ON COMPARISON OF EXPERTS

DRAFT - NOT FOR CIRCULATION

Abstract

A policy maker faces a sequence of unknown outcomes. At each stage two (self-proclaimed) experts provide probabilistic forecasts on the outcome in the next stage. A comparison test is a protocol for the policy maker to (eventually) decide which of the two experts is better informed. The protocol takes as input the sequence of pairs of forecasts and actual realizations and (weakly) ranks the two experts. We propose two natural properties that such a comparison test must adhere to and show that these essentially uniquely determine the comparison test. This test is a function of the derivative of the induced pair of measures at the realization.

KEYWORDS: Forecasting, probability, testing.

1 Introduction

The literature on expert testing has, by and large, treated the question of whether a self-proclaimed expert can be identified as such, while also not allowing for charlatans to pass the test. A striking result due to Sandroni (2003) is that no such test exists without additional structural assumptions on the problem. The basic premise of this literature is the validity of the underlying question of whether a forecaster, or rather a probabilistic model, is correct or false. In a hypothetical world, where only one model exists and the tester can only entertain the services of a single expert, this may make sense. Even then, one might wonder what is the tester to do whenever she rejects the expert. Does she turn to another expert? to her own intuition? In any case she would probably, implicitly, utilize an alternative (possibly untested) model.

This motivates an alternative approach to the issue of expert testing and that is comparison of experts, which is the approach we pursue here. In this approach the tester is exposed to a few alternative models (forecasters) and a single realization of events. The tester then compares the two forecasters and decides which is the better informed one. Facing many (possibly conflicting) experts is commonplace in weather forecasting, financial forecasting, medical prognosis and more. Nevertheless, the design of comparison tests has been almost entirely ignored in the literature on expert testing. Two exceptions are

Al-Najjar and Weinstein (2008) and Feinberg and Stewart (2008) which we will discuss later.

The approach we take in this paper is axiomatic. After defining exactly what is meant by a comparison test we will turn to discuss some desirable properties for such tests. We then construct a test with all the desired properties and show it is essentially unique. The setting we focus on is that of two experts and a test which (weakly) ranks the two and hence its domain consists of three outcomes. It may either point at one of the two experts as being better informed or it may be indecisive. Let us discuss the properties that are central to our main results.

Anonymity - A test is *anonymous* if it does not depend on the identity of the agents but only on their forecasts.

Error-free - Let us assume that one of the experts has the correct model (namely, he would have passed a standard single expert test which has no type one errors). An *error-free* test will surely not point at the second expert as the superior one (albeit, it may provide a non-conclusive outcome).

Reasonable - Let us consider an event, $\mathcal{A}$, that has positive probability according to the first expert but zero probability according to the second. Conditional on the occurrence of event $\mathcal{A}$ a *reasonable* test must assign positive probability to the first expert being better than the second.

Tail test - As experts often require some initial data to calibrate their models, we would like to rule out tests whose decision is based on a prefix of the realized outcome. A tail test is one which depends only on forecasts made eventually, after the calibration phase. Whereas much of the literature emphasizes tests which provide their verdict at some finite outcome we take the opposite approach for some of our results and consider comparison tests that are based on a long-run performance.[1]

1.1 Results

We construct a specific comparison test, based on the derivative of two measures that are induced by the two forecasts derived from the likelihood ratio of the two forecasts. We prove that this test is symmetric, error-free, reasonable and is also a tail test.

We then show that the test is unique in the following sense. For any test that differs from the construction and is symmetric and reasonable there exist two forecasters for which an error will be made (the probability of reversing the order). If, in addition, the test is a tail test then this error could be made arbitrarily close to one.

Finally, our constructed test perfectly identifies the correct forecaster whenever the two measures induced by the forecasters are mutually singular with respect to each other.

[1] Consider the classical example of an IID process. A forecaster who is aware that indeed the process is such must calibrate the model to learn its parameter. Initial forecasts may be wrong, yet those made after a calibration phase become more accurate and long-run predictions are spot-on.

Requiring the test to identify the correct expert when the measures are not mutually singular is shown to be impossible.

1.2 Related Literature

Much of the literature on expert testing focuses on the single expert setting. This literature dates back to the seminal paper of Dawid (1982), who proposes the calibration test as a scheme to evaluate the validity of weather forecasters. Dawid asserts that a test must not fail a true expert. Foster and Vohra (1998) show how a charlatan, who has no knowledge of the weather, can produce forecasts which are always calibrated. The basic ingredient that allows the charlatan to fool the test is the use of random forecasts. Lehrer (2001) and Sandroni, Smorodinsky and Vohra (2003) extend this observation to a broader class of calibration-like tests.

Finally, Sandroni (2003) shows that there exists no error-free test that is immune to such random charlatans (see also extensions of Sandroni's result in Shmaya (2008) and Olszewski and Sandroni (2008)).

To circumvent the negative results various authors suggest to limit the set of models for which the test must be error-free (e.g., Al-Najjar, Sandroni, Smorodinsky and Weinstein (2010) and Pomatto (2017)), or to limit the computational power associated with the charlatan (e.g., Fortnow and Vohra (2009)) or to replace measure theoretic implausibility with topological implausibility by resorting to the notion of *category one sets* (e.g., Dekel and Feinberg (2006)).

As previously mentioned, the comparison of experts has drawn little attention in the community studying expert testing, with two exceptions we are aware of, which we now turn to discuss.

Al-Najjar and Weinstein (2008) consider a non-counterfactual likelihood test for comparing two experts. They show that if one expert knows the true process whereas the other is uninformed, then one of the following must occur: either, the test correctly identifies the informed expert, or the forecasts made by the uninformed expert are close to those made by the informed one. Note that the test they propose is anonymous and reasonable but is not error-free. An asymptotic version of this likelihood ratio, however, will play a crucial role in our construction.

Another approach was suggested by Feinberg and Stewart (2008), who study an infinite horizon model of testing multiple experts, using a cross-calibration test. In their test N experts are tested simultaneously; each expert is tested according to a calibration restricted to dates where not only does the expert have a fixed forecast but the other experts also have a fixed forecast, possibly with different values. That is to say, where the calibration test checks the empirical frequency of observed outcomes conditional on each forecast, the cross-calibration test checks the empirical frequency of observed outcomes conditional on each profile of forecasts.

They showed that if an expert predicts according to the data-generating process, the expert is guaranteed to pass the cross-calibration test with probability 1, no matter what strategies the other experts use. In addition, they prove that in the presence of an informed expert, the subset of data-generating processes under which an ignorant expert (a charlatan) will pass the Cross-Calibration test with positive probability, is topologically "small". However, this test, like the test of Al-Najjar and Weinstein (2008), is anonymous but is not reasonable (a concise example is provided in Section 4) and is error-free (proof is straightforward and omitted).

1.3 Finite or Infinite Test?

A long-standing debate in the literature on expert testing is whether a test should be finite. A test is finite if its decision is made in some finite time. In contrast, an infinite test may require the infinite sequence of forecasts and realizations prior to making a verdict. The argument for considering finite tests is that infinite tests are impractical.

Although we sympathize with the argument that infinite tests are impractical we do think they have academic merit. The construction of well-behaved infinite, possibly impractical, tests would eventually shed light on their finite counterpart. Thus, if the technical analysis underlying the understanding of infinite tests is more tractable than that of finite tests then the study of infinite tests should be the port of embarkation for this research endeavor. This is what motivates our approach in this paper.

Furthermore, in expert testing we should allow experts to calibrate their model given the data. Pushing the design of tests towards finite tests may result in tests that give a verdict before these models are refined and calibrated. In a way the recent success of 'deep learning' based on enormous data sets (paralleling our interest in long-run observations) testifies to the importance of patience in model (expert) selection and the benefit of looking at many data points.

2 Model

At the beginning of each period $t = 1, 2, \ldots$ an outcome ω_t, drawn randomly by Nature from the set $\Omega = \{A, B\}$, is realized.[2] Before ω_t is realized two self-proclaimed experts (sometimes referred to as forecasters) simultaneously announce their forecast, in the form of a probability distribution over Ω. We assume that both forecasters observe all past realizations and all previous pairs of forecasts.

An element $\omega = \{\omega_1, \omega_2, \ldots\} \in \Omega^\infty$ is called a realization of outcomes and we denote by $\omega^t = \{\omega_1, \omega_2, \ldots, \omega_t\}$ the partial history of outcomes up to period t (by convention ω^0 denotes the empty history). A cylinder $C_{\omega^t} \triangleq \{\hat{\omega} \in \Omega^\infty | \; \hat{\omega_n} = \omega_n, \; 1 \leq n \leq t\}$ is a set of realizations which share a common partial history of outcomes up to period t. Let g_t denote

[2]For expository reasons we restrict attention to a binary set $\Omega = \{A, B\}$. The results extend to any finite set.

the σ-algebra on Ω^∞ generated by the cylinder sets C_{ω^t} and let $g_\infty \triangleq \sigma(\bigcup_{t=0}^{\infty} g_t)$ denote the smallest σ-algebra which consists of all cylinders (also known as the Borel σ-algebra).[3] In addition, we endow Ω^∞ with the product topology, that is, the topology that is generated by a basis of cylinders. Let $\Delta(\Omega^\infty)$ be the set of all probability measures which are defined over the measure space $(\Omega^\infty, g_\infty)$.

Let $H^t \triangleq (\Omega \times \Delta(\Omega) \times \Delta(\Omega))^t$ (with $H^0 \triangleq \emptyset$) be the set of all partial histories of length t and define the set of all possible histories by $H \triangleq \bigcup_{t \geq 0} H^t$.

A (pure) forecasting strategy f is a function that maps finite histories to a probability distribution over Ω. Formally, $f: \ H \longrightarrow \Delta(\Omega)$.

Note that each forecast provided by one expert may depend, inter alia, on those provided by the other expert in previous stages. We denote by F the set of all forecasting strategies.

A probability measure $P \in \Delta(\Omega^\infty)$ naturally induces a (set of) corresponding forecasting strategy, denoted f_P, that satisfies any $\omega^t \in \Omega^t$ such that $P(C_{\omega^t}) > 0$ and any $\omega_{t+1} \in \Omega$

$$f_P((\omega^t, \cdot, \cdot))[\omega_{t+1}] = P(C_{(\omega^t, \omega_{t+1})} | C_{\omega^t}).$$

In other words, the forecasting strategy f_P derives its forecasts from the original measure, P, via the Bayes rule. Note that this does not restrict the forecast of f_P over cylinders, C_{ω^t}, for which $P(C_{\omega^t}) = 0$.

In the other direction, let $\vec{f} \triangleq (f_0, f_1)$ be a pair of forecasting strategies. Then it induces a pair of measures over Ω^∞, as follows. A realization ω together with $\vec{f}$ induce a unique play path:

$$h_{(\omega, f_0, f_1)} \triangleq (\omega_t, f^\omega_{0,t}, f^\omega_{1,t})_{t=0}^\infty$$

starting at the Null history, where $\omega_t, f^\omega_{i,t}$ denote the outcome at time t, and the probabilistic prediction of expert i for that period, respectively. Additionally, fixing time $n \geq 0$, the prefix (of length n) and the suffix (starting an n) of $h_{(\omega, f_0, f_1)}$ are denoted by $h_{(\omega^n, f_0, f_1)}$ and $h_{(n, \omega, f_0, f_1)}$, respectively.

Now consider a cylinder C_{ω^t} and let $\tilde{\omega} \in C_{\omega^t}$, since any measure in $\Delta(\Omega^\infty)$ is determined by its cylinders, it follows that the pair $\vec{f}$ uniquely induces a pair of measures, which are correspondingly denoted by $(P_0^{\vec{f}}, P_1^{\vec{f}})$, via

$$P_i^{\vec{f}}(C_{\omega^t}) = \prod_{n=0}^{t-1} f_i(h_{(\tilde{\omega}^n, f_0, f_1)})[\tilde{\omega}_{n+1}], \ i \in \{0, 1\}. \tag{1}$$

2.1 Comparison Test

Definition 1. A *comparison test* is a measurable function which inputs a pair of two forecasting strategies and a realization in Ω^∞ and which outputs a rank (weak order) over

[3] $g_0 \triangleq \{\emptyset, \Omega^\infty\}$.

the two experts. Formally,

$$T:\ \Omega^{\infty} \times F \times F \longrightarrow \{0, \frac{1}{2}, 1\}$$

where $T = i \neq \frac{1}{2}$ implies that expert i is claimed as better informed, while $T = \frac{1}{2}$ implies the test is inconclusive (this cannot be avoided, for example, when both experts agree on their forecasts).

Definition 2. A comparison test is called *symmetric* whenever

$$T(\omega, f_0, f_1) = 1 - T(\omega, f_1, f_0),\ \forall \omega \in \Omega^{\infty}, f_0, f_1 \in F.$$

In other words, the expert chosen by T does not depend on the expert's identity (0 or 1). Note, if $f_0 = f_1$ then any symmetric test T must output 0.5 for all $\omega \in \Omega^{\infty}$. Our definition of a comparison test T suggests that the outcome of the test may depend on predictions made over realizations that did not materialize. In this paper we study a restricted class of tests—those that do not depend on forecasts made over unobserved realizations. Formally, Let

$$h:\ \Omega^{\infty} \times F \times F \longrightarrow (\Omega \times \Delta(\Omega) \times \Delta(\Omega))^{\infty}$$

be a function that maps a triplet (ω, f_0, f_1) to its uniquely induced play path, $h_{(\omega, f_0, f_1)}$. A test T is called *prequential* if there exists a function

$$\hat{T}:\ (\Omega \times \Delta(\Omega) \times \Delta(\Omega))^{\infty} \longrightarrow \{0, \frac{1}{2}, 1\}$$

such that $T = \hat{T} \circ h$.

For any test, T, and a pair of forecasting strategies, $\vec{f} \triangleq (f_0, f_1) \in F \times F$, we denote by $A^{\vec{f}}_{T,k} \triangleq \{\omega |\ T(\omega, f_0, f_1) = k\}$ the set of realizations for which the test outputs k.

Definition 3. A test T is *error-free with respect to* $A \subseteq F$ if $\forall \vec{f} \triangleq (f_0, f_1) \in A \times A$

$$P^{\vec{f}}_0(A^{\vec{f}}_{T,1}) = 0.$$

A test T is called error-free if it is error-free with respect to the set of all forecasting strategies, F. In other words, if one of the forecasters uses P, then with P probability zero the other forecaster is identified as the better informed one. Obviously a test which constantly outputs $\frac{1}{2}$ is both anonymous, prequential and is also error-free.

Obviously this test has no value. Inspired by this we ask whether there are meaningful error-free tests. It turns out that error-free tests are very restrictive in how well they can distinguish between two forecasters. In fact, whenever the measures induced by the two forecasters are mutually absolutely continuous, then an error-free test cannot guarantee to identify the better informed expert. Formally,

Proposition 1. *Let $\vec{f} \triangleq (f_0, f_1) \in F \times F$ be such that $P_1^{\vec{f}} (\neq P_0^{\vec{f}}) \ll P_0^{\vec{f}}$. If T is error-free then*

$$P_0^{\vec{f}}(A_{T,0}^{\vec{f}}) < 1.$$

On a set of realizations that has probability one according to forecaster 0 it is not necessarily the case that the test will identify him as better informed.

Proof. Assume that

$$P_0^{\vec{f}}(A_{T,0}^{\vec{f}}) = 1. \tag{2}$$

Since $P_1^{\vec{f}} \ll P_0^{\vec{f}}$ it follows from (2) that

$$P_0^{\vec{f}}((A_{T,0}^{\vec{f}})^c) = 0 \Longrightarrow P_1^{\vec{f}}((A_{T,0}^{\vec{f}})^c) = 0.$$

Therefore

$$P_1^{\vec{f}}(A_{T,0}^{\vec{f}}) = 1,$$

which by the symmetry of T contradicts the assumption that T is error-free. $\square$

The next property of a comparison test asserts that for any set of realizations assigned zero probability by one forecaster and positive probability by the other forecaster, there must be some subset of realizations for which the other forecaster is deemed superior.

Definition 4. We say that a test T is *reasonable* if $\forall \vec{f} \triangleq (f_0, f_1) \in F \times F$, for any measurable set A, and for any $i \in \{0, 1\}$

$$P_i^{\vec{f}}(A) > 0 \text{ and } P_{1-i}^{\vec{f}}(A) = 0 \implies P_i^{\vec{f}}(A \cap A_{T,i}^{\vec{f}}) > 0. \tag{3}$$

It should be emphasized that reasonableness and error-free are not related notions. To see why error-free does not imply reasonableness just consider the constant error-free test $T \equiv \frac{1}{2}$. An example for which reasonableness does not imply error-free is left to the end of Subsection 3.3.

We now turn to our construction of a prequential, symmetric, error-free and reasonable comparison test.

3 An Error-Free and Reasonable Test

The following test is derived from the likelihood ratio of the two measures induced by the two forecasters. For $\vec{f} \triangleq (f_0, f_1) \in F \times F,\ \omega \in \Omega^\infty,\ t > 0$, let

$$D^t_{f_0} f_1(\omega) \triangleq \prod_{n=0}^{t-1} \frac{f_1(h_{(\omega^n, f_0, f_1)})[\omega_{n+1}]}{f_0(h_{(\omega^n, f_0, f_1)})[\omega_{n+1}]} \tag{4}$$

and define

$$\overline{D}_{f_0} f_1(\omega) \triangleq \begin{cases} \underset{t\to\infty}{limsup} D^t_{f_0} f_1(\omega) & if\ f_0(h_{(\omega^n, f_0, f_1)})[\omega_{n+1}] > 0,\ \forall n \geq 0 \\ +\infty & if\ f_0(h_{(\omega^n, f_0, f_1)})[\omega_{n+1}] = 0,\ for\ some\ n \end{cases}$$

$$\underline{D}_{f_0} f_1(\omega) \triangleq \begin{cases} \underset{t\to\infty}{liminf} D^t_{f_0} f_1(\omega) & if\ f_0(h_{(\omega^n, f_0, f_1)})[\omega_{n+1}] > 0,\ \forall n \geq 0 \\ +\infty & if\ f_0(h_{(\omega^n, f_0, f_1)})[\omega_{n+1}] = 0,\ for\ some\ n. \end{cases}$$

Note, given $P_0^{\vec{f}}, P_1^{\vec{f}}, t > 0$, the functions $\omega \longrightarrow P_0^{\vec{f}}(C_{\omega^t}),\ \omega \longrightarrow P_1^{\vec{f}}(C_{\omega^t})$ are g_t- measurable as well as

$$z_t(\omega) \triangleq \begin{cases} \frac{P_1^{\vec{f}}(C_{\omega^t})}{P_0^{\vec{f}}(C_{\omega^t})}, & P_0^{\vec{f}}(C_{\omega^t}) > 0 \\ +\infty, & P_0^{\vec{f}}(C_{\omega^t}) = 0. \end{cases}$$

Consequently, the functions, $\underline{D}_{f_0} f_1(\omega)$, $\overline{D}_{f_0} f_1(\omega)$ are measurable as lim inf, lim sup of z_t, respectively.

Definition 5. If $\overline{D}_{f_0} f_1(\omega) = \underline{D}_{f_0} f_1(\omega) < \infty$, we say that the forecasting strategy f_1 is differentiable with respect to the forecasting strategy f_0 at ω and write

$$D_{f_0} f_1(\omega) = \overline{D}_{f_0} f_1(\omega) = \underline{D}_{f_0} f_1(\omega).$$

$D_{f_0} f_1$ is the derivative of f_1 with respect to f_0. Consider the following non-counterfactual symmetric test

$$T_D(\omega, f_0, f_1) = \begin{cases} 1, & D_{f_1} f_0(\omega) = 0 \\ 0.5, & other \\ 0, & D_{f_0} f_1(\omega) = 0. \end{cases} \tag{5}$$

T_D will be called a derivative test. Expert i is pointed as the true forecaster at ω whenever the derivative of f_i with respect to $f_{(1-i)}$ exists and equals 0. Our next goal is to prove that T_D is error-free and a reasonable test. Let us first show that $D_{f_0}f_1$ exists and is finite $P_0^{\vec{f}}$ - a.e.

Lemma 1. *Let $B \triangleq \{B_i\}_{i\in\mathbb{N}}$ be an arbitrary sequence of cylinders and set $\overline{B} \triangleq \bigcup_{i\in\mathbb{N}} B_i$. Then, there exists an index set $J \subseteq \mathbb{N}$ such that $\{B_j\}_{j\in J}$ are pairwise disjoint, and $\overline{B} = \bigcup_{j\in J} B_j$.*

Proof. A cylinder is called maximal in $\overline{B}$ if it is not a subset of any other cylinders in B. Any cylinder in B is contained in some maximal cylinder in $\overline{B}$. Let $J \subseteq \mathbb{N}$ be such that $\{B_j\}_{j\in J}$ is the set of all distinct maximal cylinders. Since any two distinct maximal cylinders are disjoint it follows that $\overline{B} = \bigcup_{j\in J} B_j$. $\square$

Lemma 2. *Fix $0 < \alpha < \infty$ and let $A \subset \Omega^\infty$ be a measurable set. Then*

a) $A \subset \{\omega|\ \underline{D}_{f_0}f_1(\omega) \leq \alpha\} \Longrightarrow P_1^{\vec{f}}(A) \leq \alpha P_0^{\vec{f}}(A)$.

b) $A \subset \{\omega|\ \overline{D}_{f_0}f_1(\omega) \geq \alpha\} \Longrightarrow P_1^{\vec{f}}(A) \geq \alpha P_0^{\vec{f}}(A)$.

Proof. (a) Let A be a measurable set which satisfies the left side of (a) and let $U \subset \Omega^\infty$ be any open set such that $A \subset U$. Fix $\epsilon > 0$, then $\forall a \in A,\ \forall N > 0\ \exists t = t_{(a,N,\epsilon)} > N$ such that

$$D_{f_0}^t f_1(a) = \frac{\prod_{n=0}^{t-1} f_1(h_{(a^n,f_0,f_1)})[a_{n+1}]}{\prod_{n=0}^{t-1} f_0(h_{(a^n,f_0,f_1)})[a_{n+1}]} = \frac{P_1^{\vec{f}}(C_{a^t})}{P_0^{\vec{f}}(C_{a^t})} \leq (\alpha+\epsilon). \tag{6}$$

Consider the following set of cylinders

$$\mathcal{F_A} \triangleq \{C_{a^t} \subset U|\ a \in A,\ t > 0,\ P_1^{\vec{f}}(C_{a^t}) \leq (\alpha+\epsilon)P_0^{\vec{f}}(C_{a^t})\}.$$

Note, it follows from (6) that $\mathcal{F_A}$ is not empty where $sup\{t|\ C_{a^t} \in \mathcal{F_A}\} = \infty$. By Lemma 1 we are provided with an index set J and a collection of pairwise disjoint sets $\{B_j \in \mathcal{F_A}\}_{j\in J}$ such that

$$\overline{B}_{\mathcal{F_A}} \triangleq \bigcup_{B\in\mathcal{F_A}} B = \bigcup_{j\in J} B_j. \tag{7}$$

Hence

$$P_1^{\vec{f}}(A) \underset{A\subseteq\overline{B}_{\mathcal{F_A}}}{\leq} P_1^{\vec{f}}(\overline{B}_{\mathcal{F_A}}) \underset{(7)}{=} P_1^{\vec{f}}(\bigcup_{j\in J} B_j) \leq \sum_{j\in J} P_1^{\vec{f}}(B_j)$$

$$\underset{B_j\in\mathcal{F_A}}{\leq} \sum_{j\in J}(\alpha+\epsilon)P_0^{\vec{f}}(B_j) = (\alpha+\epsilon)\sum_{j\in J} P_0^{\vec{f}}(B_j) \underset{U\supset B_j's\ are\ disjoint}{\leq} (\alpha+\epsilon)P_0^{\vec{f}}(U).$$

Since the above inequalities hold for any open set U which contains A and

$$P_0^{\vec{f}}(A) = \underset{A \subset U-open}{inf} \{P_0^{\vec{f}}(U)\},$$

it follows that $\forall \epsilon > 0$

$$P_1^{\vec{f}}(A) \leq (\alpha + \epsilon) P_0^{\vec{f}}(A)$$

which completes the proof of Case (a). The proof of Case (b) is analogous and hence omitted. $\square$

We now turn to show that the derivative of one measure with respect to another exists and is finite almost surely.

Lemma 3. *Let $f_0, f_1 \in F$, then $D_{f_0} f_1$ exists and is finite $P_0^{\vec{f}}$ - a.e.*

Proof. Let $S \triangleq \{\omega | \ \overline{D}_{f_0} f_1(\omega) = +\infty\}$. Therefore $\forall \alpha > 0$

$$S \subset \{\omega | \ \overline{D}_{f_0} f_1(\omega) \geq \alpha\},$$

and it follows from Lemma 2 that $P_0^{\vec{f}}(S) \leq \frac{1}{\alpha} P_1^{\vec{f}}(S)$. Now let $\alpha \to \infty$ to obtain

$$P_0^{\vec{f}}(S) = 0, \tag{8}$$

and consequently $\overline{D}_{f_0} f_1$ is finite $P_0^{\vec{f}} - a.e.$ For the second part let

$$R(a, b) \triangleq \{\omega | \ \underline{D}_{f_0} f_1(\omega) < a < b < \overline{D}_{f_0} f_1(\omega) < \infty\}.$$

Note that

$$R(a, b) \subset \{\omega | \ \underline{D}_{f_0} f_1(\omega) \leq a\}$$

$$R(a, b) \subset \{\omega | \ \overline{D}_{f_0} f_1(\omega) \geq b\}$$

where applying Lemma 2 gives: $bP_0^{\vec{f}}(R(a, b)) \leq P_1^{\vec{f}}(R(a, b)) \leq aP_0^{\vec{f}}(R(a, b))$. Hence,

$$P_0^{\vec{f}}(R(a, b)) = 0, \ \forall 0 < a < b \tag{9}$$

where from (8), (9) we obtain

$$P_0^{\vec{f}}(\{\omega | \ \underline{D}_{f_0} f_1(\omega) < \overline{D}_{f_0} f_1(\omega) < \infty\})$$

$$= P_0^{\vec{f}}(\bigcup_{\substack{0<a<b \\ a,b \in \mathbb{Q}}} R(a, b)) \leq \sum_{\substack{0<a<b \\ a,b \in \mathbb{Q}}} P_0^{\vec{f}}(R(a, b)) = 0.$$

Therefore, $D_{f_1} f_0$ exists $P_0^{\vec{f}}$- a.e. $\square$

3.1 The Properties of T_D

Now that we have established the existence and the finiteness of the test T_D, let us prove it is a reasonable and error-free test. We do this in two separate claims:

Claim 1. T_D is a reasonable test.

Proof. Let A be a measurable set such that
$$P_0^{\vec{f}}(A) > 0 \ \ and \ \ P_1^{\vec{f}}(A) = 0. \tag{10}$$
For $a > 0$ let
$$R_a \triangleq A \cap \{\omega | \ 0 < a \leq D_{f_0} f_1(\omega) < \infty\}, \ a > 0.$$
Note that if $P_0^{\vec{f}}(R_a) > 0$ then applying Lemma 2 for the set R_a yields,
$$0 = P_1^{\vec{f}}(R_a) \geq a P_0^{\vec{f}}(R_a) > 0$$
which contradicts (10). Therefore
$$P_0^{\vec{f}}(A \cap \{\omega | \ 0 < D_{f_0} f_1(\omega) < \infty\}) = P_0^{\vec{f}}(\bigcup_{\substack{0<a \\ a \in \mathbb{Q}}} R_a) \leq \sum_{\substack{0<a \\ a \in \mathbb{Q}}} P_0^{\vec{f}}(R_a) = 0.$$
Since by Lemma 3 $D_{f_0} f_1$ exists and finite $P_0^{\vec{f}} - a.e.$, we conclude that
$$P_0^{\vec{f}}(A \cap \{\omega | \ D_{f_0} f_1(\omega) = 0\}^c) = 0.$$
Hence
$$0 < P_0^{\vec{f}}(A) = P_0^{\vec{f}}(A \cap \{\omega | \ D_{f_0} f_1(\omega) = 0\}) \underset{(5)}{=} P_0^{\vec{f}}(A \cap A_{T_D,0}^{\vec{f}}),$$
and the test is reasonable. □

Claim 2. T_D is error-free.

Proof. Note that
$$A_{T_D,1}^{\vec{f}} \underset{(5)}{=} \{\omega | \ D_{f_1} f_0(\omega) = 0\}$$
$$= \{\omega | \ \lim_{t \to \infty} D_{f_1}^t f_0(\omega) = 0 \ \ and \ \ \forall n > 0, \ f_1(h_{(\omega^n, f_0, f_1)})[\omega_{n+1}] > 0\}$$
$$\subset \{\omega | \ \lim_{t \to \infty} D_{f_0}^t f_1(\omega) = \infty\} \cup \{\omega | \ \exists n > 0 \ s.t \ f_0(h_{(\omega^n, f_0, f_1)})[\omega_{n+1}] = 0\}$$
$$\subset \{\omega | \ \underline{D}_{f_0} f_1(\omega) = \overline{D}_{f_0} f_1(\omega) = \infty\}.$$
By Lemma 3, $D_{f_0} f_1$ is finite $P_0^{\vec{f}} - a.e.$ Thus

$$P_0^{\vec{f}}(A_{T_D,1}^{\vec{f}}) \leq P_0^{\vec{f}}(\{\omega |\ \underline{D}_{f_0} f_1(\omega) = \overline{D}_{f_0} f_1(\omega) = \infty\}) = 0$$

and T_D is error-free. $\square$

3.2 The Uniqueness of T_D

The next theorem asserts that there exists a unique reasonable and error-free test. That is, any error-free test $T \nsim T_D$ which is reasonable, admits an error. Let us first establish what we mean by $T \nsim T_D$:

Definition 6. Let $\vec{f} \triangleq (f_0, f_1) \in F \times F$. We say that $T \sim_{\vec{f}} \hat{T}$ if $\forall i \in \{0, 1\}$

$$P_i^{\vec{f}}(\{\omega |\ T(\omega, f_0, f_1) \neq \hat{T}(\omega, f_0, f_1)\}) = 0.$$

We say that $T \sim \hat{T}$ if and only if $T \sim_{\vec{f}} \hat{T},\ \forall \vec{f} \in F \times F$.

Claim 3. $\sim$ is an equivalence relation on $\top = \{T |\ T - \textit{comparison test}\}$.

Proof. Let $T, T_1, T_2 \in \top,\ \vec{f} \in F \times F,\ \textit{and}\ i \in \{0, 1\}$.
Reflexivity:

$$P_i^{\vec{f}}(\{\omega |\ T(\omega, f_0, f_1) \neq T(\omega, f_0, f_1)\}) = 0 \Longrightarrow T \sim T.$$

Symmetry:

$$P_i^{\vec{f}}(\{\omega |\ T_1(\omega, f_0, f_1) \neq T_2(\omega, f_0, f_1)\}) = 0 \iff P_i^{\vec{f}}(\{\omega |\ T_2(\omega, f_0, f_1) \neq T_1(\omega, f_0, f_1)\}) = 0,$$

hence, $T_1 \sim T_2 \iff T_2 \sim T_1$.
Transitivity: Assume that $T_1 \sim T$, and $T \sim T_2$, hence

$$T_1 \sim_{\vec{f}} T \Longrightarrow P_i^{\vec{f}}(\{\omega |\ T_1(\omega, f_0, f_1) \neq T(\omega, f_0, f_1)\}^c) = 1,$$
$$T \sim_{\vec{f}} T_2 \Longrightarrow P_i^{\vec{f}}(\{\omega |\ T(\omega, f_0, f_1) \neq T_2(\omega, f_0, f_1)\}^c) = 1,$$

and so

$$1 = P_i^{\vec{f}}(\{\omega |\ T_1(\omega, f_0, f_1) \neq T(\omega, f_0, f_1)\}^c \cap \{\omega |\ T(\omega, f_0, f_1) \neq T_2(\omega, f_0, f_1)\}^c) =$$
$$P_i^{\vec{f}}(\{\omega |\ T_1(\omega, f_0, f_1) \neq T_2(\omega, f_0, f_1)\}^c),$$

yielding $P_i^{\vec{f}}(\{\omega |\ T_1(\omega, f_0, f_1) \neq T_2(\omega, f_0, f_1)\}) = 0$, and therefore $T_1 \sim_{\vec{f}} T_2$. $\square$

We now turn to show that at all pairs (f_0, f_1) where $T \nsim_{\vec{f}} T_D$, T must admit an error, and on top of that, if T is a tail test, then there exists a pair $(\tilde{f_0}, \tilde{f_1})$ such that the probability of the error term is arbitrarily large.

Theorem 1. *Let T be a reasonable test. If $T \nsim T_D$ then T is not error-free.*

Proof. Assume by contradiction that T is error-free. Let $\vec{f} \triangleq (f_0, f_1)$ be such that $T \nsim_{\vec{f}} T_D$ at $\vec{f}$, then (w.l.o.g. for $i = 0$) $\exists k, l (\neq k) \in \{0, \frac{1}{2}, 1\}$ such that

$$P_0^{\vec{f}}(A_{T,l}^{\vec{f}} \cap A_{T_D,k}^{\vec{f}}) > 0.$$

In addition, by Claim 2, T_D is error-free, therefore

$$P_0^{\vec{f}}(A_{T,1}^{\vec{f}}) = P_0^{\vec{f}}(A_{T_D,1}^{\vec{f}}) = 0,$$

and consequently,

$$P_0^{\vec{f}}(\hat{A}_1 \triangleq (A_{T,0}^{\vec{f}} \cap A_{T_D,\frac{1}{2}}^{\vec{f}})) > 0 \quad or \quad P_0^{\vec{f}}(\hat{A}_2 \triangleq (A_{T,\frac{1}{2}}^{\vec{f}} \cap A_{T_D,0}^{\vec{f}})) > 0.$$

Case 1: $P_0^{\vec{f}}(\hat{A}_1) > 0$. By Claim 1, T_D is reasonable; thus

$$P_1^{\vec{f}}(\hat{A}_1) = 0 \Longrightarrow P_0^{\vec{f}}(\hat{A}_1 \cap A_{T_D,0}^{\vec{f}}) > 0,$$

which leads to a contradiction since $A_{T_D,0}^{\vec{f}}$, $A_{T_D,\frac{1}{2}}^{\vec{f}}$ are disjoint. Thus

$$P_1^{\vec{f}}(A_{T,0}^{\vec{f}}) > 0,$$

which contradicts the assumption that T is error-free.

Case 2: $P_0^{\vec{f}}(\hat{A}_2) > 0$. By the assumption T is a reasonable test where, by Claim 2, T_D is error-free; therefore the contradiction

$$P_1^{\vec{f}}(A_{T_D,0}^{\vec{f}}) > 0$$

follows analogously from Case 1. □

Remark 1. The key properties can be usefully viewed as an implication of Lebesgue decomposition (Billingsley 1995, Section 31). Here it is applied in a symmetric way with some flexibility on how measure-zero sets are handled.

3.3 Tail Test

We now turn to introduce the notion of a tail test. We will argue that the test T_D is a tail test. In fact, our next theorem asserts that any other tail test that is reasonable will not only admit an error but will admit an arbitrarily large error.

Definition 7. T is a tail test if $\forall(\omega, f_0, f_1), (\omega', f_0', f_1') \in \Omega^\infty \times F \times F$:

If $\exists n > 1$ such that

$$h_{(n,\omega,f_0,f_1)} = h_{(n,\omega',f_0',f_1')} \;\; and \;\; f_{i,t}^{\omega}, f_{i,t}^{\prime\omega'} > 0,\; \forall 1 \leq t \leq n-1,\; i \in \{0,1\}, \tag{11}$$

then $T((\omega, f_0, f_1)) = T((\omega', f_0', f_1'))$.

In other words, a tail test must yield the same output for any given two pairs of forecasting strategies whose play paths eventually coincide after some time $n > 1$, and whose conditionals are strictly positive up to time n.

Claim 4. T_D is a tail test.

Proof. Let $(\omega, f_0, f_1), (\omega', f_0', f_1') \in \Omega^\infty \times F \times F$ and $n > 1$ for which (11) holds. Let $(\omega'', f_0'', f_1'') \in \Omega^\infty \times F \times F$ which satisfies

$$\forall t > 0: \quad (\omega_t'', f_{0,t}^{\omega''}, f_{1,t}^{\omega''}) = (\omega_{t+n}, f_{0,t+n}^{\omega}, f_{1,t+n}^{\omega}). \tag{12}$$

Since $D_{f_0}^t f_1(\omega) > 0$ by the left part of (11) it follows that[4]

$$0 = T_D((\omega'', f_0'', f_1'')) \underset{(5)}{\iff} 0 = D_{f_0}^t f_1(\omega) \cdot D_{f_0''} f_1''(\omega'') \underset{(12)}{=} D_{f_0} f_1(\omega) \iff T_D((\omega, f_0, f_1)) = 0.$$

Additionally, by the same consideration we have

$$1 = T_D((\omega'', f_0'', f_1'')) \iff 0 = D_{f_1''} f_0''(\omega'') \iff T_D((\omega, f_0, f_1)) = 1,$$

and therefore

$$T_D((\omega'', f_0'', f_1'')) = T_D((\omega, f_0, f_1)).$$

Similarly, we show that $T_D((\omega'', f_0'', f_1'')) = T_D((\omega', f_0', f_1'))$ by replacing $(\omega_t'', f_{0,t}^{\omega''}, f_{1,t}^{\omega''})$ with $(\omega_t', f_{0,t}^{\omega'}, f_{1,t}^{\omega'})$ in (12) and the Claim is concluded. $\square$

Before we address the main theorem let us prove the following claim.

Claim 5. If T is reasonable then for any measurable set A

$$P_i^{\vec{f}}(A \cap A_{T,k}^{\vec{f}}) > 0 \implies P_{1-i}^{\vec{f}}(A \cap A_{T,k}^{\vec{f}}) > 0,\; i \in \{0,1\},\; k \neq i.$$

[4] Note that $D_{f_0''} f_1''(\omega'') = 0$ if and only if $D_{f_0} f_1(\omega'')$ exists and equals 0.

Proof. Let A be a measurable set and (w.l.o.g) assume by contradiction that

$$P_1^{\vec{f}}(A \cap A_{T,k}^{\vec{f}}) > 0 \ \wedge \ P_0^{\vec{f}}(A \cap A_{T,k}^{\vec{f}}) = 0, \ k \in \{0, \frac{1}{2}\}.$$

T is reasonable thus (3) yields $P_1^{\vec{f}}(A \cap A_{T,k}^{\vec{f}} \cap A_{T,1}^{\vec{f}}) > 0$, which contradicts the fact that $A_{T,k}^{\vec{f}}, A_{T,1}^{\vec{f}}$ are disjoint sets. □

Theorem 2. *Let T be a reasonable tail test. If $T \nsim T_D$ then $\forall 0 < \epsilon < 1 \ \exists \vec{\hat{f}} \triangleq (\hat{f}_0, \hat{f}_1)$ such that*

$$P_0^{\vec{\hat{f}}}(A_{T,1}^{\vec{\hat{f}}}) > 1 - \epsilon \ \ or \ \ P_1^{\vec{\hat{f}}}(A_{T,0}^{\vec{\hat{f}}}) > 1 - \epsilon.$$

Proof. By Theorem 1 (w.l.o.g.) there exists a pair $\vec{f} = (f_0, f_1)$ such that $P_1^{\vec{f}}(A_{T,0}^{\vec{f}}) > 0$. In addition, since $A_{T,0}^{\vec{f}}$ is $g_\infty - measurable$ we can apply the Levy upwards theorem (Williams 1991, Theorem 14.2.) to obtain

$$\begin{aligned}
&\lim_{t\to\infty} P_1^{\vec{f}}(A_{T,0}^{\vec{f}} |\, g_t) = \\
&\lim_{t\longrightarrow\infty} E^{P_1^{\vec{f}}}[\mathbf{1}_{A_{T,0}^{\vec{f}}} \,|\, g_t] = E^{P_1^{\vec{f}}}[\mathbf{1}_{A_{T,0}^{\vec{f}}} \,|\, g_\infty] = \mathbf{1}_{A_{T,0}^{\vec{f}}}, \quad P_1^{\vec{f}} - a.s..
\end{aligned}$$

Therefore, $\exists B^{\vec{f}} \subset A_{T,0}^{\vec{f}}$ with $P_1^{\vec{f}}(B^{\vec{f}}) = P_1^{\vec{f}}(A_{T,0}^{\vec{f}})$ such that $\forall \omega \in B^{\vec{f}}$ we have

$$\lim_{t\to\infty} P_1^{\vec{f}}(A_{T,0}^{\vec{f}} |\, C_{\omega^t}) = 1 \ \ and \ \ f_{1,t}^{\omega} > 0, \ \forall 1 \le t \le n-1. \tag{13}$$

Let $0 < \epsilon < 1$. Fix $\tilde{\omega} \in B^{\vec{f}}$ and observe that from (13) $\exists n = n_{(\epsilon,\omega)} > 0$ such that $\forall t \ge n$

$$P_1^{\vec{f}}(A_{T,0}^{\vec{f}} \cap C_{\tilde{\omega}^n}) = P_1^{\vec{f}}(A_{T,0}^{\vec{f}} |\, C_{\tilde{\omega}^n}) P_1^{\vec{f}}(C_{\tilde{\omega}^n}) > (1 - \epsilon).$$

Thus, applying Claim 5 with the set $C_{\tilde{\omega}^n}$ yields $P_0^{\vec{f}}(A_{T,0}^{\vec{f}} \cap C_{\tilde{\omega}^n}) > 0$ and consequently $f_{0,t}^{\tilde{\omega}} > 0, \ \forall 1 \le t \le n-1$ is inferred from (1). Now, modify $\vec{f}$ to be the forecasting strategy $\vec{\hat{f}}$ whose one step ahead conditionals satisfy

$$\hat{f}_i(\omega^t)[\omega_{t+1}] = \begin{cases} 1 & \omega^t = \tilde{\omega}^t, \ t < n \\ f_i(\omega^t)[\omega_{t+1}] \ , & \omega^n = \tilde{\omega}^n, \ t \ge n \\ 0 & other. \end{cases} .$$

[5] Note, by construction, $\forall \omega \in A^{\vec{f}}_{T,0} \cap C_{\tilde{\omega}^n},\ i \in \{0,1\},\ 1 \leq t \leq n-1$ we have

$$h_{(n,\omega,f_0,f_1)} = h_{(n,\omega,\hat{f}_0,\hat{f}_1)} \ \ \textit{and} \ \ f^{\omega}_{i,t} = f^{\tilde{\omega}}_{i,t} > 0,\ \hat{f}^{\omega}_{i,t} = \hat{f}^{\tilde{\omega}}_{i,t} > 0,$$

and since T is a tail test it follows from (11) that $T(\omega, \hat{f}_0, \hat{f}_1) = T(\omega, f_0, f_1) = 0$. As a result, $\omega \in A^{\vec{\hat{f}}}_{T,0} \cap C_{\tilde{\omega}^n}$ yielding

$$P^{\vec{\hat{f}}}_1(A^{\vec{\hat{f}}}_{T,0}) \geq P^{\vec{\hat{f}}}_1(A^{\vec{\hat{f}}}_{T,0} \cap C_{\tilde{\omega}^n}) \geq P^{\vec{f}}_1(A^{\vec{f}}_{T,0} \cap C_{\tilde{\omega}^n}) > 1 - \epsilon,$$

and therefore completes the proof. □

The next example shows that the result in Theorem 2 cannot be obtained if we relax the requirement that T is a tail test; it furthermore illustrates that reasonableness does not imply error-free.

Example 1. Let

$$\overrightarrow{h} \triangleq (A, h_{0,t}, h_{1,t})_{t=1}^{\infty},\ \overleftarrow{h} \triangleq (A, h_{1,t}, h_{0,t})_{t=1}^{\infty}$$

be play paths which satisfy

$$h_{0,t}(A) \equiv 1,\ h_{1,1}(A) = \frac{1}{2},\ h_{1,t}(A) = 1,\ t \geq 2$$

and consider the following test

$$T((\omega, f_0, f_1)) = \begin{cases} T_D((\omega, f_0, f_1)), & \textit{other} \\ 0, & h_{(\omega,f_0,f_1)} = \overrightarrow{h} \\ 1, & h_{(\omega,f_0,f_1)} = \overleftarrow{h}. \end{cases}$$

Note, for every triplet (ω, f_0, f_1), whose induced play path coincides with $\overrightarrow{h}$ or $\overleftarrow{h}$, there exists $i \in \{0,1\}$ such that

$$P^{\vec{f}}_i(A^{\vec{f}}_{T,-i}) = P^{\vec{f}}_i(\{(A)_{t=1}^{\infty}\}) = \frac{1}{2} < 1 \tag{14}$$

where the left equality holds since T_D is error-free. Moreover, since $P^{\vec{f}}_i(\{(A)_{t=1}^{\infty}\}) > 0$ for $i \in \{0,1\}$ and T_D is a reasonable test, it follows that T is reasonable even as it admits a bounded error by (14).

[5]Note that the corresponding forecasting strategy $\hat{f}_i$ determines the one step ahead forecasts along ω, only through the history of outcomes and does not depend on the full histories, i.e $\hat{f}_i(\omega)[\omega_{t+1}] = P^{\vec{f}}_i(\omega_{t+1}|C_{\omega^t}),\ \forall 0 \leq t \leq n-1,\ i \in \{0,1\}$.

4 Ideal Tests

Recall that an error-free test eliminates the necessity of pointing out the less informed expert. A stronger and more appealing property is to point out the better informed expert. We consider tests that exhibit such a property as Ideal:

Definition 8. A symmetric test T is *Ideal with respect to* $A \subseteq F$ if for all $f \triangleq (f_0, f_1 \neq f_0) \in A \times A$.

$$P_0^{\vec{f}}(A_{T,0}^{\vec{f}}) = 1.$$

It is called *Ideal* if it is Ideal with respect to F.

In other words, whenever expert i knows the actual data generating process and expert $1-i$ does not, an Ideal test will surely identify the informed expert.

Trivially, any Ideal test with respect to a subset of forecasts A is also error-free with respect to the same set. The following is a straightforward corollary of Theorem 1.

Corollary 1. *There exists no Ideal test with respect to a set of forecasts A whenever it contains two forecasts which induce measures, one of which is absolutely continuous with respect to the other.*

This immediately entails:

Corollary 2. *There exists no Ideal test.*

However, whenever A contains no such pair of forecasts, then an Ideal test does exist. To prove this we must first accurately define the notion of mutually singular forecasts.

Definition 9. Two forecasting strategies, $f_0, f_1 (\neq f_0) \in F$, are said to be mutually singular with respect to each other, if there exist two disjoint sets

$$C_0^{\vec{f}}, C_1^{\vec{f}} \subset (\Omega \times \Delta(\Omega) \times \Delta(\Omega))^\infty$$

such that[6]

$$P_0^{\vec{f}}(\{\omega |\ h_{(\omega, f_0, f_1)} \in C_0^{\vec{f}}\}) = P_1^{\vec{f}}(\{\omega |\ h_{(\omega, f_0, f_1)} \in C_1^{\vec{f}}\}) = 1.$$

A set $A \subseteq F$ is pairwise mutually singular if $\forall f_0, f_1 (\neq f_0) \in A$, f_0, f_1 are mutually singular with respect to each other.

The next lemma asserts that a reasonable test is able to perfectly distinguish between far measures which are induced from forecasting strategies which are mutually singular with respect to each other.

[6]Recall that $\vec{f}$ induces a unique play path $h_{(\omega, f_0, f_1)}$.

Lemma 4. *Let $f_0, f_1 \neq f_0 \in F$ which are mutually singular with respect to each other. If T is reasonable then*

$$P_i^{\vec{f}}(A_{T,i}^{\vec{f}}) = 1,\ i \in \{0,1\}.$$

Proof. W.l.o.g. let B be such that: $P_0^{\vec{f}}(B) = 1,\ P_1^{\vec{f}}(B) = 0$. T is reasonable, therefore $P_0^{\vec{f}}(B \cap A_{T,0}^{\vec{f}}) > 0$ from (3). Let $k \in \{\frac{1}{2}, 1\}$ and assume that

$$P_0^{\vec{f}}(B \cap A_{T,k}^{\vec{f}}) > 0.$$

Applying Claim (5) with the set B yields

$$P_1^{\vec{f}}(B \cap A_{T,k}^{\vec{f}}) > 0$$

which contradicts the assumption that $P_1^{\vec{f}}(B) = 0$. Hence, $P_0^{\vec{f}}(B \cap A_{T,k}^{\vec{f}}) = 0$. As a result,

$$\begin{aligned} 1 = P_0^{\vec{f}}(B) = P_0^{\vec{f}}(B \cap A_{T,0}^{\vec{f}}) + P_0^{\vec{f}}(B \cap A_{T,\frac{1}{2}}^{\vec{f}}) \\ + P_0^{\vec{f}}(B \cap A_{T,1}^{\vec{f}}) = P_0^{\vec{f}}(B \cap A_{T,0}^{\vec{f}}) \end{aligned}$$

and therefore $P_0^{\vec{f}}(A_{T,0}^{\vec{f}}) = 1$. □

Note that Lemma 4 holds even for T which is not error-free.

We now demonstrate (in a concise way) that the Cross-Calibration test which was introduced by Feinberg and Stewart (2008) is error-free and nevertheless is not reasonable.[7]

Example 2. Consider the Cross-Calibration test

$$T_{cross}:\ \Omega^{\infty} \times (\Delta(\Omega))^{\infty} \times (\Delta(\Omega))^{\infty} \longrightarrow \{0,1\}^2,$$

and let $g : \{0,1\}^2 \longrightarrow \{0, \frac{1}{2}, 1\}$ be defined by: $g(1,1) = g(0,0) = \frac{1}{2},\ g(1,0) = 0,\ g(0,1) = 1$.[8] [9] We define the induced Cross-Calibration comparison test by

$$T = g \circ T_{cross}. \tag{15}$$

Let $P_0^{\vec{f}}, P_1^{\vec{f}}$ ($\vec{f} \triangleq (f_0, f_1)$) be the Dirac measures on $\overset{0}{\omega} \triangleq (0, A, A, , , A, , ,)$, $\overset{1}{\omega} \triangleq (A, A, A, , , A, , ,)$, respectively. Since both $P_0^{\vec{f}}$ and $P_1^{\vec{f}}$ pass the Cross-Calibration test on $\overset{0}{\omega}$ it follows from (15) that

[7]As mentioned in the Introdction, the Cross-Calibration test checks the empirical frequencies of the realization conditional on each profile of forecasts that occurs infinitely often.

[8]For simplicity, we limit our attention to the two-forecasters model.

[9]The case for which $T_{cross}(\omega, \ , \ ) = (i,i),\ i \in \{0,1\}$ may be interpreted as both experts either pass($i = 1$) or fail($i = 0$) the test on ω.

$$T_{cross}(\omega^0, P_0^{\vec{f}}, P_1^{\vec{f}}) = (1,1) \iff (g \circ T_{cross})(\omega^0, P_0^{\vec{f}}, P_1^{\vec{f}}) = T(\omega^0, P_0^{\vec{f}}, P_1^{\vec{f}}) = \frac{1}{2},$$

which implies that

$$P_0^{\vec{f}}(A_{T,\frac{1}{2}}^{\vec{f}}) = 1. \tag{16}$$

However, note that $P_0^{\vec{f}}, P_1^{\vec{f}}$ are mutually singular with respect to each other; so if T was a reasonable test then, by Lemma 4, it would satisfy

$$P_0^{\vec{f}}(\{\omega^0\} \cap A_{T,0}^{\vec{f}}) > 0$$

which contradicts (16) and therefore T is not reasonable. The fact that T is error-free follows immediately from Dawid (1982) and hence omitted.

The next theorem provides a necessary and sufficient condition for the existence of an Ideal test over sets.

Theorem 3. *$\exists$ non-counterfactual Ideal test with respect to A if and only if* A is *pairwise mutually singular.*

Proof. $\Longleftarrow$ Directly follows from Lemma 4 and Claim 1.

$\Longrightarrow$Let T be a non-counterfactual Ideal symmetric test with respect to a set A with $h, \hat{T}$ as in Definition 2. Let $\vec{f} \triangleq (f_0, f_1 \neq f_0) \in F \times F$ and denote

$$C_i^T \triangleq \{h_{(\omega, f_0, f_1)} |\ \omega \in A_{T,i}^{\vec{f}}\},\ i \in \{0,1\}.$$

Since h is 1:1 and $A_{T,0}^{\vec{f}},\ A_{T,1}^{\vec{f}}$ are disjoint, it follows that $C_0^T,\ C_1^T$ are disjoint where T Ideal yields

$$1 = P_i^{\vec{f}}(A_{T,i}^{\vec{f}}) = P_i^{\vec{f}}(\{\omega |\ h_{(\omega, f_0, f_1)} \in C_i^T\}),\ i \in \{0,1\}.$$

$\square$

We conclude the paper with an example of a test over a domain of mutually singular forecasts:

Example 3. Let

$$A_{IID} \times A_{IID} \triangleq \{f \triangleq (f_0, f_1) \in F \times F |\ \forall i \in \{0,1\}\ \exists a_{f_i} \in [0,1]\ s.t\ , f(h^t)[A] \equiv a_{f_i},\ \forall h^t \in H^t\}.$$

For $\omega \in \Omega^\infty$ denote the average realization by

$$a_\omega \triangleq \lim_{n\to\infty}\left(\frac{\sum_{t=1}^{n} 1_{\{\omega_t=A\}}}{n}\right)$$

and consider the following comparable test

$$T(\omega, f_0, f_1) = \begin{cases} 1, & f_1(h^0)[A] = a_\omega \neq f_0(h^0)[A] \\ 0.5, & other \\ 0, & f_0(h^0)[A] = a_\omega \neq f_1(h^0)[A]. \end{cases}$$

Obviously, T is a well-defined symmetric and non-counterfactual. Showing that T is Ideal with respect to A_{IID} is a mere application of the law of large numbers.

5 Finite Case

In the model considered thus far the tester's verdict is determined at the end of all times. This seems impractical and so we turn to model a tester determines in finite time which expert is better. In fact, in the model below, at each stage, t, the tester makes up his mind. In theory, a test can swing back and forth between the two experts. The analysis below focuses on the cases where this does not happen and the tester essentially settles on one expert or the other.

Apart from moving from a verdict that is determined at the end of all times to verdicts made in finite time we make two additional distinctions. First, we allow the tester to randomize his decision at any time t. Second, we do not allow the tester to be indecisive. A test must point to one of the two experts.

A tester makes decisions in finite times and a test is a sequence of such decisions. Nevertheless, the properties of such tests, which we are interested in, are determined on the infinite sequence.

Let $I = \{0, 1\}$, be the decision set of a tester. The tester decides 0 or 1 whenever he prefers expert 0 or 1, respectively. At each time $t > 0$, the tester uses a mixed strategy before announcing which expert he prefers by flipping a coin with a parameter which is determine at time t. The tester ends up with a sequence of decisions which is denoted by I^∞.

5.1 A Finite Comparison Test

A test is, therefore, a function that maps any finite history of realizations and pairs of forecasts into a (random) decision of which of the two experts is preferred. Formally,

Definition 10. A finite comparison test is a sequence $\overrightarrow{T} := (T_t)_{t>0}$, where

$$T_t : \ (\Omega \times \Delta(\Omega) \times \Delta(\Omega) \times I)^{\infty} \longrightarrow \Delta(I),$$

is g_t−measurable.

Note that $\overrightarrow{T}$ together with a pair $(\omega, \overrightarrow{f})$ induce a unique measure on the set I^{∞}, which we denote by $P_{\overrightarrow{T}}^{(\omega,\overrightarrow{f})}$, meaning $P_{\overrightarrow{T}}^{(\omega,\overrightarrow{f})} \in \Delta(I^{\infty})$. In addition, we denote an element in I^{∞} by $\vec{i} \triangleq (\mathrm{i}_t)_{t>0} \in I^{\infty}$, and for a given $i \in I$ we define by

$$dense(\vec{i})[i] := \liminf_{t\to\infty} \frac{\sum_{i=1}^{t} 1_{\{i_t=i\}}(\vec{i})}{t},$$

to be the density of a decision i along the sequence of decisions . Note that $dense(\vec{i})[i] = 1 \iff dense(\vec{i})[1-i] = 0$.

Definition 11. $\overrightarrow{T}$ is decisive on $i \in I$ at $(\omega, \overrightarrow{f})$ if

$$P_{\overrightarrow{T}}^{(\omega,\overrightarrow{f})}(\{\vec{i} \in I^{\infty} : dense(\vec{i})[i] = 1\}) = 1,$$

$\overrightarrow{T}$ is indecisive at $(\omega, \overrightarrow{f})$ if $\overrightarrow{T}$ is not decisive on i, $\forall i \in I$. [10][11]

Definition 12. $\overrightarrow{T}$ is error-free if $\forall \vec{f} \in F \times F$, [12]

$$P_0^{\vec{f}} \times P_{\overrightarrow{T}}^{(\omega,\overrightarrow{f})}(\{(\omega, \vec{i}) \in \Omega^{\infty} \times I^{\infty} : dense(\vec{i})[1] = 1\}) = 0.$$

Definition 13. $\overrightarrow{T}$ is reasonable if $\forall \vec{f} \in F \times F$, and $\forall C - measurable$ [13]

$$P_0^{\vec{f}}(C) > 0 \; and \; P_1^{\vec{f}}(C) = 0 \Longrightarrow P_0^{\vec{f}} \times P_{\overrightarrow{T}}^{(\omega,\overrightarrow{f})}(\{(\omega, \vec{i}) \in \Omega^{\infty} \times I^{\infty} : \omega \in C \wedge dense(\vec{i})[0] = 1\}) > 0.$$

[10]That is, $P_{\overrightarrow{T}}^{(\omega,\overrightarrow{f})}(\{\vec{i} \in I^{\infty} : dense(\vec{i})[0] > 0 \wedge dense(\vec{i})[1] > 0\}) > 0$.

[11]Note that any symmetric test is indecisive in case where $f_0 = f_1$ yielding the tester's coin to eventually distributed half.

[12]That is, $P_0^{\vec{f}}(\{\omega : \overrightarrow{T} \textit{ is decisive on } 1 \textit{ at } (\omega, \vec{f})\}) = 0, \ \forall \vec{f} \in F \times F$.

[13]That is, $P_0^{\vec{f}}(\{\omega : \overrightarrow{T} \textit{ is decisive on } A \textit{ at } (\omega, \vec{f})\} \cap C) > 0, \ \forall \vec{f} \in F \times F$.

5.2 The Test $\overrightarrow{T_D}$

Consider the following finite comparison test $\overrightarrow{T_D}$,

$$T_{D,t+1}(\omega, f_0, f_1)[0] = \frac{1}{1 + D^t_{f_0} f_1(\omega)},\ t > 0 \tag{17}$$

where $D^t_{f_0} f_1(\omega)$ is defined in (4) and $T_{D,t+1}(\omega, f_0, f_1)[0]$ denotes the tester's coin parameter corresponding to expert 0. That is, at each time $t > 0$, the tester uses a coin before announcing which expert he prefers, where the two parameters of the coin are chosen so that their ratio equal the likelihood ratio at time t. Note that each triplet (ω, f_0, f_1) induces a unique measure on I^∞ which we denote by $P^{(\omega,\vec{f})}_{\overrightarrow{T_D}}$, that is $P^{(\omega,\vec{f})}_{\overrightarrow{T_D}} \in \Delta(I^\infty)$.

The next observation state that in fact the tester uses a very 'smart' coin, it asserts that in the long run, the tester has no tendency to make mistakes. We exploit the fact that a realization ω together with f_0, f_1 induce a unique playpath, and recall that by definition $T_{D,t},\ \forall t > 1$ does not depends on any $t-1-history$ of decisions belonging to I^{t-1}, therefore we are able to consider the tester's one step head distribution (17) as the random variable $T_{D,t}:\ \Omega^t \longrightarrow [0,1]$.

5.3 The properties of $\overrightarrow{T_D}$

Claim 6. If $P^{\vec{f}}_1 \ll P^{\vec{f}}_0$ then $(T_{D,t})_{t>0}$ is a sub-martingale with respect to $P^{\vec{f}}_0$.

Proof. Let $X_t = D^t_{f_0} f_1(\omega),\ \forall t > 0$. We first prove that $(X_t)_{t>0}$ is a martingale with respect to $P^{\vec{f}}_0$, we need to show that

$$E[X_t|\ g_{t-1}] = X_{t-1},\ \forall t > 0,\ P^{\vec{f}}_0 - a.s.$$

Let $P^{\vec{f}}_{0,t}, P^{\vec{f}}_{1,t},\ t > 0$, be the restricted measures of $P^{\vec{f}}_0, P^{\vec{f}}_1$ to sigma algebra g_t, respectively, and observe that X_t is $g_t - adaptive$. Thus,

$$\int_C X_t dP^{\vec{f}}_A = \int_C X_t dP^{\vec{f}}_{0,t},\ \forall C \in g_t.$$

In addition, since $P^{\vec{f}}_1 \ll P^{\vec{f}}_0$ by the assumption we have $P^{\vec{f}}_{0,t} \ll P^{\vec{f}}_{1,t},\ \forall t > 0$, which yields that $\int dP^{\vec{f}}_{1,t} = \int (\frac{dP^{\vec{f}}_{1,t}}{dP^{\vec{f}}_{0,t}}) dP^{\vec{f}}_{0,t}$ and therefore,

$$P^{\vec{f}}_1(C) = P^{\vec{f}}_{1,t}(C) = \int_C dP^{\vec{f}}_{1,t} = \int_C (\frac{dP^{\vec{f}}_{1,t}}{dP^{\vec{f}}_{0,t}}) dP^{\vec{f}}_{0,t} = \int_C (D^t_{f_0} f_1) dP^{\vec{f}}_{0,t} = \int_C X_t dP^{\vec{f}}_{0,t},\ \forall C \in g_t,\ t > 0. \tag{18}$$

Now let $C \in g_{t-1},\ t > 0$, and note the filtration $(g_t)_{t>0}$ yields that $C \in g_{t-1} \subset g_t$, which from (18) we obtain

$$\int\limits_C X_{t-1} dP^{\vec{f}}_{0,t-1} = P^{\vec{f}}_{1,t-1}(C) \underset{C \in g_{t-1}}{=} P^{\vec{f}}_1(C) \underset{C \in g_t}{=} P^{\vec{f}}_{1,t}(C) = \int\limits_C X_t dP^{\vec{f}}_{0,t}.$$

Hence $\forall C \in g_t,\ t > 0$ we have $\int\limits_C X_{t-1} dP^{\vec{f}}_{0,t-1} = \int\limits_C X_t dP^{\vec{f}}_{0,t}$, and as a result

$$E[X_t |\ g_{t-1}] = X_{t-1},\ \forall t > 0,\ P^{\vec{f}}_0 - a.s. \tag{19}$$

Now, let $\varphi(t) = \frac{1}{1+t},\ t > 0$, and note that $\varphi(X_t) = T_{D,t}$. Then, since from Lemma 3

$$\underset{t \to \infty}{lim} X_t = \underset{t \to \infty}{lim} D^t_{f_0} f_1(\omega) = D_0 f_1(\omega) < \infty,\ P^{\vec{f}}_0 - a.s,$$

it follows that $E[X_t],\ E[\varphi(X_t)] < \infty$, and hence by Jensen's inequality and (19) we obtain

$$E[\varphi(X_{t+1})|\ g_t] \geq \varphi(E[X_{t+1}|\ g_t]) = \varphi(X_t),\ \forall t > 0,\ P^{\vec{f}}_0 - a.s.$$

and therefore completes the proof. □

Lemma 5. *Let* $P^{\vec{f}}_0, P^{\vec{f}}_1$ *then* $(T_{D,t})_{t>0}$ *is a sub-martingale* $P^{\vec{f}}_0 - a.s.$

Proof. By the Lebesgue decomposition theorem (Billingsley 1995, Section 31) , there exists a unique pair of $,P^{\vec{f}}_{1_r}, P^{\vec{f}}_{1_S}$ such that $,P^{\vec{f}}_1 = P^{\vec{f}}_{1_r} + P^{\vec{f}}_{1_s}$ where $P^{\vec{f}}_{1_r} \ll P^{\vec{f}}_0$ and $P^{\vec{f}}_{1_s} \perp P^{\vec{f}}_0$. Note that if$P^{\vec{f}}_0, P^{\vec{f}}_1$ are mutually singular with respect to each other then $\underset{t \to \infty}{lim} D^t_{f_0} f_1(\omega) = D_{f_0} f_1(\omega) = 0,\ P^{\vec{f}}_0 - a.s.$, which yields that $\underset{t \to \infty}{lim} T_{D,t}(\omega, f_0, f_1)[0] = 1,\ P^{\vec{f}}_0 - .a.s.$. Hence, toghether with Claim 6, $(T_{D,t})_{t>0}$ is a sub-martingale $P^{\vec{f}}_0 - .a.s.$ for both parts of the decomposition and therefore the result follows. □

In fact, a stronger result holds. Not only does the expected probability of making the correct decisions grow it is necessarily the case that the test $\overrightarrow{T_D}$ will essentially settle on the correct expert. This is what we referred to as 'error-free':

Claim 7. $\overrightarrow{T_D}$ is error-free.

Proof. Let $\vec{f} \in F \times F$, and observe that for any fixed $(\omega, \vec{f})$, and $t - 1 > 0$, the tester's one step ahead distribution $T_{D,t}$ does not depend on any past history belonging to I^{t-1} and depends solely on the unique induced playpath (ω, f_0, f_1) up to time t.

As a result, denoting $A_{t,1} = \{\vec{i} \in I^\infty |\ i_t = 1\}$, the infinite product measure $P^{(\omega, \overrightarrow{f})}_{\overrightarrow{T_D}}$ induces a unique sequence of ID-random variables $(1^{(\omega, \overrightarrow{f})}_{A_{t,1}})_{t>0}$, where

$$1_{A_{t,1}}^{(\omega,\overrightarrow{f})} \sim Bern(T_{D,t}(\omega,f_0,f_1)[1] = P_{\overrightarrow{T_D}}^{(\omega,\overrightarrow{f})}(A_{t,1}) = \frac{D_{f_0}^t f_1(\omega)}{1+D_{f_0}^t f_1(\omega)}),\ \forall t>0, \tag{20}$$

and hence

$E[(1_{A_{t,1}}^{(\omega,\overrightarrow{f})})^2] = E[1_{A_{t,1}}^{(\omega,\overrightarrow{f})}] = 1-\frac{1}{1+D_{f_0}^t f_1(\omega)} < \infty,\ \underset{t\to\infty}{lim}\frac{\sum_{i=1}^{t} Var(1_{A_{t,1}}^{(\omega,\overrightarrow{f})})}{t^2} = \frac{\sum_{i=1}^{t}\frac{D_{f_0}^t f_1(\omega)}{1+D_{f_0}^t f_1(\omega)}\cdot\frac{1}{1+D_{f_0}^t f_1(\omega)}}{t^2} < \infty,\ \forall t>0,\ P_0^{\vec{f}}-\omega-a.s.$. Therefore, applying the kolmogorov's strong law, (Jiming 2010, Chapter 6, Theorem 6.7) for the sequence $(1_{A_{t,1}}^{(\omega,\overrightarrow{f})})_{t>0}$ we obtain

$$dense(\vec{i})[1] = \underset{t\to\infty}{lim}\frac{\sum_{i=1}^{t}1_{A_{i,1}}^{(\omega,\overrightarrow{f})}(\vec{i})}{t} = \underset{t\to\infty}{lim}\frac{\sum_{i=1}^{t}E[1_{A_{i,1}}^{(\omega,\overrightarrow{f})}]}{t} = \underset{t\to\infty}{lim}\frac{\sum_{i=1}^{t}1-\frac{1}{1+D_{f_0}^i f_1(\omega)}}{t},\ P_{\overrightarrow{T_D}}^{(\omega,\overrightarrow{f})}-\vec{i}-a.s.. \tag{21}$$

Now note that from Claim 2 T_D is error-free, thus, $P_0^{\vec{f}}(\{\omega|\ \underset{t\to\infty}{lim} D_{f_0}^t f_1(\omega) = D_{f_0}f_1(\omega) = \infty\}) = 0$, which by Kronecker's lemma implies that

$$P_0^{\vec{f}} \times P_{\overrightarrow{T_D}}^{(\omega,\overrightarrow{f})}(\{(\omega,\vec{i}) \in (\Omega^\infty, I^\infty) : dense(\vec{i})[1] = 1\}) = 0.$$

□

Claim 8. $\overrightarrow{T_D}$ is reasonable.

Proof. Let $\vec{f} \in F\times F$, $C-measurale$, and suppose that

$$P_0^{\vec{f}}(C) > 0 \ and\ P_1^{\vec{f}}(C) = 0.$$

Denote $A_{t,0} = \{\overrightarrow{i} \in I^\infty|\ i_t = 1\}$, and observe that for any fixed $(\omega,\vec{f})$, and $t-1>0$, the tester's one step ahead distribution $T_{D,t}$ does not depend on any past history belonging to I^{t-1} and depends solely on the unique induced playpath (ω, f_0, f_1) up to time t. As a result, the infinite product measure $P_{\overrightarrow{T_D}}^{(\omega,\overrightarrow{f})}$ induces a unique sequence $(1_{A_{t,0}}^{(\omega,\overrightarrow{f})})_{t>0}$ of ID-random variables, where

$$1_{A_{t,0}}^{(\omega,\overrightarrow{f})} \sim Bern(T_{D,t}(\omega,f_0,f_1)[0] = P_{\overrightarrow{T_D}}^{(\omega,\overrightarrow{f})}(A_{t,0}) = \frac{1}{1+D_{f_0}^t f_1(\omega)}),\ \forall t>0, \tag{22}$$

and hence

$E[(1_{A_{t,0}}^{(\omega,\overrightarrow{f})})^2] = E[1_{A_{t,0}}^{(\omega,\overrightarrow{f})}] = \frac{1}{1+D_{f_0}^t f_1(\omega)} < \infty,\ \underset{t\to\infty}{lim}\frac{\sum_{i=1}^{t} Var(1_{A_{t,0}}^{(\omega,\overrightarrow{f})})}{t^2} = \frac{\sum_{i=1}^{t}\frac{D_{f_0}^t f_1(\omega)}{1+D_{f_0}^t f_1(\omega)}\cdot\frac{1}{1+D_{f_0}^t f_1(\omega)}}{t^2} < \infty,\ \forall t>0,\ P_0^{\vec{f}}-\omega-a.s..$

Therefore, applying the kolmogorov's strong law (Jiming 2010, Chapter 6, Theorem 6.7), for the sequence $(1^{(\omega,\overrightarrow{f})}_{A_{t,0}})_{t>0}$ we obtain

$$dense(\vec{i})[0] =$$
$$\lim_{t\to\infty}\frac{\sum_{i=1}^{t}1^{(\omega,\overrightarrow{f})}_{A_{i,0}}(\vec{i})}{t} = \lim_{t\to\infty}\frac{\sum_{i=1}^{t}E[1^{(\omega,\overrightarrow{f})}_{A_{i,0}}]}{t} = \lim_{t\to\infty}\frac{\sum_{i=1}^{t}\frac{1}{1+D^{i}_{f_0}f_1(\omega)}}{t},\ P^{(\omega,\overrightarrow{f})}_{\overrightarrow{T_D}} - \vec{i} - a.s.,$$

Now let $\epsilon > 0$ and note that $\forall\tilde{\omega}\in C$ with $\lim_{t\to\infty}\frac{1}{1+D^{i}_{f_0}f_1(\tilde{\omega})} = 1$, $\exists N > 0$ such that

$$\frac{\sum_{i=1}^{N}\frac{1}{1+D^{i}_{f_0}f_1(\omega)} - N(1-\epsilon)}{t} + 1 - \epsilon < \frac{\sum_{i=1}^{t}\frac{1}{1+D^{i}_{f_0}f_1(\omega)}}{t} \leq 1,\ \forall t > N,$$

where, taking $t\to\infty$ yields

$$dense(\vec{i})[0] > 1-\epsilon,\ P^{(\tilde{\omega},\overrightarrow{f})}_{\overrightarrow{T_D}} - \vec{i} - a.s.. \tag{23}$$

In addition, recall that from Claim 1 T_D is reasonable, thus, $P^{\vec{f}}_0(\{\omega\in\Omega^\infty : \omega\in C \ \wedge\ \lim_{t\to\infty}D^{t}_{f_0}f_1(\omega) = 0\}) = P^{\vec{f}}_0(C) > 0$, and consequently, from (23) we obtain

$$P^{\vec{f}}_0 \times P^{(\omega,\overrightarrow{f})}_{\overrightarrow{T_D}}(\{(\omega,\vec{i})\in\Omega^\infty\times I^\infty : \omega\in C \text{ and } 1-\epsilon < dense(\vec{i})[0] \leq 1\}) > 0,$$

where taking $\epsilon\to 0$ the result follows. □

5.4 The Uniqueness of $\overrightarrow{T_D}$

For a given $\overrightarrow{T}$, $\overrightarrow{f}$, $i\in I$, we denote by $A^{\overrightarrow{f}}_{\overrightarrow{T},\frac{1}{2}}, A^{\overrightarrow{f}}_{\overrightarrow{T},i}$ to be the sets of realizations for which $\overrightarrow{T}$ is either indecisive or decisive on i at $(\omega,\overrightarrow{f})$, respectively.[14]

Definition 14. Let $\vec{f} \triangleq (f_0, f_1) \in F\times F$ $\omega\in\Omega^\infty$. We say that $\overrightarrow{T}$ and $\overrightarrow{T}$ decide differently at $(\omega,\overrightarrow{f})$ if $\exists i, j(\neq i)\in\{0,1,\frac{1}{2}\}$ such that $\omega\in A^{\overrightarrow{f}}_{\overrightarrow{T},i}\cap A^{\overrightarrow{f}}_{\overrightarrow{T},j}$.[15]

[14] For example, $A^{\vec{f}}_{\vec{T},1} = \{\omega : \overrightarrow{T} \text{ is decisive on } 1 \text{ at } (\omega,\vec{f})\} =$
$\{\omega |\ P^{(\omega,\overrightarrow{f})}_{\vec{T}}(\{\vec{i}\in I^\infty : dense(\vec{i})[1] = 1\}) = 1\}$.

[15] For example: $\omega\in A^{\vec{f}}_{\vec{T},0}\cap A^{\vec{f}}_{\vec{T},1} \iff$
$P^{(\omega,\overrightarrow{f})}_{\vec{T}}(\{\vec{i}\in I^\infty : dense(\vec{i})[0] = 1\}) = 1 \text{ and } P^{(\omega,\overrightarrow{f})}_{\vec{T}}(\{\vec{i}\in I^\infty : dense(\vec{i})[1] = 1\}) = 1$.

Definition 15. Let $\vec{f} \triangleq (f_0, f_1) \in F \times F$. We say that $\overrightarrow{T} \sim_{\vec{f}} \overrightarrow{\hat{T}}$ if $\forall i \in \{0,1\}$

$$P_i^{\vec{f}}(\{\omega : \overrightarrow{T} \textit{ and } \overrightarrow{\hat{T}} \textit{ decide differently at } (\omega, \overrightarrow{f})\}) = 0.$$

We say that $\overrightarrow{T} \sim \overrightarrow{\hat{T}}$ if and only if $\overrightarrow{T} \sim_{\vec{f}} \overrightarrow{\hat{T}}, \forall \vec{f} \in F \times F$.

Claim 9. $\sim$ is an equivalence relation on $\overrightarrow{\top} = \{\overrightarrow{T} : \ \overrightarrow{T} - \textit{random comparison test}\}$.

Proof. The proof is analogously to proof of Claim 3 with minor modifications concerning density arithmetic consideration and hence omitted. □

The next theorem asserts that, up to an equivalence class representative, there exists a unique reasonable and error-free test. That is, any error-free test $\overrightarrow{T} \nsim \overrightarrow{T}_D$ which is reasonable, admits an error.

Theorem 4. *Let $\overrightarrow{T}$ be a reasonable test. If $\overrightarrow{T} \nsim \overrightarrow{T_D}$ then $\overrightarrow{T}$ is not error-free.*

Proof. Assume by contradiction that $\overrightarrow{T}$ is error-free. Let $\vec{f} \triangleq (f_0, f_1)$ be such that $\overrightarrow{T} \nsim_{\vec{f}} \overrightarrow{T_D}$ at $\vec{f}$, then $\exists i, j (\neq i) \in \{0, 1, \frac{1}{2}\}$ such that (w.l.o.g)

$$P_0^{\vec{f}}(A^{\vec{f}}_{\overrightarrow{T},i} \cap A^{\vec{f}}_{\overrightarrow{T_D},j}) > 0.$$

In addition, by Claim 7 $\overrightarrow{T_D}$ is error-free, therefore

$$P_0^{\vec{f}}(A^{\vec{f}}_{\overrightarrow{T},1}) = P_0^{\vec{f}}(A^{\vec{f}}_{\overrightarrow{T_D},1}) = 0,$$

and consequently,

$$P_0^{\vec{f}}(\hat{C}_1 \triangleq (A^{\vec{f}}_{\overrightarrow{T},0} \cap A^{\vec{f}}_{\overrightarrow{T_D},\frac{1}{2}})) > 0 \ \ \textit{or} \ \ P_0^{\vec{f}}(\hat{C}_2 \triangleq (A^{\vec{f}}_{\overrightarrow{T},\frac{1}{2}} \cap A^{\vec{f}}_{\overrightarrow{T_D},0})) > 0.$$

Case 1: $P_0^{\vec{f}}(\hat{C}_1) > 0$. By Claim 8, $\overrightarrow{T_D}$ is reasonable; thus

$$P_1^{\vec{f}}(\hat{C}_1) = 0 \Longrightarrow P_0^{\vec{f}}(\{\omega | \ \overrightarrow{T_D} \textit{ is decisive on } 0 \textit{ at } (\omega, \vec{f})\} \cap \hat{C}_1) = P_0^{\vec{f}}(A^{\vec{f}}_{\overrightarrow{T_D},0} \cap \hat{C}_1) > 0,$$

which leads to a contradiction since $\forall \omega \in A^{\vec{f}}_{\overrightarrow{T_D},0} \cap \hat{C}_1$

$$P^{(\omega,\overrightarrow{f})}_{\overrightarrow{T_D}}(\{\vec{i} \in I^\infty : \ dense(\vec{i})[0] = 1\}) = 1 \Longrightarrow P^{(\omega,\overrightarrow{f})}_{\overrightarrow{T_D}}(\{\vec{i} \in I^\infty : \ dense(\vec{i})[1] > 0\}) = 0,$$

which implies that $A^{\vec{f}}_{\overrightarrow{T_D},0}$, $A^{\vec{f}}_{\overrightarrow{T_D},\frac{1}{2}}$ are disjoint. Thus

$$P_1^{\vec{f}}(A^{\vec{f}}_{\overrightarrow{T},0}) > 0,$$

which contradicts the assumption that $\overrightarrow{T}$ is error-free.

Case 2: $P_0^{\vec{f}}(\hat{C}_2) > 0$. By the assumption $\overrightarrow{T}$ is a reasonable test where, by Claim 2, $\overrightarrow{T_D}$ is error-free; therefore the contradiction

$$P_1^{\vec{f}}(A^{\vec{f}}_{\overrightarrow{T_D},0}) > 0$$

follows analogously from Case 1. $\square$

6 Discussion

Consider a scenario where we require some expert advice on the evolution of some unknown system (e.g., the economy or a financial market). We typically entertain a few experts and would like to make sure we take the advice from the better informed one. This suggests that expert testing should be framed in comparative terms. Instead of asking whether or not a single forecaster is indeed an expert or a charlatan, as is done in the lion's share of the literature on expert testing, we advocate a different approach in which we compare a few experts; the test is designed to spot the better informed one.

We provide some natural properties for comparison tests and show that these properties uniquely characterize test that are based on the expert's likelihood ratio. We do so for infinite tests - namely tests which verdict is cast at the end of all times - and also for finite tests.

References

[1] Al-Najjar, Nabil I. and Weinstein, J. (2008). Comparative testing of experts. Econometrica, 76(3), 541–559.

[2] Al-Najjar, N.I., Sandroni, A., Smorodinsky, R. and Weinstein, J. (2010). Testing theories with learnable and predictive representations. Journal of Economic Theory, 145(6), 2203–2217.

[3] Billingsley, P. (1995). Probability and Measure, 3rd edition, John Wiley, New York.

[4] Dawid, P. (1982). The well-calibrated bayesian. Journal of the American Statistical Association, 77, 605–613.

[5] Dekel, E. and Feinberg, Y. (2006). Non-bayesian testing of a stochastic prediction. Review of Economic Studies, 73, 893–906.

[6] Feinberg, Y. and Stewart, C. (2008). Testing multiple forecasters. Econometrica, 76, 561–582.

[7] Fortnow, L. and Vohra, R.V. (2009). The complexity of forecast testing. Econometrica, 77, 93–105.

[8] Foster, D.P. and Vohra, R.V. (1998). Asymptotic calibration. Biometrika, 85, 379–390.

[9] Fudenberg, D. and Levine, D. (1999). An easier way to calibrate. Games and Economic Behavior, 29, 131–137.

[10] Jiming, J. (2010). Large Sample Techniques for Statistics. Springer-Verlag New York.

[11] Kalai, E., Lehrer E. and Smorodinsky, R. (1999). Calibrated forecasting and merging. Games and Economic Behavior, 29, 151–159.

[12] Lehrer, E. (2001). Any inspection is manipulable. Econometrica, 69, 1333–1347.

[13] Olszewski, W. and Sandroni, A. (2008). Manipulability of future-independent tests. Econometrica, 76, 1437–1466.

[14] Pomatto, L. (2016). Testable forecasts. Caltech (unpublished results).

[15] Sandroni, A. (2003). The reproducible properties of correct forecasts. International Journal of Game Theory, 32, 151–159.

[16] Sandroni, A., Smorodinsky, R. and Vohra, R. (2003). Calibration with many checking rules. Mathematics of Operations Research, 28, 141–153.

[17] Shmaya, E. (2008). Many inspections are manipulable. Theoretical Economics, 3, 367–382.

[18] Williams, M. (1991). Probability with Martingales. Cambridge University Press.