Center for Quantitative Economics

Multi-horizon uniform superior predictive ability revisited: A size-exploiting and consistent test

Verena Monschang[†], Mark Trede[†], Bernd Wilfling[†]

106/2023

[†] Department of Economics, University of Münster, Germany

wissen.leben
Universität Münster

Multi-horizon uniform superior predictive ability revisited: A size-exploiting and consistent test

Verena Monschang[a], Mark Trede[a], Bernd Wilfling[a,*]

Department of Economics (CQE), Universität Münster, Germany

November 28, 2023

Abstract

Quaedvlieg (2021, *Journal of Business & Economic Statistics*) proposes a uniform Superior Predictive Ability (uSPA) test for comparing forecasts across multiple horizons. The procedure is based on a 'minimum Diebold-Mariano' test statistic, and asymptotic critical values are obtained via bootstrapping. We show, theoretically and via simulations, that Quaedvlieg's test is subject to substantial size distortions. In this article, we establish several convergence results for the 'minimum Diebold-Mariano' statistic, revealing that appropriate asymptotic critical values are given by the quantiles of the standard normal distribution. The uSPA test modified in this manner (i) always retains the nominal size, (ii) is size-exploiting along the boundary that separates the parameter subsets of the null and the alternative uSPA hypotheses, and (iii) is consistent. Based on the closed skew normal distribution, we present a procedure for approximating the power function and demonstrate the favorable finite-sample properties of our test. In an empirical replication, we find that Quaedvlieg's (2021) results on uSPA comparisons between direct and iterative forecasting methods are statistically invalid.

Keywords: Forecast evaluation; Joint-hypothesis testing; Stochastic dominance; Closed skew normal distribution.

JEL classification: C12, C15, C52, C53.

*Correspondance to: Am Stadtgraben 9, 48143 Münster, Germany.
Email addresses: verena.monschang@wiwi.uni-muenster.de (V. Monschang).
mark.trede@uni-muenster.de (M. Trede).
bernd.wilfling@wiwi.uni-muenster.de (B. Wilfling).

1 Introduction

Almost 30 years after its publication, the Diebold-Mariano (DM) test is still the principal tool for comparing forecasts in empirical economic studies (Diebold and Mariano, 1995; Harvey et al., 1997). From a probabilistic perspective, this out-of-sample predictive ability test has been designed to compare forecasts at a single prespecified horizon. In practice, however, it is not uncommon for the individual DM test to be used in forecast comparisons across multiple horizons jointly on the same data set, thereby tacitly ignoring data-snooping concerns (White, 2000).

Recently, various multiple-horizon issues have been discussed in the economic forecasting literature (e.g. Fosten and Gutknecht, 2021), including Quaedvlieg's (2021) multi-horizon predictive ability tests, which directly address the joint-hypothesis testing problem. The author introduces two concepts for comparing forecasts across multiple horizons, termed *uniform* and *average* superior predictive ability (uSPA, aSPA), respectively. While uSPA declares one forecasting method superior to another, if it exhibits a lower expected loss at each individual horizon under consideration, aSPA compares the forecast-specific weighted expected losses aggregated across all horizons. It follows directly from the formal definitions in Section 2 that uSPA implies aSPA, but (in general) not conversely. In this article, we focus on the clear-cut uSPA concept, which—in contrast to aSPA—does not require the forecaster's (subjective) choice of any weights for the multi-horizon forecast comparisons. Because of its strict definition, uSPA might in practice lead to explicit rankings between two forecasts only in rare cases, especially when a large number of forecast horizons are analyzed. On the other hand, if the econometrician finds that one forecast exhibits uSPA to another on a prespecified set of horizons, the ranking is unambiguous. From this perspective, uSPA constitutes an important theoretical benchmark framework for multi-horizon forecast comparisons.

As the uSPA test statistic, Quaedvlieg (2021) uses the 'minimum Diebold-Mariano' statistic, which is the minimum obtained from the set of those DM statistics that are applied individually to each of the forecast horizons under consideration. Noting that this minimum statistic is nonpivotal, the author implements a bootstrap procedure to obtain asymptotic critical values. In Section 2, we argue and prove that this strategy is inappropriate, given the mathematical structure of the parameter space under the uSPA null hypothesis. As a result, Quaedvlieg's uSPA tests are associated with type-I error probabilities that are far beyond the nominal significance level. Interestingly, related issues have emerged elsewhere in the applied statistical literature. *Inter alia*, in the field of neuroimaging, Nichols et al. (2005) report the misuse of Friston et al.'s (1999) test for conjunction, where the underlying problem—considered in a probabilistic setting different from ours—is essentially the same.

We start our analysis by establishing various convergence results for the 'minimum Diebold-Mariano' test statistic under conventional regularity conditions. These results lead to an asymptotic uSPA test with critical values corresponding to the quantiles of the standard normal distribution. We prove that this uSPA test (i) retains its nominal size, and (ii) is size-exploiting along the boundary of the parameter subsets that characterize the uSPA null and alternative hypotheses. (iii) Our uSPA test is consistent in the sense that it rejects the uSPA null hypothesis with probability 1 whenever it is false, as the sample size tends to infinity. We then establish a procedure, based on the closed skew normal distribution (e.g. González-Farías et al., 2020), for approximating the power function of our test, which enables us to demonstrate its favorable finite-sample properties in a Monte Carlo study. Finally, we replicate parts of Quaedvlieg's (2021) empirical analysis and contrast his uSPA forecast comparisons with ours.

Our article is organized as follows. Section 2 formalizes the multi-horizon uSPA framework, establishes the asymptotic features of our uSPA test based on the 'minimum Diebold-Mariano' statistic, and characterizes the power function. Section 3 contains the simulations and our empirical replication, and Section 4 concludes. In the Appendix, we present proofs of some auxiliary results and establish the procedure for approximating the power function.

2 Multi-horizon uSPA testing

2.1 Formal setup

Let $\{\mathbf{y}_t\}_{t=0,\pm1,\pm2,\ldots}$ be a univariate or a multivariate stochastic process of interest. For the integer-valued horizon $h \geq 1$, we denote a forecast of $\mathbf{y}_t$ based on information available at time $t-h$ by $\widehat{\mathbf{y}}_{t|t-h}(\widehat{\boldsymbol{\theta}})$, where $\widehat{\boldsymbol{\theta}}$ is a vector of estimated parameters.[1] We consider two competing forecasting methods, indexed as 1 and 2, and compare their forecasting accuracy by taking into account the multiple horizons $h = 1, \ldots, H$. We write the competing h-step ahead forecasts as $\widehat{\mathbf{y}}^1_{t|t-h}$, $\widehat{\mathbf{y}}^2_{t|t-h}$, and the forecast losses as $L^1_{t,h} \equiv L(\mathbf{y}_t, \widehat{\mathbf{y}}^1_{t|t-h})$, $L^2_{t,h} \equiv L(\mathbf{y}_t, \widehat{\mathbf{y}}^2_{t|t-h})$, where $L(\cdot,\cdot)$ is some real-valued loss function. For the loss differential,

$$d_{t,h} \equiv L^1_{t,h} - L^2_{t,h},$$

we make the following assumption.

Assumption 1. *For each $h = 1, \ldots, H$, $\{d_{t,h}\}$ is assumed to be first-moment stationary.*

Assumption 1 allows us to consider the expected loss differentials $\mu_h \equiv \mathbb{E}(d_{t,h})$ for $h = 1, \ldots, H$, which are the main building block of our forecast comparisons.

Quaedvlieg (2021) discusses two concepts of multi-horizon superior predictive ability, termed 'average Superior Predictive Ability' (aSPA) and 'uniform Superior Predictive Ability' (uSPA), respectively. While aSPA compares the accuracy of the two methods across horizons by means of the quantity $\mu_{\text{aSPA}} \equiv \sum_{h=1}^{H} w_h \mu_h$ (with weights $w_1, \ldots, w_H$ summing

[1] Note that we compare 'forecasting methods' in the sense of Giacomini and White (2006) rather than 'forecasting models'.

to 1), we focus on the stronger uSPA concept, which postulates that the superior method yields better forecasts at all forecast horizons under consideration.

Definition 1. *We define Method 2 to exhibit uSPA to Method 1 if*

$$\mu_{\text{uSPA}} \equiv \min\{\mu_1, \ldots, \mu_H\} > 0.$$

Based on Definition 1, we consider the statistical hypotheses

$$H_0 : \mu_{\text{uSPA}} \leq 0 \quad \text{versus} \quad H_1 : \mu_{\text{uSPA}} > 0, \tag{1}$$

with H_0 stating that there is at least one horizon at which Method 1 performs at least as well as Method 2. Accordingly, H_1 indicates uSPA of Method 2 to Method 1.

A meaningful asymptotic statistic for the uSPA testing problem in Eq. (1) is based on the following technical requirement (e.g. De Jong, 1997; Gonçalves and White, 2002).

Assumption 2. *The loss-differential vector* $\mathbf{d}_t \equiv (d_{t,1}, \ldots, d_{t,H})'$ *is near-epoch dependent in* $L_{2+\delta}$*-norm on* $\{V_t\}$ *with near-epoch dependent coefficients* v_k *of size* $-2(q-1)/(q-2)$*, where* $\{V_t\}$ *is* α*-mixing of size* $-(2+\delta)(q+\delta)/(q-2)$*, for some* $q > 2$ *and* $0 < \delta \leq 2$*, and* $\mathbb{V}\text{ar}(d_{t,h}) > 0$ *for all* $h \in \{1, \ldots, H\}$*.*

Let us define $\bar{\mathbf{d}} = (\bar{d}_1, \ldots, \bar{d}_H)' \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbf{d}_t$ (T is the size of the sample used to evaluate the forecasts) and $\boldsymbol{\mu} \equiv (\mu_1, \ldots, \mu_H)'$. Then, Assumptions 1 and 2 ensure the following 'convergence-in-distribution' result towards the multivariate normal distribution:

$$\sqrt{T}\left(\bar{\mathbf{d}} - \boldsymbol{\mu}\right) \xrightarrow{\text{d}} \mathbb{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{2}$$

with $\boldsymbol{\Sigma} \equiv \lim_{T\to\infty} \mathbb{E}\left[T(\bar{\mathbf{d}} - \boldsymbol{\mu})(\bar{\mathbf{d}} - \boldsymbol{\mu})'\right]$ (e.g. Gonçalves and White, 2002). Thus, Eq. (2) suggests testing the uSPA hypotheses in Eq. (1) via the 'minimum Diebold-Mariano' statistic

$$t_{\text{uSPA}} = \min\left\{\sqrt{T}\frac{\bar{d}_1}{\widehat{\sigma}_1}, \ldots, \sqrt{T}\frac{\bar{d}_H}{\widehat{\sigma}_H}\right\}, \tag{3}$$

where $\widehat{\sigma}_h$ is a consistent, almost surely (a.s.) positive HAC-type estimator of $\sigma_h \equiv \sqrt{(\boldsymbol{\Sigma})_{hh}}$ (the main-diagonal elements of $\boldsymbol{\Sigma}$).

In order to contrast Quaedvlieg's (2021) test with our uSPA test (to be established in Theorem 2 below), we introduce the following notation. We formally label the two tests, which are both based on the t_{uSPA}-statistic from Eq. (3), by Υ_{Quaed} and Υ_{uSPA}, and—when convenient—more compactly write $\Upsilon_{\bullet}$, where $\bullet$ stands for one of the subscripts 'Quaed' or 'uSPA'. We parameterize the uSPA testing problem in Eq. (1) by $\boldsymbol{\mu} \in \mathbb{R}^H \equiv \boldsymbol{\Theta}$, and consider the parameter subsets $\boldsymbol{\Theta}_0 = \left\{\boldsymbol{\mu} \in \mathbb{R}^H : \min\{\mu_1, \ldots, \mu_H\} \leq 0\right\}$ (representing H_0) and $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}_0$ (representing H_1), respectively. Furthermore, we denote the test-specific (i) critical (H_0-rejection) region at the significance level α by $\mathrm{C}^{\alpha}_{\Upsilon_{\bullet}}$, (ii) power function by $\pi_{\Upsilon_{\bullet}}(\boldsymbol{\mu}) = \mathbb{P}_{\boldsymbol{\mu}}\left(\text{'}\Upsilon_{\bullet} \text{ rejects } H_0\text{'}\right) = \mathbb{P}_{\boldsymbol{\mu}}\left(t_{\text{uSPA}} \in \mathrm{C}^{\alpha}_{\Upsilon_{\bullet}}\right)$ for $\boldsymbol{\mu} \in \boldsymbol{\Theta}$, so that the size of the $\Upsilon_{\bullet}$-test can be written as $\sup_{\boldsymbol{\mu}\in\boldsymbol{\Theta}_0}\left\{\pi_{\Upsilon_{\bullet}}(\boldsymbol{\mu})\right\}$.

In our subsequent search for a valid uSPA test, the following theorem proves useful. It provides the asymptotic distribution of the t_{uSPA}-statistic under the condition that $\boldsymbol{\mu} = \mathbf{0}$. (Its proof is closely related to that of Theorem 4 in Hansen et al. (2011).)

Theorem 1. *Under Assumptions 1 and 2, consider the random vector* $\mathbf{X} \equiv (X_1, \ldots, X_H)' \sim \mathbb{N}(\mathbf{0}, \mathbf{R})$, *where* $\mathbf{R} = \mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}$ *with diagonal matrix* $\mathbf{D} = \operatorname{diag}(\sigma_1, \ldots, \sigma_H)$. *Let (i)* $\widehat{\sigma}_h$ *be a consistent estimator of* σ_h *for* $h = 1, \ldots, H$, *and (ii)* $F_{\mathbf{R}}$ *denote the distribution of* $\min\{X_1, ..., X_H\}$. *Then, provided that* $\boldsymbol{\mu} = \mathbf{0}$, *we have*

$$t_{\text{uSPA}} \xrightarrow{\text{d}} F_{\mathbf{R}}.$$

Proof. Given that $\boldsymbol{\mu} = \mathbf{0}$, Eq. (2) provides $\sqrt{T}\bar{\mathbf{d}} \xrightarrow{\text{d}} \mathbb{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Since $\widehat{\mathbf{D}} \equiv \operatorname{diag}(\widehat{\sigma}_1, \ldots \widehat{\sigma}_H)$ is a consistent estimator of $\mathbf{D}$, Slutsky's theorem yields $\sqrt{T}\widehat{\mathbf{D}}^{-1}\bar{\mathbf{d}} \xrightarrow{\text{d}} \mathbb{N}(\mathbf{0}, \mathbf{R})$ with $\mathbf{R} = \mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}$. Since the function $M : \mathbb{R}^H \to \mathbb{R}$, $(x_1, \ldots, x_H)' \mapsto \min\{x_1, \ldots, x_H\}$ is continuous, the continuous mapping theorem yields $t_{\text{uSPA}} \xrightarrow{\text{d}} F_{\mathbf{R}}$. □

In obtaining the critical region $C^{\alpha}_{\Upsilon_{\text{Quaed}}}$ for his (asymptotic) Υ_{Quaed}-test, Quaedvlieg (2021) stipulates $\boldsymbol{\mu} = \mathbf{0}$ (without explicitly mentioning it). The author appropriately notes that the t_{uSPA}-statistic from Eq. (3)—which he uses—is nonpivotal, since its distribution depends on the unknown covariance matrix $\boldsymbol{\Sigma}$ (cf. Theorem 1). He attempts to tackle this nuisance-parameter problem by applying the moving block bootstrap of Künsch (1989) and Liu and Singh (1992). However, the author's stipulation $\boldsymbol{\mu} = \mathbf{0}$ produces size-distorted Υ_{Quaed}-tests. Formally, $\boldsymbol{\mu} = \mathbf{0} \in \boldsymbol{\Theta}_0$, but—as will become evident from the proof of Theorem 2 below—when determining the critical region via the t_{uSPA}-statistic, there are other non-zero vectors $\boldsymbol{\mu}^* \in \boldsymbol{\Theta}_0$ with $\pi_{\Upsilon_{\text{Quaed}}}(\boldsymbol{\mu}^*) > \pi_{\Upsilon_{\text{Quaed}}}(\mathbf{0}) = \alpha$, and thus $\sup_{\boldsymbol{\mu}\in\boldsymbol{\Theta}_0} \left\{\pi_{\Upsilon_{\text{Quaed}}}(\boldsymbol{\mu})\right\} > \alpha$, meaning that Quaedvlieg's (2021) testing methodology entails substantial size distortions.

2.2 An asymptotic size-exploiting uSPA test

Since the uSPA null hypothesis in Eq. (1) is composite, the critical region of a t_{uSPA}-based decision rule should be determined under an appropriate parameter vector $\boldsymbol{\mu}^* \in \boldsymbol{\Theta}_0$, which ensures that the size of the resulting uSPA test is equal to (or less than) the prespecified significance level. The proof of the following theorem shows that such vectors have the form $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_H^*)'$ with $\mu_h^* = 0$ for exactly one index $h \in \{1, \ldots, H\}$, and $\mu_j^* > 0$ for all other indices $j \in \{1, \ldots, H\} \setminus \{h\}$.

Theorem 2. *Under Assumptions 1 and 2, consider the test statistic t_{uSPA} from Eq. (3) with critical region $(u_{1-\alpha}, +\infty)$, where $\alpha \in (0,1)$ is the prespecified significance level, and $u_{1-\alpha}$ the $(1-\alpha)$-quantile of the standard normal distribution $\mathbb{N}(0,1)$. Then, the resulting asymptotic test for the uSPA hypotheses in Eq. (1), denoted by Υ_{uSPA}, is size-α-exploiting (i.e. $\pi_{\Upsilon_{\text{uSPA}}}(\boldsymbol{\mu}) \leq \alpha$ for all $\boldsymbol{\mu} \in \boldsymbol{\Theta}_0$ and $\pi_{\Upsilon_{\text{uSPA}}}(\boldsymbol{\mu}^*) = \alpha$ for at least one $\boldsymbol{\mu}^* \in \boldsymbol{\Theta}_0$).*

The proof of Theorem 2 consists of two steps. (1) We prove that the t_{uSPA}-statistic converges in distribution towards the standard normal for those vectors $\boldsymbol{\mu}^* \in \boldsymbol{\Theta}_0$ that contain exactly one zero and $H-1$ strictly positive elements. (2) We prove that the $\mathbb{N}(0,1)$-limiting distribution stochastically dominates the t_{uSPA}-distributions for all $\boldsymbol{\mu} \in \boldsymbol{\Theta}_0$.

Proof of Theorem 2. Prior to establishing the two steps, let us reconsider the matrix $\widehat{\mathbf{D}} \equiv \text{diag}(\widehat{\sigma}_1, \dots \widehat{\sigma}_H)$, with $\widehat{\sigma}_h$ being a consistent, a.s. positive estimator of σ_h ($h = 1, \dots, H$), and define the random vector

$$\widehat{\mathbf{Z}} = (\widehat{Z}_1, \dots, \widehat{Z}_H)' \equiv \sqrt{T}\,\widehat{\mathbf{D}}^{-1}\bar{\mathbf{d}}.$$

Let $\widehat{\mathbf{e}} = (\widehat{e}_1, \dots, \widehat{e}_H)'$ be the random vector, which consists of $H-1$ zeros and has the element '1' exactly at that position where the random vector $\widehat{\mathbf{Z}}$ attains its minimum. Then, the t_{uSPA}-statistic from Eq. (3) can be written as

$$t_{\text{uSPA}} = \widehat{\mathbf{e}}'\,\widehat{\mathbf{Z}} = \sqrt{T}\,\widehat{\mathbf{e}}'\,\widehat{\mathbf{D}}^{-1}\bar{\mathbf{d}}. \tag{4}$$

Additionally, consider the following equality, which holds for any vector $\mathbf{e} = (e_1, \dots, e_H)'$ and any $\boldsymbol{\mu} = (\mu_1, \dots, \mu_H)'$:

$$\sqrt{T}\left(\widehat{\mathbf{e}}'\,\widehat{\mathbf{D}}^{-1}\bar{\mathbf{d}} - \mathbf{e}'\,\widehat{\mathbf{D}}^{-1}\boldsymbol{\mu}\right) = \widehat{\mathbf{e}}'\,\widehat{\mathbf{D}}^{-1}\sqrt{T}\left(\bar{\mathbf{d}} - \boldsymbol{\mu}\right) + \sqrt{T}\left(\widehat{\mathbf{e}}' - \mathbf{e}'\right)\widehat{\mathbf{D}}^{-1}\boldsymbol{\mu}. \tag{5}$$

Below, we specify $\mathbf{e}$ and $\boldsymbol{\mu}$ such that we are able to obtain the desired convergence result for the t_{uSPA}-statistic.

Step 1: Consider the vectors $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_H^*)' \in \boldsymbol{\Theta}_0$ with $\mu_h^* = 0$ for exactly one index $h \in \{1, \dots, H\}$, and $\mu_j^* > 0$ for all other indices $j \in \{1, \dots, H\} \setminus \{h\}$. We specify the vector $\mathbf{e} = (e_1, \dots, e_H)'$ from Eq. (5) such that $e_h = 1$ when $\mu_h^* = 0$, and $e_j = 0$ for all remaining $j \in \{1, \dots, H\} \setminus \{h\}$ (i.e. when $\mu_j^* > 0$). Then, $\mathbf{e}'\,\widehat{\mathbf{D}}^{-1}\boldsymbol{\mu}^* = 0$, and the left-hand side of Eq. (5) coincides with the t_{uSPA}-statistic for $\boldsymbol{\mu} = \boldsymbol{\mu}^*$. For the two terms on the right-hand side of Eq. (5), we next show that

$$\text{(a)}\ \ \widehat{\mathbf{e}}'\,\widehat{\mathbf{D}}^{-1}\sqrt{T}\left(\bar{\mathbf{d}} - \boldsymbol{\mu}^*\right) \xrightarrow{\text{d}} \mathbb{N}(0,1), \quad \text{and (b)}\ \ \sqrt{T}\left(\widehat{\mathbf{e}}' - \mathbf{e}'\right)\widehat{\mathbf{D}}^{-1}\boldsymbol{\mu}^* \xrightarrow{\text{p}} 0. \tag{6}$$

To prove (a), it suffices to show that $\widehat{\mathbf{e}} \xrightarrow{\text{p}} \mathbf{e}$, since then, in view of Eq. (2),

$$\widehat{\mathbf{e}}' \widehat{\mathbf{D}}^{-1} \sqrt{T} \left(\bar{\mathbf{d}} - \boldsymbol{\mu}^* \right) \xrightarrow{\text{d}} \mathbb{N}(0, \mathbf{e}' \mathbf{D}^{-1} \boldsymbol{\Sigma} \mathbf{D}^{-1} \mathbf{e}) = \mathbb{N}(0, 1).$$

Part (b) in Eq. (6) follows from the continuous-mapping theorem, if $\sqrt{T} \left(\widehat{\mathbf{e}}' - \mathbf{e}' \right) \xrightarrow{\text{p}} \mathbf{0}'$. The two claims, $\widehat{\mathbf{e}} \xrightarrow{\text{p}} \mathbf{e}$ and $\sqrt{T} \left(\widehat{\mathbf{e}}' - \mathbf{e}' \right) \xrightarrow{\text{p}} \mathbf{0}'$, are proved in the Lemmata A.1 and A.2 in Appendix A. Collecting results then establishes

$$t_{\text{uSPA}} \xrightarrow{\text{d}} \mathbb{N}(0, 1) \quad (\text{under } \boldsymbol{\mu}^*).$$

Step 2: It remains to show that the cumulative distribution function (CDF) of the $\mathbb{N}(0, 1)$ distribution, denoted by $\Phi(t)$, stochastically dominates the asymptotic CDF of the test statistic, $F_{t_{\text{uSPA}}}(t)$, under every $\boldsymbol{\mu} \in \boldsymbol{\Theta}_0$.[2] We note that all parameter vectors $\boldsymbol{\mu} \in \boldsymbol{\Theta}_0$ are included in the union of the two sets

$$\mathbf{S}_1 \equiv \{ \boldsymbol{\mu} \in \boldsymbol{\Theta}_0 : \mu_h = 0 \text{ for at least one index } h, \, \mu_j \in \mathbb{R} \text{ otherwise} \}, \quad \text{(Case 1)}$$

$$\mathbf{S}_2 \equiv \{ \boldsymbol{\mu} \in \boldsymbol{\Theta}_0 : \mu_h < 0 \text{ for at least one index } h, \, \mu_j \in \mathbb{R} \text{ otherwise} \}. \quad \text{(Case 2)}$$

Prior to treating the two cases, we note the following inequality, which holds irrespective of any specific choice of $\boldsymbol{\mu}$ and for any $h \in \{1, \ldots, H\}$:

$$\begin{aligned} F_{t_{\text{uSPA}}}(t) = \mathbb{P}_{\boldsymbol{\mu}} \left(\min \left\{ \widehat{Z}_1, \ldots, \widehat{Z}_H \right\} \le t \right) &= 1 - \mathbb{P}_{\boldsymbol{\mu}} \left(\min \left\{ \widehat{Z}_1, \ldots, \widehat{Z}_H \right\} > t \right) \\ &= 1 - \mathbb{P}_{\boldsymbol{\mu}} \left(\widehat{Z}_1 > t, \ldots, \widehat{Z}_H > t \right) \\ &\ge 1 - \mathbb{P}_{\boldsymbol{\mu}} \left(\widehat{Z}_h > t \right) \\ &= \mathbb{P}_{\boldsymbol{\mu}} \left(\widehat{Z}_h \le t \right). \end{aligned} \tag{7}$$

Case 1. Consider $\boldsymbol{\mu} \in \mathbf{S}_1$. Without loss of generality, let $\mu_1 = 0$ and $\mu_2, ..., \mu_H \in \mathbb{R}$. Then, choosing $h = 1$ in (7) and using Eq. (2), we obtain

$$F_{t_{\text{uSPA}}}(t) \ge \mathbb{P}_{\boldsymbol{\mu}} \left(\widehat{Z}_1 \le t \right) \xrightarrow{T \to \infty} \Phi(t).$$

[2]We refer to the definition of first-degree stochastic dominance between the CDFs F_X and F_Y of two random variables: $F_X(t)$ stochastically dominates $F_Y(t)$ if $F_X(t) \le F_Y(t)$ for all t.

Case 2. Consider $\boldsymbol{\mu} \in \mathbf{S}_2$. Without loss of generality, let $\mu_1 < 0$ and $\mu_2, ..., \mu_H \in \mathbb{R}$. Since $\widehat{\sigma}_1^{-1} > 0$ almost surely, we also almost surely have $\widehat{Z}_1 \leq \widehat{Z}_1 - \sqrt{T}\widehat{\sigma}_1^{-1}\mu_1$, and thus

$$\mathbb{P}_{\boldsymbol{\mu}}\left(\widehat{Z}_1 \leq t\right) \geq \mathbb{P}_{\boldsymbol{\mu}}\left(\widehat{Z}_1 - \sqrt{T}\widehat{\sigma}_1^{-1}\mu_1 \leq t\right) \quad \text{for all } t. \tag{8}$$

Choosing $h = 1$ in (7), and using Eqs. (8) and (2), we obtain

$$\begin{aligned} F_{t_{\text{uSPA}}}(t) &\geq \mathbb{P}_{\boldsymbol{\mu}}\left(\widehat{Z}_1 \leq t\right) \\ &\geq \mathbb{P}_{\boldsymbol{\mu}}\left(\widehat{Z}_1 - \sqrt{T}\widehat{\sigma}_1^{-1}\mu_1 \leq t\right) \\ &= \mathbb{P}_{\boldsymbol{\mu}}\left(\sqrt{T}\widehat{\sigma}_1^{-1}\left(\bar{d}_1 - \mu_1\right) \leq t\right) \overset{T\to\infty}{\longrightarrow} \Phi(t), \end{aligned}$$

thus completing the proof. $\square$

2.3 Power of the Υ_{uSPA}-test

The aim of this section is twofold. (1) We show that Υ_{uSPA} is a consistent test in the sense of having a power function $\pi_{\Upsilon_{\text{uSPA}}}(\boldsymbol{\mu})$ under the alternative hypothesis '$\mu_{\text{uSPA}} > 0$' that converges (pointwise) to 1 as $T \to \infty$. (2) We establish an accurate approximation of $\pi_{\Upsilon_{\text{uSPA}}}(\boldsymbol{\mu})$ based on the closed skew normal distribution. As a starting point, we recapitulate (i) that the power function $\pi_{\Upsilon_{\text{uSPA}}}(\boldsymbol{\mu}) = \mathbb{P}_{\boldsymbol{\mu}}\left(t_{\text{uSPA}} > u_{1-\alpha}\right)$ for any $\boldsymbol{\mu} \in \boldsymbol{\Theta}_1$ can be written as

$$\pi_{\Upsilon_{\text{uSPA}}}(\boldsymbol{\mu}) = \mathbb{P}_{\boldsymbol{\mu}}\left(\min\left\{\widehat{Z}_1, \ldots, \widehat{Z}_H\right\} > u_{1-\alpha}\right), \tag{9}$$

and (ii) that under Assumptions 1 and 2, the proof of Theorem 1 in Section 2.1 establishes

$$\sqrt{T}\widehat{\mathbf{D}}^{-1}\left(\bar{\mathbf{d}} - \boldsymbol{\mu}\right) \overset{\text{d}}{\longrightarrow} \mathbb{N}(\mathbf{0}, \mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}). \tag{10}$$

Theorem 3. *Under Assumptions 1 and 2, we have* $\lim_{T\to\infty} \pi_{\Upsilon_{\text{uSPA}}}(\boldsymbol{\mu}) = 1$ *for all* $\boldsymbol{\mu} \in \boldsymbol{\Theta}_1$, *i.e.* Υ_{uSPA} *is a consistent test.*

Proof. For all $\boldsymbol{\mu} \in \boldsymbol{\Theta}_1$, we obtain

$$
\begin{aligned}
\pi_{\Upsilon_{\text{uSPA}}}(\boldsymbol{\mu}) &= \mathbb{P}_{\boldsymbol{\mu}}\left(\min\{\widehat{Z}_1, \ldots, \widehat{Z}_H\} > u_{1-\alpha}\right) \\
&= \mathbb{P}_{\boldsymbol{\mu}}\left(\bigcap_{h=1}^{H}\left\{\widehat{Z}_h > u_{1-\alpha}\right\}\right) \\
&= \mathbb{P}_{\boldsymbol{\mu}}\left(\bigcap_{h=1}^{H}\left\{\widehat{Z}_h - \sqrt{T}\widehat{\sigma}_h^{-1}\mu_h > u_{1-\alpha} - \sqrt{T}\widehat{\sigma}_h^{-1}\mu_h\right\}\right) \\
&= 1 - \mathbb{P}_{\boldsymbol{\mu}}\left(\bigcup_{h=1}^{H}\left\{\widehat{Z}_h - \sqrt{T}\widehat{\sigma}_h^{-1}\mu_h \leq u_{1-\alpha} - \sqrt{T}\widehat{\sigma}_h^{-1}\mu_h\right\}\right) \\
&\geq 1 - \sum_{h=1}^{H}\mathbb{P}_{\boldsymbol{\mu}}\left(\widehat{Z}_h - \sqrt{T}\widehat{\sigma}_h^{-1}\mu_h \leq u_{1-\alpha} - \sqrt{T}\widehat{\sigma}_h^{-1}\mu_h\right) \\
&\overset{T\to\infty}{\to} 1,
\end{aligned}
$$

where we use the fact that (i) $\widehat{Z}_h - \sqrt{T}\widehat{\sigma}_h^{-1}\mu_h \overset{\text{d}}{\longrightarrow} \mathbb{N}(0,1)$ according to Eq. (10), and (ii) that $u_{1-\alpha} - \sqrt{T}\widehat{\sigma}_h^{-1}\mu_h$ diverges in probability to $-\infty$. □

Next, we obtain an approximation of the power function. For large (but finite) sample size T, and given expectation vector $\boldsymbol{\mu} \in \boldsymbol{\Theta}_1$, Eq. (10) suggests approximating the distribution of $\widehat{\mathbf{Z}} = \sqrt{T}\widehat{\mathbf{D}}^{-1}\bar{\mathbf{d}}$ by

$$
\widehat{\mathbf{Z}} \overset{\text{approx.}}{\sim} \mathbb{N}\left(\sqrt{T}\widehat{\mathbf{D}}^{-1}\boldsymbol{\mu}, \widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{D}}^{-1}\right). \tag{11}
$$

Along with recent results on the closed skew normal distribution (e.g. González-Farías et al., 2020, Lemma 2.2.1) and on the distribution of the *maximum* of a multivariate normal random vector (e.g. Arellano-Valle and Genton, 2008), Eqs. (9) and (11) allow us to obtain a straightforward approximation of the power function $\pi_{\Upsilon_{\text{uSPA}}}(\boldsymbol{\mu})$ for $\boldsymbol{\mu} \in \boldsymbol{\Theta}_1$. In Appendix B, we outline how to compute the CDF of the *minimum* of a normally distributed random vector, which gives us the approximated CDF of $\min\{\widehat{Z}_1, \ldots, \widehat{Z}_H\}$ in Eq. (9). Our Monte-Carlo simulations in Section 3 demonstrate that this technique provides accurate approximations for typical sample sizes. Note that the technique can also be used to approximate the rejection probabilities under parameter vectors $\boldsymbol{\mu} \in \boldsymbol{\Theta}_0$.

3 Monte-Carlo study and empirical results

Our simulation setup is a streamlined version of the setting from Quaedvlieg (2021). Instead of separately simulating the losses of the two forecasting methods, we directly sample their loss differentials for the forecast horizons $h = 1, \ldots, H$. Using the notation from Section 2.1, we assume i.i.d. loss-differential vectors

$$\mathbf{d}_t \sim \mathbb{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

with expected loss-differential vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_H)'$ and $H \times H$ covariance matrix $\boldsymbol{\Sigma}$. Note that relaxing the above i.i.d.-assumption is straightforward, but would require an HAC covariance matrix estimator, making the simulations more complex, but without yielding deeper insights into the mechanics of the uSPA testing procedure.

Since real-world forecast performances are likely to be positively correlated across different horizons, we model the covariance matrix as $\boldsymbol{\Sigma} = \mathbf{P} + \mathbf{P} = 2\mathbf{P}$ with $\mathbf{P}$ being the $H \times H$ matrix (see Quaedvlieg, 2021)[3]

$$(\mathbf{P})_{ij} = \begin{cases} 1 & \text{for } i = j, \\ \exp(-0.4 + 0.025 \max(i-1, j-1) - 0.125|i-j|) & \text{else.} \end{cases}$$

For prespecified $\lambda > 0$ and horizon set $\mathcal{H}_0 \subseteq \{1, \ldots, H\}$, we set the expected loss differentials to

$$\mu_h = \begin{cases} \lambda & \text{for } h \notin \mathcal{H}_0, \\ 0 & \text{for } h \in \mathcal{H}_0. \end{cases}$$

Thus, the set $\mathcal{H}_0$ contains all forecast horizons for which both methods perform equally well. When $\mu_h = \lambda > 0$, Method 2 outperforms Method 1 at horizon h. Choosing $\mathcal{H}_0$ as the empty set ($\emptyset$) implies $\boldsymbol{\mu} \in \boldsymbol{\Theta}_1$, while choosing $\mathcal{H}_0 \neq \emptyset$ implies $\boldsymbol{\mu} \in \boldsymbol{\Theta}_0$.

Figure 1 about here

[3] In contrast to Quaedvlieg (2021), who simulates the separate losses, we simulate on the basis of $\boldsymbol{\Sigma} = 2\mathbf{P}$, which reflects the correlation structure of the loss differentials

Figure 1 summarizes our simulation results for the prespecified significance level $\alpha = 0.05$ (the dotted horizontal lines in the four panels). In Panel (a), we depict ten simulation settings, each reflecting a stepwise increase in the 'Number of horizons with equal expected losses' ($\mathcal{H}_0 = \{1\}, \mathcal{H}_0 = \{1,2\}, \dots, \mathcal{H}_0 = \{1,2,\dots,10\}$) along the abscissa (where we use the same loss-differential parameter $\lambda = 1$ and sample size $T = 1000$ in all settings). The solid line represents the simulated type-I error rates of our Υ_{uSPA}-test, the dashed line the error rates of the Υ_{Quaed}-test proposed in Quaedvlieg (2021). Our Υ_{uSPA}-test is exactly size-exploiting at the 'boundary' of the null hypothesis, where both forecasting methods perform equally well at a *single* horizon (shown in the panel for $\mathcal{H}_0 = \{1\}$). The Υ_{uSPA}-test becomes progressively conservative as the number of horizons in $\mathcal{H}_0$ increases, i.e. when the forecasting methods perform equally well at multiple horizons. By contrast, the Υ_{Quaed}-test exhibits (mostly severe) size distortions, unless both forecasting methods perform equally well at all horizons under consideration (shown in the panel for $\mathcal{H}_0 = \{1, \dots, 10\}$).

Panel (b) displays type-I error rates of both uSPA tests for various settings at the boundary between $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_1$. The feature common to all our boundary settings is that both forecasting methods perform equally well at the single horizon $h = 1$, while Method 2 outperforms Method 1 at all other horizons $h = 2, \dots, H$, i.e. $\mathcal{H}_0 = \{1\}$ throughout. The boundary settings now differ in their number of H ('Maximal forecast horizon H') along the abscissa. In our simulations, we again used the uniform loss differential parameter $\lambda = 1$ and the sample size $T = 1000$ in all settings. As shown in Panel (b), the Υ_{uSPA}-test (solid line) always retains the nominal size $\alpha = 0.05$. By contrast, the type-I error rates increase drastically in H for the distorted test suggested in Quaedvlieg (2021).

Finally, Panels (c) and (d) convey an impression of the power of our Υ_{uSPA}-test. The three lines in Panel (c) display the power as a function of the loss-differential parameter λ for the sample sizes $T = 500, 1000, 2000$, where we used the same λ-value at all forecast horizons $h = 1, \dots, H$ with $H = 5$. It is important to note that we did not derive the three power lines by simulation, but calculated them via our approximation procedure from Section 2.3. The superimposed points on the power line for sample size $T = 500$ represent the power, as computed by Monte-Carlo simulations (with 100 000 replications), indicating high accuracy of our power approximation. In Panel (d), the loss-differential parameters are $\lambda_1 \in [0, 0.3]$ for horizon $h = 1$, and $\lambda = 1$ for horizons $h = 2, \ldots, 5$. Since the Υ_{Quaed}-test does not maintain the nominal size, a power comparison with our Υ_{uSPA}-test is not meaningful.

Figure 2 about here

As an empirical application, we replicate the results presented in Quaedvlieg (2021, Figure 3) who compares the performance of *direct forecasts* (Method 1) versus *iterated forecasts* (Method 2) via one-sided multiple uSPA-tests for maximal horizons ranging from 2 to 24 months ($H = 2, \ldots, 24$). Referring to technical details from Marcellino et al. (2006), we analyze the four time series `IVSRRQ` (ratio for manufacturing and trade), `FYGM6` (interest rate), `LHNAG` (civilian labour-force employment), and `FYAAAC` (bond yield).[4] Figure 2 reproduces the uSPA-plots in Quaedvlieg (2021, Figure 3), where the (red) dots represent the values of the t_{uSPA}-statistic for the distinct maximal forecast horizons H. By construction, the t_{uSPA}-realizations are positive, if the iterated forecasts outperform the direct forecasts. Thus, the dots in the four panels in Figure 2 show that the iterated forecasts are better for `IVSRRQ` (with a single exception for $H = 24$) and `FYGM6`, while direct forecasts perform better for `LHNAG` and `FYAAAC`.

In the four panels of Figure 2, the solid (blue) lines represent Quaedvlieg's (2021) critical values, the dashed (blue) lines the critical values of our Υ_{uSPA}-test, both at the 5% level. A mere visual inspection reveals the far-reaching consequences of using the correct critical

[4]We used the data, the Ox-code for the data transformations and the computation of the test statistics, and the bootstrapped critical values from the JBES-site of Quaedvlieg (2021).

region $\mathrm{C}^{\alpha}_{\Upsilon_{\text{uSPA}}}$. Quaedvlieg (2021) reports that the t_{uSPA}-statistics for `IVSRRQ` and `FYGM6` in the Panels (a) and (b) are significantly larger than his critical values for all maximal horizons H, and interprets this finding as statistical support in favour of iterated forecasts. By contrast, using the correct critical region $\mathrm{C}^{\alpha}_{\Upsilon_{\text{uSPA}}}$ shows that the test statistics are insignificant at all maximal horizons, except for $H = 2$ in Panels (a) and (b). Thus, there is actually no statistically significant evidence of uSPA. In a similar vein, for `LHNAG` in Panel (c), Quaedvlieg erroneously interprets the negative t_{uSPA}-values at long horizons ($H \geq 14$) as significant evidence in favour of iterated forecasts, even though negative values of the test statistic indicate that direct forecasts are superior. And finally, for the `FYAAAC` series in Panel (d), the one-sided test in the other direction is no longer significant at any horizon, when using the correct Υ_{uSPA} critical values.

4 Conclusion

In this article, we address the joint-hypothesis-testing problem that typically arises when comparing two competing forecasts across multiple forecast horizons. We propose a test for uniform Superior Predictive Ability, based on a 'minimum Diebold-Mariano' test statistic with asymptotic critical values corresponding to the quantiles of standard normal distribution. Our test is consistent and size-exploiting along the boundary of the parameter subsets belonging to the uSPA null and alternative hypotheses. We provide an algorithm for accurately approximating the power function and demonstrate the favorable probabilistic properties of our test.

An obvious follow-up task is to analyze multi-horizon extensions of the model confidence sets (MCS) of Hansen et al. (2011) (Fosten and Gutknecht, 2021; Quaedvlieg, 2021). Since our uSPA decision rule requires no computational effort, it is in principle a natural candidate for embedding in MCS algorithms. However, due to its strict definition, the uSPA concept often fails to provide statistically significant rankings of forecasts for multiple horizons in empirical applications, as shown in Section 3. We therefore recommend establishing alternative multi-horizon ranking concepts in future research. Here, we briefly sketch the following.

A major advantage of the uSPA concept over its aSPA counterpart is that uSPA does not require the forecaster to take an explicit stance on the choice of weights in the ranking rule. A concept with the same feature, which we call 'majorized Superior Predictive Ability' (mSPA), can be defined by comparing the forecast-specific partial sums of the ordered forecast losses across all horizons. Formally, let μ_h^i denote the expected loss of Method i $(i = 1, 2)$ for horizon h $(h = 1, \dots, H)$, and consider the vectors $(\mu_1^i, \dots \mu_H^i)'$ and their increasing rearrangements $(\mu_{(1)}^i, \dots \mu_{(H)}^i)'$. We then define Method 2 to exhibit mSPA to Method 1, if $\sum_{j=1}^k \mu_{(j)}^1 \geq \sum_{j=1}^k \mu_{(j)}^2$ for $k = 1, \dots, H$ with strict inequality for at least one k. This mSPA formulation is based on Marshall et al.'s (2011, pp. 10-12) definition of 'weak supermajorization' and is similar to the concept of generalized Lorenz dominance (Shorrocks, 1983; Chang et al., 2022). It follows by induction that uSPA implies mSPA. On the other hand, mSPA can provide explicit multi-horizon forecast rankings in many cases where uSPA does not. We leave the derivation of a multiple-horizon mSPA test and its embedding in MCS algorithms to future research.

A Proof(s) and remark(s)

Lemma A.1. *Given the assumptions and notation from Step 1 in the proof of Theorem 2, it follows that* $\widehat{\mathbf{e}} \xrightarrow{\text{p}} \mathbf{e}$.

Proof. Without loss of generality, let $\mu_1^* = 0$ and $\mu_h^* > 0$ for $h = 2, \dots, H$. The lemma is proved, if for the vectors $\widehat{\mathbf{e}}$ and $\mathbf{e}$, we have elementwise

$$\widehat{e}_h \xrightarrow{\text{p}} e_h \quad \text{for } h = 1, \dots, H.$$

We start with the first element $e_1 = 1$. The event $\{\widehat{e}_1 = 1\}$ coincides with the event that the first element in $\widehat{\mathbf{Z}} = (\widehat{Z}_1, \ldots \widehat{Z}_h)'$ is the smallest element. Thus,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{e}_1 = 1) &= \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_1 < \widehat{Z}_2, \ldots, \widehat{Z}_1 < \widehat{Z}_H) \\
&= 1 - \mathbb{P}_{\boldsymbol{\mu}^*}\left(\left\{\widehat{Z}_1 \geq \widehat{Z}_2\right\} \cup \ldots \cup \left\{\widehat{Z}_1 \geq \widehat{Z}_H\right\}\right) \\
&\geq 1 - \left[\mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_1 \geq \widehat{Z}_2) + \ldots + \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_1 \geq \widehat{Z}_H)\right]. \qquad \text{(A.1)}
\end{aligned}$$

It remains to show that $\lim_{T\to\infty} \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_1 \geq \widehat{Z}_h) = 0$ for $h = 2, \ldots, H$. Now,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_1 \geq \widehat{Z}_h) &= \mathbb{P}_{\boldsymbol{\mu}^*}\left(\sqrt{T}\frac{\bar{d}_1}{\widehat{\sigma}_1} \geq \sqrt{T}\frac{\bar{d}_h}{\widehat{\sigma}_h}\right) \\
&= \mathbb{P}_{\boldsymbol{\mu}^*}(\bar{d}_1\widehat{\sigma}_h - \bar{d}_h\widehat{\sigma}_1 \geq 0).
\end{aligned}$$

Note that $\bar{d}_1 \xrightarrow{\text{p}} 0$, $\bar{d}_h \xrightarrow{\text{p}} \mu_h^* > 0$, $\widehat{\sigma}_1 \xrightarrow{\text{p}} \sigma_1 > 0$ and $\widehat{\sigma}_h \xrightarrow{\text{p}} \sigma_h > 0$, implying $\lim_{T\to\infty} \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_1 \geq \widehat{Z}_h) = 0$. Taking the limit and inserting in Eq. (A.1) yields $\widehat{e}_1 \xrightarrow{\text{p}} e_1 = 1$.

Similarly, we show the componentwise convergence of $\widehat{\mathbf{e}}$ towards the remaining elements $e_2 = e_3 = \ldots = e_H = 0$. For $h = 2, \ldots, H$,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{e}_h = 1) &= \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_h < \widehat{Z}_1, \ldots, \widehat{Z}_h < \widehat{Z}_{h-1}, \widehat{Z}_h < \widehat{Z}_{h+1}, \ldots, \widehat{Z}_h < \widehat{Z}_H) \\
&= \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_h < \widehat{Z}_1) \cdot \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_h < \widehat{Z}_2, \ldots, \widehat{Z}_h < \widehat{Z}_{h-1}, \widehat{Z}_h < \widehat{Z}_{h+1}, \ldots, \widehat{Z}_h < \widehat{Z}_H | \widehat{Z}_h < \widehat{Z}_1) \\
&\leq \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_h < \widehat{Z}_1). \qquad \text{(A.2)}
\end{aligned}$$

Using the same reasoning as above, it follows that $\lim_{T\to\infty} \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{Z}_h < \widehat{Z}_1) = 0$. Taking the limit and inserting in Eq. (A.2) yields $\widehat{e}_h \xrightarrow{\text{p}} e_h = 0$ for $h = 2, \ldots, H$. □

Lemma A.2. *Given the assumptions and notation from Step 1 in the proof of Theorem 2, it follows that* $\sqrt{T}\left(\widehat{\mathbf{e}}' - \mathbf{e}'\right) \xrightarrow{\text{p}} \mathbf{0}'$.

Proof. Without loss of generality, let $\mu_1^* = 0$ and $\mu_h^* > 0$ for $h = 2, \ldots, H$. As in Lemma A.1, we prove the componentwise convergence of $\sqrt{T}\left(\widehat{\mathbf{e}}' - \mathbf{e}'\right)$, starting with the first element

$\sqrt{T}(\widehat{e}_1 - 1)$. For $\epsilon > 0$, we have

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\mu}^*}\left(\left|\sqrt{T}\,(\widehat{e}_1 - 1) - 0\right| > \epsilon\right) &\leq \mathbb{P}_{\boldsymbol{\mu}^*}\left(\left|\sqrt{T}\,(\widehat{e}_1 - 1) - 0\right| > 0\right) \\
&= \mathbb{P}_{\boldsymbol{\mu}^*}\left(|\widehat{e}_1 - 1| > 0\right) \\
&= \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{e}_1 = 0) = 1 - \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{e}_1 = 1). \qquad \text{(A.3)}
\end{aligned}$$

From the proof of Lemma A.1, we have $\lim_{T\to\infty} \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{e}_1 = 1) = 1$. Taking the limit and inserting in Eq. (A.3) yields $\sqrt{T}\,(\widehat{e}_1 - e_1) \xrightarrow{\text{p}} 0$.

Similarly, for $h = 2, \ldots, H$, we obtain

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\mu}^*}\left(\left|\sqrt{T}\,(\widehat{e}_h - 0) - 0\right| > \epsilon\right) &\leq \mathbb{P}_{\boldsymbol{\mu}^*}\left(\left|\sqrt{T}\,(\widehat{e}_h - 0) - 0\right| > 0\right) \\
&= \mathbb{P}_{\boldsymbol{\mu}^*}\left(|\widehat{e}_h| > 0\right) = \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{e}_h = 1). \qquad \text{(A.4)}
\end{aligned}$$

From the proof of Lemma A.1, we have $\lim_{T\to\infty} \mathbb{P}_{\boldsymbol{\mu}^*}(\widehat{e}_h = 1) = 0$. Taking the limit and inserting in Eq. (A.4) establishes $\sqrt{T}\,(\widehat{e}_h - e_h) \xrightarrow{\text{p}} 0$ for $h = 2, \ldots, H$. □

B Minimum of multivariate normal distribution

Consider a random vector $\mathbf{X} = (X_1, \ldots, X_H)' \sim \mathbb{N}(\boldsymbol{\nu}, \boldsymbol{\Psi})$ and its minimal element $X_{\min} = \min\{X_1, \ldots, X_H\}$. The CDF of the minimum, $F_{X_{\min}}(z)$, is given by

$$F_{X_{\min}}(z) = \sum_{h=1}^{H} F_{X_h|X_h=X_{\min}}(z) \cdot \mathbb{P}(X_h = X_{\min}), \qquad \text{(B.1)}$$

where $F_{X_h|X_h=X_{\min}}(z)$ is the conditional CDF of X_h given that X_h is the minimal element of $\mathbf{X}$. For the computation of the summands on the right side of Eq. (B.1), we use well-known results on the closed skew normal distribution (González-Farías et al., 2020; Arellano-Valle and Genton, 2008). Let us define the $(H \times H)$ matrices

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -1 & 0 & 0 & \dots & 1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & -1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & -1 & 0 & \dots & 1 \end{bmatrix}, \quad \dots ,$$

$$\mathbf{A}_{H-1} = \begin{bmatrix} 0 & 0 & \dots & 1 & 0 \\ 1 & 0 & \dots & -1 & 0 \\ 0 & 1 & \dots & -1 & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}, \quad \mathbf{A}_H = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{bmatrix},$$

where, at each step of the matrix sequence, the first column of $\mathbf{A}_1$ is successively shifted one column to the right. Then, for $h = 1, \dots, H$,

$$\mathbf{A}_h \mathbf{X} \sim \mathbb{N}(\boldsymbol{\nu}^{\mathbf{A}_h}, \boldsymbol{\Psi}^{\mathbf{A}_h})$$

with $\boldsymbol{\nu}^{\mathbf{A}_h} = \mathbf{A}_h \boldsymbol{\nu} = (\nu_1^{\mathbf{A}_h}, \dots, \nu_H^{\mathbf{A}_h})'$ and

$$\boldsymbol{\Psi}^{\mathbf{A}_h} = \mathbf{A}_h \boldsymbol{\Psi} \mathbf{A}_h' = \begin{bmatrix} (\boldsymbol{\Psi}^{\mathbf{A}_h})_{11} & \dots & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{1H} \\ \vdots & \ddots & \vdots \\ (\boldsymbol{\Psi}^{\mathbf{A}_h})_{H1} & \dots & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{HH} \end{bmatrix}.$$

Following González-Farías et al. (2020, Lemma 2.2.1), the H summands $F_{X_h|X_h=X_{\min}}(z) \cdot \mathbb{P}(X_h = X_{\min})$ in Eq. (B.1) can be obtained from the CDF of the multivariate normal distribution

$$\mathbb{N}\left(\begin{bmatrix} \nu_1^{\mathbf{A}_h} \\ -\nu_2^{\mathbf{A}_h} \\ -\nu_3^{\mathbf{A}_h} \\ \vdots \\ -\nu_H^{\mathbf{A}_h} \end{bmatrix}, \begin{bmatrix} (\boldsymbol{\Psi}^{\mathbf{A}_h})_{11} & -(\boldsymbol{\Psi}^{\mathbf{A}_h})_{12} & -(\boldsymbol{\Psi}^{\mathbf{A}_h})_{13} & \dots & -(\boldsymbol{\Psi}^{\mathbf{A}_h})_{1H} \\ -(\boldsymbol{\Psi}^{\mathbf{A}_h})_{21} & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{22} & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{23} & \dots & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{2H} \\ -(\boldsymbol{\Psi}^{\mathbf{A}_h})_{31} & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{32} & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{33} & \dots & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{3H} \\ \vdots & \vdots & \vdots & & \vdots \\ -(\boldsymbol{\Psi}^{\mathbf{A}_h})_{H1} & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{H2} & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{H3} & \dots & (\boldsymbol{\Psi}^{\mathbf{A}_h})_{HH} \end{bmatrix} \right),$$

evaluated at $(z, 0, \dots, 0)'$. We note that evaluating the CDF of a multivariate normal distribution is numerically challenging in high dimensions. An R code to compute the distribution of the minimum, based on the package `mvtnorm`, is available from the authors upon request.

References

Arellano-Valle, R. B. and M. G. Genton (2008). On the exact distribution of the maximum of absolutely continuous dependent random variables. *Statistics and Probability Letters 78*, 27–35.

Chang, S.-D., P. Cheng, and M. Liou (2022). Likelihood ratio tests for Lorenz dominance. *Journal of Business & Economic Statistics*.

De Jong, R. M. (1997). Central limit theorems for dependent heterogeneous random variables. *Econometric Theory 13*(3), 353–367.

Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*(3), 253–63.

Fosten, J. and D. Gutknecht (2021). Horizon confidence sets. *Empirical Economics 61*(2), 667–692.

Friston, K., A. Holmes, C. Price, C. Büchel, and K. Worsley (1999). Multisubject fMRI studies and conjunction analyses. *NeuroImage 10*(4), 385–396.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Gonçalves, S. and H. White (2002). The bootstrap of the mean for dependent heterogeneous arrays. *Econometric Theory 18*(6), 1367–1384.

González-Farías, G., A. Domínguez-Molina, and A. K. Gupta (2020). The closed skew-normal distribution. In M. G. Genton (Ed.), *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Chapter 2. Boca Raton: CRC Press.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of predicition mean squared errors. *International Journal of Forecasting 13*, 281–291.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics 17*(3), 1217–1241.

Liu, R. Y. and K. Singh (1992). Moving blocks jackknife and bootstrap capture weak dependence. In R. LePage and L. Billard (Eds.), *Exploring the Limits of Bootstrap*, pp. 225–248. John Wiley & Sons, Inc.

Marcellino, M., J. H. Stock, and M. W. Watson (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics 135*, 499–526.

Marshall, A., I. Olkin, and B. Arnold (2011). *Inequalities: Theory of Majorization and its Applications. Second Edition.* Springer.

Nichols, T., M. Brett, J. Andersson, T. Wager, and J.-B. Poline (2005). Valid conjunction inference with the minimum statistic. *NeuroImage 25*(3), 653–660.

Quaedvlieg, R. (2021). Multi-horizon forecast comparison. *Journal of Business & Economic Statistics 39*(1), 40–53.

Shorrocks, A. (1983). Ranking income distributions. *Economica 50*(197), 3–17.

White, H. (2000). A reality check for data snooping. *Econometrica 68*(5), 1097–1126.

Figures

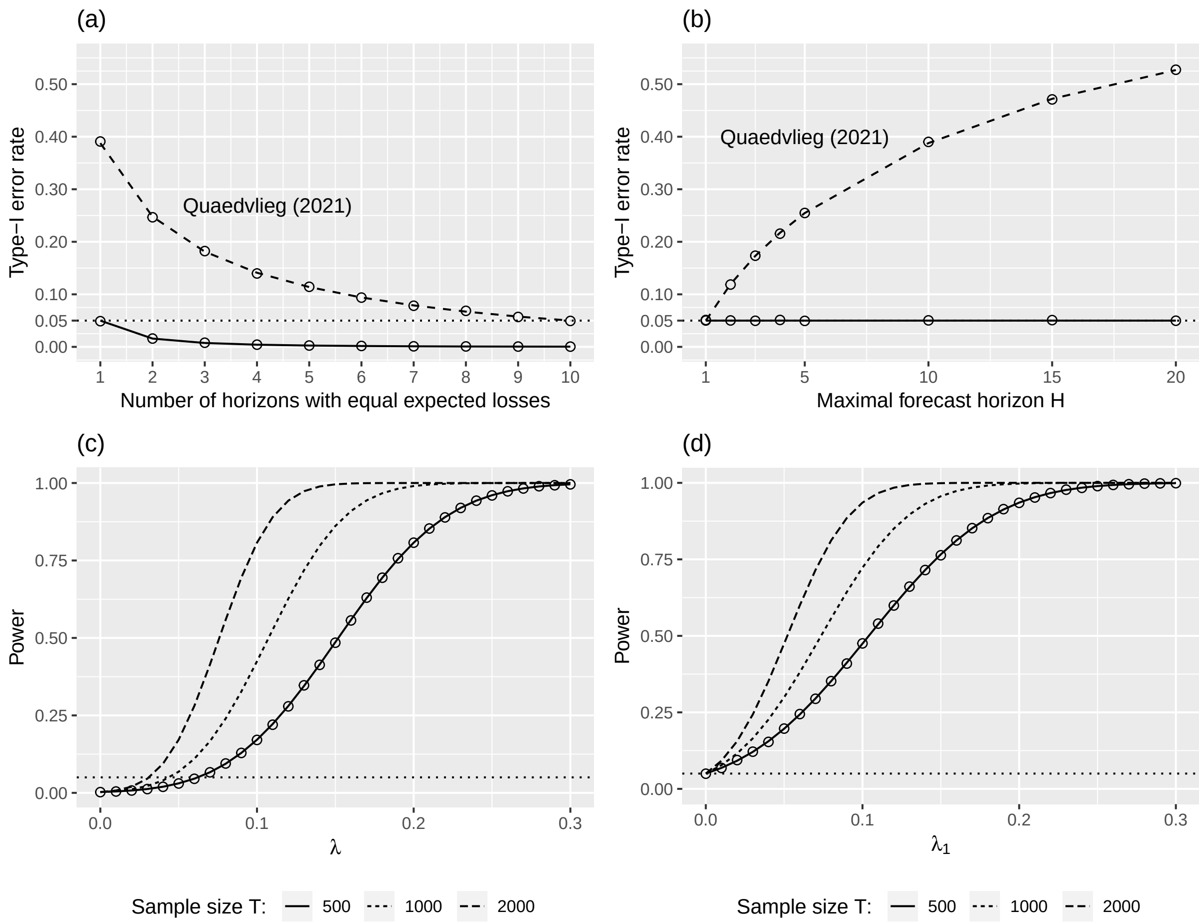

Figure 1: Size and power of tests. (a) Type-I error rate at significance level $\alpha = 0.05$ for maximal horizon $H = 10$. The abscissa indicates the number of forecast horizons with equal expected losses. For all other horizons, the expected loss differential is $\lambda = 1$. The sample size is $T = 1000$. (b) Type-I error rate at significance level $\alpha = 0.05$ for different maximal forecast horizons H. The expected loss differential is 0 at horizon $h = 1$, and $\lambda = 1$ at horizons $h = 2, \ldots, H$. The sample size is $T = 1000$. (c) Power function of the Υ_{uSPA}-test at significance level $\alpha = 0.05$ for maximal forecast horizon $H = 5$, when all expected loss differentials are λ. (d) Power function of the Υ_{uSPA}-test at significance level $\alpha = 0.05$ for maximal forecast horizon $H = 5$, when the expected loss differential is λ_1 at horizon $h = 1$, and $\lambda = 1$ at horizons $h = 2, \ldots, H$.

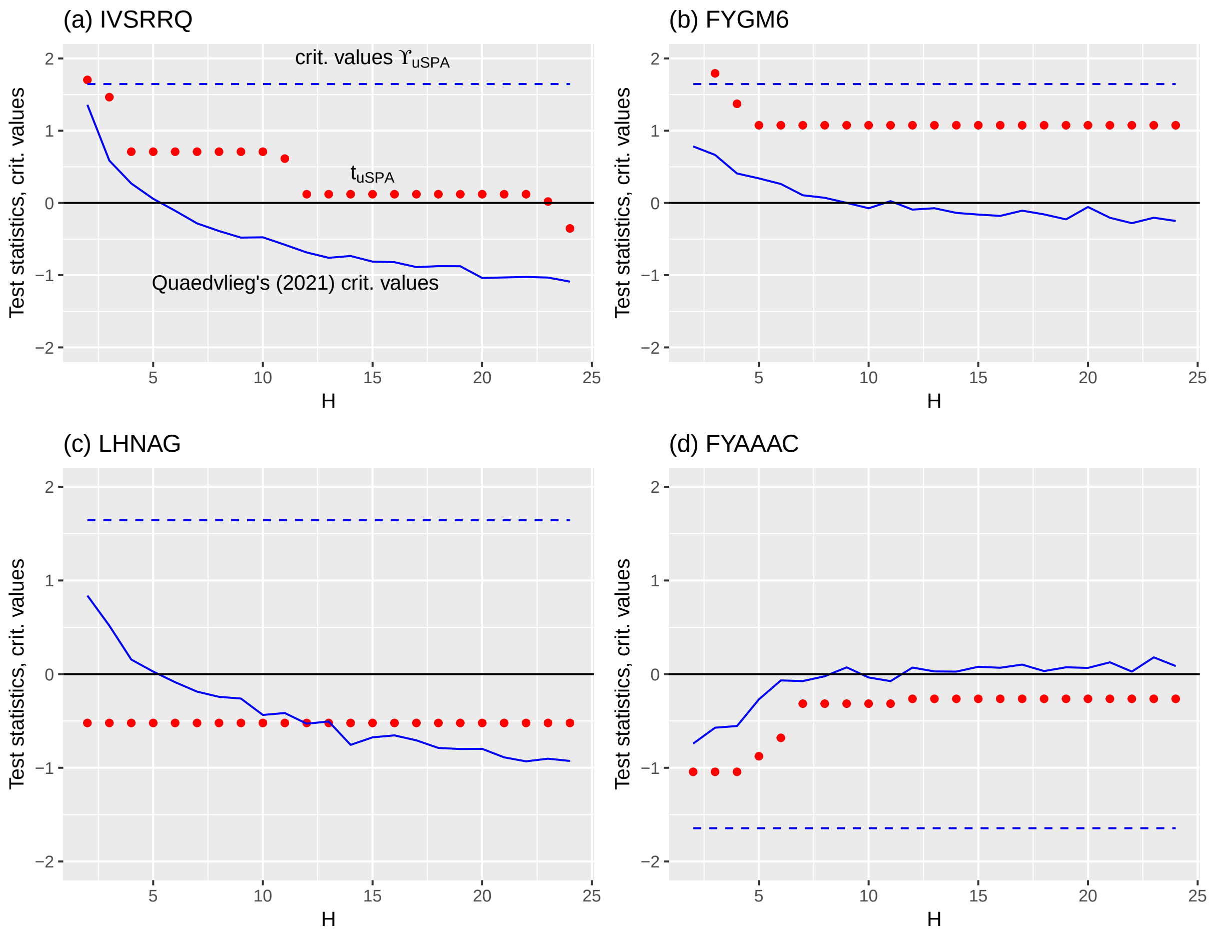

Figure 2: Multi-horizon uSPA-tests for individual series.