Robust Forecast Superiority Testing with an Application to Assessing Pools of Expert Forecasters

Valentina Corradi[1], Sainan Jin[2], and Norman R. Swanson[3]

[1]University of Surrey, [2]Singapore Management University, and [3]Rutgers University

September 2018

Abstract

Jin, Corradi and Swanson (JCS: 2017) develop a forecast superiority testing methodology which is robust to the choice of loss function. They do this by establishing a mapping between generic loss evaluation and stochastic dominance principles. However, the tests that they develop are not uniformly valid, and have correct asymptotic size only under the least favorable case. Since tests for stochastic dominance can be seen as tests for infinitely many moment inequalities, we use tools from Andrews and Shi (2013,2017) to develop new tests for robust forecast comparison which are uniformly asymptotically valid and asymptotically non-conservative. The extant (many) moment inequality results that we utilize to this end are valid for *iid* observations. However, forecast errors in our set-up may be non martingale difference sequences, because of dynamic misspecification. We thus establish uniform convergence (over error support) of HAC variance estimators and of their bootstrap counterparts. Furthermore, we extend the asymptotic validity of generalized moment selection tests to the case of non-vanishing recursive parameter estimation error. The suggested testing methodology is evaluated in a series of Monte Carlo experiments, and is used to analyze the Survey of Professional Forecasters (SPF). Our Monte Carlo experiments indicate improvements in finite sample performance, relative to the tests described in JCS (2017). Our empirical findings indicate that experience and forecast quality matters in the SPF. Namely, combining predictions from pools of expert forecasters chosen according to recent forecast "quality" yields improvement relative to mean or median consensus forecast. On the other hand, forecast combinations from expert pools chosen solely according to experience do not outperform combinations that utilize predictions from the entire expert pool.

JEL Classification: C12, C22, C53.

Keywords: Robust Forecast Evaluation, Many Moment Inequalities, Bootstrap, Estimation Error, Combination Forecasts, Survey of Professional Forecasters.

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 R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. We are grateful to Kevin Lee, Patrick Marsh, Luis Martins, Jams Mitchell, Alessia Paccagini, Paulo Parente, Ivan Petrella, Valerio Poti, Barbara Rossi, Simon Van Norden, Claudio Zoli, and to the participants at the 2018 NBER-NSF Times Series Conference, the 2016 European Meeting of the Econometric Society, Conference for 50 years of Keynes College at Kent University, and seminars at Mannheim University, the University of Nottingham, University College Dublin, Instituto Universitário de Lisboa, Universita' di Verona and the Warwick Business School for useful comments and suggestions. Additionally, many thanks are owed to Mingmian Cheng for excellent research assistance.

1 Introduction

Forecast accuracy is typically measured in terms of a given loss function. This approach has three main drawbacks. First, it is well known that the relative ranking of forecasts from misspecified models is loss function dependent. Second, sequences of forecasts evaluated with the same synthetic measure of accuracy, such as the mean square forecast error (MSFE) or mean absolute forecast error (MAFE), can be characterized by very different error distributions. Third, expected loss functions are very sensitive to outliers. This is particularly true when the loss function is convex (e.g., quadratic or linear-exponential (Linex) loss functions). Corradi and Swanson (2013) address the second and third of these limitations by introducing an alternative criterion for predictive evaluation which measures accuracy via examination of the quantiles of expected loss distributions. However, their criterion is still loss function specific. In order to develop a loss function free criterion, one must evaluate the distribution of raw errors. Diebold and Shin (2015, 2017) make important contributions in this direction by introducing so-called stochastic error distance (SED) measures, in which Cramér-von Mises type statistics are constructed using raw errors. In their framework, in which they equate SED with mean absolute error loss, one chooses the model for which the cumulative error distribution is closest to a step function which is equal to zero over the negative real line and is equal to one over the positive real line. Jin, Corradi and Swanson (JSC: 2017) make further advances in this area by developing a forecast superiority testing methodology which is robust to the choice of loss function. They do this by establishing a mapping between generic loss function forecast evaluation and stochastic dominance principles. Hence, they test forecast superiority via stochastic dominance tests. However, the suggested tests are not uniformly valid and have correct size only under the least favorable case. In the current paper, we improve on JSC (2017) by developing loss function "free" forecast superiority tests which are uniformly asymptotically valid and asymptotically non conservative. This is done in part by noting that tests for stochastic dominance can be seen as tests for infinitely many moment inequalities, hence allowing us to utilize tools recently developed by Andrews and Shi (2013, 2017) to derive asymptotically uniformly valid and non conservative forecast superiority tests.

The implementation of such tests require that sample moments are standardized by an estimator of the standard deviation. Now, forecast errors are typically non martingale difference sequences, either because they are based on dynamically misspecified models or because forecasters do not efficiently use all the available information, in the case of subjective predictions. Hence, we require heteroskedasticity and autocorrelation (HAC) robust variance estimators. In our set-up, each such variance estimator depends on a specific point in the forecasting error support. Our first methodological contribution is to establish the consistency of HAC variance estimators uniformly over the error support. We also establish uniform convergence of their bootstrap counterparts. This is important because of the presence of the lag truncation parameter, in which case uniform convergence of HAC estimators and of their bootstrap analogs does not follow straightforwardly from uniform convergence of (kernel) nonparametric estimators. To the best of our knowledge this contribution is a novel addition to the vast literature on HAC covariance matrix estimation. When forecasts are based on estimated models, one also has to take into account the contribution of parameter estimation error to the limiting distribution. As a second methodological contribution, we develop uniformly valid and asymptotically non conservative forecast superiority tests for the case of non vanishing parameter estimation error, under recursive estimation scheme. This is

accomplished by extending the recursive block bootstrap introduced in Corradi and Swanson (2007).

The forecast superiority testing methodology discussed in this paper is evaluated via a series of Monte Carlo experiments, and via a detailed empirical analysis of the Survey of Professional Forecasters (SPF). Our Monte Carlo experiments indicate clear improvement in the finite sample power associated with using the tests that we introduce, when compared with those introduced in JCS(2017). In particular, the new tests have much higher power against alternatives in which some model beats the benchmark but other are strictly dominated, even in sample of only 250 observations. Our empirical analysis builds on a very large literature studying the SPF, in which many papers find (loss function specific) evidence of the usefulness of forecast combination. For example, Zarnowitz and Braun (1992) find that using the mean or median provides a consensus forecast with lower average errors than most individual forecasts. More recently, Aiolfi, Capistrán, and Timmermann (2011) and Genre, Kenny, Meyler, and Timmermann (2013) find that equal weighted averages of SPF and ECB SPF forecasts outperform model based forecasts, although in some cases there is an improvement by averaging them with (mean of) model-based forecasts. Using our new loss function free methodology, we are able to uncover distribution based evidence that forecast averages from small pools of survey participants ranked according to recent absolute and mean square forecast error performance are preferred to forecast averages based on the entire pool of experts. For example, for forecasting U.S. GDP growth, expert pools consisting of the "top 25%" of forecasters yield loss function free dominating combination predictions, when forecasters are required to have at least 1 year of experience. In our analysis, however, solely organizing expert pools based on experience is not enough. Instead, it is crucial that pools of experts also be chosen based on the quality of their predictions. Indeed, only requiring either 1, 3, or 5 years of experience does not yield a pool of experts whose forecast combinations are superior to equal weighted averages from the entire pool of experts, when using our loss function robust methods.

The rest of the paper is organized as follows. Section 2 outlines the set-up and introduces our new tests. Section 3 establishes the asymptotic properties of the tests in the context of generalized moment selection. Section 4 establishes the asymptotic properties of the tests in the context of non-vanishing parameter estimation error, for the recursive estimation schemes. Section 5 contains the results of our Monte Carlo experiments, and Section 6 contains the results of our analysis of GDP growth forecasts from the SPF. Finally, Section 7 provides a number of concluding remarks. Proofs are gathered in an appendix.

2 Set-Up

2.1 GL and CL Forecast Superiority Tests

Let $e_{j,t}$ be a forecast error, and let there be $j = 1, ..., k$ such errors at each point in time, $t = 1, ..., n$, corresponding to k different forecast models or judgmental forecasts, for example. We begin by ignoring estimation error, such as when forecasts are judgmental or subjective. Surveys including the SPF are leading examples of judgmental forecasts. The case where predictions are based on estimated models is considered in Section 4. Hereafter, sequence $e_{1,t}, t = 1, ..., n$ is called the "benchmark". In the context of the SPF, an example of a relevant benchmark against which to compare all other sequences is the consensus forecast constructed as the simple arithmetic average of individual forecasts in the survey. Our

goal is to test whether there exists some competing forecast that is superior to the benchmark for any loss function, L, satisfying Assumption A0.

Assumption A0 (i) $L \in \mathcal{L}_G$ if $L : \mathbb{R} \to \mathbb{R}^+$ 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 $z \geq 0$. (ii) $L \in \mathcal{L}_C$ is a convex function belonging to $\mathcal{L}_G$.

Hereafter, let $F_j(x)$ denote the cumulative distribution function (CDF) of forecast error e_j. JCS (2017) establish the following two results.

1. *For any* $L \in L_G$, $E(L(e_1)) \leq E(L(e_2))$, *if and only if* $(F_2(x) - F_1(x))sgn(x) \leq 0$, *for all* $x \in \mathcal{X}$.
2. *For any* $L \in L_C$, $E(L(e_1)) \leq E(L(e_2))$, *if and only if* $\left(\int_{-\infty}^{x}(F_1(t) - F_2(t))dt1(x < 0) + \int_{x}^{\infty}(F_2(t) - F_1(t))dt1(x \geq 0)\right) \leq 0$, *for all* $x \in \mathcal{X}$.

The first statement establishes a mapping between GL forecast superiority and first order stochastic dominance (FOSD). In particular, e_1 is not GL dominated by e_2 if $F_1(x)$ lies below $F_2(x)$ on the negative real line, and lies above $F_2(x)$ on the positive real line. Indeed, this ensure we choose the forecast whose CDF has larger mass around zero. Likewise, the second statement establishes a mapping between CL superiority and second order stochastic dominance.

In this framework, it follows that testing for loss function robust forecast superiority involves testing:

$$H_0 : \max_{j=2,\ldots,k} (E(L(e_1)) - E(L(e_k))) \leq 0 \tag{2.1}$$

versus

$$H_A : \max_{j=2,\ldots,k} (E(L(e_1)) - E(L(e_k))) > 0. \tag{2.2}$$

Consider $L \in \mathcal{L}_G$. Then, these hypotheses can be restated as follows:

$$\begin{aligned} H_0^G &= H_0^{G-} \cap H_0^{G+} \\ &: \left(\max_{j=2,\ldots,k} (F_1(x) - F_j(x))) \leq 0, \text{ for } x \leq 0\right) \\ &\quad \cap \left(\max_{j=2,\ldots,k} (F_j(x) - F_1(x))) \leq 0, \text{ for } x > 0\right) \end{aligned}$$

versus

$$\begin{aligned} H_A^G &= H_A^{G-} \cup H_A^{G+} \\ &: \left(\max_{j=2,\ldots,k} (F_1(x) - F_j(x))) > 0, \text{ for some } x \leq 0\right) \\ &\quad \cup \left(\max_{j=2,\ldots,k} (F_j(x) - F_1(x))) > 0, \text{ for some } x > 0\right). \end{aligned}$$

Note that the null hypothesis is the intersection of two different null hypotheses because of a discontinuity at zero. Similarly, for the case of $L \in \mathcal{L}_C$, H_0 and H_A can be restated as:

$$\begin{aligned} H_0^C &= H_0^{C-} \cap H_0^{C+} \\ &: \left(\max_{j=2,\ldots,k} \int_{-\infty}^{x} (F_1(t) - F_j(t))dt \leq 0, \text{ for } x \leq 0\right) \\ &\quad \cap \left(\max_{j=2,\ldots,k} \int_{x}^{\infty} (F_j(t) - F_1(t))dt \leq 0, \text{ for } x > 0\right) \end{aligned}$$

versus

$$\begin{aligned} H_A^C &= H_A^{C-} \cup H_A^{C+} \\ &: \left(\max_{j=2,...,k} \int_{-\infty}^{x} (F_1(t) - F_k(t))dt > 0, \text{ for some } x \leq 0 \right) \\ &\quad \cup \left(\max_{j=2,...,k} \int_{x}^{\infty} (F_j(t) - F_1(t))dt > 0, \text{ for some } x > 0 \right). \end{aligned}$$

In order to test $H_0^{G^+}$ against $H_A^{G^+}$, and H_0^{C+} against $H_A^{C^+}$JCS (2017) utilize the following statistics:

$$\sqrt{n} G_{j,n}^+(x) = \sqrt{n} \left(\widehat{F}_{j,n}(x) - \widehat{F}_{1,n}(x) \right), \tag{2.3}$$

where $\widehat{F}_{k,n}(x)$ denotes the empirical CDF of e_k; and

$$\begin{aligned} \sqrt{n} C_{j,n}^+(x) &= \sqrt{n} \int_{x}^{\infty} \left(\widehat{F}_{j,n}(t) - \widehat{F}_{1,n}(t) \right) dt \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left([(e_{1,t} - x)]_+ - [(e_{j,t} - x)]_+ \right), \end{aligned} \tag{2.4}$$

where $[y]_+ = \max\{0, y\}$, and where the equality in the preceding expression follows from integration by parts. Now, $G_{j,n}^+$ and $C_{j,n}^+$ are similar to the test statistics discussed in LMW (2005). Also, as stated in Theorem 3 of JCS (2017),

$$\begin{aligned} \max_{j=2,...,k} \sup_{x \in \mathcal{X}^+} \sqrt{n} G_{j,n}^+ &\Rightarrow \max_{j=2,...,k} \sup_{x \in \mathcal{B}_j^{g+}} g_j(x), \text{ if } \max_{j=2,...,k} F_j(x) - F_1(x)) = 0 \\ &\Rightarrow -\infty \qquad \text{if } \max_{j=2,...,k} F_j(x) - F_1(x)) < 0, \end{aligned} \tag{2.5}$$

where $g = (g_2, ..., g_k)$ denotes a $k-1$ dimensional zero mean Gaussian process, and $\mathcal{B}_j^{g+} = \{x \in \mathcal{X}^+ : F_1(x) = F_j(x)\}$. JCS (2017) also suggest to evaluate $H_0^{G^+}$ using the percentiles of the empirical distribution of: [1]

$$\max_{j=2,...,k} \sup_{x \in \mathcal{X}^+} \sqrt{n} \left(\left(\widehat{F}_{j,n}^*(x) - \widehat{F}_{j,n}(x) \right) - \left(\widehat{F}_{1,n}^*(x) - \widehat{F}_{1,n}(x) \right) \right).$$

Hence, their tests have asymptotically correct size only under the least favorable case under the null. Namely, when $F_1(x) = F_j(x)$, for all j, and for all $x \in \mathcal{X}^+$. In this sense, their test is not asymptotically similar on the boundary, $\mathcal{B}_0^+ = \{\max_{k=2,...,j} (F_j(x) - F_1(x)) = 0, \text{ for some } x \in \mathcal{X}^+\}$; and furthermore is asymptotically biased towards certain local alternatives.[2] On the other hand, as pointed out by LMW (2005), a test based on subsampled critical values is asymptotically similar on the boundary, $\mathcal{B}_0^+$. This is because the subsampling distribution mimics the sampling distribution. A drawback is that tests which are similar on the boundary may have very little power against certain sequence of alternatives, where $F_k(x) - F_1(x) > 0$, for $x \in \mathcal{X}_A \subset \mathcal{X}^+$; and where $F_k(x) - F_1(x) < 0$, for $x \in \mathcal{X}^+ \backslash \mathcal{X}_A$. Donald and Hsu

[1] i.e., H_0^G is (not) rejected at level α, if the smallest bootstrap p-values associated to H_0^{G-} and H_0^{G+} is smaller (larger) than $\alpha/2$.

[2] Hansen (2005) shows that p-values associated with the use of stationary bootstrap tests are actually upper bounds for an asymptotically unbiased test.

(2016) provide an interesting example of this issue for the case of $k = 2$. In addition, for the case of a finite number of moment inequalities, Andrews (2012) shows that tests which are similar on the boundary have trivial power against certain alternatives. Moreover, the weak convergence of the JCS (2017) statistic, and of its bootstrap counterpart, is "only" pointwise. This is also true under subsampling. Hence, inference based on subsampling or on centered bootstrap may be not asymptotically valid, uniformly, over all probabilities under the null hypothesis. Lack of uniformity is typical of tests based on weak inequalities (see Mikusheva (2007) and Andrews and Guggenberger (2010)). In light of this, and for the case of finitely many moment inequalities, Andrews and Soares (2010) and Andrews and Barwick (2013) introduce bootstrap tests which ensure that the asymptotic size (coverage) is at most (at least) α $(1-\alpha)$, uniformly, over all probabilities under the null hypothesis.

2.2 Improved Forecast Superiority Tests

It is immediate to see that we can restate H_0^G and H_0^C in terms of infinitely many moment inequalities. Hereafter, let $\mathcal{X} = \mathcal{X}^- \cup \mathcal{X}^+$ be the union of the support of $(e_1, ..., e_k)$.Then,

$$\begin{aligned} H_0^G &= H_0^{G-} \cap H_0^{G+} \\ &: \left(F_1(x) - F_j(x) \leq 0, \text{ for } j = 2, ..., k, \text{ and for all } x \in \mathcal{X}^-\right) \\ &\quad \cap \left(F_j(x) - F_1(x) \leq 0, \text{ for } j = 2, ..., k, \text{ and for all } x \in \mathcal{X}^+\right) \end{aligned}$$

versus

$$\begin{aligned} H_A^G &= H_A^{G-} \cup H_A^{G+} \\ &: \left(F_1(x) - F_j(x) > 0, \text{ for some } j = 2, ..., k, \text{ and for some } x \in \mathcal{X}^-\right) \\ &\quad \cup \left(F_j(x) - F_1(x) > 0, \text{ for some } j = 2, ..., k, \text{ and for some } x \in \mathcal{X}^+\right). \end{aligned}$$

Analogously,

$$\begin{aligned} H_0^C &= H_0^{C-} \cap H_0^{C+} \\ &: \left(\int_{-\infty}^{x} (F_1(t) - F_j(t))dt \leq 0, \text{ for } j = 2, ..., k, \text{ and for all } x \in \mathcal{X}^-\right) \\ &\quad \cap \left(\int_{x}^{\infty} (F_j(t) - F_1(t))dt \leq 0, \text{ for } j = 2, ..., k, \text{ and for all } x \in \mathcal{X}^+\right) \end{aligned}$$

versus

$$\begin{aligned} H_A^C &= H_A^{C-} \cup H_A^{C+} \\ &: \left(\int_{-\infty}^{x} (F_1(t) - F_k(t))dt > 0, \text{ for some } j = 2, ..., k, \text{ and for some } x \in \mathcal{X}^-\right) \\ &\quad \cup \left(\max_{j=2,...,k} \int_{x}^{\infty} (F_j(t) - F_1(t))dt > 0, \text{ for some } j = 2, ..., k, \text{ and for some } x \in \mathcal{X}^+\right). \end{aligned}$$

Evidently, H_0^G and H_0^C can be written as the intersection of $(k-1)$ moment inequalities, which have to hold uniformly over $\mathcal{X}$. This gives rise to an infinite number of moment conditions. Andrews and Shi (2013) develop tests for conditional moment inequalities, and as is well known in the literature on

consistent specification testing (e.g., see Bierens (1982, 1990)) a finite number of conditional moments can be transformed into an infinite number of unconditional moments. The same is true in the case of weak inequalities. Andrews and Shi (2017) consider tests for conditional stochastic dominance, which are then characterized by an infinite number of conditional moment inequalities and so by a "twice" infinite number of unconditional inequalities. Recalling that our interest is on testing GL or CL forecast superiority as in (2.1) and (2.2), we confine our attention to unconditional testing of stochastic dominance.

Because of the discontinuity around zero in the tests discussed in JCS (2017), they separately test $H_0^{G+}\left(H_0^{C+}\right)$ and $H_0^{G-}(H_0^{C-})$, and then use the methods of Holm (1979) to control the two resulting p-values (see Rules TG and TC in JCS (2017)). The same approach is taken in this paper. Hence, without loss of generality, we focus our discussion in the sequel on testing H_0^{G+} versus H_A^{G+} and H_0^{C+} versus H_A^{C+}. Testing H_0^{G-} versus H_A^{G-} and H_0^{C-} versus H_A^{C-} follows immediately.

We begin by testing GL forecast superiority. Let $G_n^+ = \left(G_{2,n}^+, ..., G_{k,n}^+\right)$,

$$\Sigma^{G+}(x,x') = \text{acov}\left(\sqrt{n}G_n^+(x), \sqrt{n}G_n^+(x')\right) \tag{2.6}$$

and

$$\overline{\Sigma}_n^{G+}(x,x') = \widehat{\Sigma}_n^{G+}(x,x') + \varepsilon I_{k-1}, \tag{2.7}$$

where $\varepsilon \geq 0$, and where $\widehat{\Sigma}_n^{G+}(x,x')$ is the sample analog of $\Sigma^{G+}(x,x')$. Let $\widehat{u}_{j,t}(x) = 1\{e_{j,t} \leq x\} - \frac{1}{n}\sum_{t=1}^n 1\{e_{j,t} \leq x\}$, so that the $jj-$th element of $\widehat{\Sigma}_n^{G+}(x,x)$ is given by

$$\begin{aligned}\widehat{\sigma}_{jj,n}^{2,G+}(x) &= \frac{1}{n}\sum_{t=1}^n (\widehat{u}_{j,t}(x) - \widehat{u}_{1,t}(x))^2 \\ &\quad + 2\frac{1}{n}\sum_{\tau=1}^{l_n}\sum_{t=\tau+1}^n w_\tau(\widehat{u}_{j,t}(x) - \widehat{u}_{1,t}(x))(\widehat{u}_{j,t-\tau}(x) - \widehat{u}_{1,t-\tau}(x)),\end{aligned} \tag{2.8}$$

where $w_\tau = 1 - \frac{\tau}{1+l_n}$, with $l_n \to \infty$ as $n \to \infty$. In (2.7), the role of the additional εI_{k-1} term is to correct for the possible singularity of the covariance estimator, for certain values of x. This is the case when we compare forecast errors from nested models. Also, let $\overline{\sigma}_{jj,n}^{2,G^+}(x,x')$ be the $jj-$th element of $\overline{\Sigma}_n^{G+}(x,x')$, and let $\overline{\sigma}_{jj,n}^{2,C+}(x,x')$ be the $jj-$th element of $\overline{\Sigma}_n^{C+}(x,x')$. Construct the following test statistics:

$$S_n^{G+} = \int_{\mathcal{X}^+}\sum_{j=2}^k \left(\max\left\{0, \frac{\sqrt{n}G_{j,n}^+(x)}{\overline{\sigma}_{jj,n}^{G+}(x)}\right\}\right)^2 \mathrm{d}Q(x). \tag{2.9}$$

and

$$S_n^{C+} = \int_{\mathcal{X}^+}\sum_{j=2}^k \left(\max\left\{0, \frac{\sqrt{n}C_{j,n}^+(x)}{\overline{\sigma}_{jj,n}^{C+}(x)}\right\}\right)^2 \mathrm{d}Q(x), \tag{2.10}$$

where Q is a weighting function defined below; $G_{j,n}^+(x)$ and $C_{j,n}^+(x)$ are the $j-$th components of $G_n^+(x)$ and $C_n^+(x)$, as defined in (2.3) and (2.4), respectively; and $\overline{\sigma}_{jj,n}^{C+}(x)$ is constructed in analogous manner to $\overline{\sigma}_{jj,n}^{G+}(x)$, by replacing $\widehat{u}_{j,t}(x)$ in (2.8) with $\widehat{\eta}_{j,t}(x) = [e_{j,t} - x]_+ - \frac{1}{n}\sum_{t=1}^n [e_{j,t} - x]_+$.

S_n^{G+} and S_n^{C+} are "sum" functions, as in equation (3.8) in Andrews and Shi (2013), and satisfy their

Assumptions S1-S4, which are required to guarantee that convergence is uniform over the null DGPs.[3][4] If $k = 2$ and $\overline{\sigma}_{jj,n}(x) = 1$, for all j and x (i.e. no standardization), then S_n^{G+} is the statistic used in Linton, Song and Whang (2010) for testing FOSD.

Of note is that in our context, potential slackness causes a discontinuity in the pointwise asymptotic distribution of the statistic.[5] This is because the pointwise asymptotic distribution is discontinuous, unless all moment conditions hold with equality. On the other hand, the finite sample distribution is not necessarily discontinuous. Thus, in the presence of slackness, the pointwise limiting distribution is not a good approximation of the finite sample distribution, and critical values based on pointwise asymptotics may be invalid. This is why we construct tests that are uniformly asymptotically valid (i.e., this is why we study the limiting distribution of our tests under drifting sequences of probability measures belonging to the null hypothesis). Moreover, in the infinite dimensional case, there is an additional source of discontinuity. In particular, the number of moment inequalities which contributes to the statistic varies across the different values of x. For example, the key difference between the case of $k = 2$ and $k > 2$ is that in the former case, for each value of x there is only one moment inequality which can be binding (or not). On the other hand, if $k = 3$, say, then for each value of x there can be either one or two moment inequalities which may be binding (or not), and whether or not a particular inequality is binding (or not) varies over x. Under this setup, we require the following assumptions in order to analyze the asymptotic behavior of our test statistics.

Assumption A1: For $j = 1, ..., k$, $e_{t,j}$ is strictly stationary and $\beta-$mixing, with mixing coefficients, $a_m = m^{-\beta}$, where $\beta > \frac{6\delta}{1-2\delta}$, $0 < \delta < 1/2$ and $\beta\delta > 1$.

Assumption A2: The union of the supports of $e_1, .., e_k$ is the compact set, $\mathcal{X} = \mathcal{X}^- \cup \mathcal{X}^+$.

Assumption A3: $F_j(x)$ has a continuos bounded density.

Assumption A4: The weighting function Q has full support $\mathcal{X}^+$.

3 Asymptotic Properties

3.1 Uniform Convergence of the HAC Estimator

We now turn to a discussion of the estimation of the variance in our forecast superiority test statistics. If $e_{1,t}, ..., e_{k,t}$ were martingale difference sequences, then we can still use the sample second moment as a variance estimator, and uniform consistency will follow by application of an appropriate uniform law of large numbers. In our set-up we can assume that $e_1, ..., e_k$ are martingale difference sequences if either: (i) they are judgmental forecasts from professional forecasters, say, who efficiently use all available information at time t (a strong assumption, which is tested in the forecast rationality literature); or (ii) they are prediction errors from one-step ahead forecasts based on dynamically correctly specified models. With respect to (i), it is worth noting that professional forecasters may be rational, ex-post, according to some loss function (see Elliott, Komunjer and Timmermann (2005,2008), although it is not as likely

[3] Note that we could have constructed a different "sum" function, using e.g. the statistic in (3.9) in Andrews and Shi (2013).

[4] Recall that one main drawback of the $\max_{j=2,...,k} \sup_{x \in \mathcal{X}^+} \sqrt{n} G_n^+$ statistic in JCS (2017) is that it diverges to $-\infty$ under some sequence of probability measures under the null, thus ruling out uniformity.

[5] By pointwise asymptotic distribution we mean the limiting distribution under a fixed probability measure.

that they are rational according to a generalized loss function. With respect to (ii), it should be noted that at most one model can be dynamically correctly specified for a given information set, and thus e_j cannot be a martingale difference sequence, for all $j = 1, ..., k$. In light of these facts, we allow for time dependence in the forecast error sequences used in our statistics, and use a HAC variance estimator in (2.9) and (2.10). In order to ensure that the HAC estimators converge uniformly over $\mathcal{X}^+$, it suffices to establish the counterpart of Lemma A1 of Supplement A of Andrews and Shi (2013) to the case of mixing sequences. This is done below.

Lemma 1*: Let Assumptions A1-A3 hold. Then, if* $l_n \approx n^\delta$ $0 < \delta < \frac{1}{2}$, *with* δ *defined as in Assumption A1:*

(i)

$$\sup_{x \in \mathcal{X}^+} \left| \widehat{\sigma}^{2,G+}_{jj,n}(x) - \sigma^{2,G+}_{jj}(x) \right| = o_p(1),$$

with $\sigma^{2,G+}_{jj}(x) = avar\left(\sqrt{n} G^+_{j,n}(x)\right)$; *and*

(ii)

$$\sup_{x \in \mathcal{X}^+} \left| \widehat{\sigma}^{2,C+}_{jj,n}(x) - \sigma^{2,C+}_{jj}(x) \right| = o_p(1),$$

with $\sigma^{2C+}_{jj}(x) = avar\left(\sqrt{n} C^+_{j,n}(x)\right)$.

Lemma 1 establishes the uniform convergence over $\mathcal{X}^+$ of HAC estimators. It is the time series counterpart of Lemma A1 in Andrews and Shi (2013). Of note is that we require $\beta-$mixing. This differs from the stationary pointwise HAC variance estimator case studied by Andrews (1991), where $\alpha-$mixing is required, and where the mixing coefficients declines to zero slightly slower than in our Assumption 1. This is because there is a trade-off between the degree of dependence and the rate of growth of the lag truncation parameter in the HAC estimator. Indeed, in the uniform case, the covering number (e.g., see Andrews and Pollard (1994)) grows with both l_n and the degree of dependence, thus leading to a trade-off between the two. For example, in the case of exponential mixing series, δ can be arbitrarily close to 1/2.

In the proof of Lemma 1, we require $\mathcal{X}^+$ in (2.9) and (2.10) to be a compact set. However, for the case of generalized loss superiority, the union of the supports of $e_1, .., e_k$ can be unbounded. This is because S_n^{G+} is bounded, regardless of the boundedness of the support. On the other hand, $S_n^{C^+}$ is bounded only when the union of the support of the forecasting error is bounded.

For carrying out inference on our forecast superiority tests, we require a bootstrap analog of the HAC variance estimator, which can be constructed as follows. Using the block bootstrap, make b_n draws of length l_n from $e_{1,t}, ..., e_{k,t}$, in order to obtain $\left(e^*_{j,1}, ..., e^*_{j,n}\right) = \left(e_{j,I_{1+1}}, ..., e^*_{j,I_{1+l}}, ..., e^*_{j,I_b+l}\right)$, with $b_n l_n = n$, where the block size, l_n, is equal to the lag truncation parameter in the HAC estimator described above.[6] Now, let $u^*_{j,t}(x) = 1\left\{e^*_{j,t} \leq x\right\} - 1\left\{e_{j,t} \leq x\right\}$, and

$$\widehat{\sigma}^{2*G+}_{jj,n}(x) = \frac{1}{b_n} \sum_{k=1}^{b_n} \left(\frac{1}{l_n^{1/2}} \sum_{i=1}^{l_n} \left(u^*_{j,(k-1)l_n+i}(x) - u^*_{1,(k-1)l_n+i}(x) \right) \right)^2. \tag{3.1}$$

Define $\widehat{\sigma}^{*2C^+}_{jj,n}(x)$ analogously, replacing $u^*_{j,t}(x)$ with $\eta^*_{j,t}(x) = \left[e^*_{j,t} - x\right]_+ - \left[e_{j,t} - x\right]_+$. The following result holds.

[6] We thus use the same notation, l_n, for both the lag truncation parameter and the block length.

Lemma 2: *Let Assumptions A1-A3 hold. Then, if* $l_n \approx n^\delta$ $0 < \delta < \frac{1}{2}$, *with* δ *defined as in Assumption A1:*

(i)

$$\sup_{x \in \mathcal{X}^+} \left| \widehat{\sigma}^{*G+}_{jj,n}(x) - \mathrm{E}^* \left(\widehat{\sigma}^{*G+}_{jj,n}(x) \right) \right| = o^*_p(1),$$

and (ii)

$$\sup_{x \in \mathcal{X}^+} \left| \widehat{\sigma}^{*C+}_{jj,n}(x) - \mathrm{E}^* \left(\widehat{\sigma}^{*C+}_{jj,n}(x) \right) \right| = o^*_p(1),$$

where $o^*_p(1)$ *denotes convergence to zero according to the bootstrap law,* P^*, *conditional on the sample.*

3.2 Inference Using the Bootstrap and Bounding Limiting Distributions

The statistics S^{G+}_n and S^{C+}_n are highly discontinuous over x. Exactly which moment conditions, and how many of them are binding varies over x. Hence, S^{G+}_n and S^{C+}_n do not necessarily have a well defined limiting distribution; and the continuous mapping theorem cannot be applied. However, following the generalized moment selection (GMS) test approach of Andrews and Shi (2013) we can establish lower and upper bound limiting distributions. Let

$$D^{G+}(x) = \mathrm{diag}\Sigma^{G+}_n(x,x),$$

$$h^{G+}_{A,n}(x) = D^{G+}(x)^{-1/2} \left(\sqrt{n} G^+_2(x), ..., \sqrt{n} G^+_k(x) \right)', \tag{3.2}$$

$$h^{G+}_B(x,x') = D^{G+}(x)^{-1/2} \left(\Sigma^{G+}_n + \varepsilon I_{k-1} \right)(x,x') D^{G+}(x')^{-1/2}, \tag{3.3}$$

and

$$v^{G+}(.) = (v^{G+}_2(.), ..., v^{G+}_k(.))', \tag{3.4}$$

where $v^{G+}(.)$ is a $(k-1)-$dimensional zero mean Gaussian process with correlation $h^{G+}_B(x,x')$. Also, let $D^{C+}(x), h^{C+}_{A,n}(x), h^{C+}_B(x,x'), v^{C+}(.)$ be defined analogously, by replacing $\Sigma^{G+}(x,x), G^+_2(x), ..., G^+_k(x)$ with $\Sigma^{C+}(x,x), C^+_2(x), ..., C^+_k(x)$. Finally, define

$$S^{\dagger G+}_n = \int_{\mathcal{X}^+} \sum_{j=2}^{k} \left(\max \left\{ 0, \frac{v^{G+}_j(x) + h^{G+}_{j,A,n}(x)}{\sqrt{h^{G+}_{jj,B}(x)}} \right\} \right)^2 \mathrm{d}Q(x) \tag{3.5}$$

where $h^{G+}_{jj,B}(x)$ is the $jj-$th element of $h^{G+}_B(x,x)$, and let

$$S^{\dagger G+}_\infty = \int_{\mathcal{X}^+} \sum_{j=2}^{k} \left(\max \left\{ 0, \frac{v^{G+}_j(x) + h^{G+}_{j,A,\infty}(x)}{\sqrt{h^{G+}_{jj,B}(x)}} \right\} \right)^2 \mathrm{d}Q(x), \tag{3.6}$$

where $h^{G+}_{j,A,\infty}(x) = 0$, if $G_j(x) = 0$, and $h^{G+}_{j,A,\infty}(x) = -\infty$, if $G_j(x) < 0$. Also, define $S^{\dagger C+}_n$ and $S^{\dagger C+}_\infty$ analogously, by replacing $v^{G+}_j(x), h^{G+}_{j,A,n}(x), h^{G+}_{j,A,\infty}(x)$, and $h^{G+}_{jj,B}(x)$ with $v^{C+}_j(x), h^{C+}_{j,A,n}(x), h^{C+}_{j,A,\infty}(x)$, and $h^{C+}_{jj,B}(x)$. Hereafter let

$$\mathcal{P}^{G+}_0 = \left\{ P : \ H^{G+}_0 \text{ holds} \right\}$$

so that $\mathcal{P}_0^{G+}$ is the collection of DGPs under which the null hypothesis holds. Needless to say, if Assumption A0 also hold, then $H_0^G = H_0^{G+} \cap H_0^{G-}$ is equivalent to H_0, as defined in (2.1). Let $\mathcal{P}_0^{C+}$ be defined analogously, with H_0^{G+} replaced by H_0^{C+}. The following result holds.

Theorem 1: *Let Assumptions A1-A4 hold. Then:*

(i) under H_0^{G+}, *there exists an* $\delta > 0$ *such that*

$$\lim \sup_{n\to\infty} \sup_{P\in\mathcal{P}_0^{G+}} \left[P\left(S_n^{G+} > a_{h_{A,n}}^{G+}\right) - P\left(S_n^{\dagger G+} + \delta > a_{h_{A,n}}^{G+}\right)\right] \leq 0$$

and

$$\lim \inf_{n\to\infty} \inf_{P\in\mathcal{P}_0^{G+}} \left[P\left(S_n^{G+} > a_{h_{A,n}}^{G+}\right) - P\left(S_n^{\dagger G+} - \delta > a_{h_{A,n}}^{G+}\right)\right] \geq 0;$$

and (ii) under H_0^{C+}, *there exists an* $\delta > 0$ *such that*

$$\lim \sup_{n\to\infty} \sup_{P\in\mathcal{P}_0^{C+}} \left[P\left(S_n^{C+} > a_{h_{A,n}}^{C+}\right) - P\left(S_n^{\dagger C+} + \delta > a_{h_{A,n}}^{C+}\right)\right] \leq 0$$

and

$$\lim \inf_{n\to\infty} \inf_{P\in\mathcal{P}_0^{C+}} \left[P\left(S_n^{C+} > a_{h_{A,n}}^{C+}\right) - P\left(S_n^{\dagger C+} - \delta > a_{h_{A,n}}^{C+}\right)\right] \geq 0.$$

Theorem 1 provides upper and lower bounds for $P\left(S_n^{G+} > a_{h_{A,n}}^{G+}\right)$ and $P\left(S_n^{C+} > a_{h_{A,n}}^{C+}\right)$, uniformly, over the probabilities under H_0^{G+} and H_0^{C+}, respectively. Note that $h_{j,A,n}^{G+}(.)$ and $h_{j,A,n}^{C+}(.)$ depend on the degree of slackness, and do not need to converge. Indeed, S_n^{G+} and/or S_n^{C+} do not have to converge in distribution for this result to hold.

Following Andrews and Shi (2013), we can construct bootstrap critical values which properly mimic the critical values of $S_\infty^{\dagger G+}$ and $S_\infty^{\dagger C+}$. We rely on the block bootstrap to capture the dependence in the data when constructing our bootstrap statistics. Consider the case of $S_\infty^{\dagger G+}$. Let $\left(e_{j,1}^*, ..., e_{j,n}^*\right), b_n$, and l_n be defined as in the previous subsection, and let:

$$G_{j,n}^{*+}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1\left\{e_{j,i}^* \leq x\right\} - 1\left\{e_{1,i}^* \leq x\right\}\right) \tag{3.7}$$

and

$$v_n^{*G+}(x) = \sqrt{n}\widehat{D}_n^{-1/2,G+}(x)\left(G_n^{*+}(x) - G_n^{+}(x)\right) \tag{3.8}$$

with $v_n^{*G+}(x) = \left(v_{2,n}^{*G+}(x), ..., v_{k,n}^{*G+}(x)\right)$ and $\widehat{D}_n^{G+}(x) = \text{diag}\widehat{\Sigma}_n^{G+}(x,x)$. Then, define:

$$\xi_{j,n}^{G+}(x) = \kappa_n^{-1} n^{1/2} \overline{D}_{jj,n}^{-1/2,G+}(x) G_{j,n}^{+}(x), \tag{3.9}$$

with $\kappa_n \to \infty$, as $n \to \infty$. Here, $\overline{D}_{jj,n}^{G+}(x)$ is the jj-th element of $\overline{D}_n^{G+}(x) = \text{diag}\left(\overline{\Sigma}_n^{G+}(x,x)\right)$, $\xi_n^{G+}(x) = \left(\xi_{2,n}^{G+}(x), ..., \xi_{k,n}^{G+}(x)\right)$, and

$$\phi_{j,n}^{G+}(x) = c_n 1\left\{\xi_{j,n}^{G+}(x) < -1\right\}, \tag{3.10}$$

with c_n a positive sequence, which is bounded above from zero. Thus, $\phi_{j,n}^{G+}(x) = c_n$, when $G_{j,n}^{+}(x) < -\kappa_n n^{-1/2}\overline{D}_{jj,n}^{-1/2,G+}(x)$ (i.e., when the $j-$th inequality is slack at x), and is zero otherwise.

It it clear from the selection rule in (3.10), that we do need an estimator of the variance of the moment conditions, despite the fact we use bootstrap critical values. In fact, standardization does not play a crucial role in the statistics, as all positive sample moment conditions matter. On the other hand, without the scaling factor in (3.9), the number of non-slack moment conditions would depend on the scale, and hence our bootstrap critical values would no longer be scale invariant. Let

$$S_n^{*G+} = \int_{\mathcal{X}^+} \sum_{j=2}^{k} \max\left(\left\{ 0, \frac{v_{j,n}^{*G+}(x) - \phi_{j,n}^{G+}(x)}{\sqrt{\overline{h}_{B,jj}^{*G+}(x)}} \right\} \right)^2 \mathrm{d}Q(x), \tag{3.11}$$

where $\overline{h}_{B,jj}^{*G+}(x)$ is the jj element of $\widehat{D}_n^{-1/2,G+}(x)\overline{\Sigma}_n^{*G+}(x,x)\,\widehat{D}_n^{-1/2,G+}(x)$ and $\overline{\Sigma}_n^{*G+}(x,x)$ is the bootstrap analog of $\overline{\Sigma}_n^{G+}(x,x)$.[7] Note that if c_n grows with n, then all slack inequalities are discarded, asymptotically. It is immediate to see that S_n^{*G+} is the bootstrap counterpart of $S_n^{\dagger G+}$ in (3.5), with $\phi_{j,n}^{G+}(x)$ mimicking the contribution of the slackness of inequality j (i.e., of $j-$th element of $h_{A,n}^{G+}(x)$). However, $\phi_{j,n}^{G+}(x)$ is not a consistent estimator of $h_{A,n}^{G+}(x)$, since the latter cannot be consistently estimated.

Now, consider the case of $S_\infty^{\dagger C+}$. Let:

$$C_{j,n}^{*+}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\left[e_{j,t}^* - x \right]_+ - \left[e_{1,t}^* - x \right]_+ \right),$$

and define $v_n^{*C+}(x), \widehat{D}_n^{C+}(x), \xi_{j,n}^{C+}(x)$, and $\phi_{j,n}^{C+}(x)$ analogously to $v_n^{*G+}(x), \widehat{D}_n^{G+}(x), \xi_{j,n}^{G+}(x)$, and $\phi_{j,n}^{G+}(x)$, by replacing $G_n^{*+}(x), G_n^{+}(x)$ and $\widehat{\Sigma}_n^{G+}(x,x)$ with $C_n^{*+}(x), C_n^{+}(x)$ and $\widehat{\Sigma}_n^{C+}(x,x)$. Then, construct:

$$S_n^{*C+} = \int_{\mathcal{X}^+} \sum_{j=2}^{k} \max\left(\left\{ 0, \frac{v_{j,n}^{*C+}(x) - \phi_{j,n}^{C+}(x)}{\sqrt{\overline{h}_{B,jj}^{*C+}(x)}} \right\} \right)^2 \mathrm{d}Q(x). \tag{3.12}$$

By comparing (2.9) and (2.10) with (3.11) and (3.12), it is immediate to see that $G_{j,n}^{+}(x)$ does not contribute to the test statistic when $G_{j,n}^{+}(x) < 0$, while it does not contribute to the bootstrap statistic when $G_{j,n}^{+}(x) < -\kappa_n n^{-1/2} \overline{D}_{jj,n}^{-1/2,G+}(x)$, with $\kappa_n n^{-1/2} \to 0$. Heuristically, by letting κ_n grow with the sample size, we control the rejection rates in a uniform manner.

It remains to define the GMS bootstrap critical values. Let $c_{n,B,1-\alpha}^{*G+}\left(\phi_n^{G+}, \overline{h}_{B,n}^{*G+}\right)$ be the $(1-\alpha)$-th critical value of S_n^{*G+}, based on B bootstrap replications, with ϕ_n^{G+} defined as in (3.10) and $\overline{h}_{B,n}^{*G+}(x) = \widehat{D}_n^{-1/2,G+}(x)\overline{\Sigma}_n^{*G+}(x,x)\,\widehat{D}_n^{-1/2,G+}(x)$. The $(1-\alpha)$-th GMS bootstrap critical value, $c_{0,n,1-\alpha}^{*G+}\left(\phi_n^{G+}, \overline{h}_{B,n}^{*G+}\right)$, is defined as:

$$c_{0,n,1-\alpha}^{*G+}\left(\phi_n^{G+}, \overline{h}_{B,n}^{*G+}\right) = \lim_{B\to\infty} c_{n,B,1-\alpha+\eta}^{*G+}\left(\phi_n^{G+}, \overline{h}_{B,n}^{*G+}\right) + \eta,$$

for $\eta > 0$, arbitrarily small. Further, $c_{n,B,1-\alpha}^{*C+}\left(\phi_n^{C+}, \overline{h}_{B,n}^{*C+}\right)$ and $c_{0,n,1-\alpha}^{*C+}\left(\phi_n^{C+}, \overline{h}_{B,n}^{C+}\right)$ are defined analogously.

Here, the constant η is used to guarantee uniformity over the infinite dimensional nuisance parameters, $h_{A,n}^{G+}(.), h_{A,n}^{C+}(.)$, uniformly on $x \in \mathcal{X}^+$, and is termed the infinitesimal uniformity factor by Andrews and

[7] Thus, the diagonal elements of $\widehat{\Sigma}_n^{*G+}(x,x)$ are the $\widehat{\sigma}_{jj,n}^{2*G+}(x)$ described in the previous subsection, while the off-diagonal elements of $\widehat{\Sigma}_n^{*G+}(x,x)$ are defined accordingly, as $\widehat{\sigma}_{jj',n}^{2*G+}(x)$, with $j \neq j'$.

Shi (2013). Heuristically, if all moment conditions are slack, then both the statistic and its bootstrap counterpart are zero, and by having $\eta > 0$ though arbitrarily close to zero we control the asymptotic rejection rate.

Finally, let

$$\mathcal{B}^{G+} = \left\{ x \in \mathcal{X}^+ \text{ s.t. } h^{G+}_{A,j,\infty} = 0, \text{ for some } j = 2, ..., k \right\} \tag{3.13}$$

and

$$\mathcal{B}^{C+} = \left\{ x \in \mathcal{X}^+ \text{ s.t. } h^{C+}_{A,j,\infty} = 0, \text{ for some } j = 2, ..., k \right\}, \tag{3.14}$$

where $\mathcal{B}^{G+}$ and $\mathcal{B}^{C+}$ define the sets over which at least one moment condition holds with strict equality, and these sets represent the boundaries of H_0^{G+} and H_0^{C+}, respectively.

Although we require that the block length grows at the same rate as the lag truncation parameter, l_n, in Lemma 2 (i.e., we require that $l_n \approx n^\delta$ $0 < \delta < \frac{1}{2}$ with δ being the mixing coefficient in A1), for the asymptotic uniform validity of the bootstrap critical values, we require that the block length grows at a rate slower than $n^{1/3}$. This slower rate is required for the bootstrap empirical central limit theorem for mixing process to hold (see Peligrad (1998)). Needless to say, even in the construction of $\widehat{\sigma}^{2,G+}_{jj,n}(x)$, we should thus use $l_n = o(n^{1/3})$. The following result holds.

Theorem 2: *Let Assumptions A1-A4 hold, and let* $l_n \to \infty$ *and* $l_n n^{\frac{1}{3}-\varepsilon} \to 0$ *as* $n \to \infty$. *Under* H_0^{G+} :

(i) if as $n \to \infty$, $\kappa_n \to \infty$ *and* $c_n/\kappa_n \to 0$, *then*

$$\limsup_{n\to\infty} \sup_{P \in \mathcal{P}_0^{G+}} P\left(S_n^{G+} \geq c^{*G+}_{0,n,1-\alpha}\left(\phi_n^{G+}, \overline{h}^{*G+}_{B,n} \right) \right) \leq \alpha;$$

and

(ii) if as $n \to \infty$, $\kappa_n \to \infty$, $c_n \to \infty$, $\sqrt{n}/\kappa_n \to \infty$, *and* $Q\left(\mathcal{B}^{G+}\right) > 0$, *then*

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \sup_{P \in \mathcal{P}_0^{G+}} P\left(S_n^{G+} \geq c^{*G+}_{0,n,1-\alpha}\left(\phi_n^{G+}, \overline{h}^{*G+}_{B,n} \right) \right) = \alpha.$$

Also, under H_0^{C+} :

(iii) if as $n \to \infty$, $\kappa_n \to \infty$ *and* $c_n/\kappa_n \to 0$, *then*

$$\limsup_{n\to\infty} \sup_{P \in \mathcal{P}_0^{C+}} P\left(S_n^{C+} \geq c^{*C+}_{0,n,1-\alpha}\left(\phi_n^{C+}, \overline{h}^{*C+}_{B,n} \right) \right) \leq \alpha;$$

and (iv) if as $n \to \infty$, $\kappa_n \to \infty$, $c_n \to \infty$, $\sqrt{n}/\kappa_n \to \infty$, *and* $Q\left(\mathcal{B}^{C+}\right) > 0$, *then*

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \sup_{P \in \mathcal{P}_0^{C+}} P\left(S_n^{C+} \geq c^{*C+}_{0,n,1-\alpha}\left(\phi_n^{C+}, \overline{h}^{*C+}_{B,n} \right) \right) = \alpha.$$

Statements (i) and (iii) of Theorem 2 establish that inference based on GMS bootstrap critical values has uniform correct size. Statements (ii) and (iv) of the theorem establish that inference based on GMS bootstrap critical values is asymptotically non-conservative, whenever $Q\left(\mathcal{B}^{+}\right) > 0$ or $Q\left(\mathcal{B}^{C+}\right) > 0$ (i.e., whenever at least one moment condition holds with equality, over a set $x \in \mathcal{X}^+$ with non-zero $Q-$measure). Although the GMS based tests are not similar on the boundary, the degree of non similarity, which is

$$\begin{aligned}
&\lim_{\eta\to 0} \limsup_{n\to\infty} \sup_{P \in \mathcal{P}_0^{G+}} P\left(S_n^{G+} \geq c^{*G+}_{0,n,1-\alpha}\left(\phi_n^{G+}, \overline{h}^{*G+}_{B,n} \right) \right) \\
&- \lim_{\eta\to 0} \liminf_{n\to\infty} \inf_{P \in \mathcal{P}_0^{G+}} P\left(S_n^{G+} \geq c^{*G+}_{0,n,1-\alpha}\left(\phi_n^{G+}, \overline{h}^{*G+}_{B,n} \right) \right),
\end{aligned}$$

is much smaller than that associated with using the "usual" recentered bootstrap.

In the case of pairwise comparison (i.e., $k = 2$), Theorem 2(ii) of Linton, Song and Whang (2010) establishes similarity of stochastic dominance tests on a subset of the boundary.

3.3 Power against Fixed and Local Alternatives

As our statistics are weighted averages over $\mathcal{X}^+$, they have non-trivial power only if the null is violated over a subset of non zero $Q-$measure. This applies to both power against fixed alternative, as well as to power against $\sqrt{n}-$local alternatives. In particular, for power against fixed alternatives, we require the following assumption.

Assumption FA: (i) $Q(B_{FA}^{G+}) > 0$, where $B_{FA}^{G+} = \{x \in \mathcal{X}^+ : G_j(x) > 0 \text{ for some } j = 2, ..., k\}$.. (ii) $Q(B_{FA}^{C+}) > 0$ where $B_{FA}^{C+} = \{x \in \mathcal{X}^+ : C_j(x) > 0 \text{ for some } j = 2, ..., k\}$.

The following result holds.

Theorem 3: *Let Assumptions A1-A4 hold.*

(i) If Assumption FA(i) holds, then under H_A^{G+} :

$$\lim_{n\to\infty} P\left(S_n^{G+} \geq c_{0,n,1-\alpha}^{*G+}\left(\phi_n^{G+}, \overline{h}_{B,n}^{*G+}\right)\right) = 1.$$

(ii) If Assumption FA(ii) holds, then under H_A^{C+} :

$$\lim_{n\to\infty} P\left(S_n^{C+} \geq c_{0,n,1-\alpha}^{*C+}\left(\phi_n^{C+}, \overline{h}_{B,n}^{*C+}\right)\right) = 1.$$

It is immediate to see that we have unit power against fixed alternatives, provided that the null hypothesis is violated, for at least one $j = 2, ..., k$, over a subset of $\mathcal{X}^+$ of non-zero $Q-$measure. Now, if we instead used a Kolmogorov type statistic (i.e., replace the integral over $\mathcal{X}^+$ with the supremum over $\mathcal{X}^+$), then we would not need Assumption FA, and it would suffice to have violation for some x, with possibly zero $Q-$measure, or in general with zero Lebesgue measure.[8] However, as pointed out in Supplement B of Andrews and Shi (2013) the statements in parts (ii) and (iv) of Theorem 2 do not apply to Kolmogorov tests, and hence asymptotic non-conservativeness does not necessarily hold. This is because the proof of those statements use the bounded convergence theorem, which applies to integrals but not to suprema.

We now consider the following sequences of local alternatives:

$$H_{L,n}^{G+} : G_{Lj}^+(x) = G_j^+(x) + \frac{\delta_{1,j}(x)}{\sqrt{n}} + o\left(n^{-1/2}\right), \text{ for } j = 2, ..., k, \ x \in \mathcal{X}^+$$

and

$$H_{L,n}^{C+} : C_{Lj}^+(x) = C_j^+(x) + \frac{\delta_{2,j}(x)}{\sqrt{n}} + o\left(n^{-1/2}\right), \text{ for } j = 2, ..., k, \ x \in \mathcal{X}^+.$$

[8] The Kolmorogov versions of S_n^{G+} and S_n^{C+} are:

$$KS_n^{G+} = \max_{x\in\mathcal{X}^+} \sum_{j=2}^{k} \left(\max\left\{0, \frac{\sqrt{n}G_{j,n}^+(x)}{\overline{\sigma}_{jj,n}^{G+}(x)}\right\}\right)^2$$

$$KS_n^{C+} = \max_{x\in\mathcal{X}^+} \sum_{j=2}^{k} \left(\max\left\{0, \frac{\sqrt{n}C_{j,n}^+(x)}{\overline{\sigma}_{jj,n}^{G+}(x)}\right\}\right)^2$$

We have $\lim_{n\to\infty}\sqrt{n}D^{G+}(x)^{-1/2}G^{+}_{Lj}(x) \to h^{G+}_{j,A,\infty}(x)+\delta_{1,j}(x)$, and $\lim_{n\to\infty}\sqrt{n}D^{C+}(x)^{-1/2}C^{+}_{Lj}(x) \to h^{C+}_{j,A,\infty}(x)+\delta_{2,j}(x)$. Define,

$$S^{\dagger,G+}_{\infty,\delta_1,LG}=\int_{\mathcal{X}^+}\sum_{j=2}^{k}\left(\max\left\{0,\frac{v^{G+}_j(x)+h^{G+}_{j,A,\infty}(x)+\delta_{1,j}(x)}{\sqrt{h^{G+}_{jj,B}(x)}}\right\}\right)^2 \mathrm{d}Q(x)$$

and

$$S^{\dagger,C+}_{\infty,\delta_2,LC}=\int_{\mathcal{X}^+}\sum_{j=2}^{k}\left(\max\left\{0,\frac{v^{C+}_j(x)+h^{C+}_{j,A,\infty}(x)+\delta_{2,j}(x)}{\sqrt{h^{C+}_{jj,B}(x)}}\right\}\right)^2 \mathrm{d}Q(x)$$

We require the following assumption.

Assumption LA:(i) $Q(B^{G+}_{LA})>0$, where
$B^{G+}_{LA}=\left\{x:\sqrt{n}D^{G+}(x)^{-1/2}G^{+}_{j}(x)\to h^{G+}_{j,A,\infty}(x),\ 0<h^{G+}_{j,A,\infty}(x)<\infty,\ \text{for some } j=2,...,k\right\}$.
(ii) $Q(B^{C+}_{LA})>0$, where
$B^{C+}_{LA}=\left\{x:\sqrt{n}D^{C+}(x)^{-1/2}C^{+}_{j}(x)\to h^{C+}_{j,A,\infty}(x),\ 0<h^{C+}_{j,A,\infty}(x)<\infty,\ \text{for some } j=2,...,k\right\}$.

The following result holds.

Theorem 4: *Let Assumptions A1-A4 hold.*

(i) If Assumption AL(i) holds, then under $H^{G+}_{L,n}$:

$$\lim_{n\to\infty}P\left(S^{G+}_n\geq c^{*G+}_{0,n,1-\alpha}\left(\phi^{G+}_n,\overline{h}^{*G+}_{B,n}\right)\right)=P\left(S^{\dagger G+}_{\infty,\delta_1,LG}\geq c_{LG,1-\alpha}\left(h^{G+}_{A,\infty},h^{G+}_{B,\infty}\right)\right),$$

with $c_{LG,1-\alpha}\left(h^{G+}_{A,\infty},h^{G+}_{B,\infty}\right)$ *denoting the* $(1-\alpha)$*-th critical value of* $S^{\dagger G+}_{\infty,\delta_1,LG}$*, with* $0<h^{G+}_{j,A,\infty}(x)<\infty$*, for some* $j=2,...,k$.

(ii) If Assumption AL(ii) holds, then under $H^{C+}_{L,n}$:

$$\lim_{n\to\infty}P\left(S^{C+}_n\geq c^{*C+}_{0,n,1-\alpha}\left(\phi^{C+}_n,\overline{h}^{*C+}_{B,n}\right)\right)=P\left(S^{\dagger C+}_{\infty,\delta_2,LC}\geq c_{LC,1-\alpha}\left(h^{C+}_{A,\infty},h^{C+}_{B,\infty}\right)\right),$$

with $c_{LC,1-\alpha}\left(h^{C+}_{A,\infty},h^{C+}_{B,\infty}\right)$ *denoting the* $(1-\alpha)$*-th critical value of* $S^{\dagger C+}_{\infty,\delta_2,LC}$*, with* $0<h^{C+}_{j,A,\infty}(x)<\infty$*, for some* $j=2,...,k$.

Theorem 4 establishes that our tests have power against $\sqrt{n}-$alternatives, provided that the drifting sequence is bounded away from zero, over a subset of $\mathcal{X}^+$ of non-zero $Q-$measure. Note also that for given loss function, L, the sequence of local alternatives for the White reality check can be defined as:

$$H_{A,n}:\max_{j=2,...,k}\left(E(L(e_1))-E(L(e_j))\right)=\frac{\lambda}{\sqrt{n}}+o\left(n^{-1/2}\right),\ \lambda>0. \tag{3.15}$$

For sake of simplicity, suppose that $k=2$ (this is the well known Diebold-Mariano test framework). Here,

$$
\begin{aligned}
& 0 < \lambda = n^{1/2} E(L(e_1)) - E(L(e_k)) \\
= \;& n^{1/2} \int_{-\infty}^{\infty} L(x) \left(f_{1,n}(x) - f_{2,n}(x) \right) \mathrm{d}z \\
= \;& -n^{1/2} \int_{-\infty}^{0} L'(x) \left(F_{1,n}(x) - F_{2,n}(x) \right) \mathrm{d}x \\
& -n^{1/2} \int_{0}^{\infty} L'(x) \left(F_{1,n}(x) - F_{2,n}(x) \right) \mathrm{d}x \\
= \;& n^{1/2} \int_{-\infty}^{0} \left(h_{A,\infty}^{G-}(x) + \delta_1(x) \right) Q(x) \mathrm{d}x + n^{1/2} \int_{0}^{\infty} \left(h_{A,\infty}^{G+}(x) + \delta_1(x) \right) Q(x) \mathrm{d}x, \qquad (3.16)
\end{aligned}
$$

where $F_{j,n}(x) = F_j(x) + \frac{\delta_{1,j}(x)}{\sqrt{n}}$ as defined in JCS (2017) and above, and and $\delta_1 = \delta_{1,1} - \delta_{1,2}$. Hence, $H_{A,n}$ in (3.15) is equivalent to $H_{LA}^{G+} \cap H_{LA}^{G-}$, whenever $Q(x) = L'(x) sign(x)$.

Analogously, for any convex loss function, L, which satisfies Assumption A0, $H_{A,n}$ in (3.15) is equivalent to $H_{LA}^{C-} \cap H_{LA}^{C+-}$, whenever $Q(x) = L''(x) sign(x)$. In fact, it is easy to see that:

$$
\begin{aligned}
& 0 < \delta = n^{1/2} E(L(e_1)) - E(L(e_k)) \\
= \;& n^{1/2} \int_{-\infty}^{\infty} L(x) \left(f_{1,n}(x) - f_{2,n}(x) \right) \mathrm{d}x \\
= \;& -n^{1/2} \int_{-\infty}^{0} L'(x) \left(F_{1,n}(x) - F_{2,n}(x) \right) \mathrm{d}z - n^{1/2} \int_{0}^{\infty} L'(x) \left(F_{1,n}(x) - F_{2,n}(x) \right) \mathrm{d}z \\
= \;& -L'(x) n^{1/2} \int_{-\infty}^{x} \left(F_{1,n}(z) - F_{2,n}(z) \right) \mathrm{d}z \,|_{-\infty}^{0} + n^{1/2} \int_{-\infty}^{0} L''(x) \left(\int_{-\infty}^{x} \left(F_{1,n}(z) - F_{2,n}(z) \right) \mathrm{d}z \right) \mathrm{d}x \\
& + n^{1/2} L'(x) \int_{x}^{\infty} \left(F_{1,n}(z) - F_{2.n}(z) \right) \mathrm{d}z \,|_{0}^{\infty} - n^{1/2} \int_{0}^{\infty} L''(x) \left(\int_{x}^{\infty} \left(F_{1,n}(z) - F_{2,n}(z) \right) \mathrm{d}z \right) \mathrm{d}x \\
= \;& n^{1/2} \int_{-\infty}^{0} L''(x) \left(\int_{-\infty}^{x} \left(F_{1,n}(z) - F_{2,n}(z) \right) \mathrm{d}z \right) \mathrm{d}x - n^{1/2} \int_{0}^{\infty} L''(x) \left(\int_{x}^{\infty} \left(F_{1,n}(z) - F_{2,n}(z) \right) \mathrm{d}z \right) \mathrm{d}x \\
= \;& n^{1/2} \int_{-\infty}^{0} \left(\int_{-\infty}^{x} \left(h_{A,\infty}^{C-}(x) + \delta_2(x) \right) \mathrm{d}z \right) Q(x) \mathrm{d}x - n^{1/2} \int_{0}^{\infty} \left(\int_{x}^{\infty} \left(h_{A,\infty}^{C+}(x) + \delta_2(x) \right) \mathrm{d}z \right) Q(x) \mathrm{d}x.
\end{aligned}
$$

4 Forecast Superiority Tests in the Presence of Recursive Estimation Error

4.1 The Statistics

Thus far, we have considered the case of subjective or judgmental forecasts, in which the econometrician is provided with sequences of forecasts and forecast errors. Our analysis has thus far been model free. When the objective is the evaluation of forecasts generated by estimated models, parameter estimation error must be accounted for. Consider the following standard setup. Let $T = R + n$. We use the first R observations for estimation and the last n for out of sample predictive evaluation. For brevity, we outline

the case of recursive estimation. Namely, at each point in time, $t > R$, update model parameter estimates prior to the construction of each new forecast, using all the available information.[9]

Formalizing the idea of recursive estimation, for $j = 1, ..., k$, use the first R observations to compute $\widehat{\theta}_{j,R}$, and construct the first prediction error:

$$\widehat{e}_{j,R+1} = X_{R+1} - \phi_j \left(Z_{j,R}, \widehat{\theta}_{j,R} \right),$$

where $Z_{j,R}$ contains lags of X as well as other regressors. Then, use the first $R+1$ observations to construct

$$\widehat{e}_{j,R+2} = X_{R+2} - \phi_j \left(Z_{j,R+1}, \widehat{\theta}_{j,R+1} \right).$$

Proceed in the same manner until a sequence of n prediction errors has been constructed, defined as:

$$\widehat{e}_{j,t+1} = X_{t+1} - \phi_j \left(Z_{j,t}, \widehat{\theta}_{j,t} \right), \tag{4.1}$$

for $t = R, ..., R+n-1$, where $\widehat{\theta}_{j,t}$ is the estimator computed using observations up to time t. In the sequel, assume that $\widehat{\theta}_{j,t}$ is a recursive $m-$estimator, so that:

$$\widehat{\theta}_{j,t} = \arg \min_{\theta_j \in \Theta_j} \frac{1}{t} \sum_{i=2}^{t} m_j(X_i, Z_{j,i-1}, \theta_j), \quad R \leq t \leq n-1, \ j = 1, ..., k, \tag{4.2}$$

and

$$\theta_j^{\dagger} = \arg \min_{\theta_j \in \Theta_j} E(m_j(X_i, Z_{j,i-1}, \theta_j)).$$

For $x \geq 0$, define:

$$\widetilde{G}_{j,n}^{+}(x) = \frac{1}{n} \sum_{t=R}^{T-1} \left(1\left\{ \widehat{e}_{j,t+1} \leq x \right\} - 1\left\{ \widehat{e}_{1,t+1} \leq x \right\} \right) = \left(\widetilde{F}_{j,n}(x) - \widetilde{F}_{1,n}(x) \right) \tag{4.3}$$

and

$$\begin{aligned} \widetilde{C}_{j,n}^{+}(x) &= \int_x^{\infty} \left(\widetilde{F}_{j,n}(t) - \widetilde{F}_{1,n}(t) \right) dt \\ &= \frac{1}{n} \sum_{t=R}^{T-1} \left\{ \left[(\widehat{e}_{1,t+1} - x) \right]_+ - \left[(\widehat{e}_{j,t+1} - x) \right]_+ \right\}. \end{aligned} \tag{4.4}$$

As shown in the proof of Lemma 3(i) in the Appendix,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=R}^{T-1} \left(1\left\{ \widehat{e}_{j,t+1} \leq x \right\} - F_j(x) \right) \\ = \ & \frac{1}{\sqrt{n}} \sum_{t=R}^{T-1} \left(1\left\{ e_{j,t+1} \leq x \right\} - F_j(x) \right) + f_j(x) \frac{1}{\sqrt{n}} \sum_{t=R}^{T-1} PEE_{j,t} + o_p(1), \end{aligned}$$

[9] In the rolling estimation case, we use only the most recent R observations to re-estiamte the forecasdting model, for each $t > R$. The rolling case can be treated analogously, and it is omitted only for brevity.

where, for $t \geq R$,

$$\begin{aligned} & PEE_{j,t} \\ = \; & \mathrm{E}\left(\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\theta_j^\dagger\right)\right) \\ & \times\left(\mathrm{E}\left(\nabla_\theta^2 m_j(X_i,Z_{j,i-1},\theta_j^\dagger)\right)\right)^{-1}\frac{1}{t}\sum_{i=1}^{t}\nabla_\theta m_j(X_i,Z_{j,i-1},\theta_j^\dagger)+o_p(1). \end{aligned} \tag{4.5}$$

It is immediate to see that if $n = o(R)$, then $\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1} PEE_{j,t} = o_p(1)$, and thus does not contribute to the asymptotic covariance of the above statistics.

Define the following forecast superiority test statistics:

$$\widetilde{S}_n^{G+} = \int_{\mathcal{X}^+}\sum_{j=2}^{k}\left(\max\left\{0,\frac{\sqrt{n}\widetilde{G}_{j,n}^{+}(x)}{\widetilde{\sigma}_{jj,n}^{G+}(x)+\varepsilon}\right\}\right)^2 \mathrm{d}Q(x)$$

and

$$\widetilde{S}_n^{C+} = \int_{\mathcal{X}^+}\sum_{j=2}^{k}\left(\max\left\{0,\frac{\sqrt{n}\widetilde{C}_{j,n}^{+}(x)}{\widetilde{\sigma}_{jj,n}^{C+}(x)+\varepsilon}\right\}\right)^2 \mathrm{d}Q(x),$$

where $\widetilde{\sigma}_{jj,n}^{G+}(x)$ and $\widetilde{\sigma}_{jj,n}^{C+}(x)$ include terms accounting for the contribution of parameter estimation error to asymptotic covariance. Here, $\widetilde{\sigma}_{jj,n}^{2,G+}(x)$ is defined as:

$$\begin{aligned} \widetilde{\sigma}_{jj,n}^{2,G+}(x) \;=\; & \widehat{\sigma}_{jj,n}^{2,G+}(x) + 2\widehat{\Pi}\widehat{f}_{1,n,h}^2(x)\widehat{A}_1\widehat{\Sigma}_{11}\widehat{A}_1' + 2\widehat{\Pi}\widehat{f}_{j,n,h}^2(x)\widehat{A}_j\widehat{\Sigma}_{jj}\widehat{A}_j' \\ & -4\widehat{\Pi}\widehat{f}_{1,n,h}(x)\widehat{A}_1\widehat{\Sigma}_{1j}\widehat{A}_j'\widehat{f}_{j,n,h}(x) + 2\widehat{\Pi}\widehat{f}_{1,n,h}(x)\widehat{A}_1\widehat{\Sigma}_{u1}(x) - 2\widehat{\Pi}\widehat{f}_{j,n,h}(x)\widehat{A}_j\widehat{\Sigma}_{uj}(x), \end{aligned}$$

where $\widehat{\sigma}_{jj,n}^{2,G+}(x)$ is defined as in (2.8)[10], $\widehat{\Pi} = 1 - \frac{R}{n}\ln\left(1+\frac{n}{R}\right)$,

$$\widehat{f}_{j,n,h}(x) = \frac{1}{nh}\sum_{t=R+1}^{n} K\left(\frac{\widehat{e}_{j,t}-x}{h}\right),$$

$$\widehat{A}_j = \frac{1}{n}\sum_{t=R+1}^{T}\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\widehat{\theta}_{j,R}\right)'\left(\frac{1}{R}\sum_{t=1}^{R}\nabla_{\theta_j}^2 m_j(X_t,Z_{j,t-1},\widehat{\theta}_{j,R})\right)^{-1},$$

$$\begin{aligned} \widehat{\Sigma}_{jj} \;=\; & \frac{1}{n}\sum_{t=R+1}^{T}\nabla_{\theta_j}m_j(X_t,Z_{t,i-1},\widehat{\theta}_{j,R})\nabla_{\theta_j}m_j(X_t,Z_{t,i-1},\widehat{\theta}_{j,R})' \\ & +2\frac{1}{n}\sum_{\tau=1}^{l_n}\sum_{t=R+\tau+1}^{T} w_\tau\nabla_{\theta_j}m_j(X_t,Z_{t,i-1},\widehat{\theta}_{j,R})\nabla_{\theta_j}m_j(X_{t-\tau},Z_{t-\tau,i-1},\widehat{\theta}_{j,R})', \end{aligned}$$

[10] However, $e_{t,j}$ and $e_{t,1}$ are replaced by $\widehat{e}_{t,j}$ and $\widehat{e}_{t,1}$, and only the last n observations, from $R+1$ to T, are used in test statistic construction.

and

$$\begin{aligned}\widehat{\Sigma}_{uj}(x) &= \frac{1}{n}\sum_{t=R+1}^{T}\nabla_{\theta_j}m_j(X_t,Z_{t,i-1},\widehat{\theta}_{j,R})\left(\left(1\left\{\widehat{e}_{j,t}\le x\right\}-\frac{1}{n}\sum_{t=1}^{n}1\left\{\widehat{e}_{j,t}\le x\right\}\right)\right.\\ &\quad\left.-\left(1\left\{\widehat{e}_{1,t}\le x\right\}-\frac{1}{n}\sum_{t=1}^{n}1\left\{\widehat{e}_{1,t}\le x\right\}\right)\right)\\ &\quad+2\frac{1}{n}\sum_{\tau=1}^{l_n}\sum_{t=R+\tau+1}^{T}w_\tau\nabla_{\theta_j}m_j(X_t,Z_{t,i-1},\widehat{\theta}_{j,R})\left(\left(1\left\{\widehat{e}_{j,t-\tau}\le x\right\}-\frac{1}{n}\sum_{t=1}^{n}1\left\{\widehat{e}_{j,t-\tau}\le x\right\}\right)\right.\\ &\quad\left.-\left(1\left\{\widehat{e}_{1,t-\tau}\le x\right\}-\frac{1}{n}\sum_{t=1}^{n}1\left\{\widehat{e}_{1,t-\tau}\le x\right\}\right)\right),\end{aligned}$$

By noting that

$$\begin{aligned}&\widetilde{C}^{+}_{j,n}(x)\\ =\;&\frac{1}{n}\sum_{t=R}^{T-1}\left(\left[(e_{1,t+1}-x)\right]_+-\left[(e_{j,t+1}-x)\right]_+\right)\\ &+\frac{1}{n}\sum_{t=R}^{T-1}\left(\left(\widehat{e}_{1,t}-e_{1,t}\right)1\left\{e_{1,t}\ge x\right\}-\left(\widehat{e}_{j,t}-e_{j,t}\right)1\left\{e_{j,t}\ge x\right\}\right)\\ &+\frac{1}{n}\sum_{t=R}^{T-1}\left(\left(e_{1,t}-x\right)\left(1\left\{\widehat{e}_{1,t}\ge x\right\}-1\left\{\widehat{e}_{1,t}\ge x\right\}\right)-\left(e_{j,t}-x\right)\left(1\left\{\widehat{e}_{j,t}\ge x\right\}-1\left\{e_{j,t}\ge x\right\}\right)\right) \qquad (4.6)\\ &+\frac{1}{n}\sum_{t=R}^{T-1}\left(\left(\widehat{e}_{1,t}-e_{1,t}\right)\left(1\left\{\widehat{e}_{1,t}\ge x\right\}-1\left\{e_{1,t}\ge x\right\}\right)-\left(\widehat{e}_{j,t}-e_{j,t}\right)\left(1\left\{\widehat{e}_{j,t}\ge x\right\}-1\left\{e_{j,t}\ge x\right\}\right)\right)\end{aligned}$$

we see that $\widetilde{\sigma}^{2,C+}_{jj,n}(x)$ is defined as:

$$\begin{aligned}&\widetilde{\sigma}^{2,G+}_{jj,n}(x)\\ =\;&\widehat{\sigma}^{2,C+}_{jj,n}(x)+2\widehat{\Pi}\widehat{f}^{2}_{1,n,h}(x)\widetilde{A}_1(x)\widehat{\Sigma}_{11}\widetilde{A}_1'(x)+2\widehat{\Pi}\widehat{f}^{2}_{j,n,h}(x)\widetilde{A}_j(x)\widehat{\Sigma}_{jj}\widetilde{A}_j'(x)\\ &-4\widehat{\Pi}\widehat{f}_{1,n,h}(x)\widetilde{A}_1(x)\widehat{\Sigma}_{1j}\widetilde{A}_j'(x)\widehat{f}_{j,n,h}(x)+2\widehat{\Pi}\widehat{f}_{1,n,h}(x)\widetilde{A}_1(x)\widehat{\Sigma}_{u1}(x)-2\widehat{\Pi}\widehat{f}_{j,n,h}(x)\widetilde{A}_j(x)\widehat{\Sigma}_{uj}(x)\\ &+2\widehat{\Pi}\widetilde{B}_1(x)\widehat{\Sigma}_{11}\widetilde{B}_1'(x)+2\widehat{\Pi}\widetilde{B}_j'(x)\widehat{\Sigma}_{jj}\widetilde{B}_j'(x)-4\widehat{\Pi}\widetilde{B}_1(x)\widehat{\Sigma}_{1j}\widetilde{B}_j'(x)\\ &+2\widehat{\Pi}\widetilde{B}_1(x)\widehat{\Sigma}_{u1}(x)-2\widehat{\Pi}\widetilde{B}_j(x)'\widehat{\Sigma}_{uj}(x)\\ &+2\widehat{\Pi}\widehat{f}_{1,n,h}(x)\widetilde{A}_1(x)\widehat{\Sigma}_{11}\widetilde{B}_1'(x)+2\widehat{\Pi}\widehat{f}_{j,n,h}(x)\widetilde{A}_j\widehat{\Sigma}_{jj}\widetilde{B}_j'(x)-2\widehat{\Pi}\widetilde{B}_1(x)\widehat{\Sigma}_{1j}\widetilde{A}_j'(x)\widehat{f}_{j,n,h}(x)\\ &-2\widehat{\Pi}\widetilde{B}_j(x)\widehat{\Sigma}_{1j}\widetilde{A}_1'(x)\widehat{f}_{1,n,h}(x),\end{aligned}$$

where $\widehat{\sigma}^{2,C+}_{jj,n}(x)$ is defined as in the statement of Lemma 1, but computed using only the last n observations, with prediction errors replaced by estimated prediction errors. Also,

$$\widetilde{A}_j(x)=\frac{1}{n}\sum_{t=R+1}^{T}\left(\widehat{e}_{t+1,j}-x\right)\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\widehat{\theta}_{j,R}\right)'\left(\frac{1}{R}\sum_{t=1}^{R}\nabla^2_{\theta_j}m_j(X_t,Z_{j,t-1},\widehat{\theta}_{j,R})\right)^{-1}$$

and

$$\widetilde{B}_j(x)=\frac{1}{n}\sum_{t=R+1}^{T}1\left\{\widehat{e}_{t+1,j}>x\right\}\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\widehat{\theta}_{j,R}\right)'\left(\frac{1}{R}\sum_{t=1}^{R}\nabla^2_{\theta_j}m_j(X_t,Z_{j,t-1},\widehat{\theta}_{j,R})\right)^{-1}.$$

In order to formalize the case of asymptotically non-vanishing parameter estimation error, we require the following assumptions.

Assumption A5: ϕ_j is twice continuously differentiable on the interior of Θ_j and the elements of $\nabla_{\theta_j}\phi_j(Z_{j,i-1},\theta_i)$ and $\nabla^2_{\theta_j}\phi_j(Z_{j,i-1},\theta_i)$ are $p-$dominated on Θ_i, for $j=1,...,k$, with $p>4$.

Assumption A6: For $j=1,...,k$: (i) $\theta_j^\dagger$ is uniquely identified (i.e. $E(m_j(X_t,Z_{j,t-1},\theta_j))>E(m_j(X_t,Z_{j,t-1},\theta_j^\dagger))$, for any $\theta_j\neq\theta_j^\dagger$); (ii) m_j is twice continuously differentiable on the interior of Θ_j; (iii) the elements of $\nabla_{\theta_j}m_j$ and $\nabla^2_{\theta_j}m_j$ are $p-$dominated on Θ_j, with $p>4$; and (iii) $E\left(-\nabla^2_{\theta_j}m_j(\theta_j)\right)$ is positive definite, uniformly on Θ_j.[11]

Assumption A7: $T=R+n$, and as $T\to\infty$, $n/R\to\pi$, with $0\leq\pi<\infty$.

As explained earlier, it is crucial to have a consistent estimator of the variance of the moment conditions. Otherwise, bootstrap critical values are not scale invariant. Hence, we need to construct estimators which properly capture parameters estimation error, regardless the fact that we rely on bootstrap critical values. GMS tests in the presence of non-vanishing estimation error have been considered in Coroneo, Corradi and Santos-Monteiro (2017). We have the following result.

Lemma 3: *Let Assumptions A1-A3, and A5-A7 hold. If* $l_n\approx n^\delta$ $\delta<\frac{1}{2}$, *as defined in Assumption A1, then:*

(i) $\sup_{x\in\mathcal{X}^+}\left|\widetilde{\sigma}^{2,G+}_{jj,n}(x)-\omega^{2,G+}_{jj}(x)\right|=o_p(1)$, *with* $\omega^{2G+}_{jj}(x)=avar\left(\sqrt{n}\widetilde{G}^+_{j,n}(x)\right)$; *and*

(ii) $\sup_{x\in\mathcal{X}^+}\left|\widetilde{\sigma}^{2,C+}_{jj,n}(x)-\omega^{2,C+}_{jj}(x)\right|=o_p(1)$, *with* $\omega^{2C+}_{jj}(x)=avar\left(\sqrt{n}\widetilde{C}^+_{j,n}(x)\right)$.

Lemma 3 mirrors Lemma 1 for the case of non-vanishing estimation error. In order to provide the analog of Theorem 1 for the case of non vanishing estimation error, we need define the counterparts of $S_n^{\dagger G+}$ and $S_n^{\dagger C+}$ which take into account of parameter estimation error. Let $\overline{\Omega}^{G+}(x,x)=\Omega^{G+}(x,x)+\varepsilon I_{k-1}$, where $\Omega^{G+}(x,x)=[\omega^{2,G+}_{ij,n}(x)]$. Also,

$$\mathcal{D}^{G+}(x)=\mathrm{diag}\Omega^{G+}(x,x),$$

$$h^{G+}_{1,n}(x)=\mathcal{D}^{G+}(x)^{-1/2}\left(\sqrt{n}G^+_2(x),...,\sqrt{n}G^+_k(x)\right)',$$

$$h^{G+}_2(x,x')=\mathcal{D}^{G+}(x)^{-1/2}\overline{\Omega}^{G+}(x,x')\,\mathcal{D}^{G+}(x')^{-1/2},$$

and

$$v^{G+}(.)=(v^{G+}_2(.),...,v^{G+}_k(.))'.$$

Here, $v^{G+}(.)$ is a $(k-1)-$dimensional zero mean Gaussian process with correlation $h^{G+}_2(x,x')$. Also, let $\mathcal{D}^{C+}(x),h^{C+}_{1,n}(x),h^{C+}_2(x,x')$, and $v^{C+}(.)$ be defined analogously by replacing $\Omega^{G+}(x,x),G^+_2(x),...,G^+_k(x)$ with $\Omega^{C+}(x,x),C^+_2(x),...,C^+_k(x)$.

Finally, define

$$S_n^{\ddagger G+}=\int_{\mathcal{X}^+}\sum_{j=2}^{k}\left(\max\left\{0,\frac{v^{G+}_j(x)+h^{G+}_{j,1,n}(x)}{\sqrt{h^{G+}_{jj,2}(x)}}\right\}\right)^2\mathrm{d}Q(x),$$

[11] We say that $\nabla_{\theta_j}\ln f_j(y_t,Z^{t-1},\theta_j)$ is $2r-$dominated on Θ_j if its $v-th$ element, $v=1,...,\varrho(j)$, is such that $\left|\nabla_{\theta_j}\ln f_j(y_t,Z^{t-1},\theta_j)\right|_v\leq D_t$, and $E(|D_t|^{2r})<\infty$. For more details on domination conditions, see Gallant and White (1988, pp. 33).

where $h_{jj,2}^{G+}(x)$ is the $jj-$th element of $h_2^{G+}(x,x)$, and let

$$S_n^{\ddagger C+} = \int_{\mathcal{X}^+} \sum_{j=2}^{k} \left(\max \left\{ 0, \frac{v_j^{C+}(x) + h_{j,1,n}^{C+}(x)}{\sqrt{h_{jj,2}^{C+}(x)}} \right\} \right)^2 \mathrm{d}Q(x),$$

which is defined analogously, by replacing $v_j^{G+}(x), h_{j,1,n}^{G+}(x)$, and $h_{jj,2}^{G+}(x)$ with $v_j^{C+}(x), h_{j,1,n}^{C+}(x)$, and $h_{jj,2}^{C+}(x)$. The following result holds.

Theorem 5: *Let Assumptions A1-A7 hold.*

(i) Under H_0^{G+}, there exist $\delta > 0$ such that:

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0^{G+}} \left[P\left(\widetilde{S}_n^{G+} > a_{h_{A,n}}^{G+} \right) - P\left(S_n^{\ddagger G+} + \delta > a_{h_{A,n}}^{G+} \right) \right] \leq 0$$

and

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_0^{G+}} \left[P\left(\widetilde{S}_n^{G+} > a_{h_{A,n}}^{G+} \right) - P\left(S_n^{\ddagger G+} - \delta > a_{h_{A,n}}^{G+} \right) \right] \geq 0.$$

(ii) Under H_0^{C+}, there exist $\delta > 0$ such that:

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0^{C+}} \left[P\left(\widetilde{S}_n^{C+} > a_{h_{A,n}}^{C+} \right) - P\left(S_n^{\ddagger C+} + \delta > a_{h_{A,n}}^{C+} \right) \right] \leq 0$$

and

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_0^{C+}} \left[P\left(\widetilde{S}_n^{C+} > a_{h_{A,n}}^{C+} \right) - P\left(S_n^{\ddagger C+} - \delta > a_{h_{A,n}}^{C+} \right) \right] \geq 0.$$

Theorem 5 provides upper and lower bounds for $P\left(\widetilde{S}_n^{G+} > a_{h_{A,n}}^{G+} \right)$ and $P\left(\widetilde{S}_n^{C+} > a_{h_{A,n}}^{C+} \right)$, uniformly, over the probabilities under the null H_0^{G+} and H_0^{C+}, respectively. Note that $h_{j,A,n}^{G+}(.)$ and $h_{j,A,n}^{C+}(.)$ depend on the degree of slackness and do not need to converge. Indeed, $\widetilde{S}_n^{G+}$ and $\widetilde{S}_n^{C+}$ do not have to converge in distribution.

4.2 Bootstrap Estimators

When computing recursive $m-$estimators, it is important to note that earlier observations are used more frequently than temporally subsequent observations. On the other hand, in the standard block bootstrap, all blocks from the original sample have the same probability of being selected, regardless of the dates of the observations in the blocks. Thus, the bootstrap estimator, say $\widehat{\theta}_{j,t}^*$, which is constructed as a direct analog of $\widehat{\theta}_{j,t}$ in (4.2), is characterized by a location bias that can be either positive or negative, depending on the sample that we observe. In order to circumvent this problem, Corradi and Swanson (2007) suggest a re-centering of the bootstrap score which ensures that the new bootstrap estimator, which is no longer the direct analog of $\widehat{\theta}_{j,t}$, is asymptotically unbiased. It should be noted that the idea of re-centering is not new in the bootstrap literature for the case of full sample estimation. In the context of $m-$estimators using the full sample, re-centering is needed only for higher order asymptotics, but not for first order validity, in the sense that the bias term is of smaller order than $n^{-1/2}$ (see e.g. Andrews (2002)). In the case of recursive $m-$estimators, on the other hand, the bias term is of order $n^{-1/2}$, so

that it does contribute to the limiting distribution. In the sequel, we assume that the block length grows with the sample. Also, assume that $T = R + n = b_T l_T$, with $b_T = b_n \frac{T}{n}$ and $l_T = l_n \frac{T}{n}$, and define:

$$\widetilde{\theta}^*_{j,t} = \arg\min_{\theta_j \in \Theta_j} \frac{1}{t} \sum_{i=1}^{t} \left(m_j(X^*_i, Z^*_{j,i-1}, \theta_j) - \theta'_j \left(\frac{1}{T} \sum_{k=1}^{T-1} \nabla_{\theta_j} m_j(X_k, Z_{j,k-1}, \widehat{\theta}_{j,t}) \right) \right),$$

where $X^*_i, Z^*_{j,i-1}$ are resampled via the "standard" block bootstrap outlined in the previous section, but with block length b_T. Theorem 1 in Corradi and Swanson (2007) establish that $\frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(\widetilde{\theta}^*_{j,t} - \widehat{\theta}_{j,t} \right)$ has the same limiting distribution as $\frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(\widehat{\theta}_{j,t} - \theta^\dagger_j \right)$, conditional of the sample.

With a slight abuse of notation, let $u^*_{j,t}(x) = 1\{e^*_{j,t} \leq x\} - \frac{1}{T} \sum_{t=1}^{T} 1\{\widehat{e}_{j,t} \leq x\}$ and $\eta^*_{j,t}(x) = \left[e^*_{j,t} - x\right]_+ - \frac{1}{T} \sum_{t=1}^{n} \left[\widehat{e}_{j,t} - x\right]_+$, with $e^*_{j,t+1} = X^*_{t+1} - \phi_j \left(Z^*_{j,t}, \widehat{\theta}_{j,t} \right)$, and let $\widehat{u}^*_{j,t}(x) = 1\{\widehat{e}^*_{j,t} \leq x\} - \frac{1}{T} \sum_{t=1}^{T} 1\{\widehat{e}_{j,t} \leq x\}$ and $\widehat{\eta}^*_{j,t}(x) = \left[\widehat{e}^*_{j,t} - x\right]_+ - \frac{1}{T} \sum_{t=1}^{n} \left[\widehat{e}_{j,t} - x\right]_+$, with $\widehat{e}^*_{j,t+1} = X^*_{t+1} - \phi_j \left(Z^*_{j,t}, \widetilde{\theta}^*_{j,t} \right)$. Our first goal is to construct the bootstrap counterparts of $\widetilde{\sigma}^{2,G+}_{jj,n}(x)$ and $\widetilde{\sigma}^{2,C+}_{jj,n}(x)$, called $\widetilde{\sigma}^{*2,G+}_{jj,n}(x)$ and $\widetilde{\sigma}^{*2,C+}_{jj,n}(x)$. Define:

$$\begin{aligned}
&\widetilde{\sigma}^{*2,G+}_{jj,n}(x) \\
= \;& \widehat{\text{avar}^*} \left(\frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(\widehat{u}^*_{j,t}(x) - \widehat{u}^*_{1,t}(x) \right) \right) \\
= \;& \widehat{\text{avar}^*} \left(\frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(u^*_{j,t}(x) - u^*_{1,t}(x) \right) \right) + \widehat{\text{avar}^*} \left(\frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(\widehat{f}^*_{j,n,h}(x) \widehat{PEE^*}_{j,t} - \widehat{f}^*_{1,n,h}(x) \widehat{PEE^*}_{1,t} \right) \right) \\
& -2\widehat{\text{acov}^*} \left(\frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(u^*_{j,t}(x) - u^*_{1,t}(x) \right), \frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(\widehat{f}^*_{j,n,h}(x) \widehat{PEE^*}_{j,t} - \widehat{f}^*_{1,n,h}(x) \widehat{PEE^*}_{1,t} \right) \right),
\end{aligned}$$

and

$$\begin{aligned}
&\widetilde{\sigma}^{*2,C+}_{jj,n}(x) \\
= \;& \widehat{\text{avar}^*} \left(\frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(\widehat{\eta}^*_{j,t}(x) - \widehat{\eta}^*_{1,t}(x) \right) \right) \\
= \;& \widehat{\text{avar}^*} \left(\frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(\eta^*_{j,t}(x) - \eta^*_{1,t}(x) \right) \right) \\
& + \widehat{\text{avar}^*} \left(\frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(\left[\widehat{f}^*_{j,n,h} \widehat{PEE}^*_{j,t} - x \right]_+ - \left[\widehat{f}^*_{1,n,h} \widehat{PEE}^*_{1,t} - x \right]_+ \right) \right) + \\
& -2\widehat{\text{acov}^*} \left(\frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(\eta^*_{j,t}(x) - \eta^*_{1,t}(x) \right), \frac{1}{\sqrt{n}} \sum_{t=R}^{n-1} \left(\left[\widehat{f}^*_{j,n,h} \widehat{PEE}^*_{j,t} - x \right]_+ - \left[\widehat{f}^*_{1,n,h} \widehat{PEE}^*_{1,t} - x \right]_+ \right) \right),
\end{aligned}$$

where avar^* and acov^* denote asymptotic variances and covariances, with respect to the bootstrap probability measure, $\widehat{f}^*_{j,n,h}$ is an estimator of the density of e_j based on the resampled observations, and

$\widehat{PEE^*}_{j,t}$ is an estimator of:

$$\begin{aligned} PEE^*_{j,t} &= \mathrm{E}^*\left(\nabla_{\theta_j}\phi_j\left(Z^*_{j,t},\widetilde{\theta}^*_{j,t}\right)\right)\mathrm{E}^*\left(\nabla^2_\theta m_j\left(X^*_i,Z^*_{j,i-1},\widetilde{\theta}^*_{j,t}\right)\right) \\ &\quad \frac{1}{t}\sum_{i=1}^{t}\left(\nabla_\theta m_j\left(X^*_i,Z^*_{j,i-1},\widehat{\theta}_{j,t}\right)-\frac{1}{T}\sum_{i=1}^{T}\nabla_{\theta_j}m_j(X_k,Z_{j,k-1},\widehat{\theta}_{j,t}\right). \end{aligned} \tag{4.7}$$

Closed form expressions for $\widehat{PEE^*}_{j,t}$, $\widehat{\mathrm{avar}^*}$, and $\widehat{\mathrm{acov}^*}$ are given in the proof of Lemma 4. We have the following result.

Lemma 4: *Let Assumptions A1-A3 and A5-A7 hold. Then, if* $l_n \approx n^\delta$ $\delta < \frac{1}{2}$, *and* β *the mixing coefficient in Assumption A1 is such that* $\beta > \frac{6\delta}{1-2\delta}$:

(i) $\sup_{x\in\mathcal{X}^+}\left|\widetilde{\sigma}^{*2,G+}_{jj,n}(x)-\mathrm{E}^*\left(\widetilde{\sigma}^{*2,G+}_{jj,n}(x)\right)\right| = o_p(1)$ *and*

(ii) $\sup_{x\in\mathcal{X}^+}\left|\widetilde{\sigma}^{*2,C+}_{jj,n}(x)-\mathrm{E}^*\left(\widetilde{\sigma}^{*2,C+}_{jj,n}(x)\right)\right| = o_p(1)$.

4.3 Bootstrap Critical Values

The bootstrap statistics in the non-vanishing recursive parameter estimation error case are:

$$\widetilde{S}^{*G+}_n = \int_{\mathcal{X}^+}\sum_{j=2}^{k}\max\left(\left\{0,\frac{\widetilde{v}^{*G+}_{j,n}(x)-\widetilde{\phi}^{G+}_{j,n}(x)}{\sqrt{\widetilde{\overline{h}}^{*G+}_{2,jj}(x)}}\right\}\right)^2 \mathrm{d}Q(x), \tag{4.8}$$

where $\widetilde{\overline{h}}^{*G+}_{2,jj}(x)$ is the jj element of $\widetilde{D}^{-1/2,G+}_n(x)\overline{\widetilde{\Sigma}}^{*G+}_n(x,x)\,\widetilde{D}^{-1/2,G+}_n(x)$, with $\widetilde{D}^{G+}_n(x) = \mathrm{diag}\widetilde{\Sigma}^{*G+}_n(x,x)$, $\widetilde{\Sigma}^{*G+}_n(x,x) = [\widetilde{\sigma}^{*2,G+}_{ij,n}(x)]$ $i,j = 1,...,k$, and $\overline{\widetilde{\Sigma}}^{*G+}_n = \widetilde{\Sigma}^{*G+}_n + \varepsilon I_{k-1}$. Also,

$$\begin{aligned} \widetilde{v}^{*G+}_n(x) &= \sqrt{n}\widetilde{D}^{-1/2,G+}_n(x)\frac{1}{\sqrt{n}}\sum_{i=R+1}^{n}\left(\left(1\left\{\widehat{e}^*_{j,i}\le x\right\}-1\left\{\widehat{e}^*_{1,i}\le x\right\}\right)\right. \\ &\quad \left.\frac{1}{T}\sum_{t=1}^{T}\left(1\left\{\widehat{e}_{j,t}\le x\right\}-1\left\{\widehat{e}_{1,t}\le x\right\}\right)\right) \end{aligned}$$

and for $\widetilde{\xi}^{G+}_{j,n}(x) = \kappa_n^{-1}n^{1/2}\overline{\widetilde{D}}^{-1/2,G+}_{jj,n}(x)\widetilde{G}^{+}_{j,n}(x)$,

$$\widetilde{\phi}^{G+}_{j,n}(x) = c_n 1\left\{\widetilde{\xi}^{G+}_{j,n}(x) < -1\right\}. \tag{4.9}$$

Finally, also define

$$\widetilde{S}^{*C+}_n = \int_{\mathcal{X}^+}\sum_{j=2}^{k}\max\left(\left\{0,\frac{\widetilde{v}^{*C+}_{j,n}(x)-\widetilde{\phi}^{C+}_{j,n}(x)}{\sqrt{\widetilde{\overline{h}}^{*C+}_{2,jj}(x)}}\right\}\right)^2 \mathrm{d}Q(x),$$

where $\widetilde{v}^{*C+}_n(x), \widetilde{D}^{C^+}_n(x), \widetilde{\xi}^{C+}_{j,n}(x)$, and $\widetilde{\phi}^{C+}_{j,n}(x)$ are defined analogously to $\widetilde{v}^{*G+}_n(x), \widetilde{D}^{G+}_n(x), \widetilde{\xi}^{G+}_{j,n}(x)$, and $\widetilde{\phi}^{G+}_{j,n}(x)$. It is immediate to see that estimation error contributes to the bootstrap statistics not only as

a scaling factor, but also in determining which moment conditions are binding. This is why we need an estimator of the variance, even if inference is based on bootstrap critical values.

We now define the GMS bootstrap critical values for the case of non-vanishing recursive estimation error. Let $\widetilde{c}^{*G+}_{n,B,1-\alpha}\left(\widetilde{\phi}^{G+}_n, \overline{h}^{*G+}_{2,n}\right)$ be the $(1-\alpha)$-th critical value of $\widetilde{S}^{*G+}_n$, based on B bootstrap replications, with $\widetilde{\phi}^{G+}_n$ as in (4.9) and $\overline{\widetilde{h}}^{*G+}_{2,jj}(x)$ as in (4.8). The $(1-\alpha)$-th GMS bootstrap critical value, $\widetilde{c}^{*G+}_{0,n,1-\alpha}\left(\widetilde{\phi}^{G+}_n, \overline{h}^{*G+}_{2,n}\right)$ is defined as:

$$\widetilde{c}^{*G+}_{0,n,1-\alpha}\left(\widetilde{\phi}^{G+}_n, \overline{h}^{*G+}_{B,n}\right) = \lim_{B\to\infty} \widetilde{c}^{*G+}_{n,B,1-\alpha+\eta}\left(\widetilde{\phi}^{G+}_n, \overline{h}^{*G+}_{2,n}\right) + \eta,$$

for arbitrarily small $\eta > 0$. Also, $\widetilde{c}^{*C+}_{n,B,1-\alpha+\eta}\left(\widetilde{\phi}^{C+}_n, \overline{h}^{*C+}_{2,n}\right)$ and $\widetilde{c}^{*C+}_{0,n,1-\alpha}\left(\widetilde{\phi}^{C+}_n, \overline{h}^{*C+}_{B,n}\right)$ are defined analogously. The following result then holds.

Theorem 6: Let Assumptions A1-A7 hold, and let $l_n \to \infty$ and $l_n n^{\frac{1}{3}-\varepsilon} \to 0$, as $n \to \infty$. Under H_0^{G+} :
(i) if as $n \to \infty$, $\kappa_n \to \infty$ and $c_n/\kappa_n \to 0$, then

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0^{G+}} P\left(\widetilde{S}^{G+}_n \geq \widetilde{c}^{*C+}_{n,B,1-\alpha+\eta}\left(\widetilde{\phi}^{C+}_n, \overline{h}^{*C+}_{2,n}\right)\right) \leq \alpha;$$

and (ii) if as $n \to \infty$, $\kappa_n \to \infty$, $c_n \to \infty$, $\sqrt{n}/\kappa_n \to \infty$ and $Q\left(\mathcal{B}^{G+}\right) > 0$, $\mathcal{B}^{G+}$ as in (3.13), then

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0^{G+}} P\left(\widetilde{S}^{G+}_n \geq \widetilde{c}^{*C+}_{n,B,1-\alpha+\eta}\left(\widetilde{\phi}^{C+}_n, \overline{h}^{*C+}_{2,n}\right)\right) = \alpha.$$

Also, under H_0^{C+},
(iii) if as $n \to \infty$, $\kappa_n \to \infty$ and $c_n/\kappa_n \to 0$, then

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0^{C+}} P\left(\widetilde{S}^{C+}_n \geq \widetilde{c}^{*C+}_{0,n,1-\alpha}\left(\widetilde{\phi}^{C+}_n, \overline{h}^{*C+}_{B,n}\right)\right) \leq \alpha;$$

and (iv) if as $n \to \infty$, $\kappa_n \to \infty$, $c_n \to \infty$, $\sqrt{n}/\kappa_n \to \infty$ and $Q\left(\mathcal{B}^{C+}\right) > 0$, $\mathcal{B}^{C+}$ as in (3.14), then

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0^{C+}} P\left(\widetilde{S}^{C+}_n \geq \widetilde{c}^{*C+}_{0,n,1-\alpha}\left(\widetilde{\phi}^{C+}_n, \overline{h}^{*C+}_{B,n}\right)\right) = \alpha.$$

Statements (i) and (iii) of Theorem 6 establish that inference based on GMS bootstrap critical values has uniform correct size, in the parameter estimation error case. Statements (ii) and (iv) of the theorem establish that inference based on the GMS bootstrap critical values is asymptotically non-conservative, whenever $Q\left(\mathcal{B}^{+}\right) > 0$ or $Q\left(\mathcal{B}^{C+}\right) > 0$.

5 Monte Carlo Experiments

In this section, we evaluate the finite sample performance of GL and CL forecast superiority tests when there are multiple competing sequences of forecast errors, under stationarity. In addition to analyzing the performance of our tests based on S_n^{G+} and S_n^{G-}, (GL forecast superiority) as well as based on S_n^{C+} and S_n^{C-} (CL forecast superiority), we also analyze the performance of the related test statistics from

JCS (2017), here called JCS_n^{G+}, JCS_n^{G-}, JCS_n^{C+}, and JCS_n^{C-}. For the sake of brevity, these two classes of tests are called S_n and JCS_n type tests, respectively. When computing the suprema used in all of these statistics, 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.[12] For each experiment we carry out 1000 Monte Carlo replications, and the number of bootstrap resamples is $B = 300$. Additionally, four different values of the smoothing parameter, J_n are examined for the JCS_n type tests, including $J_n = \{0.20, , 0.35, 0.50, 0.60\}$; and four different values of the uniformity constant, η are examined for the S_n type tests, including $\eta = \{0.045, 0.060, 0.075, 0.090\}$.[13] For the S_n type tests, when constructing $\overline{\Sigma}_n^{G+}$(as well as $\overline{\Sigma}_n^{G-}$, etc.) we set $l_n = \text{integer}[n^{0.2}]$ and $\varepsilon = 1e-8$. Finally, when constructing bootstrap statistics for S_n type tests, we set $\kappa_n = \log(n)$ and $c_n = \log(\log(n))$. Sample sizes of $n \in \{250, 500, 1000\}$ are generated using each of the following eight data generating processes (DGPs), with independent forecast errors and *i.i.d.* observations. We do not introduce forecast error dependence or parameter estimation error into our setup, as the effects of these departures from our setup are discussed in JCS (2017). In particular, for the following eight data generating processes (DGPs), we fix $e_{1t} \sim i.i.d.N(0,1)$, and let the number of competing forecasting models vary.

DGP1: $e_{1t} \sim i.i.d.N(0,1)$ and $e_{kt} \sim i.i.d.N(0,1)$, $k = 2, 3$.
DGP2: $e_{1t} \sim i.i.d.N(0,1)$ and $e_{kt} \sim i.i.d.N(0,1)$, $k = 2, 3, 4, 5$.
DGP3: $e_{1t} \sim i.i.d.N(0,1)$, $e_{kt} \sim i.i.d.N(0,1)$, $k = 2, 3, 4, 5$ and $e_{kt} \sim i.i.d.N(0, 1.2^2)$, $k = 6, 7, 8, 9$.
DGP4: $e_{1t} \sim i.i.d.N(0,1)$, $e_{kt} \sim i.i.d.N(0, 0.8^2)$, $k = 2, 3, 4, 5$ and $e_{kt} \sim i.i.d.N(0, 1.2^2)$, $k = 6, 7, 8, 9$.
DGP5: $e_{1t} \sim i.i.d.N(0,1)$ and $e_{kt} \sim i.i.d.N(0, 0.8^2)$, $k = 2, 3$.
DGP6: $e_{1t} \sim i.i.d.N(0,1)$ and $e_{kt} \sim i.i.d.N(0, 0.6^2)$, $k = 2, 3$.
DGP7: $e_{1t} \sim i.i.d.N(0,1)$ and $e_{kt} \sim i.i.d.N(0, 0.8^2)$, $k = 2, 3, 4, 5$.
DGP8: $e_{1t} \sim i.i.d.N(0,1)$ and $e_{kt} \sim i.i.d.N(0, 0.6^2)$, $k = 2, 3, 4, 5$.

Here, DGPs 1-3 are our "null" models, while DGPs 4-8 are our "alternative" models. DGPs 1 and 2 correspond to the least favorable elements in the null. In DGP3, the benchmark model outperforms some of the competing models, while in DGP4, one half of the competing models outperform the benchmark

[12] Consider S_n^{G+} and S_n^{G-} in order to illustrate how our statistics are constructed. Namely,

$$S_n^{G+} = \int_{x \in \mathcal{X}^+} \sum_{j=2}^{k} \left(\max \left\{ 0, \sqrt{n} \frac{G_{j,n}(x)}{\overline{\sigma}_{jj,n}^G(x)} \right\} \right)^2 dQ(x) \text{ and } S_n^{G-} = \int_{x \in \mathcal{X}^-} \sum_{j=2}^{k} \left(\max \left\{ 0, \sqrt{n} \frac{G_{j,n}(x)}{\overline{\sigma}_{jj,n}^G(x)} \right\} \right)^2 dQ(x).$$

Or, using simpler notation,

$$S_n^{G+} = \int_{x \in \mathcal{X}^+} H_n(x)\, dQ(x) \text{ and } S_n^{G-} = \int_{x \in \mathcal{X}^-} H_n(x)\, dQ(x),$$

say. As an example, call the 1% and 99% percentiles of the pooled empirical distribution are $c1$ and $c2$, and set these values equal to -0.0026 and 0.0039, respectively. Then the actual 98% support is [-0.0026, 0.0039]. We form an equally spaced grid, with size $[1.5n^{0.6}]$, over this range. To approximate the integral, instead of directly using dx (in our example, $dx = 0.0002$ if $n = 200$), we use $dQ(x) = \frac{dx}{c2-c1} = \frac{1}{1.5n^{0.6}}$. Thus, $Q(\cdot)$ is still uniform. Also, $H_n(x)$ is still evaluated at the original grid of points. For inference using our tests, once B is determined, estimate bootstrap $p-$values, $p_{B,n,S_n}^{G+} = \frac{1}{B}\sum_{s=1}^{B} 1\left((S_n^{*G+} + \eta) \geq S_n^{G+} \right)$. and $p_{B,n,S_n}^{G-} = \frac{1}{B}\sum_{s=1}^{B} 1\left((S_n^{*G-} + \eta) \geq S_n^{G-} \right)$. Then, use the following rules (Holm, (1979): Reject H_0^{TG} at level α, if $\min\left\{ p_{B,n,S_n}^{G+}, p_{B,n,S_n}^{G-} \right\} \leq (\alpha - \eta)/2$. Reject H_0^{TC} at level α, if $\min\left\{ p_{B,n,S_n}^{C+}, p_{B,n,S_n}^{C-} \right\} \leq (\alpha - \eta)/2$.

[13] In JCS (2017), the constant that we call J_n is called S_n.

model and the other half underperform. In the rest of the DGPs, the competing models all outperform the benchmark model. These particular DGPs are also examined in JCS (2017), and are utilized in our experiments because they illustrate the trade-offs associated with using JCS_n and S_n forecast superiority tests.

We now discuss the experimental findings gathered in Tables 1 and 2. All reported results are rejection frequencies based on carrying out the JCS_n and S_n tests using a nominal size equal to 0.1. Turning first to Table 1, note that results in this table mirror those found in JCS (2017), as the table contains findings based on JCS_n type tests. The test is well sized, under DGPs 1 and 2, is somewhat undersized under DGP3 when used to test GL forecast superiority, and is quite undersized under DGP3 when used to test CL forecast superiority. Just as importantly, the power of the GL forecast superiority version of the JCS_n test is as low as 0.382 (under DGP4, when $n = 250$). The analogous power of the CL forecast superiority test is 0.639. Note that analogous rejection frequencies for the S_n type test are 0.958 and 0.961 (see Table 2). On the other hand, for our largest sample size of $n = 1000$, both types of tests have very good empirical power (compare Tables 1 and 2). Thus, at least for relatively small samples, there appears to be a marked improvement in finite sample empirical power when using S_n type tests for checking forecast superiority, relative to JCS_n type tests. That said, it should also be pointed out that the S_n type tests are clearly quite sensitive to the choice of η. When η is "too small", the tests are quite oversized (see Table 2). The empirical size of S_n type tests appears "best" when η is approximately in the range $[0.075, 0.090]$. For this reason, in our empirical analysis, we only report findings for $\eta = 0.075$ and 0.090.[14]

6 Robust Forecast Evaluation of SPF Expert Pools

In the real-time forecasting literature, predictions from econometric models are often compared with surveys of expert forecasters.[15] Such comparisons are important when assessing the implications associated with using econometric models in policy setting contexts, for example. One key survey dataset collecting expert predictions is the *Survey of Professional Forecasters* (SPF), which is maintained by the Philadelphia Federal Reserve Bank (see Croushore (1993)). This dataset, formerly known as the *American Statistical Association/National Bureau of Economic Research Economic Outlook Survey,* collects predictions on various key economic indicators (including, for example, nominal GDP growth, real GDP growth, prices, unemployment, and industrial production). For further discussion of the variables contained in the SPF, refer to Croushore (1993) and Aiolfi, Capistrán, and Timmermann (2011). The SPF has been examined in numerous papers in recent decades. For example, Zarnowitz and Braun (1992) comprehensively study the SPF, and find, among other things, that use of the mean or median provides

[14] For a discussion of simulation results based on application of the Diebold and Mariano (DM: 1995) test (in which specific loss functions are utilized) in our experimental setup, when applied using JCS_n type tests, refer to JCS(2017). Summarizing from that paper, it is clear that when the loss function is unknown, there is an advantage to using our approach of testing for forecast superiority. However the DM test for pairwise comparison or a reality check test for multiple comparisons might yield improved power, for a given loss function. Indeed, under quadratic loss, JCS (2017) show that when the sample size is small, the DM test has better power performance than JCS_n type tests. When the sample size increases, the power difference between the two tests becomes smaller. This is as expected.

[15] See Fair and Shiller (1990), Swanson and White (1997a,b), Aiolfi, Capistrán and Timmermann (2011), and the references cited therein for further discussion.

a consensus forecast with lower average errors than most individual forecasts. More recently, Aiolfi, Capistrán, and Timmermann (2011) consider combinations of the subjective SPF survey forecasts, and find that equal weighted averages of survey forecasts outperform model based forecasts, although in some cases these mean forecasts can be improved upon by averaging them with mean econometric model-based forecasts. When utilizing European data from the recently released ECB SPF, Genre, Kenny, Meyler, and Timmermann (2013) again find that it is very difficult to beat the simple average. This well known result pervades the macroeconometric forecasting literature, and reasons for the success of such simple forecast averaging are discussed in Timmermann (2006). For example, Timmermann notes that model misspecification related to instability (non-stationarities) and estimation error in situations where there are many models and relatively few observations may account to some degree for the success of simple forecast and model averaging.

In this section, we address the issue of forecast averaging and combination by viewing the problem through the lens of forecast superiority testing. In particular, we argue that our loss function free superiority tests can be used to shed additional light on the reasons for the success of simple averaging methods. The primary motivation for our analysis is that the majority of papers in this field focus on specific loss functions. We instead use the forecast superiority tests discussed above. Our approach of using loss function robust tests differs from the approach taken by Elliott, Timmermann, and Komunjer (2005,2008), where the rationality of sequences of forecasts are evaluated by determining whether there exists a particular loss function under which the forecasts are rational. We instead evaluate predictive accuracy irrespective of the loss function implicitly used by the forecaster, and determine whether certain forecast combinations are superior when compared against any loss function, regardless of how the forecasts were constructed. In our tests, the benchmarks against which we compare various forecast combinations are simple average and median consensus forecasts. We aim to assess whether the well documented success of these benchmark combinations remains intact when they are compared against other combinations, under generic loss.[16] To this end, we examine expert SPF predictions of nominal GDP growth.

The remainder of this section is divided into three subsections, including: (i) data, (ii) empirical setup and combination methods, and (iii) empirical findings.

6.1 SPF Dataset

The SPF is a quarterly survey, and the dataset is available at the Philadelphia Federal Reserve Bank (PFRB) website. The original survey began in 1968:Q4, and PFRB took control of it in 1990:Q2; but from that date, there are only around 100 quarterly observations. In our analysis we use the entire dataset of over 160 observations. However, it should be noted that the timing of the survey was not known with certainty prior to 1990. Still, PFRB documentation states that they believe, although are not sure, that the timing of the survey was similar before and after they took control of it.

For our analysis, we consider 5 forecast horizons (i.e., $h = 0, 1, 2, 3, 4$). The reason we use $h = 0$ for one of the horizons is that the first horizon for which survey participants predict GDP growth is the

[16] For an interesting discussion of machine learning and forecast combination methods, see Lahiri, Peng, and Zhao (2017); and for a discussion of probability forecasting and calibrated combining using the SPF, see Lahiri, Peng, and Zhao (2015). In these papers, various cases where consensus combinations do not "win" are discussed.

quarter in which they are making their predictions. In light of this, forecasts made at $h = 0$ are called nowcasts. Nowcasts are very important in policy making settings, since first release GDP data are not available until around the middle of the subsequent quarter. The nominal GDP variable that we examine is called NGDP in the SPF. All test statistics constructed in this section are based on NGDP growth rate prediction errors associated. In particular, assume that one survey participant makes a forecast of NGDP, say $y^f_{t+h}|\mathcal{F}_t$.[17] The associated forecast error is:

$$e_t = \{\ln(y_{t+h}) - \ln(y_t)\} - \left\{\ln(y^f_{t+h}|\mathcal{F}_t) - \ln(y_t)\right\} = \ln(y_{t+h}) - \ln(y^f_{t+h}|\mathcal{F}_t),$$

where the actual NGDP value, y_{t+h}, is reported in the SPF, along with the NGDP predictions of each survey participant. Note that when $h = 0$, $\mathcal{F}_t$ does not include y_t. However, for $h > 0$, $\mathcal{F}_t$ includes y_t. As discussed previously, this is due to the release dates associated with the availability of NGDP data.

6.2 Empirical Setup and Combination Methods

When construct all S_n type test statistics, including S_n^{G+}, S_n^{G-}, S_n^{C-}, and S_n^{C+}. In particular, GL forecast superiority is tested using S_n^{G+} and S_n^{G-} statistics; while CL forecast superiority is tested using S_n^{C+} and S_n^{C-} statistics. We also test for forecast superiority using the JCS_n type tests discussed above, which are not uniformly valid and have correct size only under the least favorable case under the null. In particular, we construct JCS_n^{G+}, JCS_n^{G-}, JCS_n^{C-}, and JCS_n^{C+} test statistics (see Section 2 for further details). All test statistics are calculated using the same parameter values (for B, J_n, η, l_n, and ε) as used in our Monte Carlo experiments, although results are only reported for $J_n = \{0.20, 0.35\}$ and $\eta = \{0.075, 0.090\}$, as these are the values that yielded the best results in our Monte Carlo experiments.

Two different benchmark models are considered, including (i) the arithmetic mean prediction from all participants; and (ii) the median prediction from all participants. Additionally, a variety of alternative combinations are considered.

The first group of alternative combinations allows us to answer the question: *Does experience matter?* Combinations in this group include:

∘ Combination 1a: Mean (or Median) of all participants with at least 1 year of experience.

∘ Combination 1b: Mean (or Median) of all participants with at least 3 years of experience.

∘ Combination 1c: Mean (or Median) of all participants with at least 5 years of experience.

In all of the remaining groups of combinations, individuals are ranked according to average absolute forecast errors, as well as according to average squared forecast errors. Mean (or median) predictions from these groups are then compared with our benchmark combinations.

In the first such set of combinations, we ask the following question: *Does the most highly ranked expert exhibit forecast superiority, when compared with our benchmark combinations, under various experience levels?* This leads to the following combinations:

∘ Combination 2a: Highest ranked participant in last 1 year.

∘ Combination 2b: Highest ranked participant in last 3 years.

∘ Combination 2c: Highest ranked participant in last 5 years.

[17] Here, $\mathcal{F}_t$ denotes the information set available to the expert forecaster at the time their predictions are made.

We then ask the following question: *Does the most highly ranked group of 3 experts exhibit forecast superiority, when compared with our benchmark combinations, under various experience levels?* This leads to the following combinations:

∘ Combination 3a: Highest ranked group of 3 participants in last 1 year.

∘ Combination 3b: Highest ranked group of 3 participants in last 3 years.

∘ Combination 3c: Highest ranked group of 3 participants in last 5 years.

We next ask the following question: *Do groups including the top 10% of participants, exhibit forecast superiority, when compared with our benchmark combinations, under various experience levels?* This leads to the following combinations:

∘ Combination 4a: Group of top 10% of participants in last 1 year.

∘ Combination 4b: Group of top 10% of participants in last 3 years.

∘ Combination 4c: Group of top 10% of participants in last 5 years.

Finally, we ask the following question: *Do groups including the top 25% of participants, exhibit forecast superiority, when compared with our benchmark combinations, under various experience levels?* This leads to the following combinations:

∘ Combination 5a: Group of top 25% of participants in last 1 year.

∘ Combination 5b: Group of top 25% of participants in last 3 years.

∘ Combination 5c: Group of top 25% of participants in last 5 years.

6.3 Empirical Findings

In Panels A-E of Tables 3-8, test statistics are reported under the heading "Statistic Values". Rejection (or not) of the null hypothesis of equal forecast accuracy is reported by "no" (indicating rejection) or "yes" (indicating the converse), under the headers "Rejection with p-Value Type I" (for $\eta = 0.075$ and $J_n = 0.20$) or "Rejection with p-Value Type II" (for $\eta = 0.090$ and $J_n = 0.35$). Nominal test size is 10%. In Panel F of the tables, root mean square forecast errors (RMSFEs) are reported for all of the combinations evaluated.

In the remainder of this subsection, we discuss our findings by answering the questions posed above. Due to the large volume of output associated with carrying out our analysis, most results are contained only in an online appendix. Here, we discuss selected empirical results that are presented in Tables 2-8. However, the conclusions that we draw are the qualitatively the same as those associated with inspection of our entire set of empirical results.

∘ *Does experience matter?*

Inspection of the results reported in Table 3 indicates that the answer to this question is no. The forecast superiority null hypothesis fails to reject in all but one case (i.e., $h = 1$, for S_n type GL forecast superiority). Namely, for nowcasting ($h = 0$) as well as forecasting ($h = 1, 2, 3, 4$), a benchmark that utilizes all predictions from all survey participants is not dominated by pools of experts chosen based on whether they have at least 1, 3, or 5 years of experience. Thus, when compared only with experienced pools of experts, the "best" pool contains all experts, regardless of experience. As stressed elsewhere in this paper, this result is robust to the choice of loss function. For the time being, then, our analysis tends to support utilizing predictions that take advantage of the largest pool of experts possible, in

agreement with the findings of Aiolfi, Capistrán, and Timmermann (2011) and Genre, Kenny, Meyler, and Timmermann (2013), for example.

Additionally, it is worth noting that loss functions do matter, as should be expected. This can be seen by noting that some alternative models are slightly better the entire pool of experts, when comparing quadratic loss (see the RMSFEs in Panel F of the table). For example, "Alt Model 1", which is Combination 1a above, and which pools all experts with at least 1 year of experience, yields slightly lower RMSFEs than the benchmark, for all values of h. On the other hand, results for "Alt Model 2" (i.e., Combination 1b) are mixed, with the benchmark winning for some values of h, while pooling experts with at least 5 years of experience (i.e., "Alt Model 3", which is Combination 1c) never results in a lower RMSFE than the arithmetic mean benchmark, regardless of h.[18]

Finally, it is worth noting that when the benchmark is the median rather than mean forecast, and when the alternative combinations are selectively pooled median forecasts, results are the same as those reported above (see Table 4). Namely, the forecast superiority null hypothesis fails to reject in all cases (i.e., for all values of h, and for both S_n and JCS_n type GL and CL tests).

∘ *Does the most highly ranked expert exhibit forecast superiority, when compared with our benchmark combinations, under various experience levels?*

Turning to Table 5, results when the alternative models are Combinations 2a-c are largely the same as those reported above. Namely, the forecast superiority null hypothesis fails to reject in 37 of 40 cases. Thus, the entire pool of participants is preferred to the single "best" participant, regardless of experience level. Recall that in this context, "best" refers to the lowest average least absolute deviation forecast error participant based on either 1, 3, or 5 years of experience. However, results based on ranking according to square forecast error performance rather than absolute forecast error performance are the same as those reported in Table 5. The reason why we report results from experiments where participant rankings are done using absolute forecast error performance is that RMSFEs of all of our alternative combinations are lower when ranking is done according to absolute forecast error performance than when ranking is done according to square forecast error performance. For this reason, and because our qualitative findings do not change based on ranking method, we report subsequent results only for absolute forecast error rankings.

∘ *Does the most highly ranked group of 3 experts exhibit forecast superiority, when compared with our benchmark combinations, under various experience levels?*

The story changes when the alternative models are Combinations 3a-c. In particular, in Table 6 note that the null of equal forecast accuracy is rejected in many cases, when 3 experts comprise the pool. For example, the null hypothesis is rejected in 14 of 20 cases, across all values of h, when S_n statistics are used to test forecast superiority. The same cannot be said for JCS_n type GL tests (rejection occurs in only 8 of 20 cases), although this might be expected given the lower power of JCS_n type tests when n is small. Evidence based on analysis of the RMSFEs in Panel F if the table suggests that for $h = 0$, more than 1 year of experience is preferred, while for all other values of h, 1 year of experience is enough, and delivers a pool of 3 experts that outperforms the entire pool of experts. Thus, we have direct evidence

[18] In Tables 3-4, "Alt Models 1-3" correspond to Combinations 1a-c, respectively. Also, in Table x , "Alt Model" 1-3 corresponds to Combinations xa-c, for x=5,6,7,8.

that experience coupled with selecting a small group of the very best experts leads to loss function robust forecast superiority, in many cases. This finding is in contrast to findings discussed above where the entire pool yield combinations that are preferred. However, our finding is in broad agreement with Lahiri, Peng, and Zhao (2015), who find that using "valuable" individual forecasts leads to predictions that outperform simple average predictions.[19]

∘ *Do groups including the top 10% of participants, exhibit forecast superiority, when compared with our benchmark combinations, under various experience levels?*

Results based on Combinations 4a-c are reported in Table 7, and again indicate various rejections of the null of equal forecast accuracy. Namely, we observe rejections in 12 of 20 cases, across all values of h, when S_n statistics are used to test forecast superiority. Here, the incidence of rejection of the null is not quite as pervasive as under Combinations 3a-c. Apparently, a 10% cut-off for the number of experts is not quite enough to see the benefits associated with utilizing experienced expert pools.

∘ *Do groups including the top 25% of participants, exhibit forecast superiority, when compared with our benchmark combinations, under various experience levels?*

However, Combinations 5a-c, which utilize the top 25% of experts again lead to rejection of the null hypothesis in 14 of 20 cases, across all values of h, when S_n statistics are used to test forecast superiority. Of course, this finding is clearly dependent upon the number of participants in the survey. Nevertheless, it is interesting to further note that the 25% cut-off leads to rejection in 7 of 10 cases when S_n type GL forecast superiority is tested for, as well as in 7 of 10 cases when S_n type CL forecast superiority is tested for. Evidence based on analysis of the RMSFEs in Panel F of the table suggest that for $h = 0$, all levels of experience (i.e., 1, 3, or 5 years of experience) yield improved forecast performance, while for all other values of h, requiring at least 3 year of experience is needed in order to deliver a pool of "top 25%" experts that outperforms the entire pool of experts.

Summarizing, we have direct evidence that judicious selection of a pool of the very best experienced experts can lead to loss function robust forecast superiority.

7 Concluding Remarks

We develop uniformly valid forecast superiority tests that are asymptotically correctly sized, and that are robust to the choice of loss function. Our tests are based on principles of stochastic dominance, which can be interpreted as tests for infinitely many moment inequalities. In light of this, we use tools from Andrews and Shi (2013, 2017) when developing our tests. The tests build on earlier work due to Jin, Corradi, and Swanson (2017), and are meant to provide a class of predictive accuracy tests that are not reliant on a choice of loss function, such as the Diebold and Mariano (1995) test discussed in McCracken (2000). In developing the new tests, we establish uniform convergence (over error support) of HAC variance estimators, and of their bootstrap counterparts. We also extend the theory of generalized moment selection testing to allow for the presence of non-vanishing parameter estimation error. In a series of Monte Carlo experiments, we show that finite sample performance of our tests is quite good,

[19] Interestingly, Lahiri, Peng, and Zhao (2015) find that the numbers of forecasters making valuable predictions dimishes as h increases. Our results, which are loss function robust, do not make this distinction.

and that the power of our tests dominates those proposed by JCS (2017). Additionally, we carry out an extensive empirical analysis of the well known Survey of Professional Forecasters, and show that utilizing expert pools based on past forecast quality as well as years of experience can lead to loss function robust forecast superiority, when compared with pools that include all survey participants.

8 Appendix

Proof of Lemma 1: **(i)** The proof is the same for all j. Thus, let $u_t(x) = (1\{e_{j,t} \leq x\} - F_j(x)) - (1\{e_{1,t} \leq x\} - F_1(x))$, and define

$$\widehat{\widehat{\sigma}}_n^{2,G+}(x) = \frac{1}{n}\sum_{t=1}^{n} u_t^2(x) + 2\frac{1}{n}\sum_{\tau=1}^{l_n} w_\tau u_t(x)u_{t-\tau}(x).$$

We first show that

$$\sup_{x\in\mathcal{X}^+}\left|\widehat{\widehat{\sigma}}_n^{2,G+}(x) - \sigma^{2,G+}(x)\right| = o_p(1),$$

and then we show that

$$\sup_{x\in\mathcal{X}^+}\left|\widehat{\widehat{\sigma}}_n^{2,G+}(x) - \widehat{\sigma}_n^{2,G+}(x)\right| = o_p(1). \tag{8.1}$$

Now,

$$\begin{aligned}
&\sup_{x\in\mathcal{X}^+}\left|\widehat{\widehat{\sigma}}_n^{2,G+}(x) - \sigma^{2,G+}(x)\right| \\
\leq\ &\sup_{x\in\mathcal{X}^+}\left|\frac{1}{n}\sum_{t=1}^{n}\left(u_t^2(x) - \mathrm{E}\left(u_t^2(x)\right)\right) + 2\frac{1}{n}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n}\left(u_t(x)u_{t-\tau}(x) - \mathrm{E}\left(u_t(x)u_{t-\tau}(x)\right)\right)\right| \\
&+ \sup_{x\in\mathcal{X}^+}\left|\left(\sigma^2(x) - \frac{1}{n}\sum_{t=1}^{n}\mathrm{E}\left(u_t^2(x)\right) + 2\frac{1}{n}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n}\mathrm{E}\left(u_t(x)u_{t-\tau}(x)\right)\right)\right|.
\end{aligned} \tag{8.2}$$

We begin with the first term on the RHS of (8.2). First note that,

$$\begin{aligned}
&\sup_{x\in\mathcal{X}^+}\left|\frac{1}{n}\sum_{t=1}^{n}\left(u_t^2(x) - \mathrm{E}\left(u_t^2(x)\right)\right) + 2\frac{1}{n}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n}\left(u_t(x)u_{t-\tau}(x) - \mathrm{E}\left(u_t(x)u_{t-\tau}(x)\right)\right)\right| \\
\leq\ &\sup_{x\in\mathcal{X}^+} 2\sum_{\tau=0}^{l_n}\left|\frac{1}{n}\sum_{t=1}^{n}\left(u_t(x)u_{t-\tau}(x) - \mathrm{E}\left(u_t(x)u_{t-\tau}(x)\right)\right)\right|.
\end{aligned}$$

Now,

$$\begin{aligned}
&\Pr\left(\sup_{x\in\mathcal{X}^+} 2\sum_{\tau=0}^{l_n}\left|\frac{1}{n}\sum_{t=1}^{n}\left(u_t(x)u_{t-\tau}(x) - \mathrm{E}\left(u_t(x)u_{t-\tau}(x)\right)\right)\right| \geq \varepsilon\right) \\
\leq\ &2\sum_{\tau=0}^{l_n}\Pr\left(\sup_{x\in\mathcal{X}^+}\left|\frac{1}{n}\sum_{t=1}^{n}\left(u_t(x)u_{t-\tau}(x) - \mathrm{E}\left(u_t(x)u_{t-\tau}(x)\right)\right)\right| \geq \frac{\varepsilon}{l_n}\right),
\end{aligned}$$

so that we need to show that,

$$\Pr\left(\sup_{x\in\mathcal{X}^+}\left|\frac{1}{n}\sum_{t=1}^{n}\left(u_t(x)u_{t-\tau}(x) - \mathrm{E}\left(u_t(x)u_{t-\tau}(x)\right)\right)\right| \geq \frac{\varepsilon}{l_n}\right) < \frac{\delta}{l_n}.$$

Given Assumption A2, WLOG, we can set $\mathcal{X}^+ = [0, \Delta]$, so that it can be covered by a_n^{-1} balls S_j, $j = 1, ..., \Delta a_n^{-1}$, centered at S_j, with radius a_n. Then,

$$\begin{aligned}
& \sup_{x \in \mathcal{X}^+} \left| \frac{1}{n} \sum_{t=1}^{n} \left(u_t(x) u_{t-\tau}(x) - \mathrm{E}\left(u_t(x) u_{t-\tau}(x) \right) \right) \right| \\
\leq & \max_{j=1,..,\Delta a_n^{-1}} \left| \frac{1}{n} \sum_{t=1}^{n} \left(u_t(s_j) u_{t-\tau}(s_j) - \mathrm{E}\left(u_t(s_j) u_{t-\tau}(s_j) \right) \right) \right| \\
& + \max_{j=1,..,\Delta a_n^{-1}} \sup_{x \in S_j} 2 \left| \left(\frac{1}{n} \sum_{t=1}^{n} u_{t-\tau}(s_j) \left(u_t(x) - u_t(s_j) \right) \right) \right. \\
& \left. - \left(\frac{1}{n} \sum_{t=1}^{n} \mathrm{E}\left(u_{t-\tau}(s_j) \left(u_t(x) - u_t(s_j) \right) \right) \right) \right| \\
& + \text{smaller order} \\
= & \; I_n + II_n.
\end{aligned}$$

Now,

$$\begin{aligned}
II_n \leq & \max_{j=1,..,\Delta a_n^{-1}} \sup_{x \in S_j} \left| \frac{1}{n} \sum_{t=1}^{n} u_{t-\tau}(s_j) \left(u_t(x) - u_t(s_j) \right) \right| \\
& + \max_{j=1,..,\Delta a_n^{-1}} \sup_{x \in S_j} \left| \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}\left(u_{t-\tau}(s_j) \left(u_t(x) - u_t(s_j) \right) \right) \right| \\
= & \; II_n^A + II_n^B.
\end{aligned}$$

Given Assumption A1, noting that by Cauchy - Schwarz,

$$\begin{aligned}
& \max_{j=1,..,\Delta a_n^{-1}} \sup_{x \in S_j} \left| \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}\left(u_{t-\tau}(s_j) \left(u_t(s_j) - u_{t-\tau}(s_j) \right) \right) \right| \\
\leq & \max_{j=1,..,\Delta a_n^{-1}} \sup_{x \in S_j} \sqrt{\mathrm{E}\left(u_{t-\tau}(s_j) \right)^2} \max_{j=1,..,a_n^{-1}} \sup_{x \in S_j} \sqrt{\mathrm{E}\left(u_t(s_j) - u_t(x) \right)^2} \\
= & \; O\left(a_n^{1/2} \right),
\end{aligned}$$

for some constant C. Recalling given that $u_t(x) = \left(1\left\{ e_{j,t} \leq x \right\} - F_j(x) \right) - \left(1\left\{ e_{1,t} \leq x \right\} - F_1(x) \right)$ and $u_t(s_j)$ stay between -1 and 1

$$\begin{aligned}
& \max_{j=1,..,\Delta a_n^{-1}} \sup_{x \in S_j} \left| \frac{1}{n} \sum_{t=1}^{n} u_{t-\tau}(s_j) \left(u_t(s_j) - u_t(x) \right) \right| \\
\leq & \; 2 \max_{j=1,..,\Delta a_n^{-1}} \sup_{x \in S_j} \frac{1}{n} \sum_{t=1}^{n} \left| u_t(s_j) - u_t(x) \right| \\
\leq & \; \frac{2}{n} \sum_{t=1}^{n} 1\left\{ x - a_n \leq e_{1,t} \leq x + a_n \right\} + \frac{2}{n} \sum_{t=1}^{n} 1\left\{ x - a_n \leq e_{j,t} \leq x + a_n \right\} \\
& + 2 \sup_{x \in \mathcal{X}^+} \left(f_1(x) + f_j(x) \right) \\
= & \; O_p\left(a_n \right) = o_p(a_n^{1/2})
\end{aligned}$$

Hence, by Chebyshev inequality

$$l_n \Pr\left(II_n > \frac{\varepsilon}{l_n}\right) = O\left(a_n l_n^3\right) = o(1),$$

for $a_n = o\left(l_n^{-3}\right)$.

Now, consider I_n. By the Lemma on page 739 of Hansen (2008), setting $a_n = l_n^{-4}$, $m = \frac{\Delta n}{4l_n^2}$, and $l_n = n^\delta$, with $\delta < 1/2$, and recalling that given Assumption A1, $\text{var}\left(\sum_{t=1}^m \left(u_t(s_j)u_{t-\tau}(s_j) - \text{E}\left(u_t(s_j)u_{t-\tau}(s_j)\right)\right)\right) \leq Cm$, it follows that for some constant C,

$$\begin{aligned}
& \Pr\left(\max_{j=1,..,a_n^{-1}} \left|\frac{1}{n}\sum_{t=1}^n \left(u_t(s_j)u_{t-\tau}(s_j) - \text{E}\left(u_t(s_j)u_{t-\tau}(s_j)\right)\right)\right| \geq \frac{\varepsilon}{l_n}\right) \\
\leq\ & a_n^{-1} \Pr\left(\left|\sum_{t=1}^n \left(u_t(s_j)u_{t-\tau}(s_j) - \text{E}\left(u_t(s_j)u_{t-\tau}(s_j)\right)\right)\right| \geq \frac{n\varepsilon}{l_n}\right) \\
\leq\ & 4a_n^{-1}\left(\exp\left(-\frac{\frac{n^2}{l_n^2}\varepsilon^2}{64Cn + \frac{8}{3}\frac{\Delta n^2}{4l_n^3}}\right) + \frac{16}{b} l_n^2 \left(\frac{4}{\Delta}\frac{n}{l_n^2}\right)^{-\beta}\right) \\
=\ & a_n^{-1}\exp\left(-\frac{1}{64C\frac{n}{l_n^2} + \frac{8}{3}\frac{\Delta n^2}{4l_n^3}}\right) + \frac{64}{b} a_n^{-1} l_n^2 l_n^{2\beta} n^{-\beta} \\
=\ & o(1) + O\left(n^{\delta(6+2\beta)} n^{-\beta}\right) \\
=\ & o(1) \text{ for } \beta > \frac{6\delta}{1-2\delta}.
\end{aligned}$$

We now consider the second term on the RHS of (8.2). Note that

$$\begin{aligned}
& \sup_{x\in\mathcal{X}^+} \left|\left(\sigma^{2,G+}(x) - \frac{1}{n}\sum_{t=1}^n \text{E}\left(u_t^2(x)\right) + 2\frac{1}{n}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^n \text{E}\left(u_t(x)u_{t-\tau}(x)\right)\right)\right| \\
\leq\ & 2\sup_{x\in\mathcal{X}^+}\left|\frac{1}{n}\sum_{\tau=1}^{l_n}(1-w_\tau)\sum_{t=1}^n \text{E}\left(u_t(x)u_{t-\tau}(x)\right)\right| \\
& +2\sup_{x\in\mathcal{X}^+}\left|\frac{1}{n}\sum_{\tau=l_n+1}^{n} w_\tau \sum_{t=1}^n \text{E}\left(u_t(x)u_{t-\tau}(x)\right)\right|.
\end{aligned} \tag{8.3}$$

The first term on the RHS of (8.3) is $o_p(1)$, by the same argument as that used in Theorem 2 of Newey and West (1987). Also, by Lemma 6.17 in White (1984), for $q > 2$,

$$\text{E}\left(u_t(x)u_{t-\tau}(x)\right) \leq C\tau^{-\beta/2-1/q}\text{var}\left(u_t(x)\right)^{1/2}\text{E}\left\|u_t(x)\right\|^q$$

and

$$\begin{aligned}
& \sup_{x\in\mathcal{X}^+}\left|\frac{1}{n}\sum_{\tau=l_n+1}^{n} w_\tau \sum_{t=1}^n \text{E}\left(u_t(x)u_{t-\tau}(x)\right)\right| \\
\leq\ & C\sup_{x\in\mathcal{X}^+}\text{var}\left(u_t(x)\right)^{1/2}\text{E}\left\|u_t(x)\right\|^q \sum_{\tau=l_n+1}^{n}\tau^{-\beta/2-1/q} = o(1),
\end{aligned}$$

as $\beta\delta > 1$, given Assumption A1, and noting that q can be taken arbitrarily large because of the boundedness of $u_t(x)$.

Finally, by the same argument as that used in the proof of (8.2), for all j,

$$\sup_{x\in\mathcal{X}^+}\frac{1}{n}\sum_{t=1}^{n}\left(1\left\{e_{j,t}\leq x\right\}-F_j(x)\right)=o_p\left(l_n^{-1}\right).$$

The statement in (8.1) follows immediately.

(ii) By noting that,

$$\begin{aligned}
& [e_{j,t}-s_j]_+-[e_{j,t}-x]_+ \\
= \;& (x-s_j)1\{e_t\geq x\}+(x-s_j)\left(1\{e_t\geq x\}-1\{e_t\geq s_j\}\right) \\
& +(e_t-x)\left(1\{e_t\geq s_j\}-1\{e_t\geq x\}\right),
\end{aligned}$$

the statement follows by the same argument as that used in part (i) of the proof.

Proof of Lemma 2: For notational simplicity, we suppress the jj subscript. Also, we suppress the superscripts C^+ and G^+, as the proof follows by analogous argument. Note that

$$\begin{aligned}
& \sup_{x\in\mathcal{X}^+}\left|\widehat{\sigma}_n^{*2}(x)-\mathrm{E}^*\left(\widehat{\sigma}_n^*(x)\right)\right| \\
\leq \;& \sup_{x\in\mathcal{X}^+}\frac{l_n}{b}\sum_{k=1}^{b}\left|\left(\frac{1}{l_n}\sum_{j=1}^{l_n}u^*_{(k-1)l_n+j}(x)\right)^2-\mathrm{E}^*\left(\left(\frac{1}{l_n}\sum_{j=1}^{l_n}u^*_{(k-1)l_n+j}(x)\right)^2\right)\right| \\
= \;& \sup_{x\in\mathcal{X}^+}\frac{l_n}{b}\sum_{k=1}^{b}\left|\frac{1}{l_n^2}\sum_{j=1}^{l_n}\sum_{i=1}^{l_n}u^*_{(k-1)l_n+j}(x)u^*_{(k-1)l_n+i}(x)-\mathrm{E}^*\left(u^*_{(k-1)l_n+j}(x)u^*_{(k-1)l_n+i}(x)\right)\right|
\end{aligned}$$

Now,

$$\begin{aligned}
& \Pr\left(\sup_{x\in\mathcal{X}^+}\frac{l_n}{b}\sum_{k=1}^{b}\left|\frac{1}{l_n^2}\sum_{j=1}^{l_n}\sum_{i=1}^{l_n}u^*_{(k-1)l_n+j}(x)u^*_{(k-1)l_n+i}(x)-\mathrm{E}^*\left(u^*_{(k-1)l_n+j}(x)u^*_{(k-1)l_n+i}(x)\right)\right|\geq\varepsilon_1 a_n\right) \\
\leq \;& l_n\Pr\left(\sup_{x\in\mathcal{X}^+}\frac{l_n}{b}\sum_{k=1}^{b}\left|\frac{1}{l_n^2}\sum_{j=1}^{l_n}\sum_{i=1}^{l_n}u^*_{(k-1)l_n+j}(x)u^*_{(k-1)l_n+i}(x)-\mathrm{E}^*\left(u^*_{(k-1)l_n+j}(x)u^*_{(k-1)l_n+i}(x)\right)\right|\geq\varepsilon_1\frac{a_n}{l_n}\right).
\end{aligned}$$

It suffices to show that, uniformly in k,

$$\Pr\left(\sup_{x\in\mathcal{X}^+}\left|\frac{1}{l_n^2}\sum_{j=1}^{l_n}\sum_{i=1}^{l_n}u^*_{(k-1)l+j}(x)u^*_{(k-1)l+i}(x)-\mathrm{E}^*\left(u^*_{(k-1)l+j}(x)u^*_{(k-1)l+i}(x)\right)\right|\geq\varepsilon_1\frac{a_n}{l_n}\right)<\frac{\delta}{l_n}.$$

This follows using the same "covering numbers" argument used in the proof of Lemma 1.

Proof of Theorem 1: We again suppress the superscripts G^+ and C^+, as the proof follows by the same argument. We need to show that the statement in Lemma A1 in the Supplement Appendix of Andrews and Shi (2013) holds. Then, the proof of the theorem will follow using the same arguments as those used

in the proof of their Theorem 1, as the proof is the same for independent and dependent observations. In fact, our set-up differs from Andrews and Shi (2013) only because we have dependent observations, and because we scale the statistic by a Newey-West variance estimator. For the rest of the proof, our set-up is simpler as we can fix their θ_n at a given value, say zero. It suffices to show that:
(i) $v_n(.) \Rightarrow v(.)$, as a process indexed by $x \in \mathcal{X}^+$, where $v(.)$ is a zero-mean $k-1-$dimensional Gaussian process, with covariance kernel given $\Sigma(x, x')$.
(ii) $\sup_{x,x' \in \mathcal{X}^+} \left\| \overline{h}_{B,n}(x, x') - \overline{h}_B(x, x') \right\| = o_p(1)$.
Now, statement (ii) follows directly from Lemma 1. It remains to show that (i) holds. The key difference between the independent and the dependent cases is that in the former we can rely on the concept of manageability, while in the latter we cannot. Nevertheless, (i) follows if we can show that $v_n(.)$ satisfies an empirical process. Given A1-A3, this follows from Lemma A2 in Jin, Corradi and Swanson (2017).

Proof of Theorem 2: (i) For notational simplicity, we omit the superscript G^+. The proof of this theorem mirrors the proof of Theorem 2(a) in the Supplement of Andrews and Shi (2013). Let $c_0(h_{A,n}, \alpha)$ be the α critical value of $S_n^\dagger$, as defined in (3.5). Given Theorem 1(i), it follows that for all $\delta > 0$,

$$\lim \sup_{n \to \infty} \sup_{P \in \mathcal{P}_0} P\left(S_n \geq c_0(h_{A,n}, \alpha) + \delta\right) \leq \alpha.$$

The statement follows if we can show that

$$\lim \sup_{n \to \infty} \sup_{P \in \mathcal{P}_0} P\left(c^*_{0,n,\alpha}\left(\phi_n, \overline{h}^*_{B,n}\right) \leq c_{0,n,\alpha}\left(h_{A,n}, \overline{h}^*_{B,n}\right)\right) = 0, \tag{8.4}$$

with $c_{0,n,\alpha}\left(h_{A,n}, \overline{h}^*_{B,n}\right)$ defined as $c_0(h_{A,n}, \alpha)$; but with $\overline{h}^*_{B,n}$ an argument of this function rather than $h_B(x)$; and if we can show that

$$\lim \sup_{n \to \infty} \sup_{P \in \mathcal{P}_0} P\left(c_{0,n,\alpha}\left(h_{A,n}, \overline{h}^*_{B,n}\right) \leq c_0(h_{A,n}, \alpha)\right) = 0. \tag{8.5}$$

For $c_n \to \infty$ and $c_n/\kappa_n \to 0$, $\tau_n \to \infty$ and $\tau_n/\kappa_n \to 0$,

$$\begin{aligned}
& \sup_{P \in \mathcal{P}_0} P\left(c^*_{0,n,\alpha}\left(\phi_n, \overline{h}^*_{B,n}\right) \leq c_{0,n,\alpha}\left(h_{A,n}, \overline{h}^*_{B,n}\right)\right) \\
\leq & \sup_{P \in \mathcal{P}_0} P\left(-\phi_{j,n}(x) \leq h_{A,j,n}(x), \text{ for some } x \in \mathcal{X}^+ \text{ and some } j = 2, .., k\right) \\
\leq & \sup_{P \in \mathcal{P}_0} P\left(\xi_{j,n}(x) < -1 \text{ AND } -c_n \leq h_{A,j,n}(x), \text{ for some } x \in \mathcal{X}^+ \text{ and } j = 2, .., k\right) \\
\leq & \sup_{P \in \mathcal{P}_0} P\Big(D(x)^{1/2}\overline{D}^{-1/2}_{jj,n}(x) v_{j,n}(x) + D(x)^{1/2}\overline{D}^{-1/2}_{jj,n}(x) h_{j,A,n}(x) < -\kappa_n \\
& \text{AND } -c_n \leq h_{A,j,n}(x), \text{ for some } x \in \mathcal{X}^+ \text{ and } j = 2, .., k\Big) \\
\leq & \sup_{P \in \mathcal{P}_0} P\Big(-\tau_n + D(x)^{-1/2}\overline{D}^{-1/2}_{jj,n}(x) h_{j,A,n}(x) < -\kappa_n \\
& \text{AND } -c_n \leq h_{A,j,n}(x), \text{ for some } x \in \mathcal{X}^+ \text{ and } j = 2, .., k\Big) \\
& + \sup_{P \in \mathcal{P}_0} P\left(D(x)^{1/2}\overline{D}^{-1/2}_{jj,n}(x) v_{j,n}(x) < -\tau_n, \text{ for some } x \in \mathcal{X}^+ \text{ and } j = 2, .., k\right) \\
\leq & \sup_{P \in \mathcal{P}_0} P\Big(-D(x)^{-1/2}\overline{D}^{-1/2}_{jj,n}(x) h_{j,A,n}(x) < -\kappa_n + c_n \\
& \text{AND } -c_n \leq h_{A,j,n}(x), \text{ for some } x \in \mathcal{X}^+ \text{ and } j = 2, .., k\Big) \\
= & \ o(1).
\end{aligned}$$

This establishes that (8.4) holds. Finally, (8.5) follows from Lemma 1 and Lemma 2.

(ii) Recall that $c^*_{0,n,1-\alpha}(\phi_n, h_{B,n})$ is the $(1-\alpha)-$ percentile of S^*_n, as defined in (3.11); and define $c^{GMS}_{0,n,1-\alpha}(\phi_n, \overline{h}_{B,n})$ to the $(1-\alpha)-$ percentile of S^{GMS}_n, where

$$S^{GMS}_n = \max_{x\in\mathcal{X}^+} \sum_{j=2}^{k} \max\left(\left\{0, \frac{\overline{v}_{j,n}(x) - \phi_{j,n}(x)}{\sqrt{\overline{h}_{B,jj}(x)}}\right\}\right)^2,$$

with $\overline{v}_n = (\overline{v}_{2,n}, ..., \overline{v}_{k,n})'$ is a $k-1$ dimensional Gaussian process, with mean zero and covariance $\overline{h}_B(x,x') = \widehat{D}_n^{-1/2}(x)\overline{\Sigma}(x,x')\widehat{D}_n^{-1/2}(x')$. We first need to show that

$$c^*_{0,n,1-\alpha}(\phi_n, h_{B,n}) - c^{GMS}_{0,n,1-\alpha}(\phi_n, \overline{h}_{B,n}) = o_p(1), \tag{8.6}$$

and then to prove that the statement holds when replacing $c^*_{0,n,1-\alpha}(\phi_n, h_{B,n})$ with $c^{GMS}_{0,n,1-\alpha}(\phi_n, \overline{h}_{B,n})$.

From Lemma 2, $\widehat{\Sigma}^*_n(x,x') - \widehat{\Sigma}_n(x,x') = o^*_p(1)$, and so $\overline{\Sigma}^*_n(x,x') - \overline{\Sigma}_n(x,x') = o^*_p(1)$. Then, by Theorem 2.3 in Peligrad (1998),

$$v^* \overset{d^*}{\Longrightarrow} v \text{ a.s.-}\omega,$$

where $v^* \overset{*}{\Longrightarrow} v$ denotes weak convergence, conditional on sample. As $\overline{v}_n \Longrightarrow v$, (8.6) follows.

Given Assumption A4, by Lemma B3 in the Supplement of Andrews and Shi (2013), the distribution of $S^\dagger_\infty$, as defined in (3.6), is continuous. It is also strictly increasing and its $(1-\alpha)-$quantile is strictly positive, for all $\alpha < 1/2$. The statement then follows by the same argument as that used in the proof of Theorem 2(b) in the Supplement of Andrews and Shi (2013).

(iii)-(iv) follow by the same arguments as those used in the proof of **(i)** and **(ii)**, respectively. In the case of $S^{G^+}_n$, we rely on the the stochastic equicontinuity of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(1\{e_{1,i} \le x\} - 1\{e_{1,i} \le u\}\right)$, as $|x-u| \to 0$. When considering $S^{C^+}_n$, we need to ensure the stochastic equicontinuity of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \left((e_{1,i}-x)_+ - (e_{1,i}-u)_+\right)$. Now,

$$\begin{aligned} & \frac{1}{\sqrt{n}}\sum_{i=1}^n \left((e_{1,i}-x)_+ - (e_{1,i}-u)_+\right) \\ = \; & \frac{1}{\sqrt{n}}\sum_{i=1}^n (u-x)\,1\{e_{1,i} \ge u\} + \frac{1}{\sqrt{n}}\sum_{i=1}^n (e_{1,i}-u)_+ \left(1\{e_{1,i} \ge x\} - 1\{e_{1,i} \ge u\}\right), \end{aligned}$$

which, given Assumption 2, is stochastically equicontinous, by the same argument as those used for $S^{G^+}_n$. Hence, Theorem 2.3 in Peligrad (1998) also holds in this case.

Proof of Theorem 3: (i) Without loss of generality, let $B^{G+}_{FA} = \{x \in \mathcal{X}^+ : G_2(x) > 0\}$, and note that

for all $x \in B_{FA}^{G+}$, $\max\left\{0, \frac{\sqrt{n}G_{2,n}^{+}(x)}{\overline{\sigma}_{22,n}^{G+}(x)}\right\} = \frac{\sqrt{n}G_{2,n}^{+}(x)}{\overline{\sigma}_{22,n}^{G+}(x)}$. Thus,

$$\begin{aligned}
S_n^{G+} &= \int_{B_{FA}^{G+}} \sum_{j=2}^{k} \left(\max\left\{0, \frac{\sqrt{n}G_{j,n}^{+}(x)}{\overline{\sigma}_{jj,n}^{G+}(x)}\right\}\right)^2 \mathrm{d}Q(x) + \int_{\mathcal{X}^+ \setminus B_{FA}^{G+}} \sum_{j=2}^{k} \left(\max\left\{0, \frac{\sqrt{n}G_{j,n}^{+}(x)}{\overline{\sigma}_{jj,n}^{G+}(x)}\right\}\right)^2 \mathrm{d}Q(x) \\
&= \int_{B_{FA}^{G+}} \left(\frac{\sqrt{n}G_{2,n}^{+}(x)}{\overline{\sigma}_{22,n}^{G+}(x)}\right)^2 \mathrm{d}Q(x) + \int_{B_{FA}^{G+}} \sum_{j=3}^{k} \left(\max\left\{0, \frac{\sqrt{n}G_{j,n}^{+}(x)}{\overline{\sigma}_{jj,n}^{G+}(x)}\right\} - \left(\frac{\sqrt{n}G_{2,n}^{+}(x)}{\overline{\sigma}_{22,n}^{G+}(x)}\right)\right)^2 \mathrm{d}Q(x) \\
&\quad + \int_{\mathcal{X}^+ \setminus B_{FA}^{G+}} \sum_{j=2}^{k} \left(\max\left\{0, \frac{\sqrt{n}G_{j,n}^{+}(x)}{\overline{\sigma}_{jj,n}^{G+}(x)}\right\}\right)^2 \mathrm{d}Q(x) \\
&= I_n + II_n + III_n.
\end{aligned}$$

Now, I_n diverges to infinity with probability approaching one, while Theorem 1 ensures that II_n and III_n are $O_p(1)$. Thus, S_n^{G+} diverges to infinity. As S_n^{*G+} is $O_{p^*}(1)$, conditional on the sample, the statement follows.
(ii) Note that $S_n^{C^+}$ can be treated exactly as $S_n^{G^+}$.

Proof of Theorem 4:
(i) Define, $S_{\infty,LA}^{\dagger G+}$ as in (3.6), but with the vector $h_{j,A,\infty}^{G+}(x)$ having at least one component strictly bounded away above from zero, and finite, for all $x \in B_{LA}^{G+}$. Let $\mathcal{P}_{n,LA}^{G+}$ denote the set of probabilities under the sequence of local alternatives. We have that for all $a > 0$,

$$\lim \sup_{n\to\infty} \sup_{P \in \mathcal{P}_{n,LA}^{G+}} \left[P\left(S_n^{G+} > a\right) - P\left(S_{\infty,LA}^{\dagger G+} > a\right)\right] = 0,$$

and the distribution of $S_{\infty,LA}^{\dagger G+}$ is continuous at its $(1-\alpha)+\delta$ quintile, for all $0 < \alpha < 1/2$ and $\delta \geq 0$. Also, note that for all $x \in B_{LA}^{G+}$, $\phi_n^{G+} = 0$. The statement then follows by the same argument as that used in the proof of Theorem 2(ii). **(ii)** By the same argument as in part (i).

Proof of Lemma 3: **(i)** Letting $\overline{F}_j(x) = \frac{1}{n}\sum_{t=R}^{T-1} 1\left\{\widehat{e}_{j,t+1} \leq x\right\}$, by an intermediate value expansion,

in the case of a recursive estimation scheme, we have that

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{\widehat{e}_{j,t+1}\leq x\right\}-F_j(x)\right) \\
= \;& \frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right)+\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{\widehat{e}_{j,t+1}\leq x\right\}-1\left\{e_{j,t+1}\leq x\right\}\right) \\
= \;& \frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right)+\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{e_{j,t+1}\leq x-\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\overline{\theta}_{j,t}\right)\left(\widehat{\theta}_{j,t}-\theta_j^{\dagger}\right)\right\}\right.\right. \\
& \left.\left.-F_j\left(x-\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\overline{\theta}_{j,t}\right)\left(\widehat{\theta}_{j,t}-\theta_j^{\dagger}\right)\right)\right)-\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right)\right) \\
& +\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(F_j\left(x-\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\overline{\theta}_{j,t}\right)\left(\widehat{\theta}_{j,t}-\theta_j^{\dagger}\right)\right)-F_j(x)\right) \\
= \;& \frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right)+\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(F_j\left(x-\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\overline{\theta}_{j,t}\right)\left(\widehat{\theta}_{j,t}-\theta_j^{\dagger}\right)\right)-F_j(x)\right) \\
& +o_p(1) \\
= \;& \frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right)-f_j(x)\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\overline{\theta}_{j,t}\right)\left(\widehat{\theta}_{j,t}-\theta_j^{\dagger}\right)+o_p(1) \\
= \;& \frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right) \qquad (8.7) \\
& -f_j(x)\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\overline{\theta}_{j,t}\right)'\frac{1}{t}\sum_{i=1}^{t}\left(\nabla_{\theta_j}^2 m_j(X_i,Z_{j,i-1},\theta_j^{\dagger})\right)^{-1}\left(\nabla_{\theta_j}m_j(X_i,Z_{j,i-1},\theta_j^{\dagger})\right)+o_p(1) \\
= \;& \frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right)-f_j(x)\widehat{A}_j\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{i=1}^{t}\left(\nabla_{\theta_j}m_j(X_i,Z_{j,i-1},\theta_j^{\dagger})\right)+o_p(1)
\end{aligned}
$$

where the $o_p(1)$ term on the RHS of the third equality in (8.7) comes from the fact that

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{e_{j,t+1}\leq x-\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\overline{\theta}_{j,t}\right)\left(\widehat{\theta}_{j,t}-\theta_j^{\dagger}\right)\right\}\right. \\
& \left.-F_j\left(x-\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\overline{\theta}_{j,t}\right)\left(\widehat{\theta}_{j,t}-\theta_j^{\dagger}\right)\right)\right)-\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right)=o_p(1),
\end{aligned}
$$

because of stochastic equicontinuity.

Hence,

$$
\begin{aligned}
& \operatorname{var}\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(\left(1\left\{\widehat{e}_{1,t+1}\leq x\right\}-F_1(x)\right)-\left(1\left\{\widehat{e}_{j,t+1}\leq x\right\}-F_j(x)\right)\right)\right) \\
= \; & \operatorname{var}\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(\left(1\left\{e_{1,t+1}\leq x\right\}-F_1(x)\right)-\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right)\right)\right) \\
& +f_1(x)^2\mathrm{E}\left(\nabla_{\theta_1}\phi_1\left(Z_{1,t+1},\theta_1^{\dagger}\right)\right)'\left(\mathrm{E}\left(\nabla_{\theta_1}^2 m_1(X_i,Z_{1,i-1},\theta_1^{\dagger})\right)\right)^{-1} \\
& \operatorname{var}\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{i=1}^{t}\left(\nabla_{\theta_1}m_1(X_i,Z_{1,i-1},\theta_1^{\dagger})\right)\right) \\
& \left(\mathrm{E}\left(\nabla_{\theta_1}^2 m_1(X_i,Z_{1,i-1},\theta_1^{\dagger})\right)\right)^{-1}\mathrm{E}\left(\nabla_{\theta_1}\phi_1\left(Z_{1,t+1},\theta_1^{\dagger}\right)\right) \\
& +f_j(x)^2\mathrm{E}\left(\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\theta_j^{\dagger}\right)\right)'\left(\mathrm{E}\left(\nabla_{\theta_j}^2 m_j(X_i,Z_{j,i-1},\theta_j^{\dagger})\right)\right)^{-1} \\
& \operatorname{var}\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{i=1}^{t}\left(\nabla_{\theta_j}m_j(X_i,Z_{j,i-1},\theta_j^{\dagger})\right)\right) \\
& \left(\mathrm{E}\left(\nabla_{\theta_j}^2 m_j(X_i,Z_{j,i-1},\theta_j^{\dagger})\right)\right)^{-1}\mathrm{E}\left(\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\theta_j^{\dagger}\right)\right) \\
& -2f_1(x)f_j(x)\mathrm{E}\left(\nabla_{\theta_1}\phi_1\left(Z_{1,t+1},\theta_1^{\dagger}\right)\right)'\left(\mathrm{E}\left(\nabla_{\theta_1}^2 m_j(X_i,Z_{1,i-1},\theta_1^{\dagger})\right)\right)^{-1} \\
& \operatorname{cov}\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{i=1}^{t}\left(\nabla_{\theta_1}m_1(X_i,Z_{1,i-1},\theta_1^{\dagger})\right)\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{i=1}^{t}\left(\nabla_{\theta_j}m_j(X_i,Z_{j,i-1},\theta_j^{\dagger})\right)\right) \\
& \left(\mathrm{E}\left(\nabla_{\theta_j}^2 m_j(X_i,Z_{j,i-1},\theta_j^{\dagger})\right)\right)^{-1}\mathrm{E}\left(\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\theta_j^{\dagger}\right)\right) \\
& +2f_1(x)\mathrm{E}\left(\nabla_{\theta_1}\phi_1\left(Z_{1,t+1},\theta_1^{\dagger}\right)\right)'\left(\mathrm{E}\left(\nabla_{\theta_1}^2 m_1(X_i,Z_{1,i-1},\theta_1^{\dagger})\right)\right)^{-1} \\
& \operatorname{cov}\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(\left(1\left\{e_{1,t+1}\leq x\right\}-F_1(x)\right)-\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right)\right)\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{i=1}^{t}\left(\nabla_{\theta_1}m_1(X_i,Z_{1,i-1},\theta_1^{\dagger})\right)\right) \\
& -2f_j(x)^2\mathrm{E}\left(\nabla_{\theta_j}\phi_j\left(Z_{j,t+1},\theta_j^{\dagger}\right)\right)'\left(\mathrm{E}\left(\nabla_{\theta_j}^2 m_j(X_i,Z_{j,i-1},\theta_j^{\dagger})\right)\right)^{-1} \\
& \operatorname{cov}\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\left(\left(1\left\{e_{1,t+1}\leq x\right\}-F_1(x)\right)-\left(1\left\{e_{j,t+1}\leq x\right\}-F_j(x)\right)\right)\frac{1}{\sqrt{n}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{i=1}^{t}\left(\nabla_{\theta_j}m_j(X_i,Z_{j,i-1},\theta_j^{\dagger})\right)\right)
\end{aligned}
$$

(ii) Recalling (4.6) by a similar argument as in part (i).

Proof of Theorem 5:
Given Lemma 3, the statement follows by the same argument as in Theorem 1.

Proof of Lemma 4:
(i) Note that $\widehat{\operatorname{avar}^*}\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{n-1}\left(\eta_{j,t}^*(x)-\eta_{1,t}^*(x)\right)\right)=\widehat{\sigma}_{jj,n}^{2*G+}(x)$ as defined in (3.1), $\widehat{PEE^*}_{j,t}$ is defined as

$PEE^*_{j,t}$ with E^* replaced by an average, also

$$\begin{aligned}
&\widehat{\mathrm{avar}^*}\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{n-1}\left(\widehat{PEE^*}_{j,t}-\widehat{PEE^*}_{1,t}\right)\right)\\
=\;&\frac{1}{b_n}\sum_{k=1}^{b_n}\left(\frac{1}{l_n^{1/2}}\sum_{i=1}^{l_n}\left(\widehat{PEE^*}_{j,(k-1)l_n'+i}-\widehat{PEE^*}_{j,(k-1)l_n'+i}\right)\right)^2
\end{aligned}$$

and by Theorem 1 in Corradi and Swanson (2007),

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=R}^{n-1}\left(\widehat{PEE^*}_{j,t}-\widehat{PEE^*}_{1,t}\right)\\
=\;&\frac{1}{\sqrt{n}}\sum_{t=R}^{n-1}\left(\widehat{PEE}_{j,t}-\widehat{PEE}_{1,t}\right)+o_p(1)^*.
\end{aligned}$$

$$\begin{aligned}
&\widehat{\mathrm{acov}^*}\left(\frac{1}{\sqrt{n}}\sum_{t=R}^{n-1}\left(u^*_{j,t}(x)-u^*_{1,t}(x)\right),\frac{1}{\sqrt{n}}\sum_{t=R}^{n-1}\left(\widehat{PEE^*}_{j,t}-\widehat{PEE^*}_{1,t}\right)\right)\\
=\;&\frac{1}{b_n}\sum_{k=1}^{b_n}\left(\frac{1}{l_n^{1/2}}\sum_{i=1}^{l_n}\left(\widehat{PEE^*}_{j,(k-1)l_n'+i}-\widehat{PEE^*}_{j,(k-1)l_n'+i}\right)\right.\\
&\left.\frac{1}{l_n^{1/2}}\sum_{i=1}^{l_n}\left(u^*_{j,t}(x)-u^*_{1,t}(x)\right)\right)
\end{aligned}$$

and for $h\to 0$, $nh\to\infty$, $\widehat{f}^*_{j,n,h}(x)=\widehat{f}_{j,n,h}(x)+o_{p^*}(1)=f(x)+o_p(1)+o_{p^*}(1)$. The statement then follow by the same argument as in Lemma 2 and Lemma 3.
(ii) By a similar argument as in Part (i).

Proof of Theorem 6:
(i) By a similar argument as in the proof of Theorem 2 in Corradi and Swanson (2007),

$$\begin{aligned}
\widetilde{S}^{*G+}_n&=\int_{\mathcal{X}^+}\sum_{j=2}^{k}\max\left(\left\{0,\frac{\widetilde{v}^{*G+}_{j,n}(x)-\widetilde{\phi}^{G+}_{j,n}(x)}{\sqrt{\widetilde{\overline{h}}^{*G+}_{2,jj}(x)}}\right\}\right)^2\mathrm{d}Q(x)\\
&=\int_{\mathcal{X}^+}\sum_{j=2}^{k}\max\left(\left\{0,\frac{\widetilde{v}^{G+}_{j,n}(x)-\widetilde{\phi}^{G+}_{j,n}(x)}{\sqrt{\widetilde{\overline{h}}^{G+}_{2,jj}(x)}}\right\}\right)^2\mathrm{d}Q(x)+o_{p^*}(1)
\end{aligned}$$

The statement then follows from Lemma 4 and Theorem 2.
(ii) By a similar argument as in Part (i).

9 References

Aiolfi, M., C. Capistrán, and A. Timmermann (2011). Forecast Combinations. In M.P. Clements and D.F. Hendry (eds.), **Oxford Handbook of Economic Forecasting**, Oxford University Press, Oxford.

Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Robust Covariance Matrix Estimation. *Econometrica,* 59, 817-858.

Andrews, D.W.K. (2011). Similar-on-the-Boundary Tests for Moment Inequalities Exist, but Have Very Poor Power. Cowles Foundation Working Paper 1815R.

Andrews, D.W.K. and D. Pollard (1994). An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes. *International Statistical Review,* 61, 119-132.

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

Andrews, D.W.K. and P. Guggenberger (2010). Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection. *Econometrica,* 78, 119-157.

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.

Andrews, D.W.K. and X. Shi (2013). Inference Based on Conditional Moment Inequalities. *Econometrica,* 81, 609-666.

Andrews, D.W.K. and X. Shi (2017). Inference Based on Many Conditional Moment Inequalities. *Journal of Econometrics,* 196, 275-287.

Bierens H.J. (1982). Consistent Model Specification Tests. *Journal of Econometrics,* 20, 105-134.

Bierens H.J. (1990). A Consistent Conditional Moment Tests for Functional Form. *Econometrica,* 58, 1443-1458.

Coroneo, L., V. Corradi and P. Santos-Monteiro (2017), Testing for Optimal Monetary Policy via Moment Inequalities. *Journal of Applied Econometrics,* forthcoming.

Corradi, V. (1999). Deciding Between $I(0)$ and $I(1)$ via FLIL-based Bounds. *Econometric Theory*, 15, 643-663.

Corradi, V. and N.R. Swanson (2007). Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes. *International Economic Review,* 48, 67-109.

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.R. Swanson (eds.), **Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions, Essays in honor of Halbert L. White, Jr.**, Springer, New York.

Croushore, D. (1993). Introducing: The Survey of Professional Forecasters, The Federal Reserve Bank of Philadelphia Business Review, November-December, 3-15.

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

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

Diebold, F.X. and M. Shin (2017). Assessing Point Forecast Accuracy by Stochastic Error Distance. *Econometric Reviews,* 36, 588-598.

Donald, S.G. and Y.C. Hsu (2016). Improving the Power of Tests for Stochastic Dominance. *Econometric Reviews,* 35, 553-585.

Elliott, G., I. Komunjer and A. Timmermann (2005). Estimation and Testing of Forecast Rationality under Flexible Loss. *Review of Economic Studies,* 72, 1107-1125.

Elliott, G., I. Komunjer and A. Timmermann (2008). Biases in Macroeconomic Forecasts: Irrationality of Asymmetric Loss? *Journal of the European Economic Association,* 6, 122-157.

Fair, R.C. and R.J. Shiller (1990). Comparing Information in Forecasts from Econometric Models. *American Economic Review*, 80, 375-389.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann (2013). Combining the Forecasts in the ECB Survey of Professional Forecasters: Can Anything Beat the Simple Average. *International Journal of Forecasting*, 29, 108-121.

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

Hansen, B.E. (2008). Uniform Convergence Rates for Kernel Estimators with Dependent Data. *Econometric Theory,* 24, 726-748.

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

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics,* 6, 65–70.

Jin, S., V. Corradi and N.R. Swanson (2017). Robust Forecast Comparison. *Econometric Theory,* 33, 1306-1351.

Lahiri, K., H. Peng, and Y. Zhao (2015). Testing the Value of Probability Forecasts for Calibrated Combining. *International Journal of Forecasting*, 31, 113-129.

Lahiri, K., H. Peng, and Y. Zhao (2017). Online Learning and Forecast Combination in Unbalanced Panels. *Econometric Reviews,* 36, 257-288.

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

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

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

Mikusheva, A. (2007). Uniform Inference in Autoregressive Processes. *Econometrica,* 75, 1411-1452.

Peligrad, M. (1998). On the Blockwise Bootstrap for Empirical Processes for Stationary Sequences. *Annals of Probability,* 26, 877-901.

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

Swanson, N.R. and H. White (1997a). A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. *Review of Economics and Statistics*, 79, 1997, 540-550.

Swanson, N.R. and H. White (1997b). Forecasting Economic Time Series Using Adaptive Versus Nonadaptive and Linear Versus Nonlinear Econometric Models. *International Journal of Forecasting,* 13, 1997, 439-461.

Timmermann, A. (2006). Forecast Combinations. In A. Timmermann, C.W.J. Granger, and G. Elliott (eds.), **Handbook of Forecasting Vol. 1.** North Holland, Amsterdam.

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

Zarnowitz, V. and P. Braun, (1992). Twenty-Two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance. In J.H. Stock and M.W. Watson (eds.), **Business Cycles, Indicators, and Forecasting: Studies in Business Cycles, Vol. 28**, University of Chicago Press, Chicago.

Table 1: Monte Carlo Results for JCS_n^{G+}, JCS_n^{G-}, JCS_n^{C+}, and JCS_n^{C-} Forecast Superiority Tests*

	n	$J_n = 0.20$	$J_n = 0.35$	$J_n = 0.50$	$J_n = 0.65$	$J_n = 0.20$	$J_n = 0.35$	$J_n = 0.50$	$J_n = 0.65$
		GL forecast superiority				CL Forecast Superiority			
		Empirical Size							
DGP1	250	0.107	0.101	0.077	0.088	0.096	0.112	0.087	0.097
	500	0.102	0.101	0.086	0.099	0.103	0.121	0.091	0.108
	1000	0.099	0.104	0.101	0.110	0.108	0.100	0.094	0.108
DGP2	250	0.103	0.095	0.099	0.113	0.107	0.112	0.094	0.119
	500	0.104	0.102	0.107	0.104	0.111	0.089	0.100	0.103
	1000	0.112	0.093	0.109	0.104	0.097	0.108	0.106	0.101
DGP3	250	0.080	0.076	0.065	0.072	0.036	0.035	0.032	0.035
	500	0.076	0.071	0.081	0.082	0.030	0.047	0.032	0.041
	1000	0.076	0.065	0.070	0.083	0.043	0.031	0.039	0.051
		Empirical Power							
DGP4	250	0.382	0.392	0.402	0.418	0.639	0.671	0.646	0.633
	500	0.709	0.717	0.719	0.744	0.938	0.944	0.943	0.942
	1000	0.976	0.973	0.973	0.976	0.999	1.000	1.000	1.000
DGP5	250	0.542	0.562	0.521	0.549	0.880	0.908	0.883	0.913
	500	0.838	0.846	0.836	0.848	0.991	0.992	0.995	0.993
	1000	0.996	0.989	0.991	0.994	1.000	1.000	1.000	1.000
DGP6	250	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP7	250	0.600	0.551	0.562	0.541	0.920	0.920	0.905	0.927
	500	0.845	0.872	0.846	0.840	0.991	0.997	0.992	0.992
	1000	0.995	0.996	0.993	0.994	1.000	1.000	1.000	1.000
DGP8	250	1.000	0.998	1.000	0.999	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

* Notes: Entries denote rejection frequencies of (JCS_n^{G+},JCS_n^{G-}) tests (i.e., GL forecast superiority) and (JCS_n^{C+},JCS_n^{C-}) tests (i.e., CL forecast superiority) under a variety of data generating processes denoted by DGP1-DGP8. In DGP1-DGP3, no alternative outperforms the benchmark model. In DGP4-DGP8, at least one alternative model outperfroms the benchmark model. Sample sizes include n=250, 500, and 1000 observations, as indicated in the second column of entries in the table. Nominal test size is 10%, and tests are carried out using critical values constructed for values of J_n including 0.20, 0.35, 0.50, and 0.65. See Section 6 for complete details.

Table 2: Monte Carlo Results for S_n^{G+}, S_n^{G-}, S_n^{C+}, and S_n^{C-} Forecast Superiority Tests*

	n	$\eta = 0.045$	$\eta = 0.060$	$\eta = 0.075$	$\eta = 0.090$	$\eta = 0.045$	$\eta = 0.060$	$\eta = 0.075$	$\eta = 0.090$
		GL forecast superiority				CL Forecast Superiority			
		Empirical Size							
DGP1	250	0.272	0.213	0.145	0.116	0.232	0.185	0.121	0.096
	500	0.251	0.207	0.142	0.101	0.204	0.161	0.114	0.079
	1000	0.263	0.211	0.159	0.096	0.207	0.177	0.130	0.080
DGP2	250	0.277	0.261	0.174	0.115	0.219	0.193	0.131	0.081
	500	0.276	0.238	0.167	0.111	0.230	0.191	0.130	0.091
	1000	0.266	0.245	0.187	0.114	0.203	0.199	0.142	0.094
DGP3	250	0.105	0.083	0.049	0.039	0.065	0.045	0.030	0.014
	500	0.084	0.066	0.029	0.022	0.072	0.045	0.026	0.015
	1000	0.080	0.053	0.026	0.023	0.066	0.038	0.025	0.010
		Empirical Power							
DGP4	250	0.958	0.955	0.908	0.820	0.961	0.968	0.936	0.855
	500	1.000	1.000	0.997	0.992	0.999	1.000	0.998	0.997
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP5	250	0.996	0.991	0.975	0.948	0.997	0.990	0.980	0.964
	500	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.999
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP6	250	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP7	250	0.995	0.995	0.991	0.979	0.994	0.999	0.993	0.976
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP8	250	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

* Notes: Entries denote rejection frequencies of (S_n^{G+},S_n^{G-}) tests (i.e., GL forecast superiority) and (S_n^{C+},S_n^{C-}) tests (i.e., CL forecast superiority) under a variety of data generating processes denoted by DGP1-DGP8. In DGP1-DGP3, no alternative outperforms the benchmark model. In DGP4-DGP8, at least one alternative model outperfroms the benchmark model. Sample sizes include n=250, 500, and 1000 observations, as indicated in the second column of entries in the table. Nominal test size is 10%, and tests are carried out using critical values constructed for values of η including 0.045, 0.060, 0.075, and 0.090. See Section 5 for complete details.

Table 3: Forecast Superiority Test Results Part I: Do Years of Experience Matter for Mean SPF Forecasting of GDP Growth Rates?*

Test	S_n^G		S_n^C		JCS_n^G		JCN_n^C	
	S_n^{G+}	S_n^{G-}	S_n^{C+}	S_n^{C-}	JCS_n^{G+}	JCS_n^{G-}	JCS_n^{C+}	JCS_n^{C-}
Panel A : *Forecast Horizon* : h=0								
Statistic Values	0.3186	0.1979	0.3972	0.1735	0.2268	0.0756	0.0003	0.0002
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel B : *Forecast Horizon* : h=1								
Statistics Values	0.8773	0.4182	0.8726	0.1594	0.2274	0.3032	0.0011	0.0002
Rejection with p-Value Type 1	yes		yes		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel C : *Forecast Horizon* : h=2								
Statistics Values	0.1933	0.9600	0.0333	1.0000	0.8767	0.9833	0.2900	0.9567
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel D : *Forecast Horizon* : h=3								
Statistics Values	0.2733	0.2133	0.2100	0.7133	0.6700	0.8200	0.5600	0.8233
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel E : *Forecast Horizon* : h=4								
Statistics Values	0.4092	0.2271	0.5519	0.0784	0.3881	0.0776	0.0007	0.0002
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	

Panel F : *Root Mean Square Forecast Errors*

	Benchmark	Alt Model 1	Alt Model 2	Alt Model 3
$h = 0$	0.00741	0.00740	0.00738	0.00742
$h = 1$	0.01248	0.01246	0.01243	0.01250
$h = 2$	0.01682	0.01681	0.01689	0.01694
$h = 3$	0.02088	0.02087	0.02103	0.02112
$h = 4$	0.02449	0.02447	0.02461	0.02528

* Notes: Numerical entries in Panels A-E of this table are forecast superiority test statistics, for S_n^{G+}, S_n^{G-}, S_n^{C+}, S_n^{C-}, JCS_n^{G+}, JCS_n^{G-}, JCS_n^{C+}, and JCS_n^{C-} tests. The benchmark model is the arithmetic mean of all survey participants. Three alternative models are considered, including the mean of all participants with each of 1 year, 2 years, and 3 years of experience. These are called Alt Models 1, 2, and 3, respectively, in Panel F of the table. Test critical values are calculated using $\eta = 0.075$ and 0.090 (for S_n^{G+}, S_n^{G-}, S_n^{C+}, S_n^{C-} tests), or using $J_n = 0.20$ and 0.35 (for JCS_n^{G+}, JCS_n^{G-}, JCS_n^{C+}, and JCS_n^{C-} tests). Nominal test size is 10%. Test outcomes are reported under the heading "p-Value Type 1" (for $\eta = 0.075$ and $J_n = 0.20$), and "p-Value Type 2" (for $\eta = 0.090$ and $J_n = 0.35$). "No" denotes failure to reject the null hypothesis, while "yes" denotes rejection of the null. Results are reported for forecast horizons h=0 to h=4 (h=0 denotes "nowcasts"). Additionally, in Panel F, root mean square forecast errors for all models (i.e., the benchmark and each of the three alternative models) are reported, for each forecast horizon. See Section 6 for complete details.

Table 4: Forecast Superiority Test Results Part II: Do Years of Experience Matter for Median SPF Forecasting of GDP Growth Rates?*

Test	S_n^G		S_n^C		JCS_n^G		JCN_n^C	
	S_n^{G+}	S_n^{G-}	S_n^{C+}	S_n^{C-}	JCS_n^{G+}	JCS_n^{G-}	JCS_n^{C+}	JCS_n^{C-}
Panel A : *Forecast Horizon* : h=0								
Statistic Values	0.3615	0.1597	0.2546	0.0117	0.1512	0.1512	0.0003	0.0000
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel B : *Forecast Horizon* : h=1								
Statistics Values	0.4769	0.1954	1.4752	0.1506	0.3032	0.0758	0.0017	0.0002
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel C : *Forecast Horizon* : h=2								
Statistics Values	0.5554	0.0000	0.7708	0.0000	0.3041	0.0000	0.0012	-0.0001
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel D : *Forecast Horizon* : h=3								
Statistics Values	0.5121	0.2350	0.0000	0.3050	0.1525	0.0018	0.0000	
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel E : *Forecast Horizon* : h=4								
Statistics Values	0.4827	0.0757	0.1712	0.1222	0.3881	0.1552	0.0016	0.0002
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	

Panel F : *Root Mean Square Forecast Errors*				
	Benchmark	Alt Model 1	Alt Model 2	Alt Model 3
$h = 0$	0.00745	0.00745	0.00746	0.00749
$h = 1$	0.01249	0.01249	0.01249	0.01241
$h = 2$	0.01698	0.01706	0.01699	0.01712
$h = 3$	0.02096	0.02102	0.02106	0.02117
$h = 4$	0.02470	0.02478	0.02473	0.02551

* Notes: See notes to Table 3. The benchmark model is the median of all survey participants. Three alternative models are considered, including the median of all participants with each of 1 year, 2 years, and 3 years of experience.

Table 5: Forecast Superiority Test Results Part III: Should We Use Only the Top Performer for Mean SPF Forecasting of GDP Growth Rates?*

Test	S_n^G		S_n^C		JCS_n^G		JCN_n^C	
	S_n^{G+}	S_n^{G-}	S_n^{C+}	S_n^{C-}	JCS_n^{G+}	JCS_n^{G-}	JCS_n^{C+}	JCS_n^{C-}
Panel A : *Forecast Horizon* : h=0								
Statistic Values	0.6206	0.0460	0.1788	0.0059	0.6047	0.1512	0.0014	0.0001
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel B : *Forecast Horizon* : h=1								
Statistics Values	1.3935	0.0000	1.6737	0.0000	0.7581	0.0000	0.0068	-0.0002
Rejection with p-Value Type 1	yes		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel C : *Forecast Horizon* : h=2								
Statistics Values	0.7014	0.0000	0.1704	0.0000	0.9123	0.0000	0.0051	-0.0004
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel D : *Forecast Horizon* : h=3								
Statistics Values	0.5410	0.1039	0.3500	0.0026	0.6862	0.5337	0.0080	0.0018
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel E : *Forecast Horizon* : h=4								
Statistics Values	1.4847	0.1084	0.6600	0.0000	1.3038	0.3835	0.0150	-0.0004
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	yes		yes		no		no	

Panel F : *Root Mean Square Forecast Errors*				
	Benchmark	Alt Model 1	Alt Model 2	Alt Model 3
$h = 0$	0.00741	0.00882	0.00759	0.00764
$h = 1$	0.01248	0.01326	0.01300	0.01312
$h = 2$	0.01682	0.01776	0.01889	0.01858
$h = 3$	0.02088	0.02166	0.02192	0.02317
$h = 4$	0.02461	0.02618	0.02808	0.02620

* Notes: See notes to Table 3. The benchmark model is the arithmetic mean of all survey participants. Three alternative models are considered, including the top performer based on the comparison of the mean of absolute forecast errors for all participants with each of 1 year, 2 years, and 3 years of experience.

Table 6: Forecast Superiority Test Results Part IV: Should We Use Only the Top Three Performers for Mean SPF Forecasting of GDP Growth Rates?*

Test	S_n^G		S_n^C		JCS_n^G		JCN_n^C	
	S_n^{G+}	S_n^{G-}	S_n^{C+}	S_n^{C-}	JCS_n^{G+}	JCS_n^{G-}	JCS_n^{C+}	JCS_n^{C-}
Panel A : *Forecast Horizon* : h=0								
Statistic Values	1.5834	0.0869	2.3071	0.0211	1.1959	0.0104	0.6412	0.0009
Rejection with p-Value Type 1	yes		yes		no		no	
Rejection with p-Value Type 2	no		yes		no		no	
Panel B : *Forecast Horizon* : h=1								
Statistics Values	3.1392	0.0362	1.7618	0.0956	0.9855	0.0758	0.0073	0.0002
Rejection with p-Value Type 1	yes		yes		no		no	
Rejection with p-Value Type 2	yes		yes		yes		yes	
Panel C : *Forecast Horizon* : h=2								
Statistics Values	1.6563	0.1096	4.0753	0.0171	0.7603	0.3801	0.0098	0.0013
Rejection with p-Value Type 1	yes		yes		yes		no	
Rejection with p-Value Type 2	no		no		no		yes	
Panel D : *Forecast Horizon* : h=3								
Statistics Values	1.3145	0.5176	1.1122	0.2517	0.8387	0.6862	0.0106	0.0071
Rejection with p-Value Type 1	yes		yes		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel E : *Forecast Horizon* : h=4								
Statistics Values	2.3830	0.3138	4.5114	0.0267	1.1504	0.5369	0.0210	0.0045
Rejection with p-Value Type 1	yes		yes		yes		yes	
Rejection with p-Value Type 2	yes		no		yes		yes	

Panel F : *Root Mean Square Forecast Errors*				
	Benchmark	Alt Model 1	Alt Model 2	Alt Model 3
$h = 0$	0.00741	0.00751	0.00719	0.00720
$h = 1$	0.01248	0.01249	0.01287	0.01260
$h = 2$	0.01682	0.01593	0.01705	0.01709
$h = 3$	0.02088	0.01994	0.02199	0.02152
$h = 4$	0.02461	0.02325	0.02426	0.02465

* Notes: See notes to Table 3. The benchmark model is the arithmetic mean of all survey participants. Three alternative models are considered, including the top three performers based on the comparison of the mean of absolute forecast errors for all participants with each of 1 year, 2 years, and 3 years of experience.

Table 7: Forecast Superiority Test Results Part V: Should We Use Only the Top 10% of Performers for Mean SPF Forecasting of GDP Growth Rates?*

Test	S_n^G		S_n^C		JCS_n^G		JCN_n^C	
	S_n^{G+}	S_n^{G-}	S_n^{C+}	S_n^{C-}	JCS_n^{G+}	JCS_n^{G-}	JCS_n^{C+}	JCS_n^{C-}
Panel A : *Forecast Horizon* : h=0								
Statistic Values	1.5543	0.0804	1.1251	0.0551	0.8315	0.1512	0.0029	0.0006
Rejection with p-Value Type 1	yes		yes		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel B : *Forecast Horizon* : h=1								
Statistics Values	2.9458	0.0154	1.6952	0.0000	0.9097	0.1516	0.0084	-0.0001
Rejection with p-Value Type 1	yes		yes		no		no	
Rejection with p-Value Type 2	yes		yes		yes		yes	
Panel C : *Forecast Horizon* : h=2								
Statistics Values	1.5726	0.0292	2.5697	0.0000	0.6843	0.2281	0.0081	-0.0003
Rejection with p-Value Type 1	yes		yes		yes		yes	
Rejection with p-Value Type 2	no		no		no		no	
Panel D : *Forecast Horizon* : h=3								
Statistics Values	1.0083	0.6593	0.8208	0.5017	0.6100	0.8387	0.0080	0.0114
Rejection with p-Value Type 1	no		no		no		no	
Rejection with p-Value Type 2	no		no		no		no	
Panel E : *Forecast Horizon* : h=4								
Statistics Values	2.1550	0.3916	2.7560	0.0479	1.4572	0.5369	0.0229	0.0061
Rejection with p-Value Type 1	yes		yes		yes		yes	
Rejection with p-Value Type 2	yes		yes		yes		yes	

Panel F : *Root Mean Square Forecast Errors*				
	Benchmark	Alt Model 1	Alt Model 2	Alt Model 3
$h = 0$	0.00741	0.00784	0.00708	0.00738
$h = 1$	0.01248	0.01225	0.01279	0.01304
$h = 2$	0.01682	0.01623	0.01758	0.01776
$h = 3$	0.02088	0.01973	0.02168	0.02203
$h = 4$	0.02461	0.02297	0.02618	0.02531

* Notes: See notes to Table 3. The benchmark model is the arithmetic mean of all survey participants. Three alternative models are considered, including the top 10% of performers based on the comparison of the mean of absolute forecast errors for all participants with each of 1 year, 2 years, and 3 years of experience.

Table 8: Forecast Superiority Test Results Part VI: Should We Use Only the Top 25% of Performers for Mean SPF Forecasting of GDP Growth Rates?*

Test	S_n^G		S_n^C		JCS_n^G		JCN_n^C	
	S_n^{G+}	S_n^{G-}	S_n^{C+}	S_n^{C-}	JCS_n^{G+}	JCS_n^{G-}	JCS_n^{C+}	JCS_n^{C-}
Panel A : *Forecast Horizon* : h=0								
Statistic Values	1.4563	0.1437	1.6961	0.0000	0.8315	0.2268	0.0026	0.0000
Rejection with p-Value Type 1	yes		no		yes		no	
Rejection with p-Value Type 2	no		yes		no		no	
Panel B : *Forecast Horizon* : h=1								
Statistics Values	1.9542	0.0951	1.0744	0.2393	0.5307	0.1516	0.0058	0.0003
Rejection with p-Value Type 1	yes		yes		no		no	
Rejection with p-Value Type 2	no		no		yes		yes	
Panel C : *Forecast Horizon* : h=2								
Statistics Values	2.0036	0.2440	4.3577	0.3473	0.7603	0.3801	0.0090	0.0025
Rejection with p-Value Type 1	yes		yes		yes		yes	
Rejection with p-Value Type 2	no		no		yes		yes	
Panel D : *Forecast Horizon* : h=3								
Statistics Values	2.5623	0.2940	4.2772	0.1036	0.9150	0.3812	0.0142	0.0034
Rejection with p-Value Type 1	yes		yes		yes		yes	
Rejection with p-Value Type 2	yes		yes		yes		yes	
Panel E : *Forecast Horizon* : h=4								
Statistics Values	3.4393	0.4863	6.6957	0.0565	1.0738	0.4602	0.0193	0.0054
Rejection with p-Value Type 1	yes		yes		yes		yes	
Rejection with p-Value Type 2	yes		yes		yes		yes	

Panel F : *Root Mean Square Forecast Errors*				
	Benchmark	Alt Model 1	Alt Model 2	Alt Model 3
$h = 0$	0.00741	0.00736	0.00733	0.00731
$h = 1$	0.01248	0.01228	0.01267	0.01257
$h = 2$	0.01682	0.01592	0.01694	0.01728
$h = 3$	0.02088	0.01969	0.02127	0.02109
$h = 4$	0.02461	0.02343	0.02467	0.02444

* Notes: See notes to Table 3. The benchmark model is the arithmetic mean of all survey participants. Three alternative models are considered, including the top 25% of performers based on the comparison of the mean of absolute forecast errors for all participants with each of 1 year, 2 years, and 3 years of experience.