Manipulability of Future-Independent Tests*

Wojciech Olszewski† and Alvaro Sandroni‡

April 18, 2008

Abstract

The difficulties in properly anticipating key economic variables may encourage decision makers to rely on experts' forecasts. Professional forecasters, however, may not be reliable and so their forecasts must be empirically tested. This may induce experts to forecast strategically in order to pass the test.

A test can be ignorantly passed if a false expert, with no knowledge of the data generating process, can pass the test. Many tests that are unlikely to reject correct forecasts can be ignorantly passed. Tests that cannot be ignorantly passed do exist, but these tests must make use of predictions contingent on data not yet observed at the time the forecasts are rejected. Such tests cannot be run if forecasters report only the probability of the next period's events on the basis of the actually observed data. This result shows that it is difficult to dismiss false, but strategic, experts who know how theories are tested. This result also shows an important role that can be played by predictions contingent on data not yet observed.

*We are grateful to the editor and three anonymous referees for helpful comments, discussion, and suggestions; in particular, one of the referees simplified the original proof of our main result (proposition 1).

†Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston IL 60208

‡Department of Economics, University of Pennsylvania, 3718 Locust Walk, Phildadelphia PA 19104 and Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston IL 60208

1. Introduction

Expectations of future events have long been recognized as a significant factor in economic activity (see Pigou (1927)). It is, however, difficult to accurately anticipate key economic variables. This difficulty may produce a demand for experts' forecasts. If informed, an expert knows the relevant odds and can reveal them to decision makers (i.e., the decision makers' uncertainty is replaced with common risk). If uninformed, the expert does not know the relevant odds and may mislead decision makers. Hence, experts' forecasts must be tested.

A test determines observable histories that are (jointly) unlikely under the null hypothesis that the forecasts are correct. These data sequences are deemed inconsistent with the forecasts and, if observed, lead to their rejection. This method is unproblematic if the forecasts are reported honestly. However, assume that the expert (henceforth called Bob) may strategically misrepresent what he knows. A tester (named Alice) tests Bob's forecasts. She anticipates that Bob, if uninformed, may try to manipulate her test (i.e., to produce forecasts that will not be rejected, regardless of how the data turn out in the future).

The purpose of running a manipulable test is limited if Bob knows how he will be tested. Even in the extreme case that Bob is completely uninformed about the data-generating process, the test will support the hypothesis that he is informed, no matter what data are observed.

A standard calibration test requires the empirical frequencies of an outcome (say, 1) to be close to p in the periods that 1 was forecasted with probability near p. Foster and Vohra (1998) show that the calibration test can be manipulated. So it is possible to produce forecasts that in the future will prove to be calibrated. In contrast, Dekel and Feinberg (2006) and Olszewski and Sandroni (2007b) show nonmanipulable tests that are unlikely to reject the data-generating process.[1]

Unlike the calibration test, the tests proposed by Dekel and Feinberg (2006) and Olszewski and Sandroni (2007b) require Bob to deliver, at period zero, an entire theory of the stochastic process. By definition, a theory must tell Alice, from the outset, all forecasts contingent on any possible data set. Typically, a forecaster does not announce an entire theory but instead publicizes only a forecast in each period, according to the observed data. So a natural issue is whether there exist nonmanipulable tests that do not require an entire theory, but rather use only the

[1] The existence of such a test was first demonstrated by Dekel and Feinberg (2006) under the continuum hypothesis. Subsequently, Olszewski and Sandroni (2007b) constructed a test with the required properties (dispensing with the continuum hypothesis).

data and the forecasts made along the observed histories.

Assume that, before any data are observed, Bob delivers to Alice an entire theory of the stochastic process. A test is *future-independent* if whenever a theory f is rejected at some history s_t (observed at period t), then another theory f', which makes identical predictions as f until period t, must also be rejected at history s_t. So, if instead of delivering an entire theory, Bob announces each period a forecast on the basis of the actually observed data, then Alice is limited to future-independent tests because she cannot use as an input any forecasts contingent on data not yet realized.

A future-independent test must reject a theory based only on the available data and predictions up to the current period. A future-dependent test may reject a theory depending on these factors and on predictions contingent on data not yet realized. Consider a prediction such as "If it rains tomorrow, then it will also rain the day after tomorrow." Today no data is available on the qualifier of this prediction (whether it rains tomorrow) nor on the prediction itself (whether it rains after tomorrow). Yet, a future-dependent test may use such a prediction to reject a theory today (as opposed to the day after tomorrow when there will be data on this forecast).

A statistical test is *regular* if it is future-independent and unlikely to reject the actual data-generating process. A statistical test *can be ignorantly passed* if it is possible to strategically produce theories that are unlikely to be rejected on any future realization of the data.[2]

We show that any regular statistical test can be ignorantly passed. Thus, regular tests cannot reject a potentially strategic expert who knows how he will be tested. This holds even if the tester has arbitrarily long data sets at her disposal. Unless future-dependent tests can be employed, the data cannot reveal that the expert is completely uninformed.

Although future-independent tests cannot determine whether a potentially strategic expert is informed, some future-dependent tests can. Thus, whether the expert reports an entire theory or a sequence of forecasts make a fundamental difference. If Alice knows Bob's theory, then she can use the data she has observed to evaluate predictions contingent on data she does not have. These evaluations are necessary to prevent Alice's test from being ignorantly passed because future-dependent tests make use of predictions contingent on data not yet realized.

This paper is organized as follows: In section 2, we discuss some related ideas.

[2] We allow the uninformed expert to produce theories at random at period zero. So, the term *unlikely* refers to the expert's randomization and not to the possible realizations of the data.

In sections 3 and 4, we define our basic concepts. Section 5 shows our main result. In section 6, we show examples of regular tests. Section 7 presents the conclusions.

2. Related literature

2.1. Risk and uncertainty

An important distinction in economics is that between risk and uncertainty.[3] Risk is present in cases in which the available evidence can be properly summarized by a probability. Uncertainty is present in cases in which the available evidence is too imprecise to be properly summarized by a probability. The canonical example, used by Ellsberg (1961) to show the empirical content of this distinction, involves a draw from an urn containing winning and losing balls. If the composition of the urn is known, then the odds of drawing a winning ball are known. If the composition of the urn is not known, then the odds that a winning ball will be drawn are unknown. Ellsberg's experiment shows that some subjects prefer to know the odds, e.g., they prefer a ball drawn from the urn in which half the balls are winning to a ball drawn from the urn whose composition is not known.

In our model, Bob, if informed, faces risk. Alice and Bob, if uninformed, face uncertainty.[4] If Alice knows that Bob's theory is an accurate description of the data-generating process, then Alice benefits from his theory, no matter what the actual data-generating process is, because the theory tells her the relevant odds (i.e., it would replace her uncertainty with common risk).

The difficulty is that Alice does not know whether Bob is an informed expert who can deliver the data-generating process, or an uninformed agent who knows nothing about the relevant process. Thus, Alice does not know Bob's type (whether Bob faces risk or uncertainty), whereas Bob knows his type. A test can alleviate Alice's adverse selection problem (by revealing Bob's type) only if Bob finds it easier to pass the test when informed than when uninformed. However, we show strategies that Bob can use to pass the test even if he faces complete uncertainty, i.e., even if he knows nothing about the data-generating process.

[3] The distinction is traditionally attributed to Knight (1921). However, LeRoy and Singell (1987) argue that Knight did not have this distinction in mind.

[4] This is fundamentally different from the case in which both tester and forecaster face only risk. See Morgan and Stocken (2003) and Ottaviani and Sørensen (2006) (among others) for cheap-talk games between forecasters and decision-makers. See also Dow and Gorton (1997), Laster, Bennett and Geoum (1999), and Trueman (1988) for models in which professional forecasters have incentives to report their forecasts strategically.

2.2. Empirical tests of rational expectations

The rational expectations hypothesis has been extensively tested. The careful analysis by Keane and Runkle ((1990), (1998)) did not reject the hypothesis that professional forecasters' expectations are rational.[5] This literature contrasts with our work in that the forecasts are assumed to be reported nonstrategically and honestly. These differences in assumptions are partially due to differences in objectives. The purpose of this paper is to study the properties of nonmanipulable tests.

2.3. Testing strategic experts

The idea that an ignorant expert can strategically avoid being rejected by some empirical tests is not novel and can be found in a number of papers (see Foster and Vohra (1998), Fudenberg and Levine (1999), Lehrer (2001), Sandroni (2003), Sandroni, Smorodinsky and Vohra (2003), Vovk and Shafer (2005), and Olszewski and Sandroni (2007b) and (2007c)). Some of these results are reviewed in Cesa-Bianchi and Lugosi (2006). However, what is novel is the idea that nonmanipulability requires testers to use a theory's predictions which are contingent on data not yet realized. We are not aware of any result that shows or implies that the class of tests of interest here (i.e., regular tests) can be manipulated.[6]

3. Basic Set-Up

In each period one outcome, 0 or 1, is observed.[7] Before any data are observed, an expert, named Bob, announces a theory that must be tested. Conditional on any $t-$*history* $s_t \in \{0,1\}^t$, $s_t = (s^0, ..., s^{t-1})$, of outcomes in periods $0, ..., t-1$, Bob's theory claims that the probability of 1 in period t is $f(s_t)$.

To simplify the language, we identify a theory with its predictions. That is, theories that produce identical predictions are not differentiated. Hence, we define a *theory* as an arbitrary function that takes as an input any finite history and returns as an output a probability of 1.

Formally, a theory is a function

$$f : \{s_0\} \bigcup S_\infty \longrightarrow [0,1]$$

[5] See Lowell (1986) for other results on empirical testing of forecasts.

[6] In fact, our result subsumes most of the literature on manipulable tests.

[7] It is simple to extend the results to finitely many possible outcomes in each period.

where $S_\infty = \bigcup_{t=1}^{\infty} \{0,1\}^t$ is the set of of all finite histories and s_0 is the null history.

A tester named Alice tests Bob's theory empirically. So, given a string of data, Alice must either reject or not reject Bob's theory. A *test* T is an arbitrary function that takes as an input a theory f and returns as an output a set $T(f) \subseteq S_\infty$ of finite histories considered to be inconsistent with the theory f. Alice therefore rejects Bob's theory f if she observes data that belongs to $T(f)$.

Formally, a test is a function

$$T : F \to \bar{S},$$

where F is the set of all theories and $\bar{S}$ is the set of all subsets of S_∞.[8,9]

The timing of the model is as follows: At period zero, Alice selects her test T. Bob observes the test and announces his theory f (also at period zero). In period 1 and onwards, the data are revealed and Bob's theory is either rejected or not rejected by Alice's test at some point in the future.

Bob can be an informed expert who honestly reports to Alice the data-generating process. Bob may also be an uninformed expert who knows nothing about the data-generating process. If so, Bob tries to strategically produce theories with the objective of not being rejected by the data. Anticipating this, Alice wants a test that cannot be manipulated by Bob, if he is uniformed.

Although Alice tests Bob's theory using a string of outcomes, we do not make any assumptions about the data-generating process (such as whether it is a Markov process, a stationary process, or satisfies some mixing condition). It is very difficult to demonstrate that any key economic variable (such as inflation or the GDP) follows any of these well-known processes. At best, such assumptions about the data-generating process can be tested and rejected. More importantly, if Alice knew that the actual process belonged to a parametrizable set of processes (such

[8] We assume that $s_t \in T(f)$ implies that $s_m \in T(f)$ whenever $m \geq t$ and $s_t = s_m \mid t$ (i.e., s_t are the first t outcomes of s_m). That is, if a finite history s_t is considered inconsistent with the theory f, then any longer history s_m whose first t outcomes coincide with s_t is also considered inconsistent with the theory f.

For simplicity, we also assume that $s_t \in T(f)$ whenever $s_m \in T(f)$ for some $m > t$ and every s_m with $s_t = s_m \mid t$. That is, if *any* $m-$history that is an extension of a finite history s_t is considered inconsistent with the theory f, then the history s_t is itself considered inconsistent with the theory f.

[9] Alternatively, a test can be defined as a mapping from theories to stopping times. A stopping time can take finite or infinite values; a finite value is interpreted as a rejection of the theory, and if the stopping time takes an infinite value, the theory is not rejected.

as independent, identically distributed sequences of random variables), then she could almost perfectly infer the actual process from the data. Alice could accomplish this without any help from Bob.

Given that Bob must deliver an entire theory, Alice knows, at period zero, Bob's forecast contingent on any finite history. So in principle, Alice can make use of forecasts contingent on future data.

Let $s_t = s_m \mid t$ be the first t outcomes of s_m. Let $f_{s_m} = \{f(s_t),\ s_t = s_m \mid t,\ t = 0, .., m\}$ be a *sequence of the actual forecasts* made up to period m (including the forecast made for the outcomes at period m) if s_m is observed at period m. (Recall that s_m comprises outcomes from period 0 to $m-1$). Let the pair (f_{s_m}, s_m) be the observed record at period m. If Bob is required to produce a forecast each period, then Alice observes the record up to the current period, but not the forecasts contingent on unrealized data nor the forecasts contingent on future data.

3.1. Example

We now consider a simple empirical test:

$$\widetilde{R}(f, s_m) = \frac{1}{m}\sum_{t=0}^{m-1}[f(s_t) - s^t],$$

where $s_m = (s^0, ..., s^{m-1})$, marks the difference between the average forecast of 1 and the empirical frequency of 1.

Alice rejects the theory f on sufficiently long histories such that the average forecast of 1 is not sufficiently close to the empirical frequency of 1. That is, fix $\eta > 0$ and a period $\bar{m}$. Bob's theory f is rejected on any history s_m (and longer histories s_k with $s_m = s_k \mid m$) such that

$$\left|\widetilde{R}(f, s_m)\right| \geq \eta \text{ and } m \geq \bar{m}. \tag{3.1}$$

The tests defined above (henceforth called $\widetilde{R}-$tests) are notationally undemanding and can exemplify general properties of empirical tests. Given $\varepsilon > 0$, a pair (η, $\bar{m}$) can be picked such that if the theory f is the data-generating process, then f will not be rejected with probability $1 - \varepsilon$ (i.e., (3.1) occurs with probability less than ε). In addition, an $\widetilde{R}-$test rejects a theory based only on the observed record (i.e., whether (3.1) holds depends only on the data available and the sequence of actual forecasts).

Now assume that Bob is an uninformed expert who knows nothing about the data-generating process. Still, at period zero, he could announce a theory f that satisfies:

$$\begin{aligned} f(s_t) &= 1 && \text{if} \quad \widetilde{R}(f, s_t) < 0; \\ f(s_t) &= 0.5 && \text{if} \quad \widetilde{R}(f, s_t) = 0; \\ f(s_t) &= 0 && \text{if} \quad \widetilde{R}(f, s_t) > 0. \end{aligned}$$

If $\widetilde{R}$ is negative at period $t-1$, then no matter whether 0 or 1 is realized at period t, $\widetilde{R}$ increases. If $\widetilde{R}$ is positive at period $t-1$, then no matter whether 0 or 1 is realized at period t, $\widetilde{R}$ decreases. So $\widetilde{R}$ approaches zero as the data unfolds. If $\bar{m}$ is sufficiently large, Bob can pass $\widetilde{R}-$tests without any knowledge of the data-generating process.

Intuitively, the $\widetilde{R}-$tests seem weak. This intuition is seemingly confirmed by the proof that some of them can be passed without any relevant knowledge. The natural question is thus how to construct a better empirical test such that Bob, if uninformed, cannot be assured of passing. Candidates for such a test abound. However, as our main result shows, if the test has two basic properties (namely, that it does not reject the data-generating process and that it rejects theories based on the observed record), then Bob, even if completely uninformed, can also pass this test. Therefore no matter which test is used (with the stated qualifiers), Bob can be near certain he will not be rejected.

4. Properties of Empirical Tests

Any theory f uniquely defines the probability of any set $A \subseteq S_\infty$ of finite histories (denoted by $P^f(A)$). The probability of each finite history $s_m = (s^0, ..., s^{m-1})$ is the product

$$\prod_{t=0}^{m-1} f(s_t)^{s^t} (1 - f(s_t))^{1-s^t}, \tag{4.1}$$

where $s_t = s_m \mid t$ for $t \leq m$.

Definition 1. *Fix $\varepsilon \in [0,1]$. A test T does not reject the data-generating process with probability $1-\varepsilon$ if for any $f \in F$,*

$$P^f(T(f)) \leq \varepsilon.$$

If Bob is an informed expert and announces the data-generating process, then with high probability he will not be rejected. As argued in Section 2.1, if Alice is averse to uncertainty, then she benefits from learning the data-generating process, no matter what the data-generating process might be. However, some standard tests reject some theories up front (before any data are observed). So the assumption that a test does not reject any data-generating process may seem strong. The central idea behind this assumption is that Alice will not reject a theory without testing it, and therefore tests any theory that Bob may announce. In section 5.2.1, we slightly restrict the set of permissible theories.

Two theories f and f' are *equivalent until period* m if $f(s_t) = f'(s_t)$ for any $t-$history s_t, $t \leq m$. Two theories are therefore equivalent until period m if they make the same predictions up to period m.

Definition 2. *A test T is future-independent if, given any pair of theories f and f' that are equivalent until period m, $s_t \in T(f)$, $t \leq m+1$, implies $s_t \in T(f')$.*

A test is future-independent if, whenever a theory f is rejected at an $(m+1)-$history s_{m+1}, another theory f', which makes exactly the same predictions as f until period m, must also be rejected at s_{m+1}. So a future-independent test rejects a theory at period m based only on the observed data and the predictions made until period m. If Bob is required to produce a forecast only from period to period, then Alice cannot run a test that is not future-independent.

Definition 3. *A test T is called a regular $\varepsilon-$test if T is future-independent and does not reject the data-generating process with probability $1-\varepsilon$.*

In section 6, we review well-known empirical tests (including calibration tests) that do not reject theories up front and are regular. These tests use only realized forecasts and available data and, by definition, no data are available for predictions contingent on information not yet realized.

Bob is not restricted to selecting a theory deterministically. He may randomize when selecting his theory at period 0. (Hence, he randomizes at most once.) Let a random generator of theories ζ be a probability distribution over the set F of all theories. Given any finite history $s_t \in \{0,1\}^t$ and a test T, let

$$\zeta_T(s_t) := \zeta(\{f \in F : s_t \in T(f)\})$$

be the probability that ζ selects a theory that will be rejected if s_t is observed.[10]

[10] This definition requires the set F to be equipped with a $\sigma-$algebra, as well as the sets $\{f \in F : s_t \in T(f)\}$ to be measurable with respect to that $\sigma-$algebra.

Definition 4. *A test T can be ignorantly passed with probability $1 - \varepsilon$ if there exists a random generator of theories ζ such that, for all finite histories $s_t \in S_\infty$,*

$$\zeta_T(s_t) \leq \varepsilon.$$

The random generator ζ may depend on the test T, but not on any knowledge of the data-generating process. If a test can be ignorantly passed, Bob can randomly select theories that, with probability $1-\varepsilon$ (according to Bob's randomization device), will not be rejected, no matter what data are observed in the future. Thus, Alice has no reason to run a test that can be ignorantly passed if the forecaster is potentially strategic and knows how he will be tested. Even if Bob completely ignores the data-generating process, the test is unlikely to reject his theory, *no matter how the data unfolds.*

5. Main Result

Proposition 1. *Fix $\varepsilon \in [0, 1]$ and $\delta \in (0, 1 - \varepsilon]$. Any future-independent test T that does not reject the data-generating process with probability $1 - \varepsilon$ can be ignorantly passed with probability $1 - \varepsilon - \delta$.*[11]

Proposition 1 shows a fundamental limitation of regular tests, namely, that any regular test can be ignorantly passed. In particular, if Alice cannot run future-dependent tests (e.g., the expert is required to announce forecasts only from period to period), then she essentially has no reason to run a test when confronted with a potentially strategic expert who knows how his theory will be tested. This result holds even if Alice has unboundedly large data sets at her disposal.

There is an important contrast between Proposition 1 and the results of Dekel and Feinberg (2006) and Olszewski and Sandroni (2007b), both of which exhibit tests that do not reject the data-generating process and cannot be ignorantly passed. These tests enable Alice to determine whether the expert is informed, even though she has no prior knowledge about the data-generating process. The critical factor for the contrast is whether Bob announces a theory or a forecast each period. In the former papers, Alice can run a future-dependent test and use data to discredit an uninformed expert who is aware of her test. In the present paper, she cannot.

We equip F with the pointwise (forecast-by-forecast) convergence topology and the $\sigma-$algebra of all Borel sets generated by this topology. (See section 8 for the details.) All tests considered in this paper are assumed to satisfy the measurability provision.

[11] Note that ε in proposition 1 need not be "small."

5.1. An alternative perspective on proposition 1

Consider the following contract: it specifies that Bob must deliver a theory to Alice; Alice must pay Bob for the theory at period zero; but if, in the future, Bob's theory is rejected (by a test T defined in advance), then Bob is punished.[12]

Bob's disutility for having his theory rejected is $M > 0$ (called the *rejection disutility*). Bob's utility u (called the *reward*) of the payment for the theory is a draw from a random $\tilde{u}$ that has a density and support on $[u_l, u_u]$, $u_u > u_l$, where the upper-bound u_u can be arbitrarily close to the lower bound u_l. Bob does not discount the future. Without a contract, Bob's utility is zero.

Bob considers the contract and, before observing any data, must decide whether to accept it and, if he does, which theory to announce. Alice does not know whether Bob is informed. She wants a screening contract such that Bob, if informed, accepts but, if uninformed, does not.

For simplicity, we assume that Bob, if informed, reveals the data-generating process truthfully. (This assumption can be relaxed, as argued in Section 5.2 below.) The following condition ensures that Bob, if informed, accepts the contract: For every theory $f \in F$

$$u_l \geq M \; P^f(T(f)). \tag{5.1}$$

The right-hand side of (5.1) is Bob's expected rejection disutility when Bob announces f, where f is the data-generating process. Alice does not know the data-generating process. (It is what she hopes to learn from Bob.). But she does know that if Bob's reward u_l satisfies (5.1) and Bob is informed, then he will accept the contract. We say that *the informed expert accepts test* T whenever condition (5.1) is satisfied.

Now assume that Bob is uninformed. Then, by definition, Bob faces uncertainty. So Bob may not know the odds that any theory will be rejected. We assume under uncertainty Bob values a contract based on a worst-case scenario. We say that *the uninformed expert accepts test* T if the following condition holds: for every reward $u > u_l$, there exists a random generator of theories ζ such that

$$u \geq \sup_{s_t \in S_\infty} M\zeta(s_t). \tag{5.2}$$

That is, Bob's reward is greater than Bob's expected rejection disutility evaluated for the case of having the worse possible data that may be collected in the

[12] Note that the idea of interpreting an empirical test as a contract appears in Olszewski and Sandroni (2007a), although the assumptions made there differ from those imposed here.

future.[13]

Proposition $1'$ Fix any future-independent test T. If the informed expert accepts test T, then the uninformed expert accepts test T.

Proposition $1'$ is an immediate corollary of Proposition 1. Indeed, since $P^f(T(f)) \leq \varepsilon := u_l/M$ for every theory $f \in F$, there exists a random generator of theories ζ such that $\zeta(s_t) \leq u_l/M + \delta = u/M$, where $\delta := (u - u_l)/M$, for every history $s_t \in S_\infty$.

Proposition $1'$ expresses the idea that no matter how severe the punishment for having a theory rejected, no matter how small the reward for delivering a theory, no matter how large the data set the tester eventually has at her disposal - despite all of this, the completely uninformed expert will choose to announce a theory, provided that the test is future-independent, that the test is known, and that the informed expert also chooses to announce a theory.

On the other hand, as shown by Dekel and Feinberg (2006) and Olszewski and Sandroni (2007b), there are (future-dependent) tests that the informed expert accepts and the uninformed expert does not accept. (For these tests the right-hand side of (5.1) is close to zero for every theory, and the right-hand side of (5.2) is close to M for every random generator of theories.)

Proposition $1'$ holds under the assumption that the uninformed expert evaluates his prospects based on a worse-case scenario. This is the most pessimistic behavioral rule for decision under uncertainty among those axiomatized by Gilboa and Schmeidler (1989). Like other assumptions in proposition $1'$, this assumption is not meant to be realistic. Rather, it should be interpreted as an extreme case in which our result holds, thus demonstrating that it still holds under less extreme (but perhaps more realistic) conditions.

We conclude this section with a comment on the assumption that agents do not discount future payoffs. Our result still holds if Bob discounts the future, but the undiscounted scenario is more interesting. In the discounted case, Alice might be discouraged from using large data sets because Bob may not fear rejection in the distant future. The undiscounted case is more interesting because it shows that even if Alice has unbounded data sets, she still cannot determine whether Bob is informed. Hence, no exogenous constraints encourage Alice to stop testing Bob's theory.

It also worth pointing out that the focus of the model is not on distant future behavior. The significant choice (i.e., which theory to announce, if any) occurs at

[13]Given that $\tilde{u}$ has a density, the reward u is strictly greater than u_l with probability one.

period zero. The future matters only because the fear of future rejection is the main deterrent for uninformed experts.

5.2. Remarks on the assumptions of proposition 1

There are three main assumptions in proposition 1: the test is unlikely to reject the data-generating process, the test is future-independent, and Bob knows Alice's test. We discuss each assumption separately.

5.2.1. The test is unlikely to reject the data-generating process

We see both a conceptual and a practical way of establishing the validity of this assumption. In Section 2.1, we emphasize the conceptual distinction between risk and uncertainty (or between knowing and not knowing the odds). People often prefer to know the odds, no matter what these odds might be, rather than not know the odds.

In practice, people seem to have additional reasons for caring about learning the odds (of processes that matter to them), even if they could learn that all finite-histories of the same length have equal probability. In that case, they would know that the past data are irrelevant and would not waste time and effort looking for patterns. Take, for example, Paul Samuelson's claim that "properly anticipated prices fluctuate randomly." The fact that the claim that asset prices follow a random walk (which is, of course, not the same as a claim of equal probability) was extensively tested empirically strongly suggests that there is practical interest in this claim.

The assumption that the test does not reject the data-generating process can be relaxed. In particular, the informed expert need not be truthful. It suffices that for every theory $f \in F$, there is a theory $\widetilde{f} \in F$ such that $P^f(T(\widetilde{f})) \leq \varepsilon$, i.e., the informed expert can pass the test. This result requires a slightly more complicated proof, which can be found in the working paper version of this paper.

Proposition 1 generalizes to the case where the permissible theories are restricted to some (but not any) class. For example, it is straightforward to extend Proposition 1 to the case where the probability of 1 (given the available data) is in an interval $[a, b]$, where $0 \leq a < b \leq 1$. Assume that before any data are observed, Bob claims to know the exact process in this class that will generate the data. Also assume that Alice's test is future-independent and is unlikely to reject any process in this class if it actually does generate the data.[14] If Bob knows Alice's test, he can pass it even though he does not know which process generates the data. More precisely, Bob can produce theories that are likely to pass Alice's test, no matter which process in this class generates the data.

It should, however, be mentioned that proposition 1 does not generalize to the class of processes such that the probability of 1 is always bounded away from 0.5 (or any other given value in $(0, 1)$). So our results do not generalize to the case where the probability of 1 is restricted to be outside an interval $[a, b]$, where $0 < a < b < 1$. This case is analyzed in Olszewski and Sandroni (2007d).

5.2.2. The test is future-independent

Future-independence is a restriction on Alice's test which often must be imposed. For example, this would be the case if Bob claims that he will know no earlier than at period t the probability that 1 occurs at period $t + 1$.

It follows immediately from the proof of proposition 1 that the assumption of future-independence can be partially relaxed. It suffices that for every period m, there is a period $m' > m$ such that if theory f is rejected at an $m-$history s_m, another theory f', which is equivalent to f until period m', must also be rejected at s_m. Consider the special case that $m' = m + T$, for some $T > 0$. Assume that Bob delivers to Alice the part of his theory that allows her to calculate the odds of outcomes up to T periods ahead. That is, contingent on the data at period r, Bob claims that with his partial theory, Alice can accurately compute the probabilities of future events until period $r + T$. Perhaps due to computational constraints (the cause is irrelevant for our purposes), Bob claims that his partial theory does not deliver accurate probabilities for distant future events. Then, in order to test Bob's claim, Alice must use a (distant) future-independent test.

If Bob claims that he has a complete understanding of the stochastic process (including the odds of distant future events), then future-dependent tests can be used. In this case, we do not assert that they should not.

[14] The definition of a future-independent test can be extended in a natural way to the case where theories are restricted to a class. It requires that whenever a theory f (in this class) is rejected at an $(m + 1)-$history s_{m+1}, then another theory f' (in this class) that makes exactly the same predictions as f until period m must also be rejected at s_{m+1}.

5.2.3. Bob knows Alice's test

The assumption that Bob observes Alice's test before announcing his theory can be replaced with the assumption that Bob correctly anticipates the test that Alice will use after his theory is announced. It is also straightforward to extend proposition 1 to the case in which Alice selects her theory at random, provided that Bob knows (or correctly anticipates) the probabilities Alice will use to select each test.

The assumption that Bob knows the odds Alice uses to select her test cannot be completely discarded. Let $\mp_{0.5}$ be the set of all future-independent tests that do not reject the data-generating process with probability 0.5. In section 8, we show that for any $\zeta \in \Delta(F)$,

$$\sup_{s_t \in S_\infty,\ T\ \in\ \mp_{0.5}} \zeta_T(s_t) = 1. \tag{5.3}$$

So, if Bob is uninformed, he cannot simultaneously pass all future-independent tests in $\mp_{0.5}$ with strictly positive probability.[15]

5.3. Intuition of proposition 1

Consider the following zero-sum game between Nature and Bob: Nature's pure strategy is an infinite sequence of outcomes. Bob's pure strategy is a theory. Bob's payoff is one if his theory is never rejected and zero otherwise. Both Nature and Bob are allowed to randomize.

Assume that the test T does not reject the data-generating process with probability $1-\varepsilon$. Then for every mixed strategy of Nature, there is a pure strategy for Bob (to announce the theory f that coincides with Nature's strategy) that gives him a payoff of $1-\varepsilon$ or higher. So, if the conditions of Fan's (1954) MinMax are satisfied, there is a (mixed) strategy ζ for Bob that ensures him a payoff arbitrarily close to $1-\varepsilon$, no matter what strategy Nature chooses. In particular, for any path $s \in S_\infty$ that Nature can select, Bob's payoff is arbitrarily close to $1-\varepsilon$.

The conditions of Fan's (1954) MinMax are satisfied if Bob's strategy space is compact and Nature's payoff function is lower semi-continuous with respect to

[15] In Olszewski and Sandroni (2007d), Section 3.4, a zero-sum game between Alice and Bob is considered. Bob announces a theory (perhaps at random) and Alice announces a test. Alice and Bob properly anticipate each other's strategy. (Alice knows that Bob is an uninformed expert.) Bob's payoff is one if his chosen strategy ignorantly passes Alice's chosen test. An idea related to the one used to demonstrate (5.3) is used to show that this game has no equilibrium.

Bob's strategy. In the existing literature, Bob's (mixed) strategy space is typically equipped in the weak$-*$ topology. Although this topology makes Bob's strategy space compact, Nature's payoff function now need not be lower semi-continuous with respect to Bob's strategy, unless we impose additional requirements (other than future-independence) on the test T. Other topologies make Nature's payoff function lower semi-continuous with respect to Bob's strategy, but Bob's strategy space is then not compact.

We overcome this difficulty by restricting the set of Bob's pure strategies to theories that make, in each period t, a forecast from a finite set of forecasts $R_t \subset [0,1]$. This pure strategy space is compact, if endowed in the product of discrete topologies. As a result, Bob's mixed strategy space (the set of all probability distributions over pure strategies) is compact as well, if endowed with the weak$-*$ topology. The assumption that the test T is future-independent guarantees that for every pure strategy of Nature, the set of Bob's pure strategies rejected by T is open. By a standard argument, this implies that Nature's payoff function is lower semi-continuous with respect to Bob's strategy.

If the set of Bob's pure strategies is restricted arbitrarily, then we may no longer have the property that for every mixed strategy of Nature, there is a strategy for Bob that gives him a payoff of $1-\varepsilon$ or higher. However, our key lemma, i.e., lemma 2 from section 8, shows that this property is preserved for properly chosen finite sets of forecasts $R_t \subset [0,1]$ which Bob is allowed to make in period t.

5.4. Future Research

Proposition 1 is an impossibility result, which provides motivation for investigating ways to get around it. One route is to consider future-dependent tests. As mentioned in section 5.1, there are limits to their use, but a deeper analysis of future-dependent tests may yield important insights.

The results obtained so far show that in testing theories, it is important to know more than the data and the theory's predictions along the actual history. It is also important to know the theory's distant-future predictions and to reject theories partly based on distant-future predictions at a time when no data are available for those predictions. We hope that additional research will reveal the proper way in which distant-future predictions should be used. This could be achieved by observing how future-dependent tests that cannot be ignorantly passed classify theories according to their distant-future predictions.

A second route around proposition 1 would be to restrict theories to a class

$\mathcal{C}$. This comes with a cost because potentially correct theories are now rejected by definition. However, a result showing that a test that cannot be manipulated when theories are restricted - and simultaneously showing that the test can be manipulated when the theories are unrestricted - might deliver a better understanding of how the imposition of some structure of permissible theories may help in testing them empirically.

A third route would be to impose computational bounds on the expert's ability to either randomize or produce theories. We refer the reader to Fortnow and Vohra (2006) for results in this direction. Finally, one might consider the case in which there are multiple experts. Here, we refer the reader to Al-Najjar and Weinstein (2007) and to Feinberg and Stewart (2007). We also refer the reader to a recent paper by Shmaya (2008) that provides a result related to proposition 1 in the single expert setting, for the case in which the tester has infinitely large data sets at her disposal.

6. Empirical tests

The purpose of this section is to show that the assumptions of proposition 1 are satisfied by many statistical models that do not reject theories up front. We do not analyze every statistical model, but provide a number of simple examples.

Asymptotic tests work as if Alice could decide whether to reject Bob's theory at infinity. Asymptotic tests can be approximated by tests that reject theories in finite time (as required in Section 3). We now show an example of an asymptotic test that can be approximated by regular tests.

Fix $\delta \in (0, 0.5)$. Given a theory f, let f_δ be an alternative theory defined by

$$f_\delta(s_t) = \begin{cases} f(s_t) + \delta & \text{if} \quad f(s_t) \leq 0/5; \\ \\ f_\delta(s_t) - \delta & \text{if} \quad f(s_t) > 0/5. \end{cases}$$

A straightforward martingale argument shows that

$$\frac{P^{f_\delta}(s_t)}{P^f(s_t)} \underset{t \to \infty}{\longrightarrow} 0 \; P^f - \text{almost surely.}$$

So, under the null hypothesis (that P^f is the data-generating process), the likelihood of P^{f_δ} becomes much smaller than the likelihood of P^f.[16] The likelihood

[16]The alternative hypothesis need not be a single theory. This assumption is made for simplicity only.

test rejects a theory f on the set $R(f)$ of infinite histories such that the likelihood ratio

$$\frac{P^{f_\delta}(s_t)}{P^f(s_t)}$$

does not approach zero.

Say that a test T is harder than the likelihood test if $R(f) \subseteq T(f)$.[17] Rejection by the likelihood test thus implies rejection by the test T.

Proposition 2. *Given $\varepsilon > 0$, there exists a regular $\varepsilon-$test T that is harder than the likelihood test.*

By proposition 1, any regular test T can be ignorantly passed with probability near $1-\varepsilon$. So, by proposition 2, the likelihood test can be ignorantly passed with arbitrarily high probability. It is not obvious whether the theory f or the alternative theory f_δ eventually produces a higher likelihood. However, a false expert can select a theory at random such that, no matter which data are realized, it will in the future (with arbitrarily high chance) generate a much higher likelihood than the alternative theory.

The unexpected result is proposition 1. Proposition 2 is a natural finding. An intuition for proposition 2 is as follows: P^f-almost surely, the likelihood ratio approaches zero. Hence, with arbitrarily high probability, the likelihood ratio must remain small if the string of data is long enough. Let T be the test that rejects the theory f whenever the likelihood ratio is not small and the string of data is long. By construction, the test T is harder than the likelihood test and with high probability does not reject the truth. Moreover, the test T is future-independent because the likelihood ratio depends only on the forecasts made along the observed history.

The basic idea in proposition 2 is not limited to the likelihood test. Other tests can also be associated with harder regular tests. Consider calibration tests. Let $\mathcal{I}_t$ be an indicator function that depends on the data up to period $t-1$ (i.e., it depends on s_t) and on the predictions made up to period $t-1$ including the prediction made at period $t-1$ for the outcome at period t (i.e., $f(s_k)$, $s_k = s_t \mid k$, $k \leq t$). For example, $\mathcal{I}_t$ can be equal to 1 if $f(s_t) \in [\frac{j}{n}, \frac{j+1}{n}]$ for some $j < n$ and zero otherwise. Alternatively, $\mathcal{I}_t$ can be equal to 1 if t is even and 0 if t is

[17]Formally, $T(f)$ comprises finite histories, so in the inclusion $R(f) \subset T(f)$, as well as in several other places, we identify $T(f)$ with the set of all infinite extensions of histories from $T(f)$.

odd. Consider an arbitrary countable collection $\mathcal{I}^i = (\mathcal{I}_0^i, ..., \mathcal{I}_t^i, ...)$, $i = 1, 2, ...,$ of indicator functions. The calibration test (determined by this collection) requires that for all i,

$$\frac{\sum_{t=0}^{m-1} [f(s_t) - s^t] \mathcal{I}_t^i}{\sum_{t=0}^{m-1} \mathcal{I}_t^i} \longrightarrow 0 \text{ whenever } \sum_{t=0}^{m-1} \mathcal{I}_t^i \underset{m \to \infty}{\longrightarrow} \infty; \tag{6.1}$$

where $s_m = (s^0, ..., s^{m-1})$.

These calibration tests require a match between average forecasts and empirical frequencies on specific subsequences. These subsequences could be, as in Foster and Vohra (1998), periods in which the forecasts are near $p \in [0, 1]$. Then the test requires that the empirical frequencies of 1 be close to p in the periods that follow a forecast of 1 that was close to p. Alternatively, these subsequences could also be, as in Lehrer (2001), periods in which a certain outcome was observed. In general, the calibration test rejects a theory f if (6.1) does not hold.

Proposition 3. *Given $\varepsilon > 0$, there exists a regular $\varepsilon-$test T' that is harder than the calibration test.*

The intuition of proposition 3 is the same as that of proposition 2. A sophisticated law of large numbers shows that, P^f-almost surely, the calibration scores in (6.1) eventually approach zero. Hence, with arbitrarily high probability, these calibration scores must remain small if the string of data is long enough. Let T' be the test that rejects the theory f whenever the calibration scores are not small and the string of data is long. By construction, the test T' is harder than the calibration test and with high probability does not reject the truth. Moreover, the test T' is future-independent because the calibration scores depend only on the forecasts made along the observed history.

By propositions 1 and 3, the calibration tests can be ignorantly passed with arbitrarily high probability. Hence, a false expert can produce forecasts that, when the data is eventually revealed, will prove to be calibrated. This result combines Foster and Vohra's (1998) result, where the indicator function depends only on the forecasts, and Lehrer's (2001) result, where the indicator function depends only on the data.

7. Conclusion

Consider the following three basic properties of empirical tests. 1) The test is unlikely to reject the data-generating process. 2) The test is future-independent. 3) The test cannot be ignorantly passed with arbitrarily high probability.

Several statistical tests, such as the likelihood and calibration tests, are future-independent and unlikely to reject the data-generating process. The tests developed by Dekel and Feinberg (2006) and Olszewski and Sandroni (2007b) are unlikely to reject the data-generating process and cannot be ignorantly passed. The trivial test that rejects all theories on all data sets is future-independent and cannot be ignorantly passed. Consequently, a test can satisfy any pair of properties 1, 2, and 3. However, no test satisfies properties 1, 2, and 3 simultaneously. It follows that whether a tester can effectively use data to discredit a potentially strategic expert depends crucially on whether the expert announces an entire theory of the stochastic process or a forecast each period. In the former case, the tester can use data to discredit a potentially uninformed expert. In the latter case, however, she cannot as long as the expert knows how he will be tested.

8. Proofs

We use the following terminology: Let $\Omega = \{0,1\}^\infty$ be the set of all *paths*, i.e., infinite histories. Given a path s, let $s \mid t$ be the first t coordinates of s. A *cylinder* with base on $s_t \in \{0,1\}^t$ is the set $C(s_t) \subset \{0,1\}^\infty$ of all infinite extensions of s_t. We endow Ω with the topology that comprises unions of cylinders with finite base (or, equivalently, the product of discrete topologies on $\{0,1\}$). Let $\Im_t$ be the algebra that consists of all finite unions of cylinders with base on $\{0,1\}^t$. It is convenient to define $\Im_0$ as the trivial $\sigma-$algebra consisting of Ω and the empty set. Denote by N the set of natural numbers. Let $\Im$ be the $\sigma-$algebra generated by the algebra $\Im^0 := \bigcup_{t\in N} \Im_t$, i.e., $\Im$ is the smallest $\sigma-$algebra which contains $\Im^0$.

More generally, equip every compact metric space X with the $\sigma-$algebra of Borel sets (i.e., the smallest $\sigma-$algebra which contains all open sets). Note that $\Im$ is the $\sigma-$algebra of Borel subsets of Ω equipped with the topology that comprises unions of cylinders with finite base. Let $\Delta(X)$ be the set of all probability measures on X. Equip $\Delta(X)$ with the weak$-*$ topology.

We endow the set of all theories F with the product of discrete topologies. More precisely, in this topology, an open set that contains a theory f must also contain all theories g such that $g(s_t) = f(s_t)$ for some finite set of histories $s_t \in \{s_0\} \cup S_\infty$.

We will use the following two lemmas. Let X be a metric space. Recall that a function $u : X \to R$ is *lower semi-continuous* at an $x \in X$ if for every sequence $(x_n)_{n=1}^{\infty}$ converging to x:

$$\forall_{\varepsilon>0} \quad \exists_{\overline{n}} \quad \forall_{n \geq \overline{n}} \quad u(x_n) > u(x) - \varepsilon.$$

The function u is lower semi-continuous if it is lower semi-continuous at every $x \in X$.

As in the main text, we sometimes identify $T(f)$ with the set of all infinite extensions of histories from $T(f)$.

Lemma 1. *Let $U \subset X$ be an open subset of a compact metric space X. The function $H : \Delta(X) \to [0, 1]$ defined by*

$$H(P) = P(U)$$

is lower semi-continuous.

Proof: See Dudley (1989), Theorem 11.1.1(b).■

It is well-known that for every theory $f \in F$, (4.1) determines uniquely a measure $P^f \in \Delta(\Omega)$. We refer to P^f as the probability measure associated with the theory $f \in F$.

Definition 5. *A set $F' \subseteq F$ is $\delta-$supdense in F, $\delta > 0$, if for every theory $g \in F$ there exists a theory $f \in F'$ such that*

$$\sup_{A \in \Im} \left| P^f(A) - P^g(A) \right| < \delta.$$

Let γ be a sequence of positive numbers $(\gamma_t)_{t=0}^{\infty}$. Given γ, let R be a sequence of finite sets $(R_t)_{t=0}^{\infty}$ such that $R_t \subset (0, 1)$, $t \in N$, and

$$\forall_{x \in [0,1]} \; \exists_{r \in R_t} \quad |x - r| < \gamma_t.$$

Given R, let $\overline{F}$ be a subset of F defined by

$$\overline{F} = \left\{ f \in F : \forall_{t \in N} \; \forall_{s_t \in \{0,1\}^t \text{ (or } s_t = s_0 \text{ if } t=0)} \quad f(s_t) \in R_t \right\}.$$

Lemma 2. *For every $\delta > 0$ there exists a sequence of positive numbers γ such that $\overline{F}$ is $\delta-$subdense in F.*

Proof: For now, consider an arbitrary γ. Given $g \in F$, take $f \in \overline{F}$ such that

$$\forall_{t \in N} \ \forall_{s_t \in \{0,1\}^t \text{ (or } s_t = s_0 \text{ if } t=0)} \quad |f(s_t) - g(s_t)| < \gamma_t.$$

We shall show that there exists a sequence $(\gamma_t)_{t=1}^{\infty}$ such that

$$\left|P^f(C(s_r)) - P^g(C(s_r))\right| < \frac{\delta}{2} \tag{8.1}$$

for every cylinder $C(s_r)$. Later, we generalize this statement to all open sets U. It will, however, be easier to follow the general argument when one sees it first in the simpler case in which U is a cylinder.

Let

$$h^f(s_t) = f(s_t)^{s^t}(1 - f(s_t))^{1-s^t}$$

and

$$h^g(s_t) = g(s_t)^{s^t}(1 - g(s_t))^{1-s^t}.$$

Then

$$\left|P^f(C(s_r)) - P^g(C(s_r))\right| = \left|h^f(s_0) \cdot ... \cdot h^f(s_{r-1}) - h^g(s_0) \cdot ... \cdot h^g(s_{r-1})\right|,$$

and

$$\left|h^f(s_0) \cdot ... \cdot h^f(s_{r-1}) - h^g(s_0) \cdot ... \cdot h^g(s_{r-1})\right| \leq$$
$$\leq (h^g(s_0) + \gamma_1) \cdot ... \cdot \left(h^g(s_{r-1}) + \gamma_{r-1}\right) - h^g(s_0) \cdot ... \cdot h^g(s_{r-1}) \leq$$
$$\leq \left[\prod_{t=0}^{r-1}(1 + \gamma_t) - 1\right] \leq \left[\prod_{t=0}^{\infty}(1 + \gamma_t) - 1\right],$$

where $s_k = s_r \mid k$ for $k < r$. The first inequality follows from the fact that

$$|(a_0 + b_0) \cdot ... \cdot (a_{r-1} + b_{r-1}) - a_0 \cdot ... \cdot a_{r-1}| \leq (a_0 + |b_0|) \cdot ... \cdot (a_r + |b_{r-1}|) - a_0 \cdot ... \cdot a_{r-1} \tag{8.2}$$

for any sets of numbers $a_0, ..., a_{r-1} > 0$ and $b_0, ..., b_{r-1}$. Apply (8.2) to $a_k = h^g(s_k)$ and $b_k = h^f(s_k) - h^g(s_k)$, $k = 0, ..., r-1$. The second inequality follows from the fact that the function

$$(a_0 + b_0) \cdot ... \cdot (a_{r-1} + b_{r-1}) - a_0 \cdot ... \cdot a_{r-1}$$

is increasing in $a_0, ..., a_{r-1}$ for any sets of positive numbers $a_0, ..., a_{r-1}$ and $b_0, ..., b_{r-1}$.

So, (8.1) follows if we take a sequence $(\gamma_t)_{t=0}^{\infty}$ such that

$$\prod_{t=0}^{\infty}(1+\gamma_t) < 1 + \frac{\delta}{2}.$$

We shall now show that a slightly stronger condition,

$$\prod_{t=0}^{\infty}(1+2\gamma_t) < 1 + \frac{\delta}{4},$$

guarantees that $\left|P^f(U) - P^g(U)\right| < \delta/2$ for every open set (or union of cylinders) U, and not only for every single cylinder.

Indeed, suppose first that there is an n such that U is a union of cylinders with base on s_t with $t \leq n$. Since every cylinder with base on s_t can be represented as the union of two cylinders with base on $s'_{t+1} = (s_t, 0)$ and $s''_{t+1} = (s_t, 1)$, respectively, the set U is the union of a family of cylinders $\mathcal{C}$ with base on histories of length n. Thus,

$$\left|P^f(U) - P^g(U)\right| \leq \sum_{C(s_n)\in C} \left|P^f(C(s_m)) - P^g(C(s_m))\right|$$

$$\leq \sum_{C(s_n)\in C} \left[(h^g(s_0)+\gamma_0)\cdot ... \cdot \left(h^g(s_{n-1})+\gamma_{n-1}\right) - h^g(s_0)\cdot ... \cdot h^g(s_{n-1})\right] \leq$$

$$\leq \sum_{s_n\in\{0,1\}^n} \left[(h^g(s_0)+\gamma_0)\cdot ... \cdot \left(h^g(s_{n-1})+\gamma_{n-1}\right) - h^g(s_0)\cdot ... \cdot h^g(s_{n-1})\right] =$$

$$= \sum_{s_n\in\{0,1\}^n} (h^g(s_0)+\gamma_0)\cdot ... \cdot \left(h^g(s_{n-1})+\gamma_{n-1}\right) - 1 =$$

$$= \sum_{s_{n-1}\in\{0,1\}^{n-1}} (h^g(0)+\gamma_0)\cdot(h^g(0,s_0)+\gamma_1)\cdot ... \cdot \left(h^g(0,s_{n-2})+\gamma_{n-1}\right) +$$

$$+ \sum_{s_{n-1}\in\{0,1\}^{n-1}} (h^g(1)+\gamma_0)\cdot(h^g(1,s_0)+\gamma_1)\cdot ... \cdot \left(h^g(1,s_{n-2})+\gamma_{n-1}\right) - 1 \leq$$

$$\leq \left[h^g(0)+\gamma_0+h^g(1)+\gamma_0\right]\cdot$$

$$\cdot \max\left\{ \begin{array}{c} \sum_{s_{n-1}\in\{0,1\}^{n-1}} \left(h^g(0,s_0)+\gamma_1\right)\cdot \ldots \cdot \left(h^g(0,s_{n-2})+\gamma_{n-1}\right), \\ \sum_{s_{n-1}\in\{0,1\}^{n-1}} \left(h^g(1,s_0)+\gamma_1\right)\cdot \ldots \cdot \left(h^g(1,s_{n-2})+\gamma_{n-1}\right) \end{array} \right\} - 1 =$$

$$= [1+2\gamma_0]\cdot$$

$$\cdot \max\left\{ \begin{array}{c} \sum_{s_{n-1}\in\{0,1\}^{n-1}} \left(h^g(0,s_0)+\gamma_1\right)\cdot \ldots \cdot \left(h^g(0,s_{n-2})+\gamma_{n-1}\right), \\ \sum_{s_{n-1}\in\{0,1\}^{n-1}} \left(h^g(1,s_0)+\gamma_1\right)\cdot \ldots \cdot \left(h^g(1,s_{n-2})+\gamma_{n-1}\right) \end{array} \right\} - 1.$$

We can estimate each sum in this last display in a manner similar to the one we have used to estimate $\sum_{s_n\in\{0,1\}^n} \left(h^g(s_0)+\gamma_0\right)\cdot \ldots \cdot \left(h^g(s_{n-1})+\gamma_{n-1}\right)$; we continue this way to conclude that

$$\left|P^f(U) - P^g(U)\right| \leq \left[\prod_{t=0}^{n-1}(1+2\gamma_t) - 1\right] < \frac{\delta}{4}.$$

Now, suppose that U is the union of an arbitrary family of cylinders $\mathcal{C}$. Represent U as

$$U = \bigcup_{n=1}^{\infty} U_n$$

where U_n is the union of cylinders $C \in \mathcal{C}$ with base on s_t such that $t \leq n$. Since the sequence $\{U_n : n = 1, 2, ...\}$ is ascending, $\left|P^f(U) - P^f(U_n)\right| < \delta/8$ and $|P^g(U_n) - P^g(U)| < \delta/8$ for large enough n. Thus,

$$\left|P^f(U) - P^g(U)\right| \leq \left|P^f(U) - P^f(U_n)\right| +$$
$$+ \left|P^f(U_n) - P^g(U_n)\right| + |P^g(U_n) - P^g(U)| < \delta/2.$$

Finally, observe that $\left|P^f(A) - P^g(A)\right| < \delta$ for every $A \in \Im$. Indeed, take a set $U \supset A$, which is a union of cylinders, such that $\left|P^f(U) - P^f(A)\right|$, $|P^g(U) - P^g(A)| < \delta/4$. Since $\left|P^f(U) - P^g(U)\right| < \delta/2$,

$$\left|P^f(A) - P^g(A)\right| \leq \left|P^f(U) - P^f(A)\right| +$$
$$+ \left|P^f(U) - P^g(U)\right| + |P^g(U) - P^g(A)| < \delta.$$

■

8.1. Proof of Proposition 1

Theorem 8.1. *(Fan (1953)) Let X be a compact Hausdorff space, which is a convex subset of a linear space; and let Y be a convex subset of linear space (not necessarily topologized).*[18] *Let G be a real-valued function on $X \times Y$ such that for*

[18] Fan allows for the case in which X and Y may not be subsets of linear spaces. We, however, apply his theorem only to subsets of linear spaces.

every $y \in Y$, $G(x,y)$ is lower semi-continuous with respect to x. If G is also convex with respect to x and concave with respect to y (for every $y \in Y$ and for every $x \in X$, respectively), then

$$\min_{x\in X}\sup_{y\in Y} G(x,y) = \sup_{y\in Y}\min_{x\in X} G(x,y).$$

Let $X = \Delta(\overline{F})$ be the space of all random generators of theories f from the set $\overline{F}$. Let Y be the subset of $\Delta(\Omega)$ of all probability distributions over Ω with finite support. An element P of Y can be described by a finite sequence of paths $\{\omega^1, ..., \omega^n\}$ and positive weights $\{\pi_1, ..., \pi_n\}$ that add up to one (i.e., P selects ω_i with probability π_i, $i = 1, ..., n$). Let the function $G : X \times Y \to R$ be defined by

$$G(\zeta, P) := \sum_{i=1}^{n} \pi_i \zeta\left(\left\{f \in \overline{F} : \exists_{t\in N}\ \omega_t^i \in T(f)\right\}\right). \tag{8.3}$$

That is, $G(\zeta, P)$ is the probability that the theory announced by Bob will ultimately be rejected, assuming that Bob uses the random generator of theories ζ, and Nature selects data sets according to P.

Lemma 3. *G is a lower semi-continuous function of ζ.*

Proof: Observe first that sets $\{f \in \overline{F} : \omega_t^i \in T(f)\}$ are open. Indeed, since the test T is future-independent, predictions made by a theory f up to period t determine whether a history ω_t^i belongs to $T(f)$. In addition, if any set U has the property that predictions made by a theory f up to a period t determine whether $f \in U$, then the set U is a open in $\overline{F}$, treated as a subspace of F (i.e., endowed with the product of discrete topologies).[19] Therefore, the set

$$\left\{f \in \overline{F} : \exists_{t\in N}\ \omega_t^i \in T(f)\right\} = \bigcup_{t\in N}\{f \in \overline{F} : \omega_t^i \in T(f)\}$$

is also open, as a union of open sets, and by lemma 1,

$$H(\zeta, \omega) := \zeta\left(\left\{f \in \overline{F} : \exists_{t\in N}\ \omega_t^i \in T(f)\right\}\right)$$

[19] This is the only place in the proof where we refer to the assumption that the test T is future-independent. However, this assumption is essential here. If a test T is not future-independent, then the set $\{f \in \overline{F} : \omega_t^i \in T(f)\}$ may not be open.

is a lower semi-continuous function of ζ. Thus, $G(\zeta, P)$ is a lower semi-continuous function of ζ as a weighted average of lower semi-continuous functions.$\blacksquare$

Proof of Proposition 1: We first check that the conditions of Fan's theorem are satisfied for the function G. By lemma 3, $G(\zeta, P)$ is a lower semi-continuous function of ζ. By definition, G is linear on X and Y, and so it is convex on X and concave on Y. By the Riesz and Banach-Alaoglu theorems, X is a compact space in weak$-*$ topology; it is a metric space, and therefore a Hausdorff space (see, for example, Rudin (1973), theorem 3.17).

Thus, by Fan's theorem,

$$\min_{\zeta \in \Delta(\overline{F})} \sup_{P \in Y} G(\zeta, P) = \sup_{P \in Y} \min_{\zeta \in \Delta(\overline{F})} G(\zeta, P).$$

Notice that the right-hand side of this equality falls below $\varepsilon + \delta$, as the test T is assumed not to reject the data-generating process with probability $1 - \varepsilon$. Indeed, for a given $P \in Y$, there exists a theory $g \in \overline{F}$ such that $|P(A) - P^g(A)| < \delta$ for every set $A \in \Im$. This follows from lemma 2. Let ξ be a random generator of theories such that $\xi(\{g\}) = 1$. Then

$$G(\xi, P) = P(T(g)) \leq P^g(T(g)) + |P(T(g)) - P^g(T(g))| < \varepsilon + \delta.$$

Therefore, the left-hand side does not exceed $\varepsilon + \delta$, which yields the existence of a random generator of theories $\zeta \in \Delta(\overline{F})$ such that

$$G(\zeta, P) \leq \varepsilon + \delta$$

for every $P \in Y$. Taking, for any $\omega \in \Omega$, the measure P such that $P(\{\omega\}) = 1$, we obtain:

$$\zeta(\{f : \exists_{t \in N}\ \omega_t \in T(f)\}) \leq \varepsilon + \delta.$$

$\blacksquare$

8.2. Proof of (5.3)

Take any random generator of theories ζ. Given a path $s \in \Omega$, let T_s be the test such that the rejection set $T_s(f)$ of any theory f is equal to the largest cylinder $C(s_t)$ such that $s_t = s \mid t$ and $P(C(s_t)) < 0.5$. (Assume that $T_s(f) = \emptyset$ if $P(C(s_t)) \geq 0.5$ for every t.)

By construction, T_s does not reject the data-generating process with probability 0.5, and T_s is future-independent for every $s \in \Omega$.

Given a pair of different paths (s, s'), let $D_{(s,s')}$ be the set of all theories such that P^f assigns probability 0.5 to s and to s'. There are at most countably many sets $D_{(s,s')}$ such that $\zeta\left(D_{(s,s')}\right) > 0$. Let $\bar{\Omega}$ be the (uncountable) set of paths such that if $s \in \bar{\Omega}$, $s' \in \bar{\Omega}$, $s \neq s'$, then $\zeta\left(D_{(s,s')}\right) = 0$. Given a path $s \in \bar{\Omega}$, let $D_s \subseteq F$ be the set of theories f such that P^f assigns to $\{s\}$ a probability greater than or equal to 0.5. By definition, $\zeta\left(D_s \cup D_{s'}\right) = \zeta\left(D_s\right) + \zeta(D_{s'})$ whenever $s \neq s'$. So, $\zeta\left(D_s\right) > 0$ for at most countably many paths. Let $s^* \in \Omega$ be a path such that $\zeta\left(D_{s^*}\right) = 0$. Let D_{t,s^*} be the set of theories f such that $P(C(s_t^*)) \geq 0.5$, $s_t^* = s^* \mid t$.

Given that $P(C(s_t^*)) \downarrow P(\{s^*\})$, as t goes to infinity, it follows that $D_{t,s^*} \downarrow D_{s^*}$ as t goes to infinity. Hence, $\zeta(D_{t,s^*}) \downarrow \zeta(D_{s^*}) = 0$ as t goes to infinity. By definition, $\zeta_{T_{s^*}}(s_t^*) = \zeta(F - D_{t,s^*})$. So, $\zeta_{T_{s^*}}(s_t^*)$ goes to 1 as t goes to infinity.■

8.3. Proof of propositions 2 and 3

The proofs of propositions 2 and 3 use the following three lemmas.

Lemma 4. *Let a_k, b_k, $k = 1, ..., L$, $L \in N$, be nonnegative numbers such that $\sum_{k=1}^{L} b_k \leqslant \sum_{k=1}^{L} a_k = 1$.*[20] *Then:*

1. $\sum_{k=1}^{L} a_k \log\left(\frac{b_k}{a_k}\right) \leqslant 0$.

2. $\sum_{k=1}^{L} a_k \log\left(\frac{b_k}{a_k}\right) = 0$ if and only if $a_k = b_k$, $k = 1, ..., L$.

3. For every $\varepsilon > 0$, there exists $\gamma > 0$ such that if $\sum_{k=1}^{L} a_k \log\left(\frac{b_k}{a_k}\right) > -\gamma$, then $\max_{k=1,...,L} |a_k - b_k| \leqslant \varepsilon$.

[20] We apply this lemma only to positive numbers b_k; some of the numbers a_k may be equal to zero, and then we assume that

$$a_k \log\left(\frac{b_k}{a_k}\right) = \lim_{a_k' \to 0} a_k' \log\left(\frac{b_k}{a_k'}\right) = 0.$$

.

Proof: See Smorodinsky (1971), lemma 4.5, page 20; and Lehrer and Smorodinsky (1996), lemma 2.■

Let E^P and VAR^P be the expectation and variance operator associated with some $P \in \Delta(\Omega)$. Let $(X_i)_{i=1}^{\infty}$ be a sequence of random variables such that X_i is $\Im_i-$measurable and its expectation conditional on $\Im_{i-1}$ is zero (i.e., $E^P\{X_i \mid \Im_{i-1}\} = 0$). Moreover, let the sequence of conditional variances $VAR^P\{X_i \mid \Im_{i-1}\}$ be uniformly bounded (i.e., $VAR^P\{X_i \mid \Im_{i-1}\} < M$ for some $M > 0$). Let $\mathcal{I}_i$ be a $\Im_{i-1}-$measurable function that takes values 0 or 1. We define

$$S_m := \sum_{i=0}^{m} X_i \mathcal{I}_i \text{ and } \mathcal{T}_m := \sum_{i=0}^{m} \mathcal{I}_i. \tag{8.4}$$

Lemma 5. *For every $k \in N$ and $\delta > 0$,*

$$P\left(\left\{s \in \Omega : \max_{m \ s.t.\ 1 \leq \mathcal{T}_m(s) \leq k} |S_m(s)| > \delta\right\}\right) \leq \frac{kM}{\delta^2}.$$

Proof: For every $l \in \{1,..,k\}$, let $y_l : \Omega \longrightarrow N \bigcup \{\infty\}$ be the random variable such that $\mathcal{T}_{y_l(s)}(s) = l$ and $\mathcal{T}_y(s) < l$ for every $y < y_l(s)$, $y \in N$. That is, consider the series of zeroes and ones given by $\{\mathcal{I}_0(s), \mathcal{I}_1(s), ...\}$. Then $y_l(s)$ indicates the stage i in which $\mathcal{I}_i(s) = 1$ for the $l-th$ time. If the series $\{\mathcal{I}_0(s), \mathcal{I}_1(s), ...\}$ contains strictly less than l ones, then $y_l(s)$ is arbitrarily defined as ∞.

Let $\vartheta_i = X_i \mathcal{I}_i$; $\xi_l = \vartheta_{y_l}$; $X_\infty := 0$ and $S_n^* := \sum_{l=1}^{n} \xi_l$.

For all realizations $(\dot{y}_1, ..., \dot{y}_l)$ of $(y_1, ..., y_l)$, $\dot{y}_l > \dot{y}_i$ whenever $i < l$ and $\dot{y}_i \neq \infty$. Thus,

$$E^P\left\{\xi_l \mid \xi_1, ..., \xi_{l-1},\ (y_1, ..., y_l) = (\dot{y}_1, ..., \dot{y}_l)\right\} =$$

$$E^P\{\vartheta_{\dot{y}_l} \mid \vartheta_{\dot{y}_i},\ i = 1, ..., l-1,\ \dot{y}_i \neq \infty,\ (y_1, ..., y_l) = (\dot{y}_1, ..., \dot{y}_l)\} = 0,$$

where the last equality holds because $E^P\{\vartheta_{\dot{y}_l} \mid \Im_{\dot{y}_l-1}\} = 0$ ($\mathcal{I}_{\dot{y}_l-1}$ is $\Im_{\dot{y}_l-1}-$measurable and $E^P\{X_{\dot{y}_l} \mid \Im_{\dot{y}_l-1}\} = 0$). In addition, $\vartheta_{\dot{y}_i}$ is $\Im_{\dot{y}_l-1}-$measurable for $i = 1, ..., l-1$, and $(y_i, ..., y_l) = (\dot{y}_1, ..., \dot{y}_l)$ is also in $\Im_{\dot{y}_l-1}$. Thus,

$$E^P\left\{\xi_l \mid \xi_1, ..., \xi_{l-1}\right\} = 0. \tag{8.5}$$

By Kolmogorov's inequality (see Shiryaev (1996), Chapter IV, §2),

$$P\left(\left\{s \in \Omega : \max_{n \ s.t. \ 1 \le n \le k} |S_n^*(s)| > \delta\right\}\right) \le \frac{Var(S_k^*)}{\delta^2}. \tag{8.6}$$

(Shiryaev (1996) states (8.6) for independent random variables, but the proof requires only (8.5). Now, $Var(\xi_l \mid y_l = \dot{y}_l, \dot{y}_l \ne \infty) \le Var(X_{\dot{y}_l} \mid y_l = \dot{y}_l, \dot{y}_l \ne \infty) \le M$, where the last inequality holds because $y_l = \dot{y}_l$ belongs to $\Im_{\dot{y}_l - 1}$. So, $Var(\xi_l) \le M$. This and (8.5) give

$$Var(S_k^*) \le kM. \tag{8.7}$$

Finally,

$$\left\{s \in \Omega : \max_{m \ s.t. \ 1 \le \mathcal{T}_m(s) \le k} |S_m(s)| > \delta\right\} \subseteq \left\{s \in \Omega : \max_{n \ s.t. \ 1 \le n \le k} |S_n^*(s)| > \delta\right\}.$$

This inclusion holds because if $\mathcal{T}_m(s) \le k$, then $S_m(s) = S_n^*(s)$ for some $n \le k$. The conclusion now follows from (8.6) and (8.7).■

Given the definitions in (8.4), we define $Y_m := \frac{S_m}{\mathcal{T}_m}$ if $\mathcal{T}_m \ne 0$ (and $Y_m := 0$ whenever $\mathcal{T}_m = 0$).

Lemma 6. *For every $\varepsilon' > 0$ and $j \in N$, there exists $\bar{m}(j, \varepsilon') \in N$ such that*

$$P\left(\left\{s \in \Omega : \forall_{m \text{ such that } \mathcal{T}_m(s) \ge \bar{m}(j,\varepsilon')} \quad |Y_m(s)| \le \frac{1}{j}\right\}\right) > 1 - \varepsilon'.$$

Proof: Let $M_n := \max_{m \text{ such that } 2^n < \mathcal{T}_m \le 2^{n+1}} Y_m$. We assume that $M_n = 0$ whenever no m satisfies $2^n < \mathcal{T}_m \le 2^{n+1}$. Then,

$$P\left(\left\{s \in \Omega : M_n(s) > \frac{1}{j}\right\}\right) \le P\left(\left\{s \in \Omega : \max_{m \text{ such that } 2^n < \mathcal{T}_m \le 2^{n+1}} |S_m(s)| > \frac{1}{j}2^n\right\}\right)$$

$$\le P\left(\left\{s \in \Omega : \max_{m \text{ such that } 1 \le \mathcal{T}_m \le 2^{n+1}} |S_m(s)| > \frac{1}{j}2^n\right\}\right) \le 2Mj^2\frac{2^n}{4^n} = 2Mj^2\frac{1}{2^n}.$$

(The last inequality follows from lemma 5.) Therefore,

$$\sum_{n=m^*}^{\infty} P\left(\left\{s \in \Omega : M_n(s) > \frac{1}{j}\right\}\right) \le 2Mj^2 \sum_{n=m^*}^{\infty} \frac{1}{2^n} < \varepsilon'$$

(for a sufficiently large m^*).

Let $\bar{m}(j, \varepsilon') = 2^{m^*}$ for this sufficiently large m^*. By definition,

$$\Omega - \left\{ s \in \Omega : \forall_{m \text{ such that } \mathcal{T}_m \geq \bar{m}(j,\varepsilon')} \quad |Y_m(s)| \leq \frac{1}{j} \right\} \subseteq \bigcup_{n=m^*}^{\infty} \left\{ s \in \Omega : M_n(s) > \frac{1}{j} \right\}.$$

Hence,

$$P\left(\left\{ s \in \Omega : \forall_{m \text{ such that } \mathcal{T}_m \geq \bar{m}(j,\varepsilon')} \quad |Y_m(s)| \leq \frac{1}{j} \right\} \right) > 1 - \varepsilon'.$$

■

Remark 1. The quantity $\bar{m}(j, \varepsilon')$ obtained in Lemma 6 depends on M but is independent of the theory P. This feature of Lemma 6 is used to ensure that the tests constructed below are future-independent.

Proof of Proposition 2: Let

$$Z_t(s) = \log\left(\frac{h^{f_\delta}(s_t)}{h^f(s_t)} \right), \; s_t = s \mid t,$$

where

$$h^f(s_t) = f(s_t)^{s^t}(1 - f(s_t))^{1-s^t}$$

and

$$h^{f_\delta}(s_t) = f_\delta(s_t)^{s^t}(1 - f_\delta(s_t))^{1-s^t}.$$

Then, for some $\eta > 0$ and for some $M > 0$,

$$E^{P^f}\{Z_t \mid \Im_{t-1}\} < -\eta \text{ and } VAR^{P^f}\{Z_t \mid \Im_{t-1}\} < M.$$

The first inequality (on conditional expectation) follows directly from assertion 3 in lemma 4. The second inequality (on conditional variance) follows directly from the fact that $h^{f_\delta}(s_t) \in [\delta, 1-\delta]$ and from the fact that the functions $-plog(p)$ and $p\,(log(p))^2$ are bounded on $[0, 1]$.

Let $X_i = Z_i - E^{P^f}\{Z_i \mid \Im_{i-1}\}$. Let $j \in N$ be a natural number such that $\frac{1}{j} < \frac{\eta}{4}$. Let $\bar{m}(j, \varepsilon)$ be defined as in lemma 6 (in the special case that $\mathcal{I}_i \equiv 1$ for $i = 1, ..., m$ and, hence, $\mathcal{T}_m = m$). The test T is defined by

$$C(s_m) \subseteq T(f) \text{ if } \sum_{k=0}^{m} Z_k(s) > -(m+1)\frac{\eta}{2} \text{ whenever } m \geq \bar{m}(j, \varepsilon) \text{ and } s_m = s \mid m.$$

Note that

$$\left\{ s \in \Omega : \forall_{m \geq \bar{m}(j,\varepsilon)} \quad \left| \frac{1}{m+1} \sum_{k=0}^{m} \left(Z_k(s) - E^{P^f} \{ Z_k \mid \Im_{k-1} \} (s) \right) \right| \leq \frac{1}{j} \right\} \subseteq \Omega - T(f)$$

because

$$\frac{1}{m+1} \sum_{k=0}^{m} Z_k(s) \leq \frac{1}{j} - \eta < -\frac{\eta}{2}$$

implies

$$\sum_{k=0}^{m} Z_k(s) < -(m+1)\frac{\eta}{2}.$$

By lemma 6,

$$P^f \left((T(f))^c \right) > 1 - \varepsilon.$$

Hence, the test T does not reject the truth with probability $1-\varepsilon$. By construction, the test T is future-independent. (See remark 1; note that whether the test T rejects a theory at s_t depends only on the forecasts $f(s_k)$, $s_k = s_t \mid k$, $k < t$.) Finally, observe that

$$\sum_{k=0}^{t} Z_k(s) = \log \left(\frac{P^{f_\delta}(s_t)}{P^f(s_t)} \right), \;\; s_t = s \mid t.$$

Hence, if $s \notin T(f)$, then

$$\log \left(\frac{P^{f_\delta}(s_t)}{P^f(s_t)} \right) \underset{t \to \infty}{\longrightarrow} -\infty,$$

which implies that $s \notin R(f)$.■

Proof of Proposition 3: Let

$$X_t^i(s) = [f(s_t) - s^t] \mathcal{I}_t^i, \;\; s_t = s \mid t,$$

and

$$S_m^i := \sum_{t=0}^{m} X_t^i; \;\; \mathcal{T}_m^i = \sum_{t=0}^{m} \mathcal{I}_t^i \text{ and } Y_m^i := \frac{S_m^i}{\mathcal{T}_m^i}.$$

Now, let $\varepsilon_{j,i}$, $(j,i) \in N^2$ be such that $\varepsilon_{j,i} > 0$ and

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varepsilon_{j,i} < \varepsilon.$$

Given that $E^{P_f}\left\{X_t^i \mid \Im_{t-1}\right\} = 0$ and $VAR^P\left\{X_t^i \mid \Im_{i-1}\right\}$ are uniformly bounded, let $\bar{m}(j, \varepsilon_{j,k})$ be defined as in lemma 6. The test T' is defined by

$$C(s_m) \subseteq T'(f) \text{ if } \left|Y_m^i(s)\right| > \frac{1}{j} \text{ and } \mathcal{T}_m^i(s) \geq \bar{m}(j, \varepsilon_{j,i}).$$

By lemma 6,

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P^f\left(\left\{s \in \Omega : \left|Y_m^i(s)\right| > \frac{1}{j} \text{ for some } m \text{ s.t. } \mathcal{T}_m^i(s) \geq \bar{m}(j, \varepsilon_{j,i})\right\}\right) < \varepsilon.$$

As a result,

$$P^f\left(\left(T'(f)\right)^c\right) \geq 1 - \varepsilon.$$

Hence, T' does not reject the truth with probability $1 - \varepsilon$. Moreover, notice that $s \notin T'(f)$ implies that

$$\left|Y_m^i(s)\right| \leq \frac{1}{j} \text{ for all } m \text{ such that } \mathcal{T}_m^i(s) \geq \bar{m}(j, \varepsilon_{j,i}) \text{ and } (j,i) \in N^2.$$

So, for all $i \in N$, $|Y_m^i(s)| \longrightarrow 0$ whenever $\mathcal{T}_m^i(s) \underset{m\to\infty}{\longrightarrow} \infty$. By construction, the test T' is future-independent (see remark 1).■

References

[1] Al-Najjar, N. and J. Weinstein (2007) "Comparative Testing of Experts," mimeo.

[2] Cesa-Bianchi, N. and G. Lugosi (2006): *Prediction, Learning, and Games*, Cambridge University Press.

[3] Dekel, E. and Y. Feinberg (2006) "Non-Bayesian Testing of a Stochastic Prediction," *Review of Economic Studies*, **73**, 893-906.

[4] Dow, J. and G. Gorton (1997) "Noise Trading, Delegated Portfolio Management, and Economics Welfare," *Journal of Political Economy*, **105**, 1024-1050.

[5] Dudley, R.M. (1989): *Real Analysis and Probability*, Wadsworth Inc., Belmont, California.

[6] Ellsberg, D. (1961) "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, **75**, 643-669.

[7] Fan, K. (1953) "Minimax Theorems," *Proceedings of the National Academy of Science U.S.A.*, **39**, 42-47.

[8] Feinberg, Y. and C. Stewart (2007) "Testing Multiple Forecasters," mimeo.

[9] Fortnow, L. and R. Vohra (2007) "The Complexity of Forecast Testing," mimeo.

[10] Foster, D. and R. Vohra (1998) "Asymptotic Calibration," *Biometrika*, **85**, 379-390.

[11] Fudenberg, D. and D. Levine (1999) "An Easier Way to Calibrate," *Games and Economic Behavior*, **29**, 131-137.

[12] Gilboa, I. and D. Schmeidler (1989) "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, **18**, 141–153.

[13] Keane, M. and D. Runkle (1990) "Testing the Rationality of Price Forecasts: New Evidence from Panel Data," *American Economic Review*, **80**, 714-735.

[14] Keane, M. and D. Runkle (1998) "Are Financial Analysts' Forecasts of Corporate Profits Rational?" *Journal of Political Economy*, **106**, 768-805.

[15] Knight, F. (1921): *Risk, Uncertainty and Profit*, Houghton Mifflin, Boston.

[16] Laster, D., P. Bennett, and I. Geoum (1999) "Rational Bias in Macroeconomic Forecasts," *Quarterly Journal of Economics*, **114**, 293-318.

[17] Lehrer, E. (2001) "Any Inspection Rule is Manipulable," *Econometrica* **69**, 1333-1347.

[18] Lehrer, E. and R. Smorodinsky (1996) "Compatible Measures and Merging," *Mathematics of Operations Research*, **21**, 697-706.

[19] LeRoy, S. and L. Singell (1987) "Knight on R,isk and Uncertainty," *Journal of Political Economy*, **95**, 394-406.

[20] Lowell, M. (1986) "Tests of the Rational Expectations Hypothesis," *American Economic Review*, **76**, 110-154.

[21] Morgan, J. and P. Stocken (2003) "An Analysis of Stock Recommendations," *RAND Journal of Economics*, **34**, 380-391.

[22] Olszewski, W. and A. Sandroni (2007a) "Contracts and Uncertainty" *Theoretical Economics,* **2**, 1-13.

[23] Olszewski, W. and A. Sandroni (2007b) "A Nonmanipulable Test," *The Annals of Statistics*, forthcoming.

[24] Olszewski, W. and A. Sandroni (2007c) "Falsifiability," mimeo.

[25] Olszewski, W. and A. Sandroni (2007d) "Strategic Manipulation of Empirical Tests," mimeo.

[26] Ottaviani ,M. and P. Sørensen, (2006) "The Strategy of Professional Forecasting," *Journal of Financial Economics*, **81**, 441-466.

[27] Pigou, A. (1927): *Industrial Fluctuations*, McMillan, London.

[28] Rudin, W. (1973): *Functional Analysis*, McGraw-Hill, Inc.

[29] Shiryaev, A. (1996): *Probability*, Springer Verlag, New York.

[30] Sandroni, A. (2003) "The Reproducible Properties of Correct Forecasts," *International Journal of Game Theory*, **32**, 151-159.

[31] Sandroni, A., R. Smorodinsky and R. Vohra (2003), "Calibration with Many Checking Rules," *Mathematics of Operations Research*, **28**, 141-153.

[32] Shmaya, E. (2008) "Many Inspections are Manipulable," mimeo.

[33] Smorodinsky, M. (1971): *Ergodic Theory, Entropy*, Lecture Notes in Mathematics, Springer-Verlag, New York.

[34] Trueman, B. (1988) "A Theory of Noise Trading in Securities Markets," *Journal of Finance*, **18**, 83-95.

[35] Vovk, V. and G. Shafer (2005) "Good Randomized Sequential Probability Forecasting is Always Possible," *Journal of the Royal Statistical Society Series B*, **67**, 747-763.