경제학석사 학위논문

The Level-alpha Hybrid SPA Test and Its Application to Stochastic Dominance

1종 오류가 통제된 하이브리드 예측모형비교 검정과 확률적 지배에의 응용

2016년 8월

서울대학교 대학원
경제학부 경제학 전공
김 드 보 라

The Level-alpha Hybrid SPA Test and Its Application to Stochastic Dominance

지도교수 황 윤 재

이 논문을 경제학석사 학위논문으로 제출함

2017년 3월

서울대학교 대학원

경제학부 경제학전공

김 드 보 라

김드보라의 석사학위논문을 인준함

2017년 6월

위 원 장	김 재 영
부 위 원 장	황 윤 재
위 원	서 명 환

The Level-alpha Hybrid SPA Test and Its Application to Stochastic Dominance

Deborah Kim
Department of Economics
Seoul National University

Abstract

In comparing multiple forecasting models, Song suggested a statistical test coupling two different tests, one-sided sup-test and complementary test, to robustify power against local alternative hypotheses. However, his test does not satisfy the level constraint in the limit. Accordingly, I suggest the method to modify the hybrid test to satisfy the constraint by incorporating Hansen's re-centering method. The proposed test is pointwise asymptotically level α and pointwise consistent against any fixed alternative hypothesis. I also extend the idea of coupling to be applicable to the stochastic dominance test, and suggest hybrid stochastic dominance test by incorporating Donald and Hsu's method. Monte-Carlo simulation study shows that the hybrid SPA test outperforms existing tests in certain designs while its performance is similar to that of SPA test by Hansen in most cases in finite sample. Meanwhile it seems that coupling does not enhance the power when it comes to stochastic dominance test.

Keywords: Nonparametric test, Superior predictive ability, Stochastic dominance, Hybrid test
Student Number: 2014-22287

Contents

1 Introduction

A strand of literature has been developed on comparing multiple forecasting methods. White [1] proposed a one-sided sup-test for testing superior predictive ability (SPA), which is known as the reality check (RC) for data snooping. Hansen [2] demonstrated that the one-sided sup test tests are asymptotically biased and developed the way to enhance the power. Song [3] generalized Hansen's observation on biasedness and suggested to couple White's one-sided test with a complementary test based on two-sided test with the purpose of rendering power of the test robust to local alternative hypotheses.

Difficulties in developing SPA tests lie in the fact that the null hypothesis is composite and in general the limiting distribution depends on unknown parameters. While Hansen proposed the method to invoke the null distribution, White [1] circumvented this problem by taking critical value from the least favourable distribution under the null hypothesis. This necessarily makes the test asymptotically non-similar on the boundary, and so does Song's test which incorporates the test of White. Nevertheless this does not make the tests deficient. An interesting finding of Andrews [4] supports tests that are non-similar on the boundary in the limit. He showed that there exist tests with inequality restrictions which are asymptotically similar on the boundary of the null hypothesis, but they have very poor power.

However a closer look reveals that Song's test is not pointwise asymptotically level α as well as it is not asymptotically similar on the boundary. Under certain DGPs, the rejection probability of the test exceeds the significance level. This is attributable to the fact that the least favourable distribution under the null hypothesis for one-sided SPA test does not correspond to the least favourable distribution under the null hypothesis for two-sided SPA test. In the Neyman-Pearson paradigm, to control a type 1 error to be smaller than nominal level is considered of importance. A rejection of the null hypothesis would not be the evidence advocating the alternative hypothesis if the type 1 error is out of control. (see Romano et al. [5]).

The main purpose of the paper is to clarify that Song [3]'s proposal is not a level-α test in the limit and to modify Song's hybrid test to satisfy the level-constraint. By incorporating Hansen[2]'s re-centering method, we develop a test which does not resort on the least favourable configuration. It is shown that the proposed test is pointwise asymptotically level-α and pointwise consistent against any fixed alternative hypotheses. Monte-Carlo simulation shows that there are DGPs under which our proposal outperforms existing SPA tests while its performance is similar to that of SPA test by Hansen [2] in many cases.

Another purpose is to apply the modified approach to the first order stochastic dominance (SD) test. SD is an ordering rule of distributions, which is widely used in various areas including economics, finance, medicine, and so on. SD test has the similar structure with that of SPA test; SD can be interpreted as comparing the infinite number of forecasting models. Accordingly, SD test also suffers from the low power against local alternative hypotheses, and applying the idea of coupling tests may improve the power properties.

The literature related to SD has been developed in the similar way. Barrett and Donald [6] developed the framework for one-sided sup test for SD based on the least favourable case like White's test. Linton, Maasoumi, and Whang [7] generalized the method applicable to time series data based on subsampling technique. By adopting the re-centering approach of Hansen [2], Donald and Hsu [8] proposed the method to approximate the sample distribution under the null hypothesis.

Song [3] mentioned that his idea can be embedded to SD test, a naive application yields a test not satisfying the level constraint in the limit. Hence we propose to couple two tests, one-sided sup SD test with a complementary SD test by adopting the re-centering method of Donald and Hsu [8]. Likewise, the proposed test is pointwise asymptotically level-α and pointwise consistent against any fixed alternative hypotheses. Meanwhile, simulation result suggests that coupling does not enhance the power significantly given finite sample.

The rest of the paper is organized as follows. Section 2 describes the hypotheses of interest. In Section 3, we introduce Song [3]'s hybrid test and demonstrate that the rejection probability may exceed the nominal level under the null hypothesis. In Section 4, we propose the way to modify the hybrid test to be size-controlled by adopting the re-centering approach of Hansen [2]. Section 5 expands our discussion and applies the proposed method to the stochastic dominance test. The finite sample power properties of the proposed tests are investigated Section 6 using Monte-Carlo simulation. Section 7 concludes the paper.

2 Hypotheses

Let us consider a situation where a decision must be made h periods ahead. Let $\{\delta_{k,t-h}, k = 0, 1, \cdots, m\}$ be a finite of possible decision rules which are referred to as forecasting methods. Forecasting methods are evaluated with a real-valued loss function, $L(\xi_t, \delta_{k,t-h})$ where ξ_t is a random variable of interest, which is unknown at the time that we do forecast. We evaluate forecasts in terms of their expected loss, $E[L(\xi_t, \delta_{k,t-h})]$ which can be interpreted as the risk of prediction based on kth forecasting model. An overview of notation is given in Table 1. The forecasting method can represent a point forecast, an interval forecast, or a density forecast. Depending on the types of forecasting methods, we can use various loss functions such as a mean squared loss function, the Kullback-Leibler divergence or the Kolmogorov-Smirnov statistic. See Hansen [2] or Song [3] for more examples.

Among $m+1$ forecasting models, the first one, $\delta_{0,t-h}$, plays a special role and is referred to as the benchmark. We are interested in knowing whether there is any alternative forecasting model among $\delta_{k,t-h}$ where $k = 1, \cdots, m$, whose performance strictly dominates that of benchmark in terms of expected loss. Thus we seek a test of the null hypothesis that the benchmark is not inferior to any of the alternatives. To compare the performance between two different models, we define relative performance variables as

$$d_{k,t} := L_{0,t} - L_{k,t} \quad k = 1, \cdots, m.$$

$d_{k,t}$ denotes the performance of model k relative to the benchmark at time t. Define $\mathbf{d_t} := (d_{1,t}, \cdots, d_{m,t})'$ as the vector of relative performances at time t. We formulate the hypothesis of interest as

$$H_0 : \mu_k \leq 0 \text{ for all } k = 1, \cdots, m$$
$$H_1 : \mu_k > 0 \text{ for some } k = 1, \cdots, m$$

given that $\mu := E(\mathbf{d_t}) \in \mathcal{R}^m$ is well-defined.

$t = 1, \cdots, n$	Sample period for the model comparison
$k = 1, \cdots, m$	Model index where $k = 0$ is the benchmark
$M := \{1, \cdots, m\}$	index set
ξ_t	Variable of interest
$\delta_{k,t-h}$	The $k$the decision rule
$L_{k,t} := L(\xi_t, \delta_{k,t-h})$	Observed loss of the kth decision rule
$d_{k,t} := L_{0,t} - L_{k,t}$	Performance of model k relative to the benchmark
$\bar{d}_k := 1/n\Sigma_{t=1}^n d_t$	Average relative performance of model k
$\mathbf{d}_t := (d_{1,t}, \cdots, d_{m,t})'$	Vector of relative performances at time t
$\bar{\mathbf{d}} = 1/n\Sigma_{t=1}^n \mathbf{d}_t$	Vector of average relative performance
$\mu_k := E(d_{k,t})$	Expected excess performance of model k
$\mu := (\mu_1, \cdots, \mu_m)'$	Vector of expected excess performances
$\Sigma := avar(\sqrt{n}(\bar{\mathbf{d}} - \mu))$	Asymptotic $m \times m$ covariance matrix

Table 1: Overview of notation

3 Song's Hybrid Test

In his article, Song [3] generalized Hansen [2]'s observation that the one-sided sup-test is asymptotically biased, and proposed a method to enhance power properties against the local alternatives. The main idea of the method is to couple the one-sided sup-test with a complementary test. The former test refers to one-sided sup-test based on the least favourable case, whereas the later test is the counterpart test of Linton, Maasoumi and Whang [7] in the context of finite number of inequalities test. He showed that his test outperforms the existing one-sided sup-tests under certain data generating processes through simulation study. However a close look reveals that the hybrid test he proposed is not a level-α test. To shed lights on this problem, we explain the detailed procedure of the test.

In the hybrid test, Song [3] assumed that there exists a Gaussian process Z with a continuous sample path on M such that

$$\sqrt{n}(\bar{\mathbf{d}} - \mu) \Rightarrow Z \text{ as } n \to \infty$$

where $\Rightarrow$ denotes the weak convergence of stochastic processes on M. This assumption admits infinite index set M. Define two test statistics as

$$\begin{aligned} U_n^1 &:= \sqrt{n} \max_{k \in M} \bar{d}_k \text{ and} \\ U_n^2 &:= \sqrt{n} \min(\max_{k \in M} \bar{d}_k, \max_{k \in M}(-\bar{d}_k)). \end{aligned}$$

Given the significance level $\alpha \in (0, 1)$ and a real number $\gamma \in [0, 1]$, the hybrid test is defined as follows.

$$\textit{Reject } H_0 \textit{ if } U_n^2 > v_\alpha^2(\gamma), \textit{ or if } U_n^2 \le v_\alpha^2(\gamma) \textit{ and } U_n^1 > v_\alpha^1(\gamma)$$

where $v_\alpha^1(\gamma)$ and $v_\alpha^2(\gamma)$ are threshold values such that

$$\lim_{n \to \infty} P\{U_n^2 > v_\alpha^2(\gamma)\} = \alpha\gamma$$

and

$$\lim_{n \to \infty} P\{U_n^2 \le v_\alpha^2(\gamma) \textit{ and } U_n^1 > v_\alpha^1(\gamma)\} = \alpha(1 - \gamma).$$

For convenience, we refer rejecting the null hypothesis in case of $U_n^1 \le v_\alpha^1(\gamma)$ to one-sided test and refer rejecting the null hypothesis in case of $U_n^2 \le v_\alpha^2(\gamma)$ to two-sided test. The superscript indicates whether the test statistic is from one-sided or two-sided test. In order to compute the approximated critical value, he suggested to implement existing bootstrap procedure. When observations are stationary series, he suggested to use stationary bootstrap method of Politis and Romano [9]. Given bootstrap versions $\{\bar{\mathbf{d}}_b\}_{b=1}^B$, define centred bootstrap sample $\{\tilde{\mathbf{d}}_b^*\}_{b=1}^B$ where,

$$\tilde{d}_{b,k}^* := \bar{d}_{b,k}^* - \bar{d}_k \text{ and } \tilde{\mathbf{d}}_b^* := (\tilde{d}_{b,1}^*, \cdots, \tilde{d}_{b,m}^*)'.$$

Then simulate the bootstrap distribution P^* of (U_n^1, U_n^2) by generating $\{U_{b,n}^{1*}, U_{b,n}^{2*}\}_{b=1}^B$ based on centred bootstrap sample. The critical value for the two-sided test is obtained from the bootstrap empirical distribution of $\{U_{b,n}^{2*}\}_{b=1}^B$ by

$$\frac{1}{B}\sum_{b=1}^{B} 1\{U_{b,n}^{2*} > v_{\alpha,n}^{2*}(\gamma)\} = \alpha\gamma.$$

Given $v_{\alpha,n}^{2*}(\gamma)$, we take $v_{\alpha,n}^{1*}(\gamma)$ to be the $(1-\alpha(1-\gamma))$-quantile of the sample $\{U_{b,n}^{1*} \cdot 1(U_{b,n}^{2*} \leq v_{\alpha,n}^{2*}(\gamma))\}_{b=1}^B$.

This hybrid test is not a level-α test, which means that the limiting probability to reject the null hypothesis may exceed the pre-specified significance level α even if the null hypothesis is true. To be specific, the two-sided test with the critical value obtained from centred bootstrap sample is not a level $\alpha\gamma$ test in case the null hypothesis holds with at least one strict inequality. This is because the bootstrap distribution does not mimic nor stochastically dominate the limiting distribution of the test statistic U_n^2.

Restricting our attention to the simple case helps us to clarify the problem. Suppose $\mu \leq 0$ and there exists k, k' such that $\mu_k < 0$ and $\mu_{k'} = 0$. Define the contact set as $M^* := \{k \in M : \mu_k = 0\}$. Then M^* is a non-empty proper subset of M. It holds that

$$U_n^1 := \sqrt{n} \max_{k \in M} \bar{d}_k \Rightarrow \max_{k \in M^*} Z_k \text{ as } n \to \infty$$

where $Z = (Z_1, \cdots, Z_m)' \sim N(0, \Sigma^0)$ and $\Sigma_{ij}^0 = \Sigma_{ij} 1(\mu_i = 0, \mu_j = 0)$. Also it holds as $n \to \infty$ that

$$\begin{aligned}\sqrt{n}\max_{k\in M}(-\bar{d}_k) &\geq \sqrt{n}\max_{k\in M\setminus M^*}(-\bar{d}_k)\\ &= \sqrt{n}\max_{k\in M\setminus M^*}(-\bar{d}_k+\mu_k)-\sqrt{n}\mu_k \xrightarrow{p} \infty.\end{aligned}$$

Hence as $n \to \infty$, it holds that

$$U_n^2 := \min(\sqrt{n}\max_{k\in M}\bar{d}_k, \sqrt{n}\max_{k\in M}(-\bar{d}_k)) \Rightarrow \max_{k\in M^*} Z_k.$$

Meanwhile, $v_{\alpha,n}^{2*}(\gamma)$ approximates the $(1-\alpha\gamma)$-quantile of the distribution of $\min(\max_{k\in M} Z_k, \max_{k\in M}(-Z_k))$. Thus for large n,

$$\begin{aligned}P\{\mathit{Reject}\ H_0\} &= P\{U_n^2 > v_{\alpha,n}^{2*}(\gamma),\ \mathit{or}\ U_n^2 \leq v_{\alpha,n}^{2*}(\gamma)\ \mathit{and}\ U_n^1 > v_{\alpha,n}^{1*}(\gamma)\}\\ &\geq P\{U_n^2 > v_{\alpha,n}^{2*}(\gamma)\}\\ &\simeq P\{U_n^1 > v_{\alpha,n}^{2*}(\gamma)\}\\ &\simeq P\{\max_{k\in M^*} Z_k > v_{\alpha}^{2}(\gamma)\}. \qquad (1)\end{aligned}$$

The third equality holds because $U_n^1 \simeq U_n^2$ for large n. Note that $v_\alpha^2(\gamma)$ is not obtained from the distribution of $\max_{k\in M^*} Z_k$. Also $\min(\max_{k\in M} Z_k$, $\max_{k\in M}(-Z_k))$ does not stochastically dominate $\max_{k\in M^*} Z_k$. As a result, although the probability (1) varies depending on the number of elements in the contact set and the index set or the covariance of Z, there exist cases where the probability (1) exceeds the significance level $\alpha\gamma$. The following provides such an example.

Example Consider a random vector Z following a bivariate normal distribution such that $Z = (Z_1, Z_2)' \sim N(\mu, \Sigma)$ where

$$\mu = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 2.0811 & 1.0811 \\ 1.0811 & 2.0811 \end{pmatrix}.$$

Suppose $(Z-\mu)$ follows the limiting distribution of $\sqrt{n}(\bar{d}_1-\mu_1, \bar{d}_2-\mu_2)'$. Since $E(Z_1) = 0$ and $E(Z_2) < 0$, the index set is $M = \{1, 2\}$ and the contact set is $M^* = \{1\}$. The centred bootstrap distribution approximates the

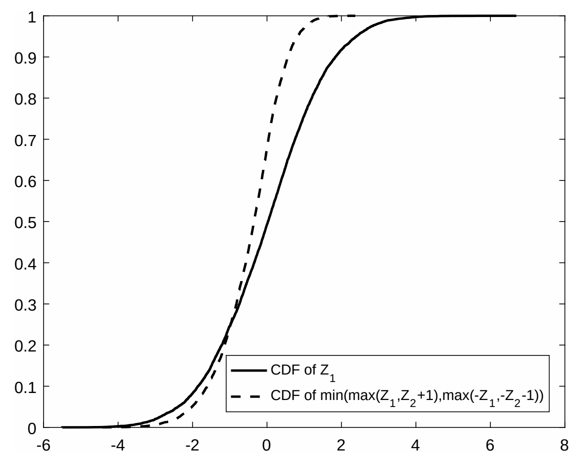

Figure 1: Cumulative distribution functions (CDFs) of Z_1 and $\min(\max(Z_1, Z_2+1), \max(-Z_1, -Z_2-1))$; two CDFs are not identical and none of them stochastically dominates the other.

distribution of $\min(\max(Z_1, Z_2+1), \max(-Z_1, -Z_2-1))$. We compute the critical value for the two-sided test $v_\alpha^{2*}(\gamma)$ from the bootstrap distribution. From the equation (1), we know that the rejection probability is determined by $\max_{k\in M^*} Z_k$ which reduces to Z_1 in this example. Figure 1 presents two CDFs of Z_1 and $\min(\max_{k=1,2}(Z_k - \mu_k), \max_{k=1,2}(-Z_k + \mu_k))$. Two CDFs are not identical and none of them stochastically dominates the other. This implies that this is not an asymptotically exact test nor an asymptotically conservative test, and thus the rejection probability under the null hypothesis is not controlled.

Figure 2 visualizes this situation. The ellipse on the parameters space (μ_1, μ_2) represents a contour of the joint distribution of (Z_1, Z_2) over which the integrated density is almost one. Since Z_1 and Z_2 are positively correlated, its major axis of the ellipse lies near the line $y = x$. (In fact, the covariance matrix Σ is designed to have $\{(1,1)', (-1,1)'\}$ as its eigenvectors.) The critical value $v_\alpha^2(\gamma)$ is determined to satisfy that integrated density over the rejection region(shaded area) equals to $\alpha\gamma$. That is, the probability on the area highlighted with slashes on the left-hand side figure is close to $\alpha\gamma$.

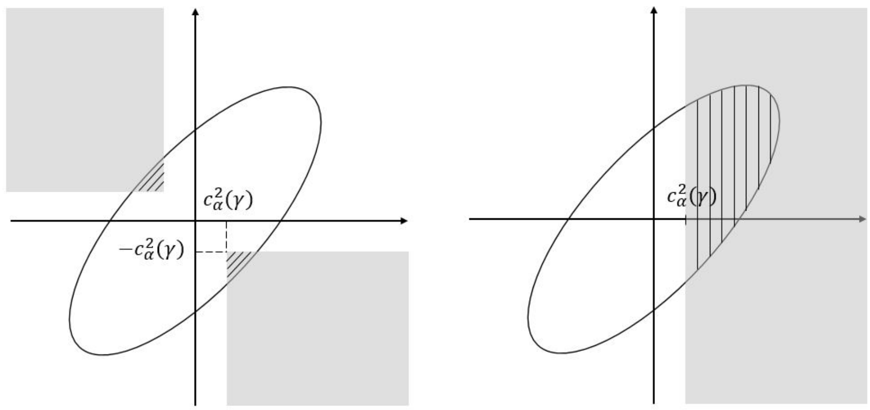

Figure 2: (left) determination of the critical value $v_\alpha^2(\gamma)$, (right) rejection probability.In both figures, the shaded area represents the rejection region of the two-sided test.

However, the rejection probability is the integrated marginal density of Z_1 over the area $\{x \in \mathcal{R} : x > v_\alpha^2(\gamma)\}$, which is represented as the shaded area on the right-hand side figure. The shaded area looks much wider than the shaded area on the left side, suggesting that the probability may exceeds the significance level-α. $\square$

4 Level-α Hybrid SPA Test

The problem that Song [3]'s hybrid test is not a level-α test occurs mainly because we compute the critical value not from the limiting distribution of the test statistic nor from the distribution stochastically dominating the limiting distribution of the test statistic. This problem can be fixed by adopting Hansen [2]'s re-centering method. He devised the way to invoke a sample-dependent null distribution using bootstrap in the one-sided SPA test setting. The adoption of re-centering method allows us to mimic the limiting distribution of the test statistics. In this section, we explain how to establish level-α hybrid test based on Hansen's re-centering method.

First let us start with introducing assumptions which are also featured in Hansen [2].

Assumption 4.1. $\{\mathbf{d}_t\}_{t=1}^n$ *is strictly stationary and* α*-mixing of size* $-(2+\delta)(r+\delta)/(r-2)$ *for some* $r > 2$ *and* $\delta > 0$ *where* $E|d_{k,t}|^{r+\delta} < \infty$ *and* $var(d_{k,t})$ *for* $k = 1, \cdots, m$*.*

By Assumption 4.1, the mean of the true distribution μ is well-defined, and it holds that $\sqrt{n}(\bar{\mathbf{d}}_n - \mu) \Rightarrow N(0, \Omega)$. Also according to Hansen, this assumption justifies the use of bootstrap techniques.

4.1 Test Statistics

As for the test statistics, we define the following:

$$T_n^1 := \sqrt{n} \max_{k \in M} \left(\frac{\bar{d}_{n,k}}{\bar{w}_{n,k}} \vee 0\right), \text{ and}$$

$$T_n^2 := \sqrt{n} \min\left(\max_{k \in M}\left(\frac{\bar{d}_{n,k}}{\bar{w}_{n,k}} \vee 0\right), \max_{k \in M}\left(-\frac{\bar{d}_{n,k}}{\bar{w}_{n,k}} \vee 0\right)\right)$$

where $a \vee b := \max(a, b)$. $\bar{w}_k^2$ is an consistent estimator of $w_{n,k}^2 := \text{var}(\sqrt{n}\bar{d}_{n,k})$ for $k = 1, \cdots, m$. T_n^1 is the test statistic which is introduced in Hansen, and T_n^2 is the two-sided version of it. These test statistics are different from those in the previous section in that $\bar{d}_k$'s are studentized and the test statistics are always greater than or equal to zero. Two alterations themselves do not change the hybrid test to be a level-α test. Yet, we adopt the statistics that Hansen proposed because the studentization typically improves the power and normalizing negative values to zero makes proofs simpler. With regard to the effect of power improvement through studentization, refer to Hansen ([2], section 2.2). To keep notation brief, we define $T_n^{1,-} := \sqrt{n} \max_{k \in M}(-\frac{\bar{d}_{n,k}}{\bar{w}_{n,k}} \vee 0))$, and represent T_n^2 as $T_n^1 \wedge T_n^{1,-}$ where $a \wedge b := \min(a, b)$.

The following lemma explains the asymptotic distribution of test statistics. To describe the distribution, define the asymptotic variance-covariance matrix $\Sigma_{ij}^0 := \Sigma_{ij} 1(\mu_i = 0, \mu_j = 0)$ and the contact set $M^* := \{k \in M : \mu_k = 0\}$.

Lemma 4.1. *Suppose Assumption 4.1 holds.*
(1) Under the null hypothesis $\mu \leq 0$, we have that

$$
\begin{aligned}
&T_n^1 \Rightarrow \max_{k \in M^*}(\frac{Z_k}{w_k} \vee 0) \\
&T_n^2 \Rightarrow \begin{cases} \min(\max_{k \in M}(\frac{Z_k}{w_k} \vee 0), \max_{k \in M}(-\frac{Z_k}{w_k} \vee 0)) \ \textit{if } \mu = 0, \\ \max_{k \in M^*}(\frac{Z_k}{w_k} \vee 0) \ \textit{otherwise.} \end{cases}
\end{aligned}
$$

In particular, if $M^+ = \emptyset$, then both statistics converges to 0 in probability.
(2) Under the alternative hypothesis $\mu \not\leq 0$, we have that

$$
\begin{aligned}
&T_n^1 \xrightarrow{p} \infty \\
&T_n^2 \Rightarrow \begin{cases} \max_{k \in M^*}(-\frac{Z_k}{w_k} \vee 0) \ \textit{if } \mu \geq 0 \\ \infty \ \textit{otherwise.} \end{cases}
\end{aligned}
$$

where $Z \sim N(0, \Sigma^0)$ and $w_k = \sqrt{\Sigma_{kk}}$.

The inequality holds if and only if it holds for all elements of a vector. That is, $\mu \leq 0$ if and only if $\mu_k \leq 0$ for all $k = 1, \cdots, m$. We use $\xrightarrow{p}$ to denote the convergence in probability.

Note that non-binding constraints, kth constraints such that $\mu_k < 0$, do not influence on the limiting distribution; binding constraints only matter. It is worth paying attention to the fact that the limiting distribution given $\mu = 0$ is different from that given $\mu \neq 0$ under the null hypothesis. As we mentioned in the previous section, the limiting distribution given $\mu = 0$ does not stochastically dominate the limiting distribution given $\mu \neq 0$. This tells us that $\mu = 0$ is not a least favourable case when it comes to the two-sided test. As a result, a test with the critical value from the centred bootstrap is not a conservative test.

4.2 Re-centering Function

Lemma 4.1 demonstrates that only the binding constraints, kth constraint with $\mu_k = 0$, matter for the asymptotic distribution under the null hypothesis. This property impedes the centred bootstrap procedure from mimicking the limiting distribution of test statistics because the centred bootstrap approximates the limiting distribution of the test statistic given $\mu = 0$. To avoid this, Hansen introduced the re-centering function $\hat{\mu}_n^c$ which is defined as follows.

$$\hat{\mu}_n^c = \bar{d}_{n,k} \cdot 1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < a_n)$$

The re-centering function converges to the negative value as the sample size n grows to infinity provided the corresponding constraint is unbinding. The following assumption imposes condition under which the re-centering function converges.

Assumption 4.2. *Let a_n be a sequence of negative numbers such that $\lim_{n\to\infty} a_n = -\infty$ and $\lim_{n\to\infty} a_n/\sqrt{n} = 0$.*

That is, the sequence of negative values a_n should converge to negative infinity with a rate slower than the rate $\sqrt{n}$.

Lemma 4.2. *Suppose Assumption 4.1 and 4.2 hold. For all $k \in M$, we have $|\hat{\mu}_{n,k}^c - \min(\mu_k, 0)| \xrightarrow{p} 0$.*

Lemma 4.2 tells us that the re-centering function converges to the mean in probability when the constraint is not binding (i.e. $\mu_k < 0$). The following lemma shows that we can generate the limiting distribution of test statistics under the null hypothesis, provided that a sequence of random vectors is centred at the re-centering function. To describe the asymptotic distribution, let us denote $\Sigma_{ij}^+ = \Sigma 1(\mu_i \geq 0, \mu_j \geq 0)$ and $M^+ := \{k \in M : \mu_k \geq 0\}$.

Lemma 4.3. *Suppose Assumption 4.1 and 4.2 hold. Consider a random vector $\mathbf{Z}_n = (Z_{n,1}, \cdots, Z_{n,m})'$ such that $\sqrt{n}(\mathbf{Z}_n - \hat{\mu}_n^c) \Rightarrow N(0, \Sigma)$, and a random vector $\mathbf{Z}^+ = (Z_1^+, \cdots, Z_m^+)'$ such that $\mathbf{Z}^+ \sim N(0, \Sigma^+)$. Then it* *holds that*

$$\sqrt{n}\max_{k\in M}(Z_{n,k}\vee 0) \Rightarrow \max_{k\in M^+}(Z_k^+\vee 0)$$

$$\sqrt{n}\min(\max_{k\in M}(Z_{n,k}\vee 0), \max_{k\in M}(-Z_{n,k}\vee 0)) \Rightarrow$$

$$\begin{cases} \min(\max_{k\in M}(Z_k^+\vee 0), \max_{k\in M}(-Z_k^+\vee 0)) \text{ if } \mu\geq 0, \\ \max_{k\in M^+}(Z_k^+\vee 0) \text{ otherwise.} \end{cases}$$

In particular, if $M^+ = \emptyset$*, then both quantities,* $\sqrt{n}\max_{k\in M}(Z_{n,k}\vee 0)$ *and* $\sqrt{n}\min(\max_{k\in M}(Z_{n,k}\vee 0), \max_{k\in M}(-Z_{n,k}\vee 0))$*, converges to 0 in probability.*

4.3 Simulated Critical Values

Following Hansen [2], we implement the stationary bootstrap of Politis and Romano [9] which is based on the pseudo-time series of the original data. The pseudo-time series $\{\mathbf{d}^*_{b,t}\} := \{\mathbf{d}_{\tau_{b,t}}\}$ for $b = 1,\cdots,B$ are resamples from $\mathbf{d}_t$ where $\{\tau_{b,1},\cdots,\tau_{b,n}\}$ is constructed by combining blocks of $\{1,\cdots,n\}$ with random lengths. Specifically, we generate B resamples from two random $B\times n$ matrices, U and V, where the elements, $u_{b,t}$ and $v_{b,t}$, are independent and uniformly distributed on (0,1]. The first element of each resample is defined by $\tau_{b,1} = \lceil nu_{b,1}\rceil$, where $\lceil x\rceil$ is the smallest integer that is larger than or equal to x. For $t = 2,\cdots,n$ the elements are given recursively by

$$\tau_{b,t} = \begin{cases} \lceil nu_{b,1}\rceil \text{ if } v_{b,t} < q \\ 1(\tau_{b,t-1} < n)\cdot\tau_{b,t-1} + 1 \text{ if } v_{b,t}\geq q. \end{cases}$$

So with the probability q, the tth element is chosen uniformly on $\{1,\cdots,n\}$ and with probability $1-q$, the tth element is chosen to be the integer that follows $\tau_{b,t-1}$, unless $\tau_{b,t-1} = n$ in which case $\tau_{b,t} = 1$.

From the pseudo-time series, we calculate their sample averages, $\bar{\mathbf{d}}^*_{n,b} = \frac{1}{n}\sum_{i=1}^n \mathbf{b}^*_{n,b,t}$, $b = 1,\cdots,B$ that can be viewed as independent draws from the distribution of $\bar{\mathbf{d}}_n$, under the bootstrap distribution. Define the re-centred bootstrap sample average as $\bar{Z}^*_{n,k,b} = \frac{1}{n}\sum_{t=1}^n Z^*_{n,k,b,t}$, and $\bar{\mathbf{Z}}^*_{n,b} = (\bar{Z}^*_{n,1,b},\cdots,\bar{Z}^*_{n,m,b})'$ where $Z^*_{n,k,b,t} = d^*_{n,k,b,t} - \bar{d}_{n,k} + \hat{\mu}^c_{n,k}$. By its definition, $\bar{\mathbf{Z}}^*_{n,b}$ is centred at $\hat{\mu}^c_n$. The following lemma provides an intermediate step to estimate the asymptotic distribution of test statistics.

Lemma 4.4. *(Theorem 3 of Hansen [2]) Let Assumption 4.1, 4.2 and 4.3 hold. Then we have*

$$\sup_{z \in \mathcal{R}^m} |P^*\{\sqrt{n}(\bar{\mathbf{Z}}^*_{n,b} - \hat{\mu}^c_n) \le z\} - P\{\sqrt{n}(\bar{d}_n - \mu) \le z\}| \xrightarrow{p} 0$$

where P^* denotes the bootstrap probability measure.

Let us define the bootstrap test statistics as follows:

$$T^{1*}_{n,b} = \sqrt{n} \max_{k \in M}(\frac{\bar{Z}^*_{n,k,b}}{\bar{w}_{n,k}} \vee 0), \text{ and}$$
$$T^{2*}_{n,b} = \sqrt{n} \min(\max_{k \in M}(\frac{\bar{Z}^*_{n,k,b}}{\bar{w}_{n,k}} \vee 0), \max_{k \in M}(-\frac{\bar{Z}^*_{n,k,b}}{\bar{w}_{n,k}} \vee 0)).$$

In order to compute the studentizing factors, we use the bootstrap population directly as Hansen recommended, which is given by

$$\bar{w}_k^2 := \hat{\beta}_{0,k} + 2\sum_{i=1}^{n-1} \kappa(n,i)\hat{\beta}_{i,k}$$

where

$$\hat{\beta}_{i,k} := \frac{1}{n}\sum_{j=1}^{n-1}(d_{k,j} - \bar{d}_k)(d_{k,j+1} - \bar{d}_k),$$

$i = 0, \cdots, n-1$ are the usual empirical covariance and the kernel weights under the stationary bootstrap are given by

$$\kappa(n,i) := \frac{n-i}{n}(1-q)^i + \frac{i}{n}(1-q)^{n-1}.$$

As a result of Lemma 4.3 and 4.4, now we can estimate the asymptotic distribution of the test statistic; the bootstrap statistics approximate the limiting distribution of the test statistics under the null hypothesis, i.e.

$$T^{1*}_{n,b} \overset{p}{\Rightarrow} \max_{k \in M^+}(\frac{Z^+_k}{w_k} \vee 0)$$
$$T^{2*}_{n,b} \overset{p}{\Rightarrow} \begin{cases} \min(\max_{k \in M^+}(\frac{Z^+_k}{w_k} \vee 0), \max_{k \in M^+}(-\frac{Z^+_k}{w_k} \vee 0)) \text{ if } \mu \ge 0 \\ \max_{k \in M^+}(\frac{Z^+_k}{w_k} \vee 0) \text{ otherwise.} \end{cases}$$

where $\overset{p}{\Rightarrow}$ denotes the weak convergence conditional on the sample path with probability approaching 1. Note that they are stochastically bounded even under the alternative hypothesis. Hence we compute critical values from this distribution. Define

$$\tilde{c}^{1*}_{\alpha,n}(\gamma) := \inf\{c \in \mathcal{R} : P^*\{T^{1*}_{n,b} \leq c\} \geq 1 - \alpha(1-\gamma)\}, \text{ and}$$
$$\tilde{c}^{2*}_{\alpha,n}(\gamma) := \inf\{c \in \mathcal{R} : P^*\{T^{2*}_{n,b} \leq c\} \geq 1 - \alpha\gamma\}$$

In the case that all constraints are non-binding (i.e. $\mu_k < 0$ for all $k \in M$), both test statistics T^1_n and T^2_n converge to zero in probability. Since the bootstrap distribution also degenerates, $\tilde{c}^{1*}_{\alpha,n}(\gamma)$ and $\tilde{c}^{2*}_{\alpha,n}(\gamma)$ converge to zero in probability. Since it is difficult to compare the convergence rate between tests statistics and critical values, we define the critical value $c^{i*}_{\alpha,n}(\gamma)$ for $i = 1, 2$ as $c^{i*}_{\alpha,n}(\gamma) = \max(\tilde{c}^{i*}_{\alpha,n}(\gamma), \eta)$ for some small $\eta > 0$ so that $c^{i*}_{\alpha,n}(\gamma)$ for $i = 1, 2$ does not converge to zero in the limit.

Out test incorporates Hansen's one-sided test. The value γ determines how much our test is similar to solely performing one-sided test which uses $(1-\alpha)$- quantile of the empirical bootstrap distribution of $\{T^{1*}_{n,b}\}$ as its critical value. As the γ is close to 0, our test becomes similar to that of one-sided test. If we let $\gamma = 0$, then the test becomes exactly what Hansen proposed.

The way we obtain the critical values is different from that of Song [3] in that he computes the critical value $\tilde{c}^{1*}_{\alpha,n}(\gamma)$ from the bootstrap sample $\{T^{1*}_{b,n} \cdot 1(T^{2*}_{b,n} \leq c^{2*}_{\alpha,n}(\gamma))\}^B_{b=1}$. In that way, he avoids redundant counts. That is, two critical values should be determined to reject $B\alpha$ bootstrap observations. Yet, we define critical values $\tilde{c}^{1*}_{\alpha,n}(\gamma)$ and $\tilde{c}^{2*}_{\alpha,n}(\gamma)$ to be $(1-\alpha(1-\gamma))$-quantile and $(1-\alpha\gamma)$-quantile of the bootstrap empirical distribution of $T^{1*}_{n,b}$ and $T^{2*}_{n,b}$. There could be a bootstrap sample $\mathbf{d}^*_{b,t}$ which makes two bootstrap statistics greater than $(1-\alpha(1-\gamma))$-quantile and $(1-\alpha\gamma)$-quantile. As a result the critical values may be chosen to reject less than $B\alpha$ bootstrap samples, which leads to reducing power of the test. We leave this point to obtain the limiting distribution of $T^{1*}_{b,n} \cdot 1(T^{2*}_{b,n} \leq c^{2*}_{\alpha,n}(\gamma))$ and to justify computing critical values from the distribution as a future research.

4.4 Asymptotic Size and Power

With test statistics and critical values obtained in the previous subsections, we define the testing rule: *reject the null hypothesis* H_0 *if* $T_n^1 > c_{\alpha,n}^{1*}(\gamma)$ *or* $T_n^2 > c_{\alpha,n}^{2*}(\gamma)$. The following theorem shows that the newly defined test has the size less than or equal to the pre-specified significance level-α. That is, the test is a level-α test.

Theorem 4.1. *Suppose Assumption 4.1, 4.2, and 4.3 hold. Under the null hypothesis* $\mu \leq 0$*, the following results are true, given the significance level* $\alpha \in (0, 0.5)$*.*
(1) If $\mu_k = 0$ *for at least one* $k \in M$*, then* $\lim_{n\to\infty} P\{T_n^1 > c_{\alpha,n}^{1*}(\gamma)$ *or* $T_n^2 > c_{\alpha,n}^{2*}(\gamma)\} \leq \alpha$*. In particular, if there exist* k *and* k' *such that* $\mu_k = 0$ *and* $\mu_{k'} < 0$*, then it holds that* $\lim_{n\to\infty} P\{T_n^1 > c_{\alpha,n}^{1*}(\gamma)$ *or* $T_n^2 > c_{\alpha,n}^{2*}(\gamma)\} \leq P\{T_n^1 > c_{\alpha,n}^{1*}(0)\}$
(2) If $\mu_k < 0$ *for all* $k \in M$*, then* $\lim_{n\to\infty} P\{T_n^2 > c_{\alpha,n}^{2*}(\gamma)\} = 0$*.*

The first statement shows that the test has size α at maximum. Also it implies that the rejection probability always be less than that of solely performing one-sided test with $\gamma = 0$. This is because the two test statistics converges to the same quantity in the limit. The second statement holds because the both of test statistics converges to zero in probability while critical values converges to η in probability.

The next theorem shows that our test is consistent; its power against any fixed local alternatives converges to 1 in the limit. This result can be easily shown because it incorporates the one-sided test which is consistent against any fixed alternatives.

Theorem 4.2. *Let Assumption 4.1, 4.2, and 4.3 hold. Given the significance level* $\alpha \in (0, 0.5)$*, the test is consistent against any fixed alternative. i.e.* $\lim_{n\to\infty} P\{T_n^1 > c_{\alpha,n}^{1*}(\gamma)$ *or* $T_n^2 > c_{\alpha,n}^{2*}(\gamma)\} = 1$*.*

5 Extension to Stochastic Dominance Test

A stochastic dominance (henceforth SD) test has a close relationship with the SPA test. By replacing the population moment in the null hypothesis into the difference of two different CDFs and replacing the index set into the common support of those CDFs, we can convert the SPA test into a SD test. Given the two different CDFs F_X and F_Y and their support $\mathcal{Z}$, the hypotheses for the SD test are written as follows.

$$H_0 : F_Y(z) \leq F_X(z) \text{ for all } z \in \mathcal{Z}$$
$$H_1 : F_Y(z) > F_X(z) \text{ for some } z \in \mathcal{Z}$$

Song [3] mentioned that his hybrid test can be embedded to SD test in his article. However it can be shown that simply coupling two SD tests yields the same problem; it is not a level-α test. In this case, the one-sided test corresponds to the Barrett and Donald [6]'s test; they proposed a SD test with the same null hypothesis and with the critical value obtained from the centred bootstrap distribution. As we do in the previous sections, we construct a hybrid SD test by adopting re-centering method of Donald and Hsu [8] who extended the Hansen's method in the setting of the SD test.

Unlike the SPA test, we focus only on the case that the two samples are independent and each observation is independently drawn from the identical distribution to make the problem simpler. The following assumptions which we borrow from Donald and Hsu [8] formally describe the conditions with respect to the support and the sampling process.

Assumption 5.1. *Assume as follows.*
(1) The common support is given as a closed interval on the real line, $\mathcal{Z} = [0, \bar{z}]$ *where* $\bar{z} < \infty$.
(2) F_X *and* F_Y *are continuous proper distribution functions defined on the real line such that* $F_X(z) = F_Y(z) = 0$ *if and only if* $z = 0$*, and* $F_X(z) = F_Y(z) = 1$ *if and only if* $z = \bar{z}$.

The continuity of F_X and F_Y is generally assumed in the literature of SD test. The second statement is assumed to make the proof simple. This assumption can be released so that two CDFs F_X and F_Y may have different supports.

Assumption 5.2. *Assume the following:*
(1) $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^m$ are independent random samples from distributions with CDF's F_X and F_Y, respectively.
(2) The sample size of Y_i, m, is a function of the sample size of X_i, n, satisfying that $m(n) \to \infty$ and $n/(n+m) \to \lambda \in (0,1)$ when $n \to \infty$.

The second statement of Assumption 5.2. requires that two sample sizes, m and n, grow at the same diverging rate.

5.1 Test Statistics

Two CDFs are estimated by the empirical CDFs as follows:

$$\bar{F}_{X,n} = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq z) \text{ and } \bar{F}_{Y,m} = \frac{1}{m}\sum_{i=1}^{m} 1(Y_i \leq z).$$

Then two test statistics are defined as the following:

$$\begin{aligned} S_n^1 &= \sqrt{\frac{nm}{n+m}} \sup_{z\in\mathcal{Z}}(\bar{F}_{Y,m}(z) - \bar{F}_{X,n}(z)) \text{ and} \\ S_n^2 &= \sqrt{\frac{nm}{n+m}} \min(\sup_{z\in\mathcal{Z}}(\bar{F}_{Y,m}(z) - \bar{F}_{X,n}(z)), \sup_{z\in\mathcal{Z}}(\bar{F}_{X,n}(z) - \bar{F}_{Y,m}(z))). \end{aligned}$$

S_n^1 is the test statistic for one-sided test, and S_n^2 is the two-sided version of S_n^1. Linton, Maasoumi, and Whang [7] introduced this symmetrized test statistic in their SD test with a different null hypothesis.

Following notation of Donald and Hsu [8], let Ψ_{h_2} denote a mean zero Gaussian process with covariance kernel equal to $h_2 \in \mathcal{H}_2$ where $\mathcal{H}_2$ denotes the collection of all covariance kernels on $\mathcal{Z} \times \mathcal{Z}$. Let $h_2^{X,Y}$ denote the covariance kernel on $\mathcal{Z} \times \mathcal{Z}$ such that $h_2^{X,Y}(z_1, z_2) = \lambda F_X(z_1)(1 - F_X(z_2)) + (1-\lambda)F_Y(z_1)(1 - F_Y(z_2))$ for $z_1 \leq z_2$ and λ is in Assumption 5.2. It is well known that

$$\sqrt{\frac{mn}{m+n}}((\bar{F}_{Y,m}(\cdot) - \bar{F}_{X,n}(\cdot)) - (F_Y(\cdot) - F_X(\cdot))) \Rightarrow \Psi_{h_2^{X,Y}}(\cdot).$$

Define the contact set as $\mathcal{Z}^* = \{z \in \mathcal{Z} : F_X(z) = F_Y(z)\}$. Note that the contact set is not empty because $0, \bar{z} \in \mathcal{Z}^*$. Like the test statistics T^1 and T_n^2 in the previous section, the test statistics S^1 and S_n^2 are always non-negative. The following lemma provides the limiting distribution of the test statistics.

Lemma 5.1. *Suppose Assumption 5.1. and 5.2. hold.*
(1) Under the null hypothesis, we have

$$S_n^1 \Rightarrow \sup_{z \in \mathcal{Z}^*} \Psi_{h_2^{X,Y}}(z)$$

$$S_n^2 \Rightarrow \begin{cases} \min(\sup_{z \in \mathcal{Z}} \Psi_{h_2^{X,Y}}(z), \sup_{z \in \mathcal{Z}}(-\Psi_{h_2^{X,Y}}(z))) \ \textit{if } \mathcal{Z} = \mathcal{Z}^* \\ \sup_{z \in \mathcal{Z}^*} \Psi_{h_2^{X,Y}}(z) \ \textit{otherwise.} \end{cases}$$

In particular, if the contact set $\mathcal{Z}^*$ is empty, then both statistics converge to zero in probability.
(2) Under the alternative hypothesis, we have

$$S_n^1 \xrightarrow{p} \infty$$

$$S_n^2 \Rightarrow \begin{cases} \sup_{z \in \mathcal{Z}^*}(-\Psi_{h_2^{X,Y}}(z)) \ \textit{if } F_Y(z) \geq F_X(z) \textit{ for all } z \in \mathcal{Z} \\ \infty \ \textit{otherwise.} \end{cases}$$

It is worth paying attention to the fact that only the binding constraints, i.e. points on the contact set, matter for the asymptotic distribution as they do in the SPA test. Non-binding constraints do not affect the limiting distribution. This prevents the bootstrap procedure which pre-supposes that $F_Y = F_X$ at all points on the support from approximating the limiting distributions of the test statistics.

5.2 Re-centering Function

Motivated by Hansen [2]'s study, Donald and Hsu [8] applied the re-centering approach on SD test. The re-centering function is defined in the same way:

$$\hat{\mu}_n^c(z) = (\bar{F}_{Y,m}(z) - \bar{F}_{X,n}(z)) \cdot 1(\sqrt{n}(\bar{F}_{Y,m}(z) - \bar{F}_{X,n}(z)) < a_n).$$

The following lemma shows that the re-centering function converges to the population quantity $F_Y(z) - F_X(z)$ uniformly over the support $\mathcal{Z}$ in probability when $F_Y(z) - F_X(z) < 0$.

Lemma 5.2. *(Lemma 3.2 of Donald and Hsu [8]) Suppose Assumption 4.2, 5.1, and 5.2 hold. Then* $\sup_{z\in\mathcal{Z}} |\hat{\mu}_n^c(z) - \min(F_Y(z) - F_X(z), 0)| \xrightarrow{p} 0$.

The re-centering function plays the similar role in SD test. The following lemma shows that we can approximate the limiting distribution of the test statistics under the null hypothesis provided the random process which is centred by the re-centering function and has the covariance kernel function $\Psi_{h_2^{X,Y}}$.

Lemma 5.3. *Consider a random process* $Z_n(\cdot)$ *such that* $\sqrt{mn/(m+n)}(Z_n(\cdot) - \hat{\mu}_n^c(\cdot)) \Rightarrow \Psi_{h_2^{X,Y}}(\cdot)$ *on* $\mathcal{Z}$*. Define* $\mathcal{Z}^+ := \{z \in \mathcal{Z} : F_Y(z) \geq F_X(z)\}$*. Then it holds that*

$$\sqrt{\frac{mn}{m+n}} \sup_{z\in\mathcal{Z}} Z_n(z) \Rightarrow \sup_{z\in\mathcal{Z}^+} \Psi_{h_2^{X,Y}}(z)$$

$$\sqrt{\frac{mn}{m+n}} \min(\sup_{z\in\mathcal{Z}} Z_n(z), \sup_{z\in\mathcal{Z}}(-Z_n(z)))$$

$$\Rightarrow \begin{cases} \min(\sup_{z\in\mathcal{Z}} \Psi_{h_2^{X,Y}}(z), \sup_{z\in\mathcal{Z}}(-\Psi_{h_2^{X,Y}}(z))) \ \textit{if } \mathcal{Z}^+ = \mathcal{Z} \\ \sup_{z\in\mathcal{Z}^+} \Psi_{h_2^{X,Y}}(z) \ \textit{otherwise.} \end{cases}$$

5.3 Simulated Critical Value

In this section, we explain the way to simulate critical values for the hybrid SD test. We take the approach to bootstrap with separate samples. Other approaches such as multiplier bootstrap or bootstrap with combined samples are applicable. See Donald and Hsu [8], or Barrett and Donald [6] for more information.

To generate the bootstrap distribution, draw a random sample of size n from $\{X_1, \cdots, X_n\}$ to form $\bar{F}^*_{X,n,b}$ and a random sample of size m from $\{Y_1, \cdots, Y_m\}$ to form $\bar{F}^*_{Y,m,b}$ where they are the empirical CDFs based on the bootstrap sample. Define

$$Z_{n,b}^{*}(z) := (\bar{F}_{Y,m,b}^{*}(z) - \bar{F}_{Y,m}(z)) - (\bar{F}_{X,n,b}^{*}(z) - \bar{F}_{X,n}(z)) + \hat{\mu}_n^c(z).$$

Then this random process satisfies the following property.

Lemma 5.4. *Suppose Assumption 4.2, 5.1, and 5.2 hold. Then we have*

$$\sqrt{\frac{mn}{m+n}}(Z_{n,b}^{*}(\cdot) - \hat{\mu}_n^c(\cdot)) \overset{p}{\Rightarrow} \Psi_{h_2^{X,Y}}(\cdot).$$

Define

$$\begin{aligned} S_{n,b}^{1*} &= \sqrt{\frac{mn}{m+n}} \sup_{z\in\mathcal{Z}} Z_{n,b}^{*}(z), \text{ and} \\ S_{n,b}^{2*} &= \sqrt{\frac{mn}{m+n}} \min(\sup_{z\in\mathcal{Z}} Z_{n,b}^{*}(z), \sup_{z\in\mathcal{Z}}(-Z_{n,b}^{*}(z))). \end{aligned}$$

The lemma 5.3 and 5.4 together imply that we can simulate a random process which approximates the limiting processes of the test statistics under the null hypothesis in the sense that they weakly converge to the sample processes as the limiting process conditional on the sample path with probability approaching 1. That is, we have that under the null hypothesis,

$$\begin{aligned} S_{n,b}^{1*} &\overset{p}{\Rightarrow} \sup_{z\in\mathcal{Z}^*} \Psi_{h_2^{X,Y}}(z) \\ S_{n,b}^{2*} &\overset{p}{\Rightarrow} \begin{cases} \min(\sup_{z\in\mathcal{Z}} \Psi_{h_2^{X,Y}}(z), \sup_{z\in\mathcal{Z}}(-\Psi_{h_2^{X,Y}}(z))) \text{ if } \mathcal{Z}^* = \mathcal{Z} \\ \sup_{z\in\mathcal{Z}^*} \Psi_{h_2^{X,Y}}(z) \text{ otherwise.} \end{cases} \end{aligned}$$

We obtain critical value from this distribution. Define

$$\begin{aligned} \hat{c}_{\alpha,n}^{1*}(\gamma) &:= \inf\{c \in \mathcal{R} : P^*\{S_{n,b}^{1*} \leq c\} \geq 1 - \alpha(1-\gamma)\}, \text{ and} \\ \hat{c}_{\alpha,n}^{2*}(\gamma) &:= \inf\{c \in \mathcal{R} : P^*\{S_{n,b}^{2*} \leq c\} \geq 1 - \alpha\gamma\}. \end{aligned}$$

Again, in the case that all constraints are non-binding (i.e. $F_Y(z) < F_X(z)$ for all $z \in \mathcal{Z}$), both test statistics S_n^1 and S_n^2 converge to zero in probability. Since the bootstrap distribution also degenerates, $\hat{c}_{\alpha,n}^{1*}(\gamma)$ and $\hat{c}_{\alpha,n}^{2*}(\gamma)$ converge to zero. Hence we define the critical value $c_{\alpha,n}^{i*}(\gamma)$ for $i = 1, 2$ as $c_{\alpha,n}^{i*}(\gamma) = \max(\tilde{c}_{\alpha,n}^{i*}(\gamma), \eta)$ for some small $\eta > 0$.

The value γ determines how much our test is similar with solely performing one-sided test which uses $(1-\alpha)$- quantile of the empirical bootstrap distribution of $\{S_{n,b}^{1*}\}$ as a critical value. If we let $\gamma = 0$, then the test becomes exactly what Donald and Hsu [8] proposed.

5.4 Asymptotic Size and Power

In this section, we present the point-wise asymptotic size and power properties of our hybrid SD test. We reject the null hypothesis if $S_n^1 > c_{\alpha,n}^{1*}(\gamma)$ or $S_n^2 > c_{\alpha,n}^{2*}(\gamma)$. The following theorems show that rejection probability of our test does not exceed the significance level under the null hypothesis, and that our test is consistent against any fixed alternatives.

Theorem 5.1. *Suppose Assumption 4.2, 5.1 and 5.2 hold. Under the null hypothesis $F_Y(z) - F_X(z) \leq 0$ over $\mathcal{Z}$, the following results are true given the significance level $\alpha \in (0, 0.5)$.*
(1) If $F_Y(z) - F_X(z) = 0$ for some $z \in \mathcal{Z}$, then $\lim_{n\to\infty} P\{S_n^1 > c_{\alpha,n}^{1*}(\gamma)$ or $S_n^2 > c_{\alpha,n}^{2*}(\gamma)\} \leq \alpha$. In particular, if $\mathcal{Z}^* \neq \mathcal{Z}$, then it holds that $\lim_{n\to\infty} P\{S_n^1 > c_{\alpha,n}^{1*}(\gamma)$ or $S_n^2 > c_{\alpha,n}^{2*}(\gamma)\} \leq P\{S_n^1 > c_{\alpha,n}^{1*}(0)\}$.
(2) If $F_Y(z) - F_X(z) < 0$ for all $z \in \mathcal{Z}$, then $\lim_{n\to\infty} P\{S_n^2 > c_{\alpha,n}^{2*}(\gamma)\} = 0$.

Theorem 5.2. *Let Assumption 4.2, 5.1, and 5.2 hold. Given the significance level $\alpha \in (0, 1/2)$, the test is consistent against any fixed alternative. i.e. $\lim_{n\to\infty} P\{S_n^1 > c_{\alpha,n}^{1*}(\gamma)$ or $S_n^2 > c_{\alpha,n}^{2*}(\gamma)\} = 1$.*

6 Monte-Carlo Simulation

6.1 Hybrid SPA Test

In this section, we study the size and power properties of the test proposed in Section 4 given finite samples. We demonstrate that the rejection probability exceeds the significance level when we perform Song's hybrid test in case the number of alternatives is small. Next, we perform three different tests, a test of White [1] (Reality Check: RC), a test of Hansen [2](Superior Predictive Ability: SPA), and this article's proposal (Hybrid SPA), and compare their performances based on the designs that Hansen [2] considered, and the design that Song [3] considered.

The test of White [1] corresponds to the one-sided test of Song's with $\gamma = 0$. It rejects the null hypothesis if $U_n^1 > v_\alpha^{1*}(0)$. The critical value $v_\alpha^{2*}(0)$ is the $(1-\alpha)$-quantile of the bootstrap distribution based on $\{U_{b,n}^{1*}\}$. The test of Hansen [2] is the one-sided test of hybrid SPA test with $\gamma = 0$. It rejects the null hypothesis if $T_n^1 > c_\alpha^{1*}(0)$. Similarly, $c_\alpha^{1*}(0)$ is the $(1-\alpha)$-quantile of the bootstrap distribution based on $\{T_{b,n}^{1*}\}$. As Hansen suggested, we choose $a_n = \sqrt{2\log\log n}$. For the Hybrid SPA test, we choose $\eta = 10^{-6}$.

In all three parts, we consider the case with the sample size $n = 200$. The rejection probabilities that we report are based on 1,000 independent samples, where we use $q = 1$ in accordance with the lack of time dependence in $\mathbf{d}_t$, $t = 1, \cdots, n$. We compute critical values from 500 bootstrap resamples. The significance level-α is chosen to be 0.10. The parameter γ is set to be 0.5 in most cases.

6.1.1 Design With a Few Alternatives

We suppose the situation where we compare two or three forecasting models (i.e. $m = 2$ or $m = 3$). We draw $\mathbf{d}_t$ independently from the identical distribution $\sim N(\mu, \Sigma)$ where

$$\mu_2 = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \Sigma = \begin{pmatrix} 2.0811 & 1.0811 \\ 1.0811 & 2.0811 \end{pmatrix} \text{ if } m = 2, \text{ and}$$

$$\mu_3 = \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix}, \Sigma = \begin{pmatrix} 1.7208 & 0.7208 & 0.7208 \\ 0.7208 & 1.7208 & 0.7208 \\ 0.7208 & 0.7208 & 1.7208 \end{pmatrix} \text{ if } m = 3.$$

First we consider the case where the means are $\mu_2 = (0, -1)'$ and $\mu_3 = (0, -1, 0)'$. Since $\mu_2 \leq 0$ and $\mu_3 \leq 0$, the data generating processes (DGPs) correspond to the null hypothesis; in both cases, the benchmark forecasting model outperforms the second forecasting model while other models have similar level of performance.

We perform Song's test in the way we describe in Section 3 based on the bootstrap procedure introduced in Section 4 with $q = 1$. We report the simulation result in Table 2. We implement four difference tests at different significance levels, 0.05 and 0.1. Observe that Song's hybrid test rejects the null hypothesis with probability significantly higher than the pre-specified level, while the type I error of other tests remain below or close to the significance level. This simulation results confirm our finding that Song's test is not a level-α test.

m	α	RC	SPA	Hybrid SPA	Song
2	0.05	0.027	0.049	0.024	0.236
2	0.1	0.068	0.112	0.050	0.288
3	0.05	0.036	0.055	0.023	0.258
3	0.1	0.080	0.112	0.056	0.302

Table 2: Rejection Probabilities Under the Null

Next, we investigate how the power of three tests changes as the mean of the DGP moves. We consider the case with $m = 2$ where μ_x is either 0.05, 0.10, or 0.15, μ_y is a positive constant running in equally spaced grid in [-0.3, 0.3], and γ is a positive constant running in equally spaced grid in [0.1, 0.9]. Figure 3 and Table 3 present the simulation result.

It is noticeable that Hybrid SPA test outperforms two tests, RC and SPA, over the region $\mu_y < 0$ whichever the value μ_x takes. This occurs because we enlarge the rejection region towards two edges of null hypothesis $\{(\mu_x, \mu_y) : \mu_x < 0, \mu_y < 0, \mu_x \cdot \mu_y = 0\}$ in the two dimensional parameter space of (μ_x, μ_y) by incorporating two-sided test. In consequence, the rejection region near the origin retreats away from the origin. This explains the superior performance of two tests over Hybrid SPA over the region $\mu_y > 0$.

As the value γ decreases to zero, Hybrid SPA test becomes similar to the one-sided test, while it becomes similar to the two-sided test as the value γ increases. Another noticeable point from Figure 3 is that even the performance of Hybrid test with $\gamma = 0.1$ conspicuously outruns that of others over the region $\mu_y < 0$, while the performance remains competitive over the region $\mu_y > 0$. Though Song [3] suggested to use $\gamma = 0.5$, any choices between 0.1 and 0.5 seems to be proper to robustify the power without significant sacrifice in power over the region $\mu_y > 0$.

6.1.2 Hansen's Design

We generate losses of kth forecasting model $L_{k,t}$ independently from the identical distribution $N(\lambda_k/\sqrt{n}, \sigma_k^2)$ for $k = 1, \cdots, m$ and $t = 1, \cdots, n$ where the benchmark model has $\lambda_0 = 0$. Positive values ($\lambda_k > 0$) correspond to alternatives that are worse than the benchmark in that their expected losses are greater than those of benchmark. Negative values ($\lambda_k < 0$) correspond to alternatives that are better than the benchmark, and zero values ($\lambda_k = 0$) correspond to the alternatives that have the same performance with the benchmark. Note that $\lambda_k = 0$ represents that kth constraint is binding. Hansen defined the vectors of λ_k such as the following.

$$\lambda = \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{m-1} \\ \lambda_m \end{bmatrix} = \begin{bmatrix} 0 \\ \Lambda_1 \\ \frac{1}{m-1}\Lambda_0 \\ \vdots \\ \frac{m-2}{m-1}\Lambda_0 \\ \Lambda_0 \end{bmatrix}$$

He designed the performances of alternatives $k = 2, \cdots, m$ such that their mean values are spread evenly between 0 and Λ_0 so that Λ_0 determines the extent to that the inequalities are binding. Whether the DGP corresponds to the null hypothesis or not depends on the signs of Λ_0 and Λ_1. If Λ_0 and Λ_1 are non-negative, then the DGP conforms to the null hypothesis. The greater the values, Λ_0 and Λ_1, are, the further the DGP is away towards the interior of the null hypothesis from the boundary $\mu = 0$. If either Λ_0 or Λ_1 is negative, then the DGP conforms to the alternative hypothesis. The

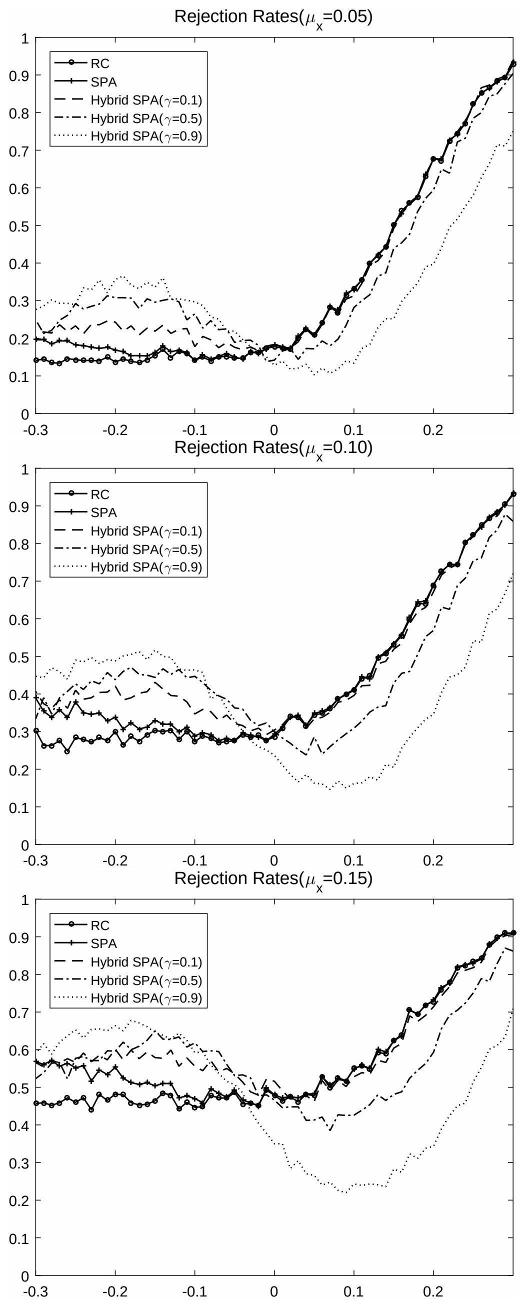

Figure 3: Rejection probabilities of three tests. The horizontal represents the value of μ_y.

μ_x	μ_y	RC	SPA	Hybrid SPA				
				γ =0.1	γ =0.3	γ =0.5	γ =0.7	γ =0.9
0.05	0.3	0.928	0.935	0.939	0.919	0.904	0.854	0.750
0.05	0.2	0.678	0.677	0.678	0.633	0.594	0.568	0.397
0.05	0.1	0.333	0.330	0.313	0.313	0.282	0.198	0.134
0.05	0	0.178	0.184	0.184	0.193	0.142	0.142	0.128
0.05	-0.1	0.142	0.141	0.178	0.241	0.265	0.268	0.297
0.05	-0.2	0.137	0.169	0.243	0.264	0.308	0.337	0.356
0.05	-0.3	0.141	0.196	0.214	0.193	0.249	0.273	0.276
0.10	0.3	0.932	0.934	0.930	0.911	0.859	0.844	0.720
0.10	0.2	0.690	0.690	0.672	0.612	0.569	0.501	0.346
0.10	0.1	0.410	0.412	0.396	0.341	0.309	0.253	0.162
0.10	0	0.286	0.293	0.304	0.291	0.299	0.270	0.237
0.10	-0.1	0.273	0.289	0.347	0.420	0.447	0.457	0.464
0.10	-0.2	0.300	0.338	0.421	0.430	0.438	0.491	0.488
0.10	-0.3	0.303	0.389	0.407	0.373	0.334	0.414	0.447
0.15	0.3	0.910	0.910	0.906	0.890	0.862	0.836	0.709
0.15	0.2	0.728	0.731	0.714	0.687	0.593	0.538	0.343
0.15	0.1	0.550	0.549	0.528	0.479	0.426	0.361	0.243
0.15	0	0.479	0.480	0.516	0.492	0.466	0.430	0.350
0.15	-0.1	0.447	0.470	0.556	0.600	0.618	0.627	0.592
0.15	-0.2	0.482	0.553	0.601	0.569	0.596	0.629	0.664
0.15	-0.3	0.457	0.568	0.566	0.526	0.523	0.563	0.590

Table 3: Rejection probabilities under the local alternatives with $\mu_x \in \{0.05, 0.10, 0.15\}$ and $\mu_y \in \{-0.3, -0.2, \cdots, 0.3\}$.

variance is designed to make the experiment more realistic;

$$\sigma_k^2 = \frac{1}{2}\exp(\arctan(\lambda_k))$$

so that a good forecasting model has a smaller variance than the poor model does.

Table 4 presents rejection probabilities under the null hypothesis. In the situation where all inequalities are binding ($\Lambda_0 = \Lambda_1$=0), the rejection probabilities of all three tests are close to the nominal level. The slight over-rejection of SPA test given $\Lambda_0 = \Lambda_1 = 0$ appears to be a small sample problem. In other cases, the rejection probabilities are all less than the nominal level.

Table 5 and 6 present rejection probabilities in different designs. Since those designs conform to the alternative hypothesis, rejection probabilities correspond to the power. In the design in Table 5, only one alternative (forecasting model 1) works better than the benchmark, while all alternatives are not worse thann the benchmark in the design in Table 6. Though Hansen [2] investigated power properties of RC and SPA under the design in Table 5 where RC performs poorly, we consider additional design in Table 6.

Rejection probabilities in the upper panel in Table 5 hardly exceeds the significance level. It can be shown that even the one-sided SPA test can be biased against local alternative because the test statistic $T_n^1 \Rightarrow \max_{z\in M}(Z_k/w_k \vee 0)$ where $Z \sim N(\lambda, \Omega)$ while $T_n^1 \overset{p}{\Rightarrow} \max_{z\in M}(Z_k/w_k \vee 0)$ where $Z \sim N(0, \Omega)$. In spite of the biasedness, we can still compare the power properties of three tests. In Table 5, Hybrid SPA always outperforms RC, but SPA outperforms Hybrid SPA. Similarly in Table 6, the performance of Hybrid SPA outruns that of RC except four cases that $\Lambda_0 = -0.5$ which is close to the null hypothesis. It is worth paying attention to that Hybrid SPA outperforms SPA provided that $\Lambda_0 \leq -2.0$. Since the design in Table 5 is closer to the null hypothesis than the other, rejection probabilities in Table 6 are higher than those in Table 5 overall.

Λ_1	Λ_0	m=50			m=100		
		RC	SPA	Hybrid SPA	RC	SPA	Hybrid SPA
0	0	0.093	0.101	0.094	0.093	0.105	0.069
0	0.5	0.064	0.080	0.044	0.054	0.068	0.038
0	1.0	0.025	0.047	0.017	0.023	0.047	0.026
0	1.5	0.022	0.038	0.032	0.009	0.033	0.016
0	2.0	0.009	0.025	0.021	0.002	0.022	0.010
0	2.5	0.004	0.019	0.016	0.007	0.021	0.011
0	3.0	0.002	0.013	0.008	0.001	0.010	0.010

Table 4: Rejection Probabilities Under H_0 (Hansen's Design)

6.1.3 Song's Design

Not only did Song [3] investigate the power properties of SPA in Hansen[2]'s design, but also he presented the design where his test outperforms RC and SPA. Following Song [3], we consider the following alternative scheme: for each $k = 1, \cdots, m$, $\lambda_k = u \times \{\Phi(-8k/m + v) - 1/2\}$ where u is a positive constant running in equally spaced grid in [1,5], v is a constant whose value is either 0.4, 0.6, or 0.8, and Φ is a standard normal distribution function. As the value v increases, the portion of alternatives which perform better than the benchmark increases. The value u controls the extent to that the constraints are binding. Figure 4 presents this alternative designs at $u = 1, 3, 5$ and $v = 0.4, 0.8$ with $m = 50$.

Figure 5 and Table 7 show the rejection probabilities under Song's designs. Unlike the result reported in Song [3], the performance of our hybrid SPA test does not dominate that of SPA. However, rejection probabilities of Hybrid SPA are higher than those of SPA for large u values in all cases. Also the extent that Hybrid SPA outperforms RC is noticeable.

Λ_1	Λ_0	m=50			m=100		
		RC	SPA	Hybrid SPA	RC	SPA	Hybrid SPA
-1	0.5	0.058	0.107	0.074	0.061	0.095	0.062
-1	1.0	0.030	0.092	0.062	0.028	0.069	0.038
-1	1.5	0.014	0.089	0.056	0.014	0.061	0.043
-1	2.0	0.008	0.079	0.055	0.010	0.053	0.037
-1	2.5	0.005	0.107	0.076	0.002	0.053	0.030
-1	3.0	0.002	0.061	0.046	0.002	0.048	0.035
-2	0.5	0.155	0.409	0.343	0.117	0.365	0.292
-2	1.0	0.105	0.432	0.367	0.041	0.328	0.261
-2	1.5	0.060	0.449	0.395	0.021	0.348	0.293
-2	2.0	0.034	0.418	0.349	0.022	0.362	0.308
-2	2.5	0.026	0.420	0.360	0.008	0.344	0.291
-2	3.0	0.017	0.395	0.338	0.007	0.325	0.257

Table 5: Rejection Probabilities Under H_1 (Hansen's Design)

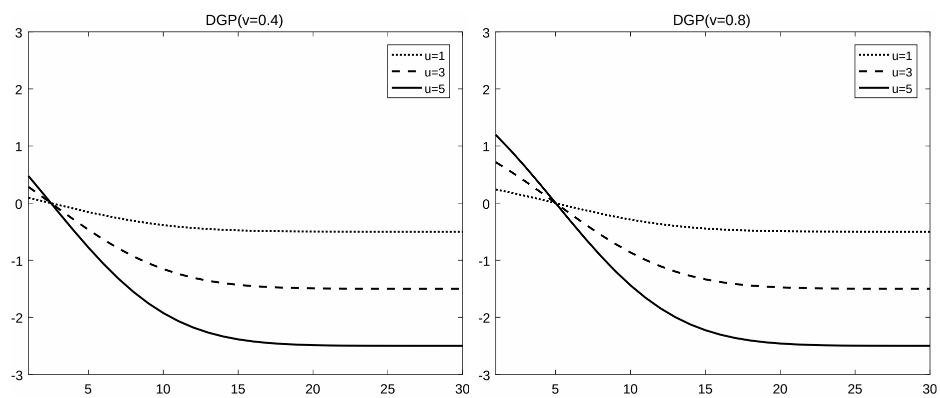

Figure 4: Designs of λ_k with $m = 50$. The horizontal axis represents the number of alternatives m while each graph represents the value of λ_k according to the value of u given $v = 0.4$(left) and $v = 0.8$(right).

Λ_1	Λ_0	m=50			m=100		
		RC	SPA	Hybrid SPA	RC	SPA	Hybrid SPA
0	-0.5	0.191	0.203	0.168	0.162	0.182	0.132
0	-1.0	0.276	0.339	0.320	0.279	0.356	0.290
0	-1.5	0.465	0.574	0.566	0.434	0.571	0.529
0	-2.0	0.674	0.780	0.807	0.645	0.757	0.784
0	-2.5	0.874	0.935	0.956	0.844	0.917	0.919
0	-3.0	0.952	0.976	0.993	0.951	0.978	0.988
-1	-0.5	0.176	0.212	0.162	0.174	0.205	0.156
-1	-1.0	0.270	0.334	0.321	0.270	0.338	0.292
-1	-1.5	0.446	0.542	0.536	0.433	0.545	0.521
-1	-2.0	0.689	0.784	0.799	0.653	0.764	0.771
-1	-2.5	0.850	0.912	0.949	0.849	0.910	0.936
-1	-3.0	0.961	0.976	0.983	0.961	0.982	0.988
-2	-0.5	0.322	0.469	0.437	0.315	0.463	0.395
-2	-1.0	0.410	0.533	0.533	0.372	0.490	0.444
-2	-1.5	0.531	0.642	0.652	0.494	0.602	0.585
-2	-2.0	0.679	0.780	0.796	0.675	0.782	0.782
-2	-2.5	0.869	0.917	0.932	0.844	0.912	0.930
-2	-3.0	0.959	0.979	0.985	0.941	0.970	0.986

Table 6: Rejection Probabilities Under H_1 (Hansen's Design)

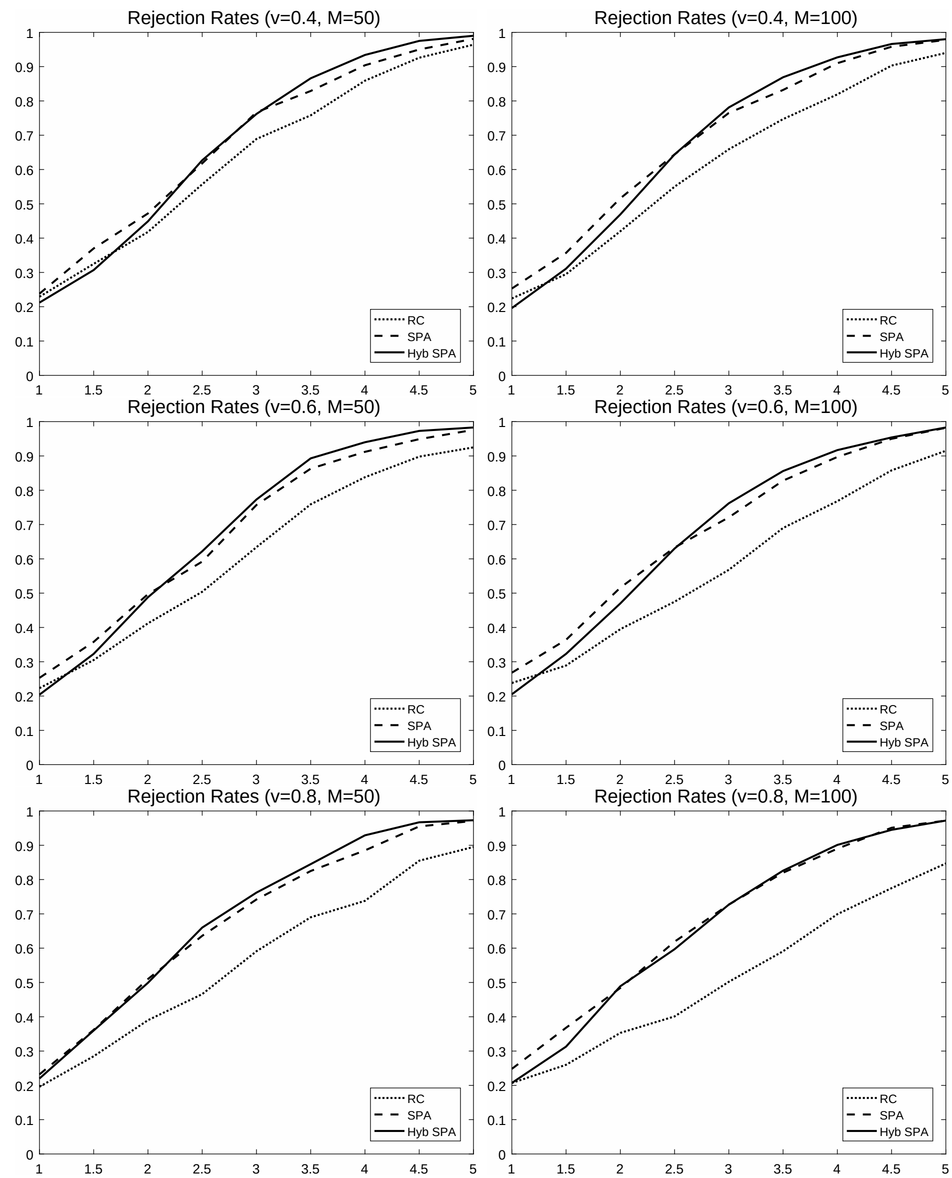

Figure 5: Rejection Probabilities Under H_1 (Song's Design). The horizontal axis represents the values of u. We report the exact quantities in Table 8.

u	v	m=50			m=100		
		RC	SPA	Hybrid SPA	RC	SPA	Hybrid SPA
1	0.4	0.229	0.238	0.212	0.224	0.253	0.196
1.5	0.4	0.325	0.370	0.307	0.295	0.357	0.311
2	0.4	0.418	0.472	0.449	0.42	0.516	0.469
2.5	0.4	0.557	0.618	0.627	0.55	0.644	0.644
3	0.4	0.689	0.767	0.762	0.659	0.765	0.781
3.5	0.4	0.758	0.829	0.866	0.747	0.832	0.869
4	0.4	0.859	0.904	0.934	0.819	0.91	0.927
4.5	0.4	0.926	0.950	0.975	0.903	0.958	0.966
5	0.4	0.964	0.981	0.990	0.940	0.978	0.98
1	0.6	0.223	0.253	0.204	0.238	0.268	0.205
1.5	0.6	0.305	0.358	0.323	0.289	0.364	0.323
2	0.6	0.412	0.497	0.488	0.395	0.516	0.47
2.5	0.6	0.504	0.592	0.622	0.475	0.633	0.63
3	0.6	0.634	0.757	0.773	0.568	0.721	0.762
3.5	0.6	0.759	0.863	0.893	0.690	0.828	0.856
4	0.6	0.838	0.912	0.940	0.768	0.897	0.917
4.5	0.6	0.898	0.949	0.973	0.858	0.95	0.954
5	0.6	0.925	0.976	0.983	0.915	0.982	0.983
1	0.8	0.196	0.232	0.220	0.207	0.248	0.207
1.5	0.8	0.285	0.362	0.360	0.260	0.368	0.313
2	0.8	0.390	0.510	0.499	0.353	0.484	0.489
2.5	0.8	0.466	0.636	0.660	0.401	0.619	0.597
3	0.8	0.591	0.742	0.762	0.502	0.727	0.727
3.5	0.8	0.690	0.825	0.845	0.591	0.820	0.826
4	0.8	0.738	0.885	0.929	0.699	0.890	0.901
4.5	0.8	0.855	0.955	0.967	0.775	0.951	0.945
5	0.8	0.895	0.971	0.973	0.847	0.972	0.972

Table 7: Rejection Probabilities Under H_1 (Song's Design)

6.2 Hybrid SD Test

In this section, we study the size and power properties of the test proposed in Section 5 given finite samples. We consider two designs that Donald and Hsu[8] considered, and compare three different tests, a test of Barrett and Donald [6] (BD), a test of Donald and Hsu [8] (DH), and this article's proposal (Hybrid SD).

We consider the case with the sample size $n = m = 200$. The rejection probabilities that we report are based on 500 independent samples. We compute critical values from 500 bootstrap resamples. The significance level-α is chosen as 0.10. As Song [3] suggested, we choose $\gamma = 0.5$ for critical variable. As Donald and Hsu[8] suggested, we choose $a_n = -0.1\sqrt{\log\log 2n}$ to compute the re-centering function. For the Hybrid SD test, we choose $\eta = 10^{-6}$.

6.2.1 Null Design

We examine the size of multiple tests including Hybrid SD. In addition, to show that naive application of coupling into SD test yields an invalid test, we perform the hybrid SD test based on the Barrett and Donald test, say Hybrid BD. The tests statistics are the same with those in Hybrid SD but we obtain critical values from the centred bootstrap distribution not from the re-centred bootstrap distribution. The critical value from the one-sided test is defined by the $(1-\alpha(1-\gamma))-$quantile of

$$\sqrt{\frac{mn}{m+n}}\sup_{z\in\mathcal{Z}}(Z^*_{n,b}(z)-\hat{\mu}^c_n(z))$$

and their two-sided critical value is defined by $(1-\alpha\gamma)$-quantile of the empirical distribution based on

$$\sqrt{\frac{mn}{m+n}}\min(\sup_{z\in\mathcal{Z}}(Z^*_{n,b}(z)-\hat{\mu}^c_n(z)),\sup_{z\in\mathcal{Z}}(-Z^*_{n,b}(z)+\hat{\mu}^c_n(z))).$$

For a given constant $L \in (0, 1]$ two samples are generated as follows:

$$X = 1(U_1 \leq L)\frac{U_1^2}{L} + 1(U_1 > L)U_1$$
$$Y = U_2$$

where U_1 and U_2 are independent random variables following the uniform distribution over [0,1]. If $L = 0$, we let $X = U_1$. Then for a specified value of L, the CDFs are given as

$$F_X(z) = \begin{cases} \sqrt{zL} \text{ if } z \in [0, L], \\ z \text{ if } z \in (L, 1] \end{cases}$$
$$F_Y(z) = z \text{ for } z \in [0, 1].$$

Note that $F_Y(z) < F_X(z)$ for $z \in (0, L)$, and $F_Y(z) = F_X(z)$ for $z = 0, L$, and 1. The design conforms to the null hypothesis and the rejection probabilities correspond to the type I errors. As the value L increases that the design is further away from the least favourable case, $F_Y(z) = F_X(z)$ for all $z \in \mathcal{Z}$.

The following Table 8 shows the rejection probabilities under the this design. It is noticeable that Hybrid BD test noticeably rejects the null hypothesis at more than the nominal level. Rejection rates of three other tests stay below the nominal level. BD which is based on the least favourable case tends to be conservative for $L \geq 0.4$, while the rejection rate of Hybrid SD lie between those of BD and DH. As the value of L increases, the overall rejection probabilities decrease. Overall, the application of coupling on the stochastic dominance test seems not to bring a considerable improvement in power.

6.2.2 Alternative Design

For a given $L \in (0, 1)$ the two samples are generated by

$$X = 1(U_1 \leq L)\frac{U_1^2}{L} + 1(U_1 > L)U_1$$
$$Y = 1(U_2 \leq L)U_2 + 1(U_2 > L)(L + \frac{(U_2 - L)^2}{1 - L})$$

L	BD	DH	Hybrid SD	Hybrid BD
0.0	0.098	0.102	0.044	0.086
0.2	0.098	0.100	0.060	0.170
0.4	0.076	0.078	0.080	0.286
0.6	0.052	0.084	0.056	0.242
0.8	0.010	0.044	0.016	0.124

Table 8: Rejection Probabilities Under H_0

L	BD	DH	Hybrid SD
0.5	0.864	0.914	0.874
0.6	0.666	0.812	0.708
0.7	0.384	0.688	0.518
0.8	0.094	0.474	0.266
0.9	0.002	0.184	0.052

Table 9: Rejection Probabilities Under H_1

where U_1 and U_2 are independent uniformly distributed random variables over [0,1]. The CDF of X and Y are given as

$$F_X(z) = \begin{cases} \sqrt{zL} \text{ if } z \in [0, L], \\ z \text{ if } z \in (L, 1] \end{cases}$$

$$F_Y(z) = \begin{cases} z \text{ if } z \in [0, L], \\ L + \sqrt{(z-L)(1-L)} \text{ if } z \in (L, 1]. \end{cases}$$

We have $F_Y(z) > F_X(z)$ for all $z \in (L, 1)$ and $F_Y(z) < F_X(z)$ for $z \in (0, L)$. That is, as the value of L increases, the design becomes closer to the null hypothesis. Thus the rejection probabilities correspond to power. Table 9 presents the rejection probabilities under the this design. As the value of L increases, the power considerably decreases in all three tests, and the power of BD and Hybrid SD ends up being smaller than the nominal level at $L = 0.9$. While Hybrid BD outperforms BD, the performance of DH outruns that of Hybrid SD.

7 Conclusion

Motivated by that Song's idea to couple one-sided and two-sided tests improves the power properties against local alternatives and by that his test is not a level-α test, we have investigated the way to make his test to be a level-α test in the context of comparing multiple forecasting models given a benchmark and in the context of setting first-order stochastic dominance relation given two distributions. To control the size of the hybrid test, we adopted to re-centre the bootstrap distribution in computing the critical value; we took the re-centering approach of Hansen[2] in SPA test, and of Donald and Hsu [8] in SD test. In consequence, we have shown that our proposals, hybrid SPA test and hybrid SD test are level-α tests and they are consistent against any fixed alternative.

The simulation study results are consistent with our theoretical result. Coupled tests which are based on the least favourable case rejects the null hypothesis with probability conspicuously higher than the nominal level even in the finite sample in both SPA test and SD test. There are cases that the hybrid SPA test outperforms existing tests including reality check of White[1] and SPA test of Hansen [2], and its power is overall similar to that of SPA test. Moreover in most cases, the performance of the hybrid SPA test outruns that of the reality check. Meanwhile, it turns out to be that coupling does not bring power improvement in stochastic dominance testing.

Though our hybrid SPA test outperforms existing tests in certain cases, it has limitation; the size is controlled in the sense of pointwise asymptotics. To guarantee that the size is smaller than the nominal level for any cases under the null hypothesis given the sample of fixed size, we needs uniformity result. Moreover, Our test does not permit the comparison of parameterized models when a recursive scheme is used to estimate the parameters according to Hansen [2]. It would make this test more practical if the test accommodating this point is constructed. In addition, it could be interesting to extend our hybrid SD test to allow high-order stochastic dominances or to permit dependency in the data.

8 Appendix

In this section, we provide proofs for Lemmas and Theorems introduced in the previous sections. Define $\psi : \mathcal{R}^m \to \mathcal{R}_+^m$ as ith element of $\psi(x_1, \cdots, x_m)$ is $x_i \vee 0$.

Lemma A 1. *Two statistics T_n^1 and T_n^2 are continuous functions from $\mathcal{R}^m$ to $\mathcal{R}$.*

Proof of Lemma A 1. Since all components of ψ is continuous, ψ is a continuous function. Next, define φ be a function from $\mathcal{R}_+^m$ to $\mathcal{R}$ such that $\varphi(x_1, \cdots, x_m) = \max_{k=1,\cdots,m}(x_k)$. Then φ is Lipschitz continuous. To show this, for two vectors $x = (x_1, \cdots, x_m)$ and $y = (y_1, \cdots, y_m)$, without loss of generality suppose that $\max_k x_k - \max_k y_k \geq 0$. And let k^* be the number satisfying $\max_k x_k = x_{k^*}$. Then it holds that

$$
\begin{aligned}
0 \leq \max_k x_k - \max_k y_k = x_{k^*} - \max_k y_k \\
&\leq x_{k^*} - y_{k^*} \\
&\leq \max_k (x_k - y_k) \\
&\leq \max_k |x_k - y_k| \\
&\leq \sqrt{(x_1 - y_1)^2 + \cdots + (x_m - y_m)^2} = ||x - y||.
\end{aligned}
$$

Because $T_n^1 = \sqrt{n}\varphi(\psi(\bar{d}_{n,1}/\bar{w}_{n,1}, \cdots, \bar{d}_{n,m}/\bar{w}_{n,m}))$ is a composite function of ψ and φ, T_n^1 is continuous. Similarly, we can show that T_n^2 is a continuous function.□

Lemma A 2. *Two statistics S_n^1 and S_n^2 are continuous functions from a function space $(\mathcal{F}, ||\cdot||_\infty)$ to $(\mathcal{R}, |\cdot|)$ where $\mathcal{F}$ is a set of bounded functions defined on $\mathcal{Z}$.*

Proof of Lemma A 2. Consider arbitrary two functions f and g in $\mathcal{F}$. Without loss of generality, we assume $\sup f \geq \sup g$. Then $0 \leq \sup f - \sup g = \sup(f - g + g) - \sup g \leq \sup(f - g) + \sup g - \sup g \leq ||f - g||_\infty$.

Thus S_n^1 is continuous, and the continuity of S_n^2 can be shown similarly. $\square$

Proof for Lemma 4.1. The proof is similar to that of Lemma 1 in Hansen . Let $A_{n,k} := \sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} \cdot 1(\mu_k = 0)$ and $B_{n,k} := \sqrt{n}\bar{d}_{n,k}/\bar{w}_k \cdot 1(\mu_k < 0)$, and let $A_n := (A_{n,1}, \cdots, A_{n,m})$ and $B_n := (B_{n,1}, \cdots, B_{n,m})$. By assumption 1, it holds that $\sqrt{n}(\bar{\mathbf{d}} - \mu) \Rightarrow N(0, \Sigma)$.

Consider the situation that $\mu \leq 0$. $\psi(A_n) \Rightarrow \psi(Z_k/w_k)$ where $Z \sim N(0, \Sigma^0)$ by continuous mapping lemma(henceforth CMT). If $\mu_k < 0$, then $\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} \xrightarrow{p} -\infty$. It's because $\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} = \sqrt{n}(\bar{d}_{n,k} - \mu_k)/\bar{w}_k + \sqrt{n}\mu_k/\bar{w}_k$. The first term is stochastically bounded while the second term diverges into negative infinity in probability. Accordingly by the continuous mapping lemma, $\psi(B_n) \xrightarrow{p} 0$. Hence, we have

$$\begin{aligned}
T_n^1 &= \sqrt{n}\max_{k\in M}(\frac{\bar{d}_{n,k}}{\bar{w}_k} \vee 0) \\
&= \max_{k\in M}\psi(A_{n,k} + B_{n,k}) \\
&= \max_{k\in M}(\psi(A_{n,k}) + \psi(B_{n,k})) \\
&= \max_{k\in M}(\psi(A_{n,k}) + o_p(1)) \\
&\Rightarrow \max_{k\in M}\psi(Z_k/w_k) \stackrel{d}{=} \max_{k\in M^*}(Z_k/w_k \vee 0).
\end{aligned}$$

The weak convergence holds by the Slutsky's lemma and by the CMT. Note that if $\mu_k < 0$ for all $k \in M$, then $T_n^1 \xrightarrow{p} 0$. Watch out the abuse of notation. Though we represent $\max(x_1, \cdots, x_m) = \max_{k\in M} x_k$, the argument of $\max_{k\in M}$ is a m-dimensional vector.

When $\mu = 0$, it holds that

$$T_n^2 \Rightarrow \min(\max_{k\in M}(Z_k/w_k \vee 0) \wedge \max_{k\in M}(-Z_k/w_k \vee 0))$$

by Lemma A1, the CMT and Slutksy's lemma. If there exists k such that $\mu_k < 0$, then $-\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} = -\sqrt{n}(\bar{d}_{n,k} - \mu_k)/\bar{w}_k - \sqrt{n}\mu_k/\bar{w}_k \xrightarrow{p} \infty$. Thus we have $-\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} \leq T_n^{1,-} = \max_{k\in M}(-\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} \vee 0) \xrightarrow{p} \infty$. Since $P\{\min(T_n^1, T_n^{1,-}) \leq x\} \leq P\{T_n^1 \leq x\} + P\{T_n^{1,-} \leq x\}$ for any $x \in R$, it holds that $0 \leq P\{\min(T_n^1, T_n^{1,-}) \leq x\} - P\{T_n^1 \leq x\} \leq P\{T_n^{1,-} \leq x\} \to 0$ as

$n \to \infty$. Hence $T_n^2 = \min(T_n^1, T_n^{1,-})$ weakly converges to the same quantity that T_n^1 converges into. The remaining part of Lemma can be shown in a similar way. □

Proof for Lemma 4.2. By assumption 1, $\bar{d}_n - \mu \xrightarrow{p} 0$. The result can be shown in the same way of showing Lemma 3.2 in Donald and Hsu [8]. □

Proof for Lemma 4.3. As we do in the proof of Lemma 1, we partition the random variable $Z_{n,k}$ into a sum of two different random variables $Z_{n,k}1(\mu_k < 0)$ and $Z_{n,k}1(\mu_k \geq 0)$ for $k = 1, \cdots, m$. And we stack these variables into two vectors, $Z_n^- = (Z_{n,1}1(\mu_1 < 0), \cdots, Z_{n,m}1(\mu_m < 0))'$ and $Z_n^+ = (Z_{n,1}1(\mu_1 \geq 0), \cdots, Z_{n,m}1(\mu_m \geq 0))'$. We can prove this lemma with the same logic used in proving Lemma 1. Since $\sqrt{n}\max_{k\in M}\psi(Z_{n,k}) = \sqrt{n}\max_{k\in M}\psi(Z_{n,k}^- + Z_{n,k}^+) = \sqrt{n}\max_{k\in M}(\psi(Z_{n,k}^-) + \psi(Z_{n,k}^+))$, it suffices to show that $\psi(Z_{n,k}^-) = o_p(1)$ and $\sqrt{n}\psi(Z_{n,k}^+)$ has the same limiting distribution with $\sqrt{n}\psi(Z_{n,k}^+ - \hat{\mu}_k^c)$.

If $\mu_k < 0$, then

$$\begin{aligned}\sqrt{n}Z_{n,k}^- = \sqrt{n}(Z_{n,k}^- - \hat{\mu}_k^c) \\ &+ \sqrt{n}(\hat{\mu}_k^c - \min(\mu_k, 0)1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < -\sqrt{2\log\log n})) \\ &+ \sqrt{n}\min(\mu_k, 0)1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < -\sqrt{2\log\log n}).\end{aligned}$$

The first term is stochastically bounded by assumption. The second term equals to $\sqrt{n}(\bar{d}_{n,k} - \mu_k)1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < -\sqrt{2\log\log n}) = O_p(1)\cdot(1 + o_p(1)) = O_p(1)$. We use the fact that $1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < -\sqrt{2\log\log n}) \xrightarrow{p} 1$, which can be deduced by Lemma 1. The third term converges to minus infinity because $-\sqrt{n}\mu_k 1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < -\sqrt{2\log\log n}) \geq -\sqrt{n}\mu_k \to \infty$ as $n \to \infty$. Multiplying -1 onto both sides of inequality yields the desired result. Thus the sum of three terms ends up converging into negative infinity in probability. Accordingly, the kth element does not affect the asymptotic distribution of $\sqrt{n}\max_{k\in M}(Z_{n,k} \vee 0)$ because $\psi(Z_{n,k}^-) = o_p(1)$.

If $\mu_k \geq 0$, then $\sqrt{n}Z_{n,k}^+ = \sqrt{n}(Z_{n,k}^+ - \hat{\mu}_k^c) + \sqrt{n}\hat{\mu}_k^c = \sqrt{n}(Z_{n,k} - \hat{\mu}_k^c) + o_p(1)$. If $\mu_k = 0$, then it is because $\sqrt{n}\hat{\mu}_k^c = \sqrt{n}\bar{d}_{n,k}1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} <$

$-\sqrt{2\log\log n}) = O_p(1)\cdot o_p(1) = o_p(1)$. If $\mu_k > 0$, then it is because

$$\begin{aligned}\sqrt{n}\hat{\mu}_k^c =& \sqrt{n}(\bar{d}_{n,k}-\mu_k)1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < -\sqrt{2\log\log n})\\ &+ \sqrt{n}\mu_k 1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < -\sqrt{2\log\log n})\\ =& O_p(1)o_p(1) + o_p(1)\\ =& o_p(1).\end{aligned}$$

Note $\sqrt{n}\mu_k 1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < -\sqrt{2\log\log n}) = o_p(1)$ holds because for any $\delta > 0$, $P\{\sqrt{n}\mu_k 1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < -\sqrt{2\log\log n}) < \delta\} = P\{1(\sqrt{n}\bar{d}_{n,k}/\bar{w}_{n,k} < -\sqrt{2\log\log n}) = 0\}$. Consequently, $\sqrt{n}\psi(Z_{n,k}^+)$ has the same limiting distribution with $\sqrt{n}\psi(Z_{n,k}^+ - \hat{\mu}_k^c)$. The rest of the proof is the same with that of Lemma 1. □

Proof for Lemma 4.4. This follows from Theorem 3 of Hansen which follows trivially from work of Goncalves and de Jong ([10], lemma 2). □

Proof for Theorem 4.1. This proof is similar to that of Lemma A.2. of Donald and Hsu [8]. By Lemma 4.4, it holds that $\sqrt{n}(\bar{Z}_b^* - \hat{\mu}^c) \overset{p}{\Rightarrow} N(0,\Sigma)$ where $\overset{p}{\Rightarrow}$ represents the weak convergence conditional on the sample path with probability approaching to 1. Thus for each subsequence n_p of n there exists a further subsequence n_{p_q} such that $\sqrt{n_{p_q}}(\bar{Z}_{n_{p_q},b}^* - \hat{\mu}^c) \overset{a.s.}{\Rightarrow} N(0,\Sigma)$. There exist a subset Ω_0 of a sample space Ω such that $P\{\Omega_0\}$=1, and for $\omega \in \Omega_0$, $T_{n_{p_q},b}^{1*}(\omega) \Rightarrow \max_{k\in M^+}(Z_k^+/w_k \vee 0)$ by Lemma 4.3. In particular, if $M^+ = M$, then $T_{n_{p_q},b}^{2*}(\omega) \Rightarrow \min(\max_{k\in M^+}(Z_k^+/w_k \vee 0), \max_{k\in M^+}(-Z_k^+/w_k \vee 0))$ on Ω_0, and if $M^+ \neq M$, then $T_{n_{p_q},b}^{2*}(\omega) \Rightarrow \max_{k\in M^+}(Z_k^+/w_k \vee 0)$ on Ω_0.

Since it holds that $P\{T_n^1 > c_{\alpha,n}^{1*}(\gamma)$ or $T_n^2 > c_{\alpha,n}^{2*}(\gamma)\} \leq P\{T_n^1 > c_{\alpha,n}^{1*}(\gamma)\} + P\{T_n^2 > c_{\alpha,n}^{2*}(\gamma)\}$, we will show each of terms does not exceed $\alpha(1-\gamma)$ and $\alpha\gamma$. Let us consider the first term, the rejection probability of the one-sided test.

If $M^+ \neq \emptyset$, then $\max_{k\in M^+}(Z_k^+/w_k \vee 0)$ does not degenerate. Since Z is a random vector following joint normal distribution, the CDF of $\max_{k\in M^+}(Z_k^+/w_k \vee 0)$ may have a discontinuity point at zero, but it is continuous and strictly increasing on $(0,\infty)$. The value of CDF at zero is less than or equal to 1/2 because $P\{\max_{k\in M^+}(Z_k^+/w_k \vee 0) \le 0\} = P\{Z_k \le 0 \text{ for all } k \in M^+\} \le 1/2$. Since the $\alpha(1-\gamma) < 1/2$, the $\tilde{c}^{1*}_{\alpha,n_{pq}}(\gamma)$ converges to $1-\alpha(1-\gamma)$-quantile of the CDF of $\max_{k\in M^+}(Z_k^+ \vee 0)$ by Lemma 21.2. in Van der vaart [11] on Ω_0. Thus it holds that $\tilde{c}^{1*}_{\alpha,n}(\gamma) \xrightarrow{p} c$, and that $c^{1*}_{\alpha,n}(\gamma) \xrightarrow{p} c$ because $c > \eta > 0$ for sufficiently small η. Hence $\lim_{n\to\infty} P\{T_n^1 > c^{1*}_{\alpha,n}(\gamma)\} = \alpha(1-\gamma)$. If $M^+ = \emptyset$, then $\max_{k\in M^+}(Z_k^+/w_k \vee 0)$ degenerates. In this case, the $\tilde{c}^{1*}_{\alpha,n_{pq}}(\gamma)$ converges to 0 almost surely, $\tilde{c}^{1*}_{\alpha,n}(\gamma)$ converges to 0 in probability, and consequently $c^{1*}_{\alpha,n}(\gamma)$ converges to η in probability. Since $T_n^1 \xrightarrow{p} 0$, it holds that $\lim_{n\to\infty} P\{T_n^1 > c^{1*}_{\alpha,n}(\gamma)\} = 0$.

Next we consider the rejection probability of the two-sided test. If $M^+ = M$, then we have $T^{2*}_{n_{pq},b} \Rightarrow \min(\max_{k\in M}(Z_k^+/w_k \vee 0), \max_{k\in M}(-Z_k^+/w_k \vee 0))$ on Ω_0. Similarly, the CDF of $\min(\max_{k\in M}(Z_k^+/w_k \vee 0), \max_{k\in M}(-Z_k^+/w_k \vee 0))$ may have a discontinuity point at zero, but it is continuous and strictly increasing on $(0,\infty)$. Denote the CDF of $\min(\max_{k\in M}(Z_k^+/w_k \vee 0), \max_{k\in M} (-Z_k^+/w_k \vee 0))$ as F_2. If $F_2(0) < 1-\alpha\gamma$, then the $\tilde{c}^{2*}_{\alpha,n_{pq}}(\gamma)$ converges to $(1-\alpha\gamma)-$quantile of F_2, say c, almost surely, $\tilde{c}^{1*}_{\alpha,n}(\gamma) \xrightarrow{p} c$, and consequently $c^{1*}_{\alpha,n}(\gamma) \xrightarrow{p} c$. Hence we have $\lim_{n\to\infty} P\{T_n^2 > c^{2*}_{\alpha,n}(\gamma)\} = \alpha\gamma$. If $F_2(0) \ge 1-\alpha\gamma$, then $\tilde{c}^{2*}_{\alpha,n_{pq}}(\gamma) \xrightarrow{a.s.} 0$, $\tilde{c}^{1*}_{\alpha,n}(\gamma) \xrightarrow{p} 0$, and consequently $c^{1*}_{\alpha,n}(\gamma) \xrightarrow{p} \eta$. Hence $\lim_{n\to\infty} P\{T_n^2 > c^{2*}_{\alpha,n}(\gamma)\} = \lim_{n\to\infty} P\{T_n^2 > \eta\} \le \lim_{n\to\infty} P\{T_n^2 > 0\} \le \alpha\gamma$. If $M^+ \ne M$, the limiting distribution of $T^{2*}_{n_{pq},b}$ is identical with that of $T^{1*}_{n_{pq},b}$. Therefore if $M \ne \emptyset$, then $\lim_{n\to\infty} P\{T_n^2 > c^{2*}_{\alpha,n}(\gamma)\} = \alpha\gamma$. If $M^+ = \emptyset$, then we have $\lim_{n\to\infty} P\{T_n^2 > c^{2*}_{\alpha,n}(\gamma)\} = 0$.

Hence, $\lim_{n\to\infty} P\{T_n^1 > c^{1*}_{\alpha,n}(\gamma) \text{ or } T_n^2 > c^{2*}_{\alpha,n}(\gamma)\} \le \lim_{n\to\infty} P\{T_n^1 > c^{1*}_{\alpha,n}(\gamma)\} + \lim_{n\to\infty} P\{T_n^2 > c^{2*}_{\alpha,n}(\gamma)\} \le \alpha(1-\gamma) + \alpha\gamma = \alpha$ under the null hypothesis. In particular, if $\mu_k < 0$ for all k, then $\lim_{n\to\infty} P\{T_n^1 > c^{1*}_{\alpha,n}(\gamma) \text{ or } T_n^2 > c^{2*}_{\alpha,n}(\gamma)\} = 0$ because both terms become zero. □

Proof for Theorem 4.2. We have that $\lim_{n\to\infty} P\{T_n^1 > c^{1*}_{\alpha,n}(\gamma) \text{ or } T_n^2 > c^{2*}_{\alpha,n}(\gamma)\} \le \lim_{n\to\infty} P\{T_n^1 > c^{1*}_{\alpha,n}(\gamma)\} = 1$ because $T_n^1 \xrightarrow{p} \infty$ under the alternative hypothesis while the critical value is bounded in probability. □

Proof for Lemma 5.1. It follows from Lemma 2.1. of Donald and Hsu[8] that $S_n^1 \Rightarrow \sup_{z\in\mathcal{Z}^*} \Psi_{h_2^{X,Y}}(z)$ under the null hypothesis, and that $S_n^1 \xrightarrow{p} \infty$ under the alternative hypothesis. We have that $S_n^2 \Rightarrow \min(\sup_{z\in\mathcal{Z}} \Psi_{h_2^{X,Y}}(z)$, $\sup_{z\in\mathcal{Z}}(-\Psi_{h_2^{X,Y}}(z)))$ by Lemma A.2. and CMT. The remaining results can be shown in a way similar to that of Lemma 4.1. □

Proof for Lemma 5.3. It follows from Lemma A.2. of Donald and Hsu [8] that $\sqrt{mn/(m+n)}\sup_{z\in\mathcal{Z}} \hat{Z}_n(z) \Rightarrow \sup_{z\in\mathcal{Z}^+} \Psi_{h_2^{X,Y}}(z)$. Then $\sqrt{mn/(m+n)}$ $\min(\sup_{z\in\mathcal{Z}} \hat{Z}_n(z), \sup_{z\in\mathcal{Z}}(-\hat{Z}_n(z))) \Rightarrow \min(\sup_{z\in\mathcal{Z}} \Psi_{h_2^{X,Y}}(z), \sup_{z\in\mathcal{Z}}(-\Psi_{h_2^{X,Y}}$ $(z)))$ follows from the CMT when $\mathcal{Z}^+ = \mathcal{Z}$. The last one holds because $\sqrt{mn/(m+n)}\ \sup_{z\in\mathcal{Z}}(-\hat{Z}_n(z)) \xrightarrow{p} \infty$ which can be shown similarly as we do in Lemma 4.2. □

Proof for Lemma 5.4. It holds by Lemma 3.1 of Donald and Hsu. □

Proof for Theorem 5.1. It holds that $\lim_{n\to\infty} P\{S_n^1 > c_{\alpha,n}^{1*}(\gamma)$ or $S_n^2 > c_{\alpha,n}^{2*}(\gamma)\} \leq \lim_{n\to\infty} P\{S_n^1 > c_{\alpha,n}^{1*}(\gamma)\} + \lim_{n\to\infty} P\{S_n^2 > c_{\alpha,n}^{2*}(\gamma)\}$. The first term is no larger than $\alpha(1-\gamma)$ by Theorem 4.2 in Donald and Hsu. In case $\mathcal{Z}^* = \mathcal{Z}$, it can be shown that $c_{\alpha,n}^{2*}(\gamma) \xrightarrow{p} c$ where c is the $(1-\alpha\gamma)$-quantile of $\min(\sup_{z\in\mathcal{Z}} \Psi_{h_2^{X,Y}}(z), \sup_{z\in\mathcal{Z}}(-\Psi_{h_2^{X,Y}}(z)))$ because the limiting distribution is absolutely continuous according to Linton, Maasoumi, and Whang ([7], p.762). Hence $\lim_{n\to\infty} P\{S_n^2 > c_{\alpha,n}^{2*}(\gamma)\}$ does not exceed $\alpha\gamma$. In case that $\mathcal{Z}^* \neq \mathcal{Z}$, it can be shown that $\lim_{n\to\infty} P\{S_n^1 > c_{\alpha,n}^{1*}(\gamma)$ or $S_n^2 > c_{\alpha,n}^{2*}(\gamma)\}$ reduces to $\lim_{n\to\infty} P\{S_n^1 > \min(c_{\alpha,n}^{1*}(\gamma), c_{\alpha,n}^{2*}(\gamma))\}$ because the limiting distributions of two test statistics are identical. The last term is less than equal to $\lim_{n\to\infty} P\{S_n^1 > c_{\alpha,n}^{1*}(0)\}$ which is less then or equal to α. This proves the first part. The remaining part can be shown in the similar way we do in proving Theorem 4.1. □

Proof for Theorem 5.2. It holds by the same argument in proof for Theorem 4.2 and by Theorem 4.2 of Donald and Hsu. □

References

[1] H. White, "A reality check for data snooping," *Econometrica*, vol. 68, no. 5, pp. 1097–1126, 2000.

[2] P. R. Hansen, "A test for superior predictive ability," *Journal of Business & Economic Statistics*, vol. 23, no. 4, pp. 365–380, 2005.

[3] K. Song, "Testing predictive ability and power robustification," *Journal of Business & Economic Statistics*, vol. 30, no. 2, pp. 288–296, 2012.

[4] D. W. Andrews, "Similar-on-the-boundary tests for moment inequalities exist, but have poor power," 2012.

[5] J. P. Romano, A. M. Shaikh, and M. Wolf, "Hypothesis testing in econometrics," *Annu. Rev. Econ.*, vol. 2, no. 1, pp. 75–104, 2010.

[6] G. F. Barrett and S. G. Donald, "Consistent tests for stochastic dominance," *Econometrica*, vol. 71, no. 1, pp. 71–104, 2003.

[7] O. Linton, E. Maasoumi, and Y.-J. Whang, "Consistent testing for stochastic dominance under general sampling schemes," *The Review of Economic Studies*, vol. 72, no. 3, pp. 735–765, 2005.

[8] S. G. Donald and Y.-C. Hsu, "Improving the power of tests of stochastic dominance," *Econometric Reviews*, vol. 35, no. 4, pp. 553–585, 2014.

[9] D. N. Politis and J. P. Romano, "The stationary bootstrap," *Journal of the American Statistical association*, vol. 89, no. 428, pp. 1303–1313, 1994.

[10] S. Gonçalves and R. de Jong, "Consistency of the stationary bootstrap under weak moment conditions," *Economics Letters*, vol. 81, no. 2, pp. 273–278, 2003.

[11] A. W. Van der Vaart, *Asymptotic statistics*, vol. 3. Cambridge university press, 2000.

1종 오류가 통제된 하이브리드 예측모형비교 검정과 확률적 지배에의 응용

김드보라

서울대학교 경제학부

2016년 8월

국 문 초 록

다양한 예측모형을 비교할 때, 지역 대안 가설들에 대한 검정력을 강건화하기 위하여, Song 은 서로 다른 두 개의 검정, 단측 sup검정과 양측 검정을 동시에 사용할 것을 제안하였다. 그러나 그가 제시한 검정의 경우 1종 오류에 대한 제약조건을 만족시키지 못한다. 이에 따라 본 저자는 Hansen의 방법을 이용하여 Song의 하이브리드 검정방법이 1종 오류에 대한 제약조건을 만족시킬 수 있도록 변형하는 방법을 제안한다. 또한 동시에 사용이라는 아이디어를 확률적 지배검정에 적용할 수 있도록 확장하여, Donald and Hsu의 방법을 바탕으로하는 하이브리트 확률적 지배검정 방법을 제안한다. 시뮬레이션 실험 결과에 따르면, 특정 자료 생성 과정 하에서 본 논문에서 제시하는 하이브리드 예측모형비교 검정이 기존의 검정들 보다 검정력이 좋으며, 대부분의 경우에 Hansen의 예측모형비교 검정과 검정력이 비슷하다. 반면, 확률적지배검정에의 적용은 검정력을 크게 증가시키진 않는 것으로 관찰된다.

주요어: 비모수검정, 예측모형비교, 확률적 지배, 하이브리드 검정
학번: 2014-22287