Estimating the Variance of a Combined Forecast: Bootstrap-Based Approach *

Ulrich Hounyo[†] **Kajal Lahiri**[‡]

September 23, 2020

Abstract

This paper considers bootstrap inference in model averaging for predictive regressions. We first show that a naïve bootstrap approach, which consists of stacking all residuals at time t into a vector, and then resampling these cross-sectional vectors of residuals over time is invalid in the context of model averaging. The naïve approach induces an unwarranted bias-related term in the bootstrap variance of averaging estimators. We then propose and justify two general fixed-design residual-based bootstrap resampling approaches for model averaging in predicting regressions. In a local asymptotic framework, we show the validity of the bootstrap in estimating the variance of a combined forecast and the asymptotic covariance matrix of a combined parameter vector with fixed weights. Our two proposed methods - the general blocking-based residual resampling and the general dependent wild-based residual resampling - can preserve nonparametrically the cross-sectional dependence between different models and the time series dependence in the errors simultaneously. The finite sample performance of these methods are assessed via Monte Carlo simulation. We illustrate our approach using an empirical study of the Taylor rule equation with 24 alternative specifications.

JEL Classification: C33, C53, C80.

Keywords: Bootstrap, Local asymptotic theory, Model average estimators, Wild bootstrap, Variance of consensus forecast.

*Please address correspondence to:

[†]Department of Economics, University at Albany – State University of New York, Albany, NY 12222; CREATES, Aarhus, Denmark; e-mail: khounyo@albany.edu.

[‡]Department of Economics, University at Albany – State University of New York, Albany, NY 12222; e-mail: klahiri@albany.edu.

1 Introduction

The idea of forecast combination was introduced by Bates and Granger (1969), extended by Granger and Ramanathan (1984), and spawned a large literature. For a recent overview of forecast combination literature, see Elliott and Timmermann (2016). Granger and Jeon (2004) introduced the concept of "thick modeling", which consists of making inference based on combined outputs from alternative models.

In this paper we use bootstrap to consistently estimate the variance of a combined forecast and the asymptotic covariance matrix of a weighted average of an estimated parameter vector using alternative models with *fixed weights*. Our theoretical framework follows Hansen (2014) and Liu and Kuo (2016) in generating forecasts by using weighted average of the predictions from a set of candidate models that vary by the choice of auxiliary regressors adopted by forecasters. Thus, there is a panel of forecasting models with different sets of predictors.

We first show that a naïve bootstrap approach, which consists of stacking all residuals at time t into a vector, and then resampling these cross-sectional vectors of residuals over time, is invalid in the context of model averaging.[1] Note that this naïve bootstrap approach is a common and natural way to preserve cross-sectional dependence and is valid in other contexts, see for example Maddala and Wu (1999), Gonçalves (2011) and Gospodinov and Ng (2013). See also the related work of Kilian and Lütkepohl (2017 cf. Ch 12) in the context of bootstrapping VAR models, among others. In our context of model averaging, the failure of this common approach is due to its inability to mimic appropriately the behavior of the regression residuals from the full model. Due to the omitted variable biases in approximating models, the naïve bootstrap approach induces an unwarranted additional term in the bootstrap variance of averaging estimators. We then propose and theoretically justify two alternative fixed-design residual-based bootstrap approaches for model averaging in predictive regressions. The two proposed methods, the general blocking-based residual resampling and the general dependent wild-based residual resampling, can preserve nonparametrically the cross-sectional dependence over different models and the time series dependence in the error term simultaneously.

Following Hjort and Claeskens (2003), Elliott et al. (2013), Hansen (2014), and Liu (2015), we study the asymptotic properties of averaging estimators in a local asymptotic framework, where the true regression coefficients associated with the auxiliary regressors are in a local $T^{-1/2}$ neighborhood of zero. This framework ensures the consistency of the averaging estimators, while, in general, it presents an asymptotic bias. We analyze the asymptotic distribution of averaging estimator with both fixed weights and data-dependent weights. As discussed in Liu (2015), we find that for the averaging estimator with fixed weights the asymptotic bias is a function of the local parameters, whereas the asymptotic variance is not. For the averaging estimator with data-dependent weights, both the asymptotic bias and the asymptotic variance are functions of the local parameters. Given

[1]Recently, Gonçalves and Perron (2020) show that a common approach of resampling cross-sectional vectors over time is invalid in the context of factor-augmented regressions with cross-sectional dependence among idiosyncratic errors.

that in the local asymptotic framework, the local parameters cannot be estimated consistently (see e.g., Liu (2015)), it is not possible to provide a consistent estimator of the asymptotic mean squared error (AMSE) of averaging estimator (with fixed weights and/or with data-dependent weights). So the bootstrap estimate of the AMSE will be inconsistent, under drifting sequence of parameters.

For this reason, in order to be able to carry out our bootstrap analysis, we focus only on the part of the AMSE of averaging estimator with fixed weights, which is consistently estimable, i.e., the asymptotic variance. Our results support the findings of Hjort and Claeskens (2003) (cf. Section 10.6), who showed that it is not possible to use bootstrap methods to consistently estimate the asymptotic distribution of averaging estimators. This is because, in the local asymptotic framework, the asymptotic distribution of the averaging estimator is function of the local asymptotic parameters which are not consistently estimable. Similarly, Liu (2015) showed that the asymptotic distribution of averaging estimator with data-dependent weights cannot be approximated by simulation. In a related work, Pötscher (2006) showed that the finite sample distribution of the averaging estimator cannot be consistently estimated. It should be pointed out that the proposed bootstrap approach analyzed in our paper is not for model selection purposes. Furthermore, the bootstrap theory presented in our paper (in a local asymptotic framework) is only applicable for averaging estimators based on *fixed weights*. Nevertheless, in professional forecasting, reporting equally weighted averages under the name "consensus forecasts" has been the norm rather than the exception, cf. Blue Chip Forecasts, the Survey of Professional Forecasters etc..

In this paper, we show that although bootstrapping methods do not work to estimate consistently the whole distribution of the weighted averaging estimator, it can be used to consistently estimate the variance of the estimator with fixed weights. We show the validity of the bootstrap in estimating the variance of a combined forecast and the asymptotic covariance matrix of the estimated combined parameter based on different models. We study and illustrate the general resampling residual-based bootstrap approaches for a blocking-based and a dependent wild-based method. Specifically, regression residuals are resampled by either the moving blocks bootstrap (MBB) of Künsch (1989) and Liu and Singh (1992), the non-overlapping block bootstrap (NBB) of Carlstein (1986), the dependent wild bootstrap (DWB) of Shao (2010), or the BEB method of Yeh (1998) and Shao (2011).

Gonçalves and White (2005) proved the consistency of the bootstrap covariance matrix estimator in a time series regression context, but without model averaging. Hansen and Racine (2018) propose a bootstrap model averaging procedure for testing unit roots. Recently Gonçalves et al. (2019) studied conditions under which block bootstrap can be used to obtain valid standard errors of parameters estimated via multi-stage QMLE estimators. In related work, Hahn and Liao (2019) studied the relation between bootstrap consistency and consistency of bootstrap standard errors.

The bagging, also known as bootstrap aggregation or bootstrap smoothing introduced by Breiman (1996), is a model-averaging device that reduces the variability and eliminates discontinuities of a combined predictor. Even though the bagging method uses bootstrap, it was originally introduced to improve the accuracy of the estimators – rather than to approximate the distributions or improve the confidence interval of predictions. See e.g., the work of Bühlmann and Yu (2002) and Inoue and Kilian (2008). Here, we are using the bootstrap to estimate the variance of a combined estimator with fixed weights based on different models.[2]

Our paper is organized as follows. Section 2 introduces the forecasting model, approximating models and review the asymptotic results. In Section 3, we introduce the bootstrap method and prove its consistency. Section 4 presents the simulation results. Section 5 provides an empirical illustration, reexamining the Taylor rule estimates reported by Granger and Jeon (2004), based on bagging using 24 alternative models. Finally, Section 6 concludes. The mathematical proofs are relegated to the Appendix.

2 Approximating Models

We consider the following h-step-ahead forecasting model

$$y_{t+h} = \mathbf{x}_t'\beta + \mathbf{z}_t'\gamma + e_{t+h} \equiv \mathbf{h}_t'\theta + e_{t+h},\ t = 1, \ldots, T-h, \tag{1}$$

$$E\left(\mathbf{h}_t e_{t+h}\right) = 0, \tag{2}$$

where $h = 1, 2, 3\ldots,$ is the forecast horizon, y_{t+h} is real-valued variable of interest, for example, inflation, GDP growth, unemployment rate and the like. $\mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{pt})'$ $(p \times 1)$ and $\mathbf{z}_t = (z_{1t}, z_{2t}, \ldots, z_{qt})'$ $(q \times 1)$ are vectors of predictors such that $\mathbf{h}_t = \left(h_{1t}, h_{2t}, \ldots, h_{(p+q)t}\right)' = (\mathbf{x}_t', \mathbf{z}_t')'$ $((p+q) \times 1)$, $\theta = (\beta', \gamma')'$ is the $((p+q) \times 1)$ vector of parameters and e_{t+h} is an unobservable error term. We allow e_{t+h} to be heteroskedastic and serially correlated (formal assumptions are given in Section 2.2).

We follow Liu (2015) and Liu and Kuo (2016), and interpret $\mathbf{x}_t$ and $\mathbf{z}_t$ as the core regressors and the auxiliary regressors, respectively. The core regressors $\mathbf{x}_t$ are of primary interest to researchers and must be included in the model, while the auxiliary regressors $\mathbf{z}_t$ may or may not included in the model. Then researchers want $\mathbf{x}_t$ in the model irrespective of the estimated t-ratios of the β-parameters, while they are less certain in including regressors $\mathbf{z}_t$. The auxiliary regressors could be lags of y_t, any nonlinear transformations of the original variables, or the interaction terms between the regressors, see e.g., Liu and Kuo (2016). As discussed in Magnus et al. (2010), Liang et al. (2011) and Liu (2015), the core regressors $\mathbf{x}_t$ may only include a constant term or even an empty matrix.

Suppose we have a set of N approximating models $\{i : 1, \ldots, N\}$ that are not necessarily nested. Each model uses a particular set of auxiliary regressors $\mathbf{z}_t^{(i)}$ $(q_i \times 1)$ (i.e., selects q_i regressors from the available set of auxiliary regressors) but all use the same core regressors $\mathbf{x}_t$. Let $\mathbf{\Pi}_i$ be a $q_i \times q$ selection matrix that selects the included (potentially relevant) predictors used in the ith model by the

[2] D'Agostino et al. (2012) use bootstrap approach to test equal forecast ability in an unbalanced panel of experts, without requiring imputation of missing observations.

forecaster. For example suppose that $q = 5$ and the ith model includes the following three auxiliary regressors: z_{1t}, z_{3t} and z_{4t}. Then, we have $q_i = 3$,

$$\mathbf{\Pi}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \text{ such that } \mathbf{\Pi}_i'\mathbf{\Pi}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The ith model includes all core regressors $\mathbf{x}_t$ and a subset of auxiliary regressors $\mathbf{z}_t^{(i)} = \mathbf{\Pi}_i \mathbf{z}_t$. The goal is to provide a h-step-ahead forecast of y_{T+h} or its conditional mean $y_{T+h|T} = E\left(y_{T+h}|\mathbf{h}_T, \mathbf{h}_{T-1}, \ldots\right) = \mathbf{x}_T'\beta + \mathbf{z}_T'\gamma = \mathbf{h}_T'\theta$, based on the core regressors $\mathbf{x}_t$, the selected subset of auxiliary regressors $\mathbf{z}_t^{(i)}$ and using the available data $\{\left(y_t, \mathbf{x}_t, \mathbf{z}_t^{(i)}\right) : 1, \ldots, T\}$ at time T. The i'th approximating model is

$$y_{t+h} = \mathbf{x}_t'\beta + \mathbf{z}_t^{(i)\prime}\gamma_i + e_{t+h}^{(i)} \equiv \mathbf{h}_t^{(i)\prime}\theta_i + e_{t+h}^{(i)}, \text{ for } i = 1, \ldots, N, \ t = 1, \ldots, T-h \tag{3}$$

where $\mathbf{h}_t^{(i)} = \left(\mathbf{x}_t', \mathbf{z}_t^{(i)\prime}\right)'$ is the selected regressors of dimension $((p+q_i) \times 1)$, $\theta_i = (\beta', \gamma_i')'$ is an $((p+q_i) \times 1)$ vector of coefficients and $e_{t+h}^{(i)}$ is the approximating error in the i'th model. Thus, the i'th model uses $p + q_i$ regressors. In matrix notation, (1) can be written as follows

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\gamma + \mathbf{e} \equiv \mathbf{H}\theta + \mathbf{e}, \tag{4}$$

where $\mathbf{y} = (y_{1+h}, \ldots, y_T)'$, $\mathbf{X} = \left(\mathbf{x}_1', \ldots, \mathbf{x}_{T-h}'\right)'$, $\mathbf{Z} = \left(\mathbf{z}_1', \ldots, \mathbf{x}_{T-h}'\right)'$, $\mathbf{e} = (e_{1+h}, \ldots, e_T)'$, and $\mathbf{H} = (\mathbf{X}, \mathbf{Z})$. Similarly, we write (3) in matrix notation as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}_i\gamma_i + \mathbf{e}^{(i)} \equiv \mathbf{H}_i\theta_i + \mathbf{e}^{(i)}, \tag{5}$$

where $\mathbf{Z}_i = \mathbf{Z}\mathbf{\Pi}_i' = \left(\mathbf{z}_1^{(i)\prime}, \ldots, \mathbf{z}_{T-h}^{(i)\prime}\right)'$, $\mathbf{H}_i = (\mathbf{X}, \mathbf{Z}_i)$ and $\mathbf{e}^{(i)} = \left(e_{1+h}^{(i)}, \ldots, e_T^{(i)}\right)'$.

Given (4) and (5), we can write

$$\mathbf{e}^{(i)} = \mathbf{H}\theta - \mathbf{H}_i\theta_i + \mathbf{e} = \mathbf{Z}\left(\mathbf{I}_q - \mathbf{\Pi}_i'\mathbf{\Pi}_i\right)\gamma + \mathbf{e}. \tag{6}$$

Following Hansen (2014), we can see equation (5) as having omitted variables. Let $\mathbf{I}$ denote an identity matrix and $\mathbf{0}$ a zero matrix. We also let

$$\mathbf{S}_i = \begin{pmatrix} \mathbf{I}_p & \mathbf{0}_{p \times q_i} \\ \mathbf{0}_{q \times p} & \Pi_i' \end{pmatrix}$$

be a selection matrix of dimension $(p+q) \times (p+q_i)$. We can also write $\theta_i = \mathbf{S}_i'\theta$, and similarly $\mathbf{H}_i = \mathbf{H}\mathbf{S}_i$. In the full model where all auxiliary regressors are included in the model (i.e., $q_i = q$), we have $\mathbf{\Pi}_i' = \mathbf{I}_q$, and the ordinary least-square (OLS) estimator of θ is

$$\hat{\theta} = \left(\mathbf{H}'\mathbf{H}\right)^{-1}\mathbf{H}'\mathbf{y} = \left(\hat{\beta}', \hat{\gamma}'\right)'. \tag{7}$$

The OLS estimator in the ith submodel is

$$\hat{\theta}_i = \left(\mathbf{H}_i'\mathbf{H}_i\right)^{-1}\mathbf{H}_i'\mathbf{y}, \tag{8}$$

whereas in the narrowest model (i.e., the smallest model among all possible submodels used by forecasters), $\Pi_i' = \mathbf{0}_q$, and the OLS estimator is given by

$$\hat{\theta}_i = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}. \tag{9}$$

The h-step-ahead point forecast of y_{T+h} from the ith approximating model is given by

$$\hat{y}^{(i)}_{T+h|T} = \mathbf{h}^{(i)\prime}_T\hat{\theta}_i = \mathbf{h}'_T\mathbf{S}_i\hat{\theta}_i. \tag{10}$$

We form with these individual forecasts $\hat{y}^{(i)}_{T+h|T}$, $i = 1, \ldots, N$ the $N \times 1$-dimensional vector $\hat{\mathbf{y}}_{T+h|T} = \left(\hat{y}^{(1)}_{T+h|T}, \ldots, \hat{y}^{(N)}_{T+h|T}\right)'$. We want to linearly combine these N forecasts using weights ω_i, $i = 1, \ldots, N$, such that $\omega = (\omega_1, \ldots, \omega_N)'$ is a weight vector in the unit simplex in $\mathbb{R}^N$,

$$\mathcal{W} = \{\omega \in [0,1]^N : \sum_{i=1}^{N}\omega_i = 1\}. \tag{11}$$

Model selection is the process of identifying which submodel is the best approximating model where the practitioner applies weight 1 to a particular single model ($\omega_i = 1$) and weight 0 to all other models. When many competing models are available for estimation, and without enough guidance from theory, model averaging may represent a feasible alternative to model selection. Forecast combination generalizes forecasting method when many competing forecasts are available from alternative models.

2.1 Combination of forecasts

Define the average forecast estimator of $y_{T+h|T}$ as

$$\hat{y}_{T+h|T}(\omega) = \omega'\hat{\mathbf{y}}_{T+h|T} = \sum_{i=1}^{N}\omega_i\hat{y}^{(i)}_{T+h|T} = \sum_{i=1}^{N}\omega_i\mathbf{h}'_T\mathbf{S}_i\hat{\theta}_i = \mathbf{h}'_T\hat{\theta}(\omega), \tag{12}$$

where

$$\hat{\theta}(\omega) = \sum_{i=1}^{N}\omega_i\mathbf{S}_i\hat{\theta}_i. \tag{13}$$

It follows that in approximating linear models, the combined forecast is the same as the forecast based on the weighted average of the parameter estimates across different models.

Some practitioners, who adopt the combination of forecasts approach, may choose *optimally* the weight ω by using a statistical procedure having known properties. For instance, one may select the forecast weights to minimize the asymptotic risk over the set of all possible forecast combinations. Alternatively, among many other choices, the mean square forecast error (MSFE) or the Mallows Model Averaging as in Hansen (2007, 2008) can be used to choose ω, resulting to a data-dependent weight, which may be random, cf. Elliott and Timmermann (2016, ch.14). In Section 2.2.2, we discuss the impact of using data-dependent weights on the variance of the averaging estimators $\hat{\theta}(\omega)$ and $\hat{y}_{T+h|T}(\omega)$.

One of our goals in this paper is to measure the uncertainty of the average forecast estimator defined in (12), for a given weight ω, whether optimal or not. In particular, we propose a bootstrap based-approach to compute the variance of the average forecast estimator $\hat{y}_{T+h|T}(\omega)$.

2.2 Assumptions and asymptotic results

We need to put some structure on the problem. Following Hjort and Claeskens (2003), Elliott et al. (2013), Hansen (2014), and the more recent work of Liu (2015), we examine the asymptotic distribution of $\hat{\theta}(\omega)$ and $\hat{y}_{T+h|T}(\omega)$ in a local asymptotic framework, where the parameters γ are in a root-T neighborhood of 0. More specifically, we make the following assumption.

Assumption 1. $\gamma = \gamma_T = \delta/\sqrt{T}$, *where δ is an unknown constant.*

Throughout, for a matrix $\mathbf{A}$, $\mathbf{A} > 0$ denotes $\mathbf{A}$ is positive definite. $\|\mathbf{A}\| = (trace(\mathbf{A}'\mathbf{A}))^{1/2}$ denotes the Euclidean norm. C represents a generic finite constant. We also impose the following assumption:

Assumption 2. (a) $\{(\mathbf{h}_t', e_{t+h})\}$ *is a strictly stationary and ergodic time series with finite $r > 4$ moments and $E(e_{t+h}|\mathcal{F}_t) = 0$, where $\mathcal{F}_t = \sigma(\mathbf{h}_t, \mathbf{h}_{t-1}, \ldots; e_t, e_{t-1}, \ldots)$.*

(b) $\mathbf{Q} = \lim_{T\to\infty} E\left(T^{-1}\mathbf{H}'\mathbf{H}\right) > 0$ *or equivalently* $\mathbf{Q} = \lim_{T\to\infty} E\left(T^{-1}\sum_{t=1}^{T-h}(\mathbf{h}_t\mathbf{h}_t')\right) > 0.$

(c) $\lim_{T\to\infty} Var\left(T^{-1/2}\mathbf{H}'\mathbf{e}\right) > 0$ *and* $\mathbf{\Omega} = \lim_{T\to\infty} \frac{1}{T}\sum_{s=1}^{T-h}\sum_{t=1}^{T-h} E\left(\mathbf{h}_s\mathbf{h}_t' e_{s+h}e_{t+h}\right) > 0.$

Assumption 1 ensures that the AMSE of the averaging estimators $\hat{\theta}(\omega)$ and $\hat{y}_{T+h|T}(\omega)$ remain finite. The $O\left(1/\sqrt{T}\right)$ ensures that both squared model biases and estimator variances have the same order $O(1/T)$. The least squares estimator (given by (9)) for the submodel has omitted variable bias. As we will see below (see equation (14)), by Assumption 1, $\sqrt{T}\left(\mathbf{S}_i\hat{\theta}_i - \theta\right)$ does not diverge despite the presence of the asymptotic bias.

Assumption 2 imposes moment conditions on $\{e_{t+h}\}$, $\{\mathbf{h}_t\}$ and the score vector $\{\mathbf{h}_t e_{t+h}\}$, and assume that data are strictly stationary. Assumption 2(a) is identical to Assumption 3.2' of Liu and Kuo (2016 cf. footnote 14). The latter is a modification of Assumption 3.2 of Liu and Kuo (2016) for h-step-ahead forecasting model. Assumption 2 is similar to Assumption R of Cheng and Hansen (2014), see also Assumption 5 of Djogbenou et al. (2015), and Assumption 5 of Gonçalves and Perron (2014). Assumption 2(a) implies that e_{t+h} is conditionally unpredictable at time t. As discussed by Cheng and Hansen (2014), when $h > 1$, it implies that e_{t+h} can be serially correlated. This is in line with the fact that for h-step-ahead forecasting model, the error e_{t+h} typically follows a moving average process of order $h-1$ (see e.g., Brown and Maital (1981) and Diebold, 2007, pp. 256-257). Assumption 2 is sufficient to imply that $T^{-1}\mathbf{H}'\mathbf{H} \xrightarrow{\mathbf{p}} \mathbf{Q}$ and $T^{-1/2}\mathbf{H}'\mathbf{e} \xrightarrow{\mathbf{d}} \mathbf{R} \sim \mathbf{N}(\mathbf{0}_{(p+q)\times 1}, \mathbf{\Omega})$.

Before stating the next results, it is convenient to introduce some more notations, which also will be needed later. We define

$$\mathbf{P}_i = p\lim_{T\to\infty}\mathbf{P}_{i,T} \text{ where } \mathbf{P}_{i,T} = \mathbf{S}_i\left(\frac{1}{T}\mathbf{S}_i'\mathbf{H}'\mathbf{H}\mathbf{S}_i\right)^{-1}\mathbf{S}_i', \text{ and } \mathbf{S}_0 = \begin{pmatrix} \mathbf{0}_{p\times q} \\ \mathbf{I}_q \end{pmatrix}.$$

Following the proof of Theorem 1 of Liu and Kuo (2016 cf. (A.1)), under Assumptions 1 and 2, as $T \to \infty$,

$$\sqrt{T}\left(\mathbf{S}_i\hat{\theta}_i - \theta\right) \xrightarrow{d} \mathbf{A}_i\delta + \mathbf{N}(\mathbf{0}_{(p+q)\times 1}, \mathbf{V}_{ii}) = \mathbf{A}_i\delta + \mathbf{P}_i\mathbf{R} \equiv \mathbf{\Lambda}_i, \tag{14}$$

where $\mathbf{A}_i = (\mathbf{P}_i\mathbf{Q} - \mathbf{I}_{p+q})\,\mathbf{S_0}$, and $\mathbf{V}_{ij} \equiv Cov\,(\mathbf{\Lambda}_i, \mathbf{\Lambda}_j) = \mathbf{P}_i\mathbf{\Omega}\mathbf{P}'_j$.

2.2.1 Distribution of averaging estimators with fixed weights

In this section we discuss the asymptotic distribution of averaging estimator with fixed weights. Given (13) and (14), it follows that under Assumptions 1 and 2, as $T \to \infty$,

$$\sqrt{T}\left(\hat{\theta}\,(\omega) - \theta\right) \xrightarrow{d} \mathbf{A}\,(\mathbf{w})\,\delta + \mathbf{N}(\mathbf{0}_{(p+q)\times 1}, \mathbf{V}\,(\mathbf{w})) = \sum_{i=1}^{N}\omega_i\mathbf{\Lambda}_i \equiv \mathbf{\Lambda}, \tag{15}$$

where

$$\mathbf{A}\,(\mathbf{w}) \equiv \sum_{i=1}^{N}\omega_i\mathbf{A}_i,$$

and

$$\mathbf{V}\,(\mathbf{w}) \equiv \mathbf{V}^{(1)}\,(\mathbf{w}) + \mathbf{V}^{(2)}\,(\mathbf{w}) = \sum_{i=1}^{N}\omega_i^2\mathbf{V}_{ii} + \sum_{i\neq j}\omega_i\omega_j\mathbf{V}_{ij}. \tag{16}$$

In (14), $\mathbf{A}_i\delta$ is the asymptotic bias that arises in estimating θ in model i, whereas when we use the weighted average of the parameter estimates across the different models $\hat{\theta}\,(\omega)$ to estimate θ, the asymptotic bias becomes $\mathbf{A}\,(\mathbf{w})\,\delta$, as given in (15). The asymptotic bias $\mathbf{A}_i\delta$ is nonzero for all possible models except the full model where all auxiliary regressors are included and such that $q_i = q$, $\Pi'_i = \mathbf{I}_q$, $\mathbf{S}_i = \mathbf{I}_{p+q}$ implying that $\mathbf{P}_i = \mathbf{Q}^{-1}$.

Given (15), it follows that the AMSE of the averaging estimator $\hat{\theta}\,(\omega)$ (based on fixed weights) is

$$\text{AMSE}\left(\hat{\theta}\,(\omega)\right) = \mathbf{A}\,(\mathbf{w})\,\delta\delta'\mathbf{A}'\,(\mathbf{w}) + \mathbf{V}\,(\mathbf{w})\,, \tag{17}$$

which is a function of the local parameter δ. As is well known, in the local asymptotic framework (see e.g., Liu (2015)), the local parameter δ cannot be consistently estimated. this implies that we cannot provide a consistent estimator of $\text{AMSE}\left(\hat{\theta}\,(\omega)\right)$. In particular, the bootstrap estimate of $\text{AMSE}\left(\hat{\theta}\,(\omega)\right)$ will be inconsistent.

Similarly, given (12) and (15), it follows that the asymptotic bias and variance of the combined forecast $\hat{y}_{T+h|T}\,(\omega)$ are $\mathbf{h}'_T\mathbf{A}\,(\mathbf{w})\,\delta$ and $\mathbf{\Sigma}_{y_{T+h|T}} \equiv \mathbf{h}'_T\mathbf{V}\,(\mathbf{w})\,\mathbf{h}_T$, respectively, implying that

$$\text{AMSE}\left(\hat{y}_{T+h|T}\,(\omega)\right) = \mathbf{h}'_T\mathbf{A}\,(\mathbf{w})\,\delta\delta'\mathbf{A}'\,(\mathbf{w})\,\mathbf{h}_T + \mathbf{h}'_T\mathbf{V}\,(\mathbf{w})\,\mathbf{h}_T = \mathbf{h}'_T\text{AMSE}\left(\hat{\theta}\,(\omega)\right)\mathbf{h}_T, \tag{18}$$

. Thus, the AMSE of the average forecast $\hat{y}_{T+h|T}\,(\omega)$ is also function of the local parameter δ and cannot be consistently estimated. For this reason, we focus on the estimation of the part of the AMSE of the averaging estimators $\hat{\theta}\,(\omega)$ and $\hat{y}_{T+h|T}\,(\omega)$ which are consistently estimable, i.e., the asymptotic variances $\mathbf{V}\,(\mathbf{w})$ and $\mathbf{\Sigma}_{y_{T+h|T}}$, respectively.

Note that the decomposition of the variance given in (16) has two components: the first component $\mathbf{V}^{(1)}(\mathbf{w})$ is a weighted average of the variances of the estimated parameter from each model and the second component $\mathbf{V}^{(2)}(\mathbf{w})$ is a weighted average of their covariances. As is evident in (6) the error $\mathbf{e}^{(i)}$ from each model has a common component $\mathbf{e}$, which drives the non-zero covariances across models. Our aim in this paper is to use bootstrap approach to consistently estimate the asymptotic variance $\mathbf{V}(\mathbf{w})$ and/or $\mathbf{\Sigma}_{y_{T+h|T}}$. As we show later, any valid bootstrap should mimic both components of $\mathbf{V}(\mathbf{w})$, as well as the behavior of the regression residuals from the full model. We accomplish this in Section 3.2.

2.2.2 Distribution of averaging estimators with data-dependent weights

The models chosen for the forecast combination often result in practice from model selection tests. Hence, in this section, we follow Claeskens and Hjort (2003) and Liu (2015) and study the asymptotic distributions of averaging estimators with data-dependent weights. Specifically, as in Claeskens and Hjort (2003) we assume that the weight is a smooth function of the asymptotic distribution of $\hat{\delta}$, where $\hat{\delta} = \sqrt{T}\hat{\gamma}$, such that $\hat{\gamma}$ is given in (7) and is the estimate from the full model. Before stating the asymptotic distribution of the averaging estimators, it is useful to state the distribution of $\hat{\delta}$. Given (14), under Assumptions 1 and 2, as $T \to \infty$, in the full model we have,

$$\sqrt{T}\left(\hat{\theta} - \theta\right) \xrightarrow{d} \mathbf{N}(\mathbf{0}_{(p+q)\times 1}, \mathbf{Q}^{-1}\mathbf{\Omega}\mathbf{Q}^{-1}) = \mathbf{Q}^{-1}\mathbf{R}, \tag{19}$$

implying that

$$\hat{\delta} = \sqrt{T}\hat{\gamma} \xrightarrow{d} \mathbf{R}_{\delta} = \delta + \mathbf{S}_{\mathbf{0}}'\mathbf{Q}^{-1}\mathbf{R}. \tag{20}$$

Next, let $\omega\left(i|\hat{\delta}\right)$ denote a data-dependent weight function for the ith model. As for the fixed weight case, we assume that for $i = 1, \ldots, N$, the weights $\omega\left(i|\hat{\delta}\right)$ take the values in the interval $[0, 1]$ and the sum of the weights is required to be one. Given (14), and following the proof of Theorem 6 of Liu (2015), if $\omega\left(i|\hat{\delta}\right) \xrightarrow{d} \omega\left(i|\mathbf{R}_{\delta}\right)$ and Assumptions 1 and 2 hold, as $T \to \infty$,

$$\sqrt{T}\left(\hat{\theta}(\omega) - \theta\right) = \sum_{i=1}^{N}\omega\left(i|\hat{\delta}\right)\sqrt{T}\left(\mathbf{S}_i\hat{\theta}_i - \theta\right) \xrightarrow{d} \sum_{i=1}^{N}\omega\left(i|\mathbf{R}_{\delta}\right)\left(\mathbf{A}_i\delta + \mathbf{P}_i\mathbf{R}\right) \equiv \mathbf{R}_1 + \mathbf{R}_2. \tag{21}$$

Hence, the asymptotic variance of the averaging estimator $\hat{\theta}(\omega)$ (based on data-dependent weights) is

$$Var\left(\mathbf{R}_1 + \mathbf{R}_2\right) = Var\left(\mathbf{R}_1\right) + Var\left(\mathbf{R}_2\right) + 2Cov\left(\mathbf{R}_1, \mathbf{R}_2\right),$$

which is function of the local parameter δ, where

$$\begin{aligned} Var(\mathbf{R}_1) &= \sum_{i=1}^{N} Var[\omega(i|\mathbf{R}_\delta)]\,\mathbf{A}_i\delta\delta'\mathbf{A}_i' + \sum_{i\neq j} Cov(\omega(i|\mathbf{R}_\delta),\omega(j|\mathbf{R}_\delta))\,\mathbf{A}_i\delta\delta'\mathbf{A}_j', \\ Var(\mathbf{R}_2) &= \sum_{i=1}^{N} Var[\omega(i|\mathbf{R}_\delta)\,\mathbf{P}_i\mathbf{R}] + \sum_{i\neq j} Cov(\omega(i|\mathbf{R}_\delta)\,\mathbf{P}_i\mathbf{R},\omega(j|\mathbf{R}_\delta)\,\mathbf{P}_j\mathbf{R}) \text{ and} \\ Cov(\mathbf{R}_1,\mathbf{R}_2) &= \sum_{i=1}^{N}\sum_{j=1}^{N} Cov[\omega(i|\mathbf{R}_\delta)\,\mathbf{A}_i\delta,\omega(j|\mathbf{R}_\delta)\,\mathbf{P}_j\mathbf{R}]. \end{aligned}$$

To gain further insight, let us consider a simple example where there is no core regressor $\mathbf{x}_t$ and $\mathbf{z}_t = 1$. Thus, we have $N = 2$ approximating models: the narrow model with no predictor ($\mathbf{\Pi}_i = 0$) and the full model having only a constant term. The OLS estimators in the narrow and full models are 0 and $\bar{y}_{T+h} = (T-h)^{-1}\sum_{t=1}^{T-h} y_{t+h}$, respectively. Consequently, the averaging estimator $\hat{\theta}(\omega) = \omega\left(i|\hat{\delta}\right)\bar{y}_{T+h}$, where we use the weights $1-\omega\left(i|\hat{\delta}\right)$ and $\omega\left(i|\hat{\delta}\right)$ for the narrow and full models, respectively. It follows that

$$\sqrt{T}\left(\hat{\theta}(\omega)-\theta\right) = \omega\left(i|\hat{\delta}\right)\sqrt{T}\bar{y}_{T+h} - \delta \xrightarrow{d} \omega(i|\mathbf{R}_\delta)\,\mathbf{R}_\delta - \delta,$$

such that

$$\mathbf{R}_\delta = \delta + \mathbf{R},\ \mathbf{R} \sim \mathbf{N}(\mathbf{0},\sigma_\infty^2) \text{ with } \sigma_\infty^2 = \lim_{T\to\infty}\frac{1}{T}\sum_{s=1}^{T-h}\sum_{t=1}^{T-h} E(e_{s+h}e_{t+h}). \tag{22}$$

Therefore, the asymptotic variance of the averaging estimator $\hat{\theta}(\omega)$ (based on data-dependent weights) is

$$\begin{aligned} Var(\omega(i|\mathbf{R}_\delta)\,\mathbf{R}_\delta - \delta) &= Var(\omega(i|\mathbf{R}_\delta)\,\mathbf{R}_\delta) \\ &= Var(\omega(i|\mathbf{R}_\delta))\,Var(\mathbf{R}_\delta) + Var(\omega(i|\mathbf{R}_\delta))\,E(\mathbf{R}_\delta)^2 + Var(\mathbf{R}_\delta)\,E(\omega(i|\mathbf{R}_\delta))^2, \end{aligned}$$

where the second equality uses the formula of variance of product of two random variables, and for simplicity, we assume that $Cov(\omega(i|\mathbf{R}_\delta),\mathbf{R}_\delta) = Cov\left(\omega^2(i|\mathbf{R}_\delta),\mathbf{R}_\delta^2\right) = 0$. Given (22), we have $Var(\mathbf{R}_\delta) = \sigma_\infty^2$ and $E(\mathbf{R}_\delta) = \delta$. Therefore, even in the (very simple) case where we impose that the mean and variance of the data-dependent weight $\omega(i|\mathbf{R}_\delta)$ are not function of δ, the asymptotic variance of the averaging estimator $\hat{\theta}(\omega)$ and $\hat{y}_{T+h|T}(\omega)$ will be function of the local parameter δ.

We emphasize that in contrast to the fixed weights case, the asymptotic variance of averaging estimators based on data-dependent weights is function of the local parameter δ. Under the local-to-zero assumption, the local parameter δ cannot be consistently estimated. thus we cannot provide a consistent estimator of the asymptotic variance of $\hat{\theta}(\omega)$ and $\hat{y}_{T+h|T}(\omega)$ when the weights are data-dependent. In particular, when the weights are data-dependent, in the local asymptotic framework, we cannot rely on bootstrapping to provide a consistent estimate of the asymptotic variances of weighted average estimators such as $\hat{\theta}(\omega)$ and $\hat{y}_{T+h|T}(\omega)$. This negative result is related to the finding in Hjort and Claeskens (2003) (cf. Section 10.6) regarding the invalidity of bootstrapping method on the weighted average of $\hat{\gamma}_i$ using data-dependent weights. In a drifting asymptotic framework using data-dependent weights (and likelihood-based model), Hjort and Claeskens (2003) argued that bootstrapping does not work because the asymptotic distribution of weighted average estimator is a function of the local parameter δ, and unfortunately, the estimator $\hat{\delta} = \sqrt{T}\hat{\gamma}$ does not go to δ in probability.

Given the impossibility to consistently estimate the asymptotic variances of $\hat{\theta}(\omega)$ and $\hat{y}_{T+h|T}(\omega)$ based on data-dependent weights, in the local asymptotic framework, we are interested in establishing valid bootstrap methods to compute the variances of $\hat{\theta}(\omega)$ and the average forecast estimator $\hat{y}_{T+h|T}(\omega)$ based on fixed (non-estimated) weights.

Notice that combination of forecasts based on fixed weights encompasses the equal-weighted ($\omega_i = 1/N$, $i = 1, \ldots, N$,) forecast combinations. Empirical studies often find a surprising result that simple equal-weighted forecast combinations perform very well compared with more sophisticated schemes that rely on estimated combination weights. Stock and Watson (1999) first reported this finding and called it "forecast combination" puzzle. Theoretical research during last 20 years has identified several reasons: (i) The gains from data-based combination weights critically depend to the heteroskedasticity and negative correlations in forecast errors between models; (ii) often bad models get weeded out, resulting in similar error variances and positive error covariances; (iii) errors introduced by the estimation of weights could overwhelm any gain from using optimal weights, and (iv) weights seldom stay the same and estimation of varying weights over the sample introduces more sampling variability. See e.g., Smith and Wallis (2009), Elliott and Timmermann (2016), Genre et al. (2013), and Lahiri et al. (2017).

3 Residual-based bootstrap inference

The goal of this section is to introduce and discuss bootstrap schemes that resample residuals in the model averaging context. Our proposed bootstrap methods resample the regression residuals [3] $\left\{\hat{e}^{(i)}_{t+h}\right\}$ over time $t = 1, \ldots, T-h$ for each model $i = 1, \ldots, N$. More specifically, we consider a fixed-design residual-based bootstrap procedure which takes the regressors in the sample as fixed, and apply an appropriate resampling method to the estimated residuals. The fixed-design (wild bootstrap) was originally suggested by Kreiss (1997), Hansen (2000) used a fixed-regressor bootstrap approach in the context of testing for structural change in regression models, whereas Gonçalves and Kilian (2004, 2007) studied fixed-design wild bootstrap for dynamic models (without model averaging). As usual, we will denote with asterisks quantities in the bootstrap world.

The regression residuals are

$$\hat{e}^{(i)}_{t+h} = y_{t+h} - \mathbf{h}^{(i)\prime}_t \hat{\theta}_i \text{ for } i = 1, \ldots, N, \ t = 1, \ldots, T-h. \tag{23}$$

[3] See also the recent work of Gonçalves and Perron (2014) and Djogbenou et al. (2015), who consider residuals-based bootstrap inference in factor-augmented regression context without model averaging; and the residual-based block bootstrap approach studied by Paparoditis and Politis (2003) and Carsten et al. (2015), in the context of unit root testing and multivariate cointegrated processes, respectively.

Let $\left\{e^{*(i)}_{t+h}, \quad t=1,\ldots,T-h\right\}$ denote a bootstrap sample from $\left\{\hat{e}^{(i)}_{t+h}, \quad t=1,\ldots,T-h\right\}$. We consider the following bootstrap DGP

$$y^{*(i)}_{t+h} = \mathbf{x}'_t\hat{\beta} + \mathbf{z}^{(i)\prime}_t\hat{\gamma}_i + e^{*(i)}_{t+h} \equiv \mathbf{h}^{(i)\prime}_t\hat{\theta}_i + e^{*(i)}_{t+h}, \text{ for } i=1,\ldots,N, \; t=1,\ldots,T-h. \tag{24}$$

We can equivalently write (24) as

$$\mathbf{y}^{*(i)} = \mathbf{H}_i\hat{\theta}_i + \mathbf{e}^{*(i)}, \tag{25}$$

where $\mathbf{y}^{*(i)} = \left(y^{*(i)}_{1+h},\ldots,y^{*(i)}_T\right)'$ and $e^{*(i)} = \left(e^{*(i)}_{1+h},\ldots,e^{*(i)}_T\right)'$. Next we refit the model using the fictitious response variables, and retain the bootstrap regression parameter estimator $\hat{\theta}^*_i$ analog of $\hat{\theta}_i$. In other words, based on the bootstrap dataset $\left\{\left(y^{*(i)}_{t+h}, \mathbf{h}^{(i)\prime}_t\right), \quad t=1,\ldots,T-h\right\}$ we compute $\hat{\theta}^*_i$. In particular, the bootstrap OLS estimator analog of $\hat{\theta}_i$ in the ith submodel is

$$\hat{\theta}^*_i = \left(\mathbf{H}'_i\mathbf{H}_i\right)^{-1}\mathbf{H}'_i\mathbf{y}^{*(i)}. \tag{26}$$

In the full model where all auxiliary regressors are included, the bootstrap OLS estimator analog of $\hat{\theta}$ is

$$\hat{\theta}^* = \left(\mathbf{H}'\mathbf{H}\right)^{-1}\mathbf{H}'\mathbf{y}^*, \text{ with } \mathbf{y}^* = \mathbf{H}\hat{\theta} + \mathbf{e}^*. \tag{27}$$

Note that because the residual-based bootstrap scheme used to generate $\mathbf{y}^{*(i)}$ is a *fixed-design*, we keep the regressors $\mathbf{H}_i$ fixed in the bootstrap regressions. Next, we can similarly compute the bootstrap analog of $\hat{y}^{(i)}_{T+h|T}$ given by (10) (i.e., the least-squares forecast of $y_{T+h|T}$ in model i) as follows

$$\hat{y}^{*(i)}_{T+h|T} = \mathbf{h}^{(i)\prime}_T\hat{\theta}^*_i. \tag{28}$$

Hence, the bootstrap average forecast estimator $\hat{y}^*_{T+h|T}(\omega)$ analog of $\hat{y}_{T+h|T}(\omega)$ is

$$\hat{y}^*_{T+h|T}(\omega) = \sum_{i=1}^{N}\omega_i\hat{y}^{*(i)}_{T+h|T} = \sum_{i=1}^{N}\omega_i\mathbf{h}'_T\mathbf{S}_i\hat{\theta}^*_i = \mathbf{h}'_T\hat{\theta}^*(\omega), \tag{29}$$

where

$$\hat{\theta}^*(\omega) = \sum_{i=1}^{N}\omega_i\mathbf{S}_i\hat{\theta}^*_i. \tag{30}$$

In the following, we let

$$\hat{\mathbf{a}}_T(\mathbf{w}) \equiv \sum_{i=1}^{N}\omega_i\hat{\mathbf{a}}_{i,T},$$

such that

$$\hat{\mathbf{a}}_{i,T} \equiv \sqrt{T}\left[\frac{1}{T}\mathbf{P}_{i,T}\mathbf{H}' - \left(\mathbf{H}'\mathbf{H}\right)^{-1}\mathbf{H}'\right]\mathbf{y}.$$

It is useful to rewrite $\sqrt{T}\left(\hat{\theta}^*(\omega)-\hat{\theta}\right)$ as:

$$\sqrt{T}\left(\hat{\theta}^*(\omega)-\hat{\theta}\right)=\underbrace{\sum_{i=1}^{N}\omega_i\hat{\mathbf{a}}_{i,T}}_{=\hat{\mathbf{a}}_T(\mathbf{w})}+\sum_{i=1}^{N}\omega_i\mathbf{P}_{i,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^{*(i)}\right) \tag{31}$$

(for further details, see equation (A.3) in the appendix).

In the following and throughout this paper, P^* (E^* and Var^*) denotes the probability measure (expected value and variance) induced by the bootstrap resampling, conditional on a realization of the original time series. In addition, for a sequence of bootstrap statistics Z_T^*, we write $Z_T^*=o_{P^*}(1)$ in probability, or $Z_T^*\to^{P^*}0$, as $n\to\infty$, in probability, if for any $\varepsilon>0$, $\iota>0$, $\lim_{T\to\infty}P\left[P^*\left(|Z_T^*|>\iota\right)>\varepsilon\right]=0$. Similarly, we write $Z_T^*=O_{P^*}(1)$ as $T\to\infty$, in probability if for all $\varepsilon>0$ there exists a $M_\varepsilon<\infty$ such that $\lim_{T\to\infty}P\left[P^*\left(|Z_T^*|>M_\varepsilon\right)>\varepsilon\right]=0$. Finally, we write $Z_T^*\to^{d^*}Z$ as $T\to\infty$, in probability, if conditional on the sample, Z_T^* weakly converges to Z under P^*, for all samples contained in a set with probability P converging to one.

Next, we let

$$\mathbf{V}^*(\mathbf{w})=p\lim_{T\to\infty}\mathbf{V}_T^*(\mathbf{w})\ \text{ where }\mathbf{V}_T^*(\mathbf{w})\equiv Var^*\left[\sqrt{T}\left(\hat{\theta}^*(\omega)-\hat{\theta}\right)\right],$$

and

$$\mathbf{V}_{ij,T}^*\equiv Cov^*\left[\mathbf{P}_{i,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^{*(i)}\right),\mathbf{P}_{j,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^{*(j)}\right)\right]. \tag{32}$$

Given (31) and (32), it follows that the bootstrap variance $\mathbf{V}_T^*(\mathbf{w})$ can be written as

$$\mathbf{V}_T^*(\mathbf{w})=\mathbf{V}_T^{*(1)}(\mathbf{w})+\mathbf{V}_T^{*(2)}(\mathbf{w})\equiv\sum_{i=1}^{N}\omega_i^2\mathbf{V}_{ii,T}^*+\sum_{i\neq j}\omega_i\omega_j\mathbf{V}_{ij,T}^*. \tag{33}$$

3.1 Failure of bootstrap methods that resample naïvely the whole vector of residuals over t

Let $\hat{\mathbf{e}}_t=\left(\hat{e}_t^{(1)},\ldots,\hat{e}_t^{(N)}\right)'$ denote an $(N\times 1)$-vector of residuals at time t from (all) models $i=1,\ldots,N$. As it is evident, we stack all residuals at time t into $\hat{\mathbf{e}}_t$. Our goal in this section is to show that a naïve application of the (fixed-design) residual-based bootstrap, which resamples the whole vector of regression residuals $\hat{\mathbf{e}}_t$ over t, fails to work in the context of model averaging. In particular, one cannot use a naïve bootstrap methods that resample $\hat{\mathbf{e}}_t$ to compute a consistent estimator of $\mathbf{V}(\mathbf{w})$ (i.e., the asymptotic variance covariance matrix of the weighted estimator $\hat{\theta}(\omega)$ given by (13)).

In this section, for simplicity we assume that $h=1$. We discuss the invalidity of two standard bootstrap methods applied on the regression residuals: the nonparametric i.i.d. bootstrap and the wild bootstrap (WB). The nonparametric i.i.d. bootstrap was first proposed by Efron (1979). The WB was originally developed by Wu (1986), Liu (1988) and Mammen (1993) in the context of static linear regression models with (unconditionally) heteroskedastic errors. Gonçalves and Kilian (2004) studied fixed-design and recursive-design WB for dynamic models, whereas Gonçalves and Kilian (2004) consider fixed-design WB for AR(∞) processes. Note that when $h=1$, under Assumption 2, e_{t+h} becomes a martingale difference sequence (m.d.s.), and, as a result, WB is an appropriate method to use.

It is well-known in the bootstrap literature that when dealing with a vector of correlated residuals for a given time period one should not treat these residuals as mutually independent when resampling, see e.g., Kilian and Lütkepohl (2017 cf. Ch 12) in the context of VAR models, among others. Similarly, in the context of panel data models with presence of cross-sectional dependence, in order to preserved cross-sectional dependence when resampling, Maddala and Wu (1999), Kapetanios (2008), and Gonçalves (2011) to name few, suggested to resample cross-sectional units as wholes rather than resampling within the units. See also the related works by Mark (1995), Rapach and Zhou (2013), Gospodinov and Ng (2013), Brüggemann, Jentsch and Trenkler (2016) and Montiel Olea and Plagborg-Moller (2020).

In our context, a naïve but "natural" way to preserve the contemporaneous correlation across model residuals is to stack all residuals at time t into a vector and resample over t, i.e., resample the whole vector $\hat{\mathbf{e}}_t$. As we will see below, this approach which is valid in other contexts, fails to work in the context of model averaging. The bootstrap sample from $\{\hat{\mathbf{e}}_{t+h}, \quad t=1,\ldots,T-h\}$ is $\left\{\mathbf{e}^*_{t+h}, \quad t=1,\ldots,T-h\right\}$ where $\mathbf{e}^*_{t+h}=\left(e^{*(1)}_{t+h},\ldots,e^{*(N)}_{t+h}\right)'$. For the WB, we let

$$\mathbf{e}^*_{t+h}=\hat{\mathbf{e}}_{t+h}\upsilon^*_{t+h}, \qquad t=1,\ldots,T-h, \tag{34}$$

where $\upsilon^*_{t+h}\sim$i.i.d.$(0,1)$ across t and such that $E^*\left|\upsilon^*_{t+h}\right|^{2+\varepsilon}<\infty$, for some $\varepsilon>0$.

The naïve application of Efron's i.i.d. bootstrap method applied on the vector $\hat{\mathbf{e}}_{t+h}$ of regression residuals generates at time $t+h$ the bootstrap residuals as:

$$\mathbf{e}^*_{t+h} \text{ i.i.d.} \sim \left\{\hat{\mathbf{e}}_{t+h}-\bar{\hat{e}}_{T-h}, \quad t=1,\ldots,T-h\right\}, \tag{35}$$

where $\bar{\hat{e}}_{T-h}=(T-h)^{-1}\sum_{t=1}^{T-h}\hat{\mathbf{e}}_{t+h}$. Note that resampling on the recentered residuals ensures that $E^*\left(\mathbf{e}^*_{t+h}\right)=\mathbf{0}_{N\times 1}$.

In the following, we let

$$\hat{b}^{(i)}_t=\hat{e}^{(i)}_t-\hat{e}_t, \tag{36}$$

where $\hat{e}^{(i)}_t=y_t-\mathbf{h}^{(i)\prime}_t\hat{\theta}_i$ and $\hat{e}_t=y_t-\mathbf{h}'_t\hat{\theta}$. We can rewrite $\hat{b}^{(i)}_t$ as follows:

$$\hat{b}^{(i)}_t=\mathbf{h}'_t\left[\mathbf{I}_{p+q}-\mathbf{P}_{i,T}\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)\right]\theta+\mathbf{h}'_t\left[\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)^{-1}-\mathbf{P}_{i,T}\right]\left(\frac{1}{T}\mathbf{H}'\mathbf{e}\right)\equiv\hat{b}^{(i)}_{t,1}+\hat{b}^{(i)}_{t,2}. \tag{37}$$

Notice that in the full model, we have $\hat{b}^{(i)}_t=0$. To present our results of invalidity of the WB and the nonparametric i.i.d. bootstrap applied to the whole vector $\hat{\mathbf{e}}_t$, as given in (34) and (35), respectively, it is helpful to observe that we can also rewrite $\mathbf{V}^*_{ij,T}$ as follows

$$\mathbf{V}^*_{ij,T} = \mathbf{P}_{i,T} \left[T^{-1} \sum_{t=1}^{T-h} \sum_{s=1}^{T-h} \mathbf{h}_s \mathbf{h}_t' Cov^* \left(e^{*(i)}_{s+h}, e^{*(j)}_{t+h} \right) \right] \mathbf{P}'_{j,T}, \tag{38}$$

(for further details, see equation (A.5) in the appendix). Next, note that for both methods: WB and i.i.d. WB, if $s \neq t$, we have $Cov^* \left(e^{*(i)}_{s+h}, e^{*(j)}_{t+h} \right) = 0$ (since $e^{*(i)}_{t+h}$ is independent across t conditionally on the observed time series), whereas if $s = t$, we have

$$Cov^* \left(e^{*(i)}_{s+h}, e^{*(j)}_{t+h} \right) = \begin{cases} \underbrace{\left(\hat{e}_t + \hat{b}^{(i)}_t \right)}_{=\hat{e}^{(i)}_t} \underbrace{\left(\hat{e}_t + \hat{b}^{(j)}_t \right)}_{=\hat{e}^{(j)}_t} & \text{if } \mathbf{e}^*_{t+h} \text{ is obtained by (34)} \\ c_{ij,T}, & \text{if } \mathbf{e}^*_{t+h} \text{ is obtained by (35)} \end{cases} \tag{39}$$

such that

$$c_{ij,T} = (T-h)^{-1} \sum_{t=1}^{T-h} \left(\hat{e}_t + \hat{b}^{(i)}_t \right) \left(\hat{e}_t + \hat{b}^{(j)}_t \right) - \left((T-h)^{-1} \sum_{t=1}^{T-h} \left(\hat{e}_t + \hat{b}^{(i)}_t \right) \right) \left((T-h)^{-1} \sum_{t=1}^{T-h} \left(\hat{e}_t + \hat{b}^{(j)}_t \right) \right).$$

Denote

$$\begin{aligned} c_{ij,1,T} &= (T-h)^{-1} \sum_{t=1}^{T-h} \hat{e}^2_t \\ c_{ij,2,T} &= (T-h)^{-1} \sum_{t=1}^{T-h} \hat{b}^{(i)}_{t,1} \hat{b}^{(j)}_{t,1} \\ c_{ij,3,T} &= \left((T-h)^{-1} \sum_{t=1}^{T-h} \hat{b}^{(i)}_{t,1} \right) \left((T-h)^{-1} \sum_{t=1}^{T-h} \hat{b}^{(j)}_{t,1} \right), \end{aligned}$$

and

$$\begin{aligned} c_{ij,1} &= \lim_{T\to\infty} \left[(T-h)^{-1} \sum_{t=1}^{T-h} E\left(\hat{e}^2_t \right) \right] \\ c_{ij,2} &= \lim_{T\to\infty} \left[(T-h)^{-1} \sum_{t=1}^{T-h} E\left(\hat{b}^{(i)}_{t,1} \hat{b}^{(j)}_{t,1} \right) \right] \\ c_{ij,3} &= \lim_{T\to\infty} \left[(T-h)^{-1} \sum_{t=1}^{T-h} E\left(\hat{b}^{(i)}_{t,1} \right) \right] \left[(T-h)^{-1} \sum_{t=1}^{T-h} E\left(\hat{b}^{(j)}_{t,1} \right) \right]. \end{aligned}$$

Theorem 3.1. *Suppose that Assumptions 1 and 2 hold.*

(a) *If $\mathbf{e}_{t+h}^{*}$ is obtained by (34), then we have*

$$\begin{aligned}\mathbf{V}_{ij,T}^{*} &= \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\underbrace{\left(\hat{e}_t+\hat{b}_t^{(i)}\right)}_{=\hat{e}_t^{(i)}}\underbrace{\left(\hat{e}_t+\hat{b}_t^{(j)}\right)}_{=\hat{e}_t^{(j)}}\right]\mathbf{P}_{j,T}' \\ &= \underbrace{\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\hat{e}_t^2\right]\mathbf{P}_{j,T}'}_{\to^P \mathbf{P}_i\boldsymbol{\Omega}\mathbf{P}_j'=\mathbf{V}_{ij}}+\widetilde{\mathbf{V}}_{ij,T}^{WB}+o_p(1),\end{aligned}$$

where

$$\widetilde{\mathbf{V}}_{ij,T}^{WB}=\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\hat{b}_{t,1}^{(i)}\hat{b}_{t,1}^{(j)}\right]\mathbf{P}_{j,T}'.$$

If in addition $\underset{T\to\infty}{p\lim}\widetilde{\mathbf{V}}_{ij,T}^{WB}=\widetilde{\mathbf{V}}_{ij}^{WB}$, *then*

$$\mathbf{V}_T^{*}(\mathbf{w})\to^P\mathbf{V}(\mathbf{w})+\widetilde{\mathbf{V}}^{WB}(\mathbf{w}),$$

as $T\to\infty$, where

$$\widetilde{\mathbf{V}}^{WB}(\mathbf{w})\equiv\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j\widetilde{\mathbf{V}}_{ij}^{WB}, \tag{40}$$

with

$$\widetilde{\mathbf{V}}_{ij}^{WB}=\mathbf{P}_i\lim_{T\to\infty}\left[T^{-1}\sum_{t=1}^{T-h}E\left[\mathbf{h}_t\mathbf{h}_t'\left[\theta'\left(\mathbf{I}_{p+q}-\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)\mathbf{P}_{i,T}'\right)\mathbf{h}_t\mathbf{h}_t'\left(\mathbf{I}_{p+q}-\mathbf{P}_{j,T}\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)\right)\theta\right]\right]\right]\mathbf{P}_j'.$$

(b) *If $\mathbf{e}_{t+h}^{*}$ is obtained by (35), then we have*

$$\begin{aligned}\mathbf{V}_{ij,T}^{*} &= c_{ij,T}\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\right]\mathbf{P}_{j,T}' \\ &= \underbrace{c_{ij,1,T}\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\right]\mathbf{P}_{j,T}'}_{\to^P c_{ij,1}\mathbf{P}_i\mathbf{Q}\mathbf{P}_j'}+\widetilde{\mathbf{V}}_{ij,T}^{i.i.d.B}+o_p(1),\end{aligned}$$

where

$$\widetilde{\mathbf{V}}_{ij,T}^{i.i.d.B}=\left(c_{ij,2,T}+c_{ij,3,T}\right)\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\right]\mathbf{P}_{j,T}'.$$

If in addition $\underset{T\to\infty}{p\lim}\widetilde{\mathbf{V}}_{ij,T}^{i.i.d.B}=\widetilde{\mathbf{V}}_{ij}^{i.i.d.B}$, *then*

$$\mathbf{V}_T^{*}(\mathbf{w})\to^P\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j c_{ij,1}\mathbf{P}_i\mathbf{Q}\mathbf{P}_j'+\widetilde{\mathbf{V}}^{i.i.d.B}(\mathbf{w}),$$

as $T \to \infty$, where

$$\widetilde{\mathbf{V}}^{i.i.d.B}(\mathbf{w}) \equiv \sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j\widetilde{\mathbf{V}}_{ij}^{i.i.d.B}, \tag{41}$$

with

$$\widetilde{\mathbf{V}}_{ij}^{i.i.d.B} = (c_{ij,2} + c_{ij,3})\,\mathbf{P}_i\mathbf{Q}\mathbf{P}_j'.$$

According to Theorem 3.1, one cannot use the naïve residual-based bootstrap method to approximate the asymptotic covariance matrix of a combined estimators, more specifically, $\underset{T\to\infty}{p\lim}\mathbf{V}_T^*(\mathbf{w}) \neq \mathbf{V}(\mathbf{w})$. The validity of any bootstrap method in the context of model averaging depends crucially on the ability of the bootstrap to allow consistent estimation of the asymptotic covariance matrix $\mathbf{V}(\mathbf{w})$. Bootstrap methods which resample naïvely the whole vector of regression residuals $\hat{\mathbf{e}}_t$ over t, fails to do so by not correctly mimicking the behavior of the regression residuals from the full model.

Remark 1. *The problem is not that the naïve residual-based bootstrap does not capture cross-sectional dependence of $\hat{e}_{t+h}^{(i)}$ over i. Rather the main problem is that it induces an additional term in the bootstrap variance (i.e., $\widetilde{\mathbf{V}}_{ij}^{WB}$ and $\widetilde{\mathbf{V}}_{ij}^{i.i.d.B}$ for the WB and the i.i.d. bootstrap, respectively), which should not be there. This additional term in the bootstrap variance is present, even in the simple context without model averaging, where we consider only one approximating model ($N = 1$), which is not the full model. Notice that in this latter simple case, the vector $\hat{\mathbf{e}}_{t+h}$ boils down to $\hat{e}_{t+h}^{(i)}$, which contains regression residuals from the full model $\hat{e}_{t+h}$ but also the term $\hat{b}_{t+h}^{(i)} \neq 0$ (see (36)). Although Theorem 3.1 considers two special cases: the WB and the i.i.d. bootstrap method that resample as in (34) and (35), respectively, the result extends to any bootstrap method that resamples the vector $\hat{\mathbf{e}}_t$ over t.*

As it is evident from Theorem 3.1, the term $\hat{b}_{t+h}^{(i)}$ (more precisely its component $\hat{b}_{t+h,1}^{(i)}$ (defined in (37))) drives the asymptotic behavior of the non-desirable additional term in the bootstrap variance. Furthermore, notice that if the regression residuals were resampled from the full model, then $\hat{b}_{t+h}^{(i)}$ would be identically zero, and consequently there will be no additional term in the bootstrap variance, i.e., $\widetilde{\mathbf{V}}_{ij}^{\mathrm{WB}} = 0$ and $\widetilde{\mathbf{V}}_{ij}^{\mathrm{i.i.d.B}} = 0$ for the WB and the i.i.d. bootstrap, respectively. Finally, note that for the i.i.d. bootstrap even if the regression residuals were resampled from the full model, such that we result with $\widetilde{\mathbf{V}}_{ij}^{\mathrm{i.i.d.B}} = 0$, the asymptotic limit of the bootstrap variance estimator $\mathbf{V}_T^*(\mathbf{w})$ would be $\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_i\omega_j c_{ij,1}\mathbf{P}_i\mathbf{Q}\mathbf{P}_j'$, the latter is equal to the asymptotic variance $\mathbf{V}(\mathbf{w})$ only when the error term is assume to be i.i.d. and homoscedastic.

Given the failure of the naïve residual-based bootstrap, which resamples the entire $(N \times 1)$-vector of regression residuals $\hat{\mathbf{e}}_t$ over t, we are interested in establishing valid bootstrap methods in this environment of combination of estimators.

3.2 General residual-based bootstrap approach for model averaging

For the bootstrap method to be valid in our framework, it should reproduce the three main characteristics of our model averaging simultaneously : (1) the possible serial correlation (dependence) in the error term $e_{t+h}^{(i)}$ over t (in particular, when $h > 1$), (2) the cross-sectional dependence of $e_{t+h}^{(i)}$ over i and (3) the behavior of the regression residuals from the full model. In general, when the forecasting horizon h is such that $h > 1$, the residuals $e_{t+h}^{(i)}$ will be correlated and may follow a moving average process (see e.g., Brown and Maital (1981) and Diebold, 2007, pp. 256-257).

When the forecasting horizon is larger than one, and the error term is correlated over t, it is well known in the bootstrap literature that one can capture time series dependence nonparametrically by applying blocking methods. For instance, the moving blocks bootstrap (MBB) of Künsch(1989) and Liu and Singh (1992), the nonoverlapping block bootstrap (NBB) of Carlstein (1986), and the stationary bootstrap (SB) of Politis and Romano (1994), among others, are suitable under these circumstances. However, as discussed in Remark 1, a naïve application of any blocking bootstrap method which resamples blocks of the vector of regression residuals $\hat{\mathbf{e}}_t$ over t will induce an additional term in the bootstrap variance.

We propose two general residual-based bootstrap approach for model averaging, which can preserve the cross-sectional dependence of $e_{t+h}^{(i)}$ over i, capture time series dependence nonparametrically in the error term, and at the same time mimick the behavior of the regression residuals from the full model (avoiding the additional term in the bootstrap variance). They are: (i) blocking-based residual resampling in model averaging, and (ii) dependent wild-based residual resampling in model averaging. These two general residual-based bootstrap methods resample a *recentered* version of the regression residuals $\left\{\hat{e}_{t+h}^{(i)}\right\}$.

3.2.1 Blocking-based residual resampling

The blocking-based residual resampling in model averaging applies blocking methods to the *recentered* version of the regression residuals by using *the same set of random draws* chosen by the blocking bootstrap method in all models $i = 1, \ldots, N$. Resampling the *recentered* version of regression residuals $\left\{\hat{e}_{t+h}^{(i)}\right\}$ over i with the same set of random indices is important to preserve the cross-sectional dependence of $e_{t+h}^{(i)}$ over different models $i = 1, \ldots, N$. To avoid the unwarranted result obtained for the naïve residual-based bootstrap in Theorem 3.1, we set the *recentered* regression residuals to be equal to the regression residuals from the full model.

More formally, in the following let $\ell = \ell_T \in \mathbb{N}$ $(1 \leq \ell < T - h)$ be a block length for a given block bootstrap. For simplicity, we assume that $(T-h)/\ell_T = k_T$ is an integer and denotes the number of blocks of size ℓ_T one have to draw. Let $\{\tau_t, t = 1, \ldots, T-h\}$ denote a sequence of random indices chosen by the blocking bootstrap taking values on $\{1, \ldots, T-h\}$. For instance for the MBB

$$\{\tau_t, t = 1, \ldots, T-h\} \equiv \{I_1 + 1, \ldots, I_1 + \ell, \ldots, I_k + 1, \ldots, I_k + \ell\}, \tag{42}$$

where I_j, $j = 1, \ldots, k$, are i.i.d. random variables distributed uniformly on $\{0, \ldots, T-h-\ell\}$. Note that $\ell = 1$ corresponds to the standard i.i.d. bootstrap. Similarly, for the NBB

$$\{\tau_t, t = 1, \ldots, T-h\} = \{J_1\ell + 1, \ldots, J_1\ell + \ell, \ldots, J_k\ell + 1, \ldots, J_k\ell + \ell\}, \tag{43}$$

where J_j are i.i.d. random variables distributed uniformly on $\{0, \ldots, k-1\}$.

Below, B is the number of bootstrap replications (e.g., $B = 999$). The steps for obtaining an estimator of the variance of a weighted average of parameter estimates across different models and/or an estimator of the variance of a combined forecast are as follows.

Algorithm 1. The general blocking-based residual resampling in model averaging.

1. For each model, $i = 1, \ldots, N$, fit the model and retain the fitted values and the residuals, as in (23).

2. Choose a blocking based bootstrap method (for instance MBB, NBB or SB) and a block length ℓ. Then, apply the blocking method and obtain the sequence of random indices $\{\tau_t, t = 1, \ldots, T-h\}$. For $i = 1$ (i.e., for the model used by forecaster number 1) the bootstrap residual samples is given by $e^{*(1)}_{t+h} = \hat{e}^{(1)}_{\tau_t+h} - \hat{b}^{(1)}_{\tau_t+h} - E^*\left(\hat{e}^{(1)}_{\tau_t+h} - \hat{b}^{(1)}_{\tau_t+h}\right) + \hat{b}^{(1)}_{(j-1)\ell+s+h}$, $t = 1, \ldots, T-h$. Store the set of index $\{\tau_t, t = 1, \ldots, T-h\}$. Next, for all the remaining models $i = 2, \ldots, N$, *use the same set of index* $\{\tau_t, t = 1, \ldots, T-h\}$ drawn for model 1 to obtain their bootstrap residual samples. More specifically, construct bootstrap residual samples as follows

$$\hat{e}^{*(i)}_{(j-1)\ell+s+h} = \underbrace{\hat{e}^{(i)}_{\tau_{(j-1)\ell+s}+h} - \hat{b}^{(i)}_{\tau_{(j-1)\ell+s}+h}}_{=\hat{e}_{\tau_{(j-1)\ell+s}+h}} - E^*\left(\hat{e}^{(i)}_{\tau_{(j-1)\ell+s}+h} - \hat{b}^{(i)}_{\tau_{(j-1)\ell+s}+h}\right) + \hat{b}^{(i)}_{(j-1)\ell+s+h}, \tag{44}$$

for $j = 1, \ldots, k, s = 1, \ldots, \ell$, $i = 1, \ldots, N$, where $E^*\left(\hat{e}^{(i)}_{\tau_{(j-1)\ell+s}+h} - \hat{b}^{(i)}_{\tau_{(j-1)\ell+s}+h}\right)$ is the bootstrap expected value of the resampling version in *the full model* of the raw regression residuals $\left\{\hat{e}_{(j-1)\ell+s+h}\right\}$. For example with the MBB,

$$\begin{aligned} E^*\left(\hat{e}^{(i)}_{\tau_{(j-1)\ell+s}+h} - \hat{b}^{(i)}_{\tau_{(j-1)\ell+s}+h}\right) &= \frac{1}{T-h-\ell+1}\sum_{j=1}^{T-h-\ell+1}\left(\hat{e}^{(i)}_{j-1+s+h} - \hat{b}^{(i)}_{j-1+s+h}\right) \\ &= \frac{1}{T-h-\ell+1}\sum_{j=1}^{T-h-\ell+1}\hat{e}_{j-1+s+h}, \end{aligned}$$

whereas for the NBB,

$$E^*\left(\hat{e}^{(i)}_{\tau_{(j-1)\ell+s}+h} - \hat{b}^{(i)}_{\tau_{(j-1)\ell+s}+h}\right) = \frac{1}{k}\sum_{j=1}^{k}\left(\hat{e}^{(i)}_{(j-1)\ell+s+h} - \hat{b}^{(i)}_{(j-1)\ell+s+h}\right) = \frac{1}{k}\sum_{j=1}^{k}\hat{e}_{(j-1)\ell+s+h}$$

for $i = 1, \ldots, N$.

3. Formulate the bootstrap version of y_{t+h} as in (24).

4. Compute $\hat{\theta}^*_i$, $\hat{y}^{*(i)}_{T+h|T}$, $\hat{\theta}^*(\omega)$ and $\hat{y}^*_{T+h|T}(\omega)$ as given by (26) and (28), (30) (29), respectively.

5. Repeat steps 2, 3 and 4 B times, resulting in statistics:

$$\left\{\hat{\theta}^{*1}(\omega), \ldots, \hat{\theta}^{*B}(\omega)\right\} \text{ and/or } \left\{\hat{y}^{*1}_{T+h|T}(\omega), \ldots, \hat{y}^{*B}_{T+h|T}(\omega)\right\},$$

then store the values of $\hat{\theta}^{*b}(\omega)$ and $\hat{y}^{*b}_{T+h|T}(\omega)$, $b = 1, \ldots, B$.

6. As will be shown shortly, the bootstrap variance estimator $\mathbf{V}^{*}_{T}(\mathbf{w})$ of the weighted average of the parameter estimates across the different models can be evaluated by simulation using

$$T\frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}^{*b}(\omega) - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b}(\omega)\right)\left(\hat{\theta}^{*b}(\omega) - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b}(\omega)\right)', \tag{45}$$

where $B = \infty$ in theory. In practice, $B = 999$ tends to provide a reasonable approximation.

Similarly, the bootstrap variance estimator of the average forecast can be evaluated by simulation

$$Var^{*}\left(\sqrt{T}\hat{y}^{*}_{T+h|T}(\omega)\right) = T\frac{1}{B}\sum_{b=1}^{B}\left(\hat{y}^{*b}_{T+h|T}(\omega) - \frac{1}{B}\sum_{b=1}^{B}\hat{y}^{*b}_{T+h|T}(\omega)\right)^{2}. \tag{46}$$

Alternatively one can also use

$$Var^{*}\left(\sqrt{T}\hat{y}^{*}_{T+h|T}(\omega)\right) = \mathbf{h}'_{T}Var^{*}\left(\sqrt{T}\hat{\theta}^{*}(\omega)\right)\mathbf{h}_{T}, \tag{47}$$

where $Var^{*}\left(\sqrt{T}\hat{\theta}^{*}(\omega)\right)$ is computed by using (45).

Note that the centering of the bootstrap sample in (44) ensures that in the full model $E^{*}\left(e^{*}_{t+h}\right) = 0$ (whereas in the model i, we have $E^{*}\left(e^{*(i)}_{t+h}\right) = \hat{b}^{(i)}_{t+h}$). See Section 3.2.3 below, where we used the MBB approach to implement blocking resampling method.[4] In practice (as in our empirical application), to compute the estimated residual from the full model $\hat{e}_{t+h}$, one may consider the model within which all of the approximations models are nested. If such a model does not exist (i.e., is not one of the approximation models $i = 1, \ldots, N$), a comprehensive model can be easily created by including all available regressors in it.

3.2.2 Dependent wild-based residual resampling

We now describe the second general bootstrap algorithm that can also be used to obtain an estimator of the variance of a weighted average of a parameter estimates across different models and/or an estimator of the variance of a combined forecast. The generic algorithm for the dependent wild-based residual resampling in model averaging reads as follows.

Algorithm 2. The general dependent wild-based residual resampling in model averaging.

1. Identical to Algorithm 1.

2. For $i = 1, \ldots, N$ construct bootstrap residual samples $e^{*(i)}_{t+h}$, $t = 1, \ldots, T-h$ by multiplying each *recentered* regression residuals (which is equal to the regression residuals from the full model) by

[4] As pointed out by one referee, an alternative to the blocking approach would be to "whiten" the serially correlated residuals before applying the bootstrap, taking advantage of the fact that residuals have an MA$(h-1)$ structure when forecasting at the h-step ahead horizon. We leave a rigorous proof for future research.

a possibly dependent variable of external draws, and *use the same variable for all* $i = 1, \ldots, N$. More specifically, construct bootstrap residual samples as follows

$$e_{t+h}^{*(i)} = \left(\underbrace{\hat{e}_{t+h}^{(i)} - \hat{b}_{t+h}^{(i)}}_{=\hat{e}_{t+h}} \right) \cdot \eta_{t+h}^{*(i)} + \hat{b}_{t+h}^{(i)}, \qquad t = 1, \ldots, T-h, \tag{48}$$

where $\eta_{t+h}^{*(i)} = \eta_{t+h}^{*}$ (for all $i = 1, \ldots, N$) is a typical element of a vector $\eta^* = \left(\eta_{1+h}^*, \ldots, \eta_T^*\right)'$ of random draws (possibly dependent across t) with mean $\mathbf{0}_{(T-h)\times 1}$. Note that in (48), η^* is the same across $i = 1, \ldots, N$.

The rest of the steps 3-6 of Algorithm 2 is same as those in Algorithm 1.

Note that imposing 0 mean to η_{t+h}^* ensures that in the full model $E^*\left(e_{t+h}^{*(i)}\right) = \hat{b}_{t+h}^{(i)}$ (as in Algorithm 1). By using the same value of random draws for all $i = 1, \ldots, N$, we preserve the dependence across models.

Remark 2. *In Algorithms 1 and 2, the bootstrap residuals* $\mathbf{e}^{*(i)}$ *(in model* i*) and* $\mathbf{e}^*$ *(in the full model) satisfy the following equation*

$$\mathbf{e}^{*(i)} = \hat{\mathbf{b}}^{(i)} + \mathbf{e}^*, \tag{49}$$

such that $E^*\left(\mathbf{e}^*\right) = 0$ *and* $E^*\left(\mathbf{e}^{*(i)}\right) = \hat{\mathbf{b}}^{(i)}$ *with* $\hat{\mathbf{b}}^{(i)} = \left(\hat{b}_{1+h}^{(i)}, \ldots, \hat{b}_T^{(i)}\right)'$. *Thus, our bootstrap approaches mimic the non-zero mean property of the error* $\mathbf{e}^{(i)}$ *in model* i *(see equation (6)). Both schemes (Algorithms 1 and 2) also have the advantage that they retain the cross-sectional dependence of* $e_{t+h}^{(i)}$ *over different models* $i = 1, \ldots, N$ *and at the same time preserves the time series dependence over* $t = 1, \ldots, T-h$ *nonparametrically in the error term.*

Remark 3. *Given (38), note that the value of* $\mathbf{V}_{ij,T}^*$ *remains unchanged if instead of using (44) in step 2 of Algorithm 1, we construct bootstrap residual samples as follows*

$$\hat{e}_{(j-1)\ell+s+h}^{*(i)} = \hat{e}_{\tau_{(j-1)\ell+s}+h}, \tag{50}$$

for $j = 1, \ldots, k, s = 1, \ldots, \ell,$ $i = 1, \ldots, N$. *Similarly,* $\mathbf{V}_{ij,T}^*$ *remains unchanged if instead of using (48) in step 2 of Algorithm 2, we simply generate bootstrap residual samples as follows*

$$e_{t+h}^{*(i)} = \hat{e}_{t+h}\eta_{t+h}^*, \tag{51}$$

for $t = 1, \ldots, T-h$, $i = 1, \ldots, N$. *Consequently, our bootstrap procedures yield exactly the same bootstrap variance estimator of averaging estimators as the simple bootstrap scheme, which consist of resampling the errors from the full model, i.e., obtained bootstrap residual sample* $\left\{e_{t+h}^{*(i)}, \quad t = 1, \ldots, T-h\right\}$ *from* $\{\hat{e}_{t+h}, \quad t = 1, \ldots, T-h\}$ *and, using these, building* y_{t+h}^* *keeping the regressors* fixed. *However, this equivalence is not necessarily true for other moments.*

Given (31) and (49), if follows that

$$\sqrt{T}\left(\hat{\theta}^*(\omega)-\hat{\theta}\right)=\underbrace{\sum_{i=1}^{N}\omega_i\hat{\mathbf{A}}_{i,T}}_{=\hat{\mathbf{A}}_T(\mathbf{w})}+\sum_{i=1}^{N}\omega_i\mathbf{P}_{i,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^*\right), \tag{52}$$

where we let $\hat{\mathbf{A}}_{i,T}=\hat{\mathbf{a}}_{i,T}+\mathbf{P}_{i,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\hat{\mathbf{b}}^{(i)}\right)$. Therefore, $\mathbf{V}^*_{ij,T}$ (as defined in (32)) can be written as $\mathbf{V}^*_{ij,T}=\mathbf{P}_{i,T}\mathbf{\Omega}^*_T\mathbf{P}'_{j,T}$, where $\mathbf{\Omega}^*_T\equiv Var^*\left(T^{-1/2}\mathbf{H}'\mathbf{e}^*\right)=Var^*\left(T^{-1/2}\sum_{t=1}^{T-h}\mathbf{h}_te^*_{t+h}\right)$. Thus, conditional on the observed data, the dependence structure of the scaled average of the bootstrap regression scores$\{\mathbf{h}_te^*_{t+h}\}$ dictates the consistency of the bootstrap variance $\mathbf{V}^*_T(\mathbf{w})$ toward the asymptotic variance $\mathbf{V}(\mathbf{w})$.

Next, we provide a set of high level conditions on $\{\mathbf{h}_te^*_{t+h}\}$ that will allow us to characterize the bootstrap distribution of $\hat{\theta}^*(\omega)$.

Condition (A*): $\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^*\xrightarrow{\mathbf{d}^*}\mathbf{N}(\mathbf{0}_{(p+q)\times 1},\mathbf{\Omega}^*)$, in probability, such that $\mathbf{\Omega}^*>0$ with $\mathbf{\Omega}^*\equiv p\lim_{T\to\infty}\mathbf{\Omega}^*_T=\mathbf{\Omega}$.

Condition (B*): For $i,j=1,\ldots,N$, $\mathbf{V}^*_{ij}\equiv p\lim_{T\to\infty}\mathbf{V}^*_{ij,T}=\mathbf{V}_{ij}$.

Condition A* requires the bootstrap regression scores to obey a central limit theorem in the bootstrap world. This condition is rather standard in bootstrapping model selection context. More specifically, when we apply weight 1 to the full model and weight 0 to all other models, we have $\hat{\theta}(\omega)=\hat{\theta}$, $\mathbf{A}_T(\mathbf{w})=\hat{\mathbf{A}}_T(\mathbf{w})=0$, and therefore in such a context, under Assumption 2, Condition A* is sufficient to show the first-order asymptotic validity of the bootstrap, i.e.,

$$\sup_{x\in\mathbb{R}^{p+q}}\left|P^*\left(\sqrt{T}\left(\hat{\theta}^*(\omega)-\hat{\theta}\right)\leq x\right)-P\left(\left(\sqrt{T}\left(\hat{\theta}(\omega)-\theta\right)\right)\leq x\right)\right|\to^P 0, \tag{53}$$

as $T\to\infty$.

Condition B* mimics the cross-sectional dependence of $e^{(i)}_{t+h}$ over models $i=1,\ldots,N$. It is useful to note that once Condition A* is satisfied (in particular, when $p\lim_{T\to\infty}\mathbf{\Omega}^*_T=\mathbf{\Omega}$), we only need to show that $\mathbf{P}_{i,T}\to^P\mathbf{P}_i$ to conclude that Condition B* holds.

Theorem 3.2. *Let Assumptions 1 and 2 hold. Assume (24) where $e^{*(i)}_{t+h}$ is obtained either by (44) or by (48) for which Conditions A* and B* are satisfied, then as $T\to\infty$*

$$\sqrt{T}\left(\hat{\theta}^*(\omega)-\hat{\theta}\right)-\hat{\mathbf{A}}_T(\mathbf{w})\to^{d^*}\mathbf{N}(\mathbf{0}_{(p+q)\times 1},\mathbf{V}(\mathbf{w})), \tag{54}$$

in prob-P.

Theorem 3.2 implies that

$$\sup_{x\in\mathbb{R}^{p+q}}\left|P^*\left(\sqrt{T}\left(\hat{\theta}^*(\omega)-\hat{\theta}\right)-\hat{\mathbf{A}}_T(\mathbf{w})\leq x\right)-P\left(\left(\sqrt{T}\left(\hat{\theta}(\omega)-\theta\right)-\mathbf{A}_T(\mathbf{w})\right)\leq x\right)\right|\to^P 0, \tag{55}$$

as $T\to\infty$, thus justifying the use of the bootstrap distribution of $\sqrt{T}\left(\hat{\theta}^*(\omega)-\hat{\theta}\right)-\hat{\mathbf{A}}_T(\mathbf{w})$ as a consistent estimator of the distribution of $\sqrt{T}\left(\hat{\theta}(\omega)-\theta\right)-\mathbf{A}_T(\mathbf{w})=\sqrt{T}\left(\hat{\theta}(\omega)-\left(\theta+\frac{\mathbf{A}_T(\mathbf{w})}{\sqrt{T}}\right)\right)$.

In particular, the bootstrap can be used to construct percentile-type intervals for $\theta + \frac{\mathbf{A}_T(\mathbf{w})}{\sqrt{T}}$. A $100\,(1-\alpha)\,\%$ nominal level symmetric bootstrap percentile confidence interval for $\theta + \frac{\mathbf{A}_T(\mathbf{w})}{\sqrt{T}}$ is given by

$$\hat{\theta}\,(\omega) \pm T^{-1/2} c^*_{1-\alpha}, \tag{56}$$

where $c^*_{1-\alpha}$ is such that $P^*\left(\left|\sqrt{T}\left(\hat{\theta}^*\,(\omega) - \hat{\theta}\right) - \hat{\mathbf{A}}_T\,(\mathbf{w})\right| \leq c^*_{1-\alpha}\right) = 1-\alpha$. Unfortunately, our parameter of interest is θ and not $\theta + \frac{\mathbf{A}_T(\mathbf{w})}{\sqrt{T}}$. One may construct an *unbiased* bootstrap percentile confidence interval for θ by using

$$\hat{\theta}\,(\omega) \pm T^{-1/2} c^*_{1-\alpha} - \frac{\tilde{\mathbf{A}}_T\,(\mathbf{w})}{\sqrt{T}}, \tag{57}$$

where $\tilde{\mathbf{A}}_T\,(\mathbf{w}) \equiv \sum_{i=1}^{N} \omega_i \tilde{\mathbf{A}}_{i,T}$, with $\tilde{\mathbf{A}}_{i,T} \equiv \left[\mathbf{P}_{i,T}\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right) - \mathbf{I}_{p+q}\right] \mathbf{S}_0 \hat{\delta}$, and where $\hat{\delta} = \sqrt{T}\hat{\gamma}$, such that $\hat{\gamma}$ is given in (7) and is the estimate from the full model.

Note that $\tilde{\mathbf{A}}_T\,(\mathbf{w})$ is not a consistent estimator of $\mathbf{A}_T\,(\mathbf{w})$ but is asymptotically unbiased. This is the reason why we call (57) "*unbiased* bootstrap percentile confidence interval for θ" and not simply as usual in the bootstrap literature "bootstrap percentile confidence interval for θ". Result in Theorem 3.2 does not imply that

$$\sqrt{T}\left(\hat{\theta}^*\,(\omega) - \hat{\theta}\right) \to^{d^*} \mathbf{N}(\mathbf{A}_T\,(\mathbf{w})\,, \mathbf{V}\,(\mathbf{w})).$$

Remark 4. *Theorem 3.2 does not imply that the distributions of* $\sqrt{T}\left(\hat{\theta}^*\,(\omega) - \hat{\theta}\right)$ *and* $\sqrt{T}\left(\hat{\theta}\,(\omega) - \theta\right)$ *are close, in the sense that (53) holds. This negative result does not contradict Hjort and Claeskens (2003) (cf. Section 10.6) regarding the invalidity of bootstrapping method on the weighted average of* $\hat{\gamma}_i$ *in the framework of local alternative.*

As discussed by Shao and Tu (1995) (pp 79), Gonçalves and White (2004) and lucidly pointed out by Gonçalves et al. (2019), convergence in distribution of a random sequence does not imply convergence of moments. Therefore, Theorem 3.2 does not by itself justify using the covariance matrix of the bootstrap distribution of $\sqrt{T}\left(\hat{\theta}^*\,(\omega) - \hat{\theta}\right) - \hat{\mathbf{A}}_T\,(\mathbf{w})$, given by

$$\lim_{B\to\infty} (1/B) \sum_{b=1}^{B} T\left(\hat{\theta}^{*(b)}\,(\omega) - \overline{\hat{\theta}^*\,(\omega)}\right)\left(\hat{\theta}^{*(b)}\,(\omega) - \overline{\hat{\theta}^*\,(\omega)}\right)', \tag{58}$$

where $\overline{\hat{\theta}^*\,(\omega)} = (1/B)\sum_{b=1}^{B} \hat{\theta}^{*(b)}\,(\omega)$ with B the number of bootstrap replications, to consistently estimate the asymptotic covariance matrix of $\hat{\theta}\,(\omega)$. Nevertheless, given that

$$Var^*\left[\sqrt{T}\left(\hat{\theta}^*\,(\omega) - \hat{\theta}\right) - \hat{\mathbf{A}}_T\,(\mathbf{w})\right] = Var^*\left[\sqrt{T}\left(\hat{\theta}^*\,(\omega)\right)\right],$$

and given Theorem 3.2, a sufficient condition for the consistency of the bootstrap covariance estimator in (58) is that $\left\{\left[\sqrt{T}\left(\hat{\theta}^*\,(\omega) - \hat{\theta}\right) - \hat{\mathbf{A}}_T\,(\mathbf{w})\right]\left[\sqrt{T}\left(\hat{\theta}^*\,(\omega) - \hat{\theta}\right) - \hat{\mathbf{A}}_T\,(\mathbf{w})\right]'\right\}$ is uniformly integrable, which is implied by the condition that

$$E^*\left|\left[\sqrt{T}\left(\hat{\theta}^*\,(\omega) - \hat{\theta}\right) - \hat{\mathbf{A}}_T\,(\mathbf{w})\right]\right|^{2+\delta'} = O_P\,(1) \tag{59}$$

for some small $\delta' > 0$.

We now consider the special case of the MBB and the NBB schemes to generate $e^{*(i)}_{t+h}$ in step 2 of Algorithm 1. Thereafter we will consider the special case of the DWB and the blocking external bootstrap (BEB) method schemes to generate $e^{*(i)}_{t+h}$ in step 2 of Algorithm 2.

3.2.3 Special case for Algorithm 1

The first scheme we consider is the MBB. In step 2 of Algorithm 1, with the MBB method the set of indices are formally given by (42). Therefore, the bootstrap residuals are given by

$$\hat{e}^{*(i)}_{(j-1)\ell+s+h} = \hat{e}^{(i)}_{I_j+s+h} - \hat{b}^{(i)}_{I_j+s+h} - \frac{1}{T-h-\ell+1}\sum_{j=1}^{T-h-\ell+1}\left(\underbrace{\hat{e}^{(i)}_{(j-1)+s+h} - \hat{b}^{(i)}_{(j-1)+s+h}}_{=\hat{\epsilon}_{(j-1)+s+h}}\right) + \hat{b}^{(i)}_{(j-1)\ell+s+h}, \tag{60}$$

for $j = 1, \ldots, k, s = 1, \ldots, \ell$, $i = 1, \ldots, N$, where I_j are i.i.d random variables distributed uniformly on $\{0, \ldots, T-h-\ell\}$.[5]

Theorem 3.3. *Suppose that a blocking-based residual resampling is used to generate bootstrap residual samples $\left\{e^{*(i)}_{t+h}\right\}$, such that in step 2 of Algorithm 1 $e^{*(i)}_{t+h}$ is given by (60). Let Assumptions 1 and 2 be true, and $\mathbf{\Sigma}_T^{-1} = O(1)$, where $\mathbf{\Sigma}_T = \sum_{t=1}^{T-h}\sum_{s=1}^{T-h} Cov(\mathbf{h}_t e_{t+h}, \mathbf{h}_s e_{s+h})$. If $\ell_T \to \infty$ such that $\ell_T = o\left(T^{1/2}\right)$, as $T \to \infty$, then the conclusions of Theorem 3.2 follow. If in addition, for some $\delta' > 0$, $\lambda_{\max}^{2+\delta'}(\mathbf{P}_{i,T}) = O_p(1)$, where $\lambda_{\max}(\mathbf{P}_{i,T})$ denotes the largest eigenvalue of $\mathbf{P}_{i,T}$, then (59) holds.*

3.2.4 Special case for Algorithm 2

The dependent wild bootstrap (DWB) was proposed by Shao (2010) for smooth function of the sample mean with time series observations.[6] The DWB differs from the BEB by smoothing the external draw across blocks. When specialized in our context, in step 2 of Algorithm 2, we construct bootstrap

[5] Similarly for the NBB method, in step 2 of Algorithm 1, the set of indices are formally given by (43). Thus the NBB analog of (60) is given as follows

$$\hat{e}^{*(i)}_{(j-1)\ell+s+h} = \hat{e}^{(i)}_{J_j+s+h} - \hat{b}^{(i)}_{J_j+s+h} - \frac{1}{k}\sum_{j=1}^{k}\left(\hat{e}^{(i)}_{(j-1)\ell+s+h} - \hat{b}^{(i)}_{(j-1)\ell+s+h}\right) + \hat{b}^{(i)}_{(j-1)\ell+s+h}, \tag{61}$$

for $j = 1, \ldots, k, s = 1, \ldots, \ell$, $i = 1, \ldots, N$, where J_j are i.i.d random variables distributed uniformly on $\{0, \ldots, k-1\}$.

[6] The BEB method, which was first proposed by Yeh (1998) for a linear regression with fixed scalar regressor and strong mixing errors has been analyzed in other contexts by Shao (2011), Smeekes and Urbain (2013) and Djogbenou et al. (2015). The related wild block bootstrap method of Hounyo (2017) and Hounyo et al. (2017) can also be used as well in step 2 of Algorithm 2. For instance, specializing to BEB in our context, in step 2 of Algorithm 2, first we form non-overlapping blocks of size ℓ of consecutive residuals, then construct bootstrap residual samples as follows

$$e^{*(i)}_{(j-1)\ell+s+h} = \left(\hat{e}^{(i)}_{(j-1)\ell+s+h} - \hat{b}^{(i)}_{(j-1)\ell+s+h}\right) \cdot \eta^*_{(j-1)\ell+s+h} + \hat{b}^{(i)}_{(j-1)\ell+s+h}, \tag{62}$$

with $\eta^*_{(j-1)\ell+s+h} = v^*_j$, $j = 1, \ldots, k, s = 1, \ldots, \ell$, where v^*_j is an external random variable such that $v^*_j \sim$i.i.d.$(0,1)$ across $j = 1, \ldots, k$. Then, the bootstrap residuals are obtained by multiplying each residual by an external random variable that is the same for all observations within a block j. More importantly, in our context, we impose the vector $\eta^* = (\eta^*_{1+h}, \ldots, \eta^*_T)$ to be the same for all $i = 1, \ldots, N$.

residual samples as follows

$$e_{t+h}^{*(i)} = \underbrace{\left(\hat{e}_{t+h}^{(i)} - \hat{b}_{t+h}^{(i)}\right)}_{=\hat{e}_{t+h}} \cdot \eta_{t+h}^{*} + \hat{b}_{t+h}^{(i)}, \qquad t = 1, \ldots, T-h, \tag{63}$$

where $\eta^* = \left(\eta_{1+h}^*, \ldots, \eta_T^*\right)'$ is a random vector with mean $\mathbf{0}_{(T-h)\times 1}$ and covariance matrix K, with typical element $K_{st} = E^*\left(\eta_s^* \cdot \eta_t^*\right) = k_{DWB}\left(\frac{t-s}{l_T}\right)$, where $k_{DWB}(\cdot)$ is a kernel function and l_T a bandwidth parameter. Following Shao (2010), in this paper we assume that η^* is ℓ_T-dependent. In Section 5 we set $\eta^* = K\eta$, with $\eta \sim N\left(0, I_{T-h}\right)$. Then η_{t+h}^* is a local weighted average of external draws, thereby making the neighbouring observations time dependent. In addition, since the vector η^* is set to be the same for all $i = 1, \ldots, N$, this preserves the cross-sectional dependence of $\hat{e}_{t+h}^{(i)}$ over different models $i = 1, \ldots, N$ as well.

In order to state our result for the DWB, we follow Djogbenou et al (2015) and require a slightly stronger dependence and moment conditions than Assumption 2. Specifically, we impose:

Assumption 2':

(a) For some $r > 2$, $\{(\mathbf{h}_t', e_{t+h})\}$ is a fourth-order stationary strong mixing sequence of size $-\frac{3r}{r-2}$ and $E\left(e_{t+h}|\mathcal{F}_t\right) = 0$, where $\mathcal{F}_t = \sigma\left(\mathbf{h}_t, \mathbf{h}_{t-1}, \ldots; e_t, e_{t-1}, \ldots\right); E\left\|\mathbf{h}_t\right\|^{4r} < C$ and $E\left\|e_{t+h}\right\|^{4r} < C$.

The other parts of this assumption remain as before. Assumption 2' is analogous to the assumptions made in Andrews (1991, Lemma 1) to prove consistency of the HAC estimator.

We also follow Shao (2010) and make the following restriction on the class of kernels.

Assumption 3. *$k_{DWB} : \mathbb{R} \to [0,1]$ is symmetric with compact support on $[-1,1]$, $k_{DWB}(0) = 1$, $\lim_{x\to 0}\left(1 - k_{DWB}(x)\right)/|x|^q \neq 0$ for some $q \in (0,1]$ such that $\psi(\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} k_{DWB}(x)\exp(i\xi x)\,dx \geq 0$, for all $\xi \in \mathbb{R}$.*

The condition $\psi(\xi) \geq 0$ ensures that the matrix K is positive definite. These assumptions are satisfied by the Bartlett and Parzen kernels but not by the truncated, quadratic spectral and the Tukey-Hanning kernels (see e.g., Andrews (1991) pp. 822-823). By imposing Assumptions 2' and 3, we are able to build on results in Andrews (1991) and Shao (2010) when proving our result.

Theorem 3.4. *Suppose that a dependent wild-based residual resampling is used to generate bootstrap residual samples $\left\{e_{t+h}^{*(i)}\right\}$, such that in step 2 of Algorithm 2 $e_{t+h}^{*(i)}$ is given by (63) with $E^*\left|\eta_{t+h}^*\right|^{2r} \leq \Delta < \infty$ for some $r > 2$. Under Assumptions 1, 2', and 3, if $l_T \to \infty$ such that $T_T^{-1}\ell^{2(r+1)/r} \to 0$, as $T \to \infty$, then the conclusions of Theorem 3.2 follow. If in addition, for some $\delta' > 0$, $\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right) = O_p(1)$, where $\lambda_{\max}\left(\mathbf{P}_{i,T}\right)$ denotes the largest eigenvalue of $\mathbf{P}_{i,T}$, then (59) holds.*

This result is the DWB analog of Theorem 3.3.[7] Both theorems allow us to use the two methods to estimate the asymptotic variance of a combined estimator as stated in part 6 of Algorithms 1 and 2.

[7] For the NBB and BEB methods, similar results as in Theorems 3.3 and 3.4 hold, respectively.

4 Monte Carlo simulations

In this section we assess the finite sample properties of the bootstrap methods discussed in Section 3. The data-generating process is similar to the one used by Liu and Kuo (2016). Specifically, we consider the linear regression model:

$$y_{t+h} = \sum_{j=1}^{k} \beta_j x_{jt} + e_{t+h}, \tag{64}$$

$$x_{jt} = \rho_x x_{jt-1} + u_{jt} \text{ for } j \geq 2, \tag{65}$$

where x_{jt} are AR(1) processes with $\rho_x = 0.5$ and 0.9 and we set $x_{1t} = 1$ to be the intercept. We draw $(u_{2t}, \ldots, u_{kt})'$ from a joint normal distribution $N(\mathbf{0}, \mathbf{Q}_u)$, where the diagonal elements of $\mathbf{Q}_u$ are 1 and the off-diagonal elements are ρ_u, such that $\rho_u \in \{0.25, 0.50, 0.75, 0.9\}$. To obtain the error term e_t, we first generate an AR(1) process $\epsilon_t = 0.5\epsilon_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim N(0, 0.75)$. Then the error term is constructed by $e_t = 3^{-1/2}\left(1 - \rho_x^2\right) x_{kt}^2 \epsilon_t$. We determined the regression coefficients and the local parameters as follows:

$$\beta = \frac{c}{\sqrt{T}} \left(1, \frac{k-1}{k}, \ldots, \frac{1}{k}\right)',$$

and

$$\delta_j = \sqrt{T}\beta_j = \frac{c\,(k-j+1)}{k},$$

for $j \geq 2$. The parameter c is selected to vary the population $R^2 = \widetilde{\beta}'\mathbf{Q}_x\widetilde{\beta} / \left(1 + \widetilde{\beta}'\mathbf{Q}_x\widetilde{\beta}\right)$, where $\widetilde{\beta} = (\beta_2, \ldots, \beta_k)'$ and $\mathbf{Q}_x = \left(1 - \rho_x^2\right)\mathbf{Q}_u$. The population R^2 is set to vary on a grid between 0.1 and 0.9. We set $k = 5$ and the sample size $T = 200$. We consider all possible models, and hence the number of models is $N = 32$. We consider two forecasting horizons, $h = 1$ and $h = 4$. We use the equal-weighted ($\omega_i = 1/N$, $i = 1, \ldots, N$,) forecast combinations. In the simulations, we consider the following four approaches to compute the variance of the combined forecast:

(i) naïve bootstrap approach, that resample the entire $(N \times 1)$-vector of regression residuals over time, (labelled naïve); [8]

(ii) our proposed blocking-based residual method (see Algorithm 1), using the MBB to resample residuals (labelled MBB);

(iii) our proposed general dependent wild-based residual resampling (see Algorithm 2), using the DWB to resample residuals (labelled DWB);

(iv) a plug-in approach, based on a direct estimator of $\mathbf{\Sigma}_{y_{T+h|T}}$, defined below and given by (66) (labelled Plug-in);

For the plug-in approach, we compute $\hat{\mathbf{\Sigma}}_{y_{T+h|T}}$, a (consistent) plug-in estimator of the asymptotic

[8] For the naïve approach, we use the DWB method to obtain bootstrap residuals.

variance $\boldsymbol{\Sigma}_{y_{T+h|T}}$, as follows

$$\hat{\boldsymbol{\Sigma}}_{y_{T+h|T}} = \mathbf{h}_T' \hat{\mathbf{V}}_T(\mathbf{w})\, \mathbf{h}_T = \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_i \omega_j \hat{\mathbf{V}}_{ij,T}, \tag{66}$$

where

$$\hat{\mathbf{V}}_T(\mathbf{w}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_i \omega_j \hat{\mathbf{V}}_{ij,T}, \text{ with } \hat{\mathbf{V}}_{ij,T} = \mathbf{P}_{i,T}\hat{\boldsymbol{\Omega}}_T \mathbf{P}_{j,T}',$$

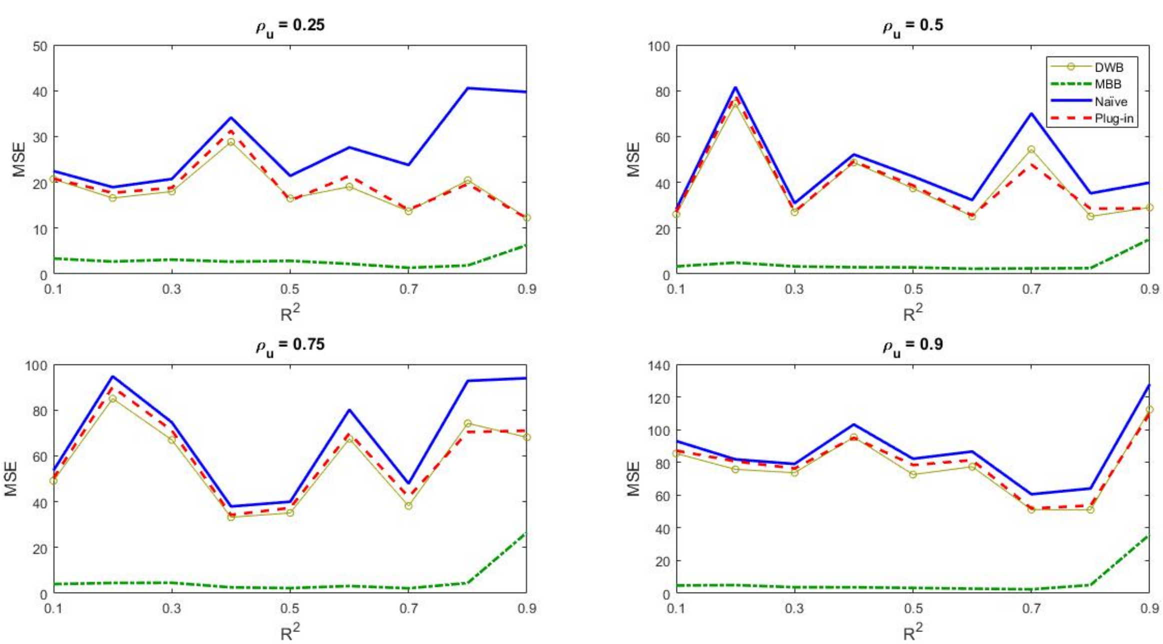
such that

$$\hat{\boldsymbol{\Omega}}_T = T^{-1}\sum_{t=1}^{T} \hat{s}_t \hat{s}_t' + T^{-1}\sum_{h=1}^{\ell_T}\left(1 - \frac{h}{\ell_T + 1}\right)\sum_{t=h+1}^{T}\left(\hat{s}_t \hat{s}_{t-h}' + \hat{s}_{t-h}\hat{s}_t'\right), \tag{67}$$

where $\hat{s}_{t+h} = \mathbf{h}_t \hat{e}_{t+h}$. More specifically, in our simulations to compute $\hat{\boldsymbol{\Omega}}_T$, we use a HAC estimator of $\boldsymbol{\Omega}$ using a Bartlett kernel with bandwidth ℓ_T selected by the data-based rule from Andrews (1991). For the DWB, we use the same bandwidth ℓ_T selected to compute $\hat{\boldsymbol{\Omega}}_T$. Similarly, to select the block size, for the MBB, we rely on the asymptotic equivalence between the MBB and the Bartlett kernel variance estimators, and then choose the block size equal to the bandwidth ℓ_T chosen by Andrews's automatic procedure for the Bartlett kernel.

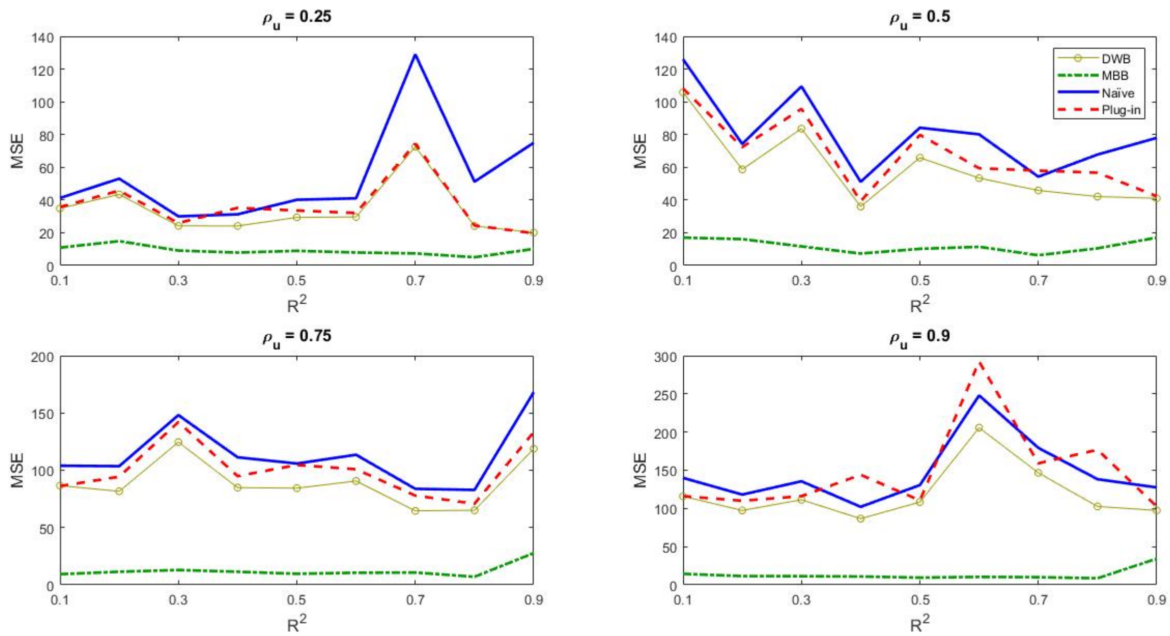
We compare the (four) estimators of the asymptotic variance $\boldsymbol{\Sigma}_{y_{T+h|T}}$ by looking at their MSE over 1000 replications. We use 499 bootstrap replications.

Figure 1: MSE for heteroscedastic linear regression models ($\rho_x = 0.5$, $h = 4$).



We first compare the MSE when the AR(1) coefficient of the predictor equal 0.5. The results are presented in Figure 1, for $h = 4$. The results for $h = 1$ (not reported) are qualitatively similar to those

Figure 2: MSE for heteroscedastic linear regression models ($\rho_x = 0.9$, $h = 4$).



reported for $h = 4$. The naïve bootstrap-based estimator has much larger MSE than other estimators. In particular, our proposed procedures (MBB and DWB) outperform the naïve bootstrap approach. Although all three methods MBB, DWB and Plug-in are asymptotically equivalent, the estimator based on the MBB is quite robust to different values of R^2 and has much lower MSE than those based on the DWB and the Plug-in approaches. In most cases, the Plug-in and the DWB estimators have quite similar performance.

Figure 2 displays the corresponding results of Figure 1, but now with $\rho_x = 0.9$. Overall, results presented in Figure 2, suggest that the ranking of estimators when $\rho_x = 0.9$ is qualitatively quite similar to that for $\rho_x = 0.5$. However, for $\rho_u = 0.50$, 0.75, and 0.9, the Plug-in and the DWB estimators do no longer have similar performance. The gains associated with the DWB method over the Plug-in approach are now more distinguishable and can be quite substantial.

5 Empirical illustration

In this section we illustrate the desirability of using our bootstrapping approaches to compute the variance of combined estimators. In particular, we follow Granger and Jeon (2004) and revisit the empirical findings of Kozicki (1999) who investigated the usefulness of the Taylor rule recommendations to policymakers based on combined estimates. Specifically, Kozicki (1999) estimated Taylor-types rules for 24 combinations from reasonable variations in the alternative definitions of inflation and output gap with monthly data from 1983-1997. In their pioneering approach Granger and Jeon (2004) reported estimates of the variance of the combined parameters using bootstrap technique. They reported that the variance from their bootstrap based-approach were considerably smaller than that from simple average over individual models.

We follow Kozicki (1999) and Granger and Jeon (2004) (cf. Section 8) and consider four inflation measures and six different measures of the output gap (which amounts to 24 different models). As the inflation measure, we use CPI inflation, core CPI inflation, GDP price inflation, and expected inflation collected from the Survey of Professional Forcasters. For the output gap variable, we consider output gap measures from the Congressional Budget Office (CBO), the International Monetary Fund (IMF), the Organization for Economic Cooperation and Development (OECD), Standard and Poor (DRI), an approximation of the definition of the output gap used by Taylor (Taylor), and a recursive version of the Taylor method (Recursive). As emphasized by Kozicki (1999), these six alternative measures are reasonable approximations to the definition of true output gap for use in a Taylor Rule equation.[9] The precise definition of the variables and data sources can be found in Kozicki (1999). We estimate the following equation for all 24 combinations from different measure of inflation and output gap:

$$r_t = c + (1 + \alpha)\,\pi_{t-1} + \beta y^g_{t-1}, \tag{68}$$

where r_t is the federal funds rate at time t, c is a constant, π_{t-1} and y^g_{t-1} are the inflation and output gap at $t-1$, respectively, cf. Granger and Jeon (2004).

We consider different methods to compute $\sqrt{var\left(\sum_{i=1}^{N}\omega_i\hat{\alpha}_i\right)}$ and $\sqrt{var\left(\sum_{i=1}^{N}\omega_i\hat{\beta}_i\right)}$, with $\omega_i = \frac{1}{N}$, $i = 1, \ldots, N = 24$. First, we implement the naïve residual-based bootstrap approach, which consists of stacking all residuals at time t into a vector, and then resample these cross-sectional vectors of residuals over time. Thus, it is not valid in our present empirical context (of combined estimators) as shown in Section 3.1. Second, we consider our proposed new procedure.

Figure 3 reveals that the residuals from our full model have significant autocorrelation. Hence, a simple i.i.d bootstrap or the WB may not be appropriate to capture the observed serial dependence in the residuals. For our proposed resampling method, we consider MBB. We use $B = 9999$ bootstrap replications. The choice of the block size for the MBB is important. As in the simulation study, we consider the full model where we included all available regressors (a constant term, the above six measures of inflation and four measures of output gaps). Then, we use Andrews's (1991) automatic procedure to compute a data-driven block size ℓ^* to implement our proposed procedure. Table 1 reports our results.

In the first two columns of Table 1, we report our replication of Kozicki's (1999) 24 individual Taylor rule equations. These estimates are seen to be very similar to those reported in Granger and Jeon (2004). We calculated the average values of the inflation and output gap coefficients to be 0.637 and 0.131, respectively. Granger and Jeon (2004) estimated these coefficients to be 0.539 and 0.191, respectively, and are close, given that we regenerated the original sample. Other estimates in

[9]We thank Sharon Kozicki for guiding us in reconstructing her data set. Since DRI has merged with IHS (now called IHS-Markit), we use the IHS-Markit output gap instead of the original DRI measure.

Table 1 (see the two right hand side columns) are obtained using the MBB method as explained in Section 3.2.3 (see also Algorithm 1).[10] On average the selected data-driven block size for the MBB was $\ell^* = 5$. Based on our proposed resampling approach (using MBB), the estimated standard error of the combined coefficients estimated for inflation and output are 0.090 and 0.042, respectively (see last row of Table 1). Granger and Jeon (2004) obtained these values to be 0.045 and 0.021, respectively. Our estimates are significantly (two times) larger than those reported by Granger and Jeon (2004). The simple averages of the standard errors of the two parameters over the 24 individual models are very close to those obtained using our new proposed bootstrapping approach.

We also computed the standard errors of the combined coefficients estimated (i.e., $\sum_{i=1}^{N} \omega_i \hat{\alpha}_i$ (for inflation) and $\sum_{i=1}^{N} \omega_i \hat{\beta}_i$ (for output gap)) using an i.i.d. bootstrap procedure, resampling regression residuals independently across models $i = 1, \ldots, N = 24$ (not reported in Table 1). We found that they are very close to those obtained by Granger and Jeon (2004), and were 0.045 and 0.020, respectively. Hence, our replication results suggest that in Granger and Jeon (2004) the bootstrap procedure did not take into account the dependence across models.

We also report the standard errors of the two coefficients using a non-robust cross-sectional resampling approach, which accommodates the serial correlation in the errors (by using the MBB) but not the cross-sectional dependence between models (by resampling regression residuals independently across models). Those estimates were found to be 0.021 and 0.012, respectively, and are significatively less than those using our resampling approach. Thus, the primarily source of underestimation of the standard errors is not due to the lack of adjustment for serial correlation but due to the failure of the bootstrap procedure in Granger and Jeon (2004) to capture the cross-sectional dependence.

As expected, the naïve bootstrap procedure which uses the MBB to resample the whole vector of regression residuals over time, overestimates the standard error of the combined coefficients quite significantly by inducing an additional term in the bootstrap variance of averaging estimators (as explained in Section 3.1). These estimates were found to be 0.376 and 0.169, respectively.

In summary, these results suggest that a resampling approach which imposes independence across models underestimates the standard error of the combined coefficients quite significantly by failing to take into account the correlation between models. In contrast, a naïve residual-based bootstrap approach which resamples the entire vector of regression residuals over time t, overestimates the standard error of the combined coefficients. Our replication results also suggest that in Granger and Jeon (2004) the bootstrap procedure did not take into account the dependence across models.

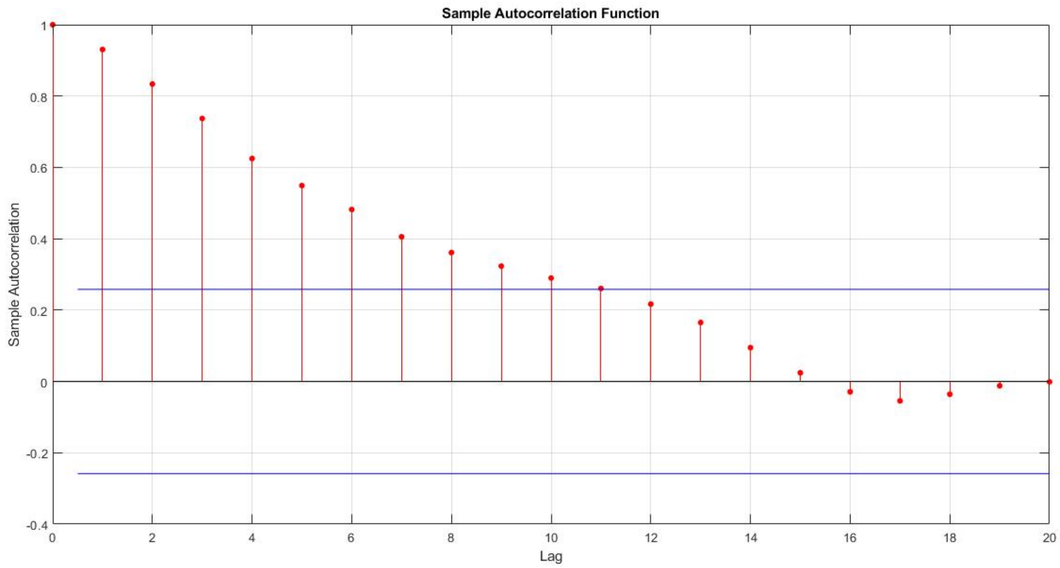
[10]For sake of brevity, we only report in Table 1 results based on MBB using Algorithm 1. Alternatively, we might have used Algorithm 2, but note that in implementing step 2 of either of the algorithms, a number of resampling methods are available (e.g., NBB of Carlstein (1986), SB of Politis and Romano (1994) or BEB method of Yeh (1998), among others).

Table 1: Taylor rule: combined estimators from different models, resampling based on the MBB.

Output gap measure	Inflation measure	Kozicki (1999)		Moving Blocks Bootstrap S.D.	
		Inflation	Output	Inflation	Output
CBO	CPI inflation	-0.004	0.023	0.081	0.046
CBO	Core CPI inflation	0.420	0.125	0.099	0.046
CBO	GDP price inflation	0.949	0.326	0.103	0.048
CBO	Expected inflation	1.197	0.328	0.098	0.047
OECD	CPI inflation	0.016	-0.066	0.081	0.044
OECD	Core CPI inflation	0.438	0.088	0.100	0.044
OECD	GDP price inflation	0.951	0.258	0.105	0.046
OECD	Expected inflation	1.278	0.326	0.101	0.046
IMF	CPI inflation	-0.022	0.099	0.082	0.047
IMF	Core CPI inflation	0.419	0.193	0.099	0.047
IMF	GDP price inflation	0.942	0.377	0.102	0.049
IMF	Expected inflation	1.164	0.353	0.097	0.048
HIS	CPI inflation	0.050	-0.236	0.081	0.041
HIS	Core CPI inflation	0.375	-0.059	0.102	0.042
HIS	GDP price inflation	0.855	0.098	0.108	0.044
HIS	Expected inflation	1.307	0.246	0.106	0.045
Taylor	CPI inflation	-0.005	0.016	0.082	0.042
Taylor	Core CPI inflation	0.396	0.083	0.099	0.042
Taylor	GDP price inflation	0.865	0.248	0.101	0.042
Taylor	Expected inflation	1.136	0.262	0.097	0.042
Recursive	CPI inflation	0.045	-0.166	0.081	0.033
Recursive	Core CPI inflation	0.382	-0.061	0.100	0.033
Recursive	GDP price inflation	0.863	0.091	0.107	0.035
Recursive	Expected inflation	1.263	0.184	0.104	0.036
Combined estimators					
Simple average over 24 models		0.637	0.131	0.096	0.043
Not allowing cross-sectional dependence				0.021	0.012
Resampling vector of residuals as whole over time				0.376	0.169
Our proposed resampling approach				0.090	0.042

Notes: This table provides the estimated coefficients and standard errors from the estimation of the Taylor rule (see equation (68)) for all 24 combinations from different measure of inflation and output gap. 'Not allowing cross-sectional dependence' means resampling independently across models, but allowing for serial correlation (by using the MBB to obtain bootstrap errors). 'Resampling vector of residuals as whole over time' is the naïve residual-based bootstrap approach, which consists to stack all residuals at time t into a vector, then resample these cross-sectional vectors of residuals over time as discussed in Section 3.1. We use 9999 bootstrap replications.

Figure 3: Sample autocorrelation function of estimated residual.



6 Conclusion

The aim of this paper has been to provide conditions under which a residual-based bootstrap method can provide a consistent estimator of the asymptotic variance of a combined forecast and/or the asymptotic covariance matrix of a weighted average of a parameter estimates across different models with fixed weights. Our results show that a naïve residual-based bootstrap approach, which consists of stacking all residuals at time t into a vector, and then resampling these cross-sectional vectors of residuals over time is invalid in the context of model averaging. We propose and theoretically justify two general residual-based bootstrap resampling approaches for model averaging in predictive regressions to estimate the variance of a combined estimator. We discuss the application of the two general methods when regression residuals are resampled by either MBB of Künsch (1989) and Liu and Singh (1992), NBB of Carlstein (1986), DWB of Shao (2010) and BEB method of Yeh (1998) and Shao (2011).

We illustrate our methods using the bootstrap estimates of the Taylor rule parameters reported by Granger and Jeon (2004), and show that underestimation of the sampling variability of the combined estimator can be substantial if the cross-sectional dependence between the models is not properly accounted for while resampling. We also show that overestimation of the sampling variability of the combined estimator can be substantial if one relies on a common approach of resampling cross-sectional vectors over time (in order to preserve the cross-sectional dependence between models).

A Appendix: Proofs

Proof of Theorem 3.1 part (a). Given (25) and (26), note that we can decompose the bootstrap OLS estimator for the ith submodel as

$$\hat{\theta}_i^* - \hat{\theta}_i = \left(\mathbf{H}_i'\mathbf{H}_i\right)^{-1}\mathbf{H}_i'\mathbf{e}^{*(i)}.$$

Given that $\mathbf{H}_i = \mathbf{H}\mathbf{S}_i$, it follows that

$$\sqrt{T}\mathbf{S}_i\left(\hat{\theta}_i^* - \hat{\theta}_i\right) = \mathbf{P}_{i,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^{*(i)}\right). \tag{A.1}$$

$$\sqrt{T}\left(\mathbf{S}_i\hat{\theta}_i^* - \hat{\theta}\right) = \underbrace{\sqrt{T}\left[\frac{1}{T}\mathbf{P}_{i,T}\mathbf{H}' - \left(\mathbf{H}'\mathbf{H}\right)^{-1}\mathbf{H}'\right]\mathbf{y}}_{=\hat{\mathbf{a}}_{i,T}} + \mathbf{P}_{i,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^{*(i)}\right). \tag{A.2}$$

and

$$\begin{aligned}
\sqrt{T}\left(\hat{\theta}^*(\omega) - \hat{\theta}\right) &= \sum_{i=1}^{N}\omega_i\sqrt{T}\left(\mathbf{S}_i\hat{\theta}_i^* - \hat{\theta}\right) \\
&= \underbrace{\sum_{i=1}^{N}\omega_i\hat{\mathbf{a}}_{i,T}}_{=\hat{\mathbf{a}}_T(\mathbf{w})} + \sum_{i=1}^{N}\omega_i\mathbf{P}_{i,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^{*(i)}\right). \qquad \text{(A.3)}
\end{aligned}$$

Given (A.2), we can write

$$\begin{aligned}
Cov^*\left[\sqrt{T}\left(\mathbf{S}_i\hat{\theta}_i^* - \hat{\theta}\right), \sqrt{T}\left(\mathbf{S}_j\hat{\theta}_j^* - \hat{\theta}\right)\right] &= Cov^*\left[\hat{\mathbf{a}}_{i,T} + \mathbf{P}_{i,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^{*(i)}\right), \hat{\mathbf{a}}_{j,T} + \mathbf{P}_{j,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^{*(j)}\right)\right] \\
&= Cov^*\left[\mathbf{P}_{i,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^{*(i)}\right), \mathbf{P}_{j,T}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^{*(j)}\right)\right] = \mathbf{V}_{ij,T}^*. \qquad \text{(A.4)}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathbf{V}_{ij,T}^* &= \mathbf{P}_{i,T}Cov^*\left[\frac{1}{\sqrt{T}}\sum_{s=1}^{T-h}\mathbf{h}_s e_{s+h}^{*(i)}, \frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mathbf{h}_t e_{t+h}^{*(j)}\right]\mathbf{P}_{j,T}' \\
&= \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\mathbf{h}_s\mathbf{h}_t' Cov^*\left(e_{s+h}^{*(i)}, e_{t+h}^{*(j)}\right)\right]\mathbf{P}_{j,T}'. \qquad \text{(A.5)}
\end{aligned}$$

Next, remark that by definition

$$Cov^*\left(e_{s+h}^{*(i)}, e_{t+h}^{*(j)}\right) = E^*\left(e_{s+h}^{*(i)}e_{t+h}^{*(j)}\right) - E^*\left(e_{s+h}^{*(i)}\right)E^*\left(e_{t+h}^{*(j)}\right).$$

Given (34), if $s \neq t$, we have $Cov^*\left(e_{s+h}^{*(i)}, e_{t+h}^{*(j)}\right) = 0$ (since $e_{t+h}^{*(i)}$ is independent across t conditionally on the observed time series), whereas if $s=t$, we have

$$\begin{aligned}
& Cov^*\left(e^{*(i)}_{s+h}, e^{*(j)}_{t+h}\right) \\
= \; & E^*\left(\left(\hat{e}^{(i)}_{t+h}\upsilon^*_{t+h}\right)\left(\hat{e}^{(j)}_{t+h}\upsilon^*_{t+h}\right)\right) - E^*\left(\hat{e}^{(i)}_{t+h}\upsilon^*_{t+h}\right)E^*\left(\hat{e}^{(j)}_{t+h}\upsilon^*_{t+h}\right) \\
= \; & \hat{e}^{(i)}_{t+h}\hat{e}^{(j)}_{t+h}\underbrace{\left[E^*\left(\upsilon^{*2}_{t+h}\right) - E^*\left(\upsilon^*_{t+h}\right)^2\right]}_{=1} \\
= \; & \left(\hat{e}_{t+h} + \hat{b}^{(i)}_{t+h}\right)\left(\hat{e}_{t+h} + \hat{b}^{(j)}_{t+h}\right) \\
= \; & \left(\hat{e}_{t+h} + \hat{b}^{(i)}_{t+h,1} + \hat{b}^{(i)}_{t+h,2}\right)\left(\hat{e}_{t+h} + \hat{b}^{(j)}_{t+h,1} + \hat{b}^{(j)}_{t+h,2}\right) \\
= \; & \hat{e}^2_{t+h} + \hat{b}^{(i)}_{t+h,1}\hat{b}^{(j)}_{t+h,1} \\
& + \hat{e}_{t+h}\hat{b}^{(j)}_{t+h,1} + \hat{e}_{t+h}\hat{b}^{(j)}_{t+h,2} + \hat{b}^{(i)}_{t+h,1}\hat{e}_{t+h} + \hat{b}^{(i)}_{t+h,1}\hat{b}^{(j)}_{t+h,2} + \hat{b}^{(i)}_{t+h,2}\hat{e}_{t+h} + \hat{b}^{(i)}_{t+h,2}\hat{b}^{(j)}_{t+h,1} + \hat{b}^{(i)}_{t+h,2}\hat{b}^{(j)}_{t+h,2},
\end{aligned}$$

where the second equality uses the fact that $Var^*\left(\upsilon^*_{t+h}\right) = 1$, and the third and fourth equalities follows given (36) and (37), respectively. Thus, $\mathbf{V}^*_{ij,T}$ can be written as follows

$$\mathbf{V}^*_{ij,T} = \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}'_t\hat{e}^2_{t+h}\right]\mathbf{P}'_{j,T} + \widetilde{\mathbf{V}}^{\mathrm{WB}}_{ij,T} + \check{\mathbf{V}}^{\mathrm{WB}}_{ij,T}, \tag{A.6}$$

such that

$$\widetilde{\mathbf{V}}^{\mathrm{WB}}_{ij,T} = \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}'_t\hat{b}^{(i)}_{t+h,1}\hat{b}^{(j)}_{t+h,1}\right]\mathbf{P}'_{j,T},$$

and

$$\begin{aligned}
\check{\mathbf{V}}^{\mathrm{WB}}_{ij,T} \; = \; & \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}'_t\hat{e}_{t+h}\hat{b}^{(j)}_{t+h,1}\right]\mathbf{P}'_{j,T} + \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}'_t\hat{e}_{t+h}\hat{b}^{(j)}_{t+h,2}\right]\mathbf{P}'_{j,T} \\
& + \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}'_t\hat{b}^{(i)}_{t+h,1}\hat{e}_{t+h}\right]\mathbf{P}'_{j,T} + \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}'_t\hat{b}^{(i)}_{t+h,1}\hat{b}^{(j)}_{t+h,2}\right]\mathbf{P}'_{j,T} \\
& + \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}'_t\hat{b}^{(i)}_{t+h,2}\hat{e}_{t+h}\right]\mathbf{P}'_{j,T} + \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}'_t\hat{b}^{(i)}_{t+h,2}\hat{b}^{(j)}_{t+h,1}\right]\mathbf{P}'_{j,T} \\
& + \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}'_t\hat{b}^{(i)}_{t+h,2}\hat{b}^{(j)}_{t+h,2}\right]\mathbf{P}'_{j,T}.
\end{aligned}$$

The desired result follows given the definitions of $\hat{b}^{(i)}_{t+h,1}$, $\hat{b}^{(i)}_{t+h,2}$ (see (37)) and Assumptions 1 and 2.

More specifically, we can write

$$\begin{aligned}
\check{\mathbf{V}}_{ij,T}^{\mathrm{WB}} &= \mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\hat{e}_{t+h}\mathbf{h}_t'\left(\mathbf{I}_{p+q}-\mathbf{P}_{j,T}\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)\right)\theta\right]\mathbf{P}_{j,T}' \\
&+\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\hat{e}_{t+h}\mathbf{h}_t'\left(\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)^{-1}-\mathbf{P}_{j,T}\right)\left(\frac{1}{T}\mathbf{H}'\mathbf{e}\right)\right]\mathbf{P}_{j,T}' \\
&+\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\theta'\left(\mathbf{I}_{p+q}-\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)\mathbf{P}_{i,T}'\right)\mathbf{h}_t\hat{e}_{t+h}\right]\mathbf{P}_{j,T}' \\
&+\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\theta'\left(\mathbf{I}_{p+q}-\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)\mathbf{P}_{i,T}'\right)\mathbf{h}_t\mathbf{h}_t'\left(\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)^{-1}-\mathbf{P}_{j,T}\right)\left(\frac{1}{T}\mathbf{H}'\mathbf{e}\right)\right]\mathbf{P}_{j,T}' \\
&+\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\left(\left(\frac{1}{T}\mathbf{H}'\mathbf{e}\right)'\left(\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)^{-1}-\mathbf{P}_{i,T}'\right)\right)\mathbf{h}_t\hat{e}_{t+h}\right]\mathbf{P}_{j,T}' \\
&+\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\left(\left(\frac{1}{T}\mathbf{H}'\mathbf{e}\right)'\left(\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)^{-1}-\mathbf{P}_{i,T}'\right)\right)\mathbf{h}_t\mathbf{h}_t'\left(\mathbf{I}_{p+q}-\mathbf{P}_{j,T}\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)\right)\theta\right]\mathbf{P}_{j,T}' \\
&+\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\left(\left(\frac{1}{T}\mathbf{H}'\mathbf{e}\right)'\left(\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)^{-1}-\mathbf{P}_{i,T}'\right)\right)\mathbf{h}_t\mathbf{h}_t'\left(\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)^{-1}-\mathbf{P}_{j,T}\right)\left(\frac{1}{T}\mathbf{H}'\mathbf{e}\right)\right]\mathbf{P}_{j,T}' \\
&= o_p(1).
\end{aligned}$$

Similarly, $\widetilde{\mathbf{V}}_{ij,T}^{\mathrm{WB}}$ can be written as

$$\widetilde{\mathbf{V}}_{ij,T}^{\mathrm{WB}}=\mathbf{P}_{i,T}\left[T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\left[\theta'\left(\mathbf{I}_{p+q}-\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)\mathbf{P}_{i,T}'\right)\mathbf{h}_t\mathbf{h}_t'\left(\mathbf{I}_{p+q}-\mathbf{P}_{j,T}\left(\frac{1}{T}\mathbf{H}'\mathbf{H}\right)\right)\theta\right]\right]\mathbf{P}_{j,T}'.$$

Finally, result follows immediately by using (33), (A.6), and recalling that under our assumptions $\underset{T\to\infty}{p\lim}\left[\mathbf{P}_{i,T}\left(T^{-1}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\hat{e}_t^2\right)\mathbf{P}_{j,T}'\right]=\mathbf{V}_{ij}$, and $\underset{T\to\infty}{p\lim}\widetilde{\mathbf{V}}_{ij,T}^{\mathrm{WB}}=\widetilde{\mathbf{V}}_{ij}^{\mathrm{WB}}$.

Proof of Theorem 3.1 part (b). The proof for the i.i.d. bootstrap follows similarly the same arguments provided in the proof of part (a) of Theorem 3.1.

Proof of Theorem 3.2. Given (31) and (49), to obtain the desired result, we need to show that (a) $\mathbf{P}_{i,T}\xrightarrow{p}\mathbf{P}_i$, (b) $\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{e}^*\xrightarrow{d^*}\mathbf{N}(\mathbf{0}_{(p+q)\times 1},\mathbf{\Omega})$ in probability, and (c) $\mathbf{V}_{ij,T}^*\xrightarrow{p}\mathbf{V}_{ij}$. Part (a) holds directly under Assumption 2, because the selection matrix $\mathbf{S}_i$ is not random with element either 0 or 1. Part (b) follows under Condition A*, whereas part (c) holds under Condition B*.

Proof of Theorem 3.3. We proceed as follows: We first show the first part of Theorem 3.3, next we verify condition (59). To show the first part of Theorem 3.3, we need to verify Conditions A* and B*.

Starting with Condition A*, we use Theorem 3.1 of Fitzenberger (1998) by verifying his assumptions. Given Assumption 2 and the additional condition in the statement of Theorem 3.3 i.e., $\mathbf{\Sigma}_T^{-1}=O(1)$, where $\mathbf{\Sigma}_T=\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}Cov(\mathbf{h}_te_{t+h},\mathbf{h}_se_{s+h})$, Fitzenberger's (1998) Assumptions (A1), (A2), (A3), (A4) and (A5) hold directly.

Condition B* follows by noting that by Condition A*, $\mathbf{\Omega}_T^*=Var^*\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mathbf{h}_te_{t+h}^*\right]\xrightarrow{p}\mathbf{\Omega}$, and under Assumption 2 $\mathbf{P}_{i,T} \xrightarrow{p} \mathbf{P}_i$.

Finally, we verify condition (59). For this purpose, we need to introduce some additional notations. In the following, for any matrix $\mathbf{A}$, $\|\mathbf{A}\|_1$ denotes the matrix norm defined by $\|\mathbf{A}\|_1^2 = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}}$. Notice that for $\mathbf{A}$ symmetric, $\|\mathbf{A}\|_1$ is equal to the largest eigenvalue of $\mathbf{A}$, i.e., $\|\mathbf{A}\|_1 = \lambda_{\max}(\mathbf{A})$.

For some small $\delta' > 0$, we can write

$$\begin{aligned}
E^*\left|\left[\sqrt{T}\left(\hat{\theta}^*(\omega)-\hat{\theta}\right)-\hat{\mathbf{A}}_T(\mathbf{w})\right]\right|^{2+\delta'} &= E^*\left|\sum_{i=1}^{N}\omega_i\left[\sqrt{T}\left(\mathbf{S}_i\hat{\theta}_i^*-\hat{\theta}\right)-\hat{\mathbf{A}}_{i,T}\right]\right|^{2+\delta'} \\
&\leq N^{1+\delta'}\sum_{i=1}^{N}\omega_i^{2+\delta'}E^*\left|\left[\sqrt{T}\left(\mathbf{S}_i\hat{\theta}_i^*-\hat{\theta}\right)-\hat{\mathbf{A}}_{i,T}\right]\right|^{2+\delta'} \\
&= N^{1+\delta'}\sum_{i=1}^{N}\omega_i^{2+\delta'}E^*\left|\mathbf{P}_{i,T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mathbf{h}_t e_{t+h}^*\right)\right|^{2+\delta'} \\
&\leq N^{1+\delta'}\sum_{i=1}^{N}\omega_i\underbrace{\|\mathbf{P}_{i,T}\|_1^{2+\delta'}}_{=\lambda_{\max}^{2+\delta'}}T^{-(2+\delta')/2}E^*\left|\sum_{t=1}^{T-h}\left(\mathbf{h}_t e_{t+h}^*-\underbrace{E^*\left(\mathbf{h}_t e_{t+h}^*\right)}_{=0}\right)\right|^{2+\delta'} \\
&\equiv C\sum_{i=1}^{N}\mathbf{B}_i,
\end{aligned}$$

where the first inequality uses the c_r-inequality. The last inequality uses the fact that for any $\delta' > 0$, and $0 \leq \omega_i \leq 1$, we have $0 \leq \omega_i^{2+\delta'} \leq \omega_i \leq 1$.

Because N is finite, it follows that to prove condition (59), it suffices to show that $\mathbf{B}_i = O_p(1)$. Thus, we have

$$\begin{aligned}
\mathbf{B}_i &\leq \omega_i\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right)T^{-(2+\delta')/2}E^*\left|\sum_{t=1}^{T-h}\left|\mathbf{h}_t e_{t+h}^*-E^*\left(\mathbf{h}_t e_{t+h}^*\right)\right|^2\right|^{(2+\delta')/2} \\
&\leq \omega_i\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right)T^{-(2+\delta')/2}E^*\left|\left(\sum_{t=1}^{T-h}\left|\mathbf{h}_t e_{t+h}^*-E^*\left(\mathbf{h}_t e_{t+h}^*\right)\right|^{2+\delta'}\right)^{2/(2+\delta')}(T-h)^{1-2/(2+\delta')}\right|^{(2+\delta')/2} \\
&= \omega_i\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right)T^{-(2+\delta')/2}(T-h)^{(2+\delta')/2-1}\sum_{t=1}^{T-h}E^*\left|\mathbf{h}_t e_{t+h}^*-E^*\left(\mathbf{h}_t e_{t+h}^*\right)\right|^{2+\delta'} \\
&\leq 2^{2+\delta'}\omega_i\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right)\left(\frac{T-h}{T}\right)^{(2+\delta')/2}(T-h)^{-1}\sum_{t=1}^{T-h}E^*\left|\mathbf{h}_t e_{t+h}^*\right|^{2+\delta'} \\
&= 2^{2+\delta'}\omega_i\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right)\left(\frac{T-h}{T}\right)^{(2+\delta')/2}(T-h)^{-1}\sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'}E^*\left|e_{(j-1)\ell+s+h}^*\right|^{2+\delta'}, \quad \text{(A.7)}
\end{aligned}$$

where the first inequality uses the Burkholder's inequality, the second inequality follows by the Holder's inequality, whereas the last inequality uses the c_r-inequality.

Next, using the definitions of e_{t+h} and $\hat{e}_{t+h}$ yields $\hat{e}_{t+h} = e_{t+h} - \mathbf{h}'_{t+h}\left(\hat{\theta}-\theta\right)$. Note that we have

$$\begin{aligned}
& \sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'} E^{*}\left|e^{*}_{(j-1)\ell+s+h}\right|^{2+\delta'} \\
= \quad & \sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'} E^{*}\left|\hat{e}_{I_j+s+h} - E^{*}\left(\hat{e}_{I_j+s+h}\right)\right|^{2+\delta'} \\
\leq \quad & C\sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'} E^{*}\left|\hat{e}_{I_j+s+h}\right|^{2+\delta'} \\
\leq \quad & C\sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'} E^{*}\left|e_{I_j+s+h} - \mathbf{h}'_{I_j+s+h}\left(\hat{\theta}-\theta\right)\right|^{2+\delta'} \\
\leq \quad & C\left[\begin{array}{c} \sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'} E^{*}\left|e_{I_j+s+h}\right|^{2+\delta'} \\ +\sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'} E^{*}\left|\mathbf{h}'_{I_j+s+h}\left(\hat{\theta}-\theta\right)\right|^{2+\delta'} \end{array}\right] \\
\leq \quad & C\left[\begin{array}{c} \sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'} \frac{1}{T-h-\ell+1}\sum_{g=1}^{T-h-\ell+1}\left|e_{g-1+s+h}\right|^{2+\delta'} \\ +\left|\sqrt{T}\left(\hat{\theta}-\theta\right)\right|^{2+\delta'}\sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'} \frac{1}{T-h-\ell+1}\sum_{g=1}^{T-h-\ell+1}\left|\mathbf{h}_{g-1+s+h}\right|^{2+\delta'} \end{array}\right] \quad \text{(A.8)}
\end{aligned}$$

Given (A.7) and (A.8) and the fact that under our assumptions $\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right) = O_P(1)$, to prove that $\mathbf{B}_i = O_p(1)$, it suffices that $\mathbf{B}_{i,1} = O_p(1)$ and $\mathbf{B}_{i,2} = O_p(1)$ such that

$$\mathbf{B}_{i,1} \equiv \left|\sqrt{T}\left(\hat{\theta}-\theta\right)\right|^{2+\delta'}(T-h)^{-1}\sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'}\frac{1}{T-h-\ell+1}\sum_{g=1}^{T-h-\ell+1}\left|\mathbf{h}_{g-1+s+h}\right|^{2+\delta'},$$

and

$$\mathbf{B}_{i,2} \equiv (T-h)^{-1}\sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'}\frac{1}{T-h-\ell+1}\sum_{g=1}^{T-h-\ell+1}\left|e_{g-1+s+h}\right|^{2+\delta'}.$$

For $\mathbf{B}_{i,1}$, note that because $\sqrt{T}\left(\hat{\theta}-\theta\right)$ converges in distribution, it follows that $\left|\sqrt{T}\left(\hat{\theta}-\theta\right)\right|^{2+\delta'} = O_P(1)$. Thus, to prove that $\mathbf{B}_{i,1} = O_p(1)$, it suffices to show that

$$\mathbf{B}_{i,1,1} \equiv (T-h)^{-1}\sum_{j=1}^{k}\sum_{s=1}^{\ell}\left|\mathbf{h}_{(j-1)\ell+s}\right|^{2+\delta'}\frac{1}{T-h-\ell+1}\sum_{g=1}^{T-h-\ell+1}\left|\mathbf{h}_{g-1+s+h}\right|^{2+\delta'} = O_p(1).$$

We have

$$
\begin{aligned}
& E\left|\mathbf{B}_{i,1,1}\right| \\
= \; & \frac{(T-h)^{-1}}{T-h-\ell+1} E\left|\sum_{s=1}^{\ell}\sum_{j=1}^{k}\left|\mathbf{h}_{g-1+s+h}\right|^{2+\delta'}\sum_{g=1}^{T-h-\ell+1}\left|\mathbf{h}_{g-1+s+h}\right|^{2+\delta'}\right| \\
\leq \; & \frac{(T-h)^{-1}}{T-h-\ell+1}\sum_{s=1}^{\ell} E\left|\sum_{j=1}^{k}\left|\mathbf{h}_{g-1+s+h}\right|^{2+\delta'}\sum_{g=1}^{T-h-\ell+1}\left|\mathbf{h}_{g-1+s+h}\right|^{2+\delta'}\right| \\
\leq \; & \frac{(T-h)^{-1}}{T-h-\ell+1}\sum_{s=1}^{\ell}\left(E\left|\sum_{j=1}^{k}\left|\mathbf{h}_{g-1+s+h}\right|^{2+\delta'}\right|^{2}\right)^{1/2}\left(E\left|\sum_{g=1}^{T-h-\ell+1}\left|\mathbf{h}_{g-1+s+h}\right|^{2+\delta'}\right|^{2}\right)^{1/2} \\
\leq \; & \frac{(T-h)^{-1}}{T-h-\ell+1}\sum_{s=1}^{\ell}\left(\frac{T-h}{\ell}\sum_{j=1}^{k} E\left|\mathbf{h}_{g-1+s+h}\right|^{2(2+\delta')}\right)^{1/2}\left((T-h-\ell+1)\sum_{g=1}^{T-h-\ell+1} E\left|\mathbf{h}_{g-1+s+h}\right|^{2(2+\delta')}\right)^{1/2} \\
= \; & \frac{(T-h)^{-1}}{(T-h-\ell+1)^{1/2}}\left(\frac{T-h}{\ell}\right)^{1/2}\sum_{s=1}^{\ell}\left(\sum_{j=1}^{k} E\left|\mathbf{h}_{g-1+s+h}\right|^{2(2+\delta')}\right)^{1/2}\left(\sum_{g=1}^{T-h-\ell+1} E\left|\mathbf{h}_{g-1+s+h}\right|^{2(2+\delta')}\right)^{1/2} \\
\leq \; & \frac{(T-h)^{-1}}{(T-h-\ell+1)^{1/2}}\left(\frac{T-h}{\ell}\right)^{1/2}\left[\sum_{s=1}^{\ell}\left(\sum_{j=1}^{k} E\left|\mathbf{h}_{g-1+s+h}\right|^{2(2+\delta')}\right)\right]^{1/2}\left[\sum_{s=1}^{\ell}\left(\sum_{g=1}^{T-h-\ell+1} E\left|\mathbf{h}_{g-1+s+h}\right|^{2(2+\delta')}\right)\right]^{1/2} \\
= \; & \left[\frac{1}{T-h}\sum_{t=1}^{T-h} E\left|\mathbf{h}_t\right|^{2(2+\delta')}\right]^{1/2}\left[\frac{1}{T-h-\ell+1}\sum_{g=1}^{T-h-\ell+1}\frac{1}{\ell}\sum_{s=1}^{\ell} E\left|\mathbf{h}_{g-1+s+h}\right|^{2(2+\delta')}\right]^{1/2} \\
= \; & O(1).
\end{aligned}
$$

Thus, by Markov's inequality, we have $\mathbf{B}_{i,1,1} = O_p(1)$. For $\mathbf{B}_{i,2}$, using the same arguments as for $\mathbf{B}_{i,1,1}$, we have

$$
\begin{aligned}
E\left|\mathbf{B}_{i,2}\right| \leq \; & \left[\frac{1}{T-h}\sum_{t=1}^{T-h} E\left|\mathbf{h}_t\right|^{2(2+\delta')}\right]^{1/2}\left[\frac{1}{T-h-\ell+1}\sum_{g=1}^{T-h-\ell+1}\frac{1}{\ell}\sum_{s=1}^{\ell} E\left|e_{g-1+s+h}\right|^{2(2+\delta')}\right]^{1/2} \\
= \; & O(1),
\end{aligned}
$$

This completes the proof.

Proof of Theorem 3.4. The strategy of the proof follows closely that of Theorem 3.3. However, we highlight the main differences. As in that proof, we first show the first part of Theorem 3.4, next we verify condition (59). To show the first part of Theorem 3.4, we need to verify Conditions A* and B*.

Starting with Condition A*, as in the proof of Theorem 3 of Djogbenou et al. (2015), we use Theorem 3.1 of Shao (2010) by verifying his assumptions. In particular, under Assumption 2, $\{\mathbf{h}_t e_{t+h}\}$ are strong mixing of size $-\frac{3r}{r-2}$ for some $r > 2$ with $E\left\|\mathbf{h}_t e_{t+h}\right\|^{2r} < C$, implying that $\sum_{j=1}^{\infty}\alpha(j)^{\frac{r}{r+2}} < \infty$ verifying his Assumption 3.1. Next, by using Lemma 1 of Andrews (1991), we also have that

$\sum_{j=1}^{\infty} j^2 \alpha(j)^{\frac{r-2}{r}} < \infty$ and $E\left\|\mathbf{h}_t e_{t+h}\right\|^{2r} < C < \infty$, thus verifying his Assumption 3.2.

Condition B* follows by noting that by Condition A*, $\mathbf{\Omega}_T^* = Var^*\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mathbf{h}_t\hat{e}_{t+h}\eta_{t+h}^*\right] \xrightarrow{p} \mathbf{\Omega}$, and under Assumption 2 $\mathbf{P}_{i,T} \xrightarrow{\mathbf{p}} \mathbf{P}_i$. Specifically, we have

$$\mathbf{V}_{ij,T}^* = \mathbf{P}_{i,T}\underbrace{\left[T^{-1}\sum_{t=1}^{T-h}\sum_{s=1}^{T-h}\mathbf{h}_t\mathbf{h}_s'\hat{e}_{t+h}\hat{e}_{s+h}k_{DWB}\left(\frac{t-s}{l_T}\right)\right]}_{=\mathbf{\Omega}_T^* = Var^*\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mathbf{h}_t\hat{e}_{t+h}\eta_{t+h}^*\right]}\mathbf{P}_{j,T}' \xrightarrow{p} \mathbf{V}_{ij}.$$

Finally we verify condition (59). Given (A.3), for some small $\delta' > 0$, we can write

$$\begin{aligned}
& E^*\left|\left[\sqrt{T}\left(\hat{\theta}^*(\omega) - \hat{\theta}\right) - \hat{\mathbf{A}}_T(\mathbf{w})\right]\right|^{2+\delta'} \\
= \; & E^*\left|\sum_{i=1}^{N}\omega_i\left[\sqrt{T}\left(\mathbf{S}_i\hat{\theta}_i^* - \hat{\theta}\right) - \hat{\mathbf{A}}_{i,T}\right]\right|^{2+\delta'} \\
\leq \; & N^{1+\delta'}\sum_{i=1}^{N}\omega_i^{2+\delta'}E^*\left|\left[\sqrt{T}\left(\mathbf{S}_i\hat{\theta}_i^* - \hat{\theta}\right) - \hat{\mathbf{A}}_{i,T}\right]\right|^{2+\delta'} \\
= \; & N^{1+\delta'}\sum_{i=1}^{N}\omega_i^{2+\delta'}E^*\left|\mathbf{P}_{i,T}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mathbf{h}_t e_{t+h}\eta_{t+h}^* - \left(\frac{1}{T}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\eta_{t+h}\right)\sqrt{T}\left(\hat{\theta} - \theta\right)\right]\right|^{2+\delta'} \\
\leq \; & N^{1+\delta'}2^{1+\delta'}\sum_{i=1}^{N}\omega_i\underbrace{\left\|\mathbf{P}_{i,T}\right\|_1^{2+\delta'}}_{=\lambda_{\max}^{2+\delta'}}E^*\left|\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mathbf{h}_t e_{t+h}\eta_{t+h}\right)\right|^{2+\delta'} \\
& + N^{1+\delta'}\sum_{i=1}^{N}\omega_i^{2+\delta'}\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right)\left|\sqrt{T}\left(\hat{\theta} - \theta\right)\right|^{2+\delta'}E^*\left|\left(\frac{1}{T}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\eta_{t+h}^*\right)\right|^{2+\delta'} \\
\equiv \; & C\sum_{i=1}^{N}\left(\mathbf{D}_{i,1} + \mathbf{D}_{i,2}\right).
\end{aligned}$$

The last inequality uses the c_r-inequality and the fact that for any $\delta' > 0$, and $0 \leq \omega_i \leq 1$, we have $0 \leq \omega_i^{2+\delta'} \leq \omega_i \leq 1$.

Thus, it suffices to show that $\mathbf{D}_{i,1} + \mathbf{D}_{i,2} = O_p(1)$, since N is finite. Note that

$$\begin{aligned}
\mathbf{D}_{i,1} &= \omega_i\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right)E^*\left|\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-h}\mathbf{h}_t e_{t+h}\eta_{t+h}^*\right)\right|^{2+\delta'} \\
&= \omega_i\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right)T^{-(2+\delta')/2}E^*\left|\left(\sum_{t=1}^{T-h}\left(\mathbf{h}_t e_{t+h}\eta_{t+h} - \underbrace{E^*\left(\mathbf{h}_t e_{t+h}\eta_{t+h}^*\right)}_{=0}\right)\right)\right|^{2+\delta'}.
\end{aligned}$$

Then we have

$$
\begin{aligned}
\mathbf{D}_{i,1} &\leq \omega_i \lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right) T^{-(2+\delta')/2} E^* \left| \sum_{t=1}^{T-h} \left| \mathbf{h}_t e_{t+h} \eta_{t+h}^* - E^*\left(\mathbf{h}_t e_{t+h} \eta_{t+h}^*\right) \right|^2 \right|^{(2+\delta')/2} \\
&\leq \omega_i \lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right) T^{-(2+\delta')/2} E^* \left| \left(\sum_{t=1}^{T-h} \left| \mathbf{h}_t e_{t+h} \eta_{t+h}^* - E^*\left(\mathbf{h}_t e_{t+h} \eta_{t+h}^*\right) \right|^{2+\delta'} \right)^{2/(2+\delta')} (T-h)^{1-2/(2+\delta')} \right|^{(2+\delta')/2} \\
&= \omega_i \lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right) T^{-(2+\delta')/2} (T-h)^{(2+\delta')/2-1} \sum_{t=1}^{T-h} E^* \left| \mathbf{h}_t e_{t+h} \eta_{t+h}^* - E^*\left(\mathbf{h}_t e_{t+h} \eta_{t+h}^*\right) \right|^{2+\delta'} \\
&\leq 2^{2+\delta'} \omega_i \lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right) T^{-(2+\delta')/2} (T-h)^{(2+\delta')/2} (T-h)^{-1} \sum_{t=1}^{T-h} \left| \mathbf{h}_t e_{t+h} \right|^{2+\delta'} E^* \left| \eta_{t+h}^* \right|^{2+\delta'},
\end{aligned}
$$

where the first inequality uses the Burholder's inequality, the second inequality follows by the Holder's inequality, whereas the last inequality uses the c_r-inequality. Because $\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right) = O_P(1)$, and given that under our assumptions we have $E^* \left|\eta_{t+h}\right|^{2+\delta'} \leq \Delta < \infty$ for some $\delta' > 0$, to prove that $\mathbf{D}_{i,1} = O_p(1)$, it suffices that $E\left|\mathbf{D}_{i,1,1}\right| = O(1)$ where $\mathbf{D}_{i,1,1} = (T-h)^{-1} \sum_{t=1}^{T-h} \left|\mathbf{h}_t e_{t+h}\right|^{2+\delta'}$. Thus, by using the Cauchy-schartz inequality, we have

$$
\begin{aligned}
E\left|\mathbf{D}_{i,1,1}\right| &\leq \left((T-h)^{-1} \sum_{t=1}^{T-h} E\left|\mathbf{h}_t\right|^{2(2+\delta')} \right)^{1/2} \left((T-h)^{-1} \sum_{t=1}^{T-h} E\left|e_{t+h}\right|^{2(2+\delta')} \right)^{1/2} \\
&= O(1).
\end{aligned}
$$

For $D_{i,2}$, note that

$$
\mathbf{D}_{i,2} = \omega_i \lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right) \left| \sqrt{T}\left(\hat{\theta} - \theta\right) \right|^{2+\delta'} E^* \left[\lambda_{\max}^{2+\delta'} \left(\frac{1}{T} \sum_{t=1}^{T-h} \mathbf{h}_t \mathbf{h}_t' \eta_{t+h}^* \right) \right].
$$

Because $\sqrt{T}\left(\hat{\theta} - \theta\right)$ converges in distribution, it follows that $\left| \sqrt{T}\left(\hat{\theta} - \theta\right) \right|^{2+\delta'} = O_P(1)$. In addition, $\lambda_{\max}^{2+\delta'}\left(\mathbf{P}_{i,T}\right) = O_P(1)$ under our assume conditions. Thus, to prove that $\mathbf{D}_{i,2} = O_P(1)$, it suffices that $E^* \left[\lambda_{\max}^{2+\delta'} \left(\frac{1}{T} \sum_{t=1}^{T-h} \mathbf{h}_t \mathbf{h}_t' \eta_{t+h}^* \right) \right] = O_P(1)$. To show this, observe that we can write $\mathbf{h}_t = \left(h_{1t}, h_{2t}, \ldots, h_{(p+q)t}\right)'$. Then, we have

$$
\begin{aligned}
E^* \left[\lambda_{\max}^{2+\delta'} \left(\frac{1}{T} \sum_{t=1}^{T-h} \mathbf{h}_t \mathbf{h}_t' \eta_{t+h}^* \right) \right] &\leq E^* \left[\left| tr \left(\frac{1}{T} \sum_{t=1}^{T-h} \mathbf{h}_t \mathbf{h}_t' \eta_{t+h}^* \right) \right|^{2+\delta'} \right] \\
&\leq T^{-q} \sum_{i=1}^{p+q} \left[E^* \left(\left| \sum_{t=1}^{T-h} \left(h_{it}^2 \eta_{t+h}^* - \underbrace{E^*\left(h_{it}^2 \eta_{t+h}^*\right)}_{=0} \right) \right|^{2+\delta'} \right) \right] \\
&\leq T^{-q} \sum_{i=1}^{p+q} \left[E^* \left(\left| \sum_{t=1}^{T-h} \left(h_{it}^2 \eta_{t+h}^* - E^*\left(h_{it}^2 \eta_{t+h}^*\right) \right)^2 \right|^{(2+\delta')/2} \right) \right],
\end{aligned}
$$

where the third inequality uses the Burholder's inequality. Next by using the Holder's inequality, follows by the c_r-inequality. We obtain

$$
\begin{aligned}
& E^*\left[\lambda_{\max}^{2+\delta'}\left(\frac{1}{T}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\eta_{t+h}^*\right)\right] \\
\leq\ & T^{-(2+\delta')}\sum_{i=1}^{p+q}E^*\left|\left(\sum_{t=1}^{T-h}\left|h_{it}^2\eta_{t+h}^*-E^*\left(h_{it}^2\eta_{t+h}^*\right)\right|^{2+\delta'}\right)^{2/(2+\delta')}(T-h)^{1-2/(2+\delta')}\right|^{(2+\delta')/2} \\
\leq\ & T^{-(2+\delta')}\sum_{i=1}^{p+q}(T-h)^{(2+\delta')/2-1}\sum_{t=1}^{T-h}E^*\left|h_{it}^2\eta_{t+h}^*\right|^{2+\delta'} \\
\leq\ & T^{-(2+\delta')}\sum_{i=1}^{p+q}(T-h)^{(2+\delta')/2-1}\sum_{t=1}^{T-h}\left|h_{it}\right|^{2(2+\delta')}E^*\left|\eta_{t+h}^*\right|^{2+\delta'}.
\end{aligned}
$$

Given that under our assumptions we have $E^*\left|\eta_{t+h}^*\right|^{2+\delta'}\leq\Delta<\infty$, it follows that

$$
\begin{aligned}
E\left|E^*\left[\lambda_{\max}^{2+\delta'}\left(\frac{1}{T}\sum_{t=1}^{T-h}\mathbf{h}_t\mathbf{h}_t'\eta_{t+h}^*\right)\right]\right| &\leq CT^{-(2+\delta')/2}\sum_{i=1}^{p+q}(T-h)^{-1}\sum_{t=1}^{T-h}E\left|h_{it}\right|^{2(2+\delta')} \\
&= O\left(T^{-(2+\delta')/2}\right),
\end{aligned}
$$

since (under Assumption 2) $\sum_{i=1}^{p+q}(T-h)^{-1}\sum_{t=1}^{T-h}E\left|h_{it}\right|^{2(2+\delta')}=O(1)$. This concludes the proof.

References

[1] Andrews DWK. (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica,* 59: 817-858.

[2] Andrews DWK, and C.J. Monahan (1992). "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica,* 60: 953-966.

[3] Bates, J.M. and C.M.W. Granger (1969). "The combination of forecasts," *Operations Research Quarterly,* 20, 451-468.

[4] Brillinger, D. R. (1975). "Time Series, Data Analysis and Theory," *New York: Holt, Rinehart and Winston.*

[5] Breiman, L. (1996). "Bagging predictors," *Machine Learning,* 24, 123-140.

[6] Brown B. W., and S. Maital (1981). "What do Economists Know? An Empirical Study of Experts' Expectations," *Econometrica,* 49, 491-504

[7] Brüggemann, P., Jentsch, C., and Trenkler, C. (2016). "Inference in VARs with conditional volatility of unknown form," *Journal of Econometrics,* 191, 69-85.

[8] Bühlmann, P. and H. R. Künsch (1995). "The blockwise bootstrap for general parameters of a stationary time series," *Scandinavian Journal of Statistics*, 22(1), 35-54.

[9] Bühlmann, P., Yu, B. (2002). “Analyzing bagging,” *Annals of Statistics* 30, 927-961.

[10] Carlstein, E. (1986). “The Use of Subseries Values for Estimating the Variance of a General Statistic From a Stationary Time Series,” *The Annals of Statistics*, 14, 1171-1179.

[11] Cheng, X., and B. Hansen (2015). “Forecasting with factor-augmented regression: A frequentist model averaging approach,” *Journal of Econometrics,* 186, 280-293.

[12] D’Agostino, A., McQuinn, K., and K. Whelan (2015), “Are Some Forecasters Really Better than Others?,”*Journal of Money, Credit and Banking,* 44, 715-732.

[13] Diebold, F.X. (2007). “Elements of Forecasting. Fourth Edition,” *Mason, OH: Thomson South-Western.*

[14] Djogbenou, A., Gonçalves, S., and B. Perron (2015), “Bootstrap Inference for Regressions with Estimated Factors and Serial Correlation,” *Journal of Time Series Analysis,* 36, 481-502.

[15] Efron, B. (1979). “Bootstrap Methods: Another look at the jackknife,” *The Annals of Statistics,* 7, 1-26.

[16] Elliottt, G. and A. Timmermann (2016). “Economic Forecasting,” *Princeton: Princeton University Press.*

[17] Elliott, G., Gargano, A., Timmermann, A., (2013). “Complete subset regressions,” *Journal of Econometrics,* 177, 357-373.

[18] Fitzenberger, B. (1998). “The moving blocks bootstrap and robust inference for linear least squares and quantile regressions,” *Journal of Econometrics,* 82, 235-287.

[19] Genre, V., G. Kenny, A. Meyler and A. Timmermann (2013). “Combining expert forecasts: can anything beat the simple average?,” *International Journal of Forecasting* 29, 108-121

[20] Gonçalves, S. (2011). “The Moving Blocks Bootstrap for Panel Linear Regression Models With Individual Fixed Effects,” *Econometric Theory*, 27 (5), 1048-1082.

[21] Gonçalves, S., Hounyo, U., Patton, A.J., Sheppard, K., (2019). “Bootstrapping two-stage quasi-maximum likelihood estimators of time series models,” *Working paper, McGill University.*

[22] Gonçalves, S. and L. Kilian (2004). “Bootstrapping autoregressions with conditional heteroscedasticity of unknown form,” *Journal of Econometrics* 123, 89-120.

[23] Gonçalves, S. and L. Kilian (2007). “Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity,” *Econometric Reviews* 26, 609-641.

[24] Gonçalves, S. and B. Perron (2014). “Bootstrapping factor-augmented regression models,” *Journal of Econometrics* 182, 156-173.

[25] Gonçalves, S. and B. Perron (2020). “Bootstrapping factor models with cross sectional dependence,” *Journal of Econometrics* 218, 476-495.

[26] Gonçalves, S., and H. White (2002). "The bootstrap of the mean for dependent heterogeneous arrays," *Econometric Theory*, 18, 1367-1384.

[27] Gonçalves, S. and H. White (2005). "Bootstrap standard error estimates for linear regressions," *Journal of the American Statistical Association* Vol. 100, 970-979.

[28] Gospodinov, N. and S. Ng, (2013). "Commodity Prices, Convenience Yields, and Inflation," *Review of Economics and Statistics,* 95, 206-219.

[29] Granger, C.W.J. and Y. Jeon (2004). "Thick modeling," *Economic Modelling* 21, 323-343.

[30] Granger, C.W.J. and R. Ramanathan (1984). "Improved methods of combining forecast accuracy," *Journal of Forecasting,* 3, 197-204

[31] Hahn J. and Z. Liao (2019). "Bootstrap Standard Error Estimates and Inference," *Working paper, UCLA*

[32] Hansen, B.E., (2000). "Testing for structural change in conditional models," *Journal of Econometrics,* 97, 93-115.

[33] Hansen, B.E., (2007). "Least squares model averaging," *Econometrica,* 75, 1175-1189.

[34] Hansen, B.E., (2008). "Least squares forecasting averaging," *Journal of Econometrics,* 146, 342-350.

[35] Hansen, B. E., (2014). "Model averaging, asymptotic risk, and regressor groups," *Quantitative Economics,* 5, 495-530.

[36] Hansen, B. E. and J. S. Racine (2018). "Bootstrap Model Averaging Unit Root Inference," *Working paper, University of Wisconsin*.

[37] Hjort, N. L. and G. Claeskens (2003). "Frequentist model average estimators," *Journal of the American Statistical Association,* 98, 879-899.

[38] Hounyo, U. (2017). "Bootstrapping integrated covariance matrix estimators in noisy jump-diffusion models with non-synchronous trading," *Journal of Econometrics, 197(1), 130-152.*

[39] Hounyo, U. , Gonçalves, S., and N. Meddahi (2017). "Bootstrapping pre-averaged realized volatility under market microstructure noise," *Econometric Theory*, 33 (4), 791-838.

[40] Kapetanios, G., (2008). "A bootstrap procedure for panel data sets with many cross-sectional units," *Econometrics Journal,* 11, 377-395.

[41] Kilian, L. and H. Lüetkepohl (2017). "Structural vector autoregressive analysis," *Themes in Modern Econometrics, Cambridge University Press.*

[42] Kozicki, S., (1999). "How useful are Taylor rules for monetary policy?," *Economic Review, Federal Reserve Bank of Kansas City.*

[43] Inoue, A. and L. Kilian (2008). "How useful is bagging in forecasting economic time series? A case study of U.S. consumer inflation," *Journal of the American Statistical Association*, 103, 511-522.

[44] Kreiss, J. P. (1997). "Asymptotic properties of residual bootstrap for autoregressions," *Working paper, Institute for Mathematical Stochastics, Technical University of Braunschweig, Germany.*

[45] Künsch, H.R. (1989). "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics* 17, 1217-1241.

[46] Lahiri, K., H. Peng and Y. Zhao (2017). "Online learning and forecast combination in unbalanced panels," *Econometric Reviews* 36, 257-288.

[47] Lahiri, S. N. (2003). "Resampling Methods for Dependent Data," *New York: Springer.*

[48] Liu, C-A. (2015). "Distribution theory of the least-squares averaging estimator," *Journal of Econometrics* 186, 142-59.

[49] Liu, R.Y. (1988). "Bootstrap procedure under some non-i.i.d. models," *Annals of Statistics* 16, 1696-1708.

[50] Liu, C. A. and B. S. Kuo. (2016). "Model averaging in predictive regressions," *The Econometrics Journal* 19, 203-31.

[51] Liu, R. Y., and Singh, K. (1992). "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," *in Exploring the Limits of Bootstrap, eds. R. LePage and L. Billard, New York: Wiley*, pp. 225-248.

[52] Maddala, G.S., and S. Wu, (1999). "A comparative study for unit root tests with panel data and a simple new test," *Oxford Bulletin of Economics and Statistics,* 61, 631-652.

[53] Mark, Nelson C., (1995). "Exchange rates and fundamentals: evidence on long-horizon predictability," *American Economic Review*, 85, 201-21.

[54] Mammen, E. (1993). "Bootstrap and wild bootstrap for high dimensional linear models," *Annals of Statistics*, 21, 255-285.

[55] Montiel Olea, J. L. and M. Plagborg-Moller (2019) "Local Projection Inference is Simpler and More Robust Than You Think," *Working paper, Princeton University.*

[56] Nordman, D. J. and Lahiri, S. N. (2012). "Block bootstrap methods for time series regression with fixed regressors," *Journal of the American Statistical Association*, 107, 233-246.

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

[58] Pötscher, B. (2006). "The distribution of model averaging estimators and an impossibility result regarding its estimation," *Lecture Notes-Monograph Series*, 52, 113-129.

[59] Rapach, David E., and Guofu Zhou, (2013). “Forecasting stock returns,” In: Graham Elliott and Allan Timmermann, eds: *Handbook of Economic Forecasting,* Volume 2, Part A, (Elsevier, Amsterdam), 328-383

[60] Shao, X. (2010). “The dependent wild bootstrap,” *Journal of the American Statistical Association,* 105, 218-235.

[61] Shao, X. (2011). “A bootstrap-assisted spectral test of white noise under unknown dependence,” *Journal of Econometrics,* 162, 213-224.

[62] Smith, J. and K.F. Wallis (2009). “A Simple Explanation of the Forecast Combination Puzzle,” *Oxford Bulletin of Economics and Statistics,* 71, 331-355.

[63] Stock, J.H., Watson, M.W., (1999). “A comparison of linear and non-linear univariate models for forecasting macroeconomic time series,” *In Engle R, White H. eds. Cointegration,Causality and Forecasting: A Festschrift in Honor of Clive W. J. Granger. Oxford University Press. pp. 1-44.*

[64] Timmermann, A. (2006). “Forecast combinations,” In Elliott, G., C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting.* Elsevier, 135-196.

[65] Wu, C.F.J., (1986). “Jackknife, bootstrap and other resampling methods in regression analysis,” *Annals of Statistics,* 14, 1261-1295.

[66] Yeh AB., (1998). “A bootstrap procedure in linear regression with nonstationary errors,” *The Canadian Journal of Statistics Association,* 26(1): 149-160.