Robust Forecast Comparison*

Sainan Jin[1] , Valentina Corradi[2] and Norman R. Swanson[3]
[1]Singapore Management University
[2]University of Surrey
[3]Rutgers University

September 2016

Abstract

Forecast accuracy is typically measured in terms of a given loss function. However, as a consequence of the use of misspecified models in multiple model comparisons, relative forecast rankings are loss function dependent. In order to address this issue, a novel criterion for forecast evaluation that utilizes the entire distribution of forecast errors is introduced. In particular, we introduce the concepts of general-loss (GL) forecast superiority and convex-loss (CL) forecast superiority; and we develop tests for GL (CL) superiority that are based on an out-of-sample generalization of the tests introduced by Linton, Maasoumi, and Whang (2005). Our test statistics are characterized by non standard limiting distributions, under the null, necessitating the use of resampling procedures to obtain critical values. Additionally, the tests are consistent and have nontrivial local power, under a sequence of local alternatives. The above theory is developed for the stationary case, as well as for the case of heterogeneity that is induced by distributional change over time. Monte Carlo simulations suggest that the tests perform reasonably well in finite samples, and an application in which we examine exchange rate data indicates that our tests can help identify superior forecasting models, regardless of loss function.

JEL Classification: C12, C22.

Keywords: Convex loss function, Empirical processes, Forecast superiority, General loss function.

*Sainan Jin, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, snjin@smu.edu.sg; Valentina Corradi, School of Economics, University of Surrey, Guildford, Surrey, GU2 7XH, UK, v.corradi@surrey.ac.uk; and Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. We are grateful to the editor, Peter C. B. Phillips, the co-editor, Robert Taylor, two referees, Xu Cheng, Frank Diebold, Jesus Gonzalo, Simon Lee, Federico Martellosio, Chris Martin, Vito Polito, Barbara Rossi, Olivier Scaillet, Minchul Shin, Tang Srisuma, Liangjun Su; and to seminar participants at the University of Surrey, the University of Bath, the 2015 Princeton/QUT/SJTU/SMU Conference on the Frontier in Econometrics, the 2015 UK Econometric Study Group, 4th International Conference in Applied Econometrics, Universita' Milano Bicocca, and the 2016 Winter Econometric Society Winter Meeting for useful comments and suggestions. For research support, Jin thanks the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021.

1 Introduction

Forecast comparison has a long history in econometrics. When forecast comparison is based upon the evaluation of forecast errors, loss functions are usually specified, and are defined in terms of (conditional) moments of forecast errors, such as mean squared forecast error (MSFE) and mean absolute forecast error (MAFE). Unfortunately, the forecast superiority of one model, relative to other models, is dependent on the loss function that is specified. To circumvent this issue, Granger (1999) proposes the use of a generalized loss function, $L(\cdot)$, with the following properties: (1) $L(e) = 0$, if the forecast error $e = 0$; (2) $L(e) \geq 0$ and $Min_e L(e) = 0$; and (3) $L(e)$ is monotonically non-decreasing, as e moves away from zero, i.e. $L(e_1) \geq L(e_2)$, if $e_1 > e_2 \geq 0$, or if $e_1 < e_2 \leq 0$. We term the class of loss functions that satisfy the above three properties as general loss (GL or $\mathcal{L}_G$) functions. A second class of loss functions are defined as convex loss (CL or $\mathcal{L}_C$) functions, if in addition to satisfying the above three properties, they are convex. Some examples of convex functions include MSFE and MAFE, as well as several asymmetric functions, such as lin-lin and linex functions. For further discussion, refer to Elliott and Timmermann (2004).

A natural question arising from the above discussion is the following. How can we assess different forecasts under generalized loss functions? In particular, suppose that there are l sets of forecasts, with corresponding sequences of one-step-ahead forecast errors, $\{e_{1t}\}$, $\{e_{2t}\}, ..., \{e_{lt}\}$, such that the forecasts are to be ranked the same way, regardless of loss function. In this context, the objective is to introduce forecast evaluation procedures that are robust to the choice of loss function. In order to do this, we introduce two concepts: GL forecast superiority and CL forecast superiority. Simply put, a forecast error sequence GL outperforms other sequences if an economic agent with a GL loss function prefers the former to the latter. Similarly, a forecast error sequence CL outperforms other sequences if an economic agent with a CL loss function prefers the former to the latter. In the sequel, we establish links between tests for GL (CL) forecast superiority and tests for first (second) order stochastic dominance. This allows us to develop a forecast evaluation procedure that is based on an out-of-sample generalization of the tests introduced by Linton, Maasoumi and Whang (2005, hereafter LMW). As our tests are based on the empirical distributions of raw forecast errors, they are robust not only to the choice of the loss function, but also to the possible presence of outliers. The approach of using stochastic dominance to rank distributions of forecast errors has previously been discussed by Corradi and Swanson (2013), although they provide no theory, and their proposed tests are loss function specific. An alternative somewhat related measure called stochastic error loss is discussed in Diebold and Shin (2015).

Since the influential work of Meese and Rogoff (1983), it has become standard practice to select models using out-of-sample forecast comparison. For this reason, much attention in recent years has been given in the econometrics literature to the issue of out-of-sample predictive accuracy testing. See Corradi and Swanson (2006), and the references cited therein for a comprehensive survey of recent developments in forecast comparison methodology. Several features characterize much of the extant literature in this area. First, most of the tests are based upon moments, or conditional moments, of forecast errors, and researchers must specify the objective function (e.g., loss function or likelihood function) in order to carry out forecast evaluation. Second, all of the tests are out-of-sample based, although some (e.g., Diebold and Mariano, 1995, hereafter DM) assume that parameter estimation error is asymptotically negligible, while others account for it when constructing test critical values (e.g., Corradi and Swanson, 2005). Finally, most of them assume that the underlying stochastic process is stationary, which is restrictive in many empirical applications (e.g., in labor economics and macroeconomics). Indeed, we argue that it is fundamentally important to consider the possibly heterogeneous nature of economic variables, and develop corresponding evaluation techniques (see, e.g., Giacomini and White, 2006).

As alluded to above, our objective in this paper is to propose alternatives to moment-based predictive accuracy tests. Namely, we develop distribution-based tests. These tests are viewed as important, because moment-based criteria only look in particular directions when used to examine forecast errors. GL and CL forecast superiority, on the other hand, is based on the evaluation of the entire forecast error distribution, and does not require knowledge or specification of a loss function. Indeed, our forecast evaluation testing procedures are based directly on the evaluation of $F(e)$, the CDF of the forecast error. Moreover, our procedures take parameter estimation error and data dependence into account. Given null hypotheses specified in terms of inequality constraints, we develop limit theory establishing that our tests are consistent and have nontrivial local power under a sequence of local alternatives. Additionally, the asymptotic null distributions of the test statistics are nonstandard, and resampling procedures are used to obtain the critical values.

A preponderance of tests based on (conditional) moments of forecast errors requires an assumption that the underlying stochastic process is stationary. One possible explanation for this assumption is the relative ease with which asymptotic properties of corresponding test statistics can be derived. At the same time, one popular explanation for systematic out-of-sample forecast failure in economics is the prevalence of time varying underlying data generating processes. In light of this fact, we provide a generalization of our testing procedures to a particular type of nonstationarity (i.e., heterogeneity), which is induced by distributional change over time. In the case of stationarity, the pseudo true parameters of all competing models can be estimated consistently, and parameter estimation error is taken into account when deriving the asymptotic properties of our tests. In the case of heterogeneity, there is no need for consistent estimation of the parameters, which may change over time.

It is worth stressing that our testing procedures can be adapted to forecast combination. This is relevant, given that it has become attractive in the literature to combine competing professional forecasts or survey predictions, to aggregate crowd wisdom collected from different sources, and to combine forecasts generated by different econometric models, for example. The reason for this is that combined forecasts often outperform the "best" individual forecasts (see Elliott and Timmermann, 2004, for detailed discussion). However, in standard procedures used in this literature, optimal forecast weights are generally loss function dependent. In our context, one can evaluate different forecast combinations, and select combination weights based on GL and CL forecast superiority. This line of investigation is left to future research.

The rest of the paper is organized as follows. Section 2 introduces hypotheses and test statistics under the assumption that the underlying stochastic process is stationary. In Section 3, we derive the asymptotic distribution of these test statistics, and establish the first order asymptotic validity of a bootstrap procedure used to construct test critical values. Section 4 studies the power properties of the test statistics, and of their associated bootstrap analogs, under local and global alternatives. In Section 5, we extend our results to heterogeneous processes. We examine the finite sample performance of the tests in a series of Monte Carlo simulations, and report findings from these simulations in Section 6. An empirical illustration in which we examine exchange rate data for six industrialized countries is discussed in Section 7. Concluding remarks are gathered in Section 8. All technical details are in the Appendix and in an online supplement.

2 Hypotheses and Tests

In this section, we discuss testing for GL and CL forecast superiority. The tests allow for parameter estimation error, data dependence, and comparison of multiple models; but require the underlying processes to be strictly stationary. We first make the following loss function (L) assumption.

Assumption A.0. $L : \mathbb{R} \rightarrow \mathbb{R}^+$ is continuously differentiable, except for finitely many points, with derivative L', such that $L'(z) \leq 0$, for all $z \leq 0$, and $L'(z) \geq 0$, for all $z \geq 0$.

Definition 2.1 *: e_1 General-Loss (GL) outperforms e_2, denoted as $e_1 \succeq_G e_2$, if and only if*

$$E(L(e_1)) \leq E(L(e_2)),$$

for all $L \in \mathcal{L}_G$. e_1 Convex-Loss (CL) outperforms e_2, denoted as $e_1 \succeq_C e_2$, if and only if

$$E(L(e_1)) \leq E(L(e_2)),$$

for all $L \in \mathcal{L}_C$.

As discussed below, in this paper we assume that e_1 and e_2 are sequences of forecast errors, in which case the above definition pertains to our notion of GL/CL forecast superiority. In order to connect this notion of forecast superiority to stochastic dominance principles, we now establish a mapping between GL forecast superiority and first order stochastic dominance. We also establish linkages between CL forecast superiority and second order stochastic dominance. These results are instrumental for deriving direct tests for GL/CL forecast superiority. Define

$$G(x) = (F_2(x) - F_1(x))sgn(x), \tag{2.1}$$

where $sgn(x) = 1$, if $x \geq 0$, and $= -1$, if $x < 0$; and

$$C(x) = \int_{-\infty}^{x} (F_1(t) - F_2(t))dt1(x < 0) + \int_{x}^{\infty} (F_2(t) - F_1(t))dt1(x \geq 0), \tag{2.2}$$

where $1(\cdot)$ denotes the indicator function, which takes the value 1 if the condition is met, and 0 otherwise.

Proposition 2.2 *: Suppose that Assumption A.0 holds. Then, $E(L(e_1)) \leq E(L(e_2))$, for all $L \in \mathcal{L}_G$, if and only if*

$$G(x) \leq 0, \text{ for all } x \in \mathcal{X}. \tag{2.3}$$

Proposition 2.3 *: Suppose that $\int_{-\infty}^{x}(F_1(t) - F_2(t))dt1(x < 0)$ and $\int_{x}^{\infty}(F_2(t) - F_1(t))dt1(x \geq 0)$ are well defined, for each $x \in \mathcal{X}$. Suppose also that Assumption A.0 holds. Then, $E(L(e_1)) \leq E(L(e_2))$, for all $L \in \mathcal{L}_C$, if and only if*

$$C(x) \leq 0, \text{ for all } x \in \mathcal{X}. \tag{2.4}$$

Remarks. First, before implementing formal tests of GL forecast superiority, we can construct a graph that contains a plot of $G(x)$ against x. When $e_1 \succeq_G e_2$, we expect all points to lie below or on the zero line. In other words, a crossing of the zero line in the graph indicates a violation of GL forecast superiority. Similarly, we can construct a graph that contains a plot of $C(x)$ against x. When $e_1 \succeq_C e_2$, we expect all points to lie below or on the zero line. In other words, a crossing of the zero line in the graph indicates a violation of CL forecast superiority.

Second, if $G(x) = 0$, for all x, in principle it is irrelevant whether one specifies the null as $G(x) \leq 0$ or $G(x) \geq 0$. This is a well known feature, or drawback, of tests for stochastic dominance. We adopt the weak concept of forecast superiority in the above propositions, in order to facilitate our specification of appropriate null hypotheses in the sequel. Strong GL or CL forecast superiority requires that strict inequality holds in (2.3) or (2.4), for some $x \in \mathcal{X}$.

Third, the above propositions only offer a partial ordering between forecast errors. One can generalize the concepts discussed in this paper to third or higher order stochastic dominance (as used in finance, for example). Naturally, higher order stochastic dominance relations correspond to increasingly smaller subsets of $\mathcal{L}_C$, and careful interpretation is needed to justify such generalizations.

Fourth, we can equivalently define the above forecast superiority concepts in terms of quantiles. We do not pursue this further in this paper, for the sake of brevity. Finally, it should be noted that econometric tests for the existence of "ordered" forecast superiority involve composite hypotheses on inequality restrictions. These restrictions may be equivalently formulated in terms of distribution functions, quantiles, or moments.

2.1 Basic framework and test statistics

Suppose that there are l sets of forecast errors, $e_1, ..., e_l$, resulting from l forecasting models. Predictions are made for n periods, indexed from R to T, so that $n = T - R + 1$. The predictions are made for a given forecast horizon, τ.

With a little abuse of notation, we denote $\mathcal{X}$ to be the union of the supports of all forecast errors. Let $\{e_{k,t+\tau} : t = 1, ..., T\}$ be realizations of e_k, for $k = 1, ..., l$. Suppose further that $\{e_{k,t+\tau} : t = 1, ..., T\}$ depends on an unknown finite dimensional parameter, $\beta_{k0} \in \Theta_k \subset \mathbb{R}^{L_k}$, and

$$\begin{aligned} e_{k,t+\tau} &= Y_{t+\tau} - m_k(Z_{k,t+\tau}, \beta_{k0}) \\ &= Y_{t+\tau} - \widetilde{m}_k(Z_{t+\tau}, \beta_0), \end{aligned}$$

where the random variables, $Y_t \in \mathbb{R}$, $Z_{k,t} \in \mathbb{R}^{P_k}$, Z_t is a $\overline{P} \times 1$ random vector comprised of the collection of all predictive regressors, $\beta_0 = (\beta_{10}', ..., \beta_{l0}')'$ is the pseudo true parameter vector on the parameter space $\Theta = \Pi_{k=1}^{l} \Theta_k$, $m_k : \mathbb{R}^{P_k} \times \Theta_k \rightarrow \mathbb{R}$, and $\widetilde{m}_k : \mathbb{R}^{\overline{P}} \times \Theta \rightarrow \mathbb{R}$. Note that $Z_{k,t+\tau}$ is observed at time t. This notation is consistent with most of the literature on forecast comparison. We allow for serial dependence of the realizations and mutual correlation across forecast errors. Let $e_{k,t+\tau}(\beta_k) = Y_{t+\tau} - m_k(Z_{k,t+\tau}, \beta_k)$, $e_{k,t+\tau} = e_{k,t+\tau}(\beta_{k0})$, and $\widehat{e}_{k,t+\tau} = e_{k,t+\tau}(\widehat{\beta}_{k,t})$, where $\widehat{\beta}_{k,t}$ is some possibly nonlinear estimator of β_{k0}, whose construction and properties are detailed below.

Following McCracken (2000), we consider three different estimation schemes, recursive, rolling, and fixed. The schemes differ in how they obtain the sequence of parameter estimates used to construct the sequence of forecasts and forecast errors. Under the recursive scheme, the sequence of forecasts is generated using updated parameter estimates. At each point in time, $t = R, ..., T$, the parameter estimate, $\widehat{\beta}_{k,t}$, depends on all observables, $(Y_s,\ Z_{k,s})$, $s = 1, ..., t$. Under the rolling scheme, however, we use only a fixed window of the most recent R observations. That is, $\widehat{\beta}_{k,t}$ is formed using observations $(Y_s,\ Z_{k,s})$ available from $s = t - R + 1$ through t. The fixed scheme is distinct from the previous two in that the parameters are not updated when new observations become available. The parameter vector is estimated only once, and all n forecasts and forecast errors are constructed using the same parameter estimate, $\widehat{\beta}_{k,t} = \widehat{\beta}_{k,R}$.

In forecasting models where there is no parameter estimation error involved, results analogous to those given below can be established using substantially simpler arguments.

Hereafter, let $\widehat{\beta}_{k,t}$ define the estimator for model k, at time t. For $k = 1, ..., l$, define:

$$\begin{aligned} F_k(x, \beta_k) &= P(e_{k,t+\tau}(\beta_k) \leq x) \text{ and} \\ \overline{F}_{k,n}\left(x, \underline{\widehat{\beta}}_{k,R:T}\right) &= n^{-1} \sum_{t=R}^{T} 1\left(e_{k,t}\left(\widehat{\beta}_{k,t}\right) \leq x\right), \end{aligned}$$

where $\underline{\widehat{\beta}}_{k,R:T} = \left(\widehat{\beta}'_{k,R}, ..., \widehat{\beta}'_{k,T}\right)'$. We denote $F_k(x) = F_k(x, \beta_{k0})$. Now define the following functionals of the joint distribution $F(x_1, ..., x_l)$ of $(e_1, ..., e_l)$:

$$TG^+ = \max_{k=2,..,l} \sup_{x\in\mathcal{X}^+} G_k(x), \qquad TG^- = \max_{k=2,..,l} \sup_{x\in\mathcal{X}^-} G_k(x) \tag{2.5}$$

and

$$TC^+ = \max_{k=2,..,l} \sup_{x\in\mathcal{X}^+} C_k(x), \qquad TC^- = \max_{k=2,..,l} \sup_{x\in\mathcal{X}^-} C_k(x), \tag{2.6}$$

where

$$G_k(x) = \ (F_k(x) - F_1(x))sgn(x), \tag{2.7}$$

and

$$C_k(x) = \int_{-\infty}^{x} (F_1(s) - F_k(s))ds1(x < 0) + \int_{x}^{\infty} (F_k(s) - F_1(s))ds1(x \geq 0). \tag{2.8}$$

In the sequel, without loss of generality, assume that the union of the supports, $\mathcal{X}$, is bounded,[1] as in Klecan, McFadden, and McFadden (1991) and LMW. Some of the recent literature on testing for stochastic dominance no longer requires bounded support. This is typically accomplished by replacing the statistics with trimmed versions thereof, in which the contribution of "extreme" values of x is downweighted. Under mild additional conditions, the differences between the original and the weighted versions of the statistics become asymptotically negligible. For further discussion of this topic, refer to Eq. (3) and Assumptions 3(iii) in Linton, Song and Whang (2010).

Notice that given the nature of our test, one only needs to verify stochastic equicontinuity for $x \in \mathcal{X}^+$ and $x \in \mathcal{X}^-$ separately, where $\mathcal{X}^+ = \mathcal{X} \cap \mathbb{R}^+$ and $\mathcal{X}^- = \mathcal{X} \cap \mathbb{R}^-$, with $\mathbb{R}^+ \equiv \{x \in \mathbb{R},\ x \geq 0\}$ and $\mathbb{R}^-$ $= \mathbb{R}\backslash\mathbb{R}^+$. The hypotheses of interest can now be stated as

$$H_0^{TG} : TG^+ \leq 0 \cap TG^- \leq 0 \text{ vs. } H_1^{TG} : TG^+ > 0 \cup TG^- > 0 \tag{2.9}$$

and

$$H_0^{TC} : TC^+ \leq 0 \cap TG^- \leq 0 \text{ vs. } H_1^{TC} : TC^+ > 0 \cup TC^- > 0. \tag{2.10}$$

In formulating the null hypothesis, H_0^{TG}, we take e_1 as the benchmark forecast error, i.e. we take the corresponding model (model 1, say) as the benchmark model. Failure to reject the null implies that e_1 GL outperforms e_k, for $k = 2, ..., l$. On the other hand, rejection means that e_1 does not GL outperform e_k, for $k = 2, ..., l$. If we do not reject H_0^{TG}, we can discard all of the $k = 2, ..., l$ competitors, as they are all GL dominated. Likewise for the CL forecast superiority test.

The test statistics that we consider are based on scaled empirical analogues of (2.5) and (2.6). They are defined to be

$$TG_n^+ = \max_{k=2,..,l} \sup_{x\in\mathcal{X}^+} \sqrt{n}G_{k,n}(x), \quad TG_n^- = \max_{k=2,..,l} \sup_{x\in\mathcal{X}^-} \sqrt{n}G_{k,n}(x)$$

[1] The boundedness assumption ensures stochastic equicontinuity of the underlying empirical processes that our theory is based upon.

and

$$TC_n^+ = \max_{k=2,..,l} \sup_{x\in\mathcal{X}^+} \sqrt{n}C_{k,n}(x), \quad TC_n^- = \max_{k=2,..,l} \sup_{x\in\mathcal{X}^-} \sqrt{n}C_{k,n}(x),$$

where $G_{k,n}(x) = \left(\overline{F}_{k,n}\left(x, \underline{\widehat{\beta}}_{k,R:T}\right) - \overline{F}_{1,n}\left(x, \underline{\widehat{\beta}}_{1,R:T}\right)\right) sgn(x)$ and $C_{k,n}(x) = \Big\{ \int_{-\infty}^{x} \Big(\overline{F}_{1,n}\left(s, \underline{\widehat{\beta}}_{1,R:T}\right)$ $-\overline{F}_{k,n}\left(s, \underline{\widehat{\beta}}_{k,R:T}\right)\Big) ds 1(x<0) + \int_{x}^{\infty}\left(\overline{F}_{k,n}\left(s, \underline{\widehat{\beta}}_{k,R:T}\right) - \overline{F}_{1,n}\left(s, \underline{\widehat{\beta}}_{1,R:T}\right)\right) ds 1(x \geq 0)\Big\}$.

We next discuss how to compute the suprema in TG_n^+ (TG_n^-) and TC_n^+ (TC_n^-), and the integrals in TC_n^+ (TC_n^-). There have been a number of suggestions in the literature that exploit the step-function nature of $F_{k,n}\left(\cdot, \underline{\widehat{\beta}}_{k,R:T}\right)$. The supremum in TG_n^+ (TG_n^-) can be exactly replaced by a maximum taken over all distinct points in the combined sample. Different methods can be applied in simulations and empirical applications to ensure good finite sample performance of the test. Regarding the computation of TC_n^+ (TC_n^-), using integration by parts, we can compute $C_{k,n}(x)$ with

$$\widehat{C}_{k,n}(x) = \frac{1}{n}\sum_{t=R}^{T}\left\{\left[\left(e_{1,t+\tau}\left(\widehat{\beta}_{1,t}\right) - x\right) sgn(x)\right]_+ - \left[\left(e_{k,t+\tau}\left(\widehat{\beta}_{k,t}\right) - x\right) sgn(x)\right]_+\right\},$$

provided that $E|e_{k,t}| < \infty$, where $[x]_+ = \max\{0, x\}$.

To reduce computation time, it may be preferable to compute approximations to the suprema in TG_n^+ (TG_n^-) and TC_n^+ (TC_n^-), by taking maxima over some smaller grid of points, $\mathcal{X}_N = \{x_1, ..., x_N\}$, where $N < n$. Theoretically, the distribution theory is unaffected by using this approximation, as the set of evaluation points becomes dense in the joint support.

Note that in principle, one can also formulate H_0^{TG} as $TG \leq 0$ versus $TG > 0$, where $TG = \max_{k=2,..,l} \sup_{x\in\mathcal{X}} (F_k(x) - F_1(x))\, sgn(x)$, and one can proceed by constructing the following statistic: $TG_n = \max_{k=2,..,l} \sup_{x\in\mathcal{X}} \sqrt{n}G_{k,n}(x) = \max_{k=2,..,l} \sup_{x\in\mathcal{X}} \sqrt{n}\left(\overline{F}_{k,n}\left(x, \underline{\widehat{\beta}}_{k,R:T}\right) - \overline{F}_{1,n}\left(x, \underline{\widehat{\beta}}_{1,R:T}\right)\right) sgn(x)$. The problem here is that there is a failure of stochastic equicontinuity around $x = 0$. Whether we can address this by replacing the sign function with a smooth function is left to future research. In the sequel, we rely on the formulation in (2.9)-(2.10).

3 Asymptotic Null Distributions

The hypotheses in (2.9) and (2.10) are composite hypotheses, since $H_0^{TG} = H_0^{TG+} \cap H_0^{TG^-}$, where $H_0^{TG+} : TG^+ \leq 0$ and $H_0^{TG-} : TG^- \leq 0$; and since $H_0^{TC} = HG_0^{TC+} \cap H_0^{TC^-}$, where $H_0^{TC+} : TC^+ \leq 0$ and $H_0^{TC-} : TC^- \leq 0$. Hence, in order to test H_0^{TG}, we separately test H_0^{TG+} vs. H_1^{TG+} and H_0^{TG-} vs. H_1^{TG-}. In this context, we (do not) reject the null at a level not higher than α, using Holm bounds (Holm, 1979). Before establishing the asymptotic distributions of our test statistics, we require a few assumptions.

3.1 Assumptions and asymptotic null distributions

Let $\|\cdot\|$ denote the Euclidean norm and let $||X||_q$ denote the L_q norm, with $(E|X|^q)^{1/q}$, for a random variable X. Let $\sup_t$ denote $\sup_{R\leq t\leq T}$, and $\sum_t$ denote $\sum_{t=R}^T$. We require the following assumptions in order to analyze the asymptotic behavior of our test statistics.

Assumption A.1. (i) $\{(Y_t, Z'_{k,t})' : t \geq 1\}$ is a strictly stationary and $\alpha-$mixing sequence with mixing coefficients $\alpha(l) = O(l^{-C_0})$, for some $C_0 > \max\{(q-1)(q+1), 1+2/\delta\}$, with $k = 1, ..., l$, where q is an even integer that satisfies $q > 3(L_{\max}+1)/2$. Here, $L_{\max} = \max\{L_1, ..., L_l\}$ and δ is a positive constant.

(ii) For $k = 1, .., l$, $m_k(Z_{k,t}, \beta_k)$ is differentiable a.s. with respect to β_k in the neighborhood Θ_{k0} of β_{k0}, with $M_k(Z_{k,t}, \beta) \equiv (\partial/\partial\beta)m_k(Z_{k,t}, \beta)$ satisfying $\sup_{\beta\in\Theta_{k0}} ||M_k(Z_{k,t}, \beta)||_2 < \infty$.

(iii) The conditional distribution, $F_k(\cdot|Z_{k,t})$, of $e_{k,t}$ given $Z_{k,t}$, has bounded density with respect to the Lebesgue measure a.s., and $||e_{k,t}||_{2+\delta} < \infty$, for $k = 1, ..., l$.

Assumption A.2. For $k = 1, ..., l$, and $t = R, ..., T$, the estimate $\widehat{\beta}_{k,t}$ satisfies $\widehat{\beta}_{k,t} - \beta_{k0} = B_k(t)H_k(t)$, where $B_k(t)$ is a $P_k \times L_k$ matrix and $H_k(t)$ is $L_k \times 1$, with:

(i) $B_k(t) \to B_k$ a.s., where B_k is a matrix of rank P_k;

(ii) $H_k(t) = t^{-1}\sum_{s=1}^t h_{k,s}$, $R^{-1}\sum_{s=t-R+1}^t h_{k,s}$ and $R^{-1}\sum_{s=1}^R h_{k,s}$ for the recursive, rolling and fixed schemes, respectively, where $h_{k,s} \equiv h_{k,s}(\beta_{k0})$;

(iii) $E(h_{k,s}(\beta_{k0})\ ) = 0$; and

(iv) $||h_{k,s}(\beta_{k0})||_{2+\delta} < \infty$, for some $\delta > 0$.

Assumption A.3. (i) The function $F_k(x, \beta_k)$ is differentiable with respect to β_k in a neighborhood Θ_{k0} of β_{k0}, for $k = 1, ..., l$.

(ii) For $k = 1, ..., l$, and for all sequence of positive constants $\{\xi_n : n \geq 1\}$, such that $\xi_n \to 0$, $\sup_{x\in\mathcal{X}} \sup_{\beta:||\beta-\beta_{k0}||<\xi_n} ||\,(\partial F_k(x, \beta)/\partial\beta\ sgn\,(x) - \Delta_{k0}(x))\,|| = O(\xi_n^\eta)$, for some $\eta > 0$, where $\Delta_{k0}(x) = \partial F_k(x, \beta_{k0})/\partial\beta sgn\,(x)$.

(iii) $\sup_{x\in\mathcal{X}} ||\Delta_{k0}(x)|| < \infty$ for $k = 1, ..., l$.

Assumption A.4. $R, n \to \infty$, as $T \to \infty$, and $\lim_{T\to\infty}(n/R) = \pi$, such that $\pi \in [0, \infty)$.

For testing H_0^{TC}, we require the following modifications of Assumptions A.1 and A.3.

Assumption A.1.* (i) $\{(Y_t, Z'_{k,t})' : t \geq 1\}$ is a strictly stationary and $\alpha-$mixing sequence with mixing coefficients $\alpha(l) = O(l^{-C_0})$, for some $C_0 > \max\{rq/(r-q), 1+2/\delta\}$, with $k = 1, ..., l$ and $r > q > L_{\max}+1$, where δ is a positive constant.

(ii) For $k = 1, .., l$, $m_k(Z_{k,t}, \beta_k)$ is differentiable a.s. with respect to β_k in the neighborhood Θ_{k0} of β_{k0}, with $M_k(Z_{k,t}, \beta) \equiv (\partial/\partial\beta)m_k(Z_{k,t}, \beta)$ satisfying $\sup_{\beta\in\Theta_{k0}} ||M_k(Z_{k,t}, \beta)||_r < \infty$.

(iii) $||e_{kt}||_r < \infty$, for $k = 1, ..., l$.

Assumption A.3.* (i) Assumption A.3(i) holds.

(ii) For $k=1,...,l$, and for all sequences of positive constants $\{\xi_n : n \geq 1\}$, such that $\xi_n \to 0$, $\sup_{x\in\mathcal{X}} \sup_{\beta:||\beta-\beta_{k0}||<\xi_n} ||(\partial/\partial\beta)\{\int_{-\infty}^{x} F_k(t,\beta)dt\ 1(x<0) + \int_{x}^{\infty}(1-F_k(t,\beta))dt1(x\geq 0)\} - \Lambda_{k0}(x)|| = O(\xi_n^{\eta})$, for some $\eta > 0$, where $\Lambda_{k0}(x) = (\partial/\partial\beta)\{\int_{-\infty}^{x} F_k(t,\beta_{k0})dt1(x<0)+\int_{x}^{\infty}(1-F_k(t,\beta_{k0}))dt1(x\geq 0)\}$.

(iii) $\sup_{x\in\mathcal{X}} ||\Lambda_{k0}(x)|| < \infty$, for $k=1,...,l$.

Remarks. First, note that the first and third assumptions parallel those imposed by LMW. The only difference is that we strengthen the uniform continuity conditions in Assumptions A.3 and A.3*. Alternatively, one can assume that the marginal distributions are second order continuously differentiable. Assumption A.1 is needed in order to verify the stochastic equicontinuity of the empirical process, for a class of bounded functions that appears in the TG_n test. Assumption A.1* introduces a trade-off between mixing sizes and moment conditions, and is used to verify the stochastic equicontinuity result for the possibly unbounded functions that appear in the TC_n test. For further details, see Hansen (1996). Assumptions A.3 and A.3* differ in the amount of smoothness required. For the CL forecast superiority test, less smoothness is required.

Second, it is worth stressing that Assumptions A.2 and A.4 are identical to Assumptions 1 and 2 in McCracken (2000), respectively. Notice that we have suppressed the dependence of $B_k(t)$ and $H_k(t)$ on the window size, R. See West (1996) and McCracken (2000) for a discussion of this assumption.

Third, when there is no parameter estimation error, we can dispense with the moment conditions for TG_n, and only need a first moment condition for TC_n. The smoothness conditions on F_k, $k=1,...,l$, and Assumption A.2 are also redundant in this case.

In order to derive the asymptotic null distributions of our test statistics, we define the following empirical processes in (x,β):

$$\begin{aligned}
v_{k,n}^{g}(x,\beta) &= \frac{1}{\sqrt{n}}\sum_{t=R}^{T}\{1(e_{k,t+\tau}(\beta)\leq x) - F_k(x,\beta)\}sgn(x) \text{ and} \\
v_{k,n}^{c}(x,\beta) &= \frac{1}{\sqrt{n}}\sum_{t=R}^{T}\left\{\int_{-\infty}^{x}[1(e_{k,t+\tau}(\beta)\leq s) - F_k(s,\beta)]ds1(x<0)\right. \\
&\quad \left. - \int_{x}^{\infty}[1(e_{k,t+\tau}(\beta)\leq s) - F_k(s,\beta)]ds1(x\geq 0)\right\}.
\end{aligned}$$

Let $(\widetilde{g}_k(\cdot), v_{k0}', v_{10}')'$ be a mean zero Gaussian process with covariance function given by

$$\Omega_k^g(x_1,x_2) = \lim_{T\to\infty} E\begin{pmatrix} v_{k,n}^{g}(x_1,\beta_{k0}) - v_{1,n}^{g}(x_1,\beta_{10}) \\ \sqrt{n}\overline{H}_{k,n} \\ \sqrt{n}\overline{H}_{1,n}\end{pmatrix}\begin{pmatrix} v_{k,n}^{g}(x_2,\beta_{k0}) - v_{1,n}^{g}(x_2,\beta_{10}) \\ \sqrt{n}\overline{H}_{k,n} \\ \sqrt{n}\overline{H}_{1,n}\end{pmatrix}', \tag{3.1}$$

where $\overline{H}_{k,n} = n^{-1}\sum_{t=R}^{T} H_k(t)$. We analogously define $(\widetilde{c}_k(\cdot), v_{k0}', v_{10}')'$ to be a mean zero Gaussian process with covariance function given by

$$\Omega_k^c(x_1,x_2)=\lim_{T\to\infty}E\begin{pmatrix} v_{k,n}^c(x_1,\beta_{k0})-v_{1,n}^c(x_1,\beta_{10}) \\ \sqrt{n}\overline{H}_{k,n} \\ \sqrt{n}\overline{H}_{1,n}\end{pmatrix}\begin{pmatrix} v_{k,n}^c(x_2,\beta_{k0})-v_{1,n}^c(x_2,\beta_{10}) \\ \sqrt{n}\overline{H}_{k,n} \\ \sqrt{n}\overline{H}_{1,n}\end{pmatrix}'. \tag{3.2}$$

It is worth mentioning that the limiting distributions for $\sqrt{n}\overline{H}_{k,n}$, $k=1,...,l$, can be different depending on the forecasting schemes and the parameter π. If we define $\Gamma_k(j)=E\left(h_{k,t}h'_{k,t-j}\right)$, we can verify that the limiting variance of $\sqrt{n}\overline{H}_{k,n}$ is given by $\gamma\sum_{j=-\infty}^{\infty}\Gamma_k(j)$, where

Scheme	γ
Recursive, $\pi=0$	0
Recursive, $0<\pi<\infty$	$2[1-\pi^{-1}\ln(1+\pi)]$
Rolling, $\pi\le 1$	$\pi-\pi^{2/3}$
Rolling, $1<\pi<\infty$	$1-(3\pi)^{-1}$
Fixed	π.

Obviously, $\gamma=0$, when $\pi=0$, indicating the case where parameter estimation error vanishes asymptotically. The limiting null distributions of our test statistics are given in the following theorem.

Theorem 3.1 *: (a) Suppose that Assumptions A.1-A.4 hold. Then, under $H_0^{TG^+}$,*

$$\begin{aligned} TG_n^+ &\Rightarrow \max_{k=2,..,l}\sup_{x\in\mathcal{B}_k^{g+}}[\widetilde{g}_k(x)+\Delta_{k0}(x)'B_kv_{k0}-\Delta_{10}(x)'B_1v_{10}], \quad \text{if } TG^+=0 \\ &\Rightarrow -\infty\ , \quad \text{if } TG^+<0, \end{aligned} \tag{3.3}$$

and under $H_0^{TG^-}$,

$$\begin{aligned} TG_n^- &\Rightarrow \max_{k=2,..,l}\sup_{x\in\mathcal{B}_k^{g-}}[\widetilde{g}_k(x)+\Delta_{k0}(x)'B_kv_{k0}-\Delta_{10}(x)'B_1v_{10}], \quad \text{if } TG^-=0 \\ &\Rightarrow -\infty, \quad \text{if } TG^-<0, \end{aligned} \tag{3.4}$$

where $\mathcal{B}_k^{g+}=\{x\in\mathcal{X}^+ : F_1(x)=F_k(x)\}$ and $\mathcal{B}_k^{g-}=\{x\in\mathcal{X}^- : F_1(x)=F_k(x)\}$.

(b) Suppose that Assumptions A.1, A.2, A.3* and A.4 hold. Then, under $H_0^{TC^+}$,*

$$\begin{aligned} TC_n^+ &\Rightarrow \max_{k=2,..,l}\sup_{x\in\mathcal{B}_k^{c+}}[\widetilde{c}_k(x)+\Lambda_{k0}(x)'B_kv_{k0}-\Lambda_{10}(x)'B_1v_{10}], \quad \text{if } TC^+=0 \\ &\Rightarrow -\infty, \quad \text{if } TC^+<0, \end{aligned} \tag{3.5}$$

and under $H_0^{TC^-}$,

$$\begin{aligned} TC_n^- &\Rightarrow \max_{k=2,..,l}\sup_{x\in\mathcal{B}_k^{c-}}[\widetilde{c}_k(x)+\Lambda_{k0}(x)'B_kv_{k0}-\Lambda_{10}(x)'B_1v_{10}], \quad \text{if } TC^-=0 \\ &\Rightarrow -\infty, \quad \text{if } TC^-<0, \end{aligned} \tag{3.6}$$

where $\mathcal{B}_k^{c+} = \{x \in \mathcal{X}^+ : \int_x^\infty (F_1(s) - F_k(s))ds1(x \geq 0) = 0\}$ *and* $\mathcal{B}_k^{c-} = \{x \in \mathcal{X}^- : \int_{-\infty}^x (F_k(s) - F_1(s))ds1(x \leq 0) = 0\}$.

The asymptotic null distributions of TG_n^+ (TG_n^-) and TC_n^+ (TC_n^-) depend on the pseudo true parameters, $\{\beta_{k0} : k = 1, ..., l\}$, and the distribution functions, $\{F_k(\cdot) : k = 1, ..., l\}$. This implies that the asymptotic critical values for TG_n^+ (TG_n^-) and TC_n^+ (TC_n^-) cannot be tabulated.

3.2 Critical values based on stationary bootstrap

The stationary bootstrap is used to approximate the asymptotic null distributions of our test statistics. In our context, the null essentially consists of an infinite number of composite hypotheses involving inequality restrictions. This negates the use of standard methods for imposing the null in bootstrapping. In addition, the mutual dependence of the forecast errors and the time series dependence in the data also complicate the issue considerably. However, it turns out that the stationary bootstrap can be applied to TG_n^+ and TG_n^-, in the sense that first order asymptotic validity of appropriate bootstrap statistics can be established. Arguments in favor of using the stationary bootstrap with TC_n^+ and TC_n^- are similar, and hence are omitted.

Our objective is to find a bootstrap procedure that mimics the asymptotic null distribution in the least favorable case, where $F_1(x) = ... = F_l(x)$, for all $x \in \mathcal{X}^+$. We use the stationary bootstrap, since it ensures that the resampled series are also stationary and mixing, conditional on the original data. Refer to Politis and Romano (1994) for complete details.

For a suitably chosen random index, $\theta(t)$, the resampled statistic is computed as follows:

$$TG_n^{*+} = \max_{k=2,..,l} \sup_{x \in \mathcal{X}^+} \sqrt{n} \left(G_{k,n}^*(x) - G_{k,n}(x) \right),$$

where

$$G_{k,n}^*(x) = \left(\overline{F}_{k,n} \left(x, \underline{\widehat{\beta}}_{k,\theta(R):\theta(T)} \right) - \overline{F}_{1,n} \left(x, \underline{\widehat{\beta}}_{1,\theta(R):\theta(T)} \right) \right) sgn(x),$$

and

$$\overline{F}_{k,n} \left(x, \underline{\widehat{\beta}}_{k,\theta(R):\theta(T)} \right) = n^{-1} \sum_{t=R}^{T} 1 \left(e_{k,\theta(t)+\tau} \left(\widehat{\beta}_{k,\theta(t)} \right) \leq x \right).$$

In the sequel, we require a smoothing parameter S_n, which satisfies Assumption A.5 below.

Assumption A.5. The smoothing parameter, S_n, satisfies the following conditions: $0 < S_n < 1$, $S_n \to 0$, and $nS_n^2 \to \infty$, as $n \to \infty$.

To implement the stationary bootstrap, follow the algorithm proposed in Politis and Romano (1994). (1) Select S_n. (2) Set $t = R$. Draw $\theta(R)$ at random, uniformly and independently from $\{R, ..., T\}$. (3) Increment t. If $t \leq T$, draw a random variable $V \sim$ Uniform (0, 1), independent of all other random variables. Stop if $t > T$. (a) If $V < S_n$, draw $\theta(t)$ at random, independently and uniformly from

$\{R, ..., T\}$; (b) If $V \geq S_n$, set $\theta(t) = \theta(t-1) + 1$; if $\theta(t) > T$, reset $\theta(t) = R$. (4) Repeat (3). This procedure delivers geometrically distributed blocks of random length, with mean block length equal to $1/S_n$.

When there is no parameter estimation error, so that β_{k0}, $k = 1, ..., l$, is used instead of $\underline{\widehat{\beta}}_{k,\theta(R):\theta(T)}$ in the definition of $G^*_{k,n}$, Theorem 3.1 in Politis and Romano (1994) applies immediately. Let $U_t = (Y_t, Z'_{1,t}, ..., Z'_{l,t})'$, for $t = 1, ..., T+\tau$. Under some regularity conditions, the distribution of $\sqrt{n}\left(G^*_{k,n}(\cdot) - G_{k,n}(\cdot)\right)$, conditional on $\{U_{R+\tau}, ..., U_{T+\tau}\}$, converges to that of $\sqrt{n}\,(G_{k,n} - G_k)$. Then by the continuous mapping theorem, we can approximate the asymptotic distribution of $\sqrt{n}G_{k,n}$, for the elements of the null least favorable to the alternative, i.e. $G_k = 0$, for all k. When $\underline{\widehat{\beta}}_{k,\theta(R):\theta(T)}$ appears in $G^*_{k,n}$, we find that $\widehat{\beta}_{k,T}$ obeys the law of the iterated logarithm. The following then holds, as discussed in White (2000).

Assumption A.6. For an arbitrary $P_k \times 1$ vector, λ_k, with $\lambda'_k\lambda_k = 1$, and for $k = 1, ..., l$, using the notation in Assumption A.2, we have

(i) $P\left[\limsup_{t\geq R} n^{1/2}\left|\lambda'_k\left(\widehat{\beta}_{k,t} - \beta_{k0}\right)\right| \Big/ \left\{\lambda'_k\Sigma_k\lambda_k \log\log(\lambda'_k\Sigma_k\lambda_k)P\right\}^{1/2} = 1\right] = 1$, for the recursive scheme, where $\Sigma_k = B_k[\lim_{T\to\infty}\text{var}(n^{-1/2}\sum_{t=R+1}^{T} H_k(t))]B'_k$.

(ii) $P\left[\limsup_{t\geq R} R^{1/2}\left|\lambda'_k\left(\widehat{\beta}_{k,t} - \beta_{k0}\right)\right| \Big/ \left\{\lambda'_k\Sigma_k\lambda_k \log\log(\lambda'_k\Sigma_k\lambda_k)R\right\}^{1/2} = 1\right] = 1$, for the rolling and fixed schemes, where $\Sigma_k = B_k[\lim_{T\to\infty}\text{var}(R^{-1/2}\sum_{t=R+1}^{T} H_k(t))]B'_k$.

We can now establish the following result.

Theorem 3.2 *: Suppose that Assumptions A.1-A.3 and A.5-A.6 hold, and that* $(n/R)\log\log R \to 0$, *as* $T \to \infty$. *Then, for* $x \in \mathcal{X}^+$ *or* $x \in \mathcal{X}^-$,

$$\rho(\mathbf{L}[\sqrt{n}\left(G^*_{k,n}(\cdot) - G_{k,n}(\cdot)\right)|U_1, ..., U_{T+\tau}], \mathbf{L}[\sqrt{n}\,(G_{k,n}(\cdot) - G_k(\cdot))]) \xrightarrow{p} 0,$$

as $T \to \infty$, *where k=2,..., l,* ρ *is any metric metrizing weak convergence, and* $\mathbf{L}[\cdot]$ *denotes the probability law of the corresponding Hilbert space valued random variable.*

The condition $(n/R)\log\log R \to 0$ appearing in the above theorem is slightly stronger than $n/R \to 0$, which is required in West (1996). However, the stationary bootstrap procedure does not require recomputing $\widehat{\beta}_{k,t}$. Note that, while estimation error is allowed for, we require that it vanishes as the sample gets large. An immediate implication of the above result is the following corollary.

Corollary 3.3 *: Suppose that Assumptions A.1-A.3 and* $A.5 - A.6$ *hold, and that* $(n/R)\log\log R \to 0$, *as* $T \to \infty$. *Then, as* $T \to \infty$,

$$\rho(\mathbf{L}[\max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\sqrt{n}(G^*_{k,n}(x) - G_{k,n}(x))|U_1, ..., U_{T+\tau}], \mathbf{L}[\max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\sqrt{n}\,(G_{k,n}(x) - G_k(x))]) \xrightarrow{p} 0 \text{ and}$$
$$\rho(\mathbf{L}[\max_{k=2,..,l}\sup_{x\in\mathcal{X}^-}\sqrt{n}(G^*_{k,n}(x) - G_{k,n}(x))|U_1, ..., U_{T+\tau}], \mathbf{L}[\max_{k=2,..,l}\sup_{x\in\mathcal{X}^-}\sqrt{n}\,(G_{k,n}(x) - G_k(x))]) \xrightarrow{p} 0.$$

The asymptotic null distribution of TG_n^+ (TG_n^-) can be approximated using $TG_n^{*+} - TG_n^+$ ($TG_n^{*-} - TG_n^-$), for the elements of the null least favorable to the alternative. To do so, specify the number of bootstrap resamples, B, and the smoothing parameter, S_n. Choose B to be a moderately large number, say 200 or 300, as B determines the accuracy of the $p-$values estimated. S_n is closely connected with data dependence. The more data dependence, the smaller S_n should be. One might select S_n to be data driven, following Hall, Horowitz, and Jing (1995), for example. In simulations and applications, we choose a set of S_n values that satisfies Assumption A.5.

Once B and S_n are determined, bootstrap critical values can be estimated. Define $q_{n,S_n}^{G+}(1-\alpha)$ to be the $(1-\alpha)$-th sample quantile of $TG_n^{*+} = \max_{k=2,..,l} \sup_{x\in\mathcal{X}^+} \sqrt{n}(G_{k,n}^*(x) - G_{k,n}(x))$, and $q_{n,S_n}^{G-}(1-\alpha)$ to be the $(1-\alpha)$-th sample quantile of $TG_n^{*-} = \max_{k=2,..,l} \sup_{x\in\mathcal{X}^-} \sqrt{n}(G_{k,n}^*(x) - G_{k,n}(x))$. Alternatively, estimate bootstrap $p-$values, $p_{B,n,S_n}^{G+} = \frac{1}{B}\sum_{s=1}^{B} 1\left(TG_n^{*+} \geq TG_n^+\right)$. Bootstrap $p-$values of TG_n^{*-}, TC_n^{*+}, and TC_n^{*-} can be defined analogously. Then, use the following rules due to Holm (1979):

Rule TG: Reject H_0^{TG} at level α, if $\min\left\{p_{B,n,S_n}^{G+}, p_{B,n,S_n}^{G-}\right\} \leq \alpha/2$.

Rule TC: Reject H_0^{TC} at level α, if $\min\left\{p_{B,n,S_n}^{C+}, p_{B,n,S_n}^{C-}\right\} \leq \alpha/2$.

It is clear that Holm bounds are equivalent to Bonferroni bounds when there are only two hypotheses.

In order to further discuss the properties of our bootstrap based test, consider the test statistic TG_n^+. From Theorem 3.2 and Corollary 3.3, it follows immediately that this test, when implemented using the stationary bootstrap, has asymptotically correct size only in the least favorable case, under the null, i.e. when $F_1(x) = F_k(x)$, for all k, and for all $x \in \mathcal{X}^+$. Hence, the test is not asymptotically similar on the boundary, $\mathcal{B}_0^+ = \{\max_{k=2,...,l}\left(F_k(x) - F_1(x)\right) = 0, \text{ for some } x \in \mathcal{X}^+\}$, and is asymptotically biased towards certain local alternatives. Hansen (2005) shows that p-values associated with the use of the stationary bootstrap are actually upper bounds for an asymptotically unbiased test. On the other hand, as pointed out by LMW, a test based on subsampled critical values is asymptotically similar on $\mathcal{B}_0^+$. This is because the subsampling distribution mimics the sampling distribution. A drawback to the use of subsampling, however, is that tests that are similar on the boundary may have very little power against certain sequences of alternatives, where $F_k(x)-F_1(x) > 0$, for $x \in \mathcal{X}_A \subset \mathcal{X}^+$, and where $F_k(x)-F_1(x) < 0$, for $x \in \mathcal{X}_A \subset \mathcal{X}^-$.[2] Moreover, only pointwise versions of the statements in Theorem 3.2 and Corollary 3.3 hold, under subsampling schemes such as that examined by LSW. Hence, inference based on subsampling or centered bootstrapping methods may be not asymptotically valid, uniformly, over all probabilities under the null hypothesis. This lack of uniformity is typical of tests based on weak inequalities, and is examined further in Corradi, Jin, and Swanson (2016).[3]

[2] In the Monte Carlo experiments conducted in this paper, we implemented subsampling as an alternative to the stationary bootstrap, but results were less satisfactory in all experiments, and so are not reported here.

[3] Note that Andrews and Guggenberger (2010) introduce bootstrap tests which ensure that the asymptotic size (coverage) is at most (at least) α $(1-\alpha)$, uniformly, over all probabilities under the null. Additionally, uniform asymptotics for stochastic dominance tests are derived by Linton, Song and Whang (2010), for the case of pairwise comparisons. In their paper, a

4 Asymptotic Power Properties

Global and local power properties of GL forecast superiority tests are investigated in this section. Analogous results can be established for CL forecast superiority tests, using arguments similar to those presented below. We begin by establishing that the TG_n^+ (TG_n^-) test is consistent against the fixed alternative hypothesis, H_1^{TG+} (H_1^{TG-}).

Theorem 4.1 *: Suppose that Assumptions A.1-A.4 hold. Then, under* H_1^{TG+},

$P(TG_n^+ > q_{n,S_n}^{G+}(1-\alpha)) \rightarrow 1,$ *as* $T \rightarrow \infty$,

and under H_1^{TG-},

$P(TG_n^- > q_{n,S_n}^{G-}(1-\alpha)) \rightarrow 1,$ *as* $T \rightarrow \infty$.

Now, consider the power of the TG_n^+ (TG_n^-) test against a sequence of contiguous local alternatives converging to the null, at rate $n^{-1/2}$. Denote $F_{k,n}(\cdot, \beta_k)$ as the distribution function of $e_{k,t}(\beta_k) \equiv e_{n,k,t}(\beta_k)$, and let $F_{k,n}(\cdot) = F_{k,n}(\cdot, \beta_{k0})$. Consider the following sequence of local alternative distribution functions:

$$F_{k,n}(x) = F_k(x) + n^{-1/2}\delta_k(x), \text{ for } k = 1, ..., l \text{ and } n = 1, 2, ..., \tag{4.1}$$

where $\delta_k(\cdot)$ are real functions, such that $F_{k,n}(\cdot)$ are distribution functions, for each k, and for each n; and where the distribution functions, $\{F_k(\cdot) : k = 1, ..., l\}$, satisfy H_0^{TG}. Let $\sup_n$ denote $\sup_{n\geq 1}$. To analyze the asymptotic behavior of the test under local alternatives, we need to modify Assumptions A.1-A.3, as follows.

Assumption B.1. (i) $\{(Y_t, Z'_{k,t})' \equiv (Y_{n,t}, Z'_{n,k,t})' : t \geq 1, n \geq 1\}$ is an $\alpha-$mixing sequence with mixing coefficients $\alpha(l) = O(l^{-C_0})$, for some $C_0 > \max\{(q-1)(q+1), 1 + 2/\delta\}$, and for $k = 1, ..., l$, where q is an even integer that satisfies $q > 3(L_{\max} + 1)/2$, with $L_{\max} = \max\{L_1, ..., L_l\}$, and δ a positive constant.

(ii) For $k = 1, .., l$, $m_k(Z_{k,t}, \beta_k)$ is differentiable a.s. with respect to β_k in the neighborhood Θ_{k0} of β_{k0}, with $M_k(Z_{k,t}, \beta) \equiv (\partial/\partial\beta)m_k(Z_{k,t}, \beta)$ satisfying $\sup_n \sup_{\beta\in\Theta_{k0}} ||M_k(Z_{k,t}, \beta)||_2 < \infty$, for all $t \geq 1$.

(iii) The conditional distribution, $F_{k,n}(\cdot|Z_{k,t})$, of $e_{k,t}$, given $Z_{k,t}$, has bounded density with respect to the Lebesgue measure a.s., and $||e_{k,t}||_{2+\delta} < \infty$, for $k = 1, ..., l$, $t \geq 1$, and for all $n \geq 1$.

Assumption B.2. For $k = 1, ..., l$ and $t = R, ..., T$, $\widehat{\beta}_{k,t}$ satisfies $\widehat{\beta}_{k,t} - \beta_{k0} = B_k(t)H_k(t)$, where $B_k(t)$ is a $P_k \times L_k$ matrix and $H_k(t)$ is $L_k \times 1$, with:

(i) $B_k(t) \rightarrow B_k$ a.s., where B_k is a matrix of rank P_k;

(ii) $H_k(t) = t^{-1}\sum_{s=1}^{t} h_{k,s}$, $R^{-1}\sum_{s=t-R+1}^{t} h_{k,s}$, and $R^{-1}\sum_{s=1}^{R} h_{k,s}$, for the recursive, rolling, and fixed schemes, respectively, where $h_{k,s} \equiv h_{k,s}(\beta_{k0})$;

key role is played by the contact set, which is the set of x over which the two CDFs are equal. However, in the multiple comparison case, the notion of a contact set has not previously been defined or used. Addressing this issue in the context of stochastic dominance testing using our framework is the topic of ongoing research.

(iii) $\sqrt{n}E(h_{k,s}(\beta_{k0})\ ) \to m_k$; and

(iv) $\sup_n ||h_{k,s}(\beta_{k0})||_{2+\delta} < \infty$, for some $\delta > 0$.

Assumption B.3. (i) The function $F_{k,n}(x, \beta)$ is differentiable with respect to β in a neighborhood Θ_{k0} of β_{k0}, for $k = 1, ..., l$.

(ii) For $k = 1, ..., l$, and for all sequences of positive constants, $\{\xi_n : n \geq 1\}$, such that $\xi_n \to 0$, $\sup_{x\in\mathcal{X}} \sup_{\beta:||\beta-\beta_{k0}||<\xi_n} ||\partial F_{k,n}(x, \beta)/\partial\beta - \Delta_{k0}(x)|| = O(\xi_n^\eta)$, for some $\eta > 0$, where $\Delta_{k0}(x) = \lim_{n\to\infty} \Delta_{k,0,n}(x)$, with $\Delta_{k,0,n}(x) = \partial F_{k,n}(x, \beta_{k0})$.

(iii) $\sup_n \sup_{x\in\mathcal{X}} ||\Delta_{k,0,n}(x)|| < \infty$, for $k = 1, ..., l$.

Note that Assumption B.2 implies that the asymptotic distribution of $\sqrt{n}\left(\widehat{\beta}_{k,t} - \beta_{k0}\right)$ has mean m_k, which might be non-zero under the local alternatives. Nevertheless, this has no effect on the asymptotic distribution of $TG_n^+(TG_n^-)$, as can be seen from the following theorem.

Theorem 4.2 *: Suppose that Assumptions B.1-B.3 and A.4 hold. Then, under the local alternatives in (4.1),*

$$\begin{aligned} TG_n^+ &\Rightarrow \max_{k=2,..,l} \sup_{x\in\mathcal{B}_k^{g+}} [\widetilde{g}_k(x) + \Delta_{k0}(x)' B_k m_k - \Delta_{10}(x)' B_1 m_1 + \mu_k(x)], \text{ and} \\ TG_n^- &\Rightarrow \max_{k=2,..,l} \sup_{x\in\mathcal{B}_k^{g-}} [\widetilde{g}_k(x) + \Delta_{k0}(x)' B_k m_k - \Delta_{10}(x)' B_1 m_1 + \mu_k(x)], \end{aligned}$$

where $\mu_k(x) = \delta_k(x) - \delta_1(x)$, using the notation defined in Section 3.

This result implies that the local asymptotic power of the TG_n^+ (TG_n^-) test, based on the use of stationary bootstrap critical values, is given by the following corollary.

Corollary 4.3 *: Suppose that Assumptions B.1-B.3 and A.5-A.6 hold, and that $(n/R)\log\log R \to 0$, as $T \to \infty$. Then, under the local alternatives,*

$$\begin{aligned} P(TG_n^+ &> q_{n,S_n}^{G+}(1-\alpha)) \to P(TG_{lc}^+ > q^{G+}(1-\alpha)), \\ P(TG_n^- &> q_{n,S_n}^{G-}(1-\alpha)) \to P(TG_{lc}^- > q^{G-}(1-\alpha)), \end{aligned}$$

as $T \to \infty$, where $q_{n,S_n}^{G+}(1-\alpha)$ and $q_{n,S_n}^{G-}(1-\alpha)$ are as defined in Section 3.2, $TG_{lc}^+ = \max_{k=2,..,l} \sup_{x\in\mathcal{B}_k^{g+}} [\widetilde{g}_k(x) + \Delta_{k0}(x)' B_k m_k - \Delta_{10}(x)' B_1 m_1 + \mu_k(x)\}$, $TG_{lc}^- = \max_{k=2,..,l} \sup_{x\in\mathcal{B}_k^{g-}} [\widetilde{g}_k(x) + \Delta_{k0}(x)' B_k m_k - \Delta_{10}(x)' B_1 m_1 + \mu_k(x)\}$, and $q^{G+}(1-\alpha)$ and $q^{G-}(1-\alpha)$ denote the $(1-\alpha)$-th quantiles of the distributions of TG_{lc}^+ and TG_{lc}^-, respectively.

We now briefly discuss the power of our loss function robust tests, relative to that of loss function specific tests. For simplicity, let us consider the case where $l = 2$, and examine the relation between DM tests and our tests. Given that our null hypotheses are stated in terms of weak inequalities, a more natural comparison would be with the White (2000) reality check. However, when $l = 2$, the reality check and DM tests are implemented using the same t-statistic. To proceed, let $L : R \to R^+$ be a loss function satisfying Assumption A.0. Suppose that we are interested in evaluating predictive accuracy according to L. Let $\{e_{1,t}\}$ and $\{e_{2,t}\}$ be the prediction errors from models 1 and 2, respectively. The DM test takes the following form:

$$H_0^{DM} : \mathrm{E}\left(L(e_{1,t})\right) - \mathrm{E}\left(L(e_{2,t})\right) \leq 0$$

vs.

$$H_1^{DM} : \mathrm{E}\left(L(e_{1,t})\right) - \mathrm{E}\left(L(e_{2,t})\right) > 0,$$

and

$$t_{DM,n} = \frac{\frac{1}{\sqrt{n}} \sum_{t=R}^{n} \left(L(e_{1,t}) - L(e_{2,t})\right)}{\widehat{\sigma}_n}, \tag{4.2}$$

where $\widehat{\sigma}_n$ is a consistent estimator of the standard error of $L(e_{1,t}) - L(e_{2,t})$. Here, we reject H_0^{DM} in favor of H_1^{DM}, at a 10% level, if $t_{DM,n} > 1.28$. The test has non trivial power against the sequence of local alternatives $n^{-1/2}\left(\mathrm{E}\left(L(e_{1,t})\right) - \mathrm{E}\left(L(e_{2,t})\right)\right) > 0$.

If H_1^{DM} is true, for any $L \in L_G$, then

$$\begin{aligned} & 0 < n^{-1/2}\mathrm{E}\left(L(e_{1,t})\right) - \mathrm{E}\left(L(e_{2,t})\right) = n^{-1/2} \int_{-\infty}^{\infty} L(z)\left(f_1(z) - f_2(z)\right) \mathrm{d}z \\ = {} & n^{-1/2} \int_{-\infty}^{0} L(z)\left(f_1(z) - f_2(z)\right) \mathrm{d}z + n^{-1/2} \int_{0}^{\infty} L(z)\left(f_1(z) - f_2(z)\right) \mathrm{d}z \\ = {} & -n^{-1/2} \int_{-\infty}^{0} L'(z)\left(F_1(z) - F_2(z)\right) \mathrm{d}z - n^{-1/2} \int_{0}^{\infty} L'(z)\left(F_1(z) - F_2(z)\right) \mathrm{d}z, \end{aligned}$$

which implies that $\max\left\{\max_{x\in\mathcal{X}^-} n^{-1/2}\left(F_1(x) - F_2(x)\right), \max_{x\in\mathcal{X}^+} n^{-1/2}\left(F_2(x) - F_1(x)\right)\right\} > 0$, because of Proposition 2.2. Furthermore, Theorem 4.2 establishes the limiting distribution under such sequences of local alternatives.

Also, if H_1^{DM} is true, for any $L \in L_C$, then

$$\begin{aligned} 0 &< n^{-1/2}\mathrm{E}\left(L(e_{1,t})\right) - \mathrm{E}\left(L(e_{2,t})\right) \\ &= n^{-1/2} \int_{-\infty}^{0} L''(z)\left(\int_{-\infty}^{z} \left(F_1(t) - F_2(t)\right) \mathrm{d}t\right) \mathrm{d}z - n^{-1/2} \int_{0}^{\infty} L''(z)\left(\int_{z}^{\infty} \left(F_1(t) - F_2(t)\right) \mathrm{d}t\right) \mathrm{d}z, \end{aligned}$$

which implies that $\max\left\{\max_{x\in\mathcal{X}^-} n^{-1/2} \int_{-\infty}^{x} \left(F_1(z) - F_2(z)\right) \mathrm{d}z, \max_{x\in\mathcal{X}^+} n^{-1/2} \int_{x}^{\infty} \left(F_1(z) - F_2(z)\right) \mathrm{d}z\right\} > 0$, because of Proposition 2.3. Hence, in large samples, the probability of rejecting H_0^{DM}, but failing to reject H_0^{TG} and/or H_0^{TC}, is zero. However, we cannot formally compare the power properties of these different tests. This is because the statistics are not asymptotically equivalent, under the null, and inference is based on different critical values. Of course, we can always conduct Monte Carlo simulations to compare finite sample power of the tests, as we do in Section 6. Nevertheless, it goes without saying that if interest lies in pairwise predictive accuracy measurement, for a given loss function, then the appropriate test to use in this example is the DM test. This follows because using our robust tests in this context may result in a small loss of power, given that our tests are implemented using Holm bounds.

5 Extensions

Previously, it has been assumed that the underlying process is stationary. However, in some applications, this assumption must be relaxed, due to the presence of heterogeneity. For this reason, asymptotic theory under heterogeneity that is induced by distributional change over time is discussed in this section.

Denote $U_t = (Y_t, Z_{1,t}^{'}, ..., Z_{l,t}^{'})'$, as before, and $Z_t = (Z_{1,t}^{'}, ..., Z_{l,t}^{'})'$. Define $\mathcal{I}_t = \sigma(Z_{t+\tau}, ..., Z_{t+1}, U_t, U_{t-1}, ...)$, where τ is the forecast horizon of interest. Consider a situation where $l \geq 2$ alternative models are used to forecast the variable of interest, τ steps ahead, say $Y_{t+\tau}$. At time t, forecasts are based on the information set, $\mathcal{I}_t$. For $t \geq R$, denote the l forecasts by $\widehat{Y}_{k,t+\tau} = m_k\left(Z_{t+\tau}; \widehat{\beta}_{k,t}\right)$, $k = 1, ..., l$, where each m_k is a measurable function, and where $\widehat{\beta}_{k,t}$ is constructed at time t, by using the most recent R observations. Let $\{e_{k,t+\tau} : t \geq R\}$ be the out-of-sample forecast errors from the k-th competing model. Namely, let $e_{k,t+\tau} = Y_{t+\tau} - \widehat{Y}_{k,t+\tau}$. Further, denote $F_{k,t}(\cdot)$ and $F_{k,t}(\cdot|\mathcal{I}_t)$ as the distribution of $e_{k,t+\tau}$ and the conditional distribution of $e_{k,t+\tau}$ given $\mathcal{I}_t$, respectively. Also, assume that predictions are made for n periods, indexed from R to T, so that $n = T - R + 1$, as above.

Now, change the definition of GL and GC forecast superiority given in Section 2 as follows.

Definition 5.1 *: A sequence of forecast errors,* $\{e_{1,t+\tau}, t \geq R\}$, *General-Loss (GL) outperforms* $\{e_{2,t+\tau}, t \geq R\}$, *denoted as* $e_1 \succeq_G e_2$, *if*

$$\lim_{n\to\infty} n^{-1} \sum_{t=R}^{T} E[L(e_{1,t+\tau}) - L(e_{2,t+\tau})] \leq 0,$$

for all $L \in \mathcal{L}_G$. *A sequence of forecast errors,* $\{e_{1,t+\tau}, t \geq R\}$, *Convex-Loss (CL) outperforms* $\{e_{2,t+\tau}, t \geq R\}$, *denoted as* $e_1 \succeq_C e_2$, *if*

$$\lim_{n\to\infty} n^{-1} \sum_{t=R}^{T} E[L(e_{1,t+\tau}) - L(e_{2,t+\tau})] \leq 0,$$

for all $L \in \mathcal{L}_C$.

Modify Propositions 2.2 and 2.3 to accommodate data heterogeneity, as follows.

Proposition 5.2 *: Suppose that Assumption A.0 holds. Then,* $\lim_{n\to\infty} n^{-1} \sum_{t=R}^{T} E[L(e_{1,t+\tau}) - L(e_{2,t+\tau})] \leq 0$, *for all* $L \in \mathcal{L}_G$, *if and only if*

$$\lim_{n\to\infty} n^{-1} \sum_{t=R}^{T} [F_{2,t}(x) - F_{1,t}(x)] sgn(x) \leq 0,$$

for all $x \in \mathcal{X}$, *where* $\mathcal{X}$ *is the union of the supports of* e_1 *and* e_2.

Proposition 5.3 *: Suppose that* $\int_{-\infty}^{x}(F_{1t}(u)-F_{2t}(u))du\ 1(x<0)$ *and* $\int_{x}^{\infty}(F_{2t}(u)-F_{1t}(u))du\ 1(x\geq 0)]$ *are well defined, for each* $x\in\mathcal{X}$. *Suppose also that Assumption A.0 holds. Then,*

$\lim_{n\to\infty} n^{-1}\sum_{t=R}^{T}E[L(e_{1,t+\tau})-L(e_{2,t+\tau})]\leq 0$, *for all* $L\in\mathcal{L}_C$, *if and only if*
$\lim_{n\to\infty}\ n^{-1}\sum_{t=R}^{T}[\int_{-\infty}^{x}(F_{1,t}(s)-F_{2,t}(s))ds\ 1(x<0)+\int_{x}^{\infty}(F_{2,t}(s)-F_{1,t}(s))ds\ 1(x\geq 0)]\leq 0$,
for all $x\in\mathcal{X}$.

Remarks. First, without the stationarity assumption, we compare the average risks for competing forecasting models, where the average is taken over all n predictions. If $\{e_{i,t+\tau},t\geq R\}$, $i=1,2$, are strictly stationary, one can denote the common marginal distributions as F_i, $i=1,2$, yielding Propositions 2.2 and 2.3.

Second, let $\overline{F}_k(\cdot)=\lim_{n\to\infty} n^{-1}\sum_{t=R}^{T}F_{k,t}(\cdot)$, $k=1,...,l$. Then $e_1\succeq_G e_2$ implies that $\overline{F}_1(0)=\overline{F}_2(0)$.

Third, consider defining conditional analogues of GL and CL forecast superiority by replacing $E[\cdot]$ with $E[\cdot|\mathcal{I}_t]$, and $F_{kt}(\cdot)$ with $F_{kt}(\cdot|\mathcal{I}_t)$, in the above definition. In this case, different sequences of forecast errors are evaluated by comparing their average conditional risks.[4]

For $k=1,...,l$, denote $F_{k,t}(x)=P(e_{k,t+\tau}\leq x)$ and $\overline{F}_{k,n}(x)=n^{-1}\sum_{t=R}^{T}1(e_{k,t+\tau}\leq x)$. Now, define the following functionals of the joint distribution, $F_t(x_1,...,x_l)$, of $(e_{1,t+\tau},...,e_{l,t+\tau})$, for $t\geq R:$

$$HTG^{+}=\max_{k=2,..,l}\sup_{x\in\mathcal{X}^{+}}HG_k(x),\qquad HTG^{-}=\max_{k=2,..,l}\sup_{x\in\mathcal{X}^{-}}HG_k(x) \tag{5.1}$$

and

$$HTC^{+}=\max_{k=2,..,l}\sup_{x\in\mathcal{X}^{+}}HC_k(x),\qquad HTC^{-}=\max_{k=2,..,l}\sup_{x\in\mathcal{X}^{-}}HC_k(x), \tag{5.2}$$

where $HG_k(x)=\lim_{n\to\infty} n^{-1}\sum_{t=R}^{T}[F_{k,t}(x)-F_{1,t}(x)]sgn(x)$, and where

$$HC_k(x)=\lim_{n\to\infty}n^{-1}\sum_{t=R}^{T}[\int_{-\infty}^{x}(F_{1,t}(s)-F_{k,t}(s))ds1(x<0)+\int_{x}^{\infty}(F_{k,t}(s)-F_{1,t}(s))ds1(x\geq 0)].$$

The hypotheses of interest can now be stated as

$$H_0^{HTG}:HTG^{+}\leq 0\cap HTG^{-}\leq 0\text{ vs. }H_1^{TG}:HTG^{+}>0\cup HTG^{-}>0 \tag{5.3}$$

and

$$H_0^{HTC}:HTC^{+}\leq 0\cap HTG^{-}\leq 0\text{ vs. }H_1^{TC}:HTC^{+}>0\cup HTC^{-}>0. \tag{5.4}$$

In formulating the null hypothesis, H_0^{HTG}, define $\{e_{1,t+\tau},t\geq R\}$ to be the benchmark forecast error, with corresponding benchmark forecasting model denoted as model 1. Interest lies in determining whether there exists some forecasting model that is superior to this model. Failure to reject the null implies that no competing model GL/CL outperforms the benchmark model.

[4] We conjecture that the asymptotic properties in this case can be derived by using the results in Harel and Puri (1999). Conditional forecast superiority is a stronger property than its unconditional analogue. However, it is difficult to find empirical support for this property, and thus the topic is left to future research.

The test statistics that we consider in this context are based on the empirical analogues of (5.1) and (5.2). They are defined as follows:

$$HTG_n^+ = \max_{k=2,..,l} \sup_{x \in \mathcal{X}+} \sqrt{n} HG_{k,n}(x), \qquad HTG_n^- = \max_{k=2,..,l} \sup_{x \in \mathcal{X}-} \sqrt{n} HG_{k,n}(x)$$

and

$$HTC_n^+ = \max_{k=2,..,l} \sup_{x \in \mathcal{X}+} \sqrt{n} HC_{k,n}(x), \qquad HTC_n^- = \max_{k=2,..,l} \sup_{x \in \mathcal{X}-} \sqrt{n} HC_{k,n}(x),$$

where $HG_{k,n}(x) = (\overline{F}_{k,n}(x) - \overline{F}_{1,n}(x))sgn(x)$ and $HC_{k,n}(x) = \int_{-\infty}^{x}(\overline{F}_{1,n}(s) - \overline{F}_{k,n}(s))ds1(x < 0) + \int_{x}^{\infty}(\overline{F}_{k,n}(s) - \overline{F}_{1,n}(s))ds1(x \geq 0)$. Theoretical analysis of these statistics requires the following modifications to the assumptions in Section 3.

Assumption HA.1. (i) $\{(Y_t, Z'_{k,t})' : t \geq 1\}$ is an $\alpha-$mixing sequence, with mixing coefficients $\alpha(l) = O(l^{-C_0})$, for some $C_0 > (q-1)(q+1)$, and for $k = 1, ..., l$, where q is an even integer that satisfies $q \geq 2$.

(ii) For all $t \geq R$, the distribution $F_{k,t}(\cdot)$ of $e_{k,t+\tau}$ has bounded density with respect to the Lebesgue measure a.s., and $\sup_{t \geq R} E|e_{k,t+\tau}| < \infty$, for $k = 1, ..., l$.

Assumption HA.4. R is fixed, so that $\lim_{T\to\infty}(n/R) = \infty$.

For the HTC_n test we additionally require the following modification of Assumption HA.1.

Assumption HA.1.* (i) $\{(Y_t, Z'_{k,t})' : t \geq 1\}$ is an $\alpha-$mixing sequence, with mixing coefficients $\alpha(l) = O(l^{-C_0})$, for some $C_0 > rq/(r-q)$, and for $k = 1, ..., l$ and $r > q \geq 2$.

(ii) $\sup_{t \geq R} ||e_{k,t+\tau}||_r < \infty$, for $k = 1, ..., l$.

To derive the asymptotic null distributions of the test statistics, define the empirical processes in x. Namely, let

$$v_{k,n}^{hg}(x) = \frac{1}{\sqrt{n}} \sum_{t=R}^{T} \{1(e_{k,t+\tau} \leq x) - F_{k,t}(x)\} sgn(x) \text{ and}$$

$$v_{k,n}^{hc}(x) = \frac{1}{\sqrt{n}} \sum_{t=R}^{T} \left\{ \int_{-\infty}^{x} [1(e_{k,t+\tau} \leq s) - F_{k,t}(s)] ds1(x < 0) - \int_{x}^{\infty} [1(e_{k,t+\tau} \leq s) - F_{k,t}(s)] ds1(x \geq 0) \right\}.$$

Let $\widetilde{hg}_k(\cdot)$ be a mean zero Gaussian process with covariance function given by

$$\Omega_k^{hg}(x_1, x_2) = \lim_{n\to\infty} E\left(v_{k,n}^{hg}(x_1) - v_{1,n}^{hg}(x_1)\right)\left(v_{k,n}^{hg}(x_2) - v_{1,n}^{hg}(x_2)\right).$$

Analogously, define $\widetilde{hc}_k(\cdot)$ to be a mean zero Gaussian process with covariance function given by

$$\Omega_k^{hc}(x_1, x_2) = \lim_{n\to\infty} E\left(v_{k,n}^{hc}(x_1) - v_{1,n}^{hc}(x_1)\right)\left(v_{k,n}^{hc}(x_2) - v_{1,n}^{hc}(x_2)\right).$$

The limiting null distributions of the test statistics are given in the following theorem.

Theorem 5.4 *: (a) Suppose that Assumptions HA.*1 *and HA.*4 *hold. Then, under* $H_0^{HTG^+}$,

$$\begin{aligned} HTG_n^+ &\Rightarrow \max_{k=2,..,l} \sup_{x \in \mathcal{B}_k^{hg+}} \widetilde{hg}_k(x), \ \ \textit{if} \ \ HTG^+ = 0 & (5.5) \\ &\Rightarrow -\infty, \textit{ if } \ HTG^+ < 0, \end{aligned}$$

and under $H_0^{HTG^-}$,

$$\begin{aligned} HTG_n^- &\Rightarrow \max_{k=2,..,l} \sup_{x \in \mathcal{B}_k^{hc-}} \widetilde{hg}_k(x), \ \ \textit{if} \ \ HTG^- = 0 & (5.6) \\ &\Rightarrow -\infty, \ \ \textit{if} \ \ HTG^- < 0, & (5.7) \end{aligned}$$

where $\mathcal{B}_k^{hg+} = \{x \in \mathcal{X}^+ : \overline{F}_1(x) = \overline{F}_k(x)\}$ *and* $\mathcal{B}_k^{hg-} = \{x \in \mathcal{X}^- : \overline{F}_1(x) = \overline{F}_k(x)\}$.

(b) Suppose that Assumptions HA.1^* *and HA.4 hold. Then, under* $H_0^{HTC^+}$,

$$\begin{aligned} HTC_n^+ &\Rightarrow \max_{k=2,..,l} \sup_{x \in \mathcal{B}_k^{hc+}} \widetilde{hc}_k(x), \ \ \textit{if} \ \ HTC^+ = 0 & (5.8) \\ &\Rightarrow -\infty, \ \ \textit{if} \ \ HTC^+ < 0, \end{aligned}$$

and under $H_0^{HTC^-}$,

$$\begin{aligned} HTC_n^- &\Rightarrow \max_{k=2,..,l} \sup_{x \in \mathcal{B}_k^{hc-}} \widetilde{hc}_k(x), \ \ \textit{if} \ \ HTC^- = 0 & (5.9) \\ &\Rightarrow -\infty, \ \ \textit{if} \ \ HTG^- < 0, & (5.10) \end{aligned}$$

where $\mathcal{B}_k^{hc^+} = \{x \in \mathcal{X}^+ : \int_x^\infty (\overline{F}_1(s) - \overline{F}_k(s))ds = 0\}$ *and* $\mathcal{B}_k^{hc^-} = \{x \in \mathcal{X}^- : \int_{-\infty}^x (\overline{F}_k(s) - \overline{F}_1(s))ds = 0\}$.

The asymptotic null distributions of HTG_n^+ (HTG_n^-) and HTC_n^+ (HTC_n^-) depend on the distribution functions, $\{\overline{F}_k(\cdot) : k = 1, ..., l\}$. This implies that the asymptotic critical values for HTG_n^+ (HTG_n^-) and HTC_n^+ (HTC_n^-) cannot be tabulated. However, Theorem 2.2 in Goncalves and White (2004) applies in this case, and their stochastic equicontinuity result for heterogeneous dependent variables can thus be used to establish the validity of block bootstrap.[5] Associated global and local power properties can also be established, as in Section 4 (for brevity, we do not repeat the arguments here).

6 Simulation Evidence

In this section, we first discuss results from simulations conducted in order to evaluate the finite sample performance of GL and CL forecast superiority tests, when there are only two competing sequences of forecast errors. We then discuss results from a Monte Carlo experiment designed to examine the

[5] We cannot use the stationary bootstrap in this case, because it assumes stationarity of the underlying process. While a subsampling procedure can be used, simulation results show that size and power properties are poor when subsampling is used.

finite sample performance of the tests, when there are more than two competing sequences of forecast errors, under stationarity. Finally, a small Monte Carlo experiment is conducted in order to evaluate the performance of the tests, when the underlying process is not stationary.

When computing the suprema in TG_n^+, TG_n^-, TC_n^+, and TC_n^-, we take the maximum over an equally spaced grid of size $\lceil 1.5n^{0.6} \rfloor$, and over a 98% range of the pooled empirical distribution. Namely, we take the 1% and 99% percentiles of this empirical distribution and then form an equally spaced grid between these two extremes. For each experiment we carry out 1000 replications. We set the number of bootstrap resamples, B, equal to 300. Additionally, six different values of the smoothing parameter, S_n, are examined for each sample size, $n \in \{100, 500, 1000\}$, for the pairwise comparison case, and four different values of S_n are examined for $n \in \{250, 500, 1000\}$, for the multiple comparison and heterogeneity cases, where values of S_n are equally spaced on the interval $[n^{-0.4}, n^{-0.1}]$. For each n, rejection probabilities of the tests with nominal size equal to 0.1 are reported. Results corresponding to different nominal sizes are qualitatively similar and are not reported.

6.1 Pairwise comparisons: stationary case

We first study three data generating processes (DGPs) that are characterized by independent forecast errors and *i.i.d.* observations. These include:

DGP1: e_{1t} ~ *i.i.d.* $N(0,1)$ and e_{2t} ~ *i.i.d.* $N(0,1)$;

DGP3: e_{1t} ~ *i.i.d.* Uniform (-2, 2) and e_{2t} ~ *i.i.d.* $N(0,1)$; and

DGP5: e_{1t} ~ *i.i.d.* Beta(1,2) and e_{2t} ~ *i.i.d.* Beta(2,4), where both forecast error sequences are recentered around their common mean of 1/3. Note that the slightly unusual numbering of these DGPs is done to simplify the presentation of the results reported in Table 1.

It is easy to verify that DGP1 allows us to examine the finite sample size properties of both forecast superiority tests, and is a "least favorable" case. DGPs 3 and 5 allow us to examine the finite sample power of both tests.

In the next three DGPs, we allow for dependence. Following Klecan, McFadden, and McFadden (1991), we generate e_{kt} according to

$$e_{kt} = (1-\lambda)(\sqrt{\rho}\widetilde{e}_{0t} + \sqrt{1-\rho}\widetilde{e}_{kt}) + \lambda e_{k,t-1}, \quad k = 1, 2,$$

where $(\widetilde{e}_{0t}, \widetilde{e}_{1t}, \widetilde{e}_{2t})$ are *i.i.d.*, but have different marginals in different DGPs. The parameters $\lambda = \rho = 0.3$ determine the mutual dependence of e_{1t} and e_{2t}, and their autocorrelations. This scheme yields autocorrelated and mutually dependent forecast errors, and we consider three such DGPs. These include:

DGP2: $(\widetilde{e}_{0t}, \widetilde{e}_{1t}, \widetilde{e}_{2t})$ are *i.i.d.* $N(0, I_3)$;

DGP4: $\widetilde{e}_{1t}$ ~ *i.i.d.* $N(0, 1.5)$ and $\widetilde{e}_{kt}$ ~ *i.i.d.* $N(0,1)$, for $k = 0$ and 2; and

DGP6: $\widetilde{e}_{0t}$ ~ *i.i.d.* Beta(1,1), $\widetilde{e}_{1t}$ ~ *i.i.d.* Beta(1,2), and $\widetilde{e}_{2t}$ ~ *i.i.d.* Beta(2,4), where all forecast error sequences are recentered around their population means of 1/2, 1/3, and 1/3, respectively.

DGP2 is our "null" model, while DGPs 4 and 6 are our "alternative" models. A comparison of the simulation results based on DGP1 and DGP2 will yield insight into the effect of autocorrelation and mutual dependence on the level of the tests. Similarly, a comparison of the simulation results based on DGP5 and DGP6 will yield insight into the effect of autocorrelation and mutual dependence on the power of the tests.

Simulation results for the above DGPs are reported in Table 1. The main entries in the table are rejection frequencies. From the left panel of the table, observe that for our small sample size ($n = 100$), the GL forecast superiority test is over-sized, for some values of S_n, and has substantial power in detecting deviations from the null. Given the nature of the testing problem considered in this paper, a sample size of 100 observations is very small indeed. A comparison between the results for DGP1 (DGP2) and DGP5 (DGP6) indicates that the level (power) of the test is somewhat sensitive to the degree of mutual and serial dependence in the data, when the sample size is small. However, test power jumps to 100% as the sample size rises, and indeed both level and power are well behaved for large sample sizes, say $n = 1000$. Similar conclusions follow for the test of CL forecast superiority, as shown in the right panel of Table 1. Qualitatively similar results also obtain when DGPs are specified using other moderate values of ρ and λ. However, some non-trivial size distortions arise when these parameters are close to unity, as shown in the online supplement to this paper.

As suggested by a referee, we also report simulation results based on application of the DM test in our experiments (see Table 1 in this paper, as well as the online supplement, for complete details). These results clearly support our earlier statement that when the loss function is unknown, there is an advantage to using our approach to testing for forecast superiority, while when a specific loss function is specified (i.e. MSFE in our experiments), the DM test has better power performance than our tests, particularly for small sample sizes. This is as expected.

We also conduct Monte Carlo simulations for DGPs with parameter estimation error. In these experiments, the DGPs are the same as those examined in Corradi and Swanson (2007). Results from these simulations, which are gathered in the online supplement, are qualitatively similar to those discussed above, when the nulls are least favorable to the alternatives, while the tests are mostly under-sized when the nulls are not least favorable to the alternatives. This verifies our theory, which predicts that the stationary bootstrap works well for least favorable nulls.

6.2 Multiple comparisons: stationary case

For the sake of brevity, we consider the case of independent forecast errors and *i.i.d.* observations. Additionally, we do not include a comparison with tests that assume a given loss function, and we do not introduce parameter estimation error into our setup, as the results were found to be qualitatively similar to those reported here.

For the following eight DGPs, we fix $e_{1t} \sim i.i.d.\ N(0,1)$, but let the number of competing forecasting models vary.

DGP7: $e_{kt} \sim i.i.d.\ N(0,1)$, $k = 2,3$.

DGP8: $e_{kt} \sim i.i.d.\ N(0,1)$, $k = 2,3,4,5$.

DGP9: $e_{kt} \sim i.i.d.\ N(0,1)$, $k = 2,3,4,5$ and $e_{kt} \sim i.i.d.\ N(0,1.2^2)$, $k = 6,7,8,9$.

DGP10: $e_{kt} \sim i.i.d.\ N(0,0.8^2)$, $k = 2,3,4,5$ and $e_{kt} \sim i.i.d.\ N(0,1.2^2)$, $k = 6,7,8,9$.

DGP11: $e_{kt} \sim i.i.d.\ N(0,0.8^2)$, $k = 2,3$.

DGP12: $e_{kt} \sim i.i.d.\ N(0,0.6^2)$, $k = 2,3$.

DGP13: $e_{kt} \sim i.i.d.\ N(0,0.8^2)$, $k = 2,3,4,5$.

DGP14: $e_{kt} \sim i.i.d.\ N(0,0.6^2)$, $k = 2,3,4,5$.

Here, DGPs 7-9 are our "null" models, while DGPs 10-14 are our "alternative" models. DGPs 7 and 8 correspond to the least favorable cases in the null, and the theory of our stationary bootstrap test applies directly to this case. In DGP9, the benchmark model outperforms all of the competing models, while in DGP10, half of the competing models outperform the benchmark model, and the other half underperform.

Table 2 summarizes results for the null hypotheses, $H_0^{TG} : TG^+ \leq 0 \cap TG^- \leq 0$, where e_{1t} is taken as the benchmark forecast error. For all sample sizes in our investigation, the test has good size performance for DGPs 7 and 8, where the nulls are least favorable, while the test is mostly under-sized for DGP 9, where the nulls are not least favorable, as expected. Now, consider the power performance of the test. Interestingly, for all cases (i.e. DGPs 10-14), the test exhibits good power performance. Note also that inclusion of "poorer" models in the design improves the power of the test, as expected.

Table 3 summarizes results for the null hypotheses, $H_0^{TC} : TC^+ \leq 0 \cap TC^- \leq 0$, where e_{1t} is again taken as the benchmark forecast error. The superiority test continues to perform well. A comparison of Tables 2 and 3 indicates that the probability of correctly rejecting $H_0^{TC}$is higher than the probability of rejecting H_0^{TG}. This is consistent with the theory that GL forecast superiority implies CL forecast superiority.

6.3 Pairwise comparisons: heterogeneous case

In this subsection, we explore the case where forecast comparison is carried out for two competing sequences of heterogeneous forecast errors. We study a small set of DGPs, again for the sake of brevity. For DGPs 15 through 18, we set $e_{it} \sim a_{it}\ N(0,1)$, for $i = 1,2$ and $t \geq 1$, where $\{a_{1t}\}$ is chosen to be the infinite repetition of the sequence {1 1 1 1.25 1.25 1.25 0.75 0.75 0.75 1 1 1}, and $\{a_{2t}\}$ to be the infinite repetition of the following sequences:

DGP15: $a_{2t} = a_{1t}$;

DGP16: {1 1 1 1.25 1.25 1.25 1 1 1 0.75 0.75 0.75 };

DGP17: {1 1 1 0.65 0.65 0.65 0.5 0.5 0.5 0.8 0.8 0.8}; and

DGP18: $a_{2t} = 0.75a_{1t}$.

Clearly, the first two designs are "null" models for both GL and CL forecast superiority tests, and they are both "least favorable" to the alternatives. The last two designs are "alternative" models.

Table 4 summarizes results for the null hypotheses, $H_0^{HTG} : HTG^+ \leq 0 \cap HTG^-$ and $H_0^{HTC} : HTC^+ \leq 0 \cap HTC^-$, where e_{1t} is taken as the benchmark forecast error. Here, we utilize the block bootstrap, where the block size is chosen to be equally spaced on the interval $[2n^{0.2}, 2n^{0.4}]$. Overall, the tests are reasonably sized, despite some upward bias, when sample sizes are small. Additionally, tests based on the use of the block bootstrap seem to have good power properties, for multiple comparison of forecasting models with heterogeneous forecast errors.

Overall, it is noteworthy that the CL test in several simulation designs has marginally better power than the GL test. Needless to say, if we reject according to CL, then we also reject according to GL, but not the other way around. Nevertheless, we cannot formally rank the relative power of the tests, because the (asymptotic) distributions of the test statistics under the (local) alternatives are not comparable.

7 Empirical Illustration

In this section, forecast superiority tests are used to evaluate forecast errors resulting from two sets of forecast models for spot exchange rates among six industrialized countries. This study is for illustrative purposes only, and all forecast models are stylized, involving no estimation.

7.1 Data and models

The data consist of six 3-month-ahead forward rates and spot rates for the Canadian Dollar (CAD), French Franc (FRF), German Mark (GEM), Japanese Yen (JPY), Swiss Franc (CHF), and British Pound (GBP), relative to the US dollar. Daily data were obtained from Datastream, for the sample period from Jan. 1, 1992 through Feb. 28, 2002, at which point the euro became the sole legal tender in all euro area countries.[6] This group of countries is the same as that studied in Hunter and Timme (1992). In summary, the dataset that we analyze is comprised of 2652 observations. However, due to national holidays and a variety of other reasons, some observations are missing. If observations for a country are missing, we simply deleted them. This results in varying numbers of observations for each country, as follows: 2556 for CAD, 2620 for FRF, 2560 for GEM, 2617 for JPY, 2579 for CHF, and 2541 for GBP.

Our forecast comparison is based upon forecast errors resulting from two sets of forecasting models. We refer to the first set of forecasting models as "*Forward*" models. In these models, forward rates are used to predict future spot rates. Namely, $E_t(X_{t+\tau}) = F_{t,\tau}$, where $X_{t+\tau}$ is the spot exchange rate at

[6] We considered various different sample periods, with little change in our empirical findings.

time $t + \tau$, $F_{t,\tau}$ is the τ-period ahead forward exchange rate observed at time t, and $E_t(X_{t+\tau})$ is the expectation of the spot rate at time period $t + \tau$, conditional on information available at time t. If the "unbiasedness hypothesis" is true, given rational expectations and risk neutrality, then we should expect that the τ-period ahead forward exchange rate is the best predictor of the future spot rate, at time $t+\tau$.[7] The second set of forecasting models are termed "*Spot*" models. It is assumed that the current spot rate is the best forecast of the future spot rate. Namely, $E_t(X_{t+\tau}) = S_t$. There is a large amount of empirical support for this model, as discussed by Chiang (1986), who shows that the current spot rate is a better predictor of the future spot rate than either the forward rate, or forecasts from structural and other time series models. In a study closely related to that carried out here, Hunter and Timme (1992) base their analysis on revenues resulting from the adoption of different forecasting models, in a hedging framework, and construct first and second order stochastic dominance tests on these revenues. Here, we take the alternative approach of carrying out forecast comparison based upon the evaluation of forecast errors. In particular, we implement tests for GL forecast superiority and CL forecast superiority.

7.2 Preliminary analysis

Before conducting forecast superiority tests, we examine the empirical distribution functions (EDFs) of the forecast errors from the two models. For all six countries, the EDFs of the *Forward* forecast errors almost coincide with the *Spot* forecast errors, with some slight differences. To save space, we do not report the results here. To further examine the differences between the forward and spot EDFs for each country, we plot $G_n(x)$ and $C_n(x)$ against x, where for x, we take 200 equally spaced values between the 1% and 99% percentiles of the pooled empirical distribution of the *Forward* and *Spot* forecast errors, and where $G_n(x)$ and $C_n(x)$ are empirical analogs of $G(x)$ and $C(x)$ defined in (2.1) and (2.2). These plots are given in Figures 1 and 2, for $G_n(x)$ and $C_n(x)$, respectively. Note that both the probability differences in $G_n(x)$ and the integrated probability differences in $C_n(x)$ have been scaled up by $\sqrt{n}$, where n is the sample size. Three cases characterize these plots.

Case 1: If $G_n(x)$ $(C_n(x))$ is significantly larger than 0 for all x, we conclude that the forward model is superior to the spot model in the sense that GL (CL) outperforms the latter model.

Case 2: If $G_n(x)$ $(C_n(x))$ is significantly smaller than 0 for all x, we conclude that the spot model is superior to the forward model.

Case 3: if $G_n(x)$ $(C_n(x))$ is positive for some values of x, and negative for other values of x, GL (CL) forecast superiority may or may not exist, depending on whether the sign changes are significant.[8]

All of the plots in Figure 1 are consistent with Case 3. Thus, it is of interest to ascertain whether

[7] We follow Hunter and Timme (1992), and use the levels of the exchange rates in all of our calculations. We also tried using logarithms of the exchange rates, with similar empirical findings.

[8] Naturally, when implementing our tests, it should be stressed that sample dependence implies that a sample function that crosses zero does not rule out the possibility of a test outcome indicating strict positivity/negativity.

the sign changes are significant or not. Turning to Figure 2, Case 2 pertains to CAD, FRF, and GEM, while Case 3 pertains to JPY, CHF, and GBP. Note that the magnitude of the integrated probability differences varies substantially from one country to another. For example, the maximum of the absolute value of $C_n(x)$ for CAD is roughly $O_p\left(1/\sqrt{n}\right)$, while that for GEM is roughly $O_p(1)$. We expect that such differences will play a role in our analysis, since the tests have nontrivial power against $O(1/\sqrt{n})$ alternatives.

7.3 Tests for forecast superiority

GL and CL test statistics and critical values are constructed as outlined in Sections 3 and 6. To be specific, in computing the suprema in $G_n(x)$ and $C_n(x)$, we take the maximum over an equally spaced grid of size $\lceil 1.5n^{0.6} \rfloor$, and over a 98% range of the pooled empirical distribution. Additionally, we choose a total of twelve different values for S_n, which are equally spaced on the interval $[n^{-0.4},\ n^{-0.1}]$.

Figure 3 reports the p-values associated with testing the null hypotheses, $H^G_{0,S}$ *: Spot GL outperforms Forward* and $H^G_{0,F}$ *: Forward GL outperforms Spot.* Small p-values (say, smaller than 0.1) suggest that the corresponding null hypothesis is false. A small p-value for one test coupled with a large p-value for the other test indicates that one model is superior to the other. Turning to our findings, first consider the Canadian Dollar. The p-values based on $H^G_{0,F}$ range from 0.11 to 0.33, and the p-values based on $H^G_{0,S}$ are larger than 0.5 for all values of S_n, suggesting a failure to reject either null. Statistically speaking, the *Forward* and *Spot* perform equally well in forecasting future spot rates, in this case. Nevertheless, if one takes into account the magnitude of the p-values, one might argue that *Spot* is "better" than *Forward*. Second, despite the sign changes in $G_n(x)$ for FRF, GEM, JPY, and CHF, our tests suggest that the *Spot* GL outperforms *Forward* for all of these countries. This finding is interesting, as it supports the finding of Hunter and Timme (1992) that *Spot* outperforms *Forward* when one directly tests for first order stochastic dominance using returns. Third, as expected from our preliminary analysis, there is no GL forecast superiority in either direction, in the case of GBP.

In Figure 4, we plot the p-values associated with implementation of our CL forecast superiority tests. Examination of this figure indicates that *Spot* CL outperforms *Forward* for CAD, FRF, GEM, JPY and CHF. The tests for CL forecast superiority in the case of GBP show that there is no CL forecast superiority in either direction.

Finally, we consider the DM test. Let $\{e_{s,t}\}$ and $\{e_{f,t}\}$ be the prediction errors from *Spot* and $Forward$, respectively. Our approach is to calculate p-values associated with testing the null hypotheses, $H_0^{DM_1} : E\left(L_1(e_{s,t})\right) - E\left(L_1(e_{f,t})\right) \leq 0$ and $H_0^{DM_2} : E\left(L_2(e_{s,t})\right) - E\left(L_2(e_{f,t})\right) \leq 0$, where L_1 denotes MSFE loss, and L_2 denotes MAFE loss. Our findings indicate that *Spot* performs better than *Forward* for CAD, FRF, GEM, JPY and CHF, in terms of both MSFE and MAFE, with p-values close to one in all cases. Also, *Forward* performs better than *Spot* for GBP, in terms of MSFE, with p-values close to zero for $H_0^{DM_1}$, but not in terms of MAFE, as the p-value is 0.155 for $H_0^{DM_2}$. Thus for GBP, we are not able to determine which model "wins".

In summary, our forecast superiority test results are broadly consistent with results based on application of the conventional DM test. Moreover, our forecast superiority tests indicate that *Spot* is superior to *Forward* for *all loss functions* in the GL class, for CAD, FRF, GEM, JPY and CHF. This is important, since moment-based criteria only look in a particular direction when evaluating forecast errors, while GL and CL forecast superiority tests are based on the entire distribution of forecast errors, and do not require knowledge of the exact form of the loss function.

8 Concluding Remarks

This paper outlines a novel approach to forecast comparison that yields forecast rankings that are robust to the choice of loss function. In particular, we introduce the concepts of general-loss (GL) forecast superiority and convex-loss (CL) forecast superiority, and we establish a mapping between GL (CL) superiority and tests for first (second) order stochastic dominance. This allows us to develop a testing procedure based on an out-of-sample generalization of the tests introduced by Linton, Maasoumi and Whang (2005). The asymptotic properties of our tests are derived, under the null and under sequences of local alternatives, and it is noted that critical values cannot be tabulated. Due to this fact, we establish the first order validity of critical values that are constructed using the stationary bootstrap. Findings from a Monte Carlo study show that the suggested tests have good properties, even for moderate sample sizes. In an empirical illustration, we find that a stylized spot exchange rate forecasting model is "preferred" to a stylized forward rate model. A limitation of our testing procedure is that our statistics have non-degenerate limiting distributions only for the least favorable case, under the null. Thus, convergence is not uniform, and the tests are asymptotically conservative, in the sense of not having exact asymptotic size. Ongoing research is tackling this issue.

Appendix

In this appendix, we first define some notation and provide some technical lemmas that are used in the proof of the main results in the text. Proofs of these lemmas are collected in the online supplement for this paper.

8.1 Notation and Some Technical Lemmas

let P denote the probability measure governing the behavior of the time series $\{U_t\}$. C or $\widetilde{C}$ is a generic constant which may vary from case to case. $||\cdot||$ denotes the Euclidean norm and $||X||_q$ denotes the norm $(E|X|^q)^{1/q}$, for a random variable X. Also, $\sup_t$ denotes $\sup_{R\leq t\leq T}$ and the summation $\sum_t$ denotes $\sum_{t=R}^{T}$. Finally, "var " and "cov" denote variance and covariance, respectively. All limits are taken as $T\to\infty$. To help present the proofs of our theorems in Sections 3 and 4, we first fix some additional notation. Denote $\beta=(\beta_1',\beta_k')'$, $\beta_0=(\beta_{10}',\beta_{k0}')'$, $\widehat{\beta}_t=\left(\widehat{\beta}_{1t}',\widehat{\beta}_{kt}'\right)'$, $N(\varepsilon)=\{\beta:\|\beta-\beta_0\|\leq\varepsilon\}$ and $N_k(\varepsilon)=\{\beta_k:\|\beta_k-\beta_{k0}\|\leq\varepsilon\}$. Further, define

$$f_{k,t+\tau}^g(x,\widehat{\beta}_t)=\left(1\left(e_{k,t+\tau}\left(\widehat{\beta}_{kt}\right)\leq x\right)-1\left(e_{1,t+\tau}\left(\widehat{\beta}_{1t}\right)\leq x\right)\right)sgn(x),\text{ and}$$

$$\begin{aligned}f_{k,t+\tau}^c(x,\widehat{\beta}_t)=&\int_{-\infty}^{x}\left(1\left(e_{1,t+\tau}\left(\widehat{\beta}_{1t}\right)\leq s\right)-1\left(e_{k,t+\tau}\left(\widehat{\beta}_{kt}\right)\leq s\right)\right)ds1(x<0)\\&+\int_{x}^{\infty}\left(1\left(e_{k,t+\tau}\left(\widehat{\beta}_{kt}\right)\leq s\right)-1\left(e_{1,t+\tau}\left(\widehat{\beta}_{1t}\right)\leq s\right)\right)ds1(x\geq 0).\end{aligned}$$

Then we can write:

$$TG_n^+=\max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\sqrt{n}D_{kn}^g(x),\ TG_n^-=\max_{k=2,..,l}\sup_{x\in\mathcal{X}^-}\sqrt{n}D_{kn}^g(x)$$

and

$$TC_n^+=\max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}D_{kn}^c(x),\ TC_n^-=\max_{k=2,..,l}\sup_{x\in\mathcal{X}^-}D_{kn}^c(x),$$

where $D_{kn}^g(x)=n^{-1}\sum_t f_{k,t+\tau}^g\left(x,\widehat{\beta}_t\right)$ and $D_{kn}^c(x)=n^{-1}\sum_t f_{k,t+\tau}^c\left(x,\widehat{\beta}_t\right)$. Further, we can decompose $D_{kn}^i(x)$ as follows:

$$\begin{aligned}\sqrt{n}D_{kn}^i(x)\ &=\ n^{-1/2}\sum_t\left\{f_{k,t+\tau}^i\left(x,\widehat{\beta}_t\right)-Ef_{k,t+\tau}^i(x,\beta)|_{\beta=\widehat{\beta}_t}\right\}\\&\quad+n^{-1/2}\sum_t\left\{Ef_{k,t+\tau}^i(x,\beta)|_{\beta=\widehat{\beta}_t}-Ef_{k,t+\tau}^i(x,\beta_0)\right\}\\&\quad+n^{1/2}Ef_{k,t+\tau}^i(x,\beta_0)\\&\equiv\ \xi_{k1}^i(x)+\xi_{k2}^i(x)+\xi_{k3}^i(x),\ \text{for } i=g,\ c,\end{aligned}\tag{A.1}$$

where we suppress the dependence of $\xi_{kj}^i(\cdot)$ on n, for $j=1,2,3$. It is clear that under the null hypotheses, $\xi_{k3}^i(x)\to-\infty$, as $T\to\infty$, for $x\notin\mathcal{B}_k^i$, $i=g$, c.

For the TG_n^+ test, our objective is to show that under $H_0^{TG^+}$, for $k = 2, ..., l$,

$$\xi_{k1}^g(\cdot) \Rightarrow \widetilde{g}_k(\cdot), \text{ and} \tag{A.2}$$

$$\xi_{k2}^g(x) = \Delta_{k0}(x)' B_k v_{k0} - \Delta_{10}(x)' B_1 v_{10} + o_p(1), \text{ uniformly in } x^+. \tag{A.3}$$

Likewise for the TG_n^-, TC_n^+ and TC_n^- tests.

Now we gather together a number of lemmas that are needed for subsequent derivations.

Lemma A.1: *Suppose that Assumptions A.2 and A.4 hold and let* $\alpha \in [0, 0.5)$*. Then, for* $k = 1, ..., l$,

(a) $\sup_t \|n^\alpha H_k(t)\| \xrightarrow{p} 0$;

(b) $\sup_t \left\| n^\alpha (\widehat{\beta}_{k,t} - \beta_{k,0}) \right\| \xrightarrow{p} 0$;

(c) $\sup_t \left\| n^{1/2} H_k(t) \right\| = O_p(1)$.

Lemma A.2: *(a) Suppose that Assumption A.1 holds. Then, for all* $k = 1, ..., l$, *and for each* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that for all* $x, \dot{x} \in \mathcal{X}^-$ or $x, \dot{x} \in \mathcal{X}^+$,

$$\overline{\lim_{T\to\infty}} \left\| \sup_{\rho_g^*((x,\beta_k),(\dot{x},\dot{\beta}_k))<\delta} \left| \nu_{k,n}^g(x, \beta_k) - \nu_{k,n}^g\left(\dot{x}, \dot{\beta}_k\right) \right| \right\|_q < \varepsilon, \tag{A.4}$$

where

$$\rho_g^*\left((x, \beta_k), \left(\dot{x}, \dot{\beta}_k\right)\right) = \left\{ E\left[\left(1(e_{kt}(\beta_k) \le x) - 1\left(e_{kt}\left(\dot{\beta}_k\right) \le \dot{x}\right) \right) \right]^2 \right\}^{1/2}. \tag{A.5}$$

(b) Suppose that Assumption A.1 holds. Then, for each* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that for all* $x, \dot{x} \in \mathcal{X}^-$ or $x, \dot{x} \in \mathcal{X}^+$,

$$\overline{\lim_{T\to\infty}} \left\| \sup_{\rho_c^*((x,\beta_k),(\dot{x},\dot{\beta}_k))<\delta} \left| \nu_{k,n}^c(x, \beta_k) - \nu_{k,n}^c\left(\dot{x}, \dot{\beta}_k\right) \right| \right\|_q < \varepsilon, \tag{A.6}$$

where

$$\rho_c^*((x, \beta_k), (\dot{x}, \dot{\beta}_k)) = \left\{ E\left| \int_{-\infty}^{x} 1(e_{k,t}(\beta_k) \le s) ds - \int_{-\infty}^{\dot{x}} 1(e_{k,t}((\dot{\beta}_k) \le s) ds \right|^r \right\}^{1/r} 1(x < 0,\ \dot{x} < 0)$$
$$+ \left\{ E\left| \int_{x}^{\infty} 1(e_{k,t}(\beta_k) > s) ds - \int_{\dot{x}}^{\infty} 1(e_{k,t}(\dot{\beta}_k) > s) ds \right|^r \right\}^{1/r} 1(x \ge 0,\ \dot{x} \ge 0). \tag{A.7}$$

Lemma A.3: *Suppose that Assumptions A.1, A.1*, and A.4 hold. Denote* $\zeta_{k,t+\tau}^i(x, \beta) = f_{k,t+\tau}^i(x, \beta) - Ef_{k,t+\tau}^i(x, \beta) - f_{k,t+\tau}^i(x, \beta_0) + Ef_{k,t+\tau}^i(x, \beta_0)$, $i = g, c$. Then, for $k = 2, ..., l$,

(a)

$$\sup_t E \sup_{\{\beta\} \in N(n^{-\alpha}\varepsilon)} \sup_{x \in \mathcal{X}^+} [\zeta_{k,t+\tau}^i(x, \beta)]^2 \le Cn^{-\alpha}\varepsilon,$$
$$\sup_t E \sup_{\{\beta\} \in N(n^{-\alpha}\varepsilon)} \sup_{x \in \mathcal{X}^-} [\zeta_{k,t+\tau}^i(x, \beta)]^2 \le Cn^{-\alpha}\varepsilon,\ i = g, c$$

(b)

$$\sup_t |E \sup_{\{\beta,\ \dot{\beta}\}\in N(n^{-\alpha}\varepsilon)} \sup_{x\in\mathcal{X}^+} \zeta^i_{k,t+\tau}(x,\beta)\zeta^i_{k,t+\tau+j}(x,\dot{\beta})| \leq \widetilde{C}\alpha(j)^d(n^{-\alpha}\varepsilon)^2,$$
$$\sup_t |E \sup_{\{\beta,\ \dot{\beta}\}\in N(n^{-\alpha}\varepsilon)} \sup_{x\in\mathcal{X}^-} \zeta^i_{k,t+\tau}(x,\beta)\zeta^i_{k,t+\tau+j}(x,\dot{\beta})| \leq \widetilde{C}\alpha(j)^d(n^{-\alpha}\varepsilon)^2,$$

where $d = 1$ and $\delta/(2+\delta)$ for $i = g$ and c, respectively.

Lemma A.4: *(a) Suppose that Assumptions A.1-A.4 hold. Then, we have for $k = 1, ..., l$,*

$$\sup_{x\in\mathcal{X}^+} |\xi^g_{k1}(x) - \nu^g_{k,n}(x,\beta_{k0}) + \nu^g_{1,n}(x,\beta_{1,0})| \xrightarrow{p} 0, \tag{A.8}$$
$$\sup_{x\in\mathcal{X}^-} |\xi^g_{k1}(x) - \nu^g_{k,n}(x,\beta_{k0}) + \nu^g_{1,n}(x,\beta_{1,0})| \xrightarrow{p} 0$$

(b) Suppose that Assumptions A.1, A.2, A.3* and A.4 hold. Then, we have for $k = 1, ..., l$,*

$$\sup_{x\in\mathcal{X}^+} |\xi^c_{k1}(x) - \nu^c_{k,n}(x,\beta_{k0}) + \nu^c_{1,n}(x,\beta_{10})| \xrightarrow{p} 0, \tag{A.9}$$
$$\sup_{x\in\mathcal{X}^-} |\xi^c_{k1}(x) - \nu^c_{k,n}(x,\beta_{k0}) + \nu^c_{1,n}(x,\beta_{10})| \xrightarrow{p} 0.$$

Lemma A.5: *(a) Suppose that Assumptions A.1-A.4 hold. Then, we have for $k = 1, ..., l$,*

$$\sup_{x\in\mathcal{X}^+} \left|\xi^g_{k2}(x) - \sqrt{n}\Delta'_{k0}(x)B_k\overline{H}_{k,n} + \sqrt{n}\Delta'_{10}(x)B_1\overline{H}_{1,n}\right| = o_p(1), \tag{A.10}$$
$$\sup_{x\in\mathcal{X}^-} \left|\xi^g_{k2}(x) - \sqrt{n}\Delta'_{k0}(x)B_k\overline{H}_{k,n} + \sqrt{n}\Delta'_{10}(x)B_1\overline{H}_{1,n}\right| = o_p(1). \tag{A.11}$$

(b) Suppose that Assumptions A.1, A.2, A.3* and A.4 hold. Then, we have for $k = 1, ..., l$,*

$$\sup_{x\in\mathcal{X}^+} \left|\xi^c_{k2}(x) - \sqrt{n}\Lambda'_{k0}(x)B_k\overline{H}_{k,n} + \sqrt{n}\Lambda'_{10}(x)B_1\overline{H}_{1,n}\right| = o_p(1), \tag{A.12}$$
$$\sup_{x\in\mathcal{X}^-} \left|\xi^c_{k2}(x) - \sqrt{n}\Lambda'_{k0}(x)B_k\overline{H}_{k,n} + \sqrt{n}\Lambda'_{10}(x)B_1\overline{H}_{1,n}\right| = o_p(1). \tag{A.13}$$

Lemma A.6: *(a) Suppose that Assumptions A.1-A.4 hold. Then, we have for $k = 2, ..., l$,*

$$\begin{pmatrix} v^g_{k,n}(\cdot,\beta_{k,0}) - v^g_{1,n}(\cdot,\beta_{1,0}) \\ \sqrt{n}\overline{H}_{k,n} \\ \sqrt{n}\overline{H}_{1,n} \end{pmatrix} \Rightarrow \begin{pmatrix} \widetilde{g}_k(\cdot) \\ v_{k0} \\ v_{10} \end{pmatrix}$$

and except at zero, the sample paths of $\widetilde{g}_k(\cdot)$ are uniformly continuous with respect to a pseudometric ρ_g on X with probability one, where for $x_1, x_2 \in X^+$ or $x_1, x_2 \in X^-$,

$$\rho_g(x_1,x_2) = \left\{E[(1(e_{1,t} \leq x_1) - 1(e_{k,t} \leq x_1)) - (1(e_{1,t} \leq x_2) - 1(e_{k,t} \leq x_2))]^2\right\}^{1/2}.$$

(b)Suppose Assumptions A.1, A.2, A.3* and A.4 hold. Then, we have for* $k = 2, ..., l$,

$$\begin{pmatrix} v^c_{k,n}(\cdot, \beta_{k0}) - v^c_{1,n}(\cdot, \beta_{10}) \\ \sqrt{n}\overline{H}_{k,n} \\ \sqrt{n}\overline{H}_{1,n} \end{pmatrix} \Rightarrow \begin{pmatrix} \widetilde{c}_k(\cdot) \\ v_{k0} \\ v_{10} \end{pmatrix}$$

and except at zero, the sample paths of $\widetilde{c}_k(\cdot)$ *are uniformly continuous with respect to a pseudometric* ρ_c *on* X *with probability one, where for* $x_1, x_2 \in X^+$ *or* $x_1, x_2 \in X^-$,

$$\rho_c(x_1, x_2) = \left\{ E\left| \int_{-\infty}^{x_1} (1(e_{1,t} \le s) - 1(e_{k,t} \le s))ds - \int_{-\infty}^{x_2} (1(e_{1,t} \le s) - 1(e_{k,t} \le s))ds \right|^r \right\}^{1/r} 1(x_1 < 0, x_2 < 0)$$

$$+ \left\{ E\left| \int_{x_1}^{\infty} (1(e_{1,t} > s) - -1(e_{k,t} > s)))ds - \int_{x_2}^{\infty} (1(e_{1,t} > s) - 1(e_{k,t} > s)))ds \right|^r \right\}^{1/r} 1(x_1 \ge 0, x_2 \ge 0).$$

Lemma HA.1: *(a) Suppose that Assumption HA.1 holds. Then, for each* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that for all* $x, \dot{x} \in \mathcal{X}^+$ or $x, \dot{x} \in \mathcal{X}^-$,

$$\overline{\lim_{T\to\infty}} \left\| \sup_{\rho^*_{hg}(x,\dot{x}))<\delta} |\nu^{hg}_{k,n}(x) - \nu^{hg}_{k,n}(\dot{x})| \right\|_q < \varepsilon, \quad \text{(A.14)}$$

where

$$\rho^*_{hg}(x, \dot{x}) = \{E[1(e_{k,t+\tau} \le x) - 1(e_{k,t+\tau} \le \dot{x})]^2\}^{1/2}. \quad \text{(A.15)}$$

(b) Suppose that Assumption HA.1 holds. Then, for each* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that for all* $x, \dot{x} \in \mathcal{X}^+$ or $x, \dot{x} \in \mathcal{X}^-$,

$$\overline{\lim_{T\to\infty}} \left\| \sup_{\rho^*_{hc}(x,\dot{x})<\delta} |\nu^{hc}_{k,n}(x) - \nu^{hc}_{k,n}(\dot{x})| \right\|_q < \varepsilon, \quad \text{(A.16)}$$

where

$$\rho^*_{hc}(x, \dot{x}) = \left\{ E\left| \int_{-\infty}^{x} 1(e_{k,t+\tau} \le s)ds - \int_{-\infty}^{\dot{x}} 1(e_{k,t+\tau} \le s)ds \right|^r \right\}^{1/r} 1(x < 0,\ \dot{x} < 0)$$

$$+ \left\{ E\left| \int_{x}^{\infty} 1(e_{k,t+\tau} > s)ds - \int_{\dot{x}}^{\infty} 1(e_{k,t+\tau} > s)ds \right|^r \right\}^{1/r} 1(x \ge 0,\ \dot{x} \ge 0). \quad \text{(A.17)}$$

Lemma HA.2: *(a) Suppose that Assumptions HA.1* and HA.4 hold. Then, we have for* $k = 2, ..., l$,

$$v^{hg}_{k,n}(\cdot) - v^{hg}_{1,n}(\cdot) \Rightarrow \widetilde{hg}_k(\cdot)$$

and except at zero, the sample paths of $\widetilde{hg}_k(\cdot)$ are uniformly continuous with respect to a pseudometric ρ_{hg} on $\mathcal{X}$ with probability one, where for $x_1, x_2 \in \mathcal{X}^+$ or $x_1, x_2 \in \mathcal{X}^-$,

$$\rho_{hg}(x_1, x_2) = \left\{ E[(1(e_{1,t+\tau} \le x_1) - 1(e_{k,t+\tau} \le x_1)) - (1(e_{1,t+\tau} \le x_2) - 1(e_{k,t+\tau} \le x_2))]^2 \right\}^{1/2}.$$

(b) Suppose that Assumptions A.1, A.2, A.3* and A.4 hold. Then, we have for* $k = 2, ..., l$,

$$v_{k,n}^{hc}(\cdot) - v_{1,n}^{hc}(\cdot) \Rightarrow \widetilde{hc}_k(\cdot)$$

and except at zero, the sample paths of $\widetilde{hc}_k(\cdot)$ *are uniformly continuous with respect to a pseudometric* ρ_{hc} *on* X *with probability one, where for* $x_1, x_2 \in X^+$ *or* $x_1, x_2 \in X^-$,

$$\begin{aligned}
&\rho_{hc}(x_1, x_2) \\
= &\left\{ E \left| \int_{-\infty}^{x_1} (1(e_{1,t+\tau} \le s) - 1(e_{k,t+\tau} \le s))ds - \int_{-\infty}^{x_2} (1(e_{1,t+\tau} \le s) - 1(e_{k,t+\tau} \le s))ds \right|^r \right\}^{1/r} 1(x_1 < 0, x_2 < 0) \\
+ &\left\{ E \left| \int_{x_1}^{\infty} (1(e_{1,t+\tau} > s) - 1(e_{k,t+\tau} > s)))ds - \int_{x_2}^{\infty} (1(e_{1,t+\tau} > s) - 1(e_{k,t+\tau} > s)))ds \right|^r \right\}^{1/r} 1(x_1 \ge 0, x_2 \ge 0).
\end{aligned}$$

8.2 Proofs of the Main Results

Proof of Proposition 2.2: Let f_1 and f_2 be the densities associated with F_1 and F_2. We first prove the IF part of the statement of the theorem. Note that:

$$\begin{aligned}
& \int_{-\infty}^{\infty} L(z)\,(f_1(z) - f_2(z))\,\mathrm{d}z \\
= \; & \int_{-\infty}^{0} L(z)\,(f_1(z) - f_2(z))\,\mathrm{d}z + \int_{0}^{\infty} L(z)\,(f_1(z) - f_2(z))\,\mathrm{d}z \\
= \; & L(z)\,(F_1(z) - F_2(z))\,|_{-\infty}^{0} + L(z)\,(F_1(z) - F_2(z))\,|_{0}^{\infty} \\
& - \int_{-\infty}^{0} L'(z)\,(F_1(z) - F_2(z))\,\mathrm{d}z - \int_{0}^{\infty} L'(z)\,(F_1(z) - F_2(z))\,\mathrm{d}z \\
= \; & - \int_{-\infty}^{0} L'(z)\,(F_1(z) - F_2(z))\,\mathrm{d}z - \int_{0}^{\infty} L'(z)\,(F_1(z) - F_2(z))\,\mathrm{d}z \\
\le \; & 0,
\end{aligned}$$

if $(F_2(z) - F_1(z))\, sgn(z) \le 0$.

We now prove the ONLY IF part of the statement of the theorem. Suppose that the two CDFs cross once. Let $\Delta_1 = (-\infty, \underline{x}]$, with $\underline{x} < 0$, and $\Delta_2 = [\overline{x}, \infty)$, with $\overline{x} > 0$. There are two possible cases, either $F_1(x) - F_2(x) > 0$, for all $x \in \Delta_1$, or $F_1(x) - F_2(x) < 0$, for $x \in \Delta_2$. Suppose that we are in the latter case. Let $L(x) = -a_1 x 1\{x \le 0\} + a_2 x 1\{0 \le x < \overline{x}\} + a_3 x 1\{x \ge \overline{x}\}$, with $a_1, a_2, a_3 > 0$, and $a_3 - \max\{a_1, a_2\} > \delta_{\Delta_2} > 0$. Note that $\widetilde{L}$ satisfies Assumption A.0. Now, for any Δ_2, there exists a δ_{Δ_2} sufficiently large, such that:

$$\begin{aligned}
& -a_3 \int_{\Delta_2} (F_1(z) - F_2(z))\,\mathrm{d}z \\
> \; & -a_1 \int_{0}^{\infty} (F_1(z) - F_2(z))\,\mathrm{d}z + a_2 \int_{0}^{\overline{x}} (F_1(z) - F_2(z))\,\mathrm{d}z.
\end{aligned}$$

Proof of Proposition 2.3: We first prove the IF part of the statement of the theorem. From the proof of Proposition 2.2, and by further integration by parts:

$$\begin{aligned}
& \int_{-\infty}^{\infty} L(z)\left(f_1(z)-f_2(z)\right) \mathrm{d} z \\
= & -\int_{-\infty}^{0} L^{\prime}(z)\left(F_1(z)-F_2(z)\right) \mathrm{d} z-\int_0^{\infty} L^{\prime}(z)\left(F_1(z)-F_2(z)\right) \mathrm{d} z \\
= & -L^{\prime}(z) \int_{-\infty}^{z}\left(F_1(t)-F_2(t)\right) \mathrm{d} t\,\Big|_{-\infty}^{0}+\int_{-\infty}^{0} L^{\prime \prime}(z)\left(\int_{-\infty}^{z}\left(F_1(t)-F_2(t)\right) \mathrm{d} t\right) \mathrm{d} z \\
& +L^{\prime}(z) \int_z^{\infty}\left(F_1(t)-F_2(t)\right) \mathrm{d} t\,\Big|_0^{\infty}-\int_0^{\infty} L^{\prime \prime}(z)\left(\int_z^{\infty}\left(F_1(t)-F_2(t)\right) \mathrm{d} t\right) \mathrm{d} z \\
= & \int_{-\infty}^{0} L^{\prime \prime}(z)\left(\int_{-\infty}^{z}\left(F_1(t)-F_2(t)\right) \mathrm{d} t\right) \mathrm{d} z-\int_0^{\infty} L^{\prime \prime}(z)\left(\int_z^{\infty}\left(F_1(t)-F_2(t)\right) \mathrm{d} t\right) \mathrm{d} z \\
\leq & \ 0,
\end{aligned}$$

since $\int_{-\infty}^{z}\left(F_1(t)-F_2(t)\right) dt \leq 0$, for all $z \leq 0$, and $\int_z^{\infty}\left(F_1(t)-F_2(t)\right) dt \geq 0$, for all $z \geq 0$.

We now prove the ONLY IF part of the statement of the theorem. Let Δ_1 and Δ_2 be defined as in the proof of Proposition 2.2. Suppose that $\int_z^{\infty}\left(F_1(t)-F_2(t)\right) dt < 0$, for all $z \in \Delta_2$. Let $L(z) = a_1 z^2 1\{z \leq 0\} + a_2 z^2 1\{0 < z < \overline{z}\} + a_3 z^2 1\{z > \overline{z}\}$, with $a_1, a_{2,} a_3 > 0$, and $a_3 - \max\{a_1, a_2\} > \delta_{\Delta_2}$. Now, for any Δ_2 there exists a δ_{Δ_2} sufficiently large, such that:

$$\begin{aligned}
& -2 a_3 \int_{\Delta^2}\left(\int_z^{\infty}\left(F_1(t)-F_2(t)\right) \mathrm{d} t\right) \mathrm{d} z \\
> & \ 2 a_2 \int_0^{\overline{z}}\left(\int_z^{\infty}\left(F_1(t)-F_2(t)\right) \mathrm{d} t\right) \mathrm{d} z-2 a_1 \int_{-\infty}^{0}\left(\int_{-\infty}^{z}\left(F_1(t)-F_2(t)\right) \mathrm{d} t\right) \mathrm{d} z .
\end{aligned}$$

Proof of Theorem 3.1: WLOG, we consider the case where $x \geq 0$. To prove part (a), note that:

$$TG_n = \max_{k=2,..,l} \sup_{x \in \mathcal{X}^+} \sqrt{n} D_{k,n}^g(x) = \max_{k=2,..,l} \sup_{x \in \mathcal{X}^+} \left\{\xi_{k1}^g(x) + \xi_{k2}^g(x) + \xi_{k3}^g(x)\right\}.$$

Recall that $TG^+ = 0$ implies that the set $\mathcal{B}_k^{g+}$ is not empty and under the null, and $\xi_{k3}^g(x) = n^{1/2}\left(F_k(x) - F_1(x)\right) sgn(x) \rightarrow -\infty$ for all $x \notin \mathcal{B}_k^{g+}$. Consequently,

$$\begin{aligned}
TG_{k,n}^+ & \equiv \sup_{x \in \mathcal{X}^+} \sqrt{n} D_{k,n}^g(x) \\
& \Rightarrow \sup_{x \in \mathcal{B}_k^{g+}} \left[\widetilde{g}_k(x) + \Delta_{k0}(x)' B_k v_{k0} - \Delta_{10}(x)' B_1 v_{10}\right],
\end{aligned}$$

by Lemmas A.4(a) through A.6(a). The result follows from the Continuous Mapping Theorem (CMT).

Suppose that $TG^+ < 0$. In this case, the set $\mathcal{B}_k^{g+}$ is empty and hence $n^{-1/2} \xi_{k3}^g(x) < 0\ \forall x \in \mathcal{X}^+$, for some $k \in \{2, ..., l\}$. Then $D_{k,n}^g(x)$ will be dominated by the term $\xi_{k3}^g(x)$, which diverges to minus infinity, for any $x \in \mathcal{X}^+$, as required.

To prove part (b), note that

$$TC_n^+ = \max_{k=2,..,l} \sup_{x \in \mathcal{X}^+} D_{k,n}^c(x) = \min_{k=2,..,l} \sup_{x \in \mathcal{X}^+} \left\{\xi_{k1}^c(x) + \xi_{k2}^c(x) + \xi_{k3}^c(x)\right\}.$$

If $TC^+ = 0$, the set $\mathcal{B}_k^{c+}$ is not empty and under the null, and $\xi_{k3}^c(x) \to -\infty$, for all $x \notin \mathcal{B}_k^{c+}$. Consequently,

$$\begin{aligned} TC_{k,n}^+ &\equiv \sup_{x\in\mathcal{X}^+} D_{k,n}^c(x) \\ &\Rightarrow \sup_{x\in\mathcal{B}_k^{c+}} \left[\widetilde{g}_k(x) + \Lambda_{k0}(x)' B_k v_{k0} - \Lambda_{10}(x)' B_1 v_{10}\right], \end{aligned}$$

by Lemmas A.4(b) through A.6(b). The result follows from the CMT.

Next suppose that $TC^+ < 0$. In this case, the set $\mathcal{B}_k^{c+}$ is empty and hence $n^{-1/2}\xi_{k3}^c(x) < 0\ \forall x \in \mathcal{X}^+$, for some $k \in \{2, ..., l\}$. Then, $D_{k,n}^c(x)$ will be dominated by the term $\xi_{k3}^c(x)$, which diverges to minus infinity, for any $x \in \mathcal{X}^+$, as required. The conclusion thus follows.

Proof of Theorem 3.2: Adding and subtracting appropriately yields:

$$\begin{aligned} \sqrt{n}(G_{k,n}^*(x) - G_{k,n}(x)) &= n^{-1/2}\sum_t \left[f_{k,\theta(t)+\tau}^g(x, \widehat{\beta}_{k,\theta(t)}) - f_{k,t+\tau}^g(x, \widehat{\beta}_{kt})\right] \\ &= n^{-1/2}\sum_t \left[f_{k,\theta(t)+\tau}^g(x, \beta_{k0}) - f_{k,t+\tau}^g(x, \beta_{k0})\right] \\ &\quad -n^{-1/2}\sum_t \left[f_{k,t+\tau}^g(x, \widehat{\beta}_{kt}) - f_{k,t+\tau}^g(x, \beta_{k0})\right] \\ &\quad +n^{-1/2}\sum_t \left[f_{k,\theta(t)+\tau}^g(x, \widehat{\beta}_{k,\theta(t)}) - f_{k,\theta(t)+\tau}^g(x, \beta_{k0})\right] \\ &\equiv \varsigma_{1,n}(x) - \varsigma_{2,n}(x) + \varsigma_{3,n}(x). \end{aligned}$$

Under Assumptions A.1-A.4 and A.6, Theorem 3.1 of Politis and Romano (1994) applies, yielding $\rho(\mathbf{L}[\varsigma_{1n}(\cdot)|U_1, ..., U_{T+\tau}], \mathbf{L}[G_{kn}(\cdot) - G_k(\cdot)]) \xrightarrow{p} 0$. Also,

$$\begin{aligned} \varsigma_{2,n}(x) &= n^{-1/2}\sum_t \left[Ef_{k,t+\tau}^g(x, \beta_k)|_{\beta_k=\widehat{\beta}_{kt}} - Ef_{k,t+\tau}^g(x, \beta_{k0})\right] \\ &\quad +n^{-1/2}\sum_t \left[f_{k,t+\tau}^g(x, \widehat{\beta}_{kt}) - Ef_{k,t+\tau}^g(x, \beta_k)|_{\beta_k=\widehat{\beta}_{kt}} - f_{k,t+\tau}^g(x, \beta_{k0}) + Ef_{k,t+\tau}^g(x, \beta_{k0})\right] \\ &= o_p(1) + o_p(1) = o_p(1), \text{ uniformly in } x \in \mathcal{X}^+, \text{ or in } x \in \mathcal{X}^-, \end{aligned}$$

where the second equality follows from Lemmas A.4(a) and A.5(a). The result follows if $P[\sup_{x\in\mathcal{X}^+} \varsigma_{3,n}(x) = o_Q(1)] \to 1$, as n increases, where Q is the probability distribution induced by the stationary bootstrap, conditional on the data $(U_1, ..., U_{T+\tau})$. Note that $\varsigma_{3,n}(x) = \varsigma_{3,n}^k(x) - \varsigma_{3,n}^1(x)$, where $\varsigma_{3,n}^k(x) = n^{-1/2}\sum_t \left[1\left(e_{k,\theta(t)+\tau}\left(\widehat{\beta}_{k,\theta(t)}\right) \le x\right) - 1(e_{k,t+\tau} \le x)\right]$, and $\sup_{x\in\mathcal{X}^+} |\varsigma_{3,n}(x)| \le \sup_x \left|\varsigma_{3,n}^k(x)\right| + \sup_x \left|\varsigma_{3,n}^1(x)\right|$. By the Markov inequality, it suffices to show that $E_Q|\sup_x \varsigma_{3n}^k(x)| = o_p(1)$, where E_Q is the expectation induced by the probability measure Q. Note that:

$$\begin{aligned}
& E_Q \left| \sup_{x \in \mathcal{X}^+} \varsigma_{3,n}^k(x) \right| \\
= & \left| \sup_{x \in \mathcal{X}^+} n^{-1/2} \sum_t \left\{ 1\left(e_{k,t+\tau}\left(\widehat{\beta}_{k,t}\right) \leq x \right) - 1(e_{k,t+\tau} \leq x) \right\} \right| \\
\leq & n^{-1/2} \sum_t \left| \sup_{x \in \mathcal{X}^+} \left\{ 1\left(e_{k,t+\tau}\left(\widehat{\beta}_{k,t}\right) \leq x \right) - 1(e_{k,t+\tau} \leq x) \right\} \right| \\
\leq & n^{-1/2} \sum_t \sup_{x \in \mathcal{X}^+} 1\left(|e_{k,t+\tau} - x| \leq \left| m_k\left(Z_{k,t+\tau}, \widehat{\beta}_{kt} \right) - m_k(Z_{k,t+\tau}, \beta_{k0}) \right| \right) \\
\equiv & \varsigma_n .
\end{aligned}$$

It suffices to show that $E\left[\varsigma_n\right] = o(1)$, by the Markov inequality and the nonnegativity of ς_n. Denote the jth elements of $\widehat{\beta}_{k,t}$ and β_{k0} as $\widehat{\beta}_{k,t}^{(j)}$ and $\beta_{k0}^{(j)}$, respectively. By Assumption A.6, for all j, $\sup_t \left| \widehat{\beta}_{k,t}^{(j)} - \beta_{k0}^{(j)}) \right| \leq R^{-1/2}\sigma_j (\log\log R\sigma_j)^{1/2}$, a.s., where σ_j is the jth diagonal element of Σ_k. The assumption $(n/R)(\log\log R) = o(1)$ trivially ensures that $\max_j \sup_t \left| n^{1/2}\left(\widehat{\beta}_{k,t}^{(j)} - \beta_{k0}^{(j)} \right) \right| = o_{a.s.}(1)$. Fix ϵ_0, $\delta > 0$. Then for all $\epsilon > 0$, there exists a T_0 such that for all $T > T_0$, $P\left[\max_j \sup_t \left| n^{1/2}\left(\widehat{\beta}_{k,t}^{(j)} - \beta_{k0}^{(j)} \right) \right| > \varepsilon \right] < \delta/2$. It is useful to note that for all $T > T_0$ and all $\varepsilon_0 > 0$,

$$\begin{aligned}
& P\left[\varsigma_n \geq \varepsilon_0\right] \\
\leq & P\left[\sup_{\{\beta_{k,t}\} \in N_k(n^{-1/2}\varepsilon)} n^{-1/2} \sum_t \sup_{x \in \mathcal{X}^+} 1\left(|e_{k,t+\tau} - x| \leq |m_k(Z_{k,t+\tau}, \beta_{k,t}) - m_k(Z_{k,t+\tau}, \beta_{k0})| \right) \geq \varepsilon_0 \right] \\
& + P\left[\max_j \sup_t \left| n^{1/2}\left(\widehat{\beta}_{k,t}^{(j)} - \beta_{k0}^{(j)} \right) \right| > \varepsilon \right] \\
\leq & \psi_n + \frac{\delta}{2},
\end{aligned}$$

where $\{\beta_{k,t}\} \equiv \{\beta_{k,t}\}_{t=R}^T$ is a nonrandom sequence and

$$\psi_n = P\left[\sup_{\{\beta_{k,t}\} \in N_k(n^{-1/2}\varepsilon)} n^{-1/2} \sum_t \sup_{x \in \mathcal{X}^+} 1\left(|e_{k,t+\tau} - x| \leq |m_k(Z_{k,t+\tau}, \beta_{k,t}) - m_k(Z_{k,t+\tau}, \beta_{k0})| \right) \geq \varepsilon_0 \right].$$

The remainder of this proof requires showing that there exists a $T_1 > T_0$, such that for all $T > T_1$,

$\psi_n < \delta/2$. Applying the Markov inequality, we have that:

$$
\begin{aligned}
& \varepsilon_0 \psi_n \\
\leq\; & E\left[\sup_{\{\beta_{k,t}\}\in N_k(n^{-1/2}\varepsilon)} n^{-1/2}\sum_t \sup_{x\in\mathcal{X}^+} 1\left(|e_{k,t+\tau}-x| \leq |m_k(Z_{k,t+\tau},\beta_{k,t})-m_k(Z_{k,t+\tau},\beta_{k0})|\right)\right] \\
\leq\; & E\left[n^{-1/2}\sum_t \sup_{\{\beta_{kt}\}\in N_k(n^{-1/2}\varepsilon)} \sup_{x\in\mathcal{X}^+} 1\left(|e_{k,t+\tau}-x| \leq |m_k(Z_{k,t+\tau},\beta_{k,t})-m_k(Z_{k,t+\tau},\beta_{k0})|\right)\right] \\
\leq\; & \widetilde{C} \sup_{\{\beta_{kt}\}\in N_k(n^{-1/2}\varepsilon)} \sup_{x\in\mathcal{X}^+} E\left|n^{1/2}\left(m_k(Z_{k,t+\tau},\beta_{k,t})-m_k(Z_{k,t+\tau},\beta_{k0})\right)\right| \\
\leq\; & \widetilde{C} \sup_{\beta_k\in\Theta_{k0}} ||M_k(Z_{k,t},\beta_k)||_2\varepsilon \\
=\; & C\varepsilon, \text{ say,}
\end{aligned}
$$

where the last inequality holds by Assumption A.1 (ii). Thus, we can choose T_1 and ε such that for all $T > T_1$, $\varepsilon < (\delta\varepsilon_0/2C)$. The result follows.

Proof of Corollary 3.3: This corollary follows immediately from Theorem 3.2 and application of the CMT.

Proof of Theorem 4.1: WLOG, we consider the case $x \geq 0$. Recall that:

$$TG_n^+ = \max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\sqrt{n}D_{k,n}^g(x) = \max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\left\{\xi_{k1}^g(x)+\xi_{k2}^g(x)+\xi_{k3}^g(x)\right\}.$$

If $TG^+ > 0$, Lemmas A.4(a)-A.6(a) continue to hold, so that $\xi_{k1}^g(x) = O_p(1)$, uniformly in $x \in \mathcal{X}^+$, and $\xi_{k2}^g(x) = o_p(1)$, uniformly in $x \in \mathcal{X}^+$. For each $k \in \{2, ..., l\}$, $\xi_{k3}^g(x) = n^{1/2}\ (F_k(x) - F_1(x))sgn(x) \to \infty$, for some $x \in \mathcal{X}^+$. Consequently, $TG_n^+ \xrightarrow{p} \infty$, as $T \to \infty$ and $n^{-1/2}TG_n^+ \xrightarrow{p} TG^+ > 0$.

Now, from Corollary 3.3:

$$\rho(\mathbf{L}[\max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\sqrt{n}(G_{k,n}^*(x)-G_{k,n}(x))|U_1,...,U_{T+\tau}], \mathbf{L}[\max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\sqrt{n}\left(G_{k,n}(x)-G_k(x)\right)]) \xrightarrow{p} 0,$$

which implies that $q_{n,S_n}^{G+}(1-\alpha) = \widetilde{q}_{n,S_n}^{G+}(1-\alpha) + o_p(1)$, where $\widetilde{q}_{n,S_n}^{G+}(1-\alpha)$ is the $(1-\alpha)$ -th sample quantile of $\widetilde{TG}^+ \equiv \max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\sqrt{n}\left(G_{k,n}(x) - G_k(x)\right)$. Furthermore,

$$
\begin{aligned}
\widetilde{TG}^+ &= \max_{k=2,..,l}\max\{\sup_{x\in\mathcal{B}_k^q+}\sqrt{n}G_{k,n}(x), \sup_{x\in\mathcal{X}^+\backslash\mathcal{B}_k^{q+}}\sqrt{n}(G_{k,n}(x)-G_k(x))\} \\
&\leq \max_{k=2,..,l}\sup_{x\in\mathcal{B}_k^{q+}}\sqrt{n}G_{k,n}(x) + \max_{k=2,..,l}\sup_{x\in\mathcal{X}^+\backslash\mathcal{B}_k^{q+}}\sqrt{n}(G_{k,n}(x)-G_k(x))\} \\
&\equiv \widetilde{TG}_{1,n}^+ + \widetilde{TG}_{2,n}^+.
\end{aligned}
$$

Here, $\widetilde{TG}_{1,n}^+ = O_p(1)$, and has the limiting distribution given in (3.3), by the proof of Theorem 3.1 (a).

Also, $\widetilde{TG}_{2,n}^{+} = o_p(1)$, by the proof of Theorem 3.2. Consequently, $\widetilde{q}^{G+}_{n,S_n}(1-\alpha) \xrightarrow{p} q^{G+}(1-\alpha)$, and,

$$\begin{aligned} P\left(TG_n^+ > q^{G+}_{n,S_n}(1-\alpha)\right) &= P\left(TG_n^+ > q^{G+}(1-\alpha) + o_p(1)\right) \\ &= P\left(n^{-1/2}TG_n^+ > n^{-1/2}q^{G+}(1-\alpha)\right) + o_p(1) \\ &= P\left(TG^+ > n^{-1/2}q^{G+}(1-\alpha)\right) + o_p(1) \\ &\rightarrow 1. \end{aligned}$$

Proof of Theorem 4.2: The proof of this theorem is similar to that of Theorem 3.1. Consider Lemmas A.1-A.6, with $v^g_{k,n}(x,\beta_k)$ now defined by:

$$v^g_{k,n}(x,\beta_k) = n^{-1/2}\sum_t \left[1\left(e_{k,t+\tau}(\beta_k) \leq x\right) - F_{k,n}(x,\beta_k)\right] sgn(x) \text{ for } k = 1, ..., l.$$

Then by contiguity, the result of Lemma A.4(a) holds, under the local alternatives. Lemma A.5(a) now changes to $\sup_{x\in\mathcal{X}^+} |\xi^g_{k2}(x) - \sqrt{n}\Delta'_{k0}(x)B_k\overline{H}_{k,n} + \sqrt{n}\Delta'_{10}(x)B_1\overline{H}_{1,n}| = o_p(1)$, because, WLOG:

$$\begin{aligned} &\sup_{x\in\mathcal{X}^+}\left|n^{-1/2}\sum_t\left[F_{k,n}(x,\widehat{\beta}_{kt}) - F_k(x,\beta_{k0}) - \sqrt{n}\Delta'_{k0}(x)B_k\overline{H}_{k,n}\right]\right| \\ = &\sup_{x\in\mathcal{X}^+}\left|n^{-1/2}\sum_t\left(\frac{\partial F_{k,n}(x,\beta^*_{k,t}(x))}{\partial\beta'_k}\right)\left(\widehat{\beta}_{k,t} - \beta_{k0}\right) - n^{1/2}\Delta_{k0}(x)'B_k\overline{H}_{k,n}\right| \\ \leq &\sup_{x\in\mathcal{X}^+}\left|n^{-1/2}\sum_t\left(\frac{\partial F_{k,n}(x,\beta^*_{k,t}(x))}{\partial\beta'_k} - \frac{\partial F_k(x,\beta_{k0})}{\partial\beta'}\right)\left(\widehat{\beta}_{k,t} - \beta_{k0}\right)\right| \\ &+ \sup_{x\in\mathcal{X}^+}\left|\Delta_{k0}(x)'\left(n^{-1/2}\sum_t\left(\widehat{\beta}_{k,t} - \beta_{k0}\right) - B_k\sqrt{n}\overline{H}_{k,n}\right)\right| \\ = &\; o_p(1) + o_p(1) = o_p(1). \end{aligned}$$

Therefore, it suffices to show that Lemma A.6 (a) holds under the local alternatives. This follows by a modification of the proof of Lemma A.6(a), and by application of the CLT of Herrndorf (1984), for $\alpha-$mixing arrays, in order to verify the finite dimensional (fidi) convergence condition of Theorem 10.2 of Pollard (1990).

Proof of Corollary 4.3: By contiguity, $q^{G+}_{n,S_n}(1-\alpha) \xrightarrow{p} q^{G+}(1-\alpha)$, under the local alternatives. The result follows immediately from Theorem **4.2**.

Proof of Theorem 5.4: The theorem follows from Lemmas HA.1 and HA.2, and by application of the CMT.

Note to readers. In an online supplement at Cambridge Journals Online, the proofs of lemmas used in the appendix to this paper are detailed. Additionally, the online supplement contains additional tabulated results from Monte Carlo experiments discussed in Section 6 of this paper. In summary, readers may refer to the supplementary material associated with this article, available at Cambridge Journals Online (journals.cambridge.org/ect).

References

[1] Andrews, D.W.K. and P. Guggenberger (2010), Asymptotic Size and a Problem with Subsampling and with the m out of n Bootstrap, *Econometric Theory* 26, 426-468.

[2] Chiang, T. (1986), Empirical Analysis of the Predictor of Future Spot Rates, *Journal of Financial Research*, 153-162.

[3] Corradi, V., S. Jin, and N. R. Swanson (2016), Improved Tests for Robust Forecast Comparison, mimeo.

[4] Corradi, V. and N. R. Swanson (2005), A Test for Comparing Multiple Misspecified Conditional Interval Models, *Econometric Theory* 21, 991-1016.

[5] Corradi, V. and N. R. Swanson (2006), Predictive Density Evaluation, in G. Elliot, C. W. J. Granger, and A. Timmermann, (eds.), *Handbook of Economic Forecasting,* Volume 1, Elsevier, Amsterdam.

[6] Corradi, V. and N. R. Swanson (2007), Nonparametric Bootstrap Procedures For Predictive Inference Based On Recursive Estimation Schemes, *International Economic Review* 48, 67–109.

[7] Corradi, V. and N. R. Swanson (2013), A Survey of Recent Advances in Forecast Accuracy Comparison Testing, with an Extension to Stochastic Dominance, in X. Chen and N. Swanson (eds.), *Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions, Essays in honor of Halbert L. White, Jr.*, Springer.

[8] Diebold, F. X. and R. S. Mariano (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* 13, 253-263.

[9] Diebold, F.X. and M. Shin (2015), Assessing Point Forecast Accuracy by Stochastic Loss Distance, *Economics Letters* 130, 37-38.

[10] Elliott, G. and A. Timmermann (2004), Optimal Forecast Combinations Under General Loss Functions and Forecast Error Distributions, *Journal of Econometrics* 122, 47–79.

[11] Giacomini, R. and H. White (2006), Tests of Conditional Predictive Ability, *Econometrica* 74, 1545-1578.

[12] Goncalves, S. and H. White (2004), Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models, *Journal of Econometrics*, 119, 199-220.

[13] Granger, C. W. J. (1999), Outline of Forecast Theory using Generalized Cost Functions, *Spanish Economic Review* 1, 161-173

[14] Hall, P., J. Horowitz, and B. Y. Jing (1995), On Blocking Rules for the Bootstrap with Dependent Data, *Biometrika* 82, 561-574.

[15] Hansen, B. E. (1996), Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays, *Econometric Theory* 12, 347-359.

[16] Hansen, P. R. (2005), A Test for Superior Predictive Ability, *Journal of Business and Economic Statistics* 23, 365–380.

[17] Harel, M. and M. L. Puri (1999), Conditional Empirical Processes Defined by Nonstationary Absolutely Regular Sequences, *Journal of Multivariate Analysis* 70, 250-285.

[18] Herrndorf, N. (1984), An Invariance Principle for Weakly Dependent Sequences of Random Variables, *Annals of Probability* 12, 141-153.

[19] Holm, S. (1979), A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics* 6, 65–70.

[20] Hunter, W. C. and S. G. Timme (1992), A Stochastic Dominance Approach to Evaluating Foreign Exchange Hedging Strategies, *Financial Management* 21, 104-112.

[21] Klecan, L., R. McFadden, and D. McFadden (1991), A Robust Test for Stochastic Dominance, Working Paper, Department of Economics, MIT.

[22] Linton, O., E. Maasoumi, and Y. J. Whang (2005), Consistent Testing for Stochastic Dominance: A Subsampling Approach, *Review of Economic Studies* 72, 735-765.

[23] Linton, O., K. Song and Y.J. Whang (2010), An Improved Bootstrap Test of Stochastic Dominance, *Journal of Econometrics* 154, 186-202.

[24] McCracken, M. W. (2000), Robust Out-of-Sample Inference, *Journal of Econometrics* 99, 195-223.

[25] Meese, R., and K. Rogoff (1983), Empirical Exchange Rate Models of the Seventies: Do They Fit out of Sample? *Journal of International Economics* 14, 3-24.

[26] Politis, D. N., J. P. Romano (1994), Limit Theorems for Weakly Dependent Hilbert Space Valued Random Variables with Application to the Stationary Bootstrap, *Statistica Sinica 4*, 461-476.

[27] Pollard, D. (1990), Empirical Processes: Theory and Applications, CBMS Conference Series in Probability and Statistic, Vol.2, Institute of Mathematical Statistics, Hayward.

[28] West, K. D. (1996), Asymptotic Inference about Predictive Ability, *Econometrica* 64, 1067-1084.

[29] White, H. (2000), A Reality Check for Data Snooping, *Econometrica* 68, 1097-1126.

Table 1: Monte Carlo Results: GL and CL Forecast Superiority (DGP1 - DGP6)

	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6
S_n	GL forecast superiority						CL forecast superiority					
	n=100						n=100					
0.63	0.105	0.130	0.836	0.723	0.862	0.385	0.089	0.133	0.761	0.856	0.908	0.508
0.54	0.097	0.113	0.830	0.734	0.856	0.349	0.111	0.145	0.777	0.875	0.926	0.521
0.44	0.112	0.107	0.850	0.726	0.871	0.373	0.103	0.128	0.770	0.861	0.938	0.515
0.35	0.099	0.108	0.824	0.730	0.874	0.348	0.106	0.126	0.780	0.853	0.932	0.481
0.25	0.123	0.119	0.841	0.726	0.882	0.412	0.103	0.118	0.796	0.865	0.940	0.528
0.16	0.121	0.123	0.859	0.748	0.887	0.384	0.120	0.136	0.809	0.870	0.936	0.540
DM	0.109	0.131	0.704	0.965	0.981	0.691						
	n=500						n=500					
0.54	0.114	0.104	1.000	1.000	1.000	0.826	0.101	0.122	1.000	1.000	1.000	0.948
0.45	0.097	0.125	1.000	1.000	1.000	0.817	0.105	0.095	1.000	1.000	1.000	0.945
0.36	0.093	0.101	1.000	1.000	1.000	0.813	0.104	0.123	1.000	1.000	1.000	0.956
0.27	0.092	0.104	1.000	0.999	1.000	0.821	0.089	0.094	1.000	1.000	1.000	0.947
0.17	0.106	0.102	1.000	0.999	1.000	0.828	0.097	0.120	1.000	1.000	1.000	0.943
0.08	0.096	0.101	1.000	1.000	1.000	0.828	0.101	0.105	1.000	1.000	1.000	0.938
DM	0.134	0.096	0.992	1.000	1.000	0.994						
	n=1000						n=1000					
0.50	0.097	0.097	1.000	1.000	1.000	0.984	0.109	0.110	1.000	1.000	1.000	0.998
0.41	0.106	0.121	1.000	1.000	1.000	0.985	0.104	0.104	1.000	1.000	1.000	0.998
0.33	0.077	0.127	1.000	1.000	1.000	0.981	0.104	0.112	1.000	1.000	1.000	0.998
0.24	0.094	0.084	1.000	1.000	1.000	0.973	0.112	0.092	1.000	1.000	1.000	1.000
0.15	0.017	0.102	1.000	1.000	1.000	0.972	0.093	0.091	1.000	1.000	1.000	0.999
0.06	0.108	0.088	1.000	1.000	1.000	0.982	0.108	0.109	1.000	1.000	1.000	0.999
DM	0.110	0.127	1.000	1.000	1.000	1.000						

*Notes: See Sections 3 and 6 for complete details. Size experiments: GP1 and DGP2. Power experiments: DGP3, DGP4, DFP5, DGP6. Entries are rejection frequencies based on 1000 Monte Carlo replications. The number of bootstrap resamples is 300 and S_n is the bootstrap smoothing parameter. Nominal test size is 10%.

Table 2: Monte Carlo Results: GL Forecast Superiority (DGP7 - DGP14)

S_n	DGP7	DGP8	DGP9	DGP10	DGP11	DGP12	DGP13	DGP14
				n=250				
0.58	0.099	0.085	0.041	0.734	0.851	1.000	0.852	1.000
0.42	0.088	0.086	0.045	0.778	0.835	1.000	0.864	1.000
0.27	0.102	0.089	0.054	0.764	0.837	1.000	0.841	1.000
0.11	0.102	0.088	0.053	0.768	0.855	1.000	0.882	1.000
				n=500				
0.54	0.102	0.101	0.057	0.971	0.988	1.000	0.993	1.000
0.39	0.113	0.081	0.033	0.971	0.988	1.000	0.991	1.000
0.23	0.111	0.091	0.050	0.975	0.980	1.000	0.989	1.000
0.08	0.102	0.106	0.059	0.973	0.978	1.000	0.993	1.000
				n=1000				
0.50	0.091	0.095	0.052	1.000	1.000	1.000	1.000	1.000
0.36	0.090	0.091	0.060	1.000	1.000	1.000	1.000	1.000
0.21	0.106	0.091	0.064	1.000	1.000	1.000	1.000	1.000
0.06	0.107	0.084	0.049	1.000	1.000	1.000	1.000	1.000

*Notes: See notes to Table 1. Size Experiments: DGP7 - DGP10. Power Experiments: DGP11 - DGP14.

Table 3: Monte Carlo Results: CL Forecast Superiority (DGP7 - DGP14)

S_n	DGP7	DGP8	DGP9	DGP10	DGP11	DGP12	DGP13	DGP14
				n=250				
0.58	0.090	0.085	0.031	0.818	0.950	1.000	0.969	1.000
0.42	0.094	0.086	0.035	0.857	0.952	1.000	0.971	1.000
0.27	0.100	0.089	0.027	0.841	0.957	1.000	0.973	1.000
0.11	0.105	0.088	0.039	0.836	0.964	1.000	0.970	1.000
				n=500				
0.54	0.083	0.087	0.041	0.993	1.000	1.000	1.000	1.000
0.39	0.101	0.093	0.034	0.996	0.998	1.000	1.000	1.000
0.23	0.089	0.079	0.027	0.990	0.999	1.000	1.000	1.000
0.08	0.097	0.105	0.044	0.996	0.998	1.000	0.998	1.000
				n=1000				
0.50	0.099	0.105	0.025	1.000	1.000	1.000	1.000	1.000
0.36	0.097	0.110	0.034	1.000	1.000	1.000	1.000	1.000
0.21	0.096	0.091	0.030	1.000	1.000	1.000	1.000	1.000
0.06	0.088	0.097	0.041	1.000	1.000	1.000	1.000	1.000

*Notes: See notes to Table 2.

Table 4: Monte Carlo Results: GL and CL Forecast Superiority (DGP15 - DGP18)

S_n	DGP15	DGP16	DGP17	DGP18	DGP15	DGP16	DGP17	DGP18
	GL forecast superiority				CL forecast superiority			
	n=250				n=250			
7	0.099	0.081	0.928	0.964	0.105	0.099	0.988	0.990
9	0.112	0.102	0.922	0.978	0.095	0.108	0.988	0.993
11	0.110	0.091	0.939	0.969	0.112	0.094	0.987	0.993
14	0.101	0.117	0.944	0.957	0.109	0.112	0.989	0.993
16	0.111	0.106	0.946	0.957	0.098	0.095	0.989	0.994
19	0.122	0.108	0.936	0.966	0.126	0.125	0.994	0.995
	n=500				n=500			
7	0.123	0.088	0.998	1.000	0.109	0.097	1.000	1.000
11	0.112	0.102	0.998	1.000	0.109	0.101	1.000	1.000
14	0.099	0.096	0.998	1.000	0.090	0.099	1.000	1.000
18	0.110	0.108	0.998	1.000	0.118	0.111	1.000	1.000
21	0.101	0.109	0.997	0.999	0.101	0.121	0.999	1.000
25	0.109	0.100	0.997	0.999	0.117	0.119	0.999	1.000
	n=1000				n=1000			
8	0.097	0.093	1.000	1.000	0.109	0.092	1.000	1.000
13	0.092	0.109	1.000	1.000	0.095	0.089	1.000	1.000
18	0.096	0.096	1.000	1.000	0.095	0.107	1.000	1.000
23	0.109	0.103	1.000	1.000	0.095	0.094	1.000	1.000
27	0.102	0.090	1.000	1.000	0.105	0.088	1.000	1.000
32	0.094	0.102	1.000	1.000	0.108	0.107	1.000	1.000

*Notes: See notes to Table 1. Size Experiments: DGP15, DGP16. Power Experiments: DGP17, DGP18.

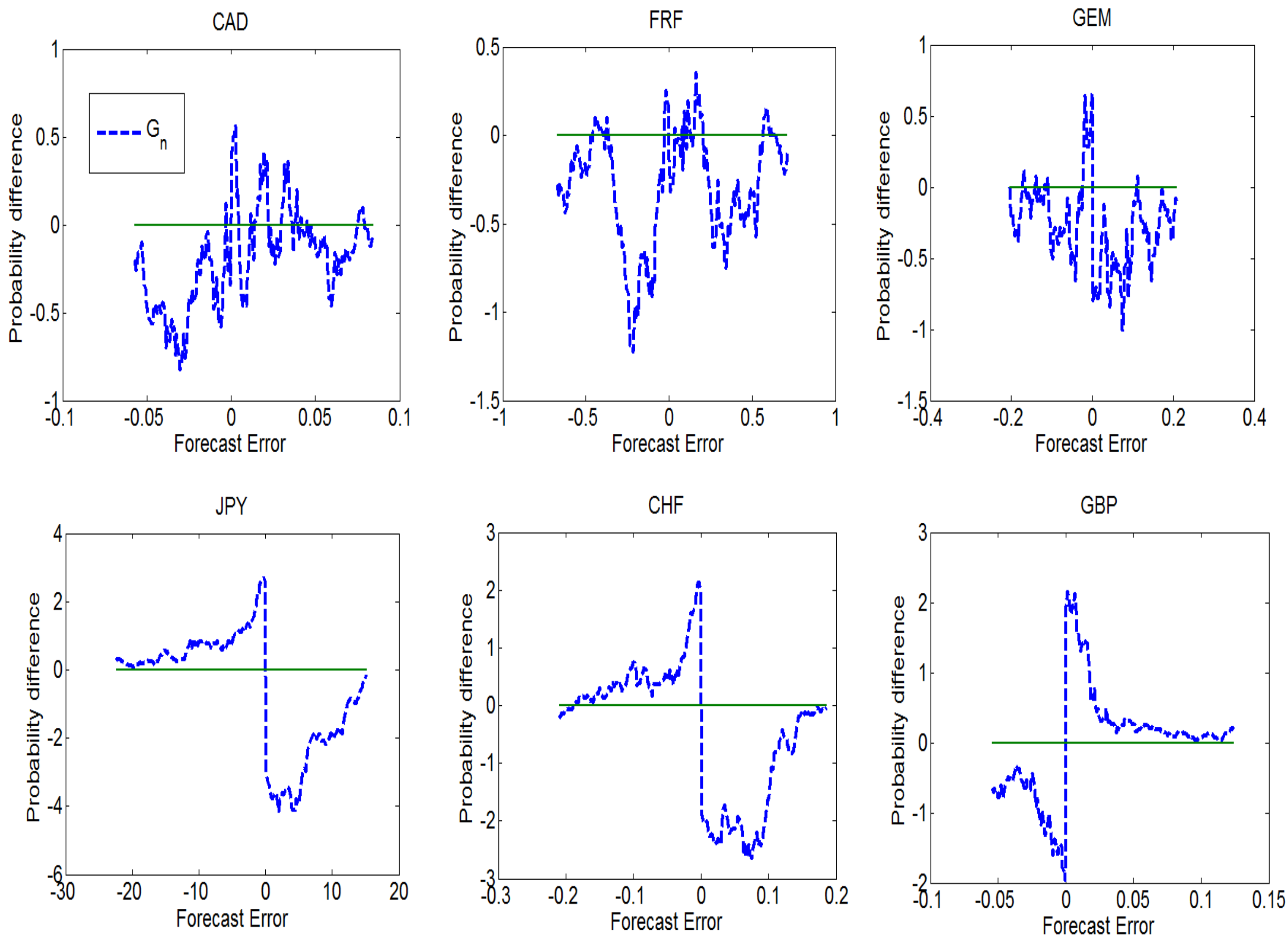

Figure 1: Plots of G_n Forecast Errors Associated with *Spot* and *Forward* Models

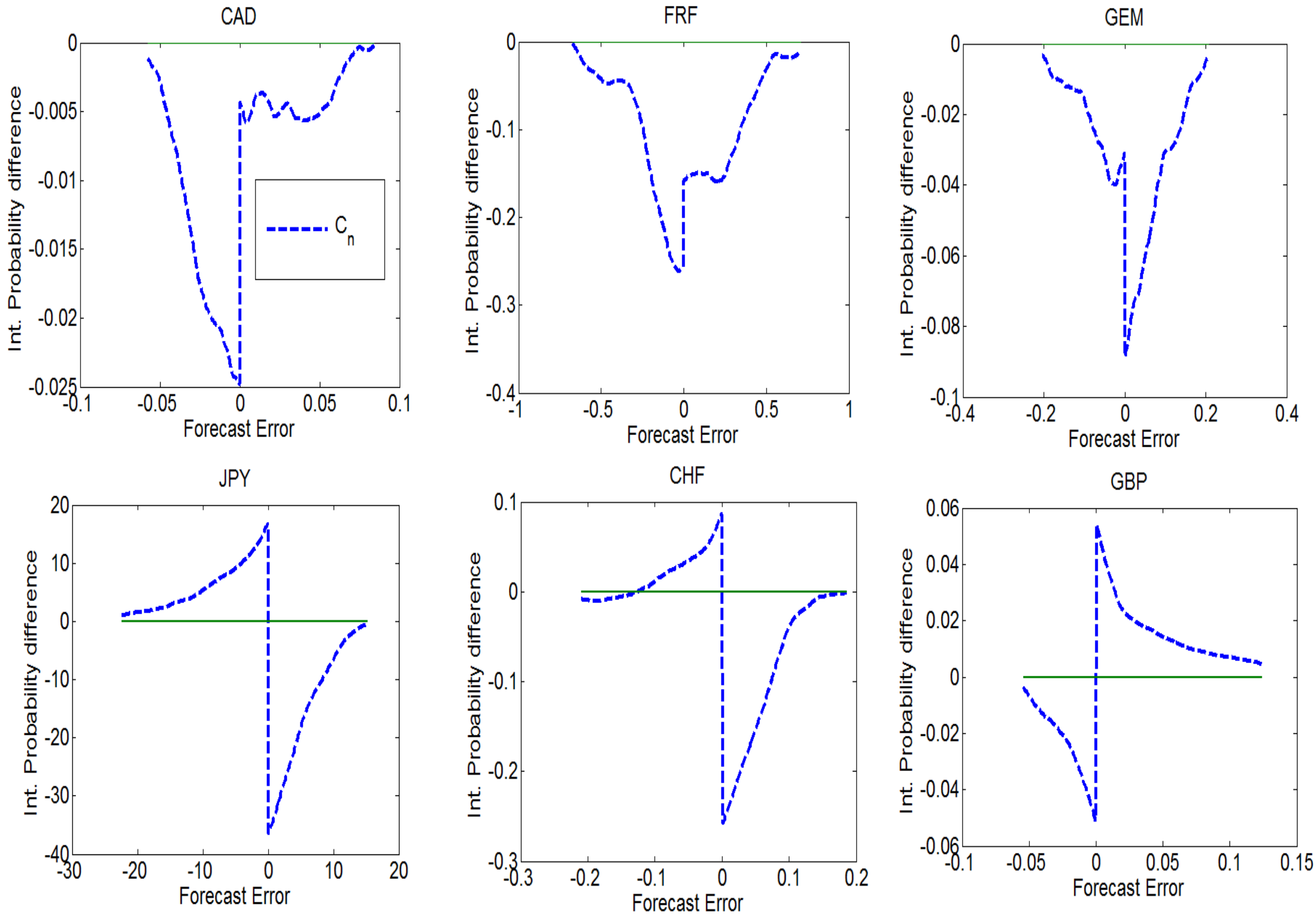

Figure 2: Plots of C_n Forecast Errors Associated with *Spot* and *Forward* Models

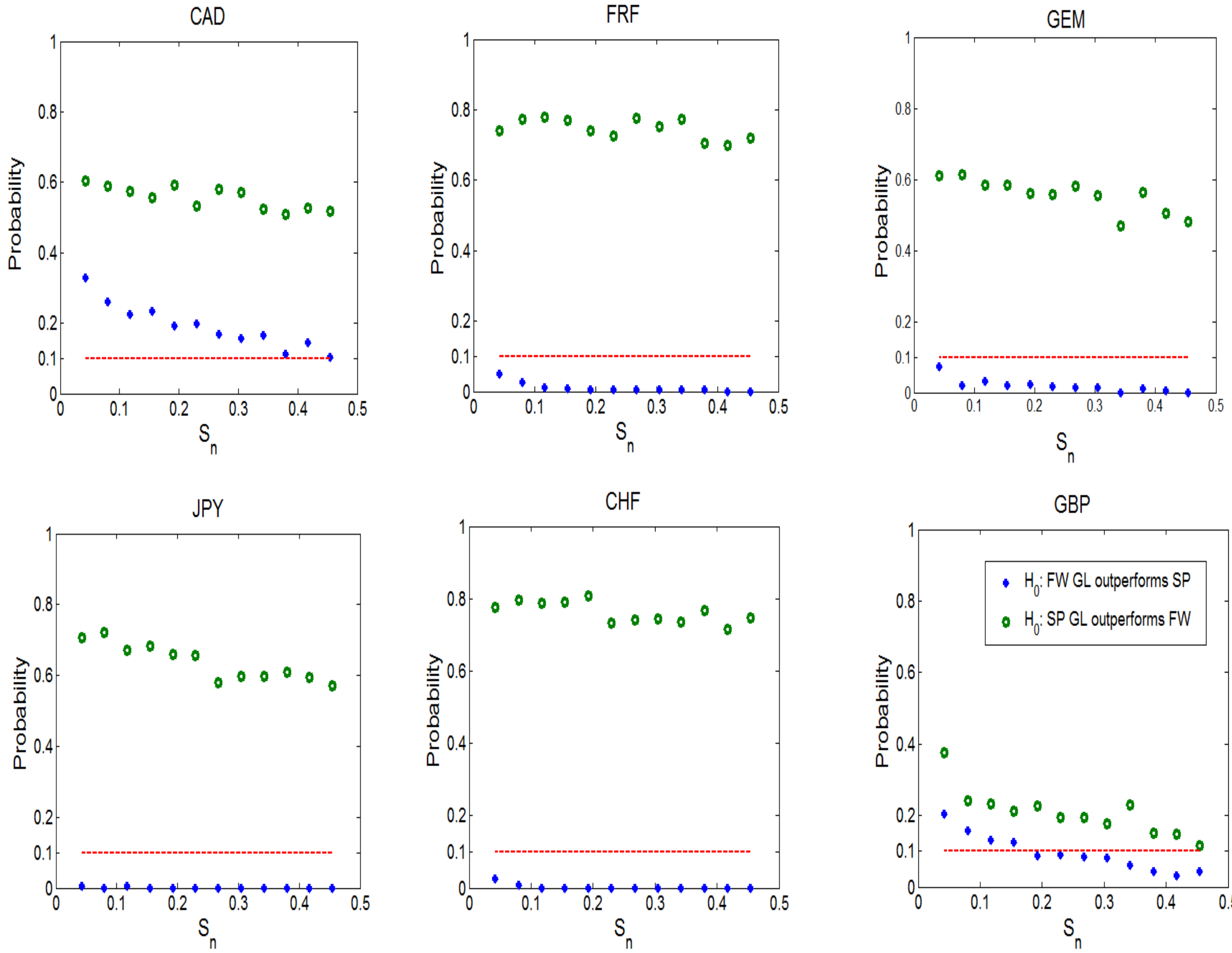

Figure 3: Plots of p-values for GL Forecast Superiority Test

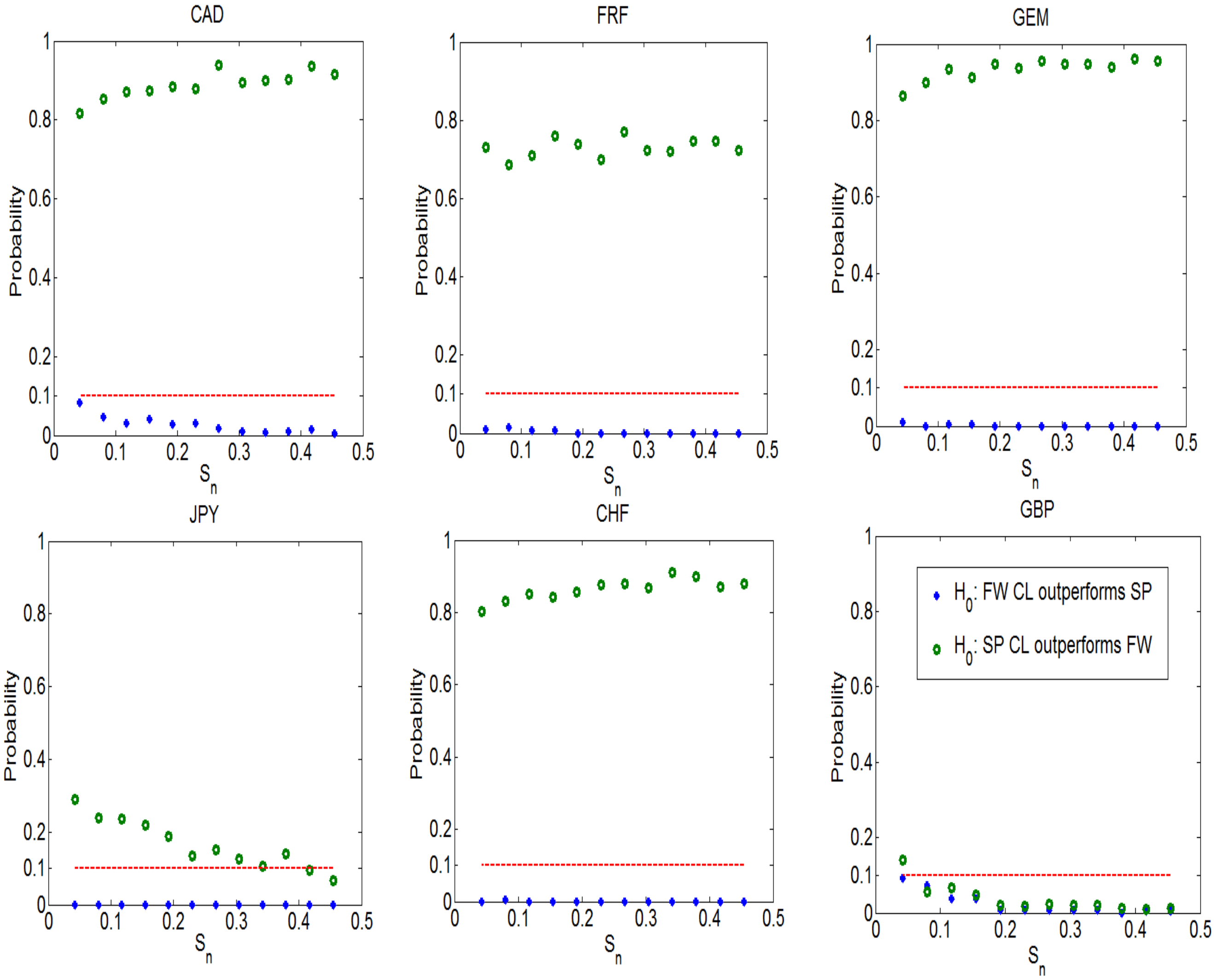

Figure 4: Plots of p-values for CL Forecast Superiority Test