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

PIER Working Paper 09-035

"Testing Predictive Ability and Power Robustification"

by

Kyungchul Song

http://ssrn.com/abstract=1487598

Testing Predictive Ability and Power Robustification

Kyungchul Song[1]

Department of Economics, University of Pennsylvania

September 8, 2009

Abstract

One of the approaches to compare forecasts is to test whether the loss from a benchmark prediction is smaller than the others. The test can be embedded into the general problem of testing functional inequalities using a one-sided Kolmogorov-Smirnov functional. This paper shows that such a test generally suffers from unstable power properties, meaning that the asymptotic power against certain local alternatives can be much smaller than the size. This paper proposes a general method to robustify the power properties. This method can also be applied to testing inequalities such as stochastic dominance and moment inequalities. Simulation studies demonstrate that tests based on this paper's approach perform quite well relative to the existing methods.

Key words and Phrases: Inequality Restrictions, Testing Predictive Ability, One-sided Nonparametric Tests, Power Robustification

JEL Classifications: C12, C14, C52, C53.

[1] I thank Werner Ploberger and Frank Schorfheide for valuable comments and advice. Part of this paper began with a draft titled, "Testing Distributional Inequalities and Asymptotic Bias." All errors are mine. Address correspondence to: Kyungchul Song, Department of Economics, University of Pennsylvania, 528 McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297.

1 Introduction

Assessing and comparing a multiple number of forecasts is important in practice, in particular, in macroeconomics and finance. Since the seminal paper by Diebold and Mariano (1995), there has been a rapidly growing interest in comparing different forecasting models in terms of their relative out-of-sample forecast performance. West (1996) offered a formal analysis of inference about the out-of-sample predictions, and White (2000) developed a framework to compare multiple forecasting models. Hansen (2005) offered a general way to improve the power of the test of predictive ability. Giacomini and White (2006) proposed out-of-sample predictive ability tests that can be applied to conditional evaluation objectives. In the meanwhile, evaluation of density forecasts has also drawn interest in the literature. See early contributions by Diebold, Gunther, and Tay (1998), Christoffersen (1998), and Diebold, Hahn, and Tay (1999). Density forecasts have been further analyzed by Amisano and Giacomini (2007), and Bao, Lee, and Saltoğlu (2007), among others.

Comparing relative predictive ability can be embedded into the general problem of testing functional inequalities of the following type:

$$\begin{aligned} H_0 &: e(m) \leq 0, \text{ for all } m \in \mathbf{M} \text{ and} \\ H_1 &: e(m) > 0, \text{ for some } m \in \mathbf{M}, \end{aligned} \tag{1}$$

where $e(m)$ is an unknown but estimable real function on a set $\mathbf{M} \subset \mathbf{R}^d$. For example, we can take

$$e(m) = \Lambda(0) - \Lambda(m)$$

where $\Lambda(m)$ denotes a risk associated with prediction based on the m-th candidate forecasting model and $\Lambda(0)$ a risk due to prediction based on a benchmark model. Then the alternative hypothesis states that there exists a model that performs strictly better than the benchmark model. More specifically, suppose that the object of forecast is a τ-ahead quantity Y_τ and there are M

number of candidate forecasts $\varphi_m(\mathcal{F}_m)$ where $\mathcal{F}_m$ is the information used by the forecast φ_m. The framework includes the situation with parameter uncertainty in forecasting models because the uncertainty is random to the extent $\mathcal{F}_m$ is. Then, we may be interested in evaluating the forecasts by comparing

$$\Lambda(m) = \mathbf{E}[\{Y_\tau - \varphi_m(\mathcal{F}_m)\}^2].$$

The expectation above is with respect to the joint distribution of variables constituting the information $\mathcal{F}_m$ and Y_τ. Alternatively, $\Lambda(m)$ can be taken to be a risk associated with a functional of a density forecast. For instance, we may take

$$\Lambda(m) = \mathbf{E}\left[\int \log(f_m(y|\mathcal{F}_m)/f(y))f(y)dy\right],$$

where $f_m(\cdot|\mathcal{F}_m)$ is the m-th density forecast using the information $\mathcal{F}_m$ and f is the true density of Y_τ. The expectation above is with respect to the distribution of variables constituting $\mathcal{F}_m$. The quantity $\Lambda(m)$ is the expected Kullback-Leibler divergence based on the m-th density forecast. Let $\Lambda(0)$ be the expected Kullback-Leibler divergence for the benchmark model. Then, testing (1) using $e(m) = \Lambda(0) - \Lambda(m)$ is tantamount to testing whether the benchmark forecast is optimal among the candidate forecasts. (See Bao, Lee, and Saltoğlu (2007) and Amisano and Giacomini (2007), and)

Although most applications assume a finite $\mathbf{M}$, the paper's proposal is not confined to this assumption. More importantly, the testing problem of (1) has many other examples beyond that of comparing forecasting models. While these examples are not pursued in this paper, it is still worth mentioning them. The first example is testing stochastic dominance of one variable by the other. (See, e.g., Davidson and Duclos (2000), Barret and Donald (2003), and Linton, Massoumi, and Whang (2005).) Testing conditional or unconditional positive dependence also falls into the testing framework of (1). For example, see Cawley and Phillipson (1999) and Chiappori and Salanié (2000) in the context of contract theory and Denuit and Scaillet (2004) in financial econometrics. Third, testing moment inequalities also belongs to this framework (e.g. Chernozhukov, Hong, and Tamer (2007), Rosen (2006), Moon and Schorfheide (2007), Andrews and Guggenberger (2006). Such tests can be used to construct a confidence set for an unknown, partially identified true parameter.

The usual method of testing (1) involves replacing e by an estimator $\hat{e}$ and taking an appropriate functional of it to form a test statistic. This paper specifically focuses on the one-sided Kolmogorov-Smirnov (KS) test statistic:

$$T^K = s_n \text{sup}_{m \in \mathbf{M}} \hat{e}(m), \tag{2}$$

where $s_n \to \infty$ is a normalizing sequence to prevent degeneracy in the limit under the null hypothesis. The testing framework is not restricted to an i.i.d. data set-up, as this paper's results concern only the limiting Gaussian experiments.

First, this paper shows that there exist a class of Pitman local alternatives against which the asymptotic power of the test in (2) is below the level of the test. When $\mathbf{M}$ is infinite, the asymptotic power is arbitrarily close to zero against certain local alternatives. When the null hypothesis is of the form: $e(m) = 0$ for all $m \in \mathbf{M}$, the poor power property of nonparametric tests against certain alternatives is well-known in the literature. For example, Janssen (2000) showed that in that situation, a nonparametric test has a nearly trivial asymptotic power, i.e. an asymptotic power close to the level α, except for a finite dimensional space of local alternatives. In contrast, this paper finds that when it comes to testing (1), the asymptotic power of the one-sided KS test is even more unstable, as the power becomes close to zero under certain local alternatives.

Given the extremely unstable power property, one may ask whether there is a way to alleviate this problem. This question leads to the main contribution of this paper. First, this paper identifies a complementary test that shows good power against alternatives under which the one-sided KS test has poor power. Then, this paper proposes a test that couples this complementary test with the one-sided KS test.

This paper is not the first to point out the poor power property of the one-sided KS test for the inequality models. Hansen (2005) studied testing predictive ability among forecasting models as in

White (2000) and showed the poor power phenomenon of the one-sided KS test through excellent illustrations. Hansen (2005) also suggested a way to improve the asymptotic power property of the tests. His idea of improving the power is general and, in fact, related to some later literatures on testing moment inequalities such as Andrews and Soares (2007) (where $\mathbf{M}$ is finite) and Linton, Song and Whang (2008) (where $\mathbf{M}$ is infinite). On the other hand, this paper's suggestion is different from his approach. While the power improvement by his approach is by transforming a test toward an asymptotic similar test, this paper suggests robustifying the power of the test by coupling it with a test that has complementary asymptotic power properties. Therefore, the suggestion of Hansen (2005) is not competitive with this paper's approach. When a test of functional inequalities is not asymptotically similar, one may first find an asymptotically similar test by applying the method of Hansen (2005) or Linton, Song, and Whang (2008), and then, robustify the power of the test by employing the hybrid test that this paper suggests. This point is exemplified in the simulation studies.

The paper is organized as follows. The next section establishes asymptotic biasedness for one-sided KS tests of functional inequalities. This section also motivates and introduces a method of hybrid tests. Section 3 applies this method to testing predictive abilities, and presents some results from simulation results. Section 4 concludes. Some technical proofs are in the Appendix.

2 Asymptotic Bias of One-Sided KS Tests

In this section, we present the result of unstable power property of the one-sided KS test. To define the scope of the result, we need to introduce some notations. For a subset $\mathbf{M} \subset \mathbf{R}^d$, let $l_\infty(\mathbf{M})$ be the space of bounded functions on $\mathbf{M}$, equipped with the sup norm $||\cdot||_\infty : ||f||_\infty = \sup_{m\in\mathbf{M}}|f(m)|$ for all bounded functions f on $\mathbf{M}$. For any functions $f, g \in l_\infty(\mathbf{M})$, we write $f \geq g$ if $f(m) \geq g(m)$ for all $m \in \mathbf{M}$. We are interested in an unknown function $e \in l_\infty(\mathbf{M})$, especially whether $e \leq 0$ or not. Let Γ^K be a one-sided Kolmogorov-Smirnov functional on $l_\infty(\mathbf{M})$: for each $\xi \in l_\infty(\mathbf{M})$,

$$\Gamma^K \xi \equiv \sup_{m \in \mathbf{M}} \xi(m).$$

We turn to the hypothesis testing of (1). For given $e_0 \in l_\infty(\mathbf{M})$ such that $e_0 \leq 0$, we take $\mathbf{M}_0$ to be the *zero-set* $\mathbf{M}_0 = \{m \in \mathbf{M} : e_0(m) = 0\}$. As long as the nondegenerate limiting distribution of the one-side KS test is concerned, it suffices to consider the collection of probabilities under the the null hypothesis (1) such that $e = e_0$ for some $e_0 \leq 0$ that has nonempty zero-set $\mathbf{M}_0$. We let Γ_0^K be Γ^K with the domain restricted to $l_\infty(\mathbf{M}_0)$, i.e., for any $\xi \in l_\infty(\mathbf{M}_0)$,

$$\Gamma_0^K \xi \equiv \sup_{m \in \mathbf{M}_0} \xi(m).$$

We introduce Pitman local alternatives. Let $A_0 = \{\delta \in l_\infty(\mathbf{M}) : \Gamma_0^K \delta > 0\}$ and $A_0^K \subset A_0$ be such that for all $\delta \in A_0$ there exists $\delta_1 \in A_0^K$ satisfying $\delta_1 \leq \delta$. For given $\delta \in A_0^K$, Pitman local alternatives in the direction δ are defined as a sequence of probabilities under which

$$e(m) = e_0(m) + \delta(m)/s_n \tag{3}$$

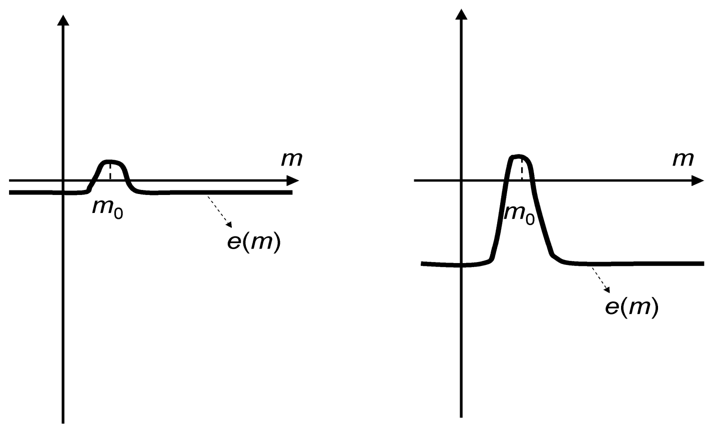
for some $e_0 \in l_\infty(\mathbf{M})$ such that $e_0 \leq 0$ and for some normalizing sequence $s_n \to \infty$. Let $\hat{e}(m)$ be an estimator of $e(m)$ such that as $n \to \infty$,

$$\begin{aligned} s_n \hat{e} &\Longrightarrow \nu \text{ in } l_\infty(\mathbf{M}_0), \text{ under } H_0, \\ s_n \hat{e} &= s_n\{\hat{e} - e\} + s_n e \Longrightarrow \nu + \delta \text{ in } l_\infty(\mathbf{M}_0), \text{ under (3)}, \end{aligned} \tag{4}$$

where ν is a Gaussian process in $l_\infty(\mathbf{M}_0)$. The weak convergence $\Longrightarrow$ can be established using the (functional) central limit theorem or the standard empirical process theory in many examples.

Hansen (2005) offers intuitive illustrations that show asymptotic bias of the one-sided KS test

Figure 1: Examples of $e(m)$ Associated with Low Power for the One-sided KS Test



for the case where $\mathbf{M} = \{1, 2\}$. However, it is not intuitively obvious whether such properties will extend to arbitrary $\mathbf{M}$. To illustrate this point, we consider two examples of alternatives for $e(m)$ shown in Figure 1. While the power is affected by the height of $e(m_0)$ at m_0, it also hinges on the values that $e(m)$ takes on the remaining area. Suppose that the height of $e(m)$ at $m = m_0$ is the same for both alternatives, and that in most of the remaining area of $\mathbf{M}_0$ outside a neighborhood of m_0, $e(m)$ takes negative values as shown in Figure 1. The extent of this negativity affects the asymptotic power in two opposite ways. A further negativity will reduce the power because the area of m's such that $\nu(m) + \delta(m) \leq \nu(m)$ tends to be larger than the area of m's such that $\nu(m) + \delta(m) > \nu(m)$. But this effect is offset by the fact that the probability of the maximizer of ν and the maximizer of $\nu + \delta$ being close to each other becomes smaller as $e(m)$ takes more negative values in the remaining area. Therefore, the total effect of negativity of $e(m)$ on the remaining area is ambiguous.

The proposition below establishes that the test T^K defined in (2) suffers from asymptotic bias against certain local alternatives in general. Therefore the observation by Hansen (2005) extends to a substantial degree of generality.

Proposition 1 : *For a subset* $\mathbf{M}_0 \subset \mathbf{M}$, *let* ν *be a separable centered Gaussian process in* $l_\infty(\mathbf{M}_0)$ *with uniformly continuous sample paths and* $Var(\nu(m) - \nu(s)) \neq 0$ *for* $m \neq s$ *in* $\mathbf{M}_0$.

(i) *For (potentially stochastic)* $c_{n,\alpha} \geq 0$ *with* $\lim_{n\to\infty} P\left\{T^K > c_{n,\alpha}\right\} \leq \alpha$ *under* H_0*, there exists* $\delta \subset A_0^K$ *such that*

$$\lim_{n\to\infty} P_\delta\left\{T^K > c_{n,\alpha}\right\} < \alpha,$$

where P_δ *denotes the sequence of probabilities under* (3).

(ii) *Furthermore, assume that there exists* $m_0 \in \mathbf{M}_0$ *such that* $P\{\nu(m_0) = 0\} = 1$. *Then, for (potentially stochastic)* $c_{n,\alpha} > 0$ *with* $\lim_{n\to\infty} P\left\{T^K > c_{n,\alpha}\right\} \leq \alpha$ *under* H_0*,*

$$\inf_{\delta\in A_0^K} \lim_{n\to\infty} P_\delta\left\{T^K > c_{n,\alpha}\right\} = 0.$$

The condition $Var(\nu(m)-\nu(s)) \neq 0,\ m \neq s$, is satisfied by almost all Gaussian processes whose sample paths are not constant functions. The separability condition for the Gaussian process ν is a technical measurability condition. (See a footnote in van der Vaart and Wellner (1996), p.98.) Proposition 1(i) tells us that the one-sided KS test is asymptotically biased against certain local alternatives. As shown in the proof of Proposition 1(i), the collection of shifts δ under which the test is asymptotically biased is not contained in a finite-dimensional subspace of $l_\infty(\mathbf{M})$ when $\mathbf{M}$ is infinite. The second result (ii) shows that the minimum asymptotic power is equal to zero when $\nu(m_0)$ is almost everywhere equal to zero at some point m_0. The latter condition is satisfied by Brownian motions and Brownian bridges, and many other Gaussian processes whose sample paths begin at zero. However, the extreme result of (ii) does not apply when ν is a nondegenerate, stationary Gaussian process or when $\mathbf{M}_0$ is a finite set.

The power property of a one-side KS test can be improved by transforming the test into one that is asymptotically similar. A test of (1) is called *asymptotically similar*, if the asymptotic rejection probability is identical to the level α whenever $\sup_{m\in\mathbf{M}} e(m) = 0$. In particular, when $\mathbf{M}_0$ is compact, a test is asymptotically similar if $\lim_{n\to\infty} P\{T^K > c_\alpha\} = \alpha$ for all the probabilities P such that $\mathbf{M}_0$ is nonempty. It is well known that asymptotically unbiased tests are asymptotically similar, or equivalently, asymptotically nonsimilar tests are asymptotically biased. (e.g. Strasser (1985) p.429. See also Hansen (2003).) In fact, one can improve the power of an asymptotically nonsimilar test by transforming it into an asymptotically similar one. (e.g. Hansen (2005), Linton, Massoumi and Whang (2005), and Linton, Song, and Whang (2008)). However, the unstable power phenomenon of the one-sided KS test in Proposition 1 still arises regardless of whether a test is asymptotically similar or not. This is because the result allows $\mathbf{M}_0$ to be a proper subset of $\mathbf{M}$.

2.1 Power Robustification via Coupling

2.1.1 A Complementary Test

The result of the previous section showed that the asymptotic power of a one-sided KS test can be very poor against certain local alternatives. We construct a hybrid test that tends to have a *robust* power property. The construction is in two steps. First, we identify a complementary test that has a better power property against local alternatives under which the one-sided KS test has a very poor power. Second, we couple the one-sided KS test with the complementary test to construct a new hybrid test.

As for a complementary test, we consider the following hypothesis testing problem:

$$\begin{aligned} H_0^S &: e(m) \leq 0 \text{ for all } m \in \mathbf{M}, \text{ or } e(m) \geq 0 \text{ for all } m \in \mathbf{M}, \\ H_1^S &: e(m) > 0 \text{ for some } m \in \mathbf{M} \text{ and } e(m) < 0 \text{ for some } m \in \mathbf{M}. \end{aligned} \tag{5}$$

We define a symmetric functional Γ^S on $l_\infty(\mathbf{M})$ as follows:

$$\Gamma^S(\xi) = \min\{\Gamma^K(\xi), \Gamma^K(-\xi)\}.$$

The use of this type of functional was proposed by Linton, Massoumi, and Whang (2005) in the context of testing stochastic dominance. Given $\hat{e}$, a test statistic for testing (5) can be constructed as

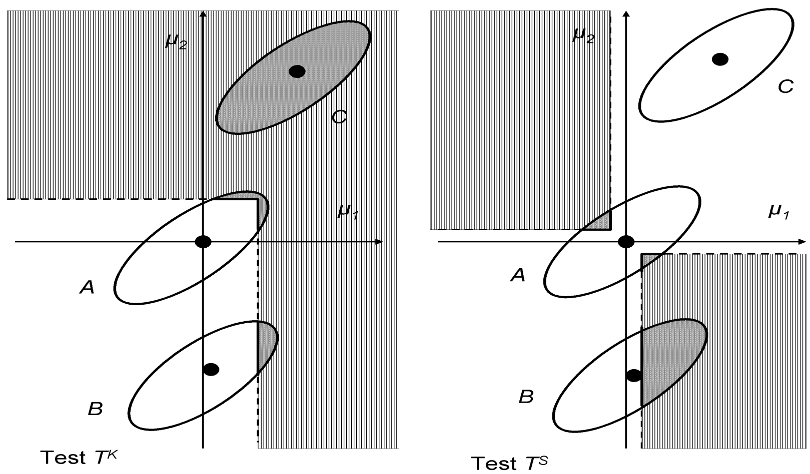
$$T^S = \Gamma^S(s_n \hat{e}). \tag{6}$$

For testing the null hypothesis in (1), the test T^S is *complementary* to T^K : T^S tends to have a greater power than T^K against local alternatives that give T^K a very poor power. We illustrate this point by considering the following example. Let $\mathbf{M} = \{1, 2\}$. Given observations Z_1 and Z_2 which are positively correlated and jointly normal with a mean vector $\mu = (\mu_1, \mu_2)$, we are interested in testing

$$\begin{aligned} H_0 &: \mu_1 \leq 0 \text{ and } \mu_2 \leq 0 \\ H_1 &: \mu_1 > 0 \text{ or } \mu_2 > 0. \end{aligned}$$

Consider $T^K = \max(Z_1, Z_2)$ and $T^S = \min\{\max(Z_1, Z_2), \max(-Z_1, -Z_2)\}$. Complementarity between the tests T^K and T^S are illustrated in Figure 2 in a form borrowed from Hansen (2005). The illustration is based on a least favorable configuration (LFC) in which we read critical values from the distribution with $\mu_1 = 0$ and $\mu_2 = 0$. Hence $\mathbf{M}_0 = \{1, 2\}$. The ellipses in Figure 2 indicate representative contours of the joint density of Z_1 and Z_2, each corresponding to different distributions denoted by A, B, and C. While A represents the null hypothesis under LFC, B and C represent alternatives. Under the alternative B, the rejection probability of the test T^K can be lower than that under A, implying the biasedness of the test. (This is illustrated by the dark area of ellipsis B in the left panel which is smaller than the dark area of ellipsis A in the same panel.) However, the rejection probability of the test T^S against this alternative B is better as indicated by a larger dark area in the corresponding ellipsis on the right panel. (This contrast may be less stark when Z_1 and Z_2 are negatively correlated.) Hence against B, test T^S has a better power than test T^K. This order of performance is reversed in the case of an alternative C where the test

Figure 2: Complementarity between T^K and T^S



T^S has a power close to zero while the test T^K has a power close to 1.

2.1.2 Coupling

The complementary test cannot replace the one-sided KS test because it will suffer from a similar kind of poor power properties as in Proposition 1 only against a different kind of local alternatives. We construct a hybrid test by coupling the complementary test and the one-sided KS test so that the resulting test may have balanced power properties. For each $\alpha \in [0, 1]$ and $\gamma \in [0, \alpha]$, we introduce the following hybrid test of the null hypothesis in (1).

$$\begin{aligned} \text{Reject } H_0 \text{ if } T^S &> c^S_{n,\alpha}(\gamma) \\ &\text{or} \\ \text{if } T^S &\leq c^S_{n,\alpha}(\gamma) \text{ and } T^K > c^K_{n,\alpha}(\gamma), \end{aligned} \tag{7}$$

where $c^S_{n,\alpha}(\gamma)$ and $c^K_{n,\alpha}(\gamma)$ are threshold values such that

$$\begin{aligned}\lim_{n\to\infty} P\{T^S &> c^S_{n,\alpha}(\gamma)\} = \gamma \text{ and} \\ \lim_{n\to\infty} P\{T^S &\leq c^S_{n,\alpha}(\gamma) \text{ and } T^K > c^K_{n,\alpha}(\gamma)\} = \alpha - \gamma.\end{aligned}$$

The hybrid test runs along a locus between the two tests T^K and T^S as we move γ between 0 and α : when γ is set to be close to α, the hybrid test becomes close to T^S, and when γ is set to be close to 0, it becomes close to T^K. The power-reducing effect of the negativity of $e(m)$ on most values of m is counteracted by the positivity of $-e(m)$ on most values of m. (Figure 3.) By coupling with T^S, the hybrid test shares this counteracting effect to an extent depending on γ, with the effect of power reduction attenuated relative to T^K. In practical situations, this paper simply proposes using $\gamma = \alpha/2$. Simulation studies in this paper suggest that this choice of γ yields a hybrid test that performs reasonably well.

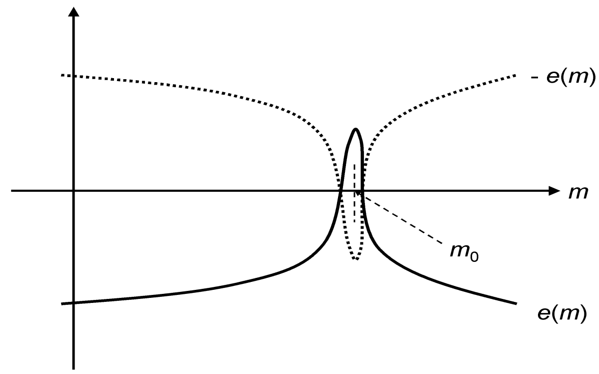
In general, we cannot evaluate the critical values $c^S_{n,\alpha}(\gamma)$ and $c^K_{n,\alpha}(\gamma)$ from the limiting distributions of test statistics T^S and T^K. This feature is not something created anew by the coupling method; in general, tests T^K and T^S are asymptotically non-pivotal except for special cases. We can compute the approximate critical values $c^{S*}_{n,\alpha}(\gamma)$ and $c^{K*}_{n,\alpha}(\gamma)$ using bootstrap or subsampling. For example, we simulate the bootstrap distribution P^* of (T^S, T^K) by generating $(T^{S*}_b, T^{K*}_b)^B_{b=1}$. Using the empirical distribution of $\{T^{S*}_b\}^B_{b=1}$, we compute $c^{S*}_{n,\alpha}(\gamma)$ such that

$$P^*\{T^{S*}_b > c^{S*}_{n,\alpha}(\gamma)\} = \gamma$$

where P^* denotes the bootstrap distribution of T^{S*}_b. Using this $c^{S*}_{n,\alpha}(\gamma)$, we find $c^{K*}_{n,\alpha}(\gamma)$ such that

$$P^*\{T^{S*}_b \leq c^{S*}_{n,\alpha}(\gamma) \text{ and } T^{K*}_b > c^{K*}_{n,\alpha}(\gamma)\} = \alpha - \gamma.$$

Figure 3: An Example of $e(m)$ and Its Reflection



Then, the bootstrap-based test can be described as

$$\begin{aligned} \text{Reject } H_0 \text{ if } T^S &> c^{S*}_{n,\alpha}(\gamma) \\ &\text{or} \\ \text{if } T^S &\leq c^{S*}_{n,\alpha}(\gamma) \text{ and } T^K > c^{K*}_{n,\alpha}(\gamma). \end{aligned}$$

The method of coupling hardly entails additional computational cost. The computational cost in most cases lies in having to compute $\hat{e}^*(m)$ using the bootstrap samples, which is a step common in the other bootstrap-based tests. Once $\hat{e}^*(m)$ is computed, finding T_b^{K*} and T_b^{S*} and obtaining bootstrap critical values are straightforward.

3 Testing Predictive Ability

3.1 Background

In this section, we apply the coupling approach to testing predictive ability. Suppose that we are given M models and want to know whether any of these models performs better than a certain benchmark model in terms of forecasting. Suppose that $e(m),\ m = 1, \cdots, M$, denotes the relative predictive performance of the m-th model to the benchmark model, so that $e(m) \leq 0$ means that the benchmark model performs better than the m-th model. For example, let $\Lambda(m)$ be a risk associated with using the m-th forecast as illustrated in the introduction, and define

$$e(m) = \Lambda(0) - \Lambda(m).$$

The null hypothesis and the alternative hypothesis are written as

$$\begin{aligned} H_0 &: \ e(m) \leq 0 \text{ for all } m = 1, \cdots, M, \text{ and} \\ H_1 &: \ e(m) > 0 \text{ for some } m = 1, \cdots, M. \end{aligned}$$

Using the data set, we estimate $\hat{e}(m)$. Under regularity conditions, $[\sqrt{n}\hat{e}(1), \cdots, \sqrt{n}\hat{e}(M)]$ is asymptotically jointly normal. As Giacomini and White (2006) showed, this is true under general conditions even when the parameter uncertainty does not vanish as the sample size increases.

White (2000) and Hansen (2005) suggested the following test statistics respectively:

$$\begin{aligned} T^{RC} &= \sqrt{n}\max_{1 \leq m \leq M} \hat{e}(m) \text{ and} \\ T^{SPA} &= \sqrt{n}\max_{1 \leq m \leq M} \hat{e}(m)/\hat{\omega}_m, \end{aligned}$$

where $\hat{\omega}_m^2$ is an estimated asymptotic variance of $\hat{e}(m)$. To obtain approximate critical values, we generate the bootstrap version $\{e_b^*(m)\}_{m=1}^M$, $b = 1, 2, \cdots, B$, from observations. Define c_α^{*RC} to be the α-th quantile of $\{\sqrt{n}\max_{1\leq m\leq M} e_b^*(m)\}_{b=1}^B$. Then an α-level one-sided KS test is obtained as:

$$\text{Reject if and only if } T^{RC} > c_\alpha^{*RC}.$$

This is a test suggested by White (2000). Next, we consider the test suggested by Hansen (2005). Given the bootstrap quantities $\{e_b^*(m)\}_{b=1}^B$, we define

$$\bar{e}_{1,b}^*(m) = e_b^*(m) - \hat{e}(m) \times 1\left\{\hat{e}(m) \geq -\sqrt{(\omega_m^2/n)2\log\log n}\right\}. \tag{8}$$

Let c_α^{*SPA} to be the α-th quantile of $\{\max_{1\leq m\leq M}\sqrt{n}\bar{e}_{1b}^*(m)/\omega_m\}_{b=1}^B$. Then an α-level test of Hansen (2005) is defined to be

$$\text{Reject if } T^{SPA} > c_\alpha^{*SPA}.$$

The construction of a hybrid test that this paper proposes proceeds as follows. First, define the complementary test:

$$T^S = \sqrt{n}\min\left\{\max_{1\leq m\leq M}\hat{e}(m)/\hat{\omega}_m, \max_{1\leq m\leq M} - \hat{e}(m)/\hat{\omega}_m\right\}.$$

We can construct a hybrid test by coupling T^S with T^K or coupling T^S with T^{SPA}. We propose using the latter approach given the result of Hansen (2000). More specifically, take $\bar{e}_{1,b}^*(m)$ as in (8), and construct

$$\begin{aligned} \bar{e}_{2,b}^*(m) &= -e_b^*(m) + \hat{e}(m) \times 1\left\{-\hat{e}(m) \geq -\sqrt{(\omega_m^2/n)2\log\log n}\right\} \text{ and} \\ T_b^{*S} &= \min\left\{\max_{1\leq m\leq M}\sqrt{n}\bar{e}_{1,b}^*(m)/\hat{\omega}_m, \max_{1\leq m\leq M}\sqrt{n}\bar{e}_{2,b}^*(m)/\hat{\omega}_m\right\}. \end{aligned}$$

Let $c_{\alpha/2}^{*S}$ be the $\alpha/2$-th quantile of $\{T_b^{*S}\}_{b=1}^B$. Define $T_b^* = T_b^{*SPA}1\{T_b^{*S} \leq c_{\alpha/2}^{*S}\}$ and take $c_{\alpha/2}^{*H}$ to be the $\alpha/2$-th quantile of $\{T_b^*\}_{b=1}^B$. Then, the hybrid test is defined as

$$\begin{aligned} \text{Reject if } T^S &> c_{\alpha/2}^{*S} \text{ or} \\ \text{if } T^S &\leq c_{\alpha/2}^{*S} \text{ and } T^{SPA} > c_{\alpha/2}^{*H}. \end{aligned}$$

In the next section, we investigate the finite sample performance of the hybrid test by simulation studies.

3.2 Simulations

The simulation design is based on Hansen (2005), and is composed of two parts. First, we focus on the local alternatives considered by Hansen (2005) and compare three types of tests, a test (Reality Check: RC) of White (2000), a test (Superior Predictive Ability: SPA) of Hansen (2005) and this paper's proposal (Hybrid). To provide a background for the simulation design, suppose that $\delta_{m,t-h}$ is a forecasting value that is made h periods in advance using the m-th model and ξ_t is the realized value. Then, the relative performance can be represented as $L(\xi_t, \delta_{m,t-h})$ for some loss function L. Suppose that $\delta_{0,t-h}$ is a forecast using a benchmark model. Let $L(\xi_t, \delta_{m,t-h})$ be the loss from the forecasting through the m-th model and we simply write $L_{m,t} = L(\xi_t, \delta_{m,t-h})$. Then we let $e(m) = \mathbf{E}\left[L_{0,t} - L_{m,t}\right]$. For $m = 1, 2, \cdots, M$ and $t = 1, 2, \cdots, n$, we draw

$$L_{m,t} \sim \text{i.i.d. } N(\lambda(m)/\sqrt{n}, \sigma_m^2)$$

for constants $\lambda(m)$ and $\sigma_m^2 = \frac{1}{2}\exp\left(\arctan(\lambda(m))\right)$. We set $\lambda(0) = 0$. The sum $\hat{e}(m) = \frac{1}{n}\sum_{m=1}^n \{L_{0,t} - L_{m,t}\}$ follows $N(\lambda(m)/\sqrt{n}, \omega_m^2)$, where $\omega_m^2 \approx 1 + \frac{1}{2}\lambda(m) + \frac{1}{4}\lambda^2(m) - \frac{7}{12}\lambda^3(m)$. (See Hansen (2005), p.373, for details.) As for $\lambda(m)$, we consider two different schemes: alternatives with local positivity and alternatives with local positivity and local negativity.

Table 1: Empirical Size of Tests of Predictive Abilities under DGP A ($k = 50, n = 200$)

		$\alpha = 0.05$			$\alpha = 0.10$		
ρ	λ_1	RC	SPA	Hybrid	RC	SPA	Hybrid
0	0	0.0530	0.0535	0.0480	0.1035	0.1065	0.1000
1	0	0.0130	0.0165	0.0360	0.0335	0.0380	0.0580
2	0	0.0075	0.0120	0.0305	0.0195	0.0255	0.0530
3	0	0.0065	0.0120	0.0315	0.0085	0.0195	0.0395
5	0	0.0020	0.0125	0.0280	0.0035	0.0175	0.0305

3.2.1 Alternatives with Local Positivity

The first scheme is identical to the experimentation of Hansen (2005), where $\lambda(m)$ is given as follows:

$$\text{DGP A: } \lambda(m) = \begin{cases} 0, & \text{if } m = 0 \\ \lambda(1), & \text{if } m = 1 \\ \frac{\rho(m-1)}{M-2}, & \text{if } m = 2, \cdots, M, \end{cases}$$

where ρ and $-\lambda(1)$ are chosen from $\{0, 1, 2, 3, 5\}$. Hence the remaining $M - 1$ models are inferior to the benchmark model with their relative performance ordered as $M \prec M - 1 \prec \cdots \prec 2$.

When $\lambda(1) = 0$, no alternative forecasting model strictly dominates the benchmark model, representing the null hypothesis. When $\lambda(1) < 0$, the model 1, having relative expected loss equal to $\lambda(1)$, performs better than the benchmark model. Hence this case corresponds to the alternative hypothesis. The magnitude ρ controls the extent to which the inequalities $\lambda(m) \geq 0$, $m = 2, \cdots, M$, lie away from binding. When $\rho = 0$, the remaining inequalities for models 2 through M are binding, i.e., $e(m) = 0$ for all $m = 2, \cdots, M$. In the simulation studies, the sample size is 200 and the number of Monte Carlo simulations and the bootstrap Monte Carlo simulations 2000.

Tables 1 and 2 show the size of simulation results with $M = 50$ and $M = 100$ under DGP A. The test RC has lower type I error as the design parameter ρ increases. For example, when $\rho = 5$, the rejection probability of the test RC is 0.2% when the nominal size is 5%. This extremely

Table 2: Empirical Size of Tests of Predictive Abilities under DGP A ($k = 100, n = 200$)

		$\alpha = 0.05$			$\alpha = 0.10$		
ρ	λ_1	RC	SPA	Hybrid	RC	SPA	Hybrid
0	0	0.0565	0.0585	0.0565	0.0930	0.0970	0.1025
1	0	0.0065	0.0090	0.0300	0.0300	0.0375	0.0595
2	0	0.0060	0.0065	0.0235	0.0145	0.0250	0.0450
3	0	0.0020	0.0085	0.0205	0.0085	0.0160	0.0365
5	0	0.0010	0.0080	0.0185	0.0080	0.0255	0.0380

conservative size of the test RC is significantly improved by the test SPA of Hansen (2005) which shows the type I error of 1.25%. The hybrid approach provides a significant improvement over the test SPA, yielding type I error of 2.8% in this case. The improvement is prominent throughout the values of $\rho = 1, 2, \cdots, 5$.

The improvement of the rejection probability by the hybrid approach is not without cost in general: it is attained by reducing the rejection probability against certain other alternatives. The theoretical results of this paper predict that the power of the hybrid test can be inferior to the test T^{SPA} of Hansen (2005) when $\rho = 0$, because the one-sided KS test has a strong power against such alternatives. The simulation results in Table 1 indeed show that the rejection probability of the hybrid test is midly lower than that of T^{SPA}. However, the reduction in the rejection probability appears to be only of minor degree. In the case of both nominal sizes 5% and 10%, the rejection probability of the hybrid test is even closer to the nominal size than the other tests.

Tables 3-4 show the power of the three tests. As expected, the reduction in power for the hybrid test is shown in the case of $\rho = 0$. However, the reduction in power appears only marginal. It is interesting to see that the rejection probability of the hybrid test is even better than the test RC when $\lambda(1) = -2, -3$, and -5 in this case. As the inequalities move farther away from binding while maintaining the violation of the null hypothesis, the performance of the hybrid test becomes prominently better. For example, when $\rho = 2$ or 3 and $\lambda(1) = -3$, the test RC of White (2000)

Table 3: Empirical Power of Tests of Predictive Abilities under DGP A ($k = 50, n = 200$)

		$\alpha = 0.05$			$\alpha = 0.10$		
ρ	λ_1	RC	SPA	Hybrid	RC	SPA	Hybrid
	−1	0.0665	0.0720	0.0675	0.1100	0.1175	0.1235
0	−2	0.1775	0.2030	0.1915	0.2600	0.2895	0.2810
	−3	0.4755	0.5400	0.5120	0.6005	0.6485	0.6260
	−5	0.9835	0.9890	0.9870	0.9910	0.9940	0.9925
	−1	0.0275	0.0320	0.0620	0.0640	0.0760	0.1120
1	−2	0.1305	0.1695	0.2185	0.1840	0.2300	0.2765
	−3	0.4395	0.5160	0.5805	0.5690	0.6370	0.6915
	−5	0.9800	0.9900	0.9905	0.9895	0.9930	0.9955
	−1	0.0285	0.0405	0.0820	0.0490	0.0685	0.1265
2	−2	0.1120	0.1765	0.2855	0.1860	0.2550	0.3540
	−3	0.4525	0.5595	0.6805	0.5455	0.6490	0.7435
	−5	0.9760	0.9880	0.9965	0.9920	0.9960	0.9975
	−1	0.0210	0.0375	0.0800	0.0340	0.0655	0.1120
3	−2	0.1175	0.1855	0.2860	0.1765	0.2870	0.3725
	−3	0.4470	0.6055	0.7275	0.5305	0.6755	0.7575
	−5	0.9755	0.9890	0.9970	0.9860	0.9960	0.9985
	−1	0.0115	0.0390	0.0750	0.0200	0.0735	0.1060
5	−2	0.0955	0.2230	0.3110	0.1620	0.3250	0.3860
	−3	0.4095	0.6300	0.7220	0.5145	0.7290	0.7765
	−5	0.9720	0.9950	0.9965	0.9870	0.9975	0.9980

Table 4: Empirical Power of Tests of Predictive Abilities under DGP A ($k = 100, n = 200$)

		$\alpha = 0.05$			$\alpha = 0.10$		
ρ	λ_1	RC	SPA	Hybrid	RC	SPA	Hybrid
	−1	0.0660	0.0690	0.0620	0.1010	0.1070	0.1120
0	−2	0.1300	0.1485	0.1355	0.2170	0.2425	0.2230
	−3	0.4210	0.4870	0.4480	0.5140	0.5680	0.5490
	−5	0.9725	0.9845	0.9805	0.9825	0.9910	0.9900
	−1	0.0230	0.0265	0.0480	0.0390	0.0515	0.0850
1	−2	0.0870	0.1180	0.1655	0.1370	0.1820	0.2400
	−3	0.3685	0.4590	0.5210	0.4675	0.5440	0.6165
	−5	0.9630	0.9800	0.9845	0.9790	0.9900	0.9915
	−1	0.0100	0.0230	0.0500	0.0265	0.0455	0.0805
2	−2	0.0990	0.1430	0.2370	0.1265	0.1910	0.2685
	−3	0.3375	0.4490	0.5925	0.4440	0.5570	0.6565
	−5	0.9630	0.9855	0.9900	0.9820	0.9925	0.9970
	−1	0.0155	0.0280	0.0525	0.0180	0.0440	0.0745
3	−2	0.0650	0.1275	0.2085	0.1140	0.2105	0.2810
	−3	0.3440	0.4695	0.6130	0.4345	0.5850	0.6635
	−5	0.9630	0.9900	0.9965	0.9765	0.9925	0.9970
	−1	0.0075	0.0340	0.0555	0.0125	0.0570	0.0750
5	−2	0.0565	0.1675	0.2375	0.1090	0.2410	0.2830
	−3	0.2925	0.5410	0.6280	0.4045	0.6540	0.6970
	−5	0.9515	0.9885	0.9940	0.9680	0.9955	0.9985

and the test SPA of Hansen (2005) reject the null hypothesis about 45% and 60% respectively. But the hybrid approach rejects the null hypothesis about 73 %. When the number of candidate models M is increased to 100, the results show a similar pattern as shown in Table 2. These results demonstrate that the approach of hybrid test proposed in this paper performs reasonably well in finite samples.

3.2.2 Alternatives with Local Positivity and Local Negativity

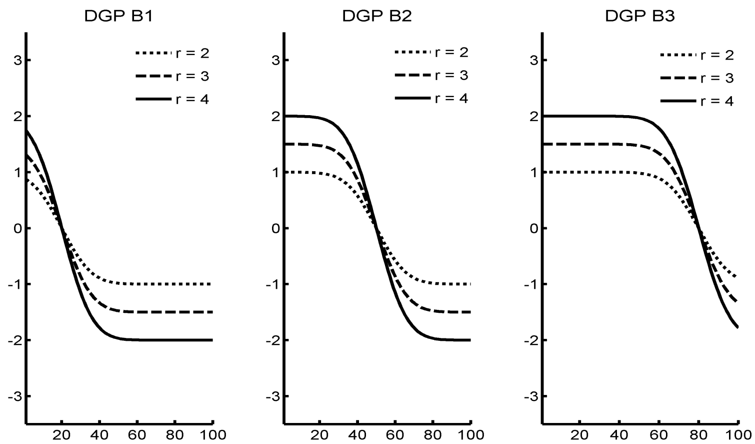
The hybrid test was shown to perform very well relative to the other two tests under Design A. However, Design A mainly focuses on alternatives such that the test RC tends to have weak power. In this section, we consider an alternative design that has alternatives in a more balanced way. The main focus in this design is on the cost of power robustification by the hybrid approach when the other tests perform well. To investigate this, we consider the following scheme: for each $m = 1, \cdots, M$,

$$\begin{aligned} &\text{DGP B1:} \qquad \lambda(m) = r \times \{\Phi(-8m/M) - 4/5\}, \\ &\text{DGP B2:} \qquad \lambda(m) = r \times \{\Phi(-8m/M) - 1/2\}, \text{ and} \\ &\text{DGP B3:} \qquad \lambda(m) = r \times \{\Phi(-8m/M) - 1/5\}, \end{aligned}$$

where Φ is a standard normal distribution function and r is a positive constant running in an equal spaced grid in $[0, 5]$. This scheme is depicted in Figure 4. DGP B1 represents the situation where there is only a small portion of models that perform better than the benchmark model and DGP B3 the situation where there is a large portion of models that perform better than the benchmark model. The general discussion of this paper predicts that the hybrid test has a relatively strong power against the alternatives under DGP B1 while it has a relatively weak power against the alternatives under DGP B3.

The results of the simulation studies are shown in Figure 5. Under DGP B3, all three tests

Figure 4: The Three Designs of $\lambda(m)$

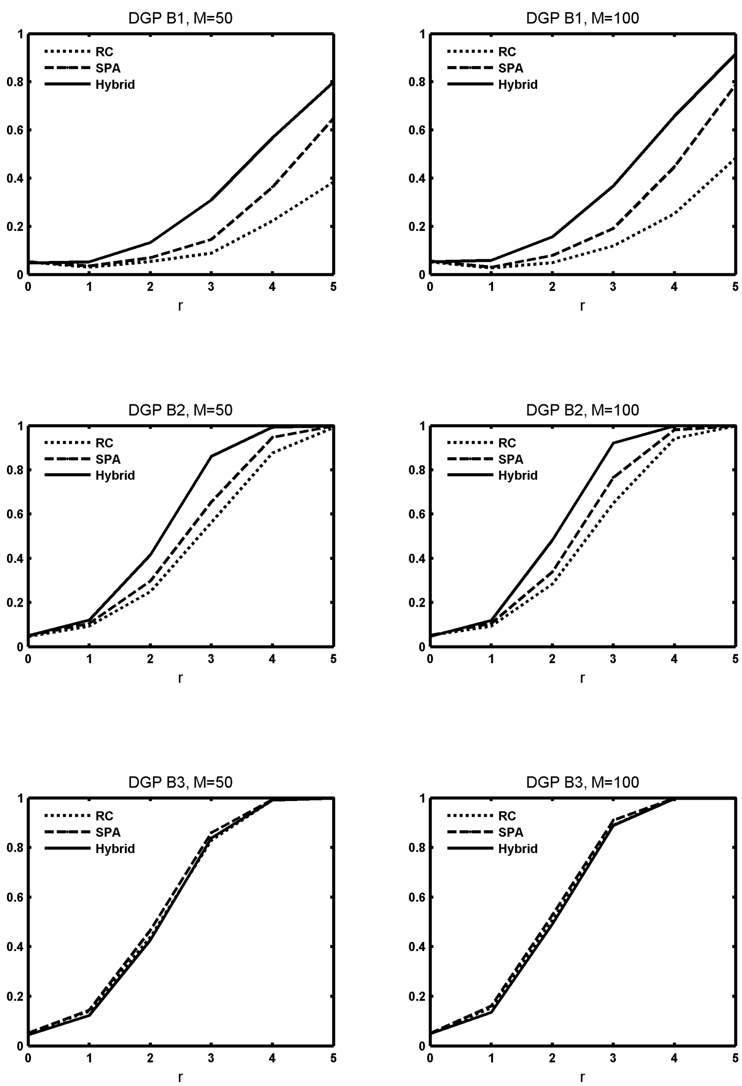


perform equally well. It is worth noting that for the power of the hybrid test (in solid line) is slightly below those of the other two tests. However, this reduction in power appears to be only marginal. This is indeed so when we compare the gain in the power of the hybrid test against the other two types of alternatives, DGP B1 and DGP B2. In this case, the test of Hansen (2005) has a better power than the test RC, as expected from Hansen (2005). Remarkably, the hybrid approach shows conspicuously better power than the other two tests. We conclude that as long as the simulation designs used so far are concerned, the power gain by the hybrid approach is considerable while its cost as a reduction in power under other alternatives is only marginal. This attests to the benefit of the hybrid approach.

4 Closing Remarks

This paper draws attention to the fact that the one-sided KS test of functional inequalities is asymptotically biased. To alleviate this problem, this paper proposes an approach of hybrid test where

Figure 5: Rejection Probabilities at 5% of Testing Predictive Ability



we couple the one-sided KS test with a symmetrized complementary test of a weaker hypothesis. Through simulations, it is shown that this approach yields a test with robust power behavior.

This paper's approach can be applied in numerous different ways. First, one may couple the one-sided KS test with a different complementary test, as far as the nature of the complementarity is made clear. Second, we may apply a similar procedure to one-sided tests involving functionals other than Kolmogorov-Smirnov functional. For example, one might consider a one-sided version of Cramér-von Mises type functional. Third, the approach can be applied to numerous other tests of inqualities beyond predictive ability tests. The question of which modification or extension is suitable often depends on the context of the application.

5 Appendix

The following lemma is crucial for the result of unstable power for the one-sided KS test.

Lemma 1: *Let ν, $\mathbf{M}_0$ and A_0^K be as in* Proposition 1. *Then, for each (potentially stochastic) $c \geq 0$, each $m_0 \in \mathbf{M}_0$, $\varepsilon \in (0, \infty)$ and $K \in \mathbf{R}$, there exists $\delta \in A_0^K$ such that*

$$P\left\{\Gamma_0^K(\nu + \delta) > c\right\} < P\left\{\nu(m_0) > c - \varepsilon\right\} + \varepsilon/K.$$

Proof of Lemma 1: Fix $\varepsilon > 0$, $K > 0$, and $m_0 \in \mathbf{M}_0$. Define $J_0^x = \{m \in \mathbf{M}_0 : ||m - m_0|| \leq x\}$, $x \in [0, \infty)$, and $J_1^x = \mathbf{M}_0 \backslash J_0^x$. Fix $\varepsilon \in (0, 1]$. Take $b \in (0, \infty)$ such that

$$P\left\{\Gamma_0^K(\nu) \leq c + b\right\} \geq 1 - \frac{\varepsilon}{4K}. \tag{9}$$

Define $\tau(\nu) = \{m \in \mathbf{M}_0 : \Gamma_0^K(\nu) = \nu(m)\}$. By Lemma 2.6 and the proof of Theorem 2.7 of Kim and Pollard (1990), $\tau(\nu)$ is a singleton and, being identified with its unique member, and it is measurable. Observe that $\Gamma_0^K(\nu - \nu(m_0))$ is a continuous random variable, which yields that

$$P\{\tau(\nu) \in J_0^0\} = P\{\tau(\nu) = m_0\} = P\{\Gamma_0^K(\nu) = \nu(m_0)\} = 0.$$

Since $P\{\tau(\nu) \in J_0^x\} = P\{||\tau(\nu) - m_0|| \leq x\}$ is increasing and right-continuous in x and $P\{\tau(\nu) \in J_0^0\} = 0$, we can find $x_1 > 0$ such that for all $x \in (0, x_1]$,

$$P\{\tau(\nu) \in J_0^x\} \leq \frac{\varepsilon}{4K}. \tag{10}$$

Since ν is a separable Gaussian process, we find $x_2 > 0$ such that for all $x \in (0, x_2]$,

$$P\{\left|\Gamma_0^K(\nu - \nu(m_0); J_0^x)\right| > \varepsilon/2\} \leq \frac{2}{\varepsilon}\mathbf{E}\left[\left|\Gamma_0^K(\nu - \nu(m_0); J_0^x)\right|\right] < \frac{\varepsilon}{4K}, \tag{11}$$

where $\Gamma_0^K(\nu - \nu(m_0); J_0^x) = \sup_{m \in J_0^x} \nu(m) - \nu(m_0)$. (e.g. Corollary 2.2.8 of van der Vaart and Wellner (1996).) We take $x = \min\{x_1, x_2\}$.

We fix this x and define $D_x(m) = (\varepsilon/2)1\{m \in J_0^x\} - b1\{m \in J_1^x\}$ and

$$\tau_x(\nu) = \{m \in \mathbf{M}_0 : \Gamma_0^K(\nu + D_x) = (\nu + D_x)(m)\}.$$

Now, observe that for any $\delta \leq D_x$,

$$P\left\{\Gamma_0^K(\nu + \delta) \leq c\right\} \geq P\left\{\Gamma_0^K(\nu + D_x) \leq c, \tau_x(\nu) \in J_1^x\right\}.$$

We can write the last term as $A_{1n} + A_{2n}$ where

$$\begin{aligned} A_{1n} &= P\{\Gamma_0^K(\nu + D_x) \leq c, \Gamma_0^K(\nu; J_0^x) + \varepsilon/2 \leq \Gamma_0^K(\nu) - b, \tau_x(\nu) \in J_1^x\} \text{ and} \\ A_{2n} &= P\{\Gamma_0^K(\nu + D_x) \leq c, \Gamma_0^K(\nu; J_0^x) + \varepsilon/2 > \Gamma_0^K(\nu) - b, \tau_x(\nu) \in J_1^x\}. \end{aligned}$$

Consider A_{1n} which we write as

$$\begin{aligned} P\{\Gamma_0^K(\nu) &\leq c + b, \Gamma_0^K(\nu; J_0^x) + \varepsilon/2 \leq \Gamma_0^K(\nu) - b, \tau_x(\nu) \in J_1^x\} \\ &= P\{\Gamma_0^K(\nu; J_0^x) + \varepsilon/2 + b \leq \Gamma_0^K(\nu) \leq c + b, \tau_x(\nu) \in J_1^x\} \\ &> P\{\nu(m_0) + \varepsilon + b \leq \Gamma_0^K(\nu) \leq c + b, \tau_x(\nu) \in J_1^x\} - \varepsilon/(4K), \end{aligned}$$

where the last inequality follows by (11). As for A_{2n}, we bound it from below by

$$\begin{aligned} P\{\Gamma_0^K(\nu; J_0^x) &\leq c - \varepsilon/2, \Gamma_0^K(\nu; J_0^x) + \varepsilon/2 > \Gamma_0^K(\nu) - b, \tau_x(\nu) \in J_1^x\} \\ &\geq P\{\nu(m_0) \leq c - \varepsilon, \Gamma_0^K(\nu) < \nu(m_0) + \varepsilon + b, \tau_x(\nu) \in J_1^x\} - \varepsilon/(4K) \\ &\geq P\{\Gamma_0^K(\nu) < \nu(m_0) + \varepsilon + b, \tau_x(\nu) \in J_1^x\} - P\{\nu(m_0) > c - \varepsilon\} - \varepsilon/(4K). \end{aligned}$$

Combining these together, we deduce that

$$\begin{aligned} & P\left\{\Gamma_0^K(\nu + D_x) \leq c, \tau_x(\nu) \in J_1^x\right\} \\ \geq\ & P\{\Gamma_0^K(\nu) \leq c + b, \tau_x(\nu) \in J_1^x\} - P\{\nu(m_0) > c - \varepsilon\} - \frac{2\varepsilon}{4K} \\ \geq\ & P\{\Gamma_0^K(\nu) \leq c + b\} - P\{\nu(m_0) > c - \varepsilon\} - \frac{3\varepsilon}{4K} \\ \geq\ & 1 - P\{\nu(m_0) > c - \varepsilon\} - \varepsilon/K \end{aligned}$$

by (10) and (9). This gives the wanted result. ■

Proof of Proposition 1 : (i) Note that

$$1 - \alpha \leq P\left\{\Gamma_0^K \nu \leq c_\alpha\right\} \leq P\left\{\nu(m_0) \leq c_\alpha\right\}.$$

Since the distribution function of $\nu(m_0)$ is strictly increasing due to Gaussianity, we can take $\varepsilon_1 > 0$

and $K > 0$ such that

$$P\left\{\nu(m_0) \leq c_\alpha - \varepsilon_1\right\} + \varepsilon_1/K > 1 - \alpha.$$

(Note that this is true even when $\nu(m_0) = 0$ because we can take $\varepsilon \in (0, c_\alpha)$.) Therefore, by Lemma 1, $P_\delta\left\{\Gamma_0^K \nu > c_\alpha\right\} < \alpha$, yielding the wanted result with $A_1 = A(\varepsilon_1)$.

(ii) By the weak convergence assumption,

$$\lim_{n\to\infty} P_\delta\left\{T^K > c_\alpha\right\} = P_\delta\left\{\Gamma_0^K \nu > c_\alpha\right\}.$$

It suffices to show that for each $\varepsilon > 0$, there exists $\delta \in A_0^K$, such that

$$P_\delta\left\{\Gamma_0^K \nu > c_\alpha\right\} < \varepsilon.$$

Due to the condition that $P\{\nu(m_0) > c_\alpha\} = 0$, this is precisely what Lemma 1 says. ■

References

[1] Amisano G. and R. Giacomini (2007), "Comparing density forecasts via weighted likelihood ratio tests," *Journal of Business and Economic Statistics*, 25, 177-190.

[2] Andrews, D. W. K. and P. Guggenberger (2007), "Hybrid and size-corrected subsample models," CFDP 1606.

[3] Andrews, D. W. K. and G. Soares (2007), "Inference for parameters defined by moment inequalities using generalized moment selection," CFDP1631.

[4] Bao, Y., T-H., Lee, and B. Saltoğlu (2007), "Comparing density forecast models," *Journal of Forecasting*, 26, 203-225.

[5] Barret, G. F. and S. G. Donald (2003), "Consistent tests for stochastic dominance," *Econometrica*, 71, 71-104.

[6] Cawley, J. and T. Phillipson (1999), "An empirical examination of information barriers to trade insurance," *American Economic Review* 89, 827-846.

[7] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models. *Econometrica,* 75, 1243-1284.

[8] Chiappori, P-A, and B. Salanié (2000), "Testing for asymmetric information in insurance markets," *Journal of Political Economy* 108, 56-78.

[9] Christoffersen, P. F. (1998), "Evaluating interval forecasts," *International Economic Review*, 39, 841-861.

[10] Davidson, R. and J. Y. Duclos (2000), "Statistical inference for stochastic dominance and measurement for the poverty and inequality," *Econometrica*, 68, 1435-1464.

[11] Denuit, M. and O. Scaillet (2004), "Nonparametric tests for positive quadrant dependence," *Journal of Financial Econometrics*, 2, 422-450.

[12] Diebold, F. X., T. A. Gunther, and A. S. Tay (1999), "Evaluating density forecasts with applications to financial risk management," *International Economic Review*, 39, 863-883.

[13] Diebold, F. X., J. Hahn, and A. S. Tay (1999), "Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange," *Review of Economics and Statistics*, 81, 661-673.

[14] Diebold, F. X. and R. S. Mariano (1995), "Comparing predictive accuracy," *Journal of Business, and Economic Statistics*, 13, 253-263.

[15] Giacomini, R. and H. White (2006), "Tests of conditional predictive ability," *Econometrica*, 74, 1545-1578.

[16] Hansen, P. R. (2003), "Asymptotic tests of composite hypotheses," Working Paper.

[17] Hansen, P. R. (2005), "Testing superior predictive ability," *Journal of Business and Economic Statististics*, 23, 365-379.

[18] Fan, Y. and S. Park (2007), "Confidence sets for some partially identified models," Working Paper.

[19] Janssen, A. (2000), "Global power functions of goodness of fit tests," *Annals of Statistics*, 28, 239-253.

[20] Kim, J. and D. Pollard (1990), "Cube root asymptotics," *Annals of Statistics* 18, 191-219.

[21] Linton, O., E. Maasoumi, and Y-J. Whang (2005), "Consistent testing for stochastic dominance under general sampling schemes," *Review of Economic Studies* 72, 735-765.

[22] Linton, O. K. Song, and Y-J. Whang (2009), "An improved bootstrap test of stochastic dominance," Forthcoming in *Journal of Econometrics*.

[23] Moon, R. H. and F. Schorfheide (2009), "Estimation with overidentifying inequality moment restrictions," Forthcoming in *Journal of Econometrics*.

[24] Rosen, A. (2008), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146, 107-117.

[25] Strasser, H. (1985). *Mathematical Theory of Statistics*. Walter de Gruyter, New York.

[26] West, K. D. (1996), "Asymptotic inference about predictive ability," *Econometrica*, 64, 1067-1084.

[27] White, H. (2000), "A reality check for data snooping," *Econometrica*, 68, 1097-1126.