Robust Forecast Comparison*

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

May 2015

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. This paper addresses this issue by using a novel criterion for forecast evaluation which is based on the entire distribution of forecast errors. 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 first (second) order stochastic dominance. This allows us to develop a forecast evaluation procedure based on an out-of-sample generalization of the tests introduced by Linton, Maasoumi and Whang (2005). The asymptotic null distributions of our test statistics are nonstandard, and resampling procedures are used to obtain the critical values. Additionally, the tests are consistent and have nontrivial local power under a sequence of local alternatives. In addition to the stationary case, we outline theory extending our tests to the case of heterogeneity induced by distributional change over time. Monte Carlo simulations suggest that the tests perform reasonably well in finite samples; and an application to 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.

*Valentina Corradi, School of Economics, University of Surrey, Guildford, Surrey, GU2 7XH, UK, v.corradi@surrey.ac.uk.; Jin Sainan, School of Economics, Singapore Management University, 90 Stamford Road, #05-30, Singapore 178903, snjin@smu.edu.sg; and Norman Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. We are grateful to Frank Diebold, Federico Martellosio, Minchul Shin, Tang Srisuma, Liangjun Su; and to seminar participants at the University of Surrey and the 2015 Princeton/QUT/SJTU/SMU Conference on the Frontier in Econometrics 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 (1999a) proposes the use of generalized loss functions $L(\cdot)$, 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 $e_1 < e_2 \leq 0$. We term the class of loss functions that satisfy the above three properties as general loss functions (GL or $\mathcal{L}_G$). A second class of loss functions are defined as convex loss functions (CL or $\mathcal{L}_C$), if in addition to satisfying the above three properties, they are convex. Indeed, convex functions include MSFE and MAFE, as well as several asymmetric functions, such as lin-lin and linex functions, see Elliott and Timmermann (2004) for details.

A natural question arising from the above discussion is the following. How do 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 (say) forecast errors, $\{e_{1t}\}$, $\{e_{2t}\}$..., $\{e_{lt}\}$, such that forecasts are to be ranked the same way, regardless of loss function. To answer this question, we introduce two concepts: general-loss (GL) forecast superiority and convex-loss (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 a mapping between GL superiority and first order stochastic dominance, and a mapping between CL superiority and second order stochastic dominance. This allows us to develop a forecast evaluation procedure to test for GL forecast superiority and CL forecast superiority, based on an out-of-sample generalization of the tests introduced by Linton, Maasoumi and Whang (2005, hereafter LMW).

Since the influential work of Meese and Rogoff (1983, 1988), it has become common 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. One of the most important contributions in this area is the seminal paper of Diebold and Mariano (1995, hereafter DM), in which a test of equal predictive accuracy between two competing models is proposed. Since then, efforts have been made to generalize DM-type tests in order to account for parameter estimation error (West, 1996; West and McCracken, 1998), to allow for non-differential loss functions together with parameter estimation error (McCracken, 2000), to test for conditional predictive ability (Giacomini and White, 2006), to allow for integrated and cointegrated variables (Clements and Hendry, 1999, 2001; Corradi, Swanson and Olivetti, 2001), to address the issue of the joint comparison of more than two competing models (Sullivan, Timmermann and White, 1999; White 2000; Hansen, 2005; Romano and Wolf, 2005; Corradi and Distaso, 2011), and to evaluate predictive intervals, conditional quantiles and predictive densities (Christoffersen, 1998; Giacomini and Komunjer, 2005; Corradi and Swanson 2005; Corradi and Swanson 2006a; Corradi and Swanson, 2006b). Other papers tackle the issue of predictive accuracy testing via the use of encompassing and related tests (Phillips, 1996; Harvey, Leybourne and Newbold, 1997; Chao, Corradi and Swanson, 2001; Clark and McCracken 2001; Corradi and Swanson, 2002; Giacomini and Komunjer, 2005). See West (2006), Clark and McCracken (2013), Corradi and Swanson (2013), and Diebold (2014) for comprehensive surveys on recent developments in forecast comparison methodology.

There are several common features of the aforementioned papers. First, most of them are based upon moments or conditional moments of the forecast errors, and researchers must specify the objective function (say, loss function or likelihood function) in order to carry out forecast evaluation. See Clements and Hendry (1993) for some limitations of this approach. Second, all of them are out-of-sample based, despite the fact that some (e.g., DM) ignore parameter estimation error. Third, 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 Giacomini and White (2006) for details.

In this paper, our objective is to extend early work that considers moment-based tests, and to instead consider distribution-based tests. A moment-based criterion only looks in a particular direction when examining forecast errors. For example, MSFE is designed for squared error loss functions and MAFE for absolute error loss functions. GL and CL forecast superiority, however, is based on evaluation of the entire forecast error distribution, and does not require knowledge of the exact form of the loss function. When implementing our evaluation procedure, the null hypothesis is specified in terms of inequality restrictions, and this delivers a direct test of forecast superiority.

In a related recent survey paper, Corradi and Swanson (2013) discuss predictive evaluation based on distributions of losses using stochastic dominance principles. They provide motivation, a basic set-up, and test statistics, without including any formal theory or Monte Carlo results. In their paper, they take the loss function as given, and propose an evaluation criterion based on comparing cumulative loss functions $F(L(e))$, where $F(L(e))$ is the CDF of $L(e)$. They consider panels or combinations of forecasts, and ignore parameter estimation error. In contrast, we provide a forecast evaluation testing procedure which is valid under generalized loss functions, and which is based directly on the evaluation of $F(e)$, the CDF of the forecast error. Moreover, our procedure takes into account parameter estimation error and data dependence. We develop limit theory for the tests under the null and show that the tests are consistent and have nontrivial local power under a sequence of local alternatives. Additionally, the asymptotic null distributions of our test statistics are nonstandard, and resampling procedures are used to obtain the critical values.

Other deviations from traditional moment-based forecast evaluation methods are available in the literature. For example, Granger and Pesaran (2000) argue in favor of a close link between the decision and the forecast evaluation problems. Pesaran and Skouras (2002) discuss a decision-based approach for evaluation and comparison of forecasts. Granger and Machina (2006) propose a class of realistic decision-based loss functions for forecast evaluation. Diebold and Shin (2014a,b) suggest choosing the model which has the cumulative distribution closest to a step function equal to zero over the negative real line and equal to one over positive real line. If one forecast is superior according to their criterion, named stochastic loss divergence, then it is also superior according to any piecewise linear loss function, such as L_1-loss or lin-lin loss. Our distribution-based forecast comparison procedure also has a link with the decision-based approach to forecast evaluation, but careful investigation of the linkages is beyond the scope of this paper.

A preponderance of tests based on (conditional) moments of forecast errors require 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. On the other hand, one popular explanation for systematic out-of-sample forecast failure in economics is the prevalence of time varying underlying data generating processes. In lieu of this fact, provide a generalization of our testing procedure to a particular type of non-stationarity (i.e., heterogeneity), which is induced by distributional change over time, see e.g. Giacomini and Rossi (2009). It should be noted that heterogeneity is plausibly of less concern in some areas of economics (say, financial economics) than in others (say, labor economics), and so we provide a procedure for heterogenous processes, and also one which assumes stationarity. 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.

Finally, it is worth stressing that our testing procedure can be adapted to forecast combination. It has become an attractive strategy to combine competing professional forecasts or survey predictions, to aggregate crowd wisdom collected from different sources, and to combine forecasts generated by econometric models, for example. The reason for this is that combined forecasts often outperform the "best" individual forecasts, see e.g., Stock and Watson (1999), Newbold and Harvey (2002), Timmermann (2006), Elliott, Gargano, and Timmermann (2013) for detailed discussions. In standard procedures used in the literature, optimal forecast weights are generally loss function dependent, see e.g. Elliott and Timmermann, (2004). In our context, one can evaluate different forecast combinations and select combination weights based on GL and CL forecast superiority. This line of research, however, is left for future work.

The rest of the paper is organized as follows. Section 2 introduces the hypotheses and test statistics under the assumption that the underlying stochastic process is stationary. In Section 3 we derive the asymptotic distribution of the test statistics, and establish the first order asymptotic validity of a bootstrap procedure used to construct critical values. Section 4 studies the power properties of the test statistics, and of their associated bootstrap analogs, under local and global alternatives. Section 5 extends 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 an appendix.

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. $L : \mathbb{R} \to \mathbb{R}^+$ 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 $L \in \mathcal{L}_C$.

We now establish a mapping between $\mathcal{L}_G$ forecast superiority and first order stochastic dominance, and between $\mathcal{L}_C$ forecast superiority and second order stochastic dominance. This mapping is instrumental for deriving direct tests for $\mathcal{L}_G/\mathcal{L}_C$ 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, 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. Namely, one forecast can outperform another forecast and at the same time be outperformed, in which case the two forecasts are equivalent in the sense that they result in the same expected loss for the loss functions in the corresponding class. Strong GL or CL forecast superiority holds by requiring that strict inequality hold 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 $\beta_{k0} \in \Theta_k \subset \mathbb{R}^{L_k}$:

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

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, say) is the collection of all predictive regressors, $\beta_0 = (\beta'_{10}, ..., \beta'_{l0})'$ is the pseudo true parameter vector on the parameter space $\Theta = \Pi_{k=1}^l \Theta_k$, $m_k : \mathbb{R}^{P_k} \times \Theta_k \to \mathbb{R}$ and $\widetilde{m}_k : \mathbb{R}^{\overline{P}} \times \Theta \to \mathbb{R}$. 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.

Like West and McCracken (1998) and McCracken (2000), we allow for three different forecasting schemes which use recursive, rolling, and fixed windows of data for estimation. However, they 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, i.e., $\widehat{\beta}_{k,t} = \widehat{\beta}_{k,R}$.

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

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 $(e_1, ..., e_l)$

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

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

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, we assume that the union of the supports, $\mathcal{X}$, is bounded,[1] as do Klecan, McFadden, and McFadden (1991) and LMW. 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

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

and

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

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 $C_{k,n}(x) = \left\{\int_{-\infty}^{x} \left(\overline{F}_{1,n}\left(s, \underline{\widehat{\beta}}_{1,R:T}\right)\right.\right.$ $\left.\left. -\overline{F}_{k,n}\left(s, \underline{\widehat{\beta}}_{k,R:T}\right)\right) ds1(x<0) + \int_{x}^{\infty} \left(\overline{F}_{k,n}\left(s, \underline{\widehat{\beta}}_{k,R:T}\right) - \overline{F}_{1,n}\left(s, \underline{\widehat{\beta}}_{1,R:T}\right)\right) ds1(x \geq 0)\right\}$.

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 the 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

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

[1] Technically speaking, this will facilitate the establishment of stochastic equicontinuity for the underlying empirical processes our theory is based upon.

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 $H_0^{TG+} : TG^+ \leq 0$, $H_0^{TG-} : TG^- \leq 0$, and since $H_0^{TC} = HG_0^{TC+} \cap H_0^{TC^-}$, where $H_0^{TC+} : TC^+ \leq 0$, $H_0^{TC-} : TC^- \leq 0$. 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-}. Then, 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 $\xi_n \to 0$, $\sup_{x\in\mathcal{X}} \sup_{\beta:||\beta-\beta_{k0}||<\xi_n} ||\left(\partial F_k(x,\beta)/\partial\beta\ sgn\left(x\right) - \Delta_{k0}(x)\right)|| = O(\xi_n^{\eta})$, for some $\eta > 0$, where $\Delta_{k0}(x) = \partial F_k(x, \beta_{k0})/\partial\beta sgn\left(x\right)$.

(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 need 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 sequence of positive constants $\{\xi_n : n \geq 1\}$, such that $\xi_n \to 0$, $\sup_{x\in\mathcal{X}} \sup_{\beta:||\beta-\beta_{k0}||<\xi_n} ||(\partial/\partial\beta)\{\int_{-\infty}^{x} F_k(t, \beta)dt\ 1(x < 0) + \int_{x}^{\infty}(1 - F_k(t, \beta))dt1(x \geq 0)\} - \Lambda_{k0}(x)|| = O(\xi_n^{\eta})$, 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. 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 appear 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. Assumptions A.3 and A.3* differ in the amount of smoothness required. For the CL forecast superiority test, less smoothness is required.

Assumption A.2 is identical to Assumption 1 in McCracken (2000). 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 discussions about this assumption. Assumption A.4 is identical to Assumption 2 of McCracken (2000).

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.

To derive the asymptotic null distributions of our test statistics, we define the empirical processes in (x, β)

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

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

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

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

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

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 $\Omega_k = \gamma\sum_{j=-\infty}^{\infty}\Gamma_k(j)$ where

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

Obviously, $\gamma = 0$ when $\pi = 0$, indicating the case when 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 $H_0^{TG^+}$,*

$$\begin{aligned} TG_n^+ &\Rightarrow \max_{k=2,..,l} \sup_{x \in \mathcal{B}_k^{g+}} [\widetilde{g}_k(x) + \Delta_{k0}(x)' B_k v_{k0} - \Delta_{10}(x)' B_1 v_{10}], \ \textit{if} \ \ TG^+ = 0 \\ &\Rightarrow -\infty \qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{if} \ \ TG^+ < 0, \end{aligned} \tag{3.3}$$

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

$$\begin{aligned} TG_n^- &\Rightarrow \max_{k=2,..,l} \sup_{x \in \mathcal{B}_k^{g-}} [\widetilde{g}_k(x) + \Delta_{k0}(x)' B_k v_{k0} - \Delta_{10}(x)' B_1 v_{10}], \ \textit{if} \ \ TG^- = 0 \\ &\Rightarrow -\infty \qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{if} \ \ TG^- < 0, \end{aligned} \tag{3.4}$$

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 $H_0^{TC^+}$,*

$$\begin{aligned} TC_n^+ &\Rightarrow \max_{k=2,..,l} \sup_{x \in \mathcal{B}_k^{c+}} [\widetilde{c}_k(x) + \Lambda_{k0}(x)' B_k v_{k0} - \Lambda_{10}(x)' B_1 v_{10}], \ \textit{if} \ \ TC^+ = 0 \\ &\Rightarrow -\infty \qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{if} \ \ TC^+ < 0, \end{aligned} \tag{3.5}$$

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

$$\begin{aligned} TC_n^- &\Rightarrow \max_{k=2,..,l} \sup_{x \in \mathcal{B}_k^{c-}} [\widetilde{c}_k(x) + \Lambda_{k0}(x)' B_k v_{k0} - \Lambda_{10}(x)' B_1 v_{10}], \ \textit{if} \ \ TC^- = 0 \\ &\Rightarrow -\infty \qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{if} \ \ TC^- < 0, \end{aligned} \tag{3.6}$$

where $\mathcal{B}_k^{c+} = \{x \in \mathcal{X}^+ : \int_x^\infty (F_1(s) - F_k(s)) ds 1(x \geq 0) = 0\}$ and $\mathcal{B}_k^{c-} = \{x \in \mathcal{X}^- : \int_{-\infty}^x (F_k(s) - F_1(s)) ds 1(x \leq 0) = 0\}$.

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 complicates 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 using the stationary bootstrap with TC_n^+ and TC_n^- are similar and hence are omitted.

More specifically, 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 $x \in \mathcal{X}^+$.[2] We use stationary bootstrap since it ensures that the resampled series are also stationary and mixing, conditional on the original data. See Politis and Romano (1994a, b) for complete details.

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

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

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: $0 < S_n < 1$, $S_n \to 0$, and $nS_n^2 \to \infty$, as $n \to \infty$.

To implement the stationary bootstrap, follow the algorithm proposed in Politis and Romano (1994b). (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). The procedure delivers geometrically distributed blocks of random length, with mean block length $1/S_n$, as discussed in Politis and Romano (1994a).

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 (1994b) 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 $\sqrt{n}\left(G_{kn} - G_k\right)$. 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 (see, e.g., 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 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 $\Sigma_k = B_k[\lim_{T\to\infty} \mathrm{var}(n^{-1/2}\sum_{t=R+1}^{T} H_k(t))]B_k'$.

[2] Note that in our setup, if the number of competing models, l, is small relative to the number of forecasts, n, size distortion is not significant (Hansen, 2003).

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

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 $x\in\mathcal{X}^-$,*

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

as $T\to\infty$, where k=2,..., l, ρ is any metric metrizing weak convergence, and $\boldsymbol{L}[\cdot]$ 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 the estimates $\widehat{\beta}_{k,t}$. 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 $T\to\infty$,*

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

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 the following simulations and applications, we choose a set of S_n that satisfies Assumption A.5.

Once B and S_n are determined, bootstrap critical values can be estimated straightforwardly. 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 (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.

Our tests do not satisfy the requirement of asymptotic similarity and thus they are asymptotically biased, like many other tests for multiple inequality restrictions. Hansen (2003) shows that p-values associated with use of the stationary bootstrap test are actually upper bounds for an asymptotically unbiased test. In our context, one can follow Hansen (2003) in order to propose an asymptotically unbiased test, but this is beyond the scope of the current paper. Moreover, simulation results show that use of the stationary bootstrap yields tests with good finite sample properties. This finding is consistent with the fact that asymptotically unbiased tests do not necessarily dominate tests carried out using the stationary bootstrap tests in finite samples.[3] In a follow up to this paper, Corradi, Jin and Swanson (2015), building on the work of Linton, Song and Whang (2010), introduce bootstrap tests for GL and CL superiority that are uniformly asymptotically valid under the null hypothesis and have exact asymptotic size over the boundary of the null hypotheses. These tests also allow for the specification of semiparametric models, and allow for non vanishing recursive and/or rolling estimation error.

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 first show 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+},*
$P(TG_n^+ > q_{n,S_n}^{G+}(1-\alpha)) \to 1$ *as* $T \to \infty$,
and under H_1^{TG-},
$P(TG_n^- > q_{n,S_n}^{G-}(1-\alpha)) \to 1$ *as* $T \to \infty$.

Next, 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:

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

[3] Simulations show that asymptotically unbiased tests are less conservative in experiments where there is "movement" away from the least favorable case, although use of the stationary bootstrap delivers tests that perform better in terms of power, in some of the cases examined. We also tried subsampling, as an alterantive approach to critical value construction; but simulation results were less satisfactory in all experiments, and so these results are not reported.

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 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) $B_k(t) \to B_k$, 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) $\sqrt{n}E(h_{k,s}(\beta_{k0})\ ) \to m_k$.

(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, 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),*

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

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

This result implies that the asymptotic local power of the TG_n^+ (TG_n^-) test based on the 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.*

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 $\widehat{\beta}_{k,t}$ 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, *i.e.*, $e_{k,t+\tau} = Y_{t+\tau} - \widehat{Y}_{k,t+\tau}$. 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 forecasting 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 forecasting 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 $\lim_{n\to\infty} n^{-1} \sum_{t=R}^{T} E[L(e_{1,t+\tau}) - L(e_{2,t+\tau})] \leq 0$, *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* $x \in \mathcal{X}$.

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 definitions. 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 $(e_{1,t+\tau}, ..., e_{l,t+\tau})$, $t \geq R$,

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

$$HTG^- = \max_{k=2,..,l} \sup_{x \in \mathcal{X}^-} HG_k(x), \tag{5.2}$$

$$HTC^+ = \max_{k=2,..,l} \sup_{x \in \mathcal{X}^+} HC_k(x), \tag{5.3}$$

$$HTC^- = \max_{k=2,..,l} \sup_{x \in \mathcal{X}^-} HC_k(x), \tag{5.4}$$

[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.

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

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

Without loss of generality, also assume that the union of the supports of all forecast error sequences, $\mathcal{X}$, is bounded. The hypotheses of interest can now be stated as

$$H_0^{HTG} : HTG^+ \leq 0 \cap HTG^- \leq 0 \text{ vs. } H_1^{TG} : HTG^+ > 0 \cup HTG^- > 0 \tag{5.5}$$

and

$$H_0^{HTC} : HTC^+ \leq 0 \cap HTG^- \leq 0 \text{ vs. } H_1^{TC} : HTC^+ > 0 \cup HTC^- > 0. \tag{5.6}$$

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

The test statistics that we consider in this context are based on the empirical analogues of (5.1) to (5.4). They are defined as follows.

$$\begin{aligned} HTG_n^+ &= \max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\sqrt{n}HG_{k,n}(x), \\ HTG_n^- &= \max_{k=2,..,l}\sup_{x\in\mathcal{X}^-}\sqrt{n}HG_{k,n}(x) \\ \\ HTC_n^+ &= \max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\sqrt{n}HC_{k,n}(x), \\ HTC_n^- &= \max_{k=2,..,l}\sup_{x\in\mathcal{X}^-}\sqrt{n}HC_{k,n}(x), \end{aligned}$$

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 $C_0 > (q-1)(q+1)$, 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 $C_0 > rq/(r-q)$, 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

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

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* $H_0^{HTG^+}$,

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

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

$$\begin{aligned} HTG_n^- &\Rightarrow \max_{k=2,..,l} \sup_{x\in\mathcal{B}_k^{hc-}} \widetilde{hg}_k(x) && \text{if } HTG^- = 0 && (5.8)\\ &\Rightarrow -\infty && \text{if } HTG^- < 0, && (5.9) \end{aligned}$$

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* $H_0^{HTC^+}$,

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

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

$$\begin{aligned} HTC_n^- &\Rightarrow \max_{k=2,..,l} \sup_{x\in\mathcal{B}_k^{hc-}} \widetilde{hc}_k(x) && \text{if } HTC^- = 0 && (5.11)\\ &\Rightarrow -\infty && \text{if } HTG^- < 0, && (5.12) \end{aligned}$$

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).

6 Simulation Evidence

In this section, we first discuss results of simulations conducted in order to evaluate the finite sample performance GL and CL forecast superiority tests when there are only two competing sequences of forecast errors. We discuss the results of a Monte Carlo experiment designed to examine the 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 to check the performance of the tests in the case where the underlying process is not stationary.

When computing the suprema in TG_n^+, TG_n^-, TC_n^+, and TC_n^-, we take a maximum over an equally spaced grid of size $\lceil 1.5n^{0.6} \rfloor$, over a 98% range of the pooled empirical distribution; that is, 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 use 1000 replications. We set the number of bootstrap resamples as $B = 300$. 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 each sample size $n \in \{250, 500, 1000\}$ 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 0.1 are reports. Results corresponding to different nominal sizes are qualitatively similar and are not reported.

6.1 Pairwise comparisons: stationary case

We first study the following three data generating processes (DGPs) with independent forecast errors and *i.i.d.* observations:

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

DGP2: e_{1t} ~ *i.i.d.* Uniform (-2, 2) and e_{2t} ~ *i.i.d.*$N(0, 1)$.

DGP3: 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.

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

It is easy to verify that the first design allows us to examine finite sample size properties for both forecast superiority tests, and is a "least favorable" case. The second and third designs allow us to examine finite sample power for both tests.

In the next three DGPs, we allow the two forecast errors to be dependent on each other with non-independent observations. Following Klecan, McFadden, and McFadden (1991), we generate e_{kt} according to

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

where $(\widetilde{e}_{0t}, \widetilde{e}_{1t}, \widetilde{e}_{2t})$ are *i.i.d.* but have different marginals in different DGPs. The parameters $\lambda = \rho = 0.1$ determine the mutual dependence of e_{1t} and e_{2t} and their autocorrelations. This scheme produces autocorrelated and mutually dependent forecast errors, and we consider three such DGPs.

DGP4: $(\widetilde{e}_{0t}, \widetilde{e}_{1t}, \widetilde{e}_{2t})$ are $i.i.d$ $N(0, I_3)$;

DGP5: $\widetilde{e}_{1t} \sim i.i.d.N(0, 1.5)$ and $\widetilde{e}_{kt} \sim i.i.d.N(0, 1)$, for $k = 0$ and 2;

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

As above, DGP4 is our "null" model, while DGPs 5 and 6 are our "alternative" models. A comparison of the simulation results based on DGP1 and DGP4 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 DGP3 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 the rejection frequencies, as discussed above. From the left panel of the table, observe that for our small sample size ($n = 100$), the 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 (3) and DGP4 (6) 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 quickly 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.

6.2 Multiple comparisons: stationary case

For the sake of brevity, we consider independent forecasts and *i.i.d.* observations. For the following eight data generating processes (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 elements in the null and the theory of our stationary bootstrap test applies directly to this case. In DGP9, the benchmark model outperform 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_1 is taken as the benchmark forecast error. For all sample sizes in our investigation, the tests have good size performance for DGPs 7 and 8, where the nulls are least favorable to the alternatives, while the tests are mostly under-sized for DGP 9, where 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. Now, consider the power performance of the test. Interestingly, for all cases (i.e. DGPs 10-14), the test has a good power performance. Note also that an inclusion of poorer models in the test 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_1 is taken as the benchmark forecast error. The superiority test continues to perform well. A comparison of Tables 2 and 3 indicates that we have more chance to correctly reject the null $H_0^{TC} : TC \leq 0$ than the null $H_0^{TG} : TG \leq 0$. 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. For the sake of brevity, we study a small set of DGPs. In particular, for DGPs 15 through 18, $e_{it} \sim a_{it}N(0,1)$, 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 sequence as follows:
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}.
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. We use the block bootstrap, where the block size is chosen to be equally spaced on the interval $[2n^{0.2}, 2n^{0.4}]$. Overall, the sizes for the testing procedure behave reasonably well, despite the fact that the test is a little upward biased when sample size is 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.

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 6 industrialized countries. This study is for illustrative purpose only, and all forecast models are stylized, involving no estimation. Due to this simplistic approach, we do not need to distinguish between in-sample and out-of sample forecasts.

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. The data were obtained from Datastream for daily 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 the observations for a country is missing, we simply deleted it. 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 the future spot rates. Namely, $E_t(X_{t+\tau}) = F_{t,\tau}$, where $X_{t+\tau}$ is the spot exchange rate at 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 conditions of 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

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

at time $t+\tau$.[7] The second set of forecast 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. See, e.g., Chiang (1986) and Meese and Rogoff (1988), who show that the current 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 conduct the 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 report two traditional measures (mean square forecast error - MSFE and mean absolute forecast error - MAFE) of forecast performance for all models under our investigation (refer to Table 5). Note that both MSFE and MAFE belong to the CL and thus also GL class of loss functions. If the forecast error from one forecasting model has both lower MSFE and MAFE than the other model, one might wonder whether it GL or CL outperforms the other. On the other hand, if one model has lower MSFE while the other model has lower MAFE, and the differences are statistically significant,[8] we would expect there to be no CL and thus no GL forecast superiority. Inspection of the results in Table 5 indicates that the spot model has lower MSFE and MAFE than the forward model for all countries, with the exception of the UK. The difference is not big though.

Now consider examining the empirical distribution functions (EDFs) for the forecast errors from the two models. For all six countries, the EDFs for the forward forecast error almost coincide with those for the spot forecast error, with some slight differences. To save space, we do not report the results here. To look at the differences between the forward and spot EDFs more clearly, 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 for 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 difference in $G_n(x)$ and the integrated probability difference in $C_n(x)$ have been scaled up by $\sqrt{n}$, where n is the sample size. Three cases may result in the examination of 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

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

[8] This frequently occurs in practice. For example, the in-sample forecast errors from a least squares regression will have a lower MSFE and higher MAFE than those from a conditional median regression model with the same structure.

is superior to the spot model in the sense that it 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.

All of the plots in Figure 1 are consistent with Case 3. Thus, it is of interest to ascertain whether the sign changes are significant or not. Turing to Figure 2, Case 2 pertains to CAD, FRF, and GEM, while Case 3 pertains to the JPY, CHF, and GBP. Note that the magnitude of the integrated probability difference varies substantially from one country to the other. 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 following 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 $\lceil 1.5n^{0.6} \rfloor$, on the 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 indicate that one model is superior to the other. Turning to our findings, first consider the Canadian Dollar. The p-values for the null $H^G_{0,F}$ range from 0.11 to 0.33, and the p-values for the null $H^G_{0,S}$ are larger than 0.5 for all values of S_n used, suggesting a failure to reject either null. Statistically speaking, the forward model and the spot model perform equally well in forecasting the 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 these countries. This finding is interesting, as it supports earlier findings due to Hunter and Timme (1992) that *Spot* outperforms *Forward* when one directly tests for first order stochastic dominance using returns resulting from a hedging based trading strategy applied to these two models. 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, i.e. we test $H^C_{0,S}$ and $H^C_{0,F}$. 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 are similar to results arising when testing GL forecast superiority.

In summary, the key finding of this empirical illustration is that our forecast superiority test results are consistent with the conventional point MSFE and MAFE criteria reported in Table 5. However, while the differences between MSFEs and MAFEs when comparing our two models are quite small and likely statistically insignificant, 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 or 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. One drawback of our GL and CL forecast superiority tests is that they only offer a partial ranking when the number of models is greater than 2.

8 Concluding Remarks

This paper outlines a novel approach to forecast comparison that yields a forecast ranking that is 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 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 under the null and under sequences of local alternatives are derived, and it is noted that critical values for the limiting distribution cannot be tabulated. In light of this finding, we show first order validity of critical values based on the stationary bootstrap. Furthermore, we have study the extension of our tests to the case of heterogeneous observations. Findings from a small Monte Carlo study show that the suggested tests have good properties, even for moderate sample sizes. Finally, an empirical illustration in which exchange rate models are used to predict future spot rates is presented. While, no clear cut conclusions can be drawn based on the inspection of mean square and mean absolute error forecast accuracy criteria, our tests indicate that a simple spot rate model is GL-superior to an alternative forward rate type model. A limitation of our testing procedure is that our statistics have non-degenerate limiting distributions only over the least favorable case, under the null. Thus, convergence is not uniform within the probability measures in the null hypotheses. As a consequence, the tests are not asymptotically similar, in the sense of not having exact asymptotic size. In a follow up of this paper, Corradi, Jin and Swanson (2015) use recent developments in testing sequences of inequality restrictions (e.g., Andrews and Barwick (2012), Andrews and Shi (2009,2014), Andrews and Soares (2010), and Linton, Song and Whang (2010)) in order to obtain tests for GL/CL superiority which are similar on the boundary of the null hypothesis. Furthermore, they analyze the trade off between similarity on the boundary and power, see e.g. Lee, Andrews (2012) and Lee, Song and Whang (2013, 2014).

Appendix

Proof of Proposition 2.2: Let f_1 and f_2 be the densities associated with F_1 and F_2. We begin with the IF part.

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

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

We now move to the ONLY IF part. We need to show that whenever, $G(x) > 0$ for all $x \in \Delta$, with $G(x) = (F_2(x)-F_1(x))\,sgn(x)$, and Δ a set of strictly positive Lebesgue measure, there exists function(s) $L \in \mathcal{L}_G$ such that $\int_{\mathcal{X}} L(z)\,(f_1(z)-f_2(z))\,\mathrm{d}z > 0$. In fact, $\int_{\Delta} L'(z)\,(F_1(z)-F_2(z))\,\mathrm{d}z > 0$ and we need a function L whose derivative L' puts enough mass on the set Δ, so that

$$\underbrace{\int_{\Delta} L'(z)\,(F_1(z)-F_2(z))\,\mathrm{d}z}_{>0} + \underbrace{\int_{\Delta^c} L'(z)\,(F_1(z)-F_2(z))\,\mathrm{d}z}_{\leq 0} > 0,$$

with Δ^c being the complement of Δ.

Proof of Proposition 2.3: We begin with the IF part. From the proof of Proposition **2.2**, and by a further integration by parts,

$$\begin{aligned}
& \int_{-\infty}^{\infty} L(z)\,(f_1(z)-f_2(z))\,\mathrm{d}z \\
= & - \int_{-\infty}^{0} L'(z)\,(F_1(z)-F_2(z))\,\mathrm{d}z - \int_{0}^{\infty} L'(z)\,(F_1(z)-F_2(z))\,\mathrm{d}z \\
= & -L'(z) \int_{-\infty}^{z} (F_1(t)-F_2(t))\,\mathrm{d}t \Big|_{-\infty}^{0} + \int_{-\infty}^{0} L''(z) \left(\int_{-\infty}^{z} (F_1(t)-F_2(t))\,\mathrm{d}t \right) \mathrm{d}z \\
& + L'(z) \int_{z}^{\infty} (F_1(t)-F_2(t))\,\mathrm{d}t \Big|_{0}^{\infty} - \int_{0}^{\infty} L''(z) \left(\int_{z}^{\infty} (F_1(t)-F_2(t))\,\mathrm{d}t \right) \mathrm{d}z \\
= & \int_{-\infty}^{0} L''(z) \left(\int_{-\infty}^{z} (F_1(t)-F_2(t))\,\mathrm{d}t \right) \mathrm{d}z - \int_{0}^{\infty} L''(z) \left(\int_{z}^{\infty} (F_1(t)-F_2(t))\,\mathrm{d}t \right) \mathrm{d}z \\
\leq & \; 0,
\end{aligned}$$

since $\int_{-\infty}^{0} (F_1(t)-F_2(t))\,\mathrm{d}t = \int_{0}^{\infty} (F_1(t)-F_2(t))\,\mathrm{d}t = 0$, $\int_{-\infty}^{z} (F_1(t)-F_2(t))\,\mathrm{d}t \leq 0$ for all $z \leq 0$, and $\int_{z}^{\infty} (F_1(t)-F_2(t))\,\mathrm{d}t \geq 0$ for all $z \geq 0$.

As for the ONLY IF part, we need to show that whenever, $C(x) > 0$ for all $x \in \Delta$, with $C(x)$ defined in (2.2) and Δ being a set of strictly positive Lebesgue measure, there exists function(s) $L \in \mathcal{L}_C$ such that $\int_{\mathcal{X}} L(z)\left(f_1(z) - f_2(z)\right) \mathrm{d}z > 0$. Without loss of generality, suppose $\Delta \subset \mathcal{X}^+$, then $\int_{\Delta} L''(z)\left(\int_z^{\infty} (F_1(t) - F_2(t))\,\mathrm{d}t\right) \mathrm{d}z < 0$, and we need to find a function $L \in \mathcal{L}_C$ having a second derivative L'' putting enough weight on Δ such that

$$-\int_{\Delta} L''(z)\left(\int_z^{\infty} (F_1(t) - F_2(t))\,\mathrm{d}t\right) \mathrm{d}z + \int_{\Delta^c} L''(z)\left(\int_{-\infty}^{z} (F_1(t) - F_2(t))\,\mathrm{d}t\right) \mathrm{d}z > 0.$$

Hereafter, we 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. $\sup_t$ denotes $\sup_{R \le t \le T}$ and the summation $\sum_t$ denotes $\sum_{t=R}^{T}$. "var " and "cov" denote variance and covariance. All limits are taken as T goes to infinity.

To help present the proofs of our theorems in Sections 3 and 4, we first fix some additional notation. Denote $\beta = (\beta_1', \beta_k')'$, $\beta_0 = (\beta_{10}', \beta_{k0}')'$, $\widehat{\beta}_t = \left(\widehat{\beta}_{1t}', \widehat{\beta}_{kt}'\right)'$, $N(\varepsilon) = \{\beta : \|\beta - \beta_0\| \le \varepsilon\}$ and $N_k(\varepsilon) = \{\beta_k : \|\beta_k - \beta_{k0}\| \le \varepsilon\}$. Further, we define

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

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

Then we can write

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

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

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

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

where we suppress the dependence of $\xi_{kj}^{i}(\cdot)$ on n for $j = 1, 2, 3$. It is clear that under the nulls $\xi_{k3}^{i}(x) \to -\infty$ 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 the null $H_0^{TG^+}$, 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.

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

Proof of Lemma A.1.

The results follow from Lemma A.1 and the proof of Lemma 2.3.2 of McCracken (2000).

The following lemma holds for all $k = 1, ..., l$.

Lemma A.2. *(a) Suppose 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, $\dot{x} \in \mathcal{X}^+$,

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

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 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

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

Proof of Lemma A.2. We first prove part (a). Without loss of generality (WLOG), we verify the conditions of Theorem 2.2 in Andrews and Pollard (1994) hold with $Q = q$ and $\gamma = 1$ for the case when x, $\dot{x} \in \mathcal{X}^+$, which is bounded on the real line. The mixing condition is implied by Assumption A.1(i). The bracketing condition also holds by the following argument. Let

$$\mathcal{F}_k^{g+} = \{1(e_{k,t}(\beta_k) \le x) : (x, \beta_k) \in \mathcal{X}^+ \times \Theta_{k0}\}.$$

We now show $\mathcal{F}_k^{g+}$ is a class of uniformly bounded functions satisfying the L^2-continuity conditions. Let $\sup_{(\dot{x},\dot{\beta}_k)}$ denote $\sup_{\{(\dot{x},\dot{\beta}_k)\in\mathcal{X}^+\times\Theta_{k0},\ |\dot{x}-x|\le r_1,||\dot{\beta}_k-\beta_k||\le r_2,\sqrt{r_1^2+r_2^2}\le\widetilde{r}\}}$, we have

$$\begin{aligned}
&\sup_t E \sup_{(\dot{x},\dot{\beta}_k)} \left|1(e_{k,t+\tau}(\dot{\beta}_k)\le\dot{x}) - 1(e_{k,t+\tau}(\beta_k)\le x)\right|^2 \\
=\ & E\sup_{(\dot{x},\dot{\beta}_k)} \Big|1(e_{k,t+\tau}\le m_k\left(Z_{k,t+\tau},\dot{\beta}_k\right) - m_k\left(Z_{k,t+\tau},\beta_{k0}\right)+\dot{x}) \\
& -1(e_{k,t+\tau}\le m_k(Z_{k,t+\tau},\beta_k) - m_k(Z_{k,t+\tau},\beta_{k0})+x)| \\
\le\ & E\sup_{(\dot{x},\dot{\beta}_k)} 1\left\{|e_{k,t+\tau} - m_k(Z_{k,t+\tau},\beta_k) + m_k(Z_{k,t+\tau},\beta_{k0}) - x| \le \right. \\
& \left. |m_k(Z_{k,t+\tau},\dot{\beta}_k) - m_k(Z_{k,t+\tau},\beta_k)+\dot{x}-x|\right\} \\
\le\ & E\sup_{(\dot{x},\dot{\beta}_k)} 1\left\{|e_{k,t+\tau} - m_k(Z_{k,t+\tau},\beta_k) + m_k(Z_{k,t+\tau},\beta_{k0}) - x| \le ||M_k(Z_{k,t+\tau},\beta_k^*)||r_2+r_1\right\} \\
\le\ & C\sup_{\beta_k\in\Theta_{k,}} E||M_k(Z_{k,t},\beta_k)||r_2+r_1 \\
\le\ & \widetilde{C}\widetilde{r}.
\end{aligned}\tag{A.8}$$

where β_k^* lies between $\dot{\beta}_k$ and β_k. The first inequality is due to the fact $|1(z\le t)-1(z\le 0)|\le 1(|z|\le|t|)$ for any scalars z and t. The second inequality follows from Assumption A.1(ii), the triangle inequality and the Cauchy-Schwartz inequality. The third inequality holds by Assumptions A.1(ii) and (iii), and $\widetilde{C}=\sqrt{2}C(\sup_{\beta_k\in\Theta_{k0}} E||M_k(Z_{k,t},\beta_k)||\vee 1)$ is finite by Assumption A.1(ii). The desired bracketing condition holds because the L^2-continuity condition implies the bracketing number satisfies

$$N(\varepsilon,\mathcal{F}_k^{g+})\le C(1/\varepsilon)^{L_k+1}.$$

The other cases can be done in the same fashion.

To prove part (b), WLOG, we only verify the case for $x_1\ge 0$ and $x_2\ge 0$. We show that the result follows from Theorem 3 of Hansen (1996) with $a=L_{\max}+1$, $\lambda=1$. Let

$$\mathcal{F}_k^{c+}=\{\int_x^\infty 1(e_{k,t}(\beta_k)>s)ds:(x,\beta_k)\in\mathcal{X}^+\times\Theta_{k0}\}.$$

Then the functions in $\mathcal{F}_k^{c+}$ satisfy the Lipschitz condition:

$$\begin{aligned}
&\left|\int_{\dot{x}}^\infty 1(e_{k,t+\tau}(\dot{\beta}_k)>s)ds - \int_x^\infty 1(e_{k,t+\tau}(\beta_k)>s)ds\right| \\
=\ & \Big|\max\left\{e_{k,t+\tau}+m_k\left(Z_{k,t+\tau},\beta_{k0}\right)-m_k\left(Z_{k,t+\tau},\dot{\beta}_k\right)-\dot{x},0\right\} \\
& -\max\left\{e_{k,t+\tau}+m_k(Z_{k,t+\tau},\beta_{k0})-m_k(Z_{k,t+\tau},\beta_k)-x,0\right\}| \\
\le\ & \left|m_k(Z_{k,t+\tau},\dot{\beta}_k)-m_k(Z_{k,t+\tau},\beta_k)\right|+|\dot{x}-x| \\
\le\ & \sqrt{2}(\sup_{\beta_k\in\Theta_{k0}}||M_k(Z_{k,t+\tau},\beta_k)||\vee 1)(||\dot{\beta}_k-\beta_k||^2+(\dot{x}-x)^2)^{1/2}
\end{aligned}$$

where the first inequality follows from the fact that $|\max\{z_1, 0\} - \max\{z_2, 0\}| \leq |z_1 - z_2|$ and the triangle inequality, and the second inequality holds by Assumption A.1*(ii) and the Cauchy-Schwartz inequality. We have $\max_k \sup_{\beta_k \in \Theta_{k0}} ||M_k(Z_{k,t+\tau}, \beta_k)||_r < \infty$ by Assumption A.1*(ii) which yields the conditions (12) and (13) of Hansen (1996). Finally, the mixing condition (11) in Hansen (1996) holds by Assumption A.1*(i).

Lemma A.3 *Suppose Assumptions A.1, A.1*, and A.4 holds. Denote* $\zeta^i_{k,t+\tau}(x, \beta) = f^i_{k,t+\tau}(x, \beta) - Ef^i_{k,t+\tau}(x, \beta) - f^i_{k,t+\tau}(x, \beta_0) + Ef^i_{k,t+\tau}(x, \beta_0)$, $i = g, c$. Then, for $k = 2, ..., l$,

(a)

$$\begin{aligned}
\sup_t E \sup_{\{\beta\} \in N(n^{-\alpha}\varepsilon)} \sup_{x \in \mathcal{X}^+} [\zeta^i_{k,t+\tau}(x, \beta)]^2 &\leq Cn^{-\alpha}\varepsilon, \\
\sup_t E \sup_{\{\beta\} \in N(n^{-\alpha}\varepsilon)} \sup_{x \in \mathcal{X}^-} [\zeta^i_{k,t+\tau}(x, \beta)]^2 &\leq Cn^{-\alpha}\varepsilon, \ i = g, c
\end{aligned}$$

(b)

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

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

Proof of Lemma A.3.

Part (a) holds directly from the proof of Lemma A.2 by taking $\dot{x} = x$ and $q = 2$ and applying the Cauchy-Schwartz inequality.

For part (b), WLOG, we consider the case $x \geq 0$.

Define $\{x^*, \gamma_1^*, \gamma_2^*\} = \operatorname{argsup}_{\{x \in \mathcal{X}^+, \{\gamma_1, \gamma_2\} \in N(n^{-\alpha}\varepsilon)\}} \zeta^i_{k,t+\tau}(x, \gamma_1) \zeta^i_{k,t+\tau+j}(x, \gamma_2)$, where we suppress the dependence of $(x^*, \gamma_1^*, \gamma_2^*)$ on $i = g$ or c. By the proof of Lemma A.2, it is easy to verify $\left\|\zeta^i_{k,t+\tau}(x^*, \gamma_1^*)\right\|_{2+\delta} \leq \left\|\sup_{\beta \in N(n^{-\alpha}\varepsilon)} \sup_{x \in \mathcal{X}^+} \zeta^i_{k,t+\tau}(x, \beta)\right\|_{2+\delta} = Cn^{-\alpha}\varepsilon$. By Assumptions A.1, A.1* and Corollary 1.1 of Bosq (1996),

$$\begin{aligned}
& |cov(\zeta^g_{k,t+\tau}(x^*, \gamma_1^*), \ \zeta^g_{k,t+\tau+j}(x^*, \gamma_2^*))| \\
\leq \ & 4\alpha(j) \left\| \sup_{\beta \in N(n^{-\alpha}\varepsilon)} \sup_{x \in \mathcal{X}^+} \zeta^g_{k,t+\tau}(x, \beta) \right\|_\infty \left\| \sup_{\beta \in N(n^{-\alpha}\varepsilon)} \sup_{x \in \mathcal{X}^+} \zeta^g_{k,t+\tau+j}(x, \beta) \right\|_\infty \\
\leq \ & C\alpha(j)(n^{-\alpha}\varepsilon)^2,
\end{aligned}$$

and

$$\begin{aligned}
& |cov(\zeta^c_{k,t+\tau}(x^*, \gamma_1^*), \ \zeta^c_{k,t+\tau+j}(x^*, \gamma_2^*))| \\
\leq \ & 2(1 + 2/\delta)(2\alpha(j))^{\delta/(2+\delta)} \left\| \sup_{\beta \in N(n^{-\alpha}\varepsilon)} \sup_{x \in \mathcal{X}^+} \zeta^c_{k,t+\tau}(x, \beta) \right\|_{2+\delta} \left\| \sup_{\beta \in N(n^{-\alpha}\varepsilon)} \sup_{x \in \mathcal{X}^+} \zeta^c_{k,t+\tau+j}(x, \beta) \right\|_{2+\delta} \\
\leq \ & C\alpha(j)^{\delta/(2+\delta)}(n^{-\alpha}\varepsilon)^2.
\end{aligned}$$

This completes the proof.

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

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

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

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

Proof of Lemma A.4. WLOG, we consider the case $x \geq 0$. Denote $\zeta^i_{k,t+\tau}\left(x, \widehat{\beta}_t\right) = f^i_{k,t+\tau}\left(x, \widehat{\beta}_t\right) - Ef^i_{k,t+\tau}(x, \beta)|_{\beta=\widehat{\beta}_t} - f^i_{k,t+\tau}(x, \beta_0) + Ef^i_{k,t+\tau}(x, \beta_0)$, $i = g$, c, then

$$\xi^i_{k1}(x) - \nu^i_{1,n}(x, \beta_{10}) + \nu^i_{k,n}(x, \beta_{k0}) = n^{-1/2} \sum_t \zeta^i_{k,t+\tau}\left(x, \widehat{\beta}_t\right).$$

Fix ε_0, $\delta > 0$. By Lemma A.1 (b), for all $\varepsilon > 0$, there exists T_0 such that for all $T > T_0$, $P\left(\sup_k \sup_t n^\alpha \left\|\widehat{\beta}_{k,t} - \beta_{k0}\right\| > \varepsilon\right) < \delta/2$. It is useful then to note that for all $T > T_0$ and $\varepsilon_0 > 0$,

$$\begin{aligned}
& P\left(\sup_{x\in\mathcal{X}^+} n^{-1/2}\left|\sum_t \zeta^i_{k,t+\tau}\left(x, \widehat{\beta}_t\right)\right| > \varepsilon_0\right) \\
\leq \;& P\left(\sup_{\{\beta_t\}\in N(n^{-\alpha}\varepsilon)} \sup_{x\in\mathcal{X}^+} n^{-1/2}\left|\sum_t \zeta^i_{k,t+\tau}(x, \beta_t)\right| > \varepsilon_0\right) + P\left(\sup_k \sup_t n^\alpha \left\|\widehat{\beta}_{k,t} - \beta_{k0}\right\| > \varepsilon\right) \\
\leq \;& P\left(\sup_{\{\beta_t\}\in N(n^{-\alpha}\varepsilon)} \sup_{x\in\mathcal{X}^+} n^{-1/2}\left|\sum_t \zeta^i_{k,t+\tau}(x, \beta_t)\right| > \varepsilon_0\right) + \delta/2 \qquad \text{(A.11)}
\end{aligned}$$

where $\{\beta_t\} \equiv \{\beta_t\}_{t=R}^T$ is a nonrandom sequence. Now we show that there exists $T_1 > T_0$ such that for all $T > T_1$, the first term on the right hand side (r.h.s.) of (A.11) is less than $\delta/2$. For the remainder of this proof only, let $\sum_j$ denote the summation $\sum_{-n+1\leq j\neq 0\leq n-1}$. Applying the Chebyshev's inequality, we have

$$
\begin{aligned}
& \varepsilon_0^2 P\left(\sup_{\{\beta_t\}\in N(n^{-\alpha}\varepsilon)} \sup_{x\in\mathcal{X}^+} n^{-1/2}\left|\sum_t \zeta^i_{k,t+\tau}(x,\beta_t)\right| > \varepsilon_0\right) \\
\leq\ & E\left(\sup_{\{\beta_t\}\in N(n^{-\alpha}\varepsilon)} \sup_{x\in\mathcal{X}^+} n^{-1/2}\left|\sum_t \zeta^i_{k,t+\tau}(x,\beta_t)\right|\right)^2 \\
=\ & E\left(\sup_{\{\beta_t\}\in N(n^{-\alpha}\varepsilon)} \sup_{x\in\mathcal{X}^+} n^{-1}\sum_t \left[\zeta^i_{k,t+\tau}(x,\beta_t)\right]^2\right) \\
& + E\left(\sup_{\{\beta_t,\beta_{t+j}\}\in N(n^{-\alpha}\varepsilon)} \sup_{x\in\mathcal{X}^+} \sum_j \left[n^{-1}\sum_{t=R}^{T-|j|} \zeta^i_{k,t+\tau}(x,\beta_t)\zeta^i_{k,t+\tau+j}\left(x,\beta_{t+j}\right)\right]\right) \\
\leq\ & E\left(\sup_{\{\beta_t\}\in N(n^{-\alpha}\varepsilon)} \sup_{x\in\mathcal{X}^+} n^{-1}\sum_t \left[\zeta^i_{k,t+\tau}(x,\beta_t)\right]^2\right) \\
& + \sum_j \left\{n^{-1}\sum_{t=R}^{T-|j|}\left|E\left[\sup_{\{\beta_t,\beta_{t+j}\}\in N(n^{-\alpha}\varepsilon)} \sup_{x\in\mathcal{X}^+} \zeta^i_{k,t+\tau}(x,\beta_t)\zeta^i_{k,t+\tau+j}(x,\beta_{t+j})\right]\right|\right\}. \qquad \text{(A.12)}
\end{aligned}
$$

For part (a), substituting the results of Lemma A.3 into (A.12), the r.h.s. of (A.12) is less than or equal to

$$
\begin{aligned}
& \widetilde{C}(n^{-\alpha}\varepsilon) + \sum_j (1-|j|/n)\widetilde{C}\alpha(j)(n^{-\alpha}\varepsilon)^2 \\
\leq\ & \widetilde{C}n^{-\alpha}\varepsilon\left\{1 + 2\sum_{j=1}^{n-1}\alpha(j)\right\} \\
\leq\ & Cn^{-\alpha}\varepsilon, \text{ say,} \qquad \text{(A.13)}
\end{aligned}
$$

provided $0 < n^{-\alpha}\varepsilon < 1$. Where $0 < C \equiv \left\{1 + 2\sum_{j=1}^{\infty}\alpha(j)\right\}\widetilde{C} < \infty$. Thus we can choose T_1 and ε such that for all $T > T_1 > T_0$, $\varepsilon < (\delta\varepsilon_0^2 n^{\alpha}/2C)$ and $0 < n^{-\alpha}\varepsilon < 1$, the result follows.

Similarly, for part (b), (A.13) holds by Lemma A.3 if we replace $\alpha(j)$ by $\alpha(j)^{\delta/(2+\delta)}$. In this case, $0 < C \equiv \left\{1 + 2\sum_{j=0}^{\infty}\alpha(j)^{\delta/(2+\delta)}\right\}\widetilde{C} < \infty$ by Assumption A.1* (see Eq. (14.6) in Davidson, 1994). Then the result follows analogously.

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

$$
\begin{aligned}
\sup_{x\in\mathcal{X}^+}\left|\xi^g_{k2}(x) - \sqrt{n}\Delta'_{k0}(x)B_k\overline{H}_{k,n} + \sqrt{n}\Delta'_{10}(x)B_1\overline{H}_{1,n}\right| &= o_p(1), \qquad \text{(A.14)} \\
\sup_{x\in\mathcal{X}^-}\left|\xi^g_{k2}(x) - \sqrt{n}\Delta'_{k0}(x)B_k\overline{H}_{k,n} + \sqrt{n}\Delta'_{10}(x)B_1\overline{H}_{1,n}\right| &= o_p(1). \qquad \text{(A.15)}
\end{aligned}
$$

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

$$
\begin{aligned}
\sup_{x\in\mathcal{X}^+}\left|\xi^c_{k2}(x) - \sqrt{n}\Lambda'_{k0}(x)B_k\overline{H}_{k,n} + \sqrt{n}\Lambda'_{10}(x)B_1\overline{H}_{1,n}\right| &= o_p(1), \qquad \text{(A.16)} \\
\sup_{x\in\mathcal{X}^-}\left|\xi^c_{k2}(x) - \sqrt{n}\Lambda'_{k0}(x)B_k\overline{H}_{k,n} + \sqrt{n}\Lambda'_{10}(x)B_1\overline{H}_{1,n}\right| &= o_p(1). \qquad \text{(A.17)}
\end{aligned}
$$

Proof of Lemma A. 5. We first prove part (a). Recall that $\Delta_{k0}(x) = (\partial F_k(x,\beta_{k0})/\partial\beta)sgn(x)$ and $\xi^g_{k2}(x) = n^{-1/2}\sum_{t=R}^{T}\left[F_k\left(x,\widehat{\beta}_{k,t}\right) - F_k(x,\beta_{k0}) - F_1\left(x,\widehat{\beta}_{1,t}\right) + F_1(x,\beta_{10})\right] sgn(x)$, WLOG, we consider the case $x \geq 0$ and prove

$$\sup_{x\in\mathcal{X}^+}\left|n^{-1/2}\sum_t\left(F_k\left(x,\widehat{\beta}_{k,t}\right) - F_k(x,\beta_{k0})\right) - \sqrt{n}\left(\frac{\partial F_k(x,\beta_{k0})}{\partial\beta_k'}\right)B_k\overline{H}_{k,n}\right| = o_p(1). \tag{A.18}$$

By Assumption A.3 (i) and the mean value theorem,

$$n^{-1/2}\sum_t\left\{F_k\left(x,\widehat{\beta}_{k,t}\right) - F_k(x,\beta_{k0})\right\} = n^{-1/2}\sum_t\left(\frac{\partial F_k(x,\beta^*_{k,t}(x))}{\partial\beta_k'}\right)\left(\widehat{\beta}_{k,t} - \beta_{k0}\right),$$

where $\beta^*_{k,t}(x)$ lies between $\widehat{\beta}_{k,t}$ and β_{k0}. By Lemma A.1 (b), for all $\alpha \in [0,1/2)$ and all $\varepsilon > 0$, there exists δ, such that $P(\sup_t \sup_{x\in\mathcal{X}^+} n^\alpha||\beta^*_{k,t}(x) - \beta_{k0}|| \leq \varepsilon) < \delta/2$ for sufficiently large n. Let

$$A_{1n} = \sup_{x\in\mathcal{X}^+}\sup_{\{\beta_k\}\in N_k(n^{-\alpha}\varepsilon)}\left\|\frac{\partial F_k(x,\beta_k)}{\partial\beta_k} - \frac{\partial F_k(x,\beta_{k0})}{\partial\beta_k}\right\|.$$

Then $A_{1n} = O(n^{-\eta\alpha})$ by Assumption A.3(ii).

$$\begin{aligned}
A_{2n} &\equiv \sup_{x\in\mathcal{X}^+}\left\|n^{-1}\sum_t\frac{\partial F_k(x,\beta^*_{k,t}(x))}{\partial\beta_k} - \frac{\partial F_k(x,\beta_{k0})}{\partial\beta_k}\right\| \\
&\leq \sup_{x\in\mathcal{X}^+}\sup_t\left\|\frac{\partial F_k(x,\beta^*_{k,t}(x))}{\partial\beta_k} - \frac{\partial F_k(x,\beta_{k0})}{\partial\beta_k}\right\| = O_p(n^{-\eta\alpha}).
\end{aligned}$$

where the last equality holds because $P(A_{2n} \leq A_{1n}) \to 1$ as $n \to \infty$ by construction. Now we have the desired result

$$\begin{aligned}
&\sup_{x\in\mathcal{X}^+}\left|n^{-1/2}\sum_t\left(F_k(x,\widehat{\beta}_{k,t}) - F_k(x,\beta_{k0})\right) - n^{1/2}\left(\frac{\partial F_k(x,\beta_{k0})}{\partial\beta'}\right)B_k\overline{H}_{k,n}\right| \\
=\ &\sup_{x\in\mathcal{X}^+}\left|n^{-1/2}\sum_t\left(\frac{\partial F_k(x,\beta^*_{k,t}(x))}{\partial\beta_k'}\right)\left(\widehat{\beta}_{k,t} - \beta_{k0}\right) - n^{1/2}\left(\frac{\partial F_k(x,\beta_{k0})}{\partial\beta'}\right)B_k\overline{H}_{k,n}\right| \\
\leq\ &\sup_{x\in\mathcal{X}^+}\left|n^{-1/2}\sum_t\left(\frac{\partial F_k(x,\beta^*_{k,t}(x))}{\partial\beta_k'} - \frac{\partial F_k(x,\beta_{k0})}{\partial\beta'}\right)\left(\widehat{\beta}_{k,t} - \beta_{k0}\right)\right| \\
&+\sqrt{n}\sup_{x\in\mathcal{X}^+}\left|\left(\frac{\partial F_k(x,\beta_{k0})}{\partial\beta'}\right)n^{-1}\sum_t\left(\widehat{\beta}_{k,t} - \beta_{k0}\right) - \left(\frac{\partial F_k(x,\beta_{k0})}{\partial\beta'}\right)B_k\overline{H}_{k,n}\right| \\
\leq\ &A_{2n}\sup_t\left\|\sqrt{n}\left(\widehat{\beta}_{k,t} - \beta_{k0}\right)\right\| + \sup_{x\in\mathcal{X}^+}\left\|\frac{\partial F_k(x,\beta_{k0})}{\partial\beta'}\right\|\left\|n^{-1/2}\sum_t\left(\widehat{\beta}_{k,t} - \beta_{k0}\right) - B_k\sqrt{n}\overline{H}_{k,n}\right\| \\
=\ &o_p(1) + o_p(1) = o_p(1)
\end{aligned}$$

where the first $o_p(1)$ follows from the fact that $A_{2n}\sup_{t=R,\ldots,T}\left\|\sqrt{n}\left(\widehat{\beta}_{k,t} - \beta_{k0}\right)\right\| = O_p(n^{-\alpha(1+\eta)+1/2}) = o_p(1)$ for all $\alpha \in (1/2(1+\eta), 1/2)$ by Lemma A.1(b), and the second $o_p(1)$ holds by Assumption A.3(iii), Lemma A.1(c) and the following argument

$$
\begin{aligned}
\left\| n^{-1/2} \sum_{t=R}^{T} \left(\widehat{\beta}_{k,t} - \beta_{k0}\right) - B_k \sqrt{n} \overline{H}_{k,n} \right\| &= \left\| n^{-1/2} \sum_{t=R}^{T} B_k(t) H_k(t) - B_k n^{-1/2} \sum_{t=R}^{T} H_k(t) \right\| \\
&= \left\| n^{-1/2} \sum_{t=R}^{T} (B_k(t) - B_k) H_k(t) \right\| \\
&\leq \sup_t \|B_k(t) - B_k\| \sup_t n^{1/2} \|H_k(t)\| \\
&= o_p(1) O_p(1) = o_p(1).
\end{aligned}
$$

The proof of part (b) is similar and thus omitted.

Lemma A.6. *(a) Suppose 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 $\mathcal{X}$ with probability one, where for $x_1, x_2 \in \mathcal{X}^+$ or $x_1, x_2 \in \mathcal{X}^-$,

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

(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 $\mathcal{X}$ with probability one, where for $x_1, x_2 \in \mathcal{X}^+$ or $x_1, x_2 \in \mathcal{X}^-$,

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

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

Proof of Lemma A.6. We first prove (a). By Theorem 10.2 of Pollard (1990), the results hold if we have (i) total boundedness of the pseudometric space $\left(\mathcal{X}, \rho_g\right)$, (ii) stochastic equicontinuity of $\left\{ v^g_{k,n}(\cdot, \beta_{k0}) - v^g_{1,n}(\cdot, \beta_{10}) : n \geq 1 \right\}$ and (iii) finite dimensional (fidi) convergence. The first two conditions follow from Lemma A.2. We now verify condition (iii), i.e., we need to show that $(v^g_{k,n}(x_1, \beta_{k0})$ $-v^g_{1,n}(x_1, \beta_{10}), ..., v^g_{k,n}(x_J, \beta_{10}) - v^g_{1,n}(x_J, \beta_{k0}), \sqrt{n}\overline{H}'_{k,n}, \sqrt{n}\overline{H}'_{1,n})'$ converges in distribution to $(\widetilde{g}_k(x_1),$ $..., \widetilde{g}_k(x_J), v'_{k0}, v'_{10})'$ $\forall x_1, ..., x_J \in \mathcal{X}^+$ or $x_1, ..., x_J \in \mathcal{X}^-$, and $\forall J \geq 1$. The central limit theorem (CLT) holds for $\sqrt{n}\overline{H}_{k,n}$ by Lemma 4.1 in West (1996). A CLT for bounded random variables under $\alpha-$mixing conditions (see Hall and Heyde, 1980) hold for $v^g_{k,n}(x_j,\beta_{k0})-v^g_{1,n}(x_j,\beta_{10})$, $j=1,...,J$. Then one obtains the above weak convergence result by the Cramer-Wold device. This establishes part (a).

For part (b), we need to verify the fidi convergence again. Note that the moment condition of Hall and Heyde (1980, Corollary 5.1) holds since (WLOG), for $x>0$,

$$E\left|\int_x^\infty (1(e_{1,t}>s)ds-1(e_{k,t}>s))ds\right|^{2+\delta}\leq E\left|e_{1,t}-e_{k,t}\right|^{2+\delta}<\infty.$$

The mixing condition also holds since we have $\sum\alpha(j)^{\delta/(2+\delta)}\leq C\sum j^{-M\delta/(2+\delta)}<\infty$ by Assumption A.1*.

Proof of Theorem 3.1

WLOG, we consider the case $x\geq 0$. To prove part (a), note that

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

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

$$\begin{aligned}TG^+_{k,n} &\equiv \sup_{x\in\mathcal{X}^+}\sqrt{n}D^g_{k,n}(x)\\ &\Rightarrow \sup_{x\in\mathcal{B}^{g+}_k}\left[\widetilde{g}_k(x)+\Delta_{k0}(x)'B_kv_{k0}-\Delta_{10}(x)'B_1v_{10}\right]\end{aligned}$$

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

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

To prove part (b), note that

$$TC^+_n=\max_{k=2,..,l}\sup_{x\in\mathcal{X}^+}D^c_{k,n}(x)=\min_{k=2,..,l}\sup_{x\in\mathcal{X}^+}\left\{\xi^c_{k1}(x)+\xi^c_{k2}(x)+\xi^c_{k3}(x)\right\}.$$

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

$$\begin{aligned}TC^+_{k,n} &\equiv \sup_{x\in\mathcal{X}^+}D^c_{k,n}(x)\\ &\Rightarrow \sup_{x\in\mathcal{B}^{c+}_k}\left[\widetilde{g}_k(x)+\Lambda_{k0}(x)'B_kv_{k0}-\Lambda_{10}(x)'B_1v_{10}\right]\end{aligned}$$

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

Next suppose $TC^+<0$. In this case, the set $\mathcal{B}^{c+}_k$ is empty and hence $n^{-1/2}\xi^c_{k3}(x)<0\ \forall x\in\mathcal{X}^+$, for some $k\in\{2,...,l\}$. Then for such k, $D^c_{k,n}(x)$ will be dominated by the term $\xi^c_{k3}(x)$ 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 gives

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

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

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

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 $\varsigma_{3,n}(x) = \varsigma^k_{3,n}(x) - \varsigma^1_{3,n}(x)$, where $\varsigma^k_{3,n}(x) = n^{-1/2}\sum_t \left[1\left(e_{k,\theta(t)+\tau}\left(\widehat{\beta}_{k,\theta(t)}\right) \le x\right) - 1(e_{k,t+\tau} \le x)\right]$, and $\sup_{x\in\mathcal{X}^+}|\varsigma_{3,n}(x)| \le \sup_x \left|\varsigma^k_{3,n}(x)\right| + \sup_x\left|\varsigma^1_{3,n}(x)\right|$. By the Markov inequality it suffices to show $E_Q|\sup_x \varsigma^k_{3n}(x)| = o_p(1)$, where E_Q is the expectation induced by the probability measure Q. Note that

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

It suffices to show $E[\varsigma_n] = o(1)$ by the Markov inequality and the nonnegativity of ς_n. Denote the jth elements of $\widehat{\beta}_{k,t}$ and β_{k0} as $\widehat{\beta}^{(j)}_{k,t}$ and $\beta^{(j)}_{k0}$ 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 $\max_j \sup_t \left|n^{1/2}\left(\widehat{\beta}_{k,t}^{(j)} - \beta_{k0}^{(j)}\right)\right| = o_{a.s.}(1)$. Fix ϵ_0, $\delta > 0$. Then for all $\epsilon > 0$, there exists 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 $\varepsilon_0 > 0$,

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

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 is to show that there exists $T_1 > T_0$ such that for all $T > T_1$, $\psi_n < \delta/2$. Applying the Markov inequality, we have

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

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 follows immediately from Theorem 3.2 and 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,

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

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

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

$\widetilde{TG}_{1,n}^+ = O_p(1)$ and it has the limiting distribution (3.3) by the proof of Theorem 3.1 (a) and $\widetilde{TG}_{2,n}^+ = o_p(1)$ by the proof of Theorem 3.2. Consequently, $\widetilde{q}_{n,S_n}^{G+}(1-\alpha) \xrightarrow{p} q^{G+}(1-\alpha)$ and

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

Proof of Theorem 4.2.

The proof is similar to that of Theorem 3.1. Consider Lemmas A.1-A.6 with $v_{k,n}^g(x, \beta_k)$ now defined by

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

Then by contiguity, the result of Lemmas 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 using the CLT of Herrndorf (1984) for $\alpha-$mixing arrays to verify the fidi convergence condition of Theorem 10.2 of Pollard (1990).

Proof of Corollary 4.3

By contiguity, $q^{G+}_{n,S_n}(1-\alpha)\overset{p}{\to}q^{G+}(1-\alpha)$ under the local alternatives. The result now follows immediately from Theorem **4.2**.

Proof of Theorem 5.4

To prove Theorem 5.4, we need the following two lemmas.

Lemma HA.1 *(a) Suppose 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, $\dot{x}\in\mathcal{X}^-$,

$$\overline{\lim_{T\to\infty}}\left\|\sup_{\rho^*_{hg}(x,\dot{x}))<\delta}|\nu^{hg}_{k,n}(x)-\nu^{hg}_{k,n}(\dot{x})|\right\|_q<\varepsilon, \tag{A.19}$$

where

$$\rho^*_{hg}(x,\dot{x})=\{E[1(e_{k,t+\tau}\leq x)-1(e_{k,t+\tau}\leq\dot{x})]^2\}^{1/2}. \tag{A.20}$$

(b) Suppose 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, $\dot{x}\in\mathcal{X}^-$,

$$\overline{\lim_{T\to\infty}}\left\|\sup_{\rho^*_{hc}(x,\dot{x})<\delta}|\nu^{hc}_{k,n}(x)-\nu^{hc}_{k,n}(\dot{x})|\right\|_q<\varepsilon, \tag{A.21}$$

where

$$\begin{aligned}\rho^*_{hc}(x,\dot{x})=&\left\{E\left|\int_{-\infty}^{x}1(e_{k,t+\tau}\leq s)ds-\int_{-\infty}^{\dot{x}}1(e_{k,t+\tau}\leq s)ds\right|^r\right\}^{1/r}1(x<0,\ \dot{x}<0)\\
&+\left\{E\left|\int_{x}^{\infty}1(e_{k,t+\tau}>s)ds-\int_{\dot{x}}^{\infty}1(e_{k,t+\tau}>s)ds\right|^r\right\}^{1/r}1(x\geq0,\ \dot{x}\geq0).\end{aligned} \tag{A.22}$$

Proof of Lemma HA.1 Assumptions HA.4 and HA.1 (i) (resp. HA.1*(i)) imply that $\{e_{k,t+\tau} : t \geq R\}$ is an $\alpha-$mixing sequence with mixing coefficients $\alpha(l) = O(l^{-C_0})$, where C_0 is as defined in HA.1 (resp. HA.1*). Note that Theorem 2.2 in Andrews and Pollard (1994) and Theorem 3 in Hansen (1996) do not require any stationarity assumption, the proof is analogous to that of Lemma A.2. For example, for part (b), Eq. (12) of Hansen (1996) is satisfied with our mixing coefficient $C_0 = 1/q - 1/r$, Eq. (12) is true by Assumption HA.1 (ii) and Equation (13) is satisfied with the dominating function b =1. Then theorem 3 in Hansen (1996) follows by taking $a = 1$ and $\lambda = 1$.

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

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

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 $x_1, x_2 \in \mathcal{X}^-$,

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

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

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

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

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

Proof of Lemma HA.2. The proof is analogous to that of Lemma A.6. The total boundedness of the pseudometric space $(\mathcal{X}, \rho_i)$, $i = hg$ and hc, and the stochastic equicontinuity of $\left\{v_{k,n}^i(\cdot) - v_{1,n}^i(\cdot) : n \geq 1\right\}$, $i = hg$ and hc follow from Lemma HA.1. The finite dimensional convergence follows from Hall and Heyde (1980). Then the result follows from Theorem 10.2 of Pollard (1990).

Proof of Theorem 5.4

The theorem follows from Lemmas HA.1 and HA.2 and the CMT.

References

[1] Andrews, D.W.K. (2012), Similar-on-The-Boundary Tests for Moment Inequalities Exist, but Have Poor Power, Cowles Foundation Working Paper 1815R.

[2] Andrews, D. W. K., and P. J. Barwick (2012), Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure, *Econometrica* 80, 2805-2826.

[3] Andrews, D. W. K. and D. Pollard (1994), An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes, *International Statistical Review* 62, 119-132.

[4] Andrews, D. W. K., and X. Shi (2009), Inference Based on Conditional Moment Inequalities, *Econometrica* 81, 609-666.

[5] Andrews, D. W. K., and X. Shi (2014), Nonparametric inference based on conditional moment inequalities, *Journal of Econometrics* 179, 31-45.

[6] Andrews, D. W. K., and G. Soares (2010), Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection, *Econometrica* 78, 119-157.

[7] Bosq, D. (1996), *Nonparametric Statistics for Stochastic Processes*, Springer-Verlag, New York.

[8] Chao, J. C., V. Corradi and N. R. Swanson (2001), An out of Sample Test for Granger Causality, *Macroeconomic Dynamics* 5, 598-620.

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

[10] Christoffersen, P. (1998), Evaluating Interval Forecasts, *International Economic Review* 39, 841-862.

[11] Clark, T. E. and M. W. McCracken (2001), Tests of Equal Forecast Accuracy and Encompassing for Nested Models, *Journal of Econometrics* 105, 85-110.

[12] Clark, T. E. and M. W. McCracken (2013), Advances in Forecast Evaluation, in G. Elliott and A. Timmerman (eds.), *Handbook of Economic Forecasting*, Volume 2, Elsevier, Amsterdam.

[13] Clements, M. P. and D. F. Hendry (1993), On the Limitations of Comparing Mean Square Forecast Errors, *Journal of Forecasting* 12, 617-637.

[14] Clements, M. P. and D. F. Hendry (1999), On Winning Forecasting Competitions in Economics, *Spanish Economic Review* 1, 123-160.

[15] Clements, M. P. and D. F. Hendry (2001), Forecasting with Difference and Trend Stationary Models, *Econometrics Journal* 4, S1–S19.

[16] Corradi, V., S. Jin, and N. R. Swanson (2015), Similar Robust Tests for Forecast Superiority, working paper.

[17] Corradi, V. and N. R. Swanson (2002), A Consistent Test for Nonlinear out of Sample Predictive Accuracy, *Journal of Econometrics 110*, 353-381.

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

[19] Corradi, V. and N. R. Swanson (2006a), Predictive Density and Conditional Confidence Interval Accuracy Tests, *Journal of Econometrics* 135, 187-228.

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

[21] Corradi, V. and W. Distaso (2011), Multiple Forecast Evaluation, in D. F. Hendry and M. P. Clements (eds), *Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford.

[22] 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.

[23] Corradi, V. , N. R. Swanson and C. Olivetti (2001), Predictive Ability with Cointegrated Variables, *Journal of Econometrics* 104, 315-358.

[24] Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press, New York.

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

[26] Diebold, F. X. (2014), Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests, (with Discussion), *Journal of Business and Economic Statistics*, in press.

[27] Diebold, F.X. and M. Shin (2014a). Assessing Point Forecast Accuracy by Stochastic Loss Distance. *Economics Letters*, forthcoming.

[28] Diebold, F.X. and M. Shin (2014b). Assessing Point Forecast Accuracy by Stochastic Error Distance. Working Paper. University of Pennsylvania.

[29] Elliott, G., A. Gargano, and A. Timmermann (2013) Complete Subset Regressions, *Journal of Econometrics* 177, 357-373.

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

[31] Frankel, J. A. and A. K. Rose (1995), Empirical Research on Nominal Exchange Rates, in G. Grossman and K. Rogoff (eds.), *Handbook of International Economics*, Volume III, 1689-1729.

[32] Giacomini, R. and I. Komunjer (2005), Evaluation and Combination of Conditional Quantile Forecasts, *Journal of Business and Economic Statistics* 23, 416-431.

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

[34] Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *Review of Economic Studies*, 76, 669-705

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

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

[37] Granger, C. W. J. (1999b), *Empirical Modeling in Economics: Specification and Evaluation*, Cambridge University Press, New York.

[38] Granger, C. W. J. and M. J. Machina (2006), Forecasting and Decision Theory, *Handbook of Economic Forecasting*, Elsevier.

[39] Granger, C. W. J. and M. H. Pesaran (2000), Economic and Statistical Measures of Forecast Accuracy, *Journal of Forecasting* 19, 537-560

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

[41] Hall, P. and C. C. Heyde (1980), Martingale Limit Theory and Its Applications, Academic Press.

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

[43] Hansen, P. R. (2003), Asymptotic Tests of Composite Hypotheses, Working paper, Dept. of Economics, Brown University.

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

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

[46] Harvey, D. I., S. J. Leybourne and P. Newbold (1997), Tests for Forecast Encompassing, *Journal of Business and Economic Statistics* 16, 254-259.

[47] Hendry, D. F. and M. P. Clements (2002), Pooling of Forecasts, *Econometrics Journal* 5, 1-26.

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

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

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

[51] Inoue, A and L. Kilian (2005), In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?, *Econometric Reviews* 23, 371-402.

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

[53] Lee, S., K. Song, and Y. J. Whang (2013), Testing Functional Inequalities, *Journal of Econometrics* 172, 14-32.

[54] Lee, S., K. Song, and Y. J. Whang (2014), Testing for a General Class of Functional Inequalities, Working paper, CWP09/14.

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

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

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

[58] 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.

[59] Meese, R., and K. Rogoff (1988), Was It Real? The Exchange Rate-Interest Differential Relation over the Modern Floating-Rate Period, *Journal of Finance*, 933-947.

[60] Newbold, P. and D. I. Harvey (2002), Forecast Combination and Encompassing, in M. P. Clements and D. F. Hendry (eds.), *A Companion to Economic Forecasting*, Oxford: Blackwell Press.

[61] Pesaran, M. H. and S. Skouras (2002), Decision-Based Methods for Forecast Evaluation, in M. P. Clements, and D. F. Hendry (eds), *A Companion to Economic Forecasting*, Oxford: Blackwell.

[62] Phillips, P. C. B. (1996), Econometric Model Determination, *Econometrica* 64, 763-812.

[63] Politis, D. N., J. P. Romano (1994a), The Stationary Bootstrap, Journal of the American *Statistical Association* 89, 1303-1313.

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

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

[66] Romano, J. P. and M. Wolf (2005), Stepwise Multiple Testing as Formalized Data Snooping, *Econometrica* 73, 1237-1282.

[67] Stock, J. H. and M. W. Watson (1999), Forecasting Inflation, *Journal of Monetary Economics* 44, 293–335.

[68] Stock, J. H. (2001), Forecasting Economic Time Series, in B. Baltagi (eds.), *Companion in Theoretical Econometrics*, Basil Blackwell.

[69] Sullivan, R., A. Timmermann and H. White (1999), Data-snooping, Technical Trading Rule Performance, and the Bootstrap, *Journal of Finance* 54, 1647-1691.

[70] Timmermann, A. (2006) Forecast Combinations, in G. Elliot, C. W. J. Granger, and A. Timmermann, (Eds.), *Handbook of economic forecasting,* Elsevier, Amsterdam.

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

[72] West, K. D. and M. W. McCracken (1998), Regression Based Tests of Predictive Ability, *International Economic Review* 39, 817-840.

[73] West, K. D. (2006), Forecast Evaluation, in G. Elliot, C. W. J. Granger, and A. Timmermann, (Eds.), *Handbook of economic forecasting,* Elsevier, Amsterdam.

[74] 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)*

S_n	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6
	GL forecast superiority						CL forecast superiority					
	n=100						n=100					
0.63	0.105	0.836	0.862	0.092	0.855	0.670	0.089	0.761	0.908	0.109	0.956	0.785
0.54	0.097	0.830	0.856	0.101	0.854	0.645	0.111	0.777	0.926	0.096	0.960	0.805
0.44	0.112	0.850	0.871	0.091	0.860	0.642	0.103	0.770	0.938	0.109	0.957	0.817
0.35	0.099	0.824	0.874	0.116	0.855	0.680	0.106	0.780	0.932	0.117	0.945	0.818
0.25	0.113	0.841	0.882	0.128	0.839	0.691	0.103	0.796	0.940	0.107	0.959	0.813
0.16	0.121	0.859	0.887	0.130	0.886	0.703	0.120	0.809	0.936	0.114	0.970	0.819
	n=500						n=500					
0.537	0.114	1.000	1.000	0.093	1.000	0.996	0.101	1.000	1.000	0.113	1.000	1.000
0.446	0.097	1.000	1.000	0.090	1.000	0.998	0.105	1.000	1.000	0.092	1.000	1.000
0.356	0.093	1.000	1.000	0.108	1.000	1.000	0.104	1.000	1.000	0.108	1.000	1.000
0.265	0.092	1.000	1.000	0.101	1.000	1.000	0.089	1.000	1.000	0.097	1.000	1.000
0.174	0.106	1.000	1.000	0.094	1.000	0.997	0.097	1.000	1.000	0.095	1.000	1.000
0.083	0.096	1.000	1.000	0.106	1.000	0.996	0.101	1.000	1.000	0.114	1.000	1.000
	n=1000						n=1000					
0.50	0.097	1.000	1.000	0.097	1.000	1.000	0.109	1.000	1.000	0.101	1.000	1.000
0.41	0.106	1.000	1.000	0.103	1.000	1.000	0.104	1.000	1.000	0.092	1.000	1.000
0.33	0.077	1.000	1.000	0.097	1.000	1.000	0.104	1.000	1.000	0.114	1.000	1.000
0.24	0.094	1.000	1.000	0.089	1.000	1.000	0.112	1.000	1.000	0.107	1.000	1.000
0.15	0.107	1.000	1.000	0.086	1.000	1.000	0.093	1.000	1.000	0.085	1.000	1.000
0.06	0.108	1.000	1.000	0.110	1.000	1.000	0.108	1.000	1.000	0.091	1.000	1.000

*Notes: See Sections 3 and 6 for complete details. Size Experiments: DGP1 and DGP4. Power Experiments: DGP2, DGP3, DGP5, 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.

Table 5: MSFE and MAFE of Spot and Forward Exchange Rate Forecasting Models*

	CAD	FRF	GEM	JPY	CHF	GBP
MSFE						
Forward Model	0.001	0.099	0.009	59.526	0.008	0.001
Spot Model	0.001	0.093	0.009	57.610	0.007	0.001
Ratio	1.054	1.059	1.051	1.033	1.054	0.974
MAFE						
Forward Model	0.026	0.249	0.077	6.079	0.070	0.019
Spot Model	0.026	0.241	0.075	5.877	0.068	0.019
Ratio*	1.022	1.033	1.024	1.034	1.031	1.005

*Notes: Ratio is the ratio of forward and spot model MSFE (MAFE). See Section 7 for complete details.