Australian National University

Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

On bootstrapping tests of equal forecast accuracy for nested models

CAMA Working Paper 27/2020
March 2020

Firmin Doko Tchatoka
School of Economics, University of Adelaide

Qazi Haque
Economics Department, University of Western Australia
Centre for Applied Macroeconomic Analysis, ANU

Abstract

The asymptotic distributions of the recursive out-of-sample forecast accuracy test statistics depend on stochastic integrals of Brownian motion when the models under comparison are nested. This often complicates their implementation in practice because the computation of their asymptotic critical values is costly. Hansen and Timmermann (2015, Econometrica) propose a Wald approximation of the commonly used recursive F-statistic and provide a simple characterization of the exact density of its asymptotic distribution. However, this characterization holds only when the larger model has one extra predictor or the forecast errors are homoscedastic. No such closed-form characterization is readily available when the nesting involves more than one predictor and heteroskedasticity is present. We first show both the recursive F-test and its Wald approximation have poor finite-sample properties, especially when the forecast horizon is greater than one. We then propose a hybrid bootstrap method consisting of a block moving bootstrap (which is nonparametric) and a residual based bootstrap for both statistics, and establish its validity. Simulations show that our hybrid bootstrap has good finite-sample performance, even in multi-step ahead forecasts with heteroscedastic or autocorrelated errors, and more than one predictor. The bootstrap method is illustrated on forecasting core inflation and GDP growth.

Keywords

Out-of-sample forecasts, HAC estimator, Moving block bootstrap, Bootstrap consistency.

JEL Classification

C12, C15, C32

Address for correspondence:

(E) cama.admin@anu.edu.au

ISSN 2206-0332

On bootstrapping tests of equal forecast accuracy for nested models *

Firmin Doko Tchatoka[†] Qazi Haque[‡]

ABSTRACT

The asymptotic distributions of the recursive out-of-sample forecast accuracy test statistics depend on stochastic integrals of Brownian motion when the models under comparison are nested. This often complicates their implementation in practice because the computation of their asymptotic critical values is costly. Hansen and Timmermann (2015, Econometrica) propose a Wald approximation of the commonly used recursive F-statistic and provide a simple characterization of the exact density of its asymptotic distribution. However, this characterization holds only when the larger model has one extra predictor or the forecast errors are homoscedastic. No such closed-form characterization is readily available when the nesting involves more than one predictor and heteroskedasticity is present. We first show both the recursive F-test and its Wald approximation have poor finite-sample properties, especially when the forecast horizon is greater than one. We then propose an hybrid bootstrap method consisting of a block moving bootstrap (which is nonparametric) and a residual based bootstrap for both statistics, and establish its validity. Simulations show that our hybrid bootstrap has good finite-sample performance, even in multi-step ahead forecasts with heteroscedastic or autocorrelated errors, and more than one predictor. The bootstrap method is illustrated on forecasting core inflation and GDP growth.

Key words: Out-of-sample forecasts; HAC estimator; Moving block bootstrap; Bootstrap consistency.

JEL classification: C12; C15; C32.

* The authors thank Prosper Dovonon, James A. Duffy, Jean-Marie Dufour, Mardi Dungey, Leandro Magnusson, Sophocles Mavroeidis, Adrian Pagan, Peter C.B. Phillips, Richard Smith, Mark Weder, and the participants of the 27th NZESG conference at the Otago Business School, Dunedin, 2-3 February 2017. This work was supported with super computing resources provided by the Phoenix HPC service at The University of Adelaide. Doko Tchatoka Acknowledges the financial support from the Australian Research Council through the Discovery Grant DP200101498. Haque acknowledges financial support from the Australian Research Council, under the grant DP170100697, and would like to thank University of Oxford for their hospitality.

[†]Corresponding author contacts: School of Economics, The University of Adelaide, 10 Pulteney St, Adelaide SA 5005, AUSTRALIA. Tel:+618 8313 1174, Fax:+618 8223 1460; e-mail: firmin.dokotchatoka@adelaide.edu.au

[‡]Economics Department, 35 Stirling Highway M251, Business School, The University of Western Australia, Crawley, WA 6009, Australia and Centre for Applied Macroeconomic Analysis, Australian National University, email: qazi.haque@uwa.edu.au

1 Introduction

Out-of-sample tests of predictive accuracy have received considerable attention in the literature.[1] Such testing procedures often involve comparing the out-of-sample mean squared forecast error (MSFE) of alternative models to select the one that minimizes this criterion. The case of nested models is particularly interesting because the test-statistics often used– such as the recursively generated F-statistic [McCracken (2007) and Clark and McCracken (2001, 2005)]– have nonstandard asymptotic distributions that depend on stochastic integrals of Brownian motion; see Clark and McCracken (2012, 2014, 2015), and Hansen and Timmermann (2015). Many studies have developed methods for approximating the quantiles of the limiting distributions of these statistics, mainly by using simulations methods; see Rossi and Inoue (2012) and Hansen and Timmermann (2012). However, these simulation methods can be computationally burdensome, especially in the multivariate setting, because it requires a discretization of both the underlying (multivariate) Brownian motion and the support of the nuisance parameters (such as the relative size of the initial estimation sample versus the out-of-sample evaluation period).

Recently, Hansen and Timmermann (2015) show that the recursively generated F-statistic of McCracken (2007) can be approximated by a Wald-type statistic whose asymptotic distribution is a convolution of dependent $\chi^2(1)$-distributed random variables, thus simplifying the computation of test critical values. When the underlying data generating process (DGP) is homoscedastic, their characterization yields a closed-form expression of the exact density of the limiting distribution of the F-statistic, even when the number of extra predictors in the larger model is greater than one; see Hansen and Timmermann (2015, Theorem 5). However, no closed-form characterization of the density of the limiting distribution of this statistic is available in the multivariate setting (i.e., when there are more than one extra predictor in the larger model) if the underlying DGP is heteroscedastic or serially correlated.

This paper contributes to this research area in two main ways. First, we show through Monte Carlo simulations that even for moderate sample sizes, both the recursively generated F-test of McCracken (2007) and its Wald approximation of Hansen and Timmermann (2015) are often oversized, especially when the forecast errors exhibit heteroscedasticity or serial correlation. The size distortions of both tests increase with the forecast horizon. For example, in a simple framework where there is only one extra predictor in the larger model, our simulations show that under serially correlated forecast errors, the rejection frequencies under the null hypothesis of the F-test (at the 5% nominal level) can jump from 10.6% when $T = 50$, 7.5% when $T = 100$,

[1] See Diebold and Mariano (1995), West (1996), White (2000), Stock and Watson (2003), Giacomini and White (2006), Corradi and Swanson (2007), McCracken (2007), Clark and McCracken (2001, 2005, 2012, 2014, 2015), Rossi and Inoue (2012), Hansen and Timmermann (2012), among others.

and 5.9% when $T = 500$ for a *1-period ahead* forecast to 27.6% when $T = 50$, 17.4% when $T = 100$, and 11.6% when $T = 500$ for a *4-period ahead* forecast.

Second, we propose an hybrid bootstrap method consisting of a block moving bootstrap (henceforth MBB, which is nonparametric) and a residual based bootstrap (which is parametric) for both the recursively generated F-test and its equivalent Wald statistic.[2] Our bootstrap method builds on earlier work by Corradi and Swanson (2007) but it differs from theirs in two important aspects. First, while Corradi and Swanson (2007) (henceforth CS) bootstrap is purely nonparametric in the sense that level data are re-sampled (pairs bootstrap), ours is semi-parametric and is based on resampling the residuals of the restricted regression that excludes the extra predictors (i.e., the null hypothesis is imposed in our bootstrap DGP). Re-sampling the residuals is paramount to recovering an eventual pattern of serial correlation in the regression errors, which is not always the case with the *pairs bootstrap*. Second, CS establish the conditions on the block length under which their MBB is consistent but there is no practical guidance on the (optimal) choice of this block length in their study. Their Monte Carlo experiments [see Tables 2-3 in Corradi and Swanson (2007)] provide a clear evidence on the importance of choosing the block length that fits the data better, as the performance of the bootstrap CS test varies largely across alternative choices of block lengths. In this paper, we suggest a data dependent approach to select the block length. Specifically, we propose setting it equal to the optimal lag length of the Newey and West's (1987) HAC estimator used in the expressions of the statistics. As the choice of the block length aims to capture the dependence structure of the data, we believe matching it to the optimal lag length of the HAC estimator is reasonable. Note, however, that we do not claim optimality of this choice, for example in the sense of maximizing test power. Rather, we follow Andrews and Monahan (1992) and the recommendations of Newey and West (1994) to select the kernel bandwidth of the HAC estimator and then use it as the block length in our bootstrap DGP. This choice satisfies the condition under which our bootstrap consistency is established, thus guaranteeing that type I error is controlled for. From this perspective, our bootstrap method can be viewed as complementary to Corradi and Swanson (2007).

We show that our proposed bootstrap is consistent under both the null hypothesis of equal forecast accuracy and the alternative hypothesis, irrespective of the forecast horizon and the underlying DGP exhibiting heteroskedasticity or serial correlation. The proof of our bootstrap is innovative and different from the one in CS. Indeed, due to nesting, the standard Gaussian approximation used in CS no longer holds, so one has to resort to the functional central limit theorem; see Davidson (1994). We present simulation evidence indicating that the bootstrap approximation performs well in small samples, even with heteroscedastic or serially correlated errors. These results are qualitatively the same across forecast horizons, confirming our theoretical

[2]See Kunsch (1989).

findings. We illustrate our theoretical results with empirical applications which look at forecasting core inflation and GDP growth.

Important contributions on residual MBB are Efron (1982, pp.35-36) and Fitzenberger (1998) but their bootstrap schemes assume the regressors to be strictly exogenous, therefore are kept fixed (not re-sampled) in the bootstrap algorithm. For weakly dependent time series with lagged dependent variables, as is the case in most applications of out-of-smple tests of equal forecast accuracy, this type of MBB will not have desired size property. Other recent contributions on bootstrapping out-of-sample tests of equal forecast accuracy include Clark and McCracken (2012, 2014, 2015). Their "wild" bootstrap algorithms also rely on the assumption that the regressors are fixed (thus is often referred to as "fixed regressor wild bootstrap (FRWB)"). Such FRWB often fails to control the size for multistep forecasts, for example, see Clark and McCracken (2012, Table 2, DGP2) and Clark and McCracken (2015, Tables 1 & 3, DGP5).

Throughout this paper, convergence almost surely is symbolized by "$a.s.$", "$\xrightarrow{p}$" stands for convergence in probability, while "$\xrightarrow{d}$" means convergence in distribution. The usual stochastic orders of magnitude are denoted by $O_p(\cdot)$ and $o_p(\cdot)$. $\mathbb{P}$ denotes the relevant probability measure and $\mathbb{E}$ is the expectation operator under $\mathbb{P}$. The "$*$" on all these symbols and other variables (for example $\mathbb{P}^*$) indicates the bootstrap world. $o_p*(1)$-$\mathbb{P}$ denotes a term converging to zero in $\mathbb{P}^*$-probability, conditional on the sample, and for all samples except a subset with probability measure approaching zero, and $O_p*(1)$-$\mathbb{P}$ is for a term that is bounded in $\mathbb{P}^*$-probability, conditional on the sample, and for all samples except a subset with probability measure approaching zero. Similarly, $o_{a.s^*}(1)$ and $O_{a.s^*}(1)$ denote the terms that approach zero almost surely and the terms that are almost surely bounded, according to the probability law $\mathbb{P}^*$, and conditional on the sample. The notation I_q stands for the identity matrix of order q, and $\|U\|$ denotes the usual Euclidian or Frobenius norm for a matrix U. Finally, $\sup_{\omega\in\Omega}|f(\omega)|$ is the supremum norm on the space of bounded continuous real functions, with topological space Ω.

The remainder of the paper is organised as follows. Section 2 presents the setup, formulates the null hypothesis as well as the assumptions used, and summarizes briefly the asymptotic properties of the tests studied. Section 3 presents our proposed bootstrap method, proves the validity of the bootstrap, and presents Monte Carlo results on the finite-sample performance of our proposed bootstrap compared to the FRWB of Clark and McCracken (2012, 2014, 2015). Section 4 applies our bootstrap test to forecasts of core inflation and real GDP growth in the US. Finally, Section 5 concludes.

2 Framework

We first introduce the setup and the testing problem of interest in Sections 2.1 and 2.2. The asymptotic properties of the test statistics are studied in Section 2.3.

2.1 Setup

Let $\{Y_t:\ 1 \leq t \leq T\}$ be a stochastic process defined on $(\Omega, \mathscr{B}, \mathcal{F})$, where $\mathscr{B}$ is a σ-algebra on Ω, $\mathcal{F}$ is the class of distributions under consideration, and Y_t has support on a compact subset of $\mathbb{R}^p$ for some positive integer p. Consider the partition $Y_t := (y_t, X_{2t}')'$, where $y_t : 1 \times 1$ and $X_{2t} : k \times 1$ $(k = p-1)$ may contains lags of y_t. By convention, we assume that a vector (or a matrix) does not appear in the model if its number of columns or rows is zero. For example, X_{2t} does not appear in the above partition of Y_t if $p = 1$. Let $s = \max\{q, \tau\} + 1$, where q denotes the maximum lag length of the variables in X_{2t} and $\tau \geq 1$ is the forecasts horizon of interest.

Consider the predictive regression model (see Hansen and Timmermann, 2015)

$$\begin{aligned} y_t &= X_{2,t-\tau}'\beta_2 + \varepsilon_{2t} \\ &= X_{1,t-\tau}'\beta_{21} + \widetilde{X}_{2,t-\tau}'\beta_{22} + \varepsilon_{2t},\ t = s, \ldots, T, \end{aligned} \tag{2.1}$$

where $X_{2t} = (X_{1,t}', \widetilde{X}_{2,t}')'$ is such that $X_{1,t} \in \mathbb{R}^{k_1}$, $\widetilde{X}_{2,t} \in \mathbb{R}^{k_2}$ $(k = k_1 + k_2)$; $\beta_2 = (\beta_{21}', \beta_{22}')' \in \mathbb{R}^k : \beta_{21} \in \mathbb{R}^{k_1}$ and $\beta_{22} \in \mathbb{R}^{k_2}$ are unknown parameter vectors; and ε_{2t} is an error term. We are interested in testing whether $\widetilde{X}_{2t}$ has predictive power in forecasting y_t at τ-periods ahead. This problem is often assessed by comparing the mean squared error (MSE) of the forecast of $y_{t+\tau}$ generated using the unrestricted model (2.1) to the one resulting from the restricted regression:

$$y_t = X_{1,t-\tau}'\beta_1 + \varepsilon_{1t},\ t = s, \ldots, T. \tag{2.2}$$

Formally, let $\beta_j^0 = \arg\min_{\beta_j} \mathbb{E}_F[(y_t - X_{j,t-\tau}'\beta_j)^2]$ denote the unknown true values of β_j $(j = 1, 2)$ in (2.2) and (2.1) respectively, for some $F \in \mathcal{F}$. The null hypothesis of equal predictive performance under the MSE loss function takes the form:

$$H_0:\ \mathbb{E}_F[(y_t - X_{2,t-\tau}'\beta_2^0)^2 - (y_t - X_{1,t-\tau}'\beta_1^0)^2] = 0 \tag{2.3}$$

for some $F \in \mathcal{F}$, where β_j^0 $(j = 1, 2)$ are defined above. The form of H_0 in (2.3) suggests building test statistics based on the MSE loss differential $\mathbb{E}_F[(y_t - X_{2,t-\tau}'\beta_2)^2 - (y_t - X_{1,t-\tau}'\beta_1)^2]$. This is usually done out-of-sample using a recursive estimation of the model parameters; see e.g. Diebold and Mariano (1995); West (1996); Clark and McCracken (2001). In the nested framework (2.1), the test statistics suggested in the above studies have limiting distributions that depend on stochastic Brownian motions, which often makes the computation of their critical values cumbersome.

In this study, we focus particularly on the recursively generated F-test of Clark and McCracken (2001), whose critical values are easier to compute in some cases due to its equivalence to a Wald-type statistic (Hansen and Timmermann, 2015). In particular, when the DGP is homoscedastic, Hansen and Timmermann (2015) provide closed-form expressions of the exact density of the limiting distributions of this F-statistic, even when the unrestricted model (2.1) contains more than one extra predictor. However, no such closed-form characterizations are available when the underlying DGP is heteroscedastic and (2.1) includes multiple extra predictors. This setting is relevant in empirical work and providing a valid statistical procedure that accounts for it is of great interest to applied researchers.

To introduce the recursively generated F-statistic, suppose that P_T out-of-sample predictions are available, where the first is based on a parameter vector estimated using data from s to R_T, the second on a parameter vector estimated using data from s to R_T+1, ..., and the last is based on a parameter vector estimated using data from s to $R_T+P_T-1 \equiv T$ (i.e., the full sample). Let $\hat{y}_{t+\tau|t}(\hat{\beta}_{2,t}) := \hat{y}_{t+\tau|t} = X'_{2t}\hat{\beta}_{2t}$ denote the τ-step ahead forecast generated from model (2.1) and $\tilde{y}_{t+\tau|t}(\hat{\beta}_{1,t}) := \tilde{y}_{t+\tau|t} = X'_{1,t}\hat{\beta}_{1,t}$ be the one that results from model (2.2), where $\hat{\beta}_{j,t}$ $(j=1,2)$ are the recursive OLS estimators of β_j from (2.1)-(2.2), i.e.

$$\hat{\beta}_{j,t} = \arg\min_{\beta_j} \frac{1}{t}\sum_{n=s}^{t}(y_n - X'_{j,n-\tau}\beta_j)^2, \quad R_T \le t \le T; j=1,2. \tag{2.4}$$

The recursively generated F-statistic for H_0 (see Hansen and Timmermann, 2015) takes the form

$$\mathscr{T}_T = \frac{1}{\hat{\sigma}^2_{\varepsilon}}\sum_{t=R_T}^{T}\left[(y_t - X'_{2,t-\tau}\hat{\beta}_{2,t})^2 - (y_t - X'_{1,t-\tau}\hat{\beta}_{1,t})^2\right], \tag{2.5}$$

where $\hat{\sigma}^2_{\varepsilon}$ is a consistent estimator of the variance of the unrestricted error in (2.1).[3]

Let $H_2 = \underset{T\to\infty}{p\lim}\big(\frac{1}{T}\sum_{t=s}^{T} X_{2,t-\tau}X'_{2,t-\tau}\big)$ (assuming that the limit exists and also X_2 includes a column vector of ones) be partitioned as:

$$H_2 = \begin{bmatrix} H_1 & H'_{21} \\ H_{21} & \widetilde{H}_2 \end{bmatrix}; \;\; H_1 : k_1 \times k_1, \;\; H_{21} : k_2 \times k_1, \;\; H_2 : k_2 \times k_2,$$

and define $\check{H}_2 = \widetilde{H}_2 - H_{21}H_1^{-1}H'_{21}$, $Z_{t-\tau} = \widetilde{X}_{2,t-\tau} - H_{21}H_1^{-1}X_{1,t-\tau}$. Also, let $\check{\Gamma}_n$ denote the nth autocovariance (suppose for now that it exists) of the stochastic process

[3]The HAC estimator with the Bartlett kernel is utilized in the simulations and empirical applications, but any kernel in class $\mathscr{K}_3$ of Andrews (1991, Eq.(7.1)) could be employed. The block length of the kernel bandwidth is selected following the recommendations of Andrews and Monahan (1992) and Newey and West (1994).

$\{Z_{t-\tau}\varepsilon_{2t}\}$, i.e.

$$\check{\Gamma}_n = p\lim_{T\to\infty}\frac{1}{T}\sum_{t=s}^{T} Z_{t-\tau}\varepsilon_{2t}\varepsilon'_{2,t-n}Z'_{t-\tau-n},$$

and define $\check{\Omega} = \sum_{n=-\tau+1}^{\tau-1}\check{\Gamma}_n$. Let $Z_{T,t-\tau}$ denote the residual from the multivariate regression of $\widetilde{X}_{2,t-\tau}$ on $X_{1,t-\tau}$, i.e.

$$Z_{T,t-\tau} = \widetilde{X}_{2,t-\tau} - \sum_{t=s}^{T}\widetilde{X}_{2,t-\tau}X'_{1,t-\tau}\left(\sum_{t=s}^{T}X_{1,t-\tau}X'_{1,t-\tau}\right)^{-1}X_{1,t-\tau}.$$

The Wald statistic for the null hypothesis $\beta_{22} = 0$ in (2.1) is given by

$$\widehat{S}_T = T\hat{\beta}'_{22\cdot T}\widehat{V}_T^{-1}\hat{\beta}_{22\cdot T}, \tag{2.6}$$

where $\hat{\beta}_{22\cdot T} = \left(\sum_{t=s}^{T} Z_{T,t-\tau}Z'_{T,t-\tau}\right)^{-1}\sum_{t=s}^{T} Z_{T,t-\tau}y_t$ and $\widehat{V}_T \equiv \widehat{V}_T(\hat{\beta}_{22\cdot T})$ is a consistent estimator of the variance of $\lim_{T\to\infty} var(\sqrt{T}\hat{\beta}_{22\cdot T})$. If the errors are homoskedastic, we have $\widehat{V}_T = \hat{\sigma}^2_\varepsilon\left(\sum_{t=s}^{T} Z_{T,t-\tau}Z'_{T,t-\tau}\right)^{-1}$ and $\widehat{S}_T$ in (2.6) can be expressed as:

$$\widehat{S}_T = \widetilde{S}_T/\hat{\sigma}^2_\varepsilon(T), \text{ where } \widetilde{S}_T \equiv \widetilde{S}(T) = \sum_{t=s}^{T} y_t Z'_{T,t-\tau}\Big(\sum_{t=s}^{T} Z_{T,t-\tau}Z'_{T,t-\tau}\Big)^{-1}\sum_{t=s}^{T} Z_{T,t-\tau}y_t. \tag{2.7}$$

We use the notation where $\hat{\sigma}^2_\varepsilon(T)$ in (2.7) to symbolize that it is a consistent estimator of $\sigma^2_\varepsilon = var(\varepsilon_t)$ based on the full sample. Similarly, let $\widehat{S}_{R_T}$ denote the Wald statistic for $\beta_{22} = 0$ computed using the first R_T observations in the sample. Hansen and Timmermann (2015) show that $\mathscr{T}_T$ in (2.5) is asymptotically equivalent to the difference between two Wald-type statistics, i.e., $\mathscr{T}_T = \mathscr{W}_T + o_p(1)$, where

$$\mathscr{W}_T = \widehat{S}_T - \widehat{S}_{R_T} + \sigma_\varepsilon^{-2}\check{\kappa}log(\rho), \tag{2.8}$$

with $\check{\kappa} = tr[\check{H}_2^{-1}\check{\Omega}]$ under H_0, where $\check{H}_2$ and $\check{\Omega}$ are defined above. The expression in (2.8) shows that $\mathscr{W}_T$ is related to the homoscedastic Wald statistics for testing $\beta_{22} = 0$, regardless of whether the underlying process is homoscedastic and regardless of whether $\beta_{22} = 0$ or not. As such, the recursive F-statistic $\mathscr{T}_T$ is not robust to heteroscedasticity. However, Hansen and Timmermann (2015) highlight the importance of using (2.8) to correct for heteroscedasticity or serial correlation because the HAC estimator is easy to implement with Wald-type statistics. However, even if such correction was implemented, the asymptotic distribution of the resulting statistic will still involve stochastic integrals of Brownian motions when (2.1) includes multiple extra predictors, thus making it cumbersome to compute critical values. This provides a strong motivation for our bootstrap method that not only alleviates the shortcomings of F-statistic $\mathscr{T}_T$, but also makes it easier to implement. Before moving on to the bootstrap procedure, we first consider the following notations and assumptions.

2.2 Notations and assumptions

Throughout the study, we use the following notations. For any $j \in \{1,2\}$, we define

$$
\begin{aligned}
s_{jt}(\beta_j) &= X_{j,t-\tau}(y_t - X'_{j,t-\tau}\beta_j) \equiv \big(s_{j,p}(t)\big)_{1\le p\le k_j}, \\
h_{jt} &= X_{j,t}X'_{j,t} \equiv \big[h_{j,pl}(t)\big]_{1\le p,l\le k_j}, \quad H_{jt} = \frac{1}{t}\sum_{n=s}^{t} h_{j,n-\tau}.
\end{aligned}
$$

Define $\beta_{2r} = (\beta_1', 0')'$ and consider the selection matrix $J = [I_{k_1\times k_1} \vdots 0_{k_1\times k_2}]'$ such that $J's_{2t}(\beta_{2r}) = s_{1t}(\beta_1)$ and $J'h_{2t}J = h_{1t}$. Also, let

$$
\begin{aligned}
\widetilde{s}_{2t}(\beta_2) &= \sigma_\varepsilon^{-1}\widetilde{A}H_2^{-1/2}s_{2t}(\beta_2), \\
\widetilde{A} &\in \mathbb{R}^{k_2\times k} : \widetilde{A}'\widetilde{A} = H_2^{1/2}(-JH_1^{-1}J' + H_2^{-1})H_2^{1/2},
\end{aligned}
\tag{2.9}
$$

where $\sigma_\varepsilon^2 = var(\varepsilon_{2,t+\tau})$ and $H_j = E_F[h_{jt}]$ for all $j \in \{1,2\}$. We denote by $B(r) = \big[B_1(r), \dots, B_{k_2}(r)\big]' \in \mathbb{R}^{k_2}$, the standard Brownian motion defined on $\mathbb{D}^{k_2}_{[0,1]}$, where $\mathbb{D}^{k_2}_{[0,1]}$ is the space of Cadlag mappings from $[0,1]$ to $\mathbb{R}^{k_2}$. For any positive definite $q\times q$ real matrix Σ, $B(\Sigma)$ stands for a q-dimensional Brownian motion having covariance matrix Σ; see e.g. (Davidson, 1994, Section 27.7). We now consider the following assumptions on the model variables and parameters.

Assumption 1.

(i) $U_{jt} = \big[s_{jt}(\beta_j)', vec\big(h_{jt} - H_j\big)'\big]'$ *is covariance stationary such that* $\mathbb{E}_F(U_{jt}) = 0$ *and* $H_j \equiv \mathbb{E}_F(h_{jt})$ *is positive definite for all* t *and* j;

(ii) U_{jt} *is* $3(2+1/\psi)$*-dominated*[4] *uniformly in* β_j *for some* $\psi > 0$ *and all* t, j;

(iii) U_{jt} *is* $L_{2+\delta}$*-NED*[5] *on some sequence* $\{V_{jt}\}$ *uniformly in* β_j *of size* $-2(1+\psi)$, *where* $\{V_{jt}\}$ *is* α*-mixing of size* $-2(2+\delta)(1+2\psi)$ *for* t, j *and some* $\delta > 0$, $\psi > 0$.

Assumption 2.
There is a kernel function $K(\cdot)$ *with bandwidth* $q_T + 1$ *satisfying*:

(i) $K(\cdot) : \mathbb{R} \to [-1,1]$, $K(0) = 0$, $K(x) = K(-x)\,\forall\, x \in \mathbb{R}$, $\int_{-\infty}^{+\infty} K(x)^2dx < \infty$, $\int_{-\infty}^{+\infty} |K(x)|dx < \infty$, $K(\cdot)$ *is continuous at 0 and at all but a number of other points in* $\mathbb{R}$, $\sup_{x\ge 0}|K(x)| < \infty$;

[4] That is, there exists $\bar{U}_{jt}$ such that $|U_{j,p}(t)| < \bar{U}_{jt}$ and $\mathbb{E}_F\big[|\bar{U}_{jt}|^{3(2+1/\psi)}\big] < \infty$, for all t, j and p, where $U_{jt} := \big(U_{j,p}(t)\big)_{1\le p\le k}$.

[5] Let $\{V_t\}$ be a stochastic process and $\mathscr{F}_{t-n}^{t+n} := \sigma(V_{t-n}, \dots, V_{t+n})$ denote the σ-field generated by $V_{t-n}, \dots, V_{t+n}$. We define a process $\{W_t\}$ to be NED (Near Epoch Dependent) on a mixing process $\{V_t\}$ if $\mathbb{E}_F[\|W_t\|^2] < \infty$ and $v_n := \sup_t \|W_t - E_{t-n}^{t+n}(W_t)\|_2 \to 0$ as $n \to \infty$, where $\|\cdot\|_p$ is the L_p norm and $E_{t-n}^{t+n}(\cdot) \equiv \mathbb{E}_F[\cdot|\mathscr{F}_{t-n}^{t+n}]$. $\{W_t\}$ is NED on $\{V_t\}$ of size $-a$ if $v_n = O(n^{-a-\delta})$ for some $\delta > 0$. We say that $\{V_t\}$ is strong mixing with coefficients $\alpha_n \equiv \sup_m \sup_{A\in\mathscr{F}_{-\infty}^m, B\in\mathscr{F}_{m+n}^{\infty}} | P(A\cap B) - P(A)P(B) |$ if $\alpha_n \to 0$ as $n \to \infty$ suitably fast.

(ii) as $T \to \infty$, $q_T \to \infty$ and $q_T/\sqrt[q]{T} \to 0$ for some $q \in [0, \infty)$ such that:

$$\|f^{(q)}\| \in [0, \infty) \text{ where } f^{(q)} = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} |j|^q \mathbb{E}_F[X_{2,t-j} X'_{2,t}];$$

(iii) $\int_0^{+\infty} \bar{K}(x)dx < \infty$ where $\bar{K}(x) = \sup_{y \geq x} |K(y)|$.

Assumption 3.

(i) $T = R_T + P_T - 1$ and $R_T = \lfloor \rho T \rfloor$ for some $\rho \in (0, 1)$;

(ii) $P_T/R_T \to \pi = (1 - \rho)/\rho$ as $T \to \infty$.

Remarks.

1. The covariance stationary condition of both the score vector $s_{jt}(\beta_j)$ and the Hessian matrix h_{jt} in Assumption 1-(i) is standard in the literature; see e.g. Clark and McCracken (2012) among others. In addition, the existence of at least the first six moments is required for both $s_{jt}(\beta_j)$ and h_{jt} (as per Assumption 1-(ii)). Moreover, the NED and strong α-mixing conditions in Assumption 1-(iii) ensures that the process $\{y_t, X_{2t}\}$ is ergodic in both the mean and the covariance, which is required for the application of the central limit theorem in this type of framework. Unlike previous studies[6], Assumption 1 allows the the score vectors $s_{jt}(\beta_j)$ and $s_{j,t-h}(\beta_j)$ to be correlated for $h \geq \tau - 1$. The idea to assume that the autocorrelations of the score vector $s_{jt}(\beta_j)$ are zeros for lag grater than $\tau - 1$ is usually sustained by the fact that the τ-period-ahead forecast errors exhibits an $MA(\tau - 1)$ serial correlation, thus vanishing between observations that are distant at least τ periods. We relax this assumption for the following reasons. As the model may include the lags of the dependent variable as predictors, researchers may not choose the optimal lags to include in the modelling process, but rather use fewer lagged predictors. This will create a pattern of serial correlation in the errors ε_{jt}, and therefore in the score vectors $s_{jt}(\beta_j)$.

2. The conditions (i) & (ii) of Assumption 2 ensures the consistency of the HAC estimator with a rate of convergence established by Andrews (1991). Under these conditions, the bandwidth parameter q_T satisfies:

$$\limsup_{T \to \infty} \sup_{0 < \nu < \nu_u} (q_T + 1)^{-1} \sum_{n=1}^{T-1} \left| K\left(\frac{n}{\nu(q_T + 1)} \right) \right| < \infty \tag{2.10}$$

for any $0 < \nu_u < \infty$– e.g., see Jansson (2002, Lemma 1). Note that condition (2.10) holds for all kernels belonging to class $\mathscr{K}_3$ of Andrews (1991, Eq.(7.1)) and those satisfying Assumptions 1 and 3 in Newey and West (1994).

3. Assumption 3 is used in most studies of out-of-sample tests of predictive accuracy. It implies that $0 < \pi < \infty$, i.e., R_T and P_T grow at the same rate as

[6] See e.g. Clark and McCracken (2001, 2012, 2014, 2015); McCracken (2007), and Hansen and Timmermann (2015) among others.

T increases. It can be extended to $\pi = 0$, i.e., P_T grows at a lower rate than R_T. Inference in this case is straightforward as it yields pivotal statistics (see McCracken, 2007, Theorem 3.2-(b)), meaning that our bootstrap method will yield a high-order refinement in this case.

Under Assumptions 1–3, Hansen and Timmermann (2015) provides a characterization of the asymptotic distribution of $\mathscr{T}_T$ under the null hypothesis $\beta_{22} = 0$ and local alternatives of the form $\beta_{22} = cT^{-1/2}b$ for some constant scalar c and vector b. More precisely, they show that:

(a) if $\beta_{22} = 0$, then

$$\mathscr{T}_T \overset{d}{\to} \sum_{l=1}^{k_2} \lambda_l \left[2\int_\rho^1 r^{-1}B_l(r)dB_l(r) - \int_\rho^1 r^{-2}B_l^2(r)d(r)\right] \tag{2.11}$$

$$\overset{d}{\equiv} \sum_{l=1}^{k_2} \lambda_l \left[B_l^2(1) - \rho^{-1}B_l^2(\rho) + log(\rho)\right]; \tag{2.12}$$

(b) and if $\beta_{22} = cT^{-1/2}b$ for some c and b (b is such that $b'\check{\Sigma}b = \sigma_\varepsilon^2\kappa$), then

$$\mathscr{T}_T \overset{d}{\to} \sum_{l=1}^{k_2} \lambda_l \Big[2\int_\rho^1 r^{-1}B_l(r)dB_l(r) - \int_\rho^1 r^{-2}B_l^2(r)d(r) + (1-\rho)c^2 + 2ca_l[B_p(1) - B_l(\rho)]\Big] \tag{2.13}$$

$$\overset{d}{\equiv} \sum_{l=1}^{k_2} \lambda_l \left[B_l^2(1) - \rho^{-1}B_l^2(\rho) + log(\rho) + (1-\rho)c^2 + 2ca_l[B_l(1) - B_l(\rho)]\right];$$

where $a = b'\check{H}_2\check{\Omega}_\infty^{-1/2}Q' \equiv \left(a_l\right)_{1\le l\le k_2}$, Q is an orthogonal matrix such that $Q'Q = I_{k_2}$ and $Q'\Lambda Q = \sigma_\varepsilon^{-2}\check{\Omega}_\infty^{1/2}\check{H}_2^{-1}\check{\Omega}_\infty^{1/2}$, $\Lambda = diag(\lambda_1, \dots, \lambda_{k_2})$, and $\check{\Omega}_\infty$ is the asymptotic variance of the stochastic process $\{Z_{t-\tau}\varepsilon_{2t}\}$ where $Z_{t-\tau} = \widetilde{X}_{2,t-\tau} - H_{21}H_1^{-1}X_{1,t-\tau}$.

Remarks. Several observations are of order.

1. The expression of the limiting distribution of $\mathscr{T}_T$ in (2.11) is well known in the literature; see e.g. McCracken (2007). Hansen and Timmermann (2015) show that this integral of stochastic Brownian motions can be expressed as a convolution of dependent $\chi^2(1)$ variables, as shown in (2.12).

2. Hansen and Timmermann (2015) show that the limiting distribution of each of the Wald statistics $\widehat{S}_T$ and $\widehat{S}_{R_T}$ in (2.8) is a convolution of dependent $\chi^2(1)$ variables, where λ_l ($l = 1, \dots k_2$) are the eigenvalues of the matrix $\sigma_\varepsilon^{-2}\check{H}_2^{-1}\check{\Omega}_\infty$. Due to the equivalence between $\mathscr{T}_T$ and $\mathscr{W}_T$, this translates to the limiting distribution of $\mathscr{T}_T$ being a convolution of dependent $\chi^2(1)$ variables, as shown in (2.12). As such, the eigenvalues λ_l ($l = 1, \dots k_2$) can be viewed a measures of heteroscedasticity in the model. Under homoscedasticity, $\lambda_l = 1$ for all l and (2.11) reduces to the earlier result in McCracken (2007). But under heteroscedasticity, $\lambda_l \neq 1$ for some $l = 1, \ldots, k_2$ and (2.12) illustrates clearly that $\mathscr{T}_T$ (thus $\mathscr{W}_T$) is not robust to heteroscedasticity.

3. When $B(r)$ is a univariate standard Brownian motion (i.e., when (2.1) contains only one extra predictor), the limiting distribution in (2.12) is identical to that of $\sqrt{1-\rho}(\mathbf{Z}_1^2 - \mathbf{Z}_2^2) + log(\rho)$, where $\mathbf{Z}_j \overset{i.i.d.}{\sim} N(0,1)$, $j = 1, 2$ (Hansen and Timmermann, 2015, Theorem 4), and can thus be simulated easily given ρ. This case is very restrictive as it implies that $\widetilde{X}_2$ in (2.1) contains only one regressor (i.e. $k_2 = 1$).

4. When $B(r)$ is a multivariate standard Brownian motion (i.e., when $\widetilde{X}_2$ contains more than one regressor) and the DGP is homoscedastic (i.e., $\lambda_1 = \ldots = \lambda_{k_2} = 1$), a closed-form expression of the pdf of the exact density of the asymptotic limit variable in (2.12) is provided in Hansen and Timmermann (2015, Theorem 5). This density can be used to simulate the asymptotic critical values of $\mathscr{T}_T$ under H_0. However, assuming homoscedasticity is restrictive in most times series applications. To the best of our knowledge, we are not aware of a closed-form characterization of the pdf of the limit variable in (2.12) under heteroscedasticity or serial correlation in the multivariate nested setting (i.e., $k_2 > 1$). Moreover, the equivalence between $\mathscr{T}_T$ and $\mathscr{W}_T$ is asymptotic in nature, so the finite-sample behavior of both statistics may differ, i.e. even in cases where the critical values of the limit variable in (2.12) can be simulated, the resulting test may still yield size distortions in small samples.

2.3 Finite-sample properties of the tests

We investigate the finite-sample performance (size) of $\mathscr{T}_T$ and $\mathscr{W}_T$ through Monte Carlo simulations. For this, we use the framework of Section 2 where $\widetilde{X}_2$ in (2.1) contains only one regressor (i.e., $k_2 = 1$). The reason we use $k_2 = 1$ is that it is the only case where the asymptotic critical values can be simulated easily upon allowing for homoscedasticity, as discussed in **Remark 3.**

Specifically, the DGP is described by the following bi-variate vector autoregression (VAR) (similar to Hansen and Timmermann, 2015):

$$y_t = 0.3y_{t-\tau} + \beta_{22}x_{t-\tau} + u_{yt}, \tag{2.14}$$

$$x_t = 0.5x_{t-\tau} + u_{xt}, \tag{2.15}$$

where u_{yt}, u_{xt} are the error terms and τ is the forecast horizon. We run the experiments with $\tau \in \{1, 4\}$ but the results are qualitatively the same for superior horizons. In all experiments, the null forecast model is no-influence of $x_{t-\tau}$ in (2.14), and the alternative (unrestricted) forecast model takes the form of (2.14)-(2.15) with $\beta_{22} \neq 0$, therefore the testing problem of interest can be framed as:

$$H_0 : \beta_{22} = 0 \ vs. \ H_1 : \beta_{22} \neq 0. \tag{2.16}$$

We cover four DGPs for the error vector $(u_{yt}, u_{xt})'$, some of which are important to account for the presence of heteroscedasticity or autocorrelation.

DGP 1 (Homoscedasticity): $(u_{yt}, u_{xt})' \overset{i.i.d.}{\sim} N(0, I_2)$;

DGP 2 (Heteroscedasticity alone): $u_{jt} \sim N(0, h_{jt})$ for $j \in \{x, y\}$, where $h_{jt} = \alpha_0 + \alpha_1 u_{jt-1} + \alpha_2 h_{jt-1}$ and $\alpha_1 = 0.1$, $\alpha_2 = 0.8$;

DGP 3 (Autocorrelation only): $u_{jt} = 0.5u_{j,t-1} + \varepsilon_{jt}$ for $j \in \{x, y\}$, where $(\varepsilon_{yt}, \varepsilon_{xt})' \overset{i.i.d.}{\sim} N(0, I_2)$ when $\tau = 1$; while $u_{jt} = 0.50u_{j,t-1} + \varepsilon_{jt} + 0.95\varepsilon_{jt-1} + 0.90\varepsilon_{jt-2} + 0.80\varepsilon_{jt-3}$, where $(\varepsilon_{yt}, \varepsilon_{xt})' \overset{i.i.d.}{\sim} N(0, I_2)$ when $\tau = 4$;[7]

DGP 4 (Heteroscedasticity & Autocorrelation): DGP 2 + DGP 3.

The simulations are run with $N = 10,000$ pseudo samples of size $T \in \{50, 100, 200, 500\}$, and the nominal level α is set at 5%.[8] We consider the following sample split points for $\pi \left(= \frac{(1-\rho)}{\rho}\right)$: $\{0.2, 0.8, 1.4, 2.0\}$. For example, when $\pi = 0.2$ we have $\rho = \frac{5}{6}$, i.e., $\left\lfloor \frac{5T}{6} \right\rfloor$ observations are used in the initial estimation. This allows us to compare our results with previous studies (see e.g., Clark and McCracken, 2001, 2005, 2001, 2005; McCracken, 2007).

Table 1 shows the rejection frequencies of the tests with $\mathscr{T}_T$ and $\mathscr{W}_T$ using the simulated asymptotic critical values, for both 1-step-ahead ($\tau = 1$) and 4-step-ahead ($\tau = 4$) forecasts. The first column of the tables presents the fraction π of the sample used in the initial estimation period. The other columns present, for each DGP, the rejections frequencies under H_0 of the tests at the nominal 5% level.

Considering first the case of 1-period-ahead forecasts, we see that both tests are oversized when the standard critical values are used and the sample size is small. In particular, when $T = 50$, the size distortions under DGP 3 (autocorrelation alone) and DGP 4 (heteroscedasticity and autocorrelation) are large. For example, the maximal rejection frequencies under DGP 4 for the F-test can be as high as twice the nominal level (i.e., 10.3%), while that of the Wald test is even worse (13.7%). More precisely, the rejection frequencies of the recursive F-test in DGP 4 range from 7.5% to 10.3% and that of the Wald test range from 12.1% to 13.7%. Similar results are observed in DGP 3 (autocorrelation alone). When $T = 100$, size distortions persist for both tests in DGP 3 and DGP 4. However, the tests show better size when $\tau = 1$ and $T \in \{200, 500\}$.

Next, considering the case of 4-period-ahead forecasts, we see that both tests over-reject the null hypothesis substantially, and the size distortions persist in DGP 3 and DGP 4 even when $T = 500$. In particular, the rejection frequency of $\mathscr{T}_T$ can be as

[7] Note that in addition to the $AR(1)$ property, the forecast errors also exhibit $MA(\tau - 1)$ when $\tau = 4$, as expected from the theory. The form of the $MA(\tau - 1)$ is identical to the one in Clark and McCracken (2012).

[8] We also run the experiments with $\alpha \in \{1\%, 10\%\}$ and the main findings remain qualitatively unchanged. These results are not included in order to shorten the exposition.

large as 27.6% when $T = 50$, 17.6% when $T = 100$, 13.9% when $T = 200$, and 12.1% when $T = 500$. Similar results are observed for $\mathscr{W}_T$ as well but its rejection frequencies under DGP 3 and DGP 4 are slightly less than that of $\mathscr{T}_T$.

Table 1: Size of $\mathscr{T}_T$ and $\mathscr{W}_T$ with asymptotic critical values, $\alpha = 5\%$

$T = 50$

	$h = 1$								$h = 4$							
	DGP 1		DGP 2		DGP 3		DGP 4		DGP 1		DGP 2		DGP 3		DGP 4	
π	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald
0.2	7.9	8.6	7.1	7.9	10.6	12.0	10.3	12.7	10.8	12.1	10.5	11.7	21.5	21.0	21.8	21.1
0.8	8.0	8.7	7.4	8.8	9.7	13.2	10.0	13.7	10.0	12.0	9.4	11.6	25.2	19.6	25.5	19.8
1.4	6.8	8.8	6.7	8.0	8.7	12.1	8.9	12.1	8.7	11.0	9.0	11.5	25.3	17.0	26.0	17.4
2.0	7.0	8.6	6.5	8.5	7.9	11.5	7.5	12.2	8.5	10.4	7.5	10.4	27.6	15.6	27.1	15.2

$T = 100$

	$h = 1$								$h = 4$							
	DGP 1		DGP 2		DGP 3		DGP 4		DGP1		DGP 2		DGP 3		DGP 4	
π	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald
0.2	6.0	6.4	5.9	6.1	7.5	8.5	7.3	8.3	7.5	7.9	7.2	7.8	14.6	13.5	14.7	13.4
0.8	6.0	6.2	6.7	6.5	7.2	8.3	8.0	9.0	7.1	7.6	7.0	7.8	16.6	13.7	16.5	13.2
1.4	6.0	5.9	6.2	6.2	7.0	8.1	6.8	8.0	7.0	7.3	6.9	7.7	17.3	12.5	17.1	12.3
2.0	6.2	5.9	6.2	6.0	6.5	7.8	6.5	7.9	6.3	6.9	6.4	7.3	17.4	11.6	17.6	11.8

$T = 200$

	$h = 1$								$h = 4$							
	DGP 1		DGP 2		DGP 3		DGP 4		DGP 1		DGP 2		DGP 3		DGP 4	
π	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald
0.2	5.8	5.9	5.3	5.6	6.0	6.6	5.9	6.4	6.4	6.8	6.3	6.9	10.2	9.4	10.5	9.9
0.8	5.6	5.4	5.5	5.5	6.0	6.5	6.0	6.7	5.9	6.3	6.0	6.4	12.2	9.6	12.1	9.6
1.4	5.6	5.7	5.6	5.6	6.2	7.0	6.1	7.2	6.6	7.1	5.3	5.9	13.9	9.8	13.1	9.5
2.0	5.5	5.5	5.4	5.1	6.1	7.2	5.7	6.6	5.8	6.5	6.1	7.0	13.8	9.1	13.9	9.5

$T = 500$

	$h = 1$								$h = 4$							
	DGP 1		DGP 2		DGP 3		DGP 4		DGP 1		DGP 2		DGP 3		DGP 4	
π	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald
0.2	5.7	5.8	5.1	5.3	5.9	6.1	5.2	5.5	5.6	5.8	5.9	6.1	8.7	7.5	7.8	7.0
0.8	5.1	5.0	4.9	5.1	5.5	5.6	5.2	5.6	5.2	5.5	5.4	5.7	9.5	7.1	9.5	7.1
1.4	5.2	5.1	4.9	5.0	5.4	6.0	5.6	6.1	5.3	5.5	5.4	5.9	10.7	7.1	11.0	7.5
2.0	5.3	5.2	5.1	5.0	5.2	5.7	5.3	5.9	5.4	5.9	5.2	5.6	11.6	7.1	12.1	7.8

3 Bootstrap tests

In this section, we propose a bootstrap procedure that alleviates some shortcomings of the previous literature[9], in addition to also providing finite-sample improvements of the F and Wald tests of equal forecast accuracy. In case of multistep-ahead forecasts (i.e. when $\tau > 1$), previous bootstrap methods often fail to control the size even for a well specified homoscedastic model due to the resulting $MA(\tau-1)$ serial correlation structure. Our bootstrap method builds on earlier work by Corradi and Swanson (2007) and performs well for both $\mathscr{T}_T$ and $\mathscr{W}_T$ irrespective of the forecast horizon τ.

3.1 Bootstrap DGP

Let $\hat{\varepsilon}_{1,t}$ denote the residuals from the OLS of (2.2), and define $\{W_t = (X^{\dagger}_{2,t}, \hat{\varepsilon}_{1,t}) : t = s, \ldots, T\}$ where $X^{\dagger}_{2,t}$ contains the variables in $X_{2,t}$ other than the lags of the dependent variables y_t. Let $\ell_T \in \mathbb{N}$ be a block length $(1 \leq \ell_T \leq T-s)$, $B_{t,\ell_T} = \{W_t, W_{t+1}, \ldots, W_{t+\ell_T-1}\}$ be the block of ℓ_T consecutive observations starting at W_t for $t = s, \ldots, T$. Assume that $T-s = b_T\ell_T$ so that the moving block bootstrap (MBB) procedure consists of drawing $b_T = (T-s)/\ell_T$ blocks, $\{B^*_{1,\ell_T}, B^*_{2,\ell_T}, \ldots, B^*_{b_T,\ell_T}\}$, randomly with replacement from the set of overlapping blocks $\{B_{s,\ell_T}, \ldots, B_{T-\ell_T+1,\ell_T}\}$. The first ℓ_T observations in the pseudo-time series are the sequence of ℓ_T values in B^*_{s,ℓ_T}, the next ℓ_T observations in the pseudo-time series are the ℓ_T values in B^*_{s+1,ℓ_T}, and so on, i.e., $W^* \equiv (W^{*'}_s, W^{*'}_{s+1}, \ldots, W^{*'}_T)' = (B^{*'}_{1,\ell_T}, B^{*'}_{2,\ell_T}, \ldots, B^{*'}_{b,\ell_T})'$, where $W^{*'}_t := [X^{\dagger^{*'}}_{2,t}, \varepsilon^*_{1,t}]$ for all $t = s, \ldots T$. Let $\mathbf{I}_1, \ldots, \mathbf{I}_b$ be i.i.d. random variables distributed uniformly on $\{s-1, s, \ldots, T-\ell_T\}$, the resulting bootstrap sample can be defined as $\{W^*_t := W_{\tau_t}, t = s, \ldots, T\}$ where τ_t is a random array, i.e., $\{\tau_t\} := \{\mathbf{I}_1 + 1, \ldots, \mathbf{I}_1 + \ell_T, \ldots, \mathbf{I}_b + 1, \ldots, \mathbf{I}_b + \ell_T\}$. To construct the bootstrap dependent variable y^*_t, we proceed as follows.

1. If $X_{2,t}$ does not contain any lag of y_t, then set $X^{\dagger^{*'}}_{2,t} := X^{*'}_{2,t} = [X^{*'}_{1,t}, \widetilde{X}^{*'}_{2,t}]$ and generate y^*_t as:

$$y^*_t = X^{*'}_{1,t-\tau}\hat{\beta}_1 + \varepsilon^*_{1,t},\ t = s, s+1, \ldots, T. \tag{3.1}$$

2. If $X_{2,t}$ contains lags of y_t, proceed as follows. First, set $y^*_t = y_t$ for all $t = 1, \ldots, s-1$ (initial values) and form $y^*_{t-\tau}$ for $t = s, \ldots, \tau+s-1$,[10] and use the bootstrap draws in $X^{\dagger^{*'}}_{2,t}$ to form $X^{*'}_{1,t}$. Then, compute y^*_t for $t = s+\tau, \ldots, T$ as:

$$y^*_t = X^{*'}_{1,t-\tau}\hat{\beta}_{1,t} + \varepsilon^*_{1,t},\ t = s+\tau, \ldots, T. \tag{3.2}$$

[9]For example, our bootstrap method applies even when the errors in the DGP are heteroscedastic or serially correlated, and the unrestricted model contains more that one predictors.

[10]Note that by definition, $s = \max\{q, \tau\} + 1$, with q being the maximum lag length of y_t included in $X_{2,t}$.

Remarks.

1. The above bootstrap scheme is an hybrid of moving block bootstrap (which is nonparametric) and residual based bootstrap. As such, it differs from previous bootstrap procedures of tests of equal forecast accuracy in a number of ways. In particular, it differs from Corradi and Swanson (2007) in three important aspects.

First, while Corradi and Swanson (2007)'s bootstrap is non-parametric in the sense that level data are re-sample (pairs bootstrap) in their DGP, ours is based on resampling the residuals from the restricted regression where the null hypothesis H_0 is imposed. Clearly, in addition to the parametric nature of the model, our bootstrap resampling scheme also exploits the null hypothesis of equal forecast accuracy. Resampling the residuals help recover the serial correlation pattern of the regression errors. For example, the Monte Carlo experiment in Corradi and Swanson (2007) illustrates that the pairs bootstrap does not always mimic well the serial correlation pattern in regression errors, as well as the persistence of the data, thus leading to valid but conservative bootstrap tests. Looking at Corradi and Swanson (2007, Tables 2-3)], we see that with moderate autocorrelation of the errors in their DGP ($a_3 = 0.3$) and mild persistence in the data ($a_2 = 0.6$), the *size2* results indicate a conservative moving block bootstrap procedure at the 10% nominal level for 1-period-ahead forecasts. The size property of this moving block bootstrap worsens as the persistence in the data increases– e.g., see Corradi and Swanson (2007, Tables 2-3, Panel: C).

Second, as our bootstrap DGP imposes the null hypothesis, the resulting bootstrap (bias-corrected) estimators and test statistics will not converge to the asymptotic distributions of the sample counterparts under the alternative hypothesis (i.e., when $\beta_{22} \neq 0$). However, we will show that the bias-corrected estimators and the resulting F-type and Wald-type bootstrap statistics will mimic well the asymptotic distributions of the sample counterparts under the null hypothesis. As such, using our proposed bootstrap critical values yields consistent tests (see Theorem 3.3).

Third, Corradi and Swanson (2007) establish the conditions on the block length ℓ_T under which the MBB is consistent. We establish similar conditions but also suggest a data dependent method to select the bootstrap block length. As the choice of the block length should capture the structure of dependence in the data, we believe equalizing it to the optimal lag length of the HAC estimator of the variance of the errors is a reasonable choice. We are not aware of any study on bootstrapping out-of-sample tests of predictive accuracy that addresses formally the problem of autocorrelation and the selection choice of the block length in a data-dependent way. For example, Corradi and Swanson (2007, Tables 2-3) provide clear evidence that the choice of block length influences the performance of the bootstrap CS test. From that perspective, our bootstrap method can be viewed as an important contribution that extends Corradi and Swanson (2007).

2. An important contribution on residual MBB is Efron (1982, pp.35-36), but its bootstrap scheme considers the regressors as strictly exogenous, and hence are not re-sampled in the bootstrap algorithm. This type of MBB is not appropriate for weakly dependent time series with lagged dependent variables. Fitzenberger (1998) proposes a MBB where the regressors are re-sampled, but as in Corradi and Swanson (2007), the choice of block length is not addressed. Other recent contributions on bootstrapping out-of-sample tests of predictive accuracy include Clark and McCracken (2012, 2014, 2015). Their bootstrap algorithm relies on a variant of the wild bootstrap but also assumes that the "regressors are fixed."

Let $\hat{\beta}^*_{j,t}$ be the recursive bootstrap estimator similar to $\hat{\beta}_{j,t}$ in (2.4), i.e.

$$\hat{\beta}^*_{j,t} \quad = \quad \arg\min_{\beta_j} \frac{1}{t} \sum_{n=s}^{t} (y^*_n - X^{*\prime}_{j,n-\tau}\beta_j)^2, \quad R_T \leq t \leq T; j = 1, 2. \tag{3.3}$$

It is important to note that in (3.3), while $\hat{\beta}^*_{1,t}$ depends on $\hat{\beta}_{1,t}$, $\hat{\beta}^*_{2,t}$ does not directly depend on $\hat{\beta}_{2,t}$ since the null hypothesis $\beta_{22} = 0$ is imposed in the bootstrap DGP. Rather, it is straightforward to show that $\hat{\beta}^*_{2,t}$ depends directly on the restricted estimator $\tilde{\beta}_{2,t} = (\hat{\beta}'_{1,t}, 0')'$. As such, letting $\hat{\theta}_{1,t} \equiv \hat{\beta}_{1,t}$ and $\hat{\theta}_{2,t} \equiv \tilde{\beta}_{2,t}$, we show in Lemma **A.3** in the appendix that $\mathbb{E}_{F^*}\left[\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\hat{\beta}^*_{j,t} - \hat{\theta}_{j,t})\right] = O_{p^*}(1)$ pr-$\mathbb{P}$ for all $j = 1, 2$, i.e., the limiting distribution of $\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\hat{\beta}^*_{j,t} - \hat{\theta}_{j,t})$ is not centered at zero but is rather characterized by a location bias. This means that a bootstrap test based on $\hat{\beta}^*_{j,t}$ may not have a desirable size property and some adjustments are required. Several studies, including Politis and Romano (1994) and Corradi and Swanson (2007), have proposed methods to eliminate this location bias from $\hat{\beta}^*_{j,t}$. Due to its simplicity, we adapt the approach by Corradi and Swanson (2007).

Define the adjusted recursive estimator

$$\tilde{\beta}^*_{j,t} \quad = \quad \arg\min_{\beta_j} \frac{1}{t} \sum_{n=s}^{t} \left[(y^*_n - X^{*\prime}_{j,n-\tau}\beta_j)^2 + 2\beta'_j \left(\mu_T \sum_{n=s}^{T} s_{j,n}(\hat{\theta}_{j,t}) \right) \right], \quad R_T \leq t \leq T, \tag{3.4}$$

where $\mu_T = 1/(T - s + 1)$ and $s_{j,n}(\hat{\theta}_{j,t}) = X_{j,n-\tau}(y_n - X'_{j,n-\tau}\hat{\theta}_{j,t})$ for all $j = 1, 2$. We can solve (3.4) explicitly for $\tilde{\beta}^*_{j,t}$ to get

$$\tilde{\beta}^*_{j,t} \quad = \quad \left(\frac{1}{t} \sum_{n=s}^{t} h^*_{j,n-\tau} \right)^{-1} \left(\frac{1}{t} \sum_{n=s}^{t} \left[X^*_{j,n-\tau} y^*_n - \mu_T \sum_{n=s}^{T} s_{j,n}(\hat{\theta}_{j,t}) \right] \right), \ R_T \leq t \leq T, \tag{3.5}$$

where $h^*_{j,n-\tau} = X^*_{j,n-\tau} X^{*\prime}_{j,n-\tau}$ is the bootstrap analog of $h_{j,n-\tau} = X_{j,n-\tau} X'_{j,n-\tau}$. We have the following result on the asymptotic behavior of $\tilde{\beta}^*_{j,t}$.

Theorem 3.1. *Suppose Assumptions 1 - 3 are satisfied and $\ell_T = o(T^{\frac{1}{4}})$. Under H_0, we have:*

$$\lim_{T\to\infty} \mathbb{P}\Big[\omega : \sup_{v_j \in \mathbb{R}^{k_j}} \Big| \mathbb{P}^*\Big(\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\tilde{\beta}^*_{j,t} - \hat{\theta}_{j,t}) \leq v_j \Big) - \mathbb{P}\Big(\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\hat{\theta}_{j,t} - \beta^0_j) \leq v_j \Big) \Big| > \zeta \Big] = 0$$

for any $\zeta > 0$*, where* $\beta_2^0 \equiv \beta_{2r}^0 = (\beta_1^{0'}, 0')'$.

Remarks.

1. Theorem 3.1 establishes the consistency under H_0 of the limiting distribution of $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\tilde{\beta}_{j,t}^* - \hat{\theta}_{j,t})$ to that of $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\theta}_{j,t} - \beta_j^0)$ for all $j = 1, 2$. Note that this result also holds under the alternative hypothesis $H_1 : \beta_{22} \neq 0$ for $j = 1$. However, it does not hold under H_1 for $j = 2$ due to the fact that H_0 is imposed under the bootstrap DGP so that the asymptotic distributions of $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\theta}_{2,t} - \beta_2^0)$ and $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\theta}_{2,t} - \beta_2) = \frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\theta}_{2,t} - \beta_2^0) - \underbrace{(\beta_2 - \beta_2^0)}_{\neq 0 \text{ under } H_1}$ are different. By the Mann and Wald's (1943) theorem, a bootstrap statistic for H_0 that is a continuous function of $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\tilde{\beta}_{j,t}^* - \hat{\theta}_{j,t})$ should possess the size control property.

2. With the exception of imposing H_0, the statement of Theorem 3.1 is similar to Corradi and Swanson (2007, Theorem 1), but the proof is slightly different because the asymptotic distribution of $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\theta}_{j,t} - \beta_j^0)$ is not a standard Gaussian random variable, but is rather a mixture of Brownian motions, as shown in Lemma **A.2-(b)** in the appendix.

Section 3.2 presents our bootstrap statistics and studies their asymptotic behavior.

3.2 Bootstrap statistics

Following Corradi and Swanson (2007), we suggest the following recursive bootstrap F-statistic and its equivalent Wald-statistic:

$$\mathscr{T}_T^* = \frac{1}{\hat{\sigma}_\varepsilon^{*2}} \sum_{t=R_T}^{T} \left[(y_t^* - X_{2,t-\tau}^{*'}\tilde{\beta}_{2,t}^*)^2 - (y_t^* - X_{1,t-\tau}^{*'}\tilde{\beta}_{1,t}^*)^2\right] \tag{3.6}$$

$$\mathscr{W}_T^* = \widehat{S}_T^* - \widehat{S}_{R_T}^* + \hat{\sigma}_\varepsilon^{*^{-2}} \breve{\kappa}^* log(\rho), \tag{3.7}$$

where $\tilde{\beta}_{j,t}^*$ is given in (3.5), and $\hat{\sigma}_\varepsilon^{*^2}$, $\widehat{S}_T^*$, $\widehat{S}_{R_T}^*$, and $\breve{\kappa}^*$ are the bootstrap sample corresponding of $\hat{\sigma}_\varepsilon^2$, $\widehat{S}_T$, $\widehat{S}_{R_T}$, and $\breve{\kappa}$ respectively (which are defined in Section 2.1).

The following theorem establishes the bootstrap validity under H_0.

Theorem 3.2. *Suppose Assumptions 1 - 3 are satisfied and* $\ell_T = o(T^{\frac{1}{4}})$*. Under* H_0*, we have:*

$$\lim_{T\to\infty} \mathbb{P}\Big[\omega : \sup_{v_j \in \mathbb{R}^{k_j}} \mid \mathbb{P}^*(\mathscr{T}_T^* \leq v_j) - \mathbb{P}(\mathscr{T}_T \leq v_j) \mid > \zeta\Big] = 0,$$

$$\lim_{T\to\infty} \mathbb{P}\Big[\omega : \sup_{v_j \in \mathbb{R}^{k_j}} \mid \mathbb{P}^*(\mathscr{W}_T^* \leq v_j) - \mathbb{P}(\mathscr{W}_T \leq v_j) \mid > \eta\Big] = 0$$

for any $\zeta > 0$ *and* $\eta > 0$.

Remarks.

1. The proof of Theorem 3.2 is given in the appendix. Since our hybrid bootstrap DGP imposes the null hypothesis, Theorem 3.2 only proves the validity of bootstrap under H_0. The results imply that the bootstrap critical values provide a good approximation of their asymptotic counterpart for both $\mathscr{T}_T$ and $\mathscr{W}_T$. This enables us to establish the consistency of the bootstrap under alternative hypothesis (see Theorem 3.3 below).

2. Theorem 3.2 holds regardless of whether the underlying data generating process is homoscedastic/weakly dependent or not. As such, the proposed bootstrap is robust heteroscedastic or weakly dependent data generating processes. This contrast with the *fixed regressors bootstrap* of Clark and McCracken (2012, 2015) that requires the regressors in (2.1) to be fixed and strictly exogenous. The theorem also holds for any forecast horizon τ. In particular, it is valid in multi-step forecasts setting (i.e., $\tau > 1$). Note that the *fixed regressor bootstrap* of the recursive F-test often fails to control the size when $\tau > 1$ due to the underlying $MA(\tau - 1)$ structure of the resulting forecast errors– e.g., see Clark and McCracken (2012, Table 2, DGP2) and Clark and McCracken (2015, Table 3, DGP5).

We shall now analyze the power property of the bootstrap test. To do this, let $c^*_{\mathscr{T}}(\alpha)$ and $c^*_{\mathscr{W}}(\alpha)$ denote the $(1-\alpha)$ quantiles under H_0 of the bootstrap statistics $\mathscr{T}^*_T$ and $\mathscr{W}^*_T$ respectively for some $\alpha \in (0,1)$. Theorem 3.3 establishes the hybrid bootstrap consistency under the alternative hypothesis $H_1 : \beta_{22} \neq 0$.

Theorem 3.3. *Suppose Assumptions 1 - 3 are satisfied and $\ell_T = o\big(T^{\frac{1}{4}}\big)$. If $\beta_{22} \neq 0$ is fixed, then we have*:

$$\mathbb{P}\big[\mathscr{T}_T > c^*_{\mathscr{T}}(\alpha)\big] \to 1, \ \ \mathbb{P}\big[\mathscr{W}_T > c^*_{\mathscr{W}}(\alpha)\big] \to 1 \ \ \textit{as}\, T \to \infty.$$

Theorems 3.2&3.3 establish the consistency of the proposed bootstrap. Note that it is possible to characterize the asymptotic power function of the bootstrap under local alternatives of the form given in (2.13) with the resulting asymptotic distributions of the statistics. This exercise is, however, of second order importance because these asymptotic distributions are unknown, and the motivation of the bootstrap at the first place was to avoid simulating them since doing so can be costly especially when $k_2 \geq 2$. We will now analyze the finite-sample performance of the proposed bootstrap through a Monte Carlo experiment. Section 3.3 presents the results.

3.3 Finite-sample performance of bootstrap tests

In this section, we study the performance (size and power) of the proposed bootstrap tests through Monte Carlo experiment. To do this, we use the simulation setting of Section 2.3. To enable comparison with recent bootstrap methods, we present the results of both our bootstrap F and Wald statistics along with the fixed regressor wild bootstrap MSE-F statistic of Clark and McCracken (2012) and Clark and McCracken (2015). Note that their fixed regressor bootstrap is different from ours because $X^{\dagger}_{2,t}$ is kept fixed in their resampling process.

3.3.1 Bootstrap test size

Table 2 shows the rejection frequencies for the proposed moving block bootstrap and the fixed-regressor MSE-F bootstrap tests for 1-step-ahead ($\tau = 1$) and 4-step-ahead ($\tau = 4$) forecasts for sample sizes $T \in \{50, 100, 200\}$ at nominal 5% level.

As seen, the moving block bootstrap performs well for both $\mathscr{T}_T$ and $\mathscr{W}_T$, while the fixed-regressor MSE-F bootstrap tends to over-reject, especially in DGPs 3 and 4. Considering first DGP 1 and DGP 2, we see that the empirical size of our moving block bootstrap is consistently around the 5% nominal level irrespective of the forecast horizon, the sample size (including when $T = 50$), and the cut-off point π. Meanwhile, the fixed-regressor MSE-F bootstrap shows some size distortions in both DGPs. Even though fixed-regressor MSE-F bootstrap size distortions are relatively small in DGP 1 and DGP 2, they can be as high as 7.4% for $\tau = 1$ and 7.6% for $\tau = 4$ when $T = 50$. Note that we should expect the (fixed-regressor) bootstrap to work in DGP 1 and DGP 2 because the regressor $X^{\dagger}_{2,t}$ is exogenous by design. Now, looking at DGP 3 and DGP 4, it is obvious that the size distortions of the fixed-regressor MSE-F bootstrap are very large, especially for 4-step-ahead forecasts and sample sizes $T = 50, 100$. For example, its maximal rejection frequencies in DGP 4 when $\tau = 4$ are 20.9% for $T = 50$ and 12.2% for $T = 100$. In most cases considered, our moving block bootstrap with both $\mathscr{T}_T$ and $\mathscr{W}_T$ outperforms the fixed-regressor MSE-F bootstrap.

Comparing the relative size performance between moving block bootstrap with $\mathscr{T}_T$ and that with $\mathscr{W}_T$, both perform equally well even when $T \in \{50, 100\}$. This is the case for all forecast horizons and cut-off points π considered, as opposed to their behavior when asymptotic critical values were used (see Table 1).

So far we have shown that the asymptotic tests suffer from size distortions in finite samples, particularly under heteroscedasticity or autocorrelation, while our proposed bootstrap provides size correction even in multi-step ahead forecasts. These results only consider the case with one extra predictor (i.e. $k_2 = 1$) in the larger model, which can be argued to be somewhat restrictive. In more general setting with $k_2 > 1$, no closed form characterization of the pdf of the density of the asymptotic limit variable in (2.12) exists under heteroscedasticity. Even though asymptotic critical values could be simulated, their use may still lead to size distortions in finite samples, as evident from the case with $k_2 = 1$. Nevertheless, bootstrapping tests can still be performed with no additional complexity and so we now investigate the finite sample performance (size) of $\mathscr{T}_T$ and $\mathscr{W}_T$ tests when there is more than one extra predictor in the larger model. Precisely, we use the following DGP:

$$y_t = 0.3y_{t-\tau} + \beta_{22}x_{1,t-\tau} + \beta_{23}x_{2,t-\tau} + u_{yt}, \tag{3.8}$$

$$x_{j,t} = 0.5x_{j,t-\tau} + u_{xjt}, \ j = 1, 2 \tag{3.9}$$

where we cover the same four DGPs for the error vector (u_{yt}, u_{xjt}). In all cases, the null forecast model is no influence of $x_{j,t-\tau}(j = 1, 2)$ in (3.8) as before and the alternative (unrestricted) forecast model takes the form of (3.8)- (3.9) with $\beta_{22}, \beta_{23} \neq 0$.

Table 3 shows the rejection frequencies for the moving block and the fixed-regressor MSE-F bootstrap tests. In line with the results with one extra predictor, our proposed moving block bootstrap has good finite-sample performance and also outperforms the fixed-regressor MSE-F bootstrap.

Table 2: Bootstrap rejection frequencies with sample sizes $T = 50; 100; 200$, $\alpha = 5\%$

$T = 50$

Moving Block

	h = 1								h = 4							
	DGP 1		DGP 2		DGP 3		DGP 4		DGP1		DGP 2		DGP 3		DGP 4	
π	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald
0.2	4.7	3.6	5.0	3.9	5.6	5.3	5.0	4.8	4.4	3.7	4.2	3.3	7.2	6.6	8.5	7.3
0.8	4.9	4.1	4.3	3.6	5.3	5.7	5.8	4.7	4.8	3.0	4.0	3.0	7.2	7.8	8.8	8.1
1.4	5.5	3.8	4.8	4.8	5.6	5.3	4.5	5.3	4.6	3.8	3.5	3.3	9.7	7.4	8.9	7.8
2.0	4.5	4.5	3.8	3.9	5.3	4.9	5.1	5.3	3.9	3.8	3.9	3.9	10.7	7.2	9.3	7.0

Fixed Regressor MSE-F

	h = 1				h = 4			
π	DGP 1	DGP 2	DGP 3	DGP4	DGP 1	DGP 2	DGP 3	DGP 4
0.2	5.8	5.9	6.3	6.9	6.7	6.6	9.8	17.8
0.8	7.4	6.7	8.1	7.6	6.7	7.0	9.7	19.9
1.4	7.1	7.1	7.3	7.9	7.6	7.0	10.2	19.4
2.0	7.2	7.2	7.7	7.9	7.4	7.1	10.4	20.9

$T = 100$

Moving Block

	h = 1								h = 4							
	DGP 1		DGP 2		DGP 3		DGP4		DGP 1		DGP 2		DGP 3		DGP 4	
π	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald
0.2	4.9	4.9	4.2	4.7	5.1	5.2	5.3	5.3	4.4	4.8	4.7	4.2	6.6	6.3	5.1	5.3
0.8	6.1	5.4	5.2	5.2	4.9	5.2	5.3	5.5	4.7	4.4	4.9	4.5	5.3	6.3	5.7	5.8
1.4	5.0	5.1	6.0	5.0	4.9	5.5	5.6	5.1	5.0	5.0	5.4	4.3	5.6	6.2	5.5	6.3
2.0	5.9	5.0	5.3	5.0	4.7	4.8	5.2	5.2	5.2	4.7	4.5	4.6	6.0	6.3	5.9	5.9

Fixed Regressor MSE-F

	h = 1				h = 4			
π	DGP 1	DGP 2	DGP 3	DGP4	DGP 1	DGP 2	DGP 3	DGP 4
0.2	5.8	5.7	6.4	6.3	6.3	6.8	8.1	12.2
0.8	6.8	6.1	6.6	7.0	7.0	6.9	7.9	10.9
1.4	6.2	5.9	6.4	6.8	6.8	5.6	8.0	11.2
2.0	6.5	6.5	6.7	6.5	6.2	5.9	7.8	11.8

$T = 200$

Moving Block

	h = 1								h = 4							
	DGP 1		DGP 2		DGP 3		DGP 4		DGP 1		DGP 2		DGP 3		DGP 4	
π	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald
0.2	4.4	4.9	4.0	4.5	5.0	4.9	5.1	5.1	5.3	5.1	4.4	4.4	4.4	5.2	6.1	5.8
0.8	5.2	4.9	4.1	4.7	4.8	5.6	5.1	5.7	3.8	5.0	4.2	4.7	4.4	5.7	5.3	5.5
1.4	5.5	4.9	5.4	5.0	5.3	5.9	5.4	4.9	4.8	5.3	5.5	5.0	4.6	4.9	6.5	5.2
2.0	5.4	5.1	5.5	4.8	4.8	5.1	5.4	5.0	4.2	4.7	4.4	5.0	6.1	4.7	4.9	5.4

Fixed Regressor MSE-F

	h = 1				h = 4			
π	DGP 1	DGP 2	DGP 3	DGP4	DGP 1	DGP 2	DGP 3	DGP 4
0.2	5.8	5.7	6.1	6.0	5.6	5.4	7.6	8.9
0.8	5.9	5.8	6.1	6.1	6.4	5.6	6.7	9.3
1.4	6.1	6.0	6.1	6.2	5.6	6.0	6.8	8.5
2.0	5.7	5.7	5.8	6.1	5.8	5.8	6.4	8.5

Table 3: Bootstrap rejection frequencies with two additional regressors in the larger model, $\alpha = 5\%$

$T = 50$

Moving Block

	h = 1								h = 4							
	DGP 1		DGP 2		DGP 3		DGP 4		DGP1		DGP 2		DGP 3		DGP 4	
π	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald
0.2	3.7	3.1	5.1	3.7	5.6	5.7	6.3	5.4	3.4	2.7	4.2	2.8	9.7	8.4	8.9	7.9
0.8	4.1	3.5	4.2	3.7	5.2	5.6	4.3	5.7	4.4	2.5	4.1	2.4	12.0	8.0	12.9	8.0
1.4	4.7	4.1	5.7	3.6	4.8	4.7	6.0	5.2	4.1	2.8	5.3	2.4	11.9	7.5	11.3	7.9
2.0	5.5	4.1	4.3	4.6	4.6	5.8	5.4	5.4	4.5	2.7	4.8	2.9	15.3	7.4	12.8	7.7

Fixed Regressor MSE-F

	h = 1				h = 4			
π	DGP 1	DGP 2	DGP 3	DGP4	DGP 1	DGP 2	DGP 3	DGP 4
0.2	5.8	6.4	6.5	7.0	6.5	6.5	10.3	21.1
0.8	7.2	7.4	8.4	8.2	6.5	7.1	11.5	24.6
1.4	7.4	7.1	8.0	7.5	7.4	7.1	11.7	26.0
2.0	6.7	6.9	7.2	8.2	6.5	7.0	12.4	28.2

$T = 100$

Moving Block

	h = 1								h = 4							
	DGP 1		DGP 2		DGP 3		DGP4		DGP 1		DGP 2		DGP 3		DGP 4	
π	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald
0.2	5.2	4.6	4.2	4.4	5.3	5.1	6.0	5.9	4.7	4.2	4.8	4.3	5.9	5.8	5.0	5.7
0.8	4.5	4.6	4.6	4.6	5.5	5.4	5.3	5.2	3.8	4.7	4.9	5.1	6.1	6.7	7.0	6.5
1.4	4.5	4.3	5.3	4.5	5.0	5.2	5.2	5.2	5.0	4.7	5.2	4.4	6.5	6.5	8.1	5.7
2.0	5.3	4.8	4.5	5.1	5.3	5.5	4.5	5.5	5.2	4.7	5.7	4.7	7.8	6.1	6.2	6.3

Fixed Regressor MSE-F

	h = 1				h = 4			
π	DGP 1	DGP 2	DGP 3	DGP4	DGP 1	DGP 2	DGP 3	DGP 4
0.2	6.2	5.8	6.6	6.4	6.5	6.1	8.9	13.7
0.8	6.3	6.3	7.1	6.7	6.1	6.0	7.5	13.5
1.4	6.6	6.1	6.9	6.7	6.9	6.3	8.1	13.5
2.0	5.9	6.0	6.2	6.2	6.5	6.7	8.2	14.0

$T = 200$

Moving Block

	h = 1								h = 4							
	DGP 1		DGP 2		DGP 3		DGP 4		DGP 1		DGP 2		DGP 3		DGP 4	
π	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald	$\mathscr{T}_T$	Wald
0.2	4.5	4.8	4.8	5.0	5.0	5.1	5.2	5.6	4.8	5.0	4.5	4.1	5.6	5.6	5.2	4.8
0.8	4.6	4.9	5.1	5.0	4.5	5.4	4.8	5.3	4.4	5.4	4.4	5.4	5.6	5.6	5.2	5.1
1.4	4.9	4.4	4.9	5.3	4.6	5.7	6.1	5.9	4.6	4.7	5.5	5.0	5.9	5.5	5.8	5.5
2.0	5.0	5.0	5.4	5.0	4.4	4.9	4.8	5.8	4.9	5.0	4.9	5.0	5.1	5.1	4.8	5.4

Fixed Regressor MSE-F

	h = 1				h = 4			
π	DGP 1	DGP 2	DGP 3	DGP4	DGP 1	DGP 2	DGP 3	DGP 4
0.2	5.7	5.4	6.2	6.1	5.1	5.4	7.9	9.0
0.8	6.0	6.1	6.3	6.1	5.9	5.9	6.9	9.3
1.4	5.8	5.9	6.1	6.1	6.0	6.0	6.5	9.1
2.0	6.1	5.8	6.1	6.1	5.8	5.6	6.4	9.2

3.3.2 Bootstrap test power

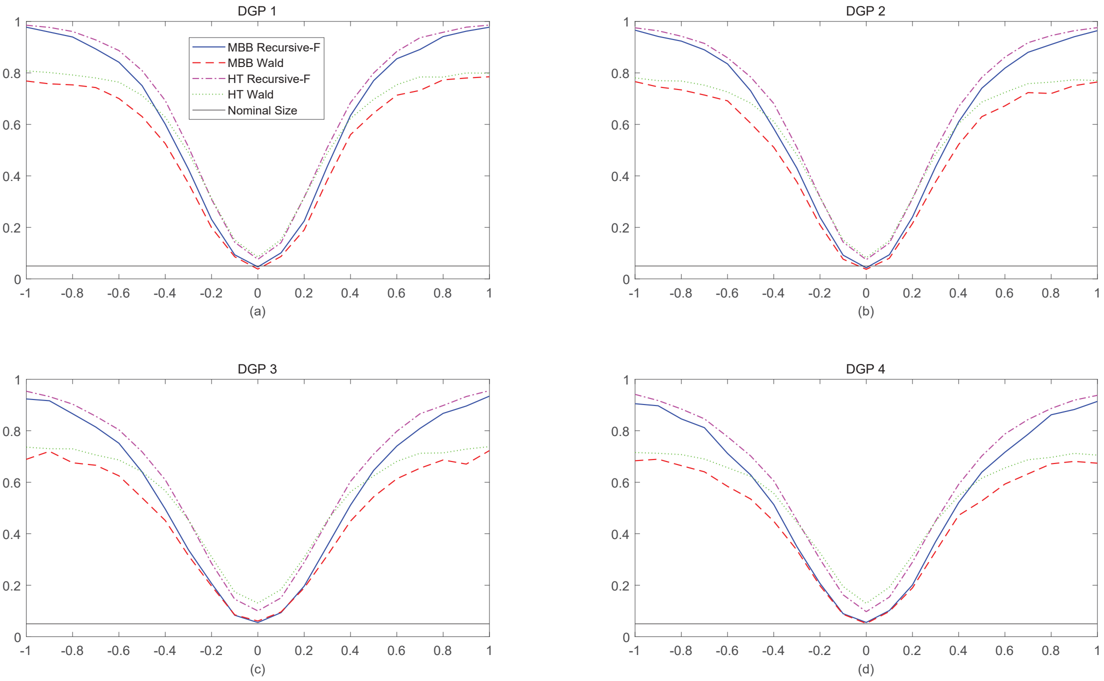
We now examine the power properties of the proposed moving block bootstrap $\mathscr{T}_T$ and $\mathscr{W}_T$ tests. To shorten the exposition, the results are shown for $T \in \{50, 200\}$ and for the cut-off point $\pi = 0.8$.

Figures 1-2 show the empirical power plots for 1-period-ahead forecasts (top four subfigures) and 4-period-ahead forecasts (bottom four subfigures). Each figure shows the empirical power of the asymptotic and the moving block bootstrap tests for both the recursive-F $\mathscr{T}_T$ and the Wald $\mathscr{W}_T$ test statistics.

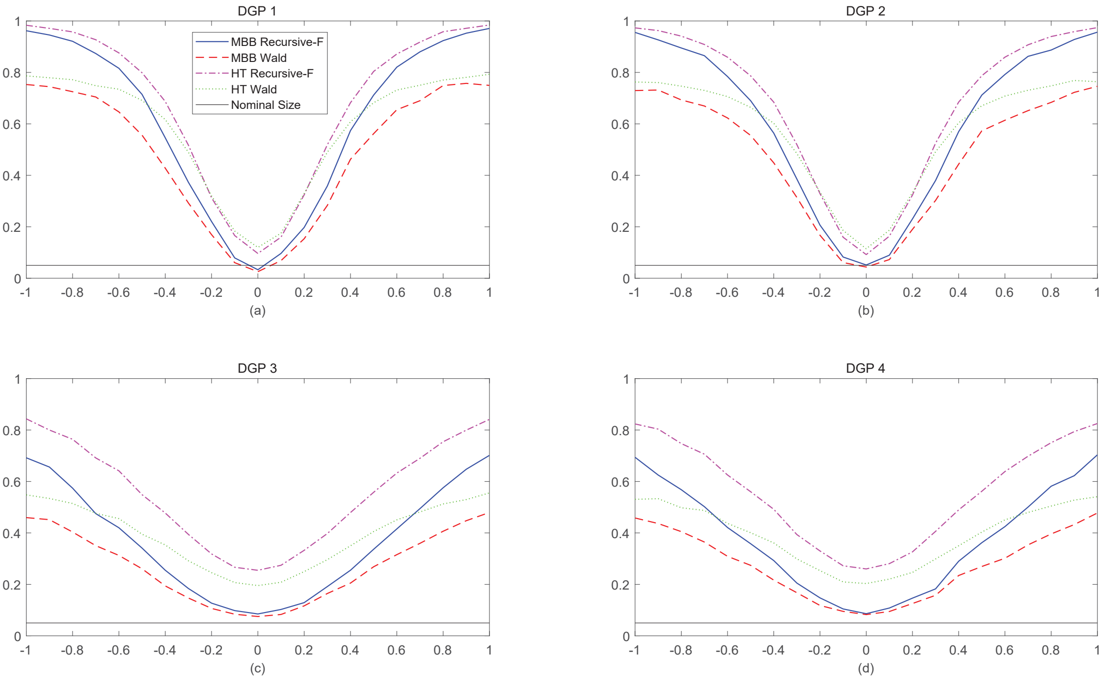
Several results stand out from these figures. First, the bootstrap empirical power is close to 1 even for moderate deviations from the null hypothesis when $T = 200$ (Figure 2), thus supporting the bootstrap consistency results in Theorem 3.2. As expected, both bootstrap tests converge faster in DGPs 1-2 than in DGPs 3-4. Also, the convergence seems faster in 1-step-ahead forecasts. Second, both tests have good power when $T = 50$ irrespective of the forecast horizon (Figure 1), and this is the case even for small deviations from the null hypothesis. Third, in all cases shown (DGPs, sample sizes, and forecast horizons), the moving block bootstrap test with $\mathscr{T}_T$ has an edge in terms of power over that with $\mathscr{W}_T$. One of the important contributions of Hansen and Timmermann's (2015) Wald approximation is that it facilitates the computation of the F-statistic critical values. Since our bootstrap method does not require simulating the asymptotic critical values of $\mathscr{T}_T$ from stochastic integrals of Brownian motions, it is more appealing than the Wald approximation. In addition, the critical values of the Wald statistic can only be simulated easily in specific cases (as discussed above in Section 2.2), while our bootstrap method applies in more general settings, including when the larger model has more than one extra predictor. Finally, the bootstrap tests do not suffer from loss of power relative to the asymptotic tests despite size correction. [11]

[11] Asymptotic power is slightly higher than bootstrap power due to the asymptotic size distortions.

Figure 1: Power of the moving block bootstrap tests, $T = 50$

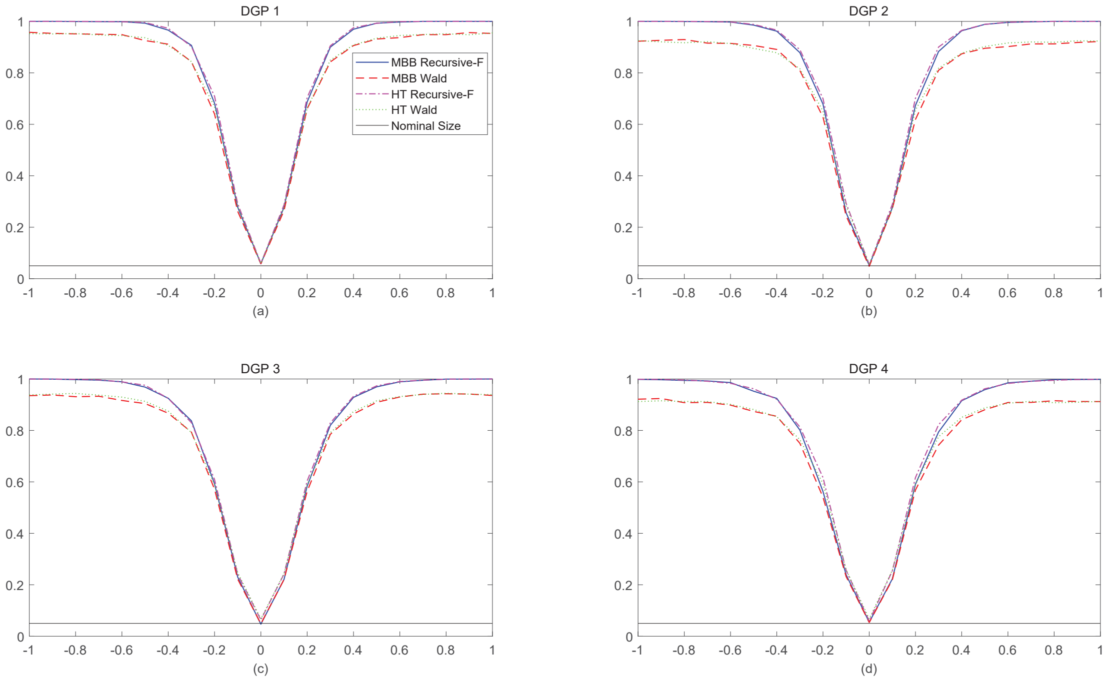


One-period ahead ($\tau = 1$) forecast

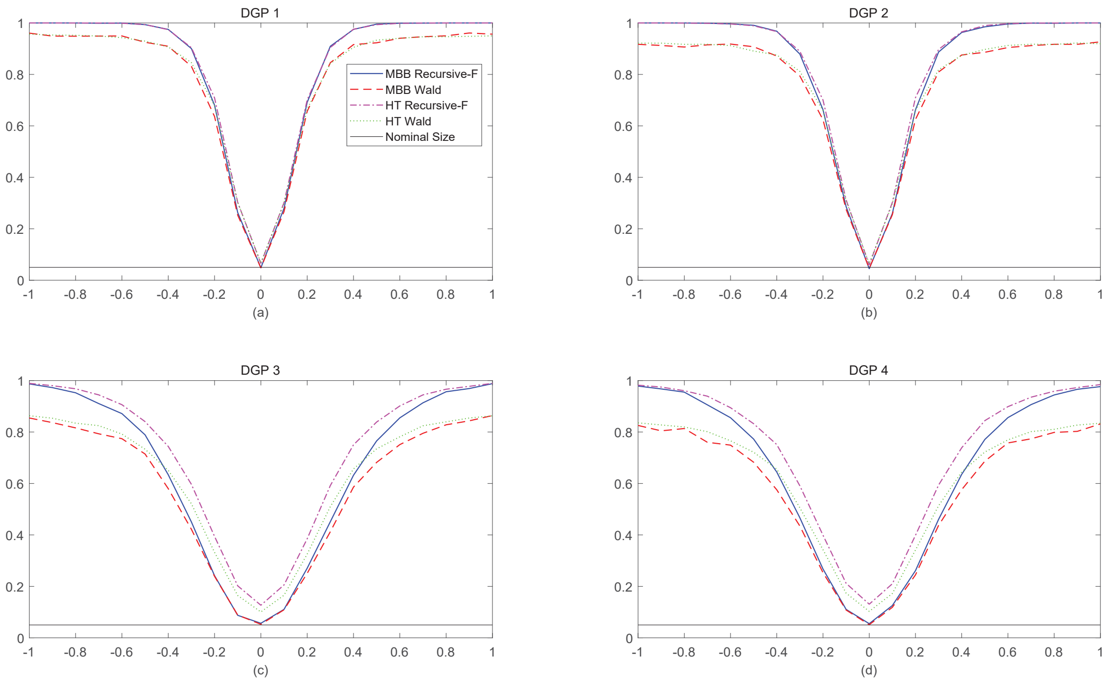


Four-period ahead ($\tau = 4$) forecast

Figure 2: Power of moving block bootstrap tests, $T = 200$



One-period ahead ($\tau = 1$) forecast



Four-period ahead ($\tau = 4$) forecast

4 Empirical applications

We illustrate our theoretical results through two applications. The first application examines the predictive ability of Chicago Fed National Activity Index (CFNAI) and other inflation measures for forecasting core PCE inflation (similar to Clark and McCracken, 2015). The second is drawn from Stock and Watson (2003) and Clark and McCracken (2012) and looks at forecasting quarterly U.S. GDP growth using a range of potential indicators.

4.1 Forecasting core inflation

In this application, we compare 1-quarter and 4-quarter ahead forecasts of inflation from two models. In the 1-quarter ahead forecasting exercise, the baseline (restricted) model relates the change in inflation at $t+1$ to current and one lagged value of inflation change, i.e.

$$y_{t+1} = b_0 + b_1 y_t + b_2 y_{t-1} + u_{y,t+1}, \tag{4.1}$$

where $y_t = \Delta\pi_t$, $\pi_t = 400ln(\frac{P_t}{P_{t-1}})$ and P_t is the aggregate price index at t. The alternative (unrestricted) model includes one lag of CFNAI, PCE food price inflation less core inflation, and import price inflation less core inflation, i.e.

$$y_{t+1} = b_0 + b_1 y_t + b_2 y_{t-1} + x_t' b_3 + u_{x,t+1}, \tag{4.2}$$

where x_t contains period t values of of CFNAI, PCE food price inflation less core inflation, and import price inflation less core inflation. In the 4-quarter ahead forecasting case, the baseline (restricted) model relates $y_{t+4}^{(4)} - y_t^{(4)}$ to a constant and $y_t^{(4)} - y_{t-4}^{(4)}$, $y_t^{(4)} = 100ln(\frac{P_t}{P_{t-4}})$. The alternative (unrestricted) model adds the period t values of CFNAI, relative food price inflation, and relative import price inflation to the baseline model.[12] In both cases, the data sample spans $1983:Q3$ through $2008:Q2$ ($T = 100$) and we use the cut-off point $\pi = 1.4$ for the initial estimation. Thus, out-of-sample forecasts from $1994:Q2 + \tau - 1$ through $2008:Q2$ ($\tau \in \{1, 4\}$) are obtained and the corresponding statistics computed. The bootstrap critical values are obtained with 9999 replications. The results are presented in Table 4. The first column of the table shows the variables included in the alternative model, while the other columns show for each forecast horizon the p-values of the proposed bootstrap test along with the *fixed regressor wild bootstrap*(FRWB) MSE-F test of Clark and McCracken (2015).

The main findings from this table can be summarized as follows. First, including CFNAI, PCE food price inflation and import price inflation do not improve forecast accuracy of core inflation at 1-quarter ahead horizon, while it does for 4-quarter ahead

[12]To simplify the lag structure, the relative food and import price inflation variables are computed as two-period averages of quarterly (relative) inflation rates; similar to Clark and McCracken (2015).

forecasts. Second, the p-value of the FRWB is 0.000 for 4-quarter ahead forecasting, while those of the MBB F- and Wald-statistics stand at 0.072 and 0.039 respectively. This means that using the FRWB leads to rejecting the baseline model at 1% nominal level, while the MBB $\mathscr{T}$ and $\mathscr{W}$ tests fail to reject the baseline model at 1% nominal level. In particular, the MBB $\mathscr{T}$ test even fails to reject the baseline model at the usual 5% nominal level.

Table 4: Test of equal accuracy for core inflation

One-quarter ahead: $\tau = 1$			
	MBB		Fixe regressor
Restricted variables	$\mathscr{T}_T$	$\mathscr{W}_T$	MSE-F
CFNAI, food, imports	0.181	0.958	0.314

Four-quarter ahead: $\tau = 4$			
	MBB		Fixed regressor
Restricted variables	$\mathscr{T}_T$	$\mathscr{W}_T$	MSE-F
CFNAI, food, imports	0.069	0.037	0.000

4.2 Forecasting real GDP growth

In this application, we examine the performance of 13 alternative models with respect to the baseline model in forecasting real GDP growth. As in the previous application, the comparison is done for τ-period ahead forecasts with $\tau \in \{1, 4\}$. The baseline model includes a constant and one lag of real GDP growth, where GDP growth between t and $t - \tau$ is measured as $y_t = (400/\tau)ln(GDP_t/GDP_{t-\tau})$, i.e.

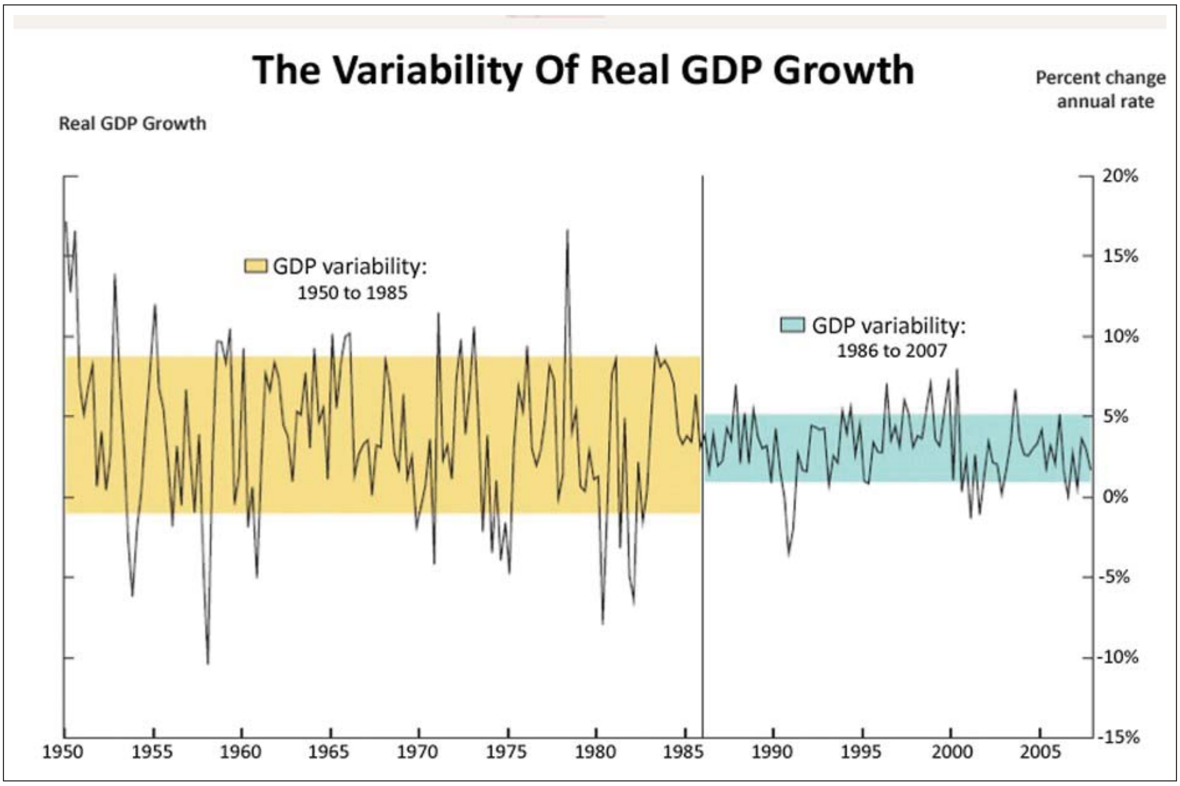
$$y_t \quad = \quad \beta_0 + \beta_1 y_{t-\tau} + u_{yt}, \tag{4.3}$$

while each of the 13 alternative models adds a potential leading indicator x_t to (4.3), i.e.

$$y_t \quad = \quad \beta_0 + \beta_1 y_{t-\tau} + \beta_2 x_{t-\tau} + u_{xt}, \tag{4.4}$$

where u_{xt} is an error term. The set of leading indicators used are shown in Table 5. It includes the change in consumption's share in GDP (measured with nominal data), weekly hours worked in manufacturing, building permits, purchasing manager indexes for supplier delivery times and orders, new claims for unemployment insurance, growth in real stock prices, change in 3-month Treasury bill rate, change in 1-year Treasury bond yield, change in 10-year Treasury bond yield, 3-month to 10-month yield spread, and spread between Aaa and Baa corporate bond yields from Moody's. The data span the period 1961 : $Q2$ through 2009 : $Q4$ ($T = 195$) and out-of-sample forecasts from 1981 : $Q4 + \tau - 1$ through 2009 : $Q4$ ($\tau \in \{1, 4\}$) are obtained and the corresponding statistics computed. Figure 3 shows the plot of the percentage change in real Gross Domestic Product from 1950 up to 2007. As seen in the figure, the volatility of the time series has altered during the period, in particular from the start of the Great Moderation, which may affect the performance of standard tests of predictive ability, including the fixed regressors MSE-F bootstrap test of Clark and McCracken (2015).

Figure 3: Quarterly real GDP growth



Note: The shaded areas of the chart show a common measure of data variability– plus and minus one standard deviation around the sample period of the data.
Source: U.S. Bureau of Economic Analysis.

The results are reported in Table 5.[13] The first column of the table shows the extra predictor added to the baseline model (thus determining alternative (or unrestricted) model), while the other columns show each test's p-values from the pairwise forecast comparisons. The top half of the table reports results for one-quarter ahead forecasts, while the bottom half shows results for four-quarter ahead forecasts.

Considering first the one-quarter ahead forecasts, we see that tests based on our MBB suggest that five models - those including change in consumption share, growth in building permits, growth in stock prices, Baa-Aaa interest rate spread and PMI new orders - improve the accuracy of forecasts relative to the benchmark AR(1) model.

[13] 9999 replications are used in computing the bootstrap critical values.

In addition to the above five models, the FRWB test finds that the alternative model that adds change in the 3-month Treasury bill rate to the baseline model also improves slightly the one-quarter ahead forecast of GDP growth (p-value of 2.6%), while our MBB fails to reject the baseline in that case at 10% nominal level (p-values of 19.3% and 47.6% for the MBB F and Wald tests respectively).

Next, looking at the four-quarter ahead forecasts, our MBB suggests that two models - those including growth in building permits and growth in stock prices - forecast better than the benchmark AR(1) model at 5% nominal level.[14] On the other hand, the FRWB also adds the models with change in consumption share and PMI orders to the above models in terms of their better forecast performance in comparison to the baseline model. According to our MBB, the higher predictive power of the change in consumption share and PMI orders seem to disappear in the four-quarter ahead forecasts even at 10% nominal level. Meanwhile, the FRWB fails to pick this up suggesting that the test over-rejects in some cases, which is in line with the Monte Carlo evidence reported earlier. Finally, weekly hours worked in manufacturing seems to exhibit a better predictive ability as the forecast horizon increases, though not statistically significant at the 10% level.

[14]In case of growth in stock prices, test based on recursive F statistics fails to reject the baseline at 10% nominal level, therefore providing mixed evidence regarding this variable.

Table 5: Test of equal accuracy for GDP

One-quarter ahead: $\tau = 1$

	MBB		Fixed regressor
Restricted variables	$\mathscr{T}_T$	$\mathscr{W}_T$	MSE-F
Δ (C/Y)	0.000	0.000	0.000
Δ ln Permits	0.002	0.000	0.000
Δ ln S & P 500	0.002	0.000	0.000
Spread, Baa-Aaa	0.067	0.085	0.072
PMI Orders	0.083	0.005	0.000
Unemployment claims	0.183	0.379	0.217
Δ 3-month treasury	0.193	0.476	0.026
Δ one-year treasury	0.233	0.280	0.451
Hours	0.342	0.531	0.420
PMI deliveries	0.331	0.647	0.905
Δ 10-year treasury	0.425	0.375	0.530
Spread, 10y - 3m	0.999	0.995	0.995
Spread, 10y - 1y	1.000	1.000	0.997

Four-quarter ahead: $\tau = 4$

	MBB		Fixed regressor
Restricted variables	$\mathscr{T}_T$	$\mathscr{W}_T$	MSE-F
Δ (C/Y)	0.203	0.266	0.059
Δ ln Permits	0.000	0.000	0.000
Δ ln S & P 500	0.187	0.016	0.000
Spread, Baa-Aaa	0.375	0.617	0.841
PMI Orders	0.300	0.123	0.011
Unemployment claims	0.336	0.391	0.307
Δ 3-month treasury	0.506	0.557	0.997
Δ one-year treasury	0.656	0.558	0.906
Hours	0.157	0.142	0.189
PMI deliveries	0.737	0.189	0.999
Δ 10-year treasury	0.933	0.864	0.889
Spread, 10y - 3m	0.999	0.998	0.999
Spread, 10y - 1y	1.000	1.000	0.998

5 Conclusion

In this paper we examine the finite-sample performance (size and power) of the recursively generated F-test of out-of-sample predictive accuracy (McCracken, 2007) and its equivalent Wald approximation (Hansen and Timmermann, 2015). We show through Monte Carlo experiments that even for moderate sample sizes, both tests can be oversized, especially when the forecast errors exhibit serial correlation and the forecast horizon is greater than one. We thus propose a bootstrap method for both statistics and establish its consistency, irrespective of the forecast horizon and the underlying data generating process. Interestingly, our bootstrap method applies even in cases where the larger model contains many extra predictors and the data generating process exhibit heteroscedasticity or serial correlation, situations under which the asymptotic critical values of the standard recursive F or Wald statistics are difficult to simulate.

The proposed bootstrap is an hybrid of a block moving bootstrap (which is nonparametric) and a residual based bootstrap (which is parametric). It is easy to implement, and we documented how to choose the block length in practice. In particular, we suggest that practitioners choose the block length that mimics the optimal lag length of the Newey and West's (1987) HAC estimator. Monte Carlo simulations show that the proposed bootstrap tests have an overall good finite-sample performance. The method is also illustrated with applications on forecasting core inflation and real GDP growth.

A Appendix

In this appendix, we present the supplemental lemmas and the proofs of the main results in the paper. In all this section, $\hat{\theta}_{j,t}, j = 1, 2$ is defined as $\hat{\theta}_{1,t} \equiv \hat{\beta}_{1,t}$ and $\hat{\theta}_{2,t} \equiv \tilde{\beta}_{2,t}$, where $\tilde{\beta}_{2,t} = (\hat{\beta}'_{1,t}, 0')'$.

A.1 Supplemental lemmas

This section establish some key results used in the main proofs of the paper.

Lemma A.1. *Suppose Assumptions 1-4 are satisfied. Then for any $r \in [\rho, 1]$, we have:*

(a) $\frac{1}{\sqrt{T}} \sum_{n=s}^{t} \widetilde{s}_{2n} \Longrightarrow B(r)$;

(b) $\frac{(T/t)}{\sqrt{T}} \sum_{n=s}^{t} \widetilde{s}_{2n} \Longrightarrow r^{-1}B(r)$;

(c) $\frac{(T/R_T)}{\sqrt{T}} \sum_{n=s+t-R_T}^{t} \widetilde{s}_{2t} \Longrightarrow \rho^{-1}[B(r) - B(r-\rho)]$,

where $\widetilde{s}_{2t}$ is defined in (2.10) and $B(r) = [B_1(r), \dots, B_{k_2}(r)]' \in \mathbb{R}^{k_2}$ is a vector of standard Brownian motions defined on $\mathbb{D}^{k_2}_{[0,1]}$.

Lemma A.2. *Let H_2 and J be given as in (2.10) and define $\Upsilon = \sigma_\varepsilon^{-1}(H_2^{-1} - JH_1^{-1}J')$. Suppose Assumptions 1-4 are satisfied. Then for any $r \in [\rho, 1]$, we have:*

(a) $\frac{(T/t)}{\sqrt{T}} \sum_{n=s}^{t} s_{2n} \Longrightarrow B\big(r^{-1}\Upsilon^{-1}\big)$;

(b) $\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\hat{\beta}_{2,t} - \beta_2^0) \Longrightarrow B\big(r^{-3}(1-\rho)H_2^{-1}\Upsilon^{-1}H_2^{-1}\big)$ *and* $\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\hat{\beta}_{1,t} - \beta_1^0) \Longrightarrow B\big(r^{-3}(1-\rho)H_1^{-1}J'\Upsilon^{-1}JH_1^{-1}\big)$,

where $\Upsilon^{-1} = \sigma_\varepsilon(H_2^{-1} - JH_1^{-1}J')^{-1}$ and for any $k_2 \times k_2$ matrix Σ, $B(\Sigma)$ is a k_2-dimensional Brownian motion having covariance Σ.

Lemma A.3. *Suppose Assumptions 1-4 are satisfied, and let $\hat{\beta}^*_{j,t}$ and $\hat{\theta}_{j,t}$ be define as in (3.3) and below Section A. If further $\ell_T = o(T^{1/2})$ as $T \to \infty$, then we have:*

$$\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\hat{\beta}^*_{j,t} - \hat{\theta}_{j,t})|\mathscr{X} \quad = \quad O_p(1), \; j = 1, 2,$$

i.e., the limiting distribution of $\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\hat{\beta}^*_{j,t} - \hat{\theta}_{j,t})$ is not centered at zero, but is rather characterized by a location bias.

Lemma A.4. *Suppose Assumptions 1-4 are satisfied, and let $\hat{\beta}^*_{j,t}$ and $\hat{\theta}_{j,t}$ be define as in (3.3) and below Section A. If further $\ell_T = o\big(T^{\frac{1}{2}}\big)$ as $T \to \infty$, then we have:*

$$\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\tilde{\beta}^*_{j,t} - \hat{\theta}_{j,t})|\mathscr{X} \quad = \quad o_p(1), \; j = 1, 2,$$

i.e., $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\tilde{\beta}^*_{j,t}-\hat{\theta}_{j,t})$ *is asymptotically unbiased.*

Lemma A.5. *Suppose Assumptions 1-4 are satisfied, and let* $\hat{\theta}_{j,t}$ *be define as below Section A. If further* $\ell_T = o\big(T^{\frac{1}{4}}\big)$ *as* $T \to \infty$*, then we have:*

$$\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}\left(\frac{1}{t}\sum_{n=s}^{t}\Big(h^*_{j,n-\tau}-\mu_T\sum_{n=s}^{T}h_{j,n-\tau}\Big)(\hat{\theta}_{j,t}-\beta_j^0)\right) \quad = \quad o_{p^*}(1)\ pr\text{-}\mathbb{P}\ \forall\ j=1,2,$$

where β_j^0 *is the true parameter value in* (2.4) *and* $\mu_T = 1/(T-s+1)$.

Proof of Lemma A.1. The proof is a direct application of the Functional Central Theorem result in Davidson (1994, corollary 29.19),[15] therefore it is omitted.

□

Proof of Lemma A.2. First, observe that As $\widetilde{s}_{2n} = \sigma_\varepsilon^{-1}\widetilde{A}H_2^{-1/2}s_{2n}$ from (2.10). Therefore we have

$$s_{2n} \quad = \quad \big[\sigma_\varepsilon^{-1}H_2^{-1/2}\widetilde{A}'\widetilde{A}H_2^{-1/2}\big]^{-1}H_2^{-1/2}\widetilde{A}'\widetilde{s}_{2n} = \Upsilon^{-1}H_2^{-1/2}\widetilde{A}'\widetilde{s}_{2n}. \qquad \text{(A.1)}$$

(**a**) From (A.1), we have

$$\frac{(T/t)}{\sqrt{T}}\sum_{n=s}^{t}s_{2n} \quad = \quad \frac{(T/t)}{\sqrt{T}}\sum_{n=s}^{t}\Upsilon^{-1}H_2^{-1/2}\widetilde{A}'\widetilde{s}_{2n} = \Upsilon^{-1}H_2^{-1/2}\widetilde{A}'\frac{(T/t)}{\sqrt{T}}\sum_{n=s}^{t}\widetilde{s}_{2n}$$
$$\Longrightarrow \quad r^{-1}\Upsilon^{-1}H_2^{-1/2}\widetilde{A}'B(r) \text{ from Lemma A.1}, \qquad \text{(A.2)}$$

and $r^{-1}\Upsilon^{-1}H_2^{-1/2}\widetilde{A}'B(r)$ is a linear transformation of a vector of standard Brownian motions, thus is also a Brownian motion. By observing that

$$var[r^{-1}\Upsilon^{-1}H_2^{-1/2}\widetilde{A}'B(r)] = r^{-2}\Upsilon^{-1}H_2^{-1/2}\widetilde{A}'(rI_{k_2})\widetilde{A}H_2^{-1/2}\Upsilon^{-1} \equiv r^{-1}\Upsilon^{-1},$$

it clear that $r^{-1}\Upsilon^{-1}H_2^{-1/2}\widetilde{A}'B(r)$ is distributed as $B\big(r^{-1}\Upsilon^{-1}\big)$.

(**b**) First, note that $\hat{\theta}_{j,t}-\beta_j^0 = \big(\frac{1}{t}\sum_{n=s}^{t}h_{j,n-\tau}\big)^{-1}\big(\frac{1}{t}\sum_{n=s}^{t}s_{j,n}\big) = H_j^{-1}\frac{1}{t}\sum_{n=s}^{t}s_{j,n} +$

[15] Also see McCracken (2007, Lemma A1).

$o_p(1)$. Hence we can approximate $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\theta}_{j,t}-\beta_j^0)$ as:

$$\begin{aligned}
\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\theta}_{j,t}-\beta_j^0) &= \frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}\frac{1}{\sqrt{T}}\frac{T}{\sqrt{T}}(\hat{\theta}_{j,t}-\beta_j^0) \\
&= H_j^{-1}\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}\frac{1}{\sqrt{T}}\frac{(T/t)}{\sqrt{T}}\sum_{n=s}^{t}s_{j,n}+o_p(P_T^{1/2}T^{-1/2}) \\
&= \sqrt{\frac{\pi}{1+\pi}}H_j^{-1}\frac{(T/t)}{\sqrt{T}}\sum_{n=s}^{t}s_{j,n}+o_p(1) \\
&= \sqrt{1-\rho}H_j^{-1}\frac{(T/t)}{\sqrt{T}}\sum_{n=s}^{t}s_{j,n}+o_p(1). \qquad (A.3)
\end{aligned}$$

For $j=2$, the last term of the RHS of (A.3) is such that $\sqrt{1-\rho}H_2^{-1}\frac{(T/t)}{\sqrt{T}}\sum_{n=s}^{t}s_{2,n}+o_p(1)\Longrightarrow\sqrt{1-\rho}r^{-1}H_2^{-1}\Upsilon^{-1}H_2^{-1/2}\widetilde{A}'B(r)$ by Lemma A.2-(**a**), which is distributed as $B\big(r^{-3}(1-\rho)H_2^{-1}\Upsilon^{-1}H_2^{-1}\big)$. The result for $j=1$ follows easily from (A.3) once we realise that $s_{1,n}=J's_{2,n}$ by the definition of J in (2.10).

□

Proof of Lemma A.3. Let $h^*_{j,n-\tau}=X^*_{j,n-\tau}X^{*'}_{j,n-\tau}$ and $s^*_{j,n}(\beta_j)=X^*_{j,n-\tau}(y_n-X^{*'}_{j,n-\tau}\beta_j)$ denote the bootstrap corresponding of $h_{j,n-\tau}$ and $s_{j,n}(\beta_j)$ in Assumptions 2-3. First, from the F.O.C of (3.3), we have for all $t=R_T,\ldots,T:$

$$\begin{aligned}
\frac{1}{t}\sum_{n=s}^{t}X^*_{j,n-\tau}(y^*_n-X^{*'}_{j,n-\tau}\hat{\beta}^*_{j,t}) &= 0=\frac{1}{t}\sum_{n=s}^{t}X^*_{j,n-\tau}[y^*_n-X^{*'}_{j,n-\tau}(\hat{\beta}^*_{j,t}-\hat{\theta}_{j,t}+\hat{\theta}_{j,t})] \\
\Leftrightarrow\ \frac{1}{t}\sum_{n=s}^{t}X^*_{j,n-\tau}X^{*'}_{j,n-\tau}(\hat{\beta}^*_{j,t}-\hat{\theta}_{j,t}) &= \frac{1}{t}\sum_{n=s}^{t}X^*_{j,n-\tau}(y^*_n-X^{*'}_{j,n-\tau}\hat{\theta}_{j,t}), \qquad (A.4)
\end{aligned}$$

so that we get

$$\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\beta}^*_{j,t}-\hat{\theta}_{j,t}) = \frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}\left(\frac{1}{t}\sum_{n=s}^{t}h^*_{j,n-\tau}\right)^{-1}\left(\frac{1}{t}\sum_{n=s}^{t}s^*_{j,n}(\hat{\theta}_{j,t})\right), \qquad (A.5)$$

for $j=1,2$. As the b_T blocks are drawn i.i.d. with probability $1/(T-s-\ell_T+1)$, we have

$$
\begin{aligned}
\frac{1}{t}\sum_{n=s}^{t}\mathbb{E}_{F^*}[s^*_{j,n}(\hat{\theta}_{j,t})|\mathscr{X}] &= \frac{1}{\ell_T(T-\ell_T+1)}\sum_{n=s}^{T-s-\ell_T}\sum_{p=1}^{\ell_T}s_{j,I_n+p}(\hat{\theta}_{j,t}) \\
&= \mu_T\sum_{n=s}^{T}s_{j,n}(\hat{\theta}_{j,t})+O_p(\ell_T T^{-1}), \qquad (A.6) \\
\frac{1}{t}\sum_{n=s}^{t}\mathbb{E}_{F^*}[h^*_{j,n-\tau}|\mathscr{X}] &= \frac{1}{\ell_T(T-s-\ell_T+1)}\sum_{n=s}^{T-\ell_T}\sum_{p=1}^{\ell_T}h_{j,I_n+p-\tau} \\
&= \mu_T\sum_{n=s}^{T}h_{j,I_n+p-\tau}+O_p(\ell_T T^{-1}), \qquad (A.7)
\end{aligned}
$$

where $\mu_T = 1/(T-s+1)$ and the I_n+p's are the uniform random variables defined in the bootstrap DGP, and the last equality in each of the equations (A.6) and (A.7) follows from the fact that the first and last ℓ_T observations have less likely to be drawn– e.g., Fitzenberger (1998, Lemma 1) and Politis and Romano (1994, Eq.(4)). As $\frac{1}{T-s+1}\sum_{n=s}^{T}h_{j,I_n+p-\tau} = H_j + o_p(1)$ for $j=1,2$, under Assumption 2, it is clear from (A.5) that

$$
\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\beta}^*_{j,t}-\hat{\theta}_{j,t}) = H_j^{-1}\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}\left(\frac{1}{t}\sum_{n=s}^{t}s^*_{j,n}(\hat{\theta}_{j,t})\right)+o_{p^*}(1)\ \text{pr-}\mathbb{P} \qquad (A.8)
$$

$$
\Rightarrow\ \mathbb{E}_{F^*}\left[\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\beta}^*_{j,t}-\hat{\theta}_{j,t})|\mathscr{X}\right] = H_j^{-1}\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}\frac{1}{t}\sum_{n=s}^{t}\mathbb{E}_{F^*}[s^*_{j,n}(\hat{\theta}_{j,t})|\mathscr{X}], \qquad (A.9)
$$

where $H_j^{-1} = (\mathbb{E}_F[h_{j,n-\tau}])^{-1}$. Now, we can write the first term of RHS of the last equality in (A.6) [scaled by $1/T$ instead of μ_T] as:

$$
\begin{aligned}
&\frac{1}{T}\sum_{n=s}^{T}s_{j,n}(\hat{\theta}_{j,t}) = \frac{1}{T}\sum_{n=s}^{T}X_{j,n-\tau}[y_n - X'_{j,n-\tau}(\hat{\theta}_{j,t}-\hat{\theta}_{j,t}+\hat{\theta}_{j,t})] \\
=\ & \frac{1}{T}\underbrace{\sum_{n=s}^{T}s_{j,n}(\hat{\theta}_{j,t})}_{=0\text{ from the F.O.C}} - \frac{1}{T}\sum_{n=s}^{T}X_{j,n-\tau}X'_{j,n-\tau}(\hat{\theta}_{j,t}-\hat{\theta}_{j,t}) \\
=\ & \underbrace{\left(\frac{1}{T}\sum_{n=s}^{T}h_{j,n-\tau}\right)}_{=O_p(1)}\underbrace{(\hat{\theta}_{j,t}-\hat{\theta}_{j,t})}_{=O_p(T^{-1/2})\text{ by Lemma A.1}} \\
=\ & O_p(1)O_p(T^{-1/2}) = O_p(T^{-1/2}). \qquad (A.10)
\end{aligned}
$$

Therefore, (A.10) entails that $\mathbb{E}_{F^*}\left[\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\beta}^*_{j,t}-\hat{\theta}_{j,t})|\mathscr{X}\right] = O_p(1)$ from (A.9) provided that $\ell_T = o(T^{1/2})$, as stated.

□

Proof of Lemma A.4. First, the F.O.C of the minimization problem in (3.4) is

$$\frac{1}{t}\sum_{n=s}^{t}\big(-2X^{*}_{j,n-\tau}(y^{*}_{n}-X^{*'}_{j,n-\tau}\tilde{\beta}^{*}_{j,t})+2\mu_T\sum_{n=s}^{T}s_{j,n}(\hat{\theta}_{j,t})\big)=0. \tag{A.11}$$

By proceeding as in (A.4)–(A.5), (A.11) implies that

$$\begin{aligned}\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\tilde{\beta}^{*}_{j,t}-\hat{\theta}_{j,t}) &= \frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}\left(\frac{1}{t}\sum_{n=s}^{t}h^{*}_{j,n-\tau}\right)^{-1}\left(\frac{1}{t}\sum_{n=s}^{t}\left[s^{*}_{j,n}(\hat{\theta}_{j,t})-\mu_T\sum_{n=s}^{T}s_{j,n}(\hat{\theta}_{j,t})\right]\right)\\ &= \frac{H_j^{-1}}{\sqrt{P_T}}\sum_{t=R_T}^{T}\left(\frac{1}{t}\sum_{n=s}^{t}\left[s^{*}_{j,n}(\hat{\theta}_{j,t})-\mu_T\sum_{n=s}^{T}s_{j,n}(\hat{\theta}_{j,t})\right]\right)+o_{p^*}(1)\text{ pr-}\mathbb{P}\end{aligned} \tag{A.12}$$

However, given the sample $\mathscr{X}$, we have $H_j^{-1}\left(\frac{1}{t}\sum_{n=s}^{t}\left[s^{*}_{j,n}(\hat{\theta}_{j,t})-\mu_T\sum_{n=s}^{T}s_{j,n}(\hat{\theta}_{j,t})\right]\right)=O_p(\ell_T T^{-1/2})$ pr-$\mathbb{P}$ from (A.6)–(A.10) in the proof of Lemma A.3. Thus we get

$$\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\tilde{\beta}^{*}_{j,t}-\hat{\theta}_{j,t})|\mathscr{X}=O_p(\ell_T T^{-1/2})\equiv o_p(1)\text{ if }\ell_T=o(T^{1/2}),$$

i.e., $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\tilde{\beta}^{*}_{j,t}-\hat{\theta}_{j,t})$ is asymptotically unbiased.

□

Proof of Lemma A.5. We have:

$$\begin{aligned}&\tfrac{1}{\sqrt{P_T}}\textstyle\sum_{t=R_T}^{T}\left(\frac{1}{t}\sum_{n=s}^{t}\left(h^{*}_{j,n-\tau}-\frac{1}{T-s+1}\sum_{n=s}^{T}h_{j,n-\tau}\right)(\hat{\theta}_{j,t}-\beta_j^0)\right)\\ \leq\;& \sup_{t\geq R_T}\sqrt{P_T}\left|\tfrac{1}{t}\textstyle\sum_{n=s}^{t}\left(h^{*}_{j,n-\tau}-\mu_T\sum_{n=s}^{T}h_{j,n-\tau}\right)(\hat{\theta}_{j,t}-\beta_j^0)\right|\\ \leq\;& \sup_{t\geq R_T}\tfrac{\sqrt{P_T}}{t^{1+\nu}}\left|\textstyle\sum_{n=s}^{t}\left(h^{*}_{j,n-\tau}-\mu_T\sum_{n=s}^{T}h_{j,n-\tau}\right)\right|\sup_{t\geq R_T}t^{\nu}\mid\hat{\theta}_{j,t}-\beta_j^0\mid\end{aligned} \tag{A.13}$$

with $1/3<\nu<1/2$. To establish Lemma A.5, it suffices to check that

$$\sup_{t\geq R_T}\frac{\sqrt{P_T}}{t^{1+\nu}}\left|\sum_{n=s}^{t}\left(h^{*}_{j,n-\tau}-\mu_T\sum_{n=s}^{T}h_{j,n-\tau}\right)\right|=o_{p^*}(1)\text{ pr-}\mathbb{P}$$

$$\text{and }\sup_{t\geq R_T}t^{\nu}\mid\hat{\theta}_{j,t}-\beta_j^0\mid=o_p(1)$$

under the conditions of the lemma. First, for any $\nu<1/2$, $\sup_{t\geq R_T}t^{\nu}\mid\hat{\theta}_{j,t}-\beta_j^0\mid=o_p(1)$ by West (1996, Lemma A3-(b)). Now, under Assumption 2, $\mu_T\sum_{n=s}^{T}h_{j,n-\tau}$ is a strong consistent estimator of $H_j=\mathbb{E}_F[h_{j,n-\tau}]$ by the Law of Iterative Logarithm (LIL), i.e., $\limsup_{T\to\infty}\left(\frac{T}{2lnlnT}\right)^{1/2}\left(\mu_T\sum_{n=s}^{T}h_{j,n-\tau}-H_j\right)=O_{as}(1)$– e.g., Lai and Wei (1983). Since $b_T\ell_T=T$ as $T\to\infty$, $\ell_T=o(T^{1/4})$, it is clear that $b_T/T^{3/4}\to\infty$ and $\left(\frac{b_T}{2lnlnb_T}\right)^{1/2}\to 0$ [note that the condition $\ell_T=o(T^{1/4})$ is necessary to have

$\left(\frac{b_T}{2lnlnb_T}\right)^{1/2} \to 0$]. Therefore, we also have

$$\sup_{t \geq R_T} \left| \frac{1}{t} \sum_{n=s}^{t} \left(h^*_{j,n-\tau} - \mu_T \sum_{n=s}^{T} h_{j,n-\tau} \right) \right| = O_{as^*}\left(\left(\frac{b_T}{2lnlnb_T} \right)^{1/2} \right), \text{a.s.-}\mathbb{P}$$

in the bootstrap world. As a result, we have

$$\sup_{t \geq R_T} \frac{\sqrt{P_T}}{t^{1+\nu}} \left| \sum_{n=s}^{t} \left(h^*_{j,n-\tau} - \mu_T \sum_{n=s}^{T} h_{j,n-\tau} \right) \right| = o_{p^*}(1) \text{ pr-}\mathbb{P}$$

for $\nu > 1/3$. Lemma A.5 then follows by combining the above results.

□

A.2 Proof of main results

Proof of Theorem 3.1. From (A.12) in the proof of Lemma A.4, we have:

$$\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\tilde{\beta}^*_{j,t} - \hat{\theta}_{j,t}) = \frac{H_j^{-1}}{\sqrt{P_T}} \sum_{t=R_T}^{T} \left(\frac{1}{t} \sum_{n=s}^{t} \left[s^*_{j,n}(\hat{\theta}_{j,t}) - \mu_T \sum_{n=s}^{T} s_{j,n}(\hat{\theta}_{j,t}) \right] \right) + o_{p^*}(1) \text{ pr-}\mathbb{P} \quad \text{(A.14)}$$

We can express $s^*_{j,n}(\hat{\theta}_{j,t})$ and $s_{j,n}(\hat{\theta}_{j,t})$ as:

$$s^*_{j,n}(\hat{\theta}_{j,t}) = X^*_{j,n-\tau}[y^*_n - X^{*\prime}_{j,n-\tau}(\hat{\theta}_{j,t} - \beta^0_j)] = s^*_{j,n}(\beta^0_j) - h^*_{j,n-\tau}(\hat{\theta}_{j,t} - \beta^0_j)$$

$$s_{j,n}(\hat{\theta}_{j,t}) = X_{j,n-\tau}[y_n - X'_{j,n-\tau}(\hat{\theta}_{j,t} - \beta^0_j)] = s_{j,n}(\beta^0_j) - h_{j,n-\tau}(\hat{\theta}_{j,t} - \beta^0_j)$$

so that (A.14) can be written as:

$$\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\tilde{\beta}^*_{j,t} - \hat{\theta}_{j,t}) = \frac{H_j^{-1}}{\sqrt{P_T}} \sum_{t=R_T}^{T} \left(\frac{1}{t} \sum_{n=s}^{t} \left[s^*_{j,n}(\beta^0_j) - \mu_T \sum_{n=s}^{T} s_{j,n}(\beta^0_j) \right] \right)$$
$$- \frac{H_j^{-1}}{\sqrt{P_T}} \sum_{t=R_T}^{T} \left(\frac{1}{t} \sum_{n=s}^{t} \left(h^*_{j,n-\tau} - \mu_T \sum_{n=s}^{T} h_{j,n-\tau} \right) (\hat{\theta}_{j,t} - \beta^0_j) \right) + o_{p^*}(1) \text{ pr-}\mathbb{P} \quad \text{(A.15)}$$

From Lemma A.5, the second term of the RHS of (A.15) is $o_{p^*}(1)$ pr-$\mathbb{P}$. Therefore we have

$$\frac{1}{\sqrt{P_T}} \sum_{t=R_T}^{T} (\tilde{\beta}^*_{j,t} - \hat{\theta}_{j,t}) = \frac{H_j^{-1}}{\sqrt{P_T}} \sum_{t=R_T}^{T} \left(\frac{1}{t} \sum_{n=s}^{t} \left[s^*_{j,n}(\beta^0_j) - \mu_T \sum_{n=s}^{T} s_{j,n}(\beta^0_j) \right] \right) + o_{p^*}(1) \text{ pr-}\mathbb{P} \quad \text{(A.16)}$$

Now, from Fitzenberger (1998, Lemma 1), we have

$$\frac{1}{t} \sum_{n=s}^{t} \mathbb{E}_{F^*}[s^*_{j,n}(\beta^0_j)] = \mu_T \sum_{n=s}^{T} s_{j,n}(\beta^0_j) + O_p\left(\ell_T T^{-1}\right),$$

thus we can express (A.16) as

$$\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\tilde{\beta}^*_{j,t}-\hat{\theta}_{j,t}) = \frac{H_j^{-1}}{\sqrt{P_T}}\sum_{t=R_T}^{T}\frac{1}{t}\sum_{n=s}^{t}\Big(s^*_{j,n}(\beta^0_j)-\mathbb{E}_{F^*}[s^*_{j,n}(\beta^0_j)]\Big)+o_{p^*}(1)\text{ pr-}\mathbb{P}\quad\text{(A.17)}$$

By following the same steps as in Lemma A.2, we can write the RHS of (A.17) as:

$$\frac{H_j^{-1}}{\sqrt{P_T}}\sum_{t=R_T}^{T}\frac{1}{\sqrt{T}}\frac{(T/t)}{\sqrt{T}}\sum_{n=s}^{t}\Big(s^*_{j,n}(\beta^0_j)-\mathbb{E}_{F^*}[s^*_{j,n}(\beta^0_j)]\Big)+o_{p^*}(1)\text{ pr-}\mathbb{P}$$

$$=\sqrt{1-\rho}H_j^{-1}\frac{(T/t)}{\sqrt{T}}\sum_{n=s}^{t}\Big(s^*_{j,n}(\beta^0_j)-\mathbb{E}_{F^*}[s^*_{j,n}(\beta^0_j)]\Big)+o_{p^*}(1)\text{ pr-}\mathbb{P}.$$

We deal with $j=2$ and $j=1$ separately. First, note as in the proof of Lemma A.2-(**b**) that for $j=2$, Lemma A.2-(**a**) along with the bootstrap sampling implies that

$$\sqrt{1-\rho}H_2^{-1}\frac{(T/t)}{\sqrt{T}}\sum_{n=s}^{t}\Big(s^*_{2,n}(\beta^0_2)-\mathbb{E}_{F^*}[s^*_{2,n}(\beta^0_2)]\Big) \Longrightarrow \sqrt{1-\rho}r^{-1}H_2^{-1}\Upsilon^{-1}H_2^{-1/2}\widetilde{A}'B^*(r) +o_{as^*}(1)\ a.s-\mathbb{P},$$

where $B^*(r)$ is a k_2 dimensional vector of standard brownian motion.

As $var\big[\sqrt{1-\rho}r^{-1}H_2^{-1}\Upsilon^{-1}H_2^{-1/2}\widetilde{A}'B^*(r)\big]=r^{-3}(1-\rho)H_2^{-1}\Upsilon^{-1}H_2^{-1}$,

it is clear that

$$\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\tilde{\beta}^*_{2,t}-\hat{\theta}_{2,t})\Longrightarrow B\big[r^{-3}(1-\rho)H_2^{-1}\Upsilon^{-1}H_2^{-1}\big]\ \ a.s-\mathbb{P},\qquad\text{(A.18)}$$

which is the distribution of $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\theta}_{2,t}-\beta^0_2)$ given in Lemma A.2. Similarly, for $j=2$, we find

$$\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\tilde{\beta}^*_{1,t}-\hat{\theta}_{1,t})\Longrightarrow B\big[r^{-3}(1-\rho)H_1^{-1}J'\Upsilon^{-1}JH_1^{-1}\big]\ \ a.s-\mathbb{P},\qquad\text{(A.19)}$$

which also is the distribution of $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\beta}_{1,t}-\beta^0_1)$ given in Lemma A.2.

Overall, this results show that $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\tilde{\beta}^*_{j,t}-\hat{\theta}_{j,t})$ converges almost surely to the asymptotic distribution of $\frac{1}{\sqrt{P_T}}\sum_{t=R_T}^{T}(\hat{\theta}_{j,t}-\beta^0_j)$ for all $j=1,2$, thus establishing Theorem **3.1**.

□

Proof of Theorem 3.2. $\mathscr{T}_T$ and $\mathscr{W}_T$ are asymptotically equivalent, it suffices to establish the result for $\mathscr{W}_T$. Also, as the MSE loss differential in the numerator of $\mathscr{T}_T$ is related to the homoskedastic Wald statistics [see Hansen and Timmermann

(2015, Corollary 1)] regardless of whether the underlying process is homoskedastic and regardless of whether the null hypothesis holds or not, we consider $\mathscr{W}_T = \widehat{S}_T - \widehat{S}_{R_T} + \sigma_\varepsilon^{-2}\check{\kappa}log(\rho)$ where $\hat{\sigma}_\varepsilon^2(\widehat{S}_T - \widehat{S}_{R_T})$ is equal to

$$\begin{aligned}\widetilde{S}_T - \widetilde{S}_{R_T} &= \check{U}'_{T,T}\check{H}_2^{-1}\check{U}_{T,T} - \frac{T}{R_T}\check{U}'_{T,R_T}\check{H}_2^{-1}\check{U}_{T,R_T} + o_p(1) \\ &\quad + \beta'_{22}\sum_{t=R_T}^{T} Z_{t-\tau}Z'_{t-\tau}\beta_{22} + 2\sqrt{T}\beta'_{22}(\check{U}_{T,T} - \check{U}_{T,R_T}), \qquad (A.20)\end{aligned}$$

$\widetilde{S}_T$ given in (2.8), $\check{U}_{T,T} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_{t-\tau}\varepsilon_{2t}$, $\widetilde{S}_{R_T}$, $\check{U}_{T,R_T}$ are the corresponding of both respectively in the sub-sample with R_T observations.

Now assume that $\beta_{22} = 0$. From the proof of Theorem 3 in Hansen and Timmermann (2015, p.2503),

$$\widetilde{S}_T - \widetilde{S}_{R_T} \xrightarrow{d} B(1)'\check{\Omega}_\infty^{1/2}\check{H}_2^{-1}\check{\Omega}_\infty^{1/2}B(1) - \rho^{-1}B(\rho)'\check{\Omega}_\infty^{1/2}\check{H}_2^{-1}\check{\Omega}_\infty^{1/2}B(\rho) \qquad (A.21)$$

where $\check{\Omega}_\infty$ is the log-run variance of the process $\{Z_{t-\tau}\varepsilon_{2t}\}$ and $B(r) \in \mathbb{D}_{[0,1]}^{k_2}$. Since $\hat{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 + o_p(1)$, then $\mathscr{W}_T$ is distributed as σ_ε^{-2} times the limit distribution of $\widetilde{S}_T - \widetilde{S}_{R_T}$ in (A.21). To establish the validity of the bootstrap for $\mathscr{W}_T$ when $\beta_{22} = 0$, it suffices to establish that $\mathscr{W}_T^*$ converges to σ_ε^{-2} times the limit distribution in (A.21) a.s.-$\mathbb{P}$.

First, note that $\mathscr{W}_T^* = \widehat{S}_T^* - \widehat{S}_{R_T}^* + \hat{\sigma}_\varepsilon^{*-2}\check{\kappa}^* log(\rho)$, and under H_0 along with the results of Lemmas A.1-A.5, we can express $\hat{\sigma}_\varepsilon^{*2}(\widehat{S}_T^* - \widehat{S}_{R_T}^*)$ as:

$$\begin{aligned}\widetilde{S}_T^* - \widetilde{S}_{R_T}^* &= \check{U}^{*\prime}_{T,T}\check{H}_2^{-1}\check{U}^*_{T,T} - \frac{T}{R_T}\check{U}^{*\prime}_{T,R_T}\check{H}_2^{-1}\check{U}^*_{T,R_T} \\ &\quad + \beta'_{22}\sum_{t=R_T}^{T} Z^*_{t-\tau}Z^{*\prime}_{t-\tau}\beta_{22} + 2\sqrt{T}\beta'_{22}(\check{U}^*_{T,T} - \check{U}^*_{T,R_T}) + o_{p^*}(1) \; pr - \mathbb{P}(A.22)\end{aligned}$$

where the various quantities in stars are the analogues to the ones in (A.20) in the bootstrap sample. It is easy to see from the model assumptions, along with the results of Lemmas A.1-A.5, that

$$\widetilde{S}_T^* - \widetilde{S}_{R_T}^* \xrightarrow{d^*} B^*(1)'\check{\Omega}_\infty^{1/2}\check{H}_2^{-1}\check{\Omega}_\infty^{1/2}B^*(1) - \rho^{-1}B^*(\rho)'\check{\Omega}_\infty^{1/2}\check{H}_2^{-1}\check{\Omega}_\infty^{1/2}B^*(\rho), \; a.s - \mathbb{P}(A.23)$$

under H_0, where $B^*(r) \in \mathbb{D}_{[0,1]}^{k_2}$. Since $B^*(r)$ in (A.23) and $B(r)$ in (A.21) have the same distribution, it is the case that (A.23) holds a.s.-$\mathbb{P}$ with $B^*(r)$ replaced by $B(r)$, i.e.

$$\widetilde{S}_T^* - \widetilde{S}_{R_T}^* \xrightarrow{d} B(1)'\check{\Omega}_\infty^{1/2}\check{H}_2^{-1}\check{\Omega}_\infty^{1/2}B(1) - \rho^{-1}B(\rho)'\check{\Omega}_\infty^{1/2}\check{H}_2^{-1}\check{\Omega}_\infty^{1/2}B(\rho), \; a.s - \mathbb{P}. \; (A.24)$$

Therefore $\mathscr{W}_T^*$ has the same asymptotic distribution as $\mathscr{W}_T$ a.s-$\mathbb{P}$ under H_0.

□

Proof of Theorem 3.3. Let $c^{\infty}_{0\cdot\mathscr{T}}(\alpha)$ and $c^{\infty}_{0\cdot\mathscr{W}}(\alpha)$ denote the $(1-\alpha)^{th}$ quantiles under H_0 of the asymptotic distributions of $\mathscr{T}_T$ and $\mathscr{W}_T$ respectively. We know from (2.11)-(2.12) that $c^{\infty}_{0\cdot\mathscr{T}}(\alpha) < \infty$ and $c^{\infty}_{0\cdot\mathscr{W}}(\alpha) < \infty$. From Theorem 3.2, we have

$$c^{*}_{\mathscr{T}}(\alpha) \xrightarrow{p} c^{\infty}_{0\cdot\mathscr{T}}(\alpha) < \infty, \; c^{*}_{\mathscr{W}}(\alpha) \xrightarrow{p} c^{\infty}_{0\cdot\mathscr{W}}(\alpha) < \infty \;\text{ as } T \to \infty. \tag{A.25}$$

From (A.20) and the model assumptions, it is not hard to show that if $\beta_{22} \neq 0$ is fixed, then $\widetilde{S}_T - \widetilde{S}_{R_T} \Longrightarrow \infty$ because the first two terms in the RHS of (A.20) are $O_p(1)$, while the last two terms diverge. As such, we also have $\widetilde{S}_T - \widetilde{S}_{R_T} \xrightarrow{d} \infty$ since weak convergence implies convergence in distribution. Since $\hat{\sigma}^2_{\varepsilon} = \sigma^2_{\varepsilon} + o_p(1)$ irrespective of whether the value of β_{22}, it follows that the above result implies that

$$\mathscr{W}_T \xrightarrow{d} \infty \;\text{ if } \beta_{22} \neq 0. \tag{A.26}$$

Theorem 3.3 follows by combining (A.25) and (A.26).

□

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Andrews, D. W. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*(4), 953–966.

Clark, T. E. and M. W. McCracken (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of econometrics 105*(1), 85–110.

Clark, T. E. and M. W. McCracken (2005). Evaluating direct multistep forecasts. *Econometric Reviews 24*(4), 369–404.

Clark, T. E. and M. W. McCracken (2012). Reality checks and comparisons of nested predictive models. *Journal of Business & Economic Statistics 30*(1), 53–66.

Clark, T. E. and M. W. McCracken (2014). Tests of equal forecast accuracy for overlapping models. *Journal of Applied Econometrics 29*(3), 415–430.

Clark, T. E. and M. W. McCracken (2015). Nested forecast model comparisons: a new approach to testing equal accuracy. *Journal of Econometrics 186*(1), 160–177.

Corradi, V. and N. R. Swanson (2007). Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. *International Economic Review 48*(1), 67–109.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians.* Oxford Uuniversity Press, Oxford.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*, Volume 38. Siam.

Fitzenberger, B. (1998). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics 82*(2), 235–287.

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

Hansen, P. and A. Timmermann (2012). Choice of sample split in out-of-sample forecast evaluation. Technical report, European University Institute.

Hansen, P. R. and A. Timmermann (2015). Equivalence between out-of-sample forecast comparisons and wald statistics. *Econometrica 83*(6), 2485–2505.

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. *Econometric Theory 18*(6), 1449–1459.

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

Lai, T. and C. Wei (1983). Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *Journal of multivariate analysis 13*(1), 1–23.

Mann, H. B. and A. Wald (1943). On stochastic limit and order relationships. *The Annals of Mathematical Statistics 14*(3), 217–226.

McCracken, M. W. (2007). Asymptotics for out of sample tests of granger causality. *Journal of Econometrics 140*(2), 719–752.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies 61*(4), 631–653.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical association 89*(428), 1303–1313.

Rossi, B. and A. Inoue (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics 30*(3), 432–453.

Stock, J. H. and M. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature 41*(3), 788–829.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica 64*(5), 1067–1084.

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