, where outcomes are generated by the independent tossing of a fair coin. This P precludes disagreement: every belief Q that is compatible with P must assign probability 1 to the sequences of observations where each block of length k appears with long-run frequency $1/2^k$. The only stationary belief Q that satisfies this condition is $Q = P$.*

Consider next $P = 1/2\mu_{2/3} + 1/2\mu_{1/3}$ introduced in Example 1. This belief does not preclude disagreement: If $Q = 3/4\mu_{2/3} + 1/4\mu_{1/3}$ then P and Q are compatible, and represent the beliefs of two agents who will disagree forever about how the initial choice of the coin was made.

3.2 Structural Uncertainty

An intuitive distinction can be made between situations where agents understand the structure of their environment and others with "structural uncertainty about long-run fundamentals."

How should such uncertainty be defined? Uncertainty about fundamentals suggests an agent who does not know all the long-run properties of the process and that additional learning about them is possible. As a simple example, suppose that there are just two candidate processes, P_1 and P_2, both of which is i.i.d. as in Examples 1 and 2. For an agent who knows that the true process is P_1, say, no amount of new observations can change his beliefs about the probability of future outcomes. In this case, there is no structural uncertainty, and new information has no predictive value. On the other hand, it is intuitive to think of an agent who believes that the true process is P_1 with probability α and P_2 with probability $1 - \alpha$ as someone who is uncertain about the structure. For such agent, additional observations will potentially change his predictions.

This simple intuition breaks down in more general settings. Suppose, that P_1 and P_2 are, instead, two non-i.i.d. Markov processes. If $\alpha \in (0,1)$, the agent is uncertain about the true process, and information is as valuable as before. The problem is that information is also valuable even if the agent knew that the true process is P_1, say. The reason is that with a Markovian process, the outcome of one period is informative about outcomes of future periods even when the process is known.

Intuitively, "structural uncertainty" suggests lack of knowledge about the long-run properties of the the process rather than outcomes in the 'near' future. In the Markovian example above, for an agent who knows P_1 additional observations are informative about the near future, but have no impact on his beliefs about the probability of outcomes in the distant future.

To make this distinction formal, fix a function $f : S^k \to \mathcal{R}^m$, with finite k, and define $f_n \equiv f(s_{n-k+1}, \ldots, s_n), n = k, k+1, \ldots$. We can think of $f_k, f_{k+1}, \ldots$ as a payoff stream where, in each period n, a payment f_n is made that depends on the realization of the past k outcomes according to the (stationary) formula f.[17] For example, f_k may represent the history-dependent dividend paid by an asset in period k and $f_k, f_{k+1}, \ldots$ is the stream of such payments. Another example is that f takes values in a finite set of actions $\{a_1, \ldots, a_m\}$ in some vector space and represents the actions of an opponent who uses some Markovian strategy.

For a sequence $(x_k, x_{k+1}, \ldots)$ of elements in $\mathcal{R}^m$, define the limiting average:

$$V(x_k, x_{k+1}, \ldots) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=k}^{k+n-1} x_i,$$

whenever the limit exists. We shall abuse notation and refer to $f_k, f_{k+1}, \ldots$ as f. Let $V(f)$ be the random vector whose value $V(f)(h)$ at a history h is the limiting average of $f_k, f_{k+1}, \ldots$ generated by f at h. This random variable is well-defined with probability 1 for any stationary P and function f (Proposition A.2).

[17] To allow history dependence of the payoff stream, payments start in period k.

As an illustration, consider the dynamic game setting of Section 2.4 with I players and strategy profiles A. Let $\Delta(A)$ denote the set of probability distributions on A viewed as a subset of $\mathcal{R}^A$ and let $f : A \to \Delta(A)$ be given by $f(a) = \delta_a$ for every $a \in A$. Then $V(f)(h)$ would simply represent the limiting empirical frequency of actions played along a history h. More complex dynamic features of a history h can be captured with suitable choices of f.

Definition 4. P displays no structural uncertainty *if for every function f and finite history h^{t-1} with $P(h^{t-1}) > 0$*

$$E_{P(\cdot\,|\,h^{t-1})} V(f) = E_P V(f).$$

When f represents payoffs, the limit-of-averages criterion can be replaced with discounted payoffs, provided the discount increases to 1. What is important is that information relevant for short-run predictions is irrelevant for the long run, thus capturing the role of information that changes beliefs about the long-run properties of the process. The limit-of-average criterion simplifies the statement of the property and the results.

3.3 Empirical Identification

In empirical studies, the implications of an economic model are usually summarized by *moment conditions*. Formally,

Definition 5. *A* moment condition *is a bounded continuous function*

$$f : Z \times S^k \to \mathcal{R}^q \tag{2}$$

where k and q are positive integers, and $Z \subset \mathcal{R}^m$ is a compact set.

We say that f identifies P *if there is a unique $\bar{z} = \bar{z}(P, f)$ such that*

$$E_P f(\bar{z}, s_1, \ldots, s_k) = 0.$$

Intuitively, a moment condition represents a finite-horizon feature of P. As an example, assume that $S = \{0, 1\}$, $Z = [0, 1]$, and $k = q = 1$. Let $f : Z \times S \to \mathcal{R}$ be the moment condition $f(z, s) \equiv z - s$. For a stationary distribution P, we have $E_P\, f(z,s) = 0$ precisely at $\bar{z} = E_P\, s_n$, the expectation of s_n under P (which does not depend on n due to stationarity). Clearly, $E_P\, s_n$ does not capture any intertemporal correlations in P. For example, if P_1 is i.i.d. 0.5 while P_2 is the process that puts equal weights on $0, 1, 0, \ldots$ and $1, 0, 1, \ldots$, then these two processes are indistinguishable from the perspective of the moment condition f since $E_{P_1}\, f(z,s) = E_{P_2}\, f(z,s) = 0$ at $\bar{z} = 1/2$.

Moment conditions underlie many statistical techniques used in econometric practice, such as the generalized method of moments, introduced in Hansen (1982), which generalizes many standard techniques. The idea is to estimate the true $\bar{z}$ by the element $\hat{z} \in Z$ that minimizes the empirical estimate of $E_P\, f$. Formally, given f and length of data $n \geq k$, define the empirical average

$$F_n(z,h) = \frac{1}{n-k} \sum_{t=k+1}^{n} f\left(z, s_{t-k}, \ldots, s_t\right),$$

viewed as $q \times 1$ column vector, with transpose denoted $F_n^\top$. The generalized moment estimator is the random variable

$$\hat{z}_n(h) \in \underset{z \in Z}{\operatorname{argmin}}\ F_n(z,h)^\top F_n(z,h).^{18} \tag{3}$$

Definition 6. *A stationary distribution P can be* empirically identified *if $\hat{z}_n \xrightarrow{P} \bar{z}$ for every moment condition that identifies P.*[19]

Hansen shows that consistency obtains for any ergodic P. For this argument, it is enough to consider k=1. But it is not in general possible to prove the converse without using higher k's.

The property that P can be empirically identified from data means that all of the implications of P can be recovered, via moment conditions, from observing the evolution of the process. Alternatively, if P' cannot be empirically identified, then there must be an implication of P', in the sense that $E_{P'} f\left(\bar{z}, s_1, \ldots, s_k\right) = 0$, such that $\bar{z}$ cannot be recovered, even asymptotically as data grows without bound.

[18] More accurately, the generalized method of moments estimator looks at the quadratic form $F(z,h)^\top\, Q\, F(z,h)$, where Q is an appropriately chosen $q \times q$ matrix. Hansen (1982) shows how to select Q optimally. We focus here on the baseline case where Q is the identity for simplicity.

[19] That is, $P\big\{h : |\, \hat{z}_n(h) - \bar{z}\,| > \alpha\big\} \to 0$ for every α as $n \to \infty$.

4 Characterization of Testable Beliefs

4.1 Main Result

Our main result relates the four concepts introduced earlier:

Theorem 1. *For any stationary process P, the following four statements are equivalent:*

1. *P is testable;*
2. *P precludes disagreement;*
3. *P precludes structural uncertainty;*
4. *P can be empirically identified.*

Theorem 1 highlights the tension between desirable properties. Testability is compelling conceptually as a minimal condition for equilibrium, while moments-based empirical methods are important for linking models to data. These reasons explain in part why the processes used in dynamic stochastic equilibrium models usually fall into the testable category. The theorem says that a commitment to testable models makes it difficult to incorporate disagreement and structural uncertainty. As discussed in the Introduction, empirical and theoretical work increasingly point to disagreement and long-run structural uncertainty as important sources of observed anomalies.

4.2 Asset Pricing and MPE Revisited

Many familiar models in the literature correspond to the testable case.

4.2.1 Irreducible and Hidden Markov Models

Our stylized Examples 1 and 2 emphasized i.i.d. processes for their simplicity. More substantive economic models usually build on distributions with complex intertemporal structures. We begin by considering classes of processes that are important in applications.

The following are standard definitions. A transition function $\pi : S \to \Delta(S)$ is *irreducible* if for every $s', s \in S$ there exists some n such that $\pi^n(s'|s) > 0$, where π^n is the n-th power of π. Every irreducible π defines a unique invariant distribution $p \in \Delta(S)$ such that

$$p[t] = \sum_s p[s]\pi(t|s).$$

We say that P is a (stationary, irreducible) Markov process if for every n:

$$P(s_0, s_1, \dots, s_n) = \pi(s_n|s_{n-1}) \cdots \pi(s_1|s_0)\, p(s_0).$$

The definition extends naturally to processes with memory k. Memory k Markov processes often appear in dynamic models in the form of vector autoregressions and can display complex intertemporal correlations.[20]

Another important class of processes is *hidden Markov models*. These are defined in terms of a Markov process on a set $\dot{S}$ of *unobservable* outcomes and a function $f : \dot{S} \to S$ where $f(s)$ is the observation made given s. The distribution induced over H is stationary but not necessarily Markovian of finite memory. Hansen (2007) proposes the hidden Markov model as a framework for uncertainty and learning in asset pricing.

The following fact will be useful in discussing the applications below:

Proposition 4.1. *The conditions of Theorem 1 are satisfied for every irreducible memory k Markov process, and every hidden Markov process where the underlying Markov process is irreducible.*

4.2.2 Asset Pricing

Assets in the consumption-based asset pricing model are priced according to the Euler equation $E_P\left[m_{n+1}\, R_{n+1} \,|\, h^n\right] = 1$, where m_{n+1} is the stochastic discount factor defined in (1) and evolving under a stationary distribution P.

One approach to model asset markets is to assume that the primitives, such as consumption, follow an exogenous stochastic structure of a specific

[20] Autoregressive processes assume a continuous state space while we assume that S is finite. See our comment on this issue in Footnote 7.

form. In the classic Lucas asset pricing model, dividends, and therefore consumption, follow a Markov process. More generally, processes with a memory k Markov structure are testable by Proposition 4.1. The theorem implies that any asset pricing model where the stochastic discount factor process $\{m_n\}$ is testable must be one that precludes disagreement and structural uncertainty.

An alternative approach is to impose no exogenous stochastic structure, but to estimate it from the data. A standard empirical practice to estimate non-linear models is to use the method of generalized method of moments. Suppose we are given a stationary model that is empirically identified for every moment condition, in the sense of Definition 6. We can again apply the theorem to conclude that the proposed model must preclude disagreement and long-run structural uncertainty.

4.2.3 Markov Perfect Equilibrium

Consider next the Markov perfect equilibrium model introduced in Section 2.4. The model generates equilibrium distributions that are irreducible Markov process whenever the transition function has full support. By Proposition 4.1 and Theorem 1, the model precludes disagreement and long-run structural uncertainty. Knowledge of the long run fundamentals (as we define them) is a strong assumption in an industry where firms expect (or fear) disruptive regulatory, technological, or demand changes. In such environments, firms are likely to believe that these changes could fundamentally impact profitability, and disagree on the nature of this impact.

4.3 Stationarity and the Need for Structure

A key restriction we impose is that processes must be stationary. An obvious motivation for stationarity is the role it plays in economic and statistical models. Here we offer more subtle motivation: the absence of stationarity may lead to significant conceptual difficulties in defining what it means to test beliefs.

The next proposition considers what happens if we remove stationarity completely from the definition of testability:

Proposition 4.2. *Extend Definition 1 by removing the stationarity assumption on P and Q. Then a distribution P is testable under this extended definition if and only if it is a Dirac measure that puts unit mass on a single sample path.*

Proof: Dirac's atomic measures on realizations are the extreme point of $\Delta(H)$. Therefore, if P is not Dirac's atomic measure then it can be written as $P = \lambda P' + (1-\lambda)P''$ for some beliefs P, P'' such that $P' \neq P''$ and some $0 < \lambda < 1$. Therefore, every Type I error-free test must have the property that $P'(T_P) = P''(T_P) = 1$. If $Q = \mu P' + (1-\mu)P''$ for some $0 < \mu < 1$ and $\mu \neq \lambda$ then $Q \neq P$ but $Q(T_P) = \mu P'(T_P) + (1-\mu)P''(T_P) = 1$ for every such test. ∎

If no restriction is imposed on the set of processes, then by Proposition 4.2 requiring beliefs to be testable implies that they must be deterministic. This is obviously unsatisfactory in modeling phenomena that are stochastic in nature, such as asset markets and games. Models with non-trivial testable beliefs require some structure restricting the processes allowed.

Stationarity is one way to introduce such structure, but it is obviously not the only one. The question of what is the "right" stochastic structure in economic and statistical models is subtle and lies outside the scope of this paper. Jackson, Kalai, and Smorodinsky (1999) propose that a stochastic structure should be coarse enough to be learnable, but fine enough so no learnable pattern remains unlearned in the long-run. They point out that the class of all distributions is too fine to be learnable, so a more restrictive structure is needed. In our companion paper, Al-Najjar and Shmaya (2012), we show that the class of stationary distributions is learnable, providing some foundation for their use in economic models. We refer the reader to these papers for elaboration on these points.

5 Learning and Equilibrium

As noted in Section 4.2, many economic models used in practice fall into the testable category of our main theorem. An implication of that theorem is that in stationary non-testable environment there is always something to learn about the long-run fundamentals. As agents learn from observing the process, their beliefs about these fundamentals change. But changing beliefs can lead to non-stationary decisions, and this undermines the stationarity of the process itself.

This informal intuition suggests a tension between stationarity and optimization. To better understand this tension, it will be convenient to assume, as in the MPE model of Section 2.4, that S can be written as a product $S = X \times A$ of action profiles A and exogenous variables X that are not affected by these actions.

The evolution of the system is described by a stochastic process P on $H = (X \times A)^{\infty}$. Let P_X and P_A denote the marginals of P on X^{∞} and A^{∞}. We will only require P_X to be stationary, but not necessarily the entire process P. This weaker assumption makes it possible to apply Theorem 1 restricted to beliefs about the exogenous variables and thus explore the implications of testability even when agents' behavior is non-stationary.

5.1 Passive Learning

Consider a variation on the MPE setting of Section 2.4 where agents are uncertain about the transition π. Pakes and Ericson (1998) study a model along these lines with the goal of empirically testing for the implications of agents' passive learning on industry dynamics.

We follow closely that section except for two departures. First, agents may be uncertain about the transition. Formally, agents have beliefs $\mu_i, i \in I$, with a finite and common support $\Pi = \{\pi_1, \ldots, \pi_L\}, L \geq 1$, representing their uncertainty about the transition. Second, agents learn about the process. Learning, however, is passive in the sense that agent's actions do not influence the evolution of the exogenous process. This rules out agents' active experimentation and keeps the theoretical and empirical analysis tractable. Formally, we require that each $\pi_l \in \Pi$ takes the form $\pi_l(x_n)$, and so depends only on x_n.

Let $P_X(\pi_l)$ be the stationary distribution on X^{∞} induced by the transitions $\pi_l \in \Pi$ as described in Section 4.2.1. Agent i believes that the exogenous variables evolve according to the stationary distribution $P_X(\mu_i) =$

$\mu_i(\pi_1)P_X(\pi_1) + \cdots + \mu_i(\pi_l)P_X(\pi_l)$ which no agent can influence. We distinguish three cases:

Case 1: *$L = 1$, so agents know the true transition π.*

Case 2: *$L > 1$ and agents share a common prior ($\mu_i = \mu$, for all i) with support $\{\pi_1, \ldots, \pi_L\}$.*

Case 3: *$L > 1$ and agents have different priors ($\mu_i \neq \mu_j$, for some i, j) with common support $\{\pi_1, \ldots, \pi_L\}$.*

Case 1 is a special case of the MPE introduced earlier.[21] The marginal process P_X is Markovian, testable and displays no structural uncertainty. The same holds for the full equilibrium process P under an MPE.

Turning to Case 2, we model decisions using Bayesian-Nash equilibrium as solution concept. Fix one such equilibrium and let P denote the distribution it generates. In any such equilibrium, the marginal process P_X on exogenous variables is the stationary process $P_X(\mu)$. Applying Theorem 1 to P_X, it is easy to verify that it is non-testable, for example since agents are uncertain about the long-run fundamentals (*i.e.,* which π is the true process).

The fact that the marginal process P_X is stationary and non-testable makes it "likely" that the equilibrium process P as a whole is not stationary. Intuitively, when agents learn about the long-run fundamentals (the true π), their beliefs and decisions will change in a non-stationary fashion. The next theorem formalizes this intuition:

Proposition 5.1. *Suppose that agents have a common prior μ with support $\{\pi_1, \ldots, \pi_L\}, L > 1$. Then there exists action sets and utilities, A^i, u^i, for $i = 1, \ldots, I$, such as the resulting game admits no Markov equilibrium profile.*

When players know the Markov transition, their beliefs about the next state depends only on their current state. Under uncertainty, however, the current state is no longer sufficient to determine agents' belief about the next outcome, and behavior may depend on the entire history of past observations. Early observations can have a persistent effect on agents' decisions.(Making

[21] Special because we require that agents's actions do not influence the state.

this argument formal requires that changes in beliefs have an impact on actions. This is why the theorem is stated in terms of *some* specification of payoffs.) Pakes and Ericson (1998) exploit this fact to construct a statistical test of whether observed industry dynamics is consistent with firms learning about their environment.

We conclude by noting a connection between testability and disagreement. In Case 2, the process P_X is non-testable yet agents agree on it. As argued earlier, justifying agreement that μ is the true process is difficult when there can be no statistical test that can, even in principle and with infinite data, disprove μ against an alternative beliefs $\mu' \neq \mu$. If one views testability as a (necessary) condition for agreement, then Case 2 seems artificial, and Case 3 where agents disagree is the more compelling alternative.

5.2 Bayesian Nash Equilibrium

Next we consider learning in a repeated game with incomplete information using the setup of Kalai and Lehrer (1993). For simplicity, assume there are just two players, each with a finite number of actions. A type of a player is his payoff matrix. Assume that there are m possible types for each player, and denote by $T^i = \{D_1^i, \ldots, D_m^i\}$ the set of types of player i. A type profile is realized from common prior distribution μ on $T^1 \times T^2$. After types are realized, each player observes his own type. The standard solution concept for this class of games is that of a Bayesian Nash equilibrium. Kalai and Lehrer (1993, Theorem 2.1) show, roughly, that conditional on (D^1, D^2) being the true type profile, the play of the game will eventually be close to that of a Nash equilibrium of the true game.

We consider an outside analyst who observes the actions of the players but not the realized types. We illustrate the relationship between testability, stationarity, and learning in three examples chosen for their tractability. In Example 3 types are completely correlated, so players have no uncertainty.

Example 3. *Assume that μ is uniform on $\{(D_j^1, D_j^2)\}_{j=1,\ldots,m}$. Then players know the type profile but an observer has uncertainty. Consider the Bayes Nash equilibrium according to which, when the type profile is (D_j^1, D_j^2) players repeatedly play some Nash equilibrium of that game.*

The process induced by the equilibrium profile in this example is stationary (in fact, exchangeable) but will typically not be testable. This is because there is uncertainty about the long-run distribution of play—unless the support of μ is concentrated on type profiles with the same Nash Equilibrium.

In our next example, players are uncertain about the game so they learn something about their opponents' type as they observe the evolution of play. However, the payoffs are such that learning does not affect the players behavior. Say that a type D of player i is *committed to action* a if playing a each period is a dominant strategy for player i in the repeated game when his payoff matrix is D.

Example 4. *Assume that μ has full support over $T^1 \times T^2$ and that all types are committed (to different actions). In a Bayes-Nash equilibrium, each player always plays his commitment strategy.*

The equilibrium profile in Example 4 is, trivially, stationary, but, when some types of a player are committed to different actions, it is not testable.

In our final example, players have uncertainty and the example is such that learning must affect behavior. In this case, no Bayesian Nash equilibrium induces a stationary play.

Example 5. *Assume that:*

1. *μ has full support over $T^1 \times T^2$.*
2. *There exists types $D, D' \in T^1$ of player 1 which are committed to different actions a, a'.*
3. *There exists type D'' of player 2 such that the set $B(a)$ of best responses of player 2 under D'' to a and the $B(a')$ of best responses of player 2 under D'' to a' are disjoint.*

We have the following result:

Proposition 5.2. *A Bayes Nash equilibrium of the game described in Example 5 cannot be stationary.*

The proof is in the Appendix. Roughly speaking, this follows from the fact that in every equilibrium in the game with payoff matrices (D, D''), the players' action profile is in $\{a\} \times B(a)$, and that in every equilibrium in the game with payoff matrices (D', D''), the players action profile is in $\{a'\} \times B(a')$. Kalai and Lehrer's theorem implies that players learn to play as in equilibrium of the game with the realized types, so that there is a positive probability that from some period onwards the action profile played is always in $\{a\} \times B(a)$ and a positive probability that from some period onwards the action profile played is always in $\{a'\} \times B(a')$. Because of stationarity, this should be the case already for the actions at period 0, and because these actions are independent given the types, it follows that at day 0 there is a positive probability that the action will be in $\{a\} \times B(a')$ which is a contradiction.

5.3 Testing Beliefs vs. Testing Behavior

Disagreement and structural uncertainty can have observable implications on agents' behavior and market outcomes. This is not in contradiction with our main theorem. That theorem concerns the difficulties associated with objectively testing whether beliefs are right and wrong. The fact that it is difficult to test whether an agent "knows the true" model makes it more plausible that his choices will reflect model uncertainty.

To make this more formal, apply our main theorem to the process P_X governing the evolution of the exogenous variables as discussed earlier. If P_X is not testable, then the theorem implies that agents have structural uncertainty. While it is not possible to devise an objective test to determine whether beliefs about P_X are correct, agents' uncertainty can have empirically measurable implications on their behavior.

The details will obviously depend on the model. Pakes and Ericson (1998) develop econometric tests in a Markovian model of industry dynamics to detect firms' learning. Firms' beliefs are not directly observable. On the other hand, subjective uncertainty has the implication that early observations have a persistent impact on posterior beliefs and, therefore, on future choices. Evidence for learning may be found by looking for non-stationary behavior, such as a correlation between early observations and the long-run evolution of the model.

Other example can be found in the context of asset pricing. Lewellen and Shanken (2002) and Brav and Heaton (2002) consider the impact of investors' learning on asset pricing. Using stylized settings where dividends are i.i.d. with unknown parameters, they show that learning can cause returns to appear predictable and excessively volatile even though prices react efficiently to information. These authors note that learning may confound empirical testing of asset pricing models.

6 Discussion and Related Literature

6.1 Testing Strategic Experts

A recent literature examines the problem of testing an expert's knowledge of the underlying process. An uninformed strategic expert may attempt to manipulate the test with a randomized forecast. A number of papers, including Sandroni (2003), Olszewski and Sandroni (2008), Shmaya (2008), provide general conditions under which a strategic expert can manipulate any test. Results establishing the existence of non-manipulable tests appear in Dekel and Feinberg (2006), Shmaya (2008), Olszewski and Sandroni (2009), Al-Najjar, Sandroni, Smorodinsky, and Weinstein (2010). See Olszewski (2012) for a survey.

In our paper, agents do not strategically randomize their forecasts to manipulate the test. Rather, they hold subjective beliefs about their environment, and choose optimally given these beliefs. Our focus is on the implications of requiring that erroneous beliefs can be rejected by data under the true model.

6.2 Bayesian Learning in Games

Section 5.2 illustrates the connection with the literature on learning in games. The concepts of testability, long-run uncertainty, and stationarity are related to Nachbar (2005)'s results for learning in games. In his motivating example (pp. 459-460) two players play an infinitely repeated game, where the stage game has just two actions A, B for each player. Player $i = 1, 2$, believes player $j \neq i$ follows an i.i.d. strategy where A is played with probability q_i each period. Player i's prior belief about the strategy of the opponent is a probability distribution μ_i over $[0, 1]$. Nachbar starting point is the observation that even ϵ-optimization will lead to the unraveling of the assumed stationary equilibrium.

In our terminology, each player faces a stationary environment described by his belief about the strategy of his opponent. In this example, there is long-run uncertainty if and only if μ does not put unit mass on one value of q_i. Proposition 5.1 shows that when players update their beliefs, their behavior will not in general be i.i.d., and is therefore not consistent with the players's priors about their opponents' strategies.

Nachbar obtains a general result for infinitely repeated game strategies where the assumption of stationarity is not natural. In this paper, we take the perspective of a single decision maker and our arguments rely heavily on stationarity. Our concepts of testability and empirical identification also have no counterpart in Nachbar's work.

6.3 Self-confirming Equilibrium

The process P in this paper describes the evolution of all variables in the model, including agents' decisions. Beliefs are in equilibrium if their predictions about future outcomes coincide with P. In our analysis, the process P puts no restrictions on agents' predictions at counter-factual events, *i.e.,* events that have zero probability under P. This is important in games, where players' decisions depend on his beliefs about what their opponents will do at all events, including those not observed in equilibrium. Although the process P describes all what an outside observer can hope to see, it is insufficient to explain what players do. This distinction is captured by the notion of self-confirming equilibrium introduced by Fudenberg and Levine (1993a) where players are assumed to know the true distribution on observed outcomes, but may have incorrect off-path conjectures.

Fudenberg and Levine (1993b) motivate the concept of self-confirming equilibrium as the result of a steady-state learning process. Related to our discussion of disagreement, Dekel, Fudenberg, and Levine (2004) argue that this steady-state interpretation of self-confirming equilibrium is difficult to reconcile with heterogenous beliefs about the state of nature.

6.4 Finite Horizon Testing

The properties in Theorem 1 are formulated in terms of infinite outcome sequences. The definitions have natural analogues for finite horizons. Take, for instance, the concept of testability. Call a test T *finite* if there exists an integer n such that T depends only on the outcome of the first n periods.[22] Tests used in statistical practice are obviously finite. It is easy to show the following:

Claim. *If P is testable then for every stationary Q and every $\alpha > 0$ there exists a finite horizon test T such that:*

1. *The Type I error of T is smaller than α.*
2. $Q(T(P)) < 1 - \alpha$

Formulating our result in a finite horizon setting would require close attention to the amount of data available to the test, an issue that does not arise in asymptotic tests. For finite tests, the horizon needed will in general depend on the complexity of the intertemporal structure of the processes being tested. To illustrate, consider the following simple example:

Example 6. *Let $S = \{0, 1\}$ and let P be the steady state distribution of a Markov process with transition*

- $\pi(0|0) = \pi(1|1) = 1 - \epsilon$ *and*
- $\pi(1|0) = \pi(0|1) = \epsilon$

[22] Formally, T is measurable with respect to $\mathcal{H}_n$.

for a small $\epsilon < \frac{1}{n}$*. Since* P *is an irreducible Markov chain, it follows from Proposition 4.1 above that* P *is testable. Let* Q *be the process that puts unit mass on the sequence* $(0, 0, \dots)$*. Then for every test* T *of horizon* n *and Type I error smaller than* $\alpha = 0.10$*, we have* $Q(T(P)) = 1$*.*

The problem in this example is that the process P is highly persistent. If we are given a very short horizon, this persistence can easily confound P with the constant process Q.

Testability with such short horizon has little bite. This should not be surprising: no statistical technique based on such limited data can distinguish the process P from Q. Relatedly, the GMM estimator of the one dimensional marginal of P will fail to converge to its correct value over such small horizon.

In summary, finite data is a limiting factor to statistical methods in general, and not specifically our model. To apply the notion of testability in finite horizon, either the class processes has to be restricted, or the horizon extended (or both). In our asymptotic testing framework, we impose no restrictions beyond stationarity, but require the horizon to be infinite. Alternatively, we could limit n to be finite, but then we must restrict the class of processes to ensure fast enough rate of convergence to make statistical testing feasible.[23] We find the asymptotic approach to give a clearer picture of the implications of testability.

[23] A natural way to proceed is to require the process to be mixing with a uniform mixing rate. This, indeed is the common assumption made in practice. Vector autoregressive processes, for example, belong to this class. See also our discussion of Pakes and Ericson (1998).

A Proofs

A.1 Mathematical Preliminaries

A function $f : H \to \mathbb{R}$ is *finitely-based* if there is an integer k such that for any history h, f depends only on the first k coordinates of h (that is, f is $\mathcal{H}^k$-measurable). Abusing notation, we express finitely-based functions in the form $f : S^k \to \mathbb{R}$. We use the following lemma:

Lemma A.1. *For any pair of distributions $P, Q \in \Delta(H), P \neq Q$ there exists a finitely-based function f such that $E_P f \neq E_Q f$.*

Given a stationary P and any (Borel)function $g : H \to \mathcal{R}$, the *ergodic theorem* states that the limit

$$\tilde{g} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g\left(s_i, s_{i+1}, \dots\right) \tag{4}$$

exists P-a.s. and that $E_P \tilde{g} = E_P g$. If the limit $\tilde{g}$ is constant (i.e., $\tilde{g} = E_P g$ almost surely) for every g then P is called *ergodic*. It is sufficient for ergodicity that the limit be constant for every finitely-based function g.[24]

The following proposition states the implication of the ergodic theorem for finitely based functions. In the terminology of Section 3, the (random) limiting average payoff of a finitely-based f is well-defined:

Proposition A.2. *Let P be stationary and $f : S^k \to \mathbb{R}$ be finitely-based. Then*

$$V(f)(h) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=k}^{k+n-1} f(s_{i-k}, \dots, s_{i-1})$$

exists for P-almost every history $h = (s_0, s_1, \dots)$, and $E_P V(f) = E_P f$. Moreover, if P is ergodic then $V(f(h))$ is P-almost surely the constant $E_P f$.

The set of all stationary distributions P over H is convex and weak*-compact. We denote by $\mathcal{E} \subset \mathcal{P}$ the set of all ergodic distributions over H.

[24] For a textbook exposition of the parts of ergodic theory relevant to this paper, see Gray (2009).

The *ergodic decomposition theorem* states that there exists a Borel function $\varepsilon : H \to \mathcal{E}$ such that $\mu\left(\varepsilon^{-1}(\mu)\right) = 1$ for every ergodic μ and such that

$$P(E) = \int \mu(E)\bar{P}(\mathrm{d}\mu) \tag{5}$$

for every event $E \subseteq H$ where $\bar{P} \in \Delta(\mathcal{E})$ is the push-forward of P under ε (i.e., $\bar{P}(E) = P\left(\varepsilon^{-1}(E)\right)$ for every Borel subset E of $\mathcal{E}$). It follows from the ergodic decomposition theorem that the extreme points of this set are the ergodic distributions. That is, P is ergodic if and only if there exists no stationary distributions R', R'' such that $R' \neq R''$ and $P = \lambda R' + (1-\lambda)R''$ for some $0 < \lambda < 1$.

A.2 Proof of Theorem 1

We will show that the properties in the theorem are equivalent to P being ergodic.

A.2.1 P is ergodic $\implies$ P displays no structural uncertainty.

Assume that P is ergodic and let $f : S^k \to \mathbb{R}$ be bounded. By Proposition A.2 $V(f)$ is almost surely the constant $E_P V(f)$. It follows that $E_P(V(f)|E) = E_P V(f)$ for every event E with $P(E) > 0$. In particular, $E_{P(\cdot|h^{t-1})}V(f) = E_P V(f)$ for every finite history h^{t-1} with $P(h^{t-1}) > 0$.

A.2.2 P displays no structural uncertainty $\implies$ P is testable.

Assume that P displays no structural uncertainty. Let $f : S^k \to \mathcal{R}$ be bounded. From the martingale convergence theorem it follows that

$$E_{P(\cdot|h^{t-1})}V(f) \xrightarrow[t\to\infty]{} V(f)(h)$$

for P-almost every h. If P displays no structural uncertainty then

$$E_{P(\cdot|h^{t-1})}V(f) = E_P V(f) = E_P f,$$

where the second equality follows from Proposition A.2. Therefore $V(f)(h) = E_P f$ for P-almost every h.

Let T_f be a test such that $T_f(P) = \{h|V(f)(h) = E_P f\}$ and $T_f(Q) = \Omega$ for every stationary $Q \neq P$. By the argument above it follows that T is Type I error free. Moreover if $Q(T_f(P)) = 1$ for some stationary distribution Q then $V(f)(h) = E_P f$ for Q-almost every h, and therefore $E_Q f = E_Q V(f) = E_P f$, where the first equality follows from Proposition A.2.

If $Q \neq P$ is any stationary distribution, then by Proposition A.1 there exists a bounded function $f : S^k \to \mathcal{R}$ such that $E_P f \neq E_Q f$. Then it follows that T_f is Type I error free and that $Q(T_f(P)) < 1$, as desired.

A.2.3 P is testable $\implies$ P precludes disagreement.

Let P be testable and let Q be any stationary belief such that $Q \neq P$. We claim that Q is not compatible with P. Indeed, let T be a Type I error free test such that $Q(T(P)) < 1$. Since T is Type I error free, we get that $P(T(P)) = 1$. Therefore P and Q are not compatible, as desired.

A.2.4 P precludes disagreement $\implies$ P is ergodic.

Assume by contradiction that P is not ergodic so that $P = \lambda R' + (1-\lambda)R''$ for some stationary beliefs $R' \neq R''$ and $0 < \lambda < 1$. Since $P(B) = \lambda R'(B) + (1-\lambda)R''(B)$ for every event B it follows that $P(B) = 1$ if and only if $R'(B) = R''(B) = 1$. Let $0 < \mu < 1$ such that $\mu \neq \lambda$ and let $Q = \mu R + (1-\mu)R''$. Then $Q(B) = 1$ if and only if $R'(B) = R''(B) = 1$ and so $Q(B) = 1$ if and only if $P(B) = 1$. Thus P and Q are compatible, and since $R \neq R''$ and $\lambda \neq \mu$ it follows that $P \neq Q$, contradicting the assumption that P precludes disagreement.

A.2.5 P is ergodic $\implies$ P can be empirically identified

This part of the theorem is a slight extension of Hansen (1982)'s result on the consistency of the GMM estimator. First we restate his result in our notation:

Proposition A.3. *Assume that some ergodic distribution P is identified by a moment condition (2) with $k = 1$. Let $\hat{z}_n$ be the GMM estimator given by (3). Then $\hat{z}_n \xrightarrow{P} \bar{z}$.*

We need to adapt this result to arbitrarily finite k. Let P be ergodic and let f be a moment condition given by (2):

$$f : Z \times S^k \to \mathcal{R}^q.$$

Consider the new set of outcomes $\bar{S} = S^k$. Define the probability distribution $\bar{P}$ over $\bar{H}$ as the push-forward of P under the map

$$(s_0, s_1, s_2, \dots) \in H \mapsto (\bar{s}_0, \bar{s}_1, \dots)$$

where

$$\bar{s}_n = (s_n, s_{n+1}, \dots, s_{n+k-1}).$$

Then $\bar{P}$ is ergodic, the moment conditions f on P translates to a moment condition on $\bar{P}$ with $k = 1$, and the GMM estimator for $\bar{P}$ under $\bar{f}$ is the same as for P under f. Thus, by Proposition A.3 it follows that the GMM estimator is consistent.

A.2.6 P can be empirically identified $\implies$ P is ergodic

Let $P \in \mathcal{P}$ be stationary and empirically identified. Fix an integer k and let $Z = \Delta(S^k)$. Define the moment condition $f : Z \times S^k \to \mathbb{R}^{S^k}$ be given by

$$f(z, h)[h'] = z[h'] - \delta_{h,h'}$$

for every $z \in \Delta(S^k)$ and $h' \in S^k$, where $\delta_{h,h'} = \begin{cases} 1, & \text{if } h = h' \\ 0, & \text{otherwise.} \end{cases}$.

If μ is a stationary distribution then f identifies μ since the unique $\bar{z}$ that satisfies $E_\mu f(\bar{z}, h) = 0$ is given by $\bar{z} = \Pi^k(\mu)$ where $\Pi^k : \mathcal{P} \to \Delta(S^k)$ is such that $\Pi^k(\mu)$ is the distribution on k-tuples induced by μ for every stationary distribution μ.

It follows from A.2.5 that for every ergodic $\mu \in \mathcal{E}$

$$\hat{z}_n \xrightarrow{\mu} \Pi^k(\mu) \tag{6}$$

under μ, where z_n is the GMM estimator.

We claim that P,

$$\hat{z}_n \xrightarrow{P} \Pi^k \circ \varepsilon. \tag{7}$$

where $\varepsilon : H \to \mathcal{E}$ is the ergodic decomposition function introduced earlier satisfying 5. Indeed, for every $\alpha > 0$

$$P\left(\left\{\omega \in H \middle| \, |\, \hat{z}_n(\omega) - \Pi^k \circ \varepsilon(\omega)\, | > \alpha\right\}\right) = \int \bar{P}(\mathrm{d}\mu)\mu\left(\left\{\omega \in H \middle| \, |\, \hat{z}_n(\omega) - \Pi^k(\mu)\, | > \alpha\right\}\right) \xrightarrow[n\to\infty]{} 0$$

where the equality follows from the definition of ε and the limit from the bounded convergence theorem and (6).

On the other hand,

$$\hat{z}_n \xrightarrow{P} \Pi^k(P). \tag{8}$$

since P can be empirically identified. It follows from (7) and (8) that $\Pi^k \circ \varepsilon = \Pi^k(P)$, P-a.s. Since this is true for every k it follows that ε is constant P-a.s., i.e., that P is ergodic.

A.3 Proof of Proposition 4.1

We proved that the conditions of Theorem 1 are satisfied for every ergodic process. Every irreducible Markov process is ergodic Durrett (2010, Example 7.1.7). Also every function of an ergodic process is ergodic Durrett (2010, Theorem 7.1.3). In particular, every hidden Markov process is irreducible when the underlying Markov process.

A.4 Proof of Proposition 5.1

Since $\pi_1, \ldots, \pi_L$ are not identical, there exists some $s^* \in S$ such that $\pi_1(s^*) \neq \pi_2(s^*)$. Therefore there exists some $u \in \mathbb{R}^S$ and $\alpha \in \mathbb{R}$ such that

$$\pi_1(s^*) \cdot u < \alpha < \pi_2(s^*) \cdot u$$

where $\cdot$ is the inner product in $\mathbb{R}^S$.

Assume that at every period n each player has two actions: a safe action that gives payoff α regardless of the next state, and a risky action that gives payoff $u[s_{n+1}]$. Thus, a player's payoff depends only on her own action and not on the opponent's actions. In this game, every equilibrium strategy

σ is such that after every history h the player chooses the safe action if the conditional expectation of $u[s_{n+1}]$ is smaller than α and the risky action if the conditional expectation of $u[s_{n+1}]$ is greater than α. Since after sufficiently long periods the player learns the ergodic transition π it follows that there are arbitrary long histories $(s_0, s_1, \ldots, s_n)$ with $s_n = s^*$ after which the player must play the safe action and arbitrary long histories after which the player must play the risky action. Thus, the equilibrium cannot be Markovian.

A.5 Proof of Proposition 5.2

The proof uses the following lemma, which follows immediately from the ergodic theorem.

Lemma A.4. *Let S be a finite set of outcomes and let P be a stationary distribution over H. Then P-almost every realization $\omega = (s_0, s_1, \ldots) \in H$ has the property that s_0 appears infinitely often in ω.*

Proof: By the ergodic decomposition we can assume w.l.o.g. that P is ergodic. Let $p \in \Delta(S)$ be the marginal one-period distribution of P, i.e. $p[s] = P(\{\omega = (s_0, s_1, \ldots) | s_0 = s\})$ for every $s \in S$. Then it follows from the ergodic theorem that, for P-almost every realization ω, every $s \in S$ appears in ω with frequency $p[s]$. In particular, if $p[s] > 0$ then s appears infinitely often in ω. This implies that s_0 appears infinitely often in ω for almost every realization $\omega = (s_0, s_1, \ldots) \in H$, as desired. ∎

To prove the proposition, let $\epsilon > 0$ be small enough such that:

1. The only subgame-perfect ϵ-equilibrium in the repeated game with incomplete information (D, D'') is such that player 1 always play a and player 2 always play an action from $B(a)$.

2. The only subgame-perfect ϵ-equilibrium in the repeated game with incomplete information (D', D'') is such that player 1 always plays a' and player 2 always play an action from $B(a')$.

Let $\eta > 0$ sufficiently small. It follows from Kalai Lehrer's Theorem that in every equilibrium of the game with incomplete information, if the types are (D, D'') there is a time from which, with probability at least $1-\eta$, player 2 always plays an action from $B(a)$. It follows from Lemma A.4 that according to the equilibrium profile, conditioned on the types being, (D, D''), player 2 plays an action in $B(a)$ with probability at least $1-\eta$ (since almost every realization of the play path on which he doesn't play an action in $B(a)$ cannot have the property that he plays an action from $B(a)$ from some point onward.) This implies that according to his equilibrium strategy, player 2 must play an action from $B(a)$ with probability $1-\eta$ at day 0 if his type is D''. Similarly, player 2 must play an action from $B(a')$ with probability $1-\eta$ at day 0 if his type is D''. Since $B(a) \cap B(a'') = \emptyset$ we get a contradiction.

References

Al-Najjar, N. I., A. Sandroni, R. Smorodinsky, and J. Weinstein (2010): "Testing Theories with Learnable and Predictive Representations," *Journal of Economic Theory*, 145, 2203–17.

Al-Najjar, N. I., and E. Shmaya (2012): "Learning and Long-run Fundamentals in Stationary Environments," Northwestern University.

Aumann, R. (1976): "Agreeing to Disagree," *The annals of statistics*, 4(6), 1236–1239.

Blackwell, D., and L. Dubins (1962): "Merging of Opinions with Increasing Information," *Ann. Math. Statist.*, 33, 882–886.

Brav, A., and J. Heaton (2002): "Competing Theories of Financial Anomalies," *Review of Financial Studies*, 15(2), 575.

Cochrane, J. (2005): *Asset pricing*. Princeton university press.

Dekel, E., and Y. Feinberg (2006): "Non-Bayesian Testing of an Expert," *Review of Economic Studies*, 73, 893–906.

Dekel, E., D. Fudenberg, and D. Levine (2004): "Learning to Play Bayesian Games," *Games and Economic Behavior*, 46(2), 282–303.

Durrett, R. (2010): *Probability: Theory and Examples*, vol. 3. Cambridge University Press.

Ericson, R., and A. Pakes (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *The Review of Economic Studies*, 62, 53–82.

Fudenberg, D., and D. K. Levine (1993a): "Self-Confirming Equilibrium," *Econometrica*, 61(3), 523–545.

——— (1993b): "Steady state learning and Nash equilibrium," *Econometrica*, 61(3), 547–573.

Gray, R. (2009): *Probability, Random Processes, and Ergodic Properties.* Springer Verlag.

Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, pp. 1029–1054.

——— (2007): "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *American Economic Review*, 97(2), 1–30.

Hong, H., and J. Stein (2007): "Disagreement and the stock market," *The Journal of Economic Perspectives*, 21(2), 109–128.

Jackson, M. O., E. Kalai, and R. Smorodinsky (1999): "Bayesian Representation of Stochastic Processes under Learning: de Finetti Revisited," *Econometrica*, 67, 875–893.

Kalai, E., and E. Lehrer (1993): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61, 1019–1045.

Lewellen, J., and J. Shanken (2002): "Learning, Asset-Pricing Tests, and Market Efficiency," *The Journal of finance*, 57(3), 1113–1145.

Lucas, R. (1978): "Asset Prices in an Exchange Economy," *Econometrica*, pp. 1429–1445.

Milgrom, P., and N. Stokey (1982): "Information, trade and common knowledge," *Journal of Economic Theory*, 26(1), 17–27.

Nachbar, J. H. (2005): "Beliefs in repeated games," *Econometrica*, 73(2), 459–480.

Olszewski, W. (2012): "Calibration and Expert Testing," in *Handbook of Game Theory*, vol. IV. North-Holland.

Olszewski, W., and A. Sandroni (2008): "Manipulability of Future-Independent Tests," *Econometrica*, 76, 1437–66.

——— (2009): "A Nonmanipulable Test," *Annals of Statistics*, 37, 1013–39.

Pakes, A., and R. Ericson (1998): "Empirical Implications of Alternative Models of Firm Dynamics," *Journal of Economic Theory*, 79, 1–45.

Pástor, L., and P. Veronesi (2009): "Learning in Financial Markets," *Annu. Rev. Financ. Econ.*, 1(1), 361–381.

Ross, S. (1989): "Intertemporal asset pricing," *Theory of Valuation: Frontiers in Modern Financial Theory*, 1, 85–96.

Sandroni, A. (2003): "The Reproducible Properties of Correct Forecasts," *Internat. J. Game Theory*, 32(1), 151–159.

Sargent, T. (2008): "Rational Expectations," in *The New Palgrave Dictionary of Economics, Second Edition*, ed. by D. S.,and L. Blume. Palgrave Macmillan.

Shmaya, E. (2008): "Many Inspections are Manipulable," *Theoretical Economics*, 3, 367–82.

Weitzman, M. (2007): "Subjective Expectations and Asset-Return Puzzles," *The American Economic Review*, 97(4), 1102–1130.