View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 08-016

"Falsifiability"

by

Wojciech Olszewski and Alvaro Sandroni

http://ssrn.com/abstract=1133128

Falsifiability*

Wojciech Olszewski[†] and Alvaro Sandroni[‡]

March 12, 2008

Abstract

We examine the fundamental concept of Popper's falsifiability within an economic model in which a tester hires a potential expert to produce a theory. Payments are made contingent on the performance of the theory vis-a-vis future realizations of the data. We show that if experts are strategic, then falsifiability has no power to distinguish legitimate scientific theories from worthless theories. We also show that even if experts are strategic there are alternative criteria that can distinguish legitimate from worthless theories.

*We are grateful to Eddie Dekel, Yossi Feinberg, Nabil Al-Najjar, Sandeep Baliga, Roger Myerson, and Rakesh Vohra for comments and conversations. We also thank workshop participants at Brown, Caltech, Columbia, Cornell, Yale, Harvard - MIT, Northwestern, Penn State, Princeton, Stanford - GSB, Rice, University of Bonn, University of British Columbia, University of California at Berkeley, UCLA, University of Chicago, Instituto de Matematica Pura e Aplicada (IMPA), University of Texas at Austin, University of Toronto, Washington University at St. Louis, and at the North American Summer Meeting of Econometric Society in Minneapolis, in 2006.

†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

The publication of "*The Logic of Scientific Discovery*" by Karl Popper (1968, first published in 1935) was a transformative event because it expressed clearly the concept of falsifiability. Popper was interested in demarcation criteria that differentiate scientific ideas from nonscientific ideas (and hence give meaning to the term *scientific*). He argued that science is not a collection of facts, but a collection of statements that can be falsified (i.e., rejected by the data). His leading example of a scientific statement was "All swans are white." This example shows the asymmetry between verification and refutation: no matter how many white swans are observed, one cannot be certain that the next one will be white, but the observation of a single nonwhite swan shows the statement to be false.

Some critics (most notably Kuhn (1962)) contend that the history of science contains many instances that seem inconsistent with Popper's criteria.[1] However, falsifiability remains a central concept for several reasons. First, it presents a guiding principle on how science should be conducted: scientists should deliver falsifiable theories that can be tested empirically. (An example of falsifiability as a guide to research is the debate on whether general equilibrium theory is testable; see Carvajal *et al.* (2004) for a review article.) Moreover, falsifiability delivers criteria for what should be taught under the rubric of science. (See ruling by U.S. District Court Judge William Overton, largely based on falsifiability, against the teaching of intelligent design as science in Arkansas public schools, "The Arkansas Balanced Treatment Act" in *McLean* v. *Arkansas Board of Education*, Act 590 of the Acts of Arkansas of 1981.) Finally, falsifiability is an important requirement in the U.S. legal system's Daubert standard, which is designed to rule as inadmissible any testimony by expert witnesses that this standard evaluates as "junk science". (See the legal precedent set in 1993 by the Supreme Court, *Daubert* v. *Merrell Dow Pharmaceuticals*, 509 U.S. 579.)

Although falsifiability has been employed as a guiding principle in legal proceedings, economics, and science in general, it has not, to our knowledge, been formally analyzed to determine whether it can distinguish useful ideas from worthless ones in a fully-fledged economic model in which agents may misreport what they know. An objective of this paper is to deliver such a model. Before continuing with a description of the model, we stress that what Popper means by falsifiability is the feasibility of *conclusive* empirical rejection. This is often regarded as too strong because it dismisses probabilistic statements that attach strictly positive probability to an event and its complement. Falsifiable probabilistic statements must attach zero probability to some event. Popper (1968) was aware of this objection. He wrote:

[1] See Lakatos' article in Lakatos and Musgrave (1970) for an attempt to reconcile Popper's view on the logic of science with Kuhn's view on its history.

"For although probability statements play such a vitally important rôle in empirical science, they turn out to be in principle *impervious* to strict *falsification*. Yet this very stumbling block will become a touchstone upon which to test my theory, in order to find out what it is worth."

An adaptation of falsifiability designed to partially accommodate probabilities is provided by Cournot's (1843) principle, which states that unlikely events must be treated as impossible.[2] However, for reasons that will become clear at the end of this introduction, we refer to falsifiability in the strict Popperian sense.

We study a contracting problem between an expert and a tester. The expert, named Bob, announces a theory which is empirically tested by the tester named Alice. Like Popper, we assume that the main purpose of a theory is to make predictions. We define a theory as a mechanism that takes the available data as an input and returns, as an output, the probabilities of future outcomes. Before data are observed, Bob decides whether to announce a theory. If he does, he cannot revise his theory later. As data unfold, Alice tests Bob's theory according to the observed history.

Alice does not have a prior over the space of theories and is too ill-informed to formulate probabilities over the relevant stochastic process (i.e., she faces Knightian uncertainty). An expert could deliver these probabilities to her. If a theory is an accurate description of the data-generating process, then she benefits from the theory because it tells her the relevant odds (i.e., it replaces her uncertainty with common risk).[3] The difficulty is that Alice does not know if Bob is an informed expert who can deliver the data-generating process, or if he is an uninformed agent who knows nothing about the relevant process and who can only deliver theories unrelated to it.

We assume that Alice takes Popper's methodology seriously and demands a falsifiable theory, i.e., a theory such that it predicts that some finite continuation of any finite history has zero probability. Alice pays Bob a small reward which gives utility $u > 0$, if he announces a falsifiable theory. In order to discourage Bob from delivering an arbitrary falsifiable theory, Alice proposes a contract that stipulates a penalty if Bob's theory is falsified in the future, i.e., if some history deemed impossible by the theory is eventually observed. This penalty gives Bob disutility $d > 0$. Bob receives no reward and no penalty if he does not announce any theory or if he announces a nonfalsifiable theory (in which case his utility is zero).

[2] Cournot (1843) was perhaps the first to relate the idea that unlikely events will not occur to the empirical meaning of probability. He wrote, "The physically impossible event is therefore the one that has infinitely small probability, and only this remark gives substance - objective and phenomenal value - to the theory of mathematical probability."

[3] Risk refers to the case where perceived likelihoods can be represented by a probability. Uncertainty refers to the case where the available information is too imprecise to be summarized by a probability. This distinction is traditionally attributed to Knight (1921). However, LeRoy and Singell (1987) argue that Knight did not have this distinction in mind.

We now make a series of assumptions that are not meant to be realistic. Rather, they should be interpreted as an extreme case in which our result holds, so that it still holds under milder and more realistic conditions. These assumptions are: Alice eventually has an unbounded data set at her disposal and never stops testing Bob's theory unless it is rejected. Bob does not discount the future and so his contingent payoffs are $u - d$ if his theory is eventually rejected and u if his theory is never falsified. Bob's liabilities are not limited and so the penalty d for having delivered a rejected theory can be made arbitrarily large, whereas the payoff u for announcing a falsifiable theory can be made arbitrarily small. Bob has no knowledge whatsoever of the data-generating process and so Bob, like Alice, also faces uncertainty and cannot determine the probability that any falsifiable theory will be rejected. Finally, Bob is an extreme pessimist and computes his payoff under the worst-case scenario, i.e., under the future realization of data that gives him minimal payoff.

Our last assumption is so extreme that it seems to settle the matter trivially. Assume that Bob announces any falsifiable theory f deterministically. There are many histories that falsify f. The worst-case scenario is the observation of one of them. Hence, the payoff for delivering any theory f deterministically is $u - d$. As long as the penalty for delivering a theory rejected by the data exceeds the reward for announcing a falsifiable theory, i.e., as long as $d > u$, Bob is better off not announcing any theory deterministically. So, it seems as if Alice can avoid the error of receiving a theory produced by an uninformed expert. However, Bob still has one remaining recourse. He can randomize (only once) at period zero and select his falsifiable theory according to this randomization. This suffices. We show that no matter how large the penalty d, and no matter how low the reward u, there exists a way to strategically select falsifiable theories at random (i.e., according to specific odds that we describe explicitly) such that for *all* possible future realizations of the data, the expected utility of the random announcement of a theory exceeds the utility of not announcing any theory at all. At the heart of argument is the demonstration that it is possible to produce falsifiable theories (at random) that are unlikely to be falsified, no matter how the data unfold in the future. Thus, even Popper's strict falsification criterion (which requires a theory to assert that some events are impossible) cannot deter even the most ignorant expert because the feasibility of conclusive empirical rejection can be removed by strategic randomization.

Our result shows a contrast between the case in which theories are exogenously given and the case in which theories may have been strategically produced. For a given theory, falsifiability makes a fundamental conceptual distinction: falsifiable theories can be conclusively rejected and nonfalsifiable theories cannot. In contrast, when theories are produced by a potentially strategic expert, falsifiability does not impose significant constraints on uninformed experts and, hence, cannot determine whether the expert is informed about the data-generating process.

The failure of falsifiability in delivering useful criteria (when experts are strategic) motivates an analysis of the merits of verification and refutation as guiding principles for empirical research. We now consider general contracts between Alice and Bob. In a *verification contract*, Alice pays Bob when the observed data is deemed consistent with Bob's theory (but Bob may pay Alice when he announces his theory). We make no restrictions on which data the contract can define as consistent with each theory. In a *refutation contract,* Bob pays Alice if the observed data is deemed inconsistent with the Bob's theory (but Alice may pay Bob when he announces his theory). So, the falsification contract is a special refutation contract in which Alice pays nothing for nonfalsifiable theories, pays positive amounts for falsifiable theories, and Bob pays Alice if his theory is conclusively rejected. In other refutation contracts, Bob may pay contingent on data that does not conclusively reject his theory.

We are interested in a screening contract that Bob, if informed, accepts and Bob, if uninformed, does not accept. If informed, Bob faces risk and evaluates his prospects by standard expected utility. If uninformed, Bob faces uncertainty and evaluates his prospects based on his expected utility computed at the worst possible future realization of the data. A contract is *accepted by the informed expert* if Bob gets positive expected utility if he announces the actual data-generating process. A contract is *accepted by the uninformed expert* if Bob can select theories at random so that no matter which data is eventually observed, his expected utility is positive.

We show that if the informed expert accepts any verification contract then the uninformed expert also accepts this contract. This result implies that it is possible to produce theories (at random) that are likely to prove, in the future, to be supported the data, no matter how the data unfold. Hence, when experts are potentially strategic, both Popper's falsifiability and verification do not provide useful criteria for the same reason: they cannot screen informed and uninformed experts. In contrast, we show a refutation contract that can screen informed and uninformed experts (i.e., informed experts accept the contract and uninformed experts do not accept it). This contract is based on an empirical test and a penalty for Bob if his theory is rejected by the test. The method determining how to refute theories is novel and is not based on falsification nor on any standard statistical test.

As argued by Popper, there is an asymmetry between verification and refutation. However, the criteria showing this asymmetry is not Popper's falsifiability. It is the existence of a screening refutation contract and the inexistence of a screening verification contract. These results deliver an original argument supporting the idea that refutation is a better maxim, for empirical research, than verification.

1.1. Related literature

The idea that an ignorant agent can strategically avoid rejection of an empirical test 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 (2006, 2007b)). Some of these results are reviewed in Cesa-Bianchi and Lugosi (2006).[4]

However, the idea that the concept of falsification can be analyzed as an empirical test, and that this test can be manipulated by ignorant experts is novel. The classes of tests considered in the literature exclude the main test of interest, i.e., the empirical test defined by falsification.

1.2. Motivating Idea: Strategic Randomization

Consider a simple two period model. In period one, a ball is drawn from an urn. The balls are of n possible colors. Alice does not know the composition of the urn. If informed, Bob has seen the composition of the urn. If uninformed, he has not.

Alice is willing to pay to become informed (i.e., to learn the composition of the urn), but she is concerned that Bob may be uninformed and would just give her an arbitrary distribution. Alice wants to discourage such a fraud. A difficulty is that if Bob tells Alice that any color is possible, then she cannot reject Bob's claim. So, Alice takes Popper's advice and proposes a contract to Bob. If he accepts he must deliver to Alice, at period zero, a falsifiable distribution (i.e., a probability measure over the n colors that assigns zero probability to at least one color). So, Bob must claim that at least one color is impossible. If none of the (allegedly) impossible colors are observed, then Bob's payoffs are $u > 0$. If an (allegedly) impossible color is observed, then Bob's payoffs are $u - d < 0$. If Bob does not accept Alice's contract, then his payoff is zero.

By requiring a falsifiable distribution, Alice may induce Bob to misrepresent what he knows (when all colors are possible). However, at least one color must have probability smaller or equal to $1/n$. So, as long as

$$u \geq \frac{d}{n}, \tag{1.1}$$

Bob, whenever informed, is better off accepting Alice's contract and asserting that a color, among those least likely to occur, is impossible (rather than not accepting Alice's contract). In addition, Bob has no incentive to misrepresent the relative odds of any colors other than the one he must claim to be impossible.

[4] See Dekel and Feinberg (2006), Kalai, Lehrer, and Smorodinsky (1999), Rustichini (1999), Lehrer and Solan (2003), Hart and Mas-Colell (2000), Olszewski and Sandroni (2007a), Al-Najjar and Weinstein (2006), Feinberg and Stuart (2006), and Fortnow and Vohra (2006) for related results.

Now assume that Bob is uninformed. Then, he cannot determine the relevant odds. Let us say that Bob determines the value of Alice's contract by the expected utility obtained under the worst possible outcome. This is the most pessimistic behavioral rule among those axiomatized by Gilboa and Schmedler (1989).

Assume that Bob announces any falsifiable distribution deterministically. In the worse-case scenario, Bob is rejected. So, in the worse-case scenario, his expected utility is negative. Hence, Bob cannot accept Alice's contract and deliver a falsifiable distribution deterministically. This suggests that Alice can screen informed and uninformed experts at least when Bob is extremely averse to uncertainty. However, this is not true. It is easy to see that Bob can produce a falsifiable distribution, at random, and obtain expected positive payoff, no matter what is the true composition of the urn.

Let p_i be a probability distribution that is falsified if and only if color i is realized (i.e., p_i assigns zero probability to i and positive probability to $j = 1, ..., n,\ j \neq i$). Assume that Bob selects each p_i, $i = 1, .., n$, with probability $1/n$. For *any* given color, Bob's realized probability measure is falsified with probability $1/n$. Hence, conditional on any composition of the urn, Bob's expected utility is nonnegative when (1.1) is satisfied. If Bob, when informed, accepts Alice's contract then Bob, even if completely uniformed and extremely averse to uncertainty, also accepts Alice's contract.

The argument above is simple, but it suggests an important implication. As pointed out in the introduction, falsifiability is a criterion proposed by Popper to differentiate scientific ideas from nonscientific ideas. Indeed, a falsifiable distribution can be conclusively rejected by the data and a nonfalsifiable one cannot. However, when theories are produced by experts who can misrepresent what they know, it is unclear whether falsifiability delivers useful criteria.

The argument in the example above can be made very simple because Alice has only one data point at her disposal. In the relevant case, Alice has many data points available to her. Then, the puzzle of which criteria can be used to distiguish useful from worthless theories (when experts are strategic) is more interesting. However, before addressing the large data set case, let us still consider the one data point example, but now assume that Bob must assert that more than one color is impossible. This stronger requirement goes beyond the concept of falsifiability and is still not able to screen informed and uninformed experts. Let us say that Bob must assert that j colors are impossible. Then, an informed expert always accepts Alice contract if (1.1) is satisfied when $1/n$ is replaced with j/n . Under this assumption, an uninformed expert can randomize and obtain a positive expected payoff no matter what is the composition of the urn (a direct proof is available upon request, but this is a corollary of a general result, proposition 2, presented in section 4). Hence, requiring that Bob must assert that multiple colors are impossible does not make the test able to screen informed and uninformed experts.

2. Basic Definitions

We now consider a model with many periods so that Alice will eventually have a large data set available to her. Each period, one outcome, out of a finite set S with n elements, is observed. Let $\Delta(S)$ denote the set of probability measures on S. An element of $\Delta(S)$ is called a *distribution over outcomes*. Let S^t denote the Cartesian product of t copies of S, let $\overline{S} = \bigcup_{t\geq 0} S^t$ be the set of all finite histories.[5] We also define $\Omega = S^\infty$ as the set of *paths*, i.e., infinite histories, and the set $\Delta(\Omega)$ of probability measures over Ω.[6]

Any function

$$f : \overline{S} \longrightarrow \Delta(S)$$

that maps finite histories into distributions over outcomes can be interpreted as follows: f takes data (outcomes up to a given period) as an input and returns a probabilistic forecast for the following period as an output. To simplify the language, any such function f is called a *theory*. Thus, a theory is defined by its predictions.

Any theory f defines a probability measure P_f. The probability of each finite history $(s^1, ..., s^m) \in S^m$ can be computed as follows: Given a finite history $\overline{s}$ and an outcome $s \in S$, let the probability of s conditional on $\overline{s}$ be denoted by $f(\overline{s})[s]$. Then, the probability P_f of $(s^1, ..., s^m)$ is equal to a product of probabilities

$$P_f(s^1, ..., s^m) = f(\varnothing)[s^1] \cdot \prod_{k=2}^{m} f\left(s^1, ..., s^{k-1}\right)[s^k].$$

Definition 1. *A theory f is falsifiable if every finite history $(s^1, ..., s^t) \in S^t$ has an extension $(s^1, ..., s^t, s^{t+1}, ..., s^m)$ such that*

$$P_f(s^1, ..., s^m) = 0.$$

A theory is falsifiable if, after any finite history, there is a finite continuation history that the theory deems impossible. Let $\mathcal{F}$ be the set of falsifiable theories. Given $f \in \mathcal{F}$, let R_f be the union of all finite histories to which P_f assigns zero probability. So, R_f is the set of all finite histories that bluntly contradict (or, equivalently, falsify) the theory $f \in \mathcal{F}$.

Assume that Alice demands a falsifiable theory. Alice pays for the theory, but if it is rejected (i.e., if Bob announces theory $f \in \mathcal{F}$ and data in R_f is observed), then Bob

---

[5] By convention, $S^0 = \{\varnothing\}$.

[6] We need a $\sigma-$algebra on which probability measures are defined. Let a *cylinder* with base on $(s^1, ..., s^m)$ be the set of all paths such that the first m elements are $(s^1, ..., s^m)$. We endow Ω with the smallest $\sigma-$algebra that contains all such cylinders. We also endow Ω with the product topology (the topology that comprises unions of cylinders).

receives a large penalty that gives disutility greater than the utility of the payment. The question is whether the falsifiability requirement dissuades an uninformed expert (who does not know the data-generating process) from announcing a theory.

We show next that an uninformed expert can strategically produce falsifiable theories, at random, with odds designed so that, with arbitrarily high probability, the realized falsifiable theory will not be falsified, regardless of which data is eventually observed. Hence, the feasibility of falsification is virtually eliminated by strategic randomization.

3. Falsification and Strategic Randomization

In this section, we maintain the basic infinite horizon model in which Alice adopts Popper's method and demands a falsifiable theory $f \in \mathcal{F}$. As an incentive, Alice pays Bob a (small) amount of money (which gives Bob utility $u > 0$) if Bob delivers a falsifiable theory. However, if a finite history in R_f is observed, then Bob's theory is falsified and he pays a penalty which gives him disutility $d > 0$. Liabilities are not limited and so d can be arbitrarily large.

Bob does not have to deliver a theory, but if he accepts Alice's conditions, he must deliver a falsifiable theory before any data are observed. We assume that Bob does not discount the future (although our result still holds if he does). So Bob's contingent payoffs are $u > 0$ if his theory is never falsified and $u - d$ if his theory is contradicted at some time t. Formally, consider the contract in which only falsifiable theories can be delivered and delivering a theory which is (later) contradicted by data is punished. If Bob accepts the contract, he announces a theory $f \in \mathcal{F}$. If a path $s = (s^1, s^2, ...) \in \Omega$ is observed, Bob's contingent net payoff is

$$U(f,s) = \begin{array}{cc} u - d & \text{if, for some period } t,\ (s^1, ..., s^t) \in R_f; \\ u & \text{otherwise.} \end{array}$$

If Bob does not accept the contract, then no theory is announced and his payoff is zero.

We assume that Bob is utterly ignorant about the relevant probabilities. So, like Alice, Bob cannot determine the odds according to which any given theory will be falsified. Bob can select his theory randomly according to a probability measure $\zeta \in \Delta(\mathcal{F})$. Given that Bob randomizes only once (at period zero), Alice cannot tell whether the theory she receives was produced deterministically or selected randomly. Finally, we assume that Bob evaluates his prospects based on the worse-case scenario, i.e., based on the path $s \in \Omega$ that gives him minimal expected utility. Formally, Bob's payoff is

$$V(\zeta) \ = \ \inf_{s \in \Omega} E^{\zeta} U(f,s),$$

where E^{ζ} is the expectation operator associated with $\zeta \in \Delta(\mathcal{F})$. We purposely consider this most pessimistic decision rule because it will show that our result holds for all other, more optimistic decision rules.

If Bob announces *any* theory $f \in \mathcal{F}$ deterministically, then his payoff is $u - d$ because for every theory $f \in \mathcal{F}$ there are many paths (i.e., those in R_f) at which f will be falsified. Hence, as long as the punishment d is greater than the reward u, Bob finds that announcing any falsifiable theory deterministically is strictly worse than not announcing any theory at all. Formally, as long as $d > u$,

$$\inf_{s \in \Omega} U(f, s) < 0 \text{ for every } f \in \mathcal{F}.$$

However, Bob can randomize and, as Proposition 1 shows, randomization alters Bob's prospects completely.

Proposition 1. *For any payoffs $u > 0$ and $d > 0$, there exists $\bar{\zeta} \in \Delta(\mathcal{F})$ such that*

$$V(\bar{\zeta}) > 0.$$

Proposition 1 shows that no matter how small the rewards for delivering a falsifiable theory, no matter how large the penalties for having a theory falsified, and no matter how much data Alice might have at her disposal, Bob is strictly better off by accepting her contract and producing theories at random. Even if Alice demands a falsifiable theory, she will not dissuade the most ignorant expert from delivering a fraudulent theory to her. This holds even in the extreme case that this ignorant expert evaluates his prospects based on the worst possible data that may be collected in the future.

The striking contrast between the case of a given theory and strategically produced theories conveys the first part of our argument. Falsifiability can be, and often is, used as a relevant criterion. In some cases, this criterion may seem intuitively weak (e.g., when there are too many possible outcomes and a single one is required to be ruled out). In other cases, this criteria may seem stronger (e.g., when a future outcome must be ruled out in every period). Still, whether intuitively weak or strong, falsifiability makes a fundamental conceptual distinction for given theories: if an expert is honest and wants his theory rejected (when false), then falsifiability is an useful criterion because only falsifiable theories can be conclusively rejected. In direct contrast, when theories are produced by a potentially strategic expert, then falsifiability cannot determine whether the expert is informed about the data-generating process. This result cast doubt on the idea that falsifiability can demarcate legitimate theories from worthless theories. As long as theories are produced by experts capable of strategic randomization, the falsification criteria cannot screen informed experts from uninformed ones.

3.1. Intuition that underlies Proposition 1

Given a random generator of falsifiable theories $\zeta \in \Delta(\digamma)$ and a path $(s^1, s^2, ...) \in \Omega$, the odds that the selected theory f is contradicted at some point in the future are

$$p_\zeta(s) = \zeta \left\{ f \in \digamma \mid \text{there exists } t \text{ such that } (s^1, ..., s^t) \in R_f \right\}.$$

The key argument is that for every $\varepsilon > 0$ there exists a random generator of falsifiable theories $\bar{\zeta} \in \Delta(\digamma)$ such that the odds of selecting a theory that will be falsified is smaller than ε, *for every path* $s \in \Omega$. That is,

$$p_{\bar{\zeta}}(s) \leq \varepsilon \text{ for every path } s \in \Omega.$$

No matter which data are realized in the future, a falsifiable theory selected by $\bar{\zeta}$ will not be falsified, with arbitrarily high probability. Hence, by randomizing according to specific probabilities, Bob is near certain that his theory will not be falsified.

To construct $\bar{\zeta}$, we first consider an increasing sequence of natural numbers Z_t, $t = 1, 2....$ Let $X_t = S^{(Z_{t+1}-Z_t)}$ be the set of $(Z_{t+1} - Z_t)$ outcomes. Given any sequence $x = (x_t)_{t=1}^{\infty}$, where each $x_t \in X_t$ consists of $(Z_{t+1} - Z_t)$ outcomes, we define a falsifiable theory f_x that is falsified if and only if x_t occurs between periods Z_t and Z_{t+1}. The random generator of theories $\bar{\zeta}$ is then defined as follows: an x_t, $t = 1, 2, ...$, is chosen from a uniform probability distribution over X_t, and the sequence $x = (x_t)_{t=1}^{\infty}$ (thus chosen) determines the theory f_x that is announced.

If Z_t grows sufficiently fast (and if z_1 is sufficiently large), the chances that x_t will be realized at least once is small. Any path s can be written in the form $s = (y_1, ..., y_t, ...)$, where y_t consists of $(Z_{t+1} - Z_t)$ outcomes. So the announced theory f_x is contradicted along s if and only if $x_t = y_t$ for some period t. By construction, this is an unlikely event.

4. Verification and Refutation Contracts

In this section, we consider general contracts. Bob decides whether to accept a contract at period zero. If Bob does not accept a contract then he does not deliver a theory and his payoff is zero. If Bob accepts a contract then he delivers a (not necessarily falsifiable) theory $f \in F$ to Alice at period zero (before any data is observed). An initial transfer may occur at period zero (after the theory is announced). This transfer gives utility $u(f, \varnothing)$ to Bob. At period t, if data $s_t \in \overline{S}$, $s_t = (s^1, ..., s^t)$, is observed, then a new transfer may occur. Bob's payoff, evaluated at period zero, for this contingent transfer is $u(f, s_t)$. So, given a path $s \in \Omega$, Bob's contingent net payoff at period zero is

$$U(f,s) = u(f,\varnothing) + \sum_{t=1}^{\infty} u(f,s_t), \quad \text{where } s = (s_t, ...).$$

Given a theory $f \in F$ and a probability measure $P \in \Delta(\Omega)$, we define

$$\bar{U}^P(f) = E^P U(f,s),$$

where E^P is the expectation operator associated with P. So, if Bob is informed at period zero and announces a theory f then his expected payoff is $\bar{U}^{P_f}(f)$. We say that an *informed expert accepts the contract* if for all $f \in F$

$$\bar{U}^{P_f}(f) \geq 0. \tag{4.1}$$

So, Bob, whenever informed about the odds of future events, prefers to announce what he knows rather than to refuse the contract.

Now assume that Bob faces uncertainty and does not know the probabilities of the future realizations of the data. As in the case of the falsification contract, Bob can select his theory by randomizing once (at period zero) according to a random generator of theories $\zeta \in \Delta(F)$ and Bob evaluates his prospects based on the worse-case scenario, i.e., based on the path $s \in \Omega$ that gives him minimal expected utility. Formally, Bob's payoff is

$$V(\zeta) = \inf_{s \in \Omega} E^{\zeta} U(f,s),$$

where E^{ζ} is the expectation operator associated with $\zeta \in \Delta(F)$. We say that the *uninformed expert accepts the contract* if there exists a random generator of theories $\bar{\zeta} \in \Delta(F)$ such that

$$V(\bar{\zeta}) \geq 0.$$

So, an uninformed expert accepts the contract if he can randomly select theories such that contingent on any possible realization of the data, the expected payoff is nonnegative. A contract screens informed and uninformed experts if the informed expert accepts the contract, but the uninformed expert does not accept it. Consider the following restrictions of Bob's payoffs:

Restriction A. For all $\zeta \in \Delta(F)$ and $P \in \Delta(\Omega)$, $E^{\zeta} E^P U(f,s)$ is finite. Moreover, for every $s_t \in \bar{S}$, $u(f,s_t)$ is a bounded function of f.

Restriction B. The function $U(f,s)$ is bounded. Moreover, for every f, $U(f,s)$ is a continuous of s, and for every s, $U(f,s)$ is a continuous function of f. That is, for every $s \in \Omega$,

$$U(f_n, s) \longrightarrow U(f,s) \text{ as } f_n \text{ converges pointwise to } f,^7$$

and for every $f \in F$

$$U(f, s^{(n)}) \longrightarrow U(f, s) \text{ as } s^{(n)} \text{ converges to } s.^8$$

Restriction A is a mild assumption, made for technical reasons. Restriction B is more demanding because it requires net payoffs in the remote future to be small.

We now consider two fundamentally different types of contracts. If $u(f, s_t) \geq 0$ for every $s_t \in \bar{S}$, $s_t \neq \varnothing$, then the contract is said to be a *verification contract*. If $u(f, s_t) \leq 0$ for every $s_t \in \bar{S}$, $s_t \neq \varnothing$, then the contract is said to be a *refutation contract*. So, Bob may receive a (positive, negative or zero) payoff at period zero when the theory is delivered, but the distinction between verification and refutation contracts depends only on the payoffs after the data is observed. Verification contracts are those in which after the data is revealed Bob receives either no payoff or positive payoff, contingent on the performance of the theory vis-a-vis the data. Refutation contracts are those in which after the data is revealed Bob receives either no payoff or negative payoffs contingent on how his theory performs vis-a-vis the data. The terminology reflects the idea that in a verification contract Bob is paid when his theory performs well and in a refutation contract Bob pays when his theory performs poorly.

Proposition 2. *Assume restriction B. If an informed expert accepts a verification contract, then an uninformed expert also accepts this verification contract.*

Proposition 2 shows that verification contracts cannot screen informed and uninformed experts. This result show a fundamental limitation of verification as a guiding principle for empirical analysis. No verification contract can mitigate Alice's adverse selection problem.

Now consider the following moral hazard problem: Bob is uninformed, but can become informed, before any data are revealed, if he acquires sufficient information that allows him to formulate probabilities (i.e., to transform his uncertainty into risk). The cost (i.e., his disutility) of acquiring this information is $c > 0$. When Bob decides whether to become informed, he does not know the data-generating process and he does not have a prior over the space of data-generating processes (otherwise he would use this prior to access the odds of the data). Hence, Bob makes a decision under uncertainty. We assume, as before, that Bob evaluates his prospects under a worst-case scenario. Therefore, at period one, Bob's net value of becoming informed is

$$V(I, c) = \inf_{f \in F} \bar{U}^{P_f}(f) - c.$$

That is, $V(I, c)$ is Bob's smallest expected utility, when informed, minus the cost c of becoming informed. On the other hand, if Bob remains uninformed and produces a theory according to ζ, then his payoff, also computed in a worse-case scenario, is $V(\zeta)$.

We say that the expert *prefers not to become informed* if there exists a random generator of theories $\bar{\zeta} \in \Delta(F)$ such that $V(\bar{\zeta}) > V(I,c)$. Then, Bob prefers to announce theories at random (selected by $\bar{\zeta}$) than to become informed at cost c.

Proposition 3. *Assume restriction A. Consider any verification contract and a positive (possibly arbitrarily small) cost $c > 0$ of acquiring information. Then, the expert prefers not to become informed.*

Proposition 3 shows that no matter how cheap is the cost of acquiring information, the expert prefers to not acquire it. Instead, the expert chooses to present a randomly selected theory.

To underscore proposition 3, assume that for every theory f, a set of finite histories A_f (called the acceptance set) is defined as consistent with f. That is, histories in A_f are deemed consistent with f. Also assume that the acceptance A_f has probability $1-\varepsilon$, $\varepsilon > 0$, under theory f (i.e., $P_f(A_f) = 1-\varepsilon$). So, the acceptance set A_f is likely under theory f.[9] Hence, if the announced theory is indeed the data-generating process, then the data is likely to be deemed consistent with the theory, but no other restrictions are placed on which histories are deemed consistent with each theory.

Now consider the contract in which Bob receives zero payoff in period zero, when the theory is announced, but Bob receives payoff 1 whenever the observed history is deemed consistent with Bob's theory. (Formally, $u(f,\varnothing) = 0$, $u(f,s_t) = 1$ if $s^t \in A_f$ and $u(f,s_t) = 0$ if $s_t \notin A_f$.) At period zero, Bob's net value of becoming informed, $V(I,c)$, is $1-\varepsilon-c$ because, if informed, Bob gets payoff 1 with probability $1-\varepsilon$. By proposition 3, $V(\bar{\zeta})$ is arbitrarily close to $1-\varepsilon$ for some $\bar{\zeta} \in \Delta(F)$. This is a strong result. It implies that no matter how the data unfolds a theory selected by $\bar{\zeta}$ will be deemed consistent with data, with probability arbitrarily close to $1-\varepsilon$. Formally, there exists $\bar{\zeta} \in \Delta(F)$ such that for all $s \in \Omega$,

$$\bar{\zeta}\{f \in F \mid s_t \in A_f \text{ for some } t\} > 1-\varepsilon-c.$$

Hence, Bob prefers not to become informed because even if he remains completely uninformed he can still ensure that his theory is likely to be deemed consistent with the data, no matter how the data unfolds. *Thus, no knowledge over the data-generating process is necessary to produce theories that will, in the future, prove to be supported by the data.*

[9]For example, if Bob's theory asserts that 1 has probability $p \in [0,1]$ in all periods, then an acceptance set could comprise all histories in which the relative frequency of 1 is between $p-\delta$ and $p+\delta$, after sufficiently many periods.

4.1. Intuition that underlies Propositions 2 and 3

The proofs of propositions 2 and 3 are similar and so we focus on the (slightly more subtle) intuition of Proposition 3. Consider the zero-sum game between Nature and the expert such that Nature's pure strategy is a path $s \in \Omega$ and the expert's pure strategy is a theory $f \in F$. The expert's payoff is $U(s, f)$. For every mixed strategy of Nature P, there exists a strategy for the expert (to announce a theory f such that $P_f = P$) that gives the expert an expected payoff of at least $V(I, c) + c$. So, if the conditions of Fan's Minmax theorem are satisfied, there is a mixed strategy for the expert, $\bar{\zeta}$, that gives him an expected payoff higher than $V(I, c)$, no matter which path $s \in \Omega$ Nature selects.

A key condition in Fan's Minmax theorem is the lower semi-continuity of Nature's payoff. We show that the positive payoffs in verification contracts suffice to deliver this condition. In the falsification contract (like other refutation contracts), Nature's payoff is not necessarily lower semi-continuous.

4.2. Refutation Contracts

Verification, like falsification, does not deliver an effective way to determine whether Bob's theory is based on any relevant knowledge of the data-generating process. So, in this section, we consider refutation contracts.

Proposition 4. *There exists a refutation contract that screens informed and uninformed experts.*

Proposition 4 shows an asymmetry between verification and refutation contracts. No verification contract can screen informed and uninformed experts, but some refutation contracts can. Popper's falsification delivers a refutation contract that cannot screen informed and uninformed experts. In this section, a screening refutation contract is constructed. We show it in the context of Popper's main example.

Assume that every period a swan is observed. This swan can be white (1) or of another color (0). Let 1_t be the $(t+1)-$history of 1 in all periods until period t and 0 at period $t+1$. Let $\bar{1}_m$ consists of 1_t, $t \geq m$. So, $\bar{1}_m$ are the histories in which only white swans are seen for at least m periods and then a nonwhite swan is observed. As Popper's pointed out, the event $\bar{1}_m$ is not impossible. Thus, a long sequence of white swans does not prove that all swans are white. However, $\bar{1}_m \downarrow \varnothing$ as m goes to infinity because $\bar{1}_{m+1}$ is contained in $\bar{1}_m$ and the interception of all $\bar{1}_m$ is empty. So, for every probability measure P, $P(\bar{1}_m) \downarrow 0$ as m goes to infinity. That is, for *any* data-generating process, the event $\bar{1}_m$ is unlikely if m is large enough.

Fix $\varepsilon > 0$. Given a theory f, let $m(f)$ be a period such that $P_f(\bar{1}_{m(f)}) \leq \varepsilon$. Consider the following contract: At period 0, Bob receives $\delta \in (\varepsilon, 0.5)$ for announcing any theory f. Bob also receives disutility 1 contingent on $\bar{1}_{m(f)}$ and no disutility otherwise. Formally, $u(f, \varnothing) = \delta$; $u(f, s_t) = -1$ if $s_t \in \bar{1}_{m(f)}$ and $u(f, s_t) = 0$ if $s_t \notin \bar{1}_{m(f)}$. We will call this contract a *simple refutation contract*.

This contract is an example of a *test-based contract*, in which a set of histories R_f (called the rejection set) is defined as inconsistent with theory f, and Bob incurs a disutility of 1 if the observed history is inconsistent with his theory. The parallel to testing is obvious: 1 is a disutility of test rejection. In the simple refutation contract, R_f is defined as $\bar{1}_{m(f)}$. The Cournot principle is relevant here because $\bar{1}_{m(f)}$ may not conclusively refute f. That is, $\bar{1}_{m(f)}$ is perhaps possible, but unlikely under f.

An informed expert accepts the simple refutation contract because it delivers positive expected payoff, whenever he is informed about the odds of future events ($\bar{U}^{P_f}(f) = \delta - P_f(\bar{1}_{m(f)}) \geq \delta - \varepsilon > 0$). We now show that an uninformed expert refuses this contract. Let $F_m \subseteq F$ be the set of all theories that are inconsistent with all finite histories in $\bar{1}_m$ (i.e., $\bar{1}_m \subset R_f$ or $m(f) \leq m$). Given that, $\bar{1}_{m+1} \subseteq \bar{1}_m$ it follows that $F_m \subseteq F_{m+1}$. Moreover, any theory $f \in F$ belongs to F_m for some m, and so, $F_m \uparrow F$ as m goes to infinity. Hence, for every random generator of theories $\zeta \in \Delta(F)$, there exists m^* such that $\zeta(F_{m^*}) \geq 1 - \delta$. So, if $s = (1_{m^*}, ...)$, then $E^{\zeta} U(f, s) = \delta - \zeta(F_{m^*}) < 0$. Hence, $V(\zeta) < 0$ for all random generator of theories $\zeta \in \Delta(F)$.

The simple refutation contract screens informed and uninformed experts. This, in conjunction with inability of verification contracts to do the same, delivers an original argument supporting the idea that refutation is a better guiding principle for empirical research than verification.

We wish to emphasize some properties of the simple refutation contract. In Olszewski and Sandroni (2006), we consider a very large class of empirical tests ordinarily used in statistics (such as calibration and likelihood tests). The contracts based on these tests, like verification contracts, cannot screen experts. Hence, the simple refutation contract is a based on an original way of testing theories. Moreover, given that ε (and, hence, δ) can be made arbitrarily small, it follows that Alice need only to make a small payments to induce an informed expert to accept the simple contract. Finally, if ε is very small then it is near-optimal for Bob, if informed, to reveal his theory truthfully. This follows because the odds that Bob will incur in any disutility can be made arbitrarily small (if he is informed and truthfully reveals his theory).

In a recent contribution, Dekel and Feinberg (2006) also show that if the continuum hypothesis holds, then there exists a test-based refutation contract that can screen informed and uninformed experts (an uninformed expert fails this test on uncountably many infinite histories). The drawback of this test is that it relies on the continuum hypothesis which cannot be shown to be true or false within standard mathematical analysis. If the contract is based on this test and Bob presents a theory (e.g., the probability of 1 is 0.5 in all periods) then, after some data is observed, it is not possible (by means of any mathematical calculation) to determine whether Bob must incur some disutility.

A weakness of the simple test is that an uninformed expert fails it only on the histories $\bar{1}_m$. However, Olszewski and Sandroni (2007a) show a complex test-based refutation contract (that can screen informed and uninformed experts) such that an uninformed expert fails the test, no matter how he randomizes, on a topologically large set of histories.

5. Conclusion

Falsifiability is a widely used guide in research and legal proceedings because it is perceived as a requirement that could disqualify nonscientific theories. Indeed, falsifiable theories can be conclusively rejected, whereas nonfalsifiable cannot. In contrast, we show that falsifiability imposes essentially no constraints when theories are produced by strategic experts. Without any knowledge, it is possible to construct falsifiable theories that are unlikely to be falsified, no matter how the data unfold in the future.

Verification suffers from the same difficulty as falsification. Strategic experts, with no knowledge of the data-generating process, can produce theories that, in the future, are likely to be consistent with the data. However, there are special ways of constructing refutation contracts (by defining which data are inconsistent with each theory) that can distinguish legitimate from worthless theories, even if experts are strategic.

6. Proofs

We use the following terminology: Let $\Omega = \{0,1\}^\infty$ be the set of all *paths*, i.e., infinite histories. 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.

Let $\Im_t$ be the algebra that consists of all finite unions of cylinders with base on $\{0,1\}^t$. Denote by N the set of natural numbers. Let $\Im$ is the $\sigma-$algebra generated by the algebra $\Im^0 := \bigcup\limits_{t \in N} \Im_t$, i.e., $\Im$ is the smallest $\sigma-$algebra which contains $\Im^0$. Let $\Delta(\Omega)$ the set of all probability measures on $(\Omega, \Im)$. It is well-known that every theory f determines uniquely a probability measure $P_f \in \Delta(\Omega)$.

6.1. Proof of Proposition 1

A cylinder with base on history $s_t = (s^1, ..., s^t)$ is denoted by $C(s^1, ..., s^t)$. Take any positive

$$\varepsilon < \min\left\{\frac{1}{n}, \frac{u}{d}\right\}.$$

We will construct a $\bar{\zeta} \in \Delta(\mathcal{F})$ such that for every path $(s^1, s^2, ...) \in \Omega$,

$$\bar{\zeta}\left\{f \in \mathcal{F} \mid \exists_t \quad (s^1, ..., s^t) \in R_f\right\} < \varepsilon.$$

This will complete the proof as

$$V(\bar{\zeta}) \geq (1-\varepsilon)u + \varepsilon(u-d) > 0.$$

Take a number $r > 0$ so small that

$$\sum_{t=1}^{\infty} r^t < \varepsilon,$$

and next take a sequence of natural numbers $\{M_t,\ t = 1, 2...\}$ such that

$$\frac{1}{n^{M_t}} < r^t.$$

Let $\hat{P} \in \Delta(\Omega)$ be the probability measure such that all outcomes $s \in S$ have equal odds in all periods. It will be convenient to denote by X_t the set S^{M_t} and by X the Cartesian product $\prod_{t=1}^{\infty} X_t$ of sets X_t; although $X = \Omega$, it will be convenient to distinguish the two spaces. Consider a sequence of independent random variables $\tilde{X}_t$, $t = 1, 2, ...$, uniformly distributed on the set X_t. Let $\tilde{X}$ be the random variable $\prod_{t=1}^{\infty} \tilde{X}_t$, distributed on X, such that $\tilde{X} = (x_1, ..., x_t, ...)$, $x_t \in X_t$, if and only if $\tilde{X}_t = x_t$ for all $t = 1, 2, ...$

Let

$$Z_t \equiv \sum_{j=1}^{t} M_j.$$

Given an $x = (x_1, ..., x_t, ...)$, $x_t \in X_t$, let

$$C_x \equiv C(x_1) \cup \bigcup_{t=1}^{\infty} \bigcup_{z_t \in S^{Z_t}} C(z_t, x_{t+1})$$

be the union of the cylinder $C(x_1)$ with base on x_1, and the cylinders with base on histories of the form (z_t, x_{t+1}), where z_t is an arbitrary element of S^{Z_t}.

Given that x_1 a sequence of M_1 outcomes and x_{t+1} is a sequence of M_{t+1} outcomes, it follows that

$$\hat{P}(\bigcup_{z_t \in S^{Z_t}} C(z_t, x_{t+1})) = \frac{1}{n^{M_{t+1}}} \text{ and } \hat{P}(C(x_1)) = \frac{1}{n^{M_1}};$$

hence,

$$\hat{P}(C_x) \leq \sum_{t=1}^{\infty} \frac{1}{n^{M_t}} < \varepsilon < 1. \tag{6.1}$$

Let $(C_x)^c$ be the complement of C_x, $x \in X$, and let $\hat{P}^x$ be the conditional probability of $\hat{P}$ on $(C_x)^c$, $x \in X$. That is,

$$\hat{P}^x(A) = \frac{\hat{P}(A \cap (C_x)^c)}{\hat{P}((C_x)^c)} \text{ for all } A \in \Im.$$

Step 1: Observe that any history $(s^1, ..., s^t) \in S^t$ has an extension

$$(s^1, ..., s^t, s^{t+1}, ..., s^m)$$

such that $\hat{P}^x(s^1, ..., s^m) = 0$.

Indeed, take a $Z_m \geq t$ and any extension $z_m \in S^{Z_m}$ of $(s^1, ..., s^m)$. Then (z_m, x_{m+1}) is also an extension of $(s^1, ..., s^t)$, and by definition, $C(z_m, x_{m+1}) \subset C_x$, and so $\hat{P}^x(C(z_m, x_{m+1})) = 0$.

Step 2: Let $C = C(s^1, ..., s^k)$ be any cylinder not contained in C_x. We will show that

$$\hat{P}(C \cap (C_x)^c) > 0.$$

Let $\hat{P}^C$ denote $\hat{P}$ conditional on C. By Bayes' rule,

$$\hat{P}(C \cap (C_x)^c) = \hat{P}^C((C_x)^c)\hat{P}((C_x)^c),$$

and by (6.1),

$$\hat{P}((C_x)^c) > 0,$$

so it suffices to show that $\hat{P}^C(C_x) < 1$.

Let $\hat{C}_x$ be the union of all cylinders with base on x_1 or with base (z_t, x_{t+1}), where z_t is an arbitrary element of S^{Z_t}, whose length is no greater than k. So, $\hat{C}_x$ is a finite union of cylinders $C(r^1,...,r^l) \subset C_x$, where $l \leq k$. The history $(s^1,...,s^k)$ cannot coincide on its first outcomes with any history from the set $\{x_1\} \cup \{(z_t, x_{t+1}) \mid z_t \in \{0,1\}^{Z_t}, t = 1,2,...\}$, because $C(s^1,...,s^k)$ would be contained in C_x. Thus, $C(r^1,...,r^l) \cap C = \varnothing$ if $C(r^1,...,r^l) \subset C_x$, where $l \leq k$, and so

$$\hat{P}^C\left(\hat{C}_x\right) = 0.$$

Let $\bar{C}_x$ be the union of all cylinders with base on x_1 or with base (z_t, x_{t+1}), where z_t is an arbitrary element of S^{Z_t}, whose length is strictly greater than k. So, if $m = 0,1,...$ is the smallest number such that $k < Z_{m+1}$, then $\bar{C}_x$ is the union of the cylinders $C(z_t, x_{t+1})$, $t \geq m$ and $z_t \in \{0,1\}^{Z_t}$ (and the set $C(x_1)$ if $m = 0$). Suppose first that $m = 0$. Then, $\hat{P}^C(C(x_1)) = n^{k-Z_1}$ if x_1 coincides on its first outcomes with $s^1,...,s^k$ or $\hat{P}^C(C(x_1)) = 0$ if x_1 does not coincide on its first outcomes with $s^1,...,s^k$, and

$$\hat{P}^C\left(\bigcup_{z_t \in \{0,1\}^{Z_t}} C(z_t, x_{t+1})\right) = \frac{1}{n^{M_{t+1}}}.$$

Hence,

$$\hat{P}^C(\bar{C}_x) \leq \frac{1}{n} + \sum_{t>1} \frac{1}{n^{M_t}} < \frac{1}{n} + \varepsilon < 1. \tag{6.2}$$

By analogous argument, (6.2) holds with $\sum_{t>k}$ replacing $\sum_{t>1}$ for $k > 0$. Obviously, $C_x = \hat{C}_x \cup \bar{C}_x$, and so

$$\hat{P}^C(C_x) \leq \hat{P}^C(\hat{C}_x) + \hat{P}^C(\bar{C}_x) < 1.$$

Step 3: We will show that C_x is the union of all cylinders $C \in \Im^0$ such that $\hat{P}^x(C) = 0$.

Let $C \in \Im^0$ be an arbitrary cylinder. If $\hat{P}^x(C) = 0$ then $\hat{P}(C \cap (C_x)^c) = 0$. By Step 2, C is contained in C_x. On the other hand, if C is contained in C_x, then $C \cap (C_x)^c = \varnothing$. Hence, $\hat{P}^x(C) = 0$.

Let $\bar{\zeta} \in \Delta(\digamma)$ be defined as follows: First a realization of the random variable $\tilde{X}$ is observed. The probability measure $\hat{P}^x$ is selected whenever $\tilde{X} = x$. The probability measure determines a theory f_x that is announced. The theory f_x is determined by conditional probabilities of $\hat{P}^x$, i.e. for any $\overline{s} \in \overline{S}$ and $s \in S$,

$$f_x(\overline{s})[s] = \frac{\hat{P}^x(C(\overline{s}, s))}{\hat{P}^x(C(\overline{s}))}$$

when $\hat{P}^x(C(\overline{s})) > 0$, and it is irrelevant how $f_x(\overline{s})[s]$ is defined when $\hat{P}^x(C(\overline{s})) = 0$.

It follows from Steps 1 that the theories f_x, $x \in X$, are falsifiable.

Fix a path $(s^1, s^2, ...) \in \Omega$; every such path can also be represented as $(\overline{s}_1, ..., \overline{s}_t, ...)$ where $\overline{s}_t \in X_t$; so, $\overline{s}_1$ are the first M_1 outcomes of $(s^1, s^2, ...)$ and $\overline{s}_{t+1}$ are the M_{t+1} outcomes that follow the first Z_t outcomes. By definition, $(\overline{s}_1, ..., \overline{s}_t, ...) \in C_x$ if and only if $\overline{s}_t = x_t$ for some $t \in 1, 2, ...$ Notice that $P_{f_x}(s^1, ..., s^m) = 0$ if and only if $\hat{P}^x(C(s^1, ..., s^m)) = 0$ for every finite history $(s^1, ..., s^m)$. Hence, by Step 3,

$$\left\{x \in X \mid \exists_m \quad (s^1, ..., s^m) \in R_{f_x}\right\} = \{x \in X \mid x_t = \bar{s}_t \text{ for some } t = 1, 2, ...\}. \quad (6.3)$$

Since

$$\hat{P}\left\{\tilde{X}_t = \bar{s}_t \text{ for some } t = 1, 2, ...\right\} \leq \sum_{t=1}^{\infty} \hat{P}\left\{\tilde{X}_t = \bar{s}_t\right\} = \sum_{t \in N} \frac{1}{2^{M_t}} \leq \sum_{t \in N} r^t < \varepsilon,$$

it follows from (6.3) that

$$\bar{\zeta}\left\{f \in F \mid \exists_m \quad (s^1, ..., s^m) \in R_f \right\} = \hat{P}\left\{x \in X \mid \exists_m \quad (s^1, ..., s^m) \in R_{f_x}\right\} < \varepsilon.$$

■

6.2. Proof of Propositions 2 and 3

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

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

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

We endow $\Delta(\Omega)$ with the weak$-*$ topology and with the $\sigma-$algebra of Borel sets, (i.e., the smallest $\sigma-$algebra which contains all open sets in weak$-*$ topology). We endow the set of all theories F with the pointwise convergence topology; in this topology, a sequence of theories $(f_n)_{n=1}^{\infty}$ converges to a theory f if $f_n(s_t) \to_n f(s_t)$ for every history $s_t \in \{\varnothing\} \cup S_{\infty}$. Let $\Delta(F)$ be the set of probability measures on F. We also endow $\Delta(F)$ with the weak$-*$ topology. It is well-known that Ω, $\Delta(\Omega)$, F, and $\Delta(F)$ are compact metrizable spaces.

Recall that if $\Delta(X)$ is endowed with the the weak$-*$ topology, then if $h : X \to R$ is a continuous function, then $H : \Delta(X) \to R$ defined by $H(\tilde{x}) = E^{\tilde{x}}(h)$, where $E^{\tilde{x}}$ is the expectation operator associated with $\tilde{x} \in \Delta(X)$, is also a continuous function.

Consider an arbitrary verification contract. Let $\mathcal{H} : \Delta(F) \times \Delta(S) \longrightarrow \Re$ be a function defined by

$$\mathcal{H}(\zeta, P) = E^{\zeta} E^P U(f, s).$$

Step 1: Assume that $\mathcal{H}(\zeta, P) < \infty$ for every $\zeta \in \Delta(F)$ and $P \in \Delta(S)$ and that for every $s_t \in \bar{S}$, $u(f, s_t)$ is a bounded function of f. Then, for every $\zeta \in \Delta(F)$, $\mathcal{H}(\zeta, P)$ is a lower semi-continuous function of P.

Proof: Let $X_t(f, s) = u(f, s_t)$ where $s = (s_t, ...)$, $t \geq 1$, and $X_0(f, s) = u(f, \varnothing)$. Then,

$$U(f, s) = \sum_{t=0}^{\infty} X_t(f, s).$$

By the monotone convergence theorem,

$$E^P U(f, s) = \sum_{t=0}^{\infty} E^P\{X_t(f, s)\} \text{ and } \mathcal{H}(\zeta, P) = \sum_{t=0}^{\infty} E^\zeta E^P X_t(f, s).$$

Fix $\varepsilon > 0$ and assume that P_n converges to P (in the weak$-*$ topology) as n goes to infinity. Then, by the definition of convergence in the weak$-*$ topology, $E^{P_n} X_t(f, s) \underset{n\to\infty}{\longrightarrow} E^P X_t(f, s)$ for every t. By the dominated convergence theorem, $E^\zeta E^{P_n} X_t(f, s) \underset{n\to\infty}{\longrightarrow} E^\zeta E^P X_t(f, s)$ for every t.

Now if $\mathcal{H}(\zeta, P) < \infty$ then there exists m^* such that

$$\sum_{t=m^*+1}^{\infty} E^\zeta E^P X_t(f, s) < \frac{\varepsilon}{2}.$$

If n is sufficiently large,

$$\sum_{t=0}^{m^*} E^\zeta E^{P_n} X_t(f, s) \geq \sum_{t=0}^{m^*} E^\zeta E^P X_t(f, s) - \frac{\varepsilon}{2}$$

and, therefore,

$$\mathcal{H}(\zeta, P_n) \geq \sum_{t=0}^{m^*} E^\zeta E^{P_n} X_t(f, s) \geq \mathcal{H}(\zeta, P) - \varepsilon.$$

Step 2: Assume that for every f, $U(f, s)$ is a continuous of s, and for every s, $U(f, s)$ is a continuous function of f. Assume, in addition, that $U(f, s)$ is a bounded function of s and f. Then, for every ζ, $\mathcal{H}(\zeta, P)$ is a continuous of P, and for every P, $\mathcal{H}(\zeta, P)$ is a continuous of ζ.

Proof: If $f_n \underset{n\to\infty}{\longrightarrow} f$ then, by assumption, $U(f_n, s) \underset{n\to\infty}{\longrightarrow} U(f, s)$. By the dominated convergence theorem, $E^P U(f_n, s) \underset{n\to\infty}{\longrightarrow} E^P U(f, s)$. Hence, $\mathcal{H}(\zeta, P)$ is continuous on ζ for every P.

If $s^n \underset{n\to\infty}{\longrightarrow} s$, then, by assumption, $U(f, s^n) \underset{n\to\infty}{\longrightarrow} U(f, s)$. Hence, if $P_n \underset{n\to\infty}{\longrightarrow} P$, then $E^{P_n} U(f, s) \underset{n\to\infty}{\longrightarrow} E^P U(f, s)$ by the definition of convergence in the weak$-*$ topology.

By the dominated convergence theorem, $\mathcal{H}(\zeta, P_n) \underset{n\to\infty}{\longrightarrow} \mathcal{H}(\zeta, P)$. Hence, for every ζ, $\mathcal{H}(\zeta, P)$ is continuous on P.

Step 3: The sets $\Delta(F)$ and $\Delta(S)$ are compact sets in the weak$-*$ topology and $\mathcal{H}(\zeta, P)$ is a linear function on ζ and P. Hence, it follows from Fan's (1953) minmax theorem that if $\mathcal{H}(\zeta, P)$ is lower semi-continuous in P, then

$$\inf_P \sup_\zeta \mathcal{H}(\zeta, P) = \sup_\zeta \inf_P \mathcal{H}(\zeta, P) \tag{6.4}$$

and if $\mathcal{H}(\zeta, P)$ is continuous in both ζ and P, then

$$\min_P \max_\zeta \mathcal{H}(\zeta, P) = \max_\zeta \min_P \mathcal{H}(\zeta, P) \tag{6.5}$$

Proof of proposition 2: For a given P take a ζ such that $\zeta(\{f\}) = 1$ for some f with $P_f = P$. Then, by (4.1), the left-hand side of (6.5) exceeds 0. So, the right-hand side of (6.5) exceeds 0. This yields $\bar{\zeta} \in \Delta(\Delta(\Omega))$ such that $\mathcal{H}(\bar{\zeta}, P) \geq 0$ for every P. Now, for any given $s \in \Omega$, take the probability measure P_s such that $P_s(\{s\}) = 1$. So, $\mathcal{H}(\bar{\zeta}, P_s) \geq 0$ for every s. Hence, $V(\bar{\zeta}) \geq 0$.■

Proof of proposition 3: For a given P take a ζ such that $\zeta(\{f\}) = 1$ for some f with $P_f = P$. Then, the left-hand side of (6.4) exceeds $V(I, c) + c$. So, the right-hand side of (6.4) exceeds $V(I, c) + c$. This yields $\bar{\zeta} \in \Delta(\Delta(\Omega))$ such that $\mathcal{H}(\bar{\zeta}, P_s) > V(I, c) + 0.5c$ for every $s \in \Omega$. Hence, $V(\bar{\zeta}) \geq V(I, c) + 0.5c > V(I, c)$.■

References

[1] Al-Najjar, N. and J. Weinstein (2006) "Comparative Testing of Experts," forthcoming *Econometrica*.

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

[3] Carvajal, A., I. Ray and S. Snyder (2004) "Equilibrium Behavior in Markets and Games: Testable Restrictions and Identification," *Journal of Mathematical Economics*, **40**, 1-40.

[4] Cournot, A. (1843): *Exposition de la Théorie des Chances et des Probabilités*, Hachette, Paris.

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

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

[7] Feinberg, Y. and C. Stewart (2006) "Testing Multiple Forecasters," forthcoming *Econometrica*.

[8] Fortnow, L. and R. Vohra (2006) "The Complexity of Forecast Testing," mimeo.

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

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

[11] Hart, S. and A. Mas-Colell (2001) "A General Class of Adaptative Strategies," *Journal of Economic Theory*, **98**, 26-54.

[12] Kuhn, T. (1962): *The Structure of Scientific Revolutions*, The University of Chicago Press.

[13] Kalai, E., E. Lehrer, and R. Smorodinsky (1999) "Calibrated Forecasting and Merging," *Games and Economic Behavior*, **29**, 151-169.

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

[15] Lakatos, I. and A. Musgrave, ed., (1970): *Criticism and the Growth of Knowledge*, Cambridge University Press.

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

[17] Lehrer, E. and E. Solan (2003) "No Regret with Bounded Computational Capacity," mimeo.

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

[19] Olszewski, W. and A. Sandroni (2006) "Manipulability of Future-Independent Tests," mimeo.

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

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

[22] Popper, K. (1968): *The Logic of Scientific Discovery*, Hutchinson & Co., London.

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

[24] Rustichini, A. (1999) "Minimizing Regret: The General Case," *Games and Economic Behavior*, **29**, 244-273.

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

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

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

[28] 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.