ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Bootstrapping out-of-sample predictability tests with real-time data

Authors	Silvia Goncalves, Michael W. McCracken, and Yongxu Yao
Working Paper Number	2023-029B
Revision Date	September 2024
Citable Link	https://doi.org/10.20955/wp.2023.029
Suggested Citation	Goncalves, S., McCracken, M.W., Yao, Y., 2024; Bootstrapping out-of-sample predictability tests with real-time data, Federal Reserve Bank of St. Louis Working Paper 2023-029. URL https://doi.org/10.20955/wp.2023.029

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

Bootstrapping out-of-sample predictability tests with real-time data*

Sílvia Gonçalves,[†] Michael W. McCracken,[‡] and Yongxu Yao[§]

September 3, 2024

Abstract

In this paper we develop a block bootstrap approach to out-of-sample inference when real-time data are used to produce forecasts. In particular, we establish its first-order asymptotic validity for West-type (1996) tests of predictive ability in the presence of regular data revisions. This allows the user to conduct asymptotically valid inference without having to estimate the asymptotic variances derived in Clark and McCracken's (2009) extension of West (1996) when data are subject to revision. Monte Carlo experiments indicate that the bootstrap can provide satisfactory finite sample size and power even in modest sample sizes. We conclude with an application to inflation forecasting that revisits the results in Ang et al. (2007) in the presence of real-time data.

*We thank participants at the Conference on Computational and Financial Econometrics (2022), Society for Financial Econometrics (2022), International Association for Applied Econometrics (2022), Asian Meeting of the Econometric Society (2022), Canadian Economics Association (2022), Dolomite Macro workshop (2023), NBER/NSF Time Series Conference (2023), CIREQ Econometrics conference in honor of Eric Ghysels (2024), Workshop on Resampling methods at the University of Warwick (2024), QFFE conference in Marseille (2024), ISNPS (2024), and seminar participants at Indiana University, Saint Mary's University, Southwestern University of Finance and Economics, Rutgers University, Central Bank of Hungary, University of Kentucky, Emory University, McGill University, University of North Carolina, University of California at Davis, ISEG-Lisbon, and the Chamberlain Seminar. The views expressed here are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

[†]McGill University, Department of Economics, 855 Sherbrooke St. W., Montréal, Québec, H3A 2T7, Canada. E-mail: silvia.goncalves@mcgill.ca.

[‡]Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; email: michael.w.mccracken@stls.frb.org.

[§]McGill University, Department of Economics, 855 Sherbrooke St. W., Montréal, Québec, H3A 2T7, Canada. E-mail: yongxu.yao@mail.mcgill.ca

1 Introduction

Real-time vintage data are often used to construct forecasts. At central banks, such as the Federal Reserve and the European Central Bank, this is due to the real-time nature of their forecasting problem. For example, at each FOMC briefing and General Council meeting, it is expected that the staff will have taken on board the latest data when constructing their forecasts. This is important not only because new data are continuously being released but also because previously released macroeconomic data are revised and the revised data are generally considered more accurate. Documenting the importance of real-time data to central bank forecasting was one of the main motivations for the development of the Real-Time Dataset for Macroeconomists (RTDSM) at the Federal Reserve Bank of Philadelphia (Croushore and Stark, 2001). In particular, while Stark and Croushore (2002) and Croushore (2006) emphasize that data revisions obviously change the conditioning information available to the forecasting agent, the revisions can affect parameter estimates and even the functional form of the predictive model.

With this in mind, there is an increasing emphasis on evaluating central bank-, survey-, and model-based forecasts using real-time vintage data. For example, Faust and Wright (2009) evaluate the efficiency of Greenbook forecasts using a sequence of historical datasets archived at the Federal Reserve Board of Governors. Faust, Rogers, and Wright (2005) investigate prospects for real-time exchange rate forecasting using vintages of data available at the OECD's database of Main Economic Indicators. Chauvet and Piger (2008) illustrate the real-time accuracy of recession forecasts using the ALFRED database hosted by the Federal Reserve Bank of St. Louis. Croushore (2011) provides a review of other issues related to real-time forecasting and monetary policy evaluation in the context of the RTDSM. While much of this empirical literature focuses on the US or the Euro area, Garratt et al. (2011) investigate real-time forecasting issues associated with data for Australia, New Zealand, and Norway.

Unfortunately, while the empirical forecasting literature has adapted to the real-time nature of macroeconomic data, the bulk of the theoretical literature on forecast evaluation largely ignores it. Examples like West (1996), Clark and McCracken (2001), Corradi, Swanson, and Olivetti (2001), Giacomini and Rossi (2010), and Odendahl, Rossi and Sekhposyan (2022) each ignore the possibility that at any given forecast origin the most recent data may be subject to revision. This is an issue because out-of-sample tests of predictive ability are operationally distinct from in-sample tests, in ways that make out-of-sample tests particularly susceptible to changes in the correlation structure of the data as the revision process unfolds. This susceptibility has three sources: (i) while parameter estimates are typically functions of only a small number of observations that remain subject to revision, out-of-sample statistics are functions of a sequence of parameter estimates (one for each forecast origin $t = R, \ldots, T$), (ii) the predictors used to generate the forecast and (iii) the realization of the dependent variable used to construct the forecast error may be subject to revision, and hence a sequence of revisions contribute to the test statistic. If data subject to revision possess a different mean and covariance structure than final revised data (e.g., Aruoba (2008)), tests of predictive ability using real-time data may have a different asymptotic distribution than tests constructed using data that are never revised.

In the context of OLS-estimated linear models, Clark and McCracken (2009) show that real-time data can affect tests of equal predictive ability under quadratic loss. In particular, they rederive the results in West (1996) but allow for forecasts that are constructed sequentially across vintages of real-time data. They find that data revisions can lead to substantial changes in asymptotic variances for asymptotically normal tests. In addition, they find that in the context of nested model comparisons, tests can be asymptotically normal even when they were not in the absence of revisions. Subsequently, the analytical results indicate that ignoring the real-time nature of the data can lead to large size distortions and substantial reductions in power. These distortions can arise when revisions are best categorized as news, in the sense of Mankiw, Runkle, and Shapiro (1984), but are most likely to occur when the revisions contain at least some element of noise which, in turn, causes forecast errors to be correlated with predictors.

The main contribution of this paper is to provide a bootstrap approach to conducting inference in a similar framework to that in Clark and McCracken (2009). When appropriately implemented, the bootstrap allows for valid out-of-sample inference without having to estimate the somewhat complicated asymptotic variances derived by Clark and McCracken (2009). More specifically, we provide analytical, Monte Carlo and empirical evidence on the effectiveness of a new block bootstrap approach to out-of-sample inference when forecasts are constructed using real-time vintage data. In many ways our bootstrap is unique. There exists no other bootstrap specifically designed for vintage data. Our block bootstrap treats those observations near the end of a given vintage of data as fundamentally distinct from older, fully revised observations. By doing so, we are able to mimic the triangular array of real-time data in each bootstrap sample.

While our bootstrap is new, it clearly builds on previous work, including Calhoun (2015) who develops a bootstrap that can be used for conducting inference on asymptotically normal out-of-sample tests of predictive ability. The procedure allows for estimation error to contribute to the asymptotic distribution as derived in West (1996) but only in the same environment as West - one that does not allow for real-time vintages of data. It is also related to that of Corradi and Swanson

(2007) who propose a block bootstrap procedure for predictive inference based on recursive estimation schemes. Like they do, we must account for the fact that as we move across forecast origins the sample size increases and hence some observations are used more frequently than others, and this affects the design of the bootstrap statistic (in particular, it requires a careful choice of the centering constant).

Monte Carlo simulations indicate that our bootstrap can provide accurately sized tests of predictive ability in the presence of revisions. Even so, the bootstrap algorithm has limitations and the assumptions underlying its validity are obvious approximations to reality. Specifically, we assume that revisions are finite lived and regular across all vintages. A few series, like non-seasonally adjusted initial and continuing claims, as well as housing starts, experience revisions that align with our assumptions. However, for many other macroeconomic variables like employment, consumption, and industrial production, our assumptions only partially hold in the sense that for most vintages, only a small handful of the most recent observations are revised. That said, these series also have a once-a-year annual benchmark revision in which many more historical observations are revised.[1] As such, the revision process is not regular across all vintages. Regardless, by abstracting from benchmark revisions we are able to take one step toward developing a bootstrap approach to inference when data revisions are present. Later we use simulations to assess the impact of this simplifying assumption on our bootstrap algorithm and its ability to provide accurately sized and powerful tests.

We then apply our bootstrap in the context of comparing the forecast accuracy of models used to forecast CPI– and PCE–based inflation. Specifically, we revisit a small subset of the results in Ang et al. (2007) in which they conclude that survey-based forecasts of CPI inflation are more accurate than a variety of model-based forecasts so long as the surveys are associated with the actual target variable (CPI rather than PCE inflation). Our goal is to see if their conclusions are robust to the presence of real-time data, which they do not consider. Over a common forecast sample, our results align closely with theirs despite the presence of real-time data. Over a longer, and more recent sample our results continue to align with theirs for CPI inflation but not for PCE inflation though a simple explanation is evident: CPI and PCE inflation are very highly correlated over the recent sample but are not in the earlier sample. We conclude with an analysis comparing the forecast accuracy of models to survey forecasts of PCE inflation over a recent but shorter sample. These too support the value of using survey forecasts rather than (admittedly simple) models when forecasting inflation and moreover, using those surveys associated with the actual target variable.

Before proceeding we first clarify that our bootstrap is designed for inference when forecasts are

[1] Consider US payroll employment. In every month other than June, there is a new initial release for the current month and the values from the previous two months are revised. In June, that same pattern is augmented by revisions to the previous five years worth of data. Hence, for 11 months of the year our assumptions are satisfied but in general are not due to the annual benchmark revision.

constructed using a common, but very specific usage of vintage data. We assume that at a given forecast origin, the current vintage of the observables is solely used to form the forecast. This implies that the parameter estimates are estimated using some data that have just been released, some that have been revised once (or more), and some that have been fully revised. Alternative approaches exist. Koenig, Dolmas, and Piger (2003) and Clements and Galvão (2013) estimate models using historical observations with a common level of revision. For example, one might estimate model parameters using only initial releases of a series. This entails using one observation each from the current and previous vintages – in contrast to our approach which uses many values from a single vintage. For a discussion on the costs and benefits of these two approaches, see Clements and Galvão (2019).

The remainder of the paper proceeds as follows. Section 2 introduces the notation and describes the forecasting environment. Section 3 presents the assumptions. Section 4 delineates the asymptotic distribution of the proposed test statistic. Section 5 describes the bootstrap algorithms and characterizes the asymptotic properties of our bootstrap approach to inference. Section 6 presents Monte Carlo evidence on the finite sample size and power of the test using both asymptotic and bootstrap critical values. Section 7 illustrates our bootstrap approach to inference in the context of real-time inflation forecasting. Section 8 concludes. An Appendix contains proofs of the theoretical results.

2 Framework

We follow the revision structure described in Clark and McCracken (2009). At each forecast origin $t = R, \ldots, T$, forecasts of a scalar target variable y are made using a finite dimensioned vector of predictors x based on the current vintage of data $\{y_s(t), x_s(t) : s = 1, \ldots, t\}$. We assume that after a finite number of r releases the revisions are final. For example, we let $y_{s|j}$ denote the j^{th} release of y_s, and hence $y_{s|1}, \ldots, y_{s|r-1}$ are all preliminary values of y_s while $y_{s|r}$ corresponds to the final value of y_s, which we write as $y_{s|r} = y_s$. For each t,

$$y_s(t) = \begin{cases} y_s & \text{for } 1 \leq s \leq t - r + 1 \\ y_{s|j} & \text{for } s = t - j + 1, \quad j = r - 1, \ldots, 1. \end{cases}$$

To illustrate the notation, consider the case of a single revision with $r = 2$. Table 1 provides a description of this data structure. It shows that the data has a triangular structure, where the last observation in each column (vintage) is updated in the following vintage.

The τ-step ahead forecast is formed using linear OLS estimated models $x_t'(t)\hat{\beta}(t)$ where

$$\hat{\beta}(t) \equiv \left(\sum_{s=1+\tau}^{t} x_{s-\tau}(t) x_{s-\tau}'(t) \right)^{-1} \sum_{s=1+\tau}^{t} x_{s-\tau}(t) y_s(t),$$

Table 1: Structure of real-time data with $r = 2$

	Vintage date (t)			
Obs. s	R	$R+1$	$\dots$	$T+1$
1	y_1	y_1		y_1
2	y_2	y_2		y_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$R-2$	y_{R-2}			
$R-1$	y_{R-1}	y_{R-1}		y_{R-1}
R	$y_{R\|1}$	y_R		y_R
$R+1$		$y_{R+1\|1}$		$\vdots$
$\vdots$				
T				y_T
$T+1$				$y_{T+1\|1}$

and the bracket notation in $\hat{\beta}(t)$ is used to emphasize dependence on vintage t data. The forecasts are evaluated against $y_{t+\tau|r'}$, the r'th release of the target variable $y_{t+\tau}$, where $r' \in \{1, \dots, r\}$. The choice of r' is context specific. For instance, if $\tau = 1$ and $r' = 1$, the forecast is evaluated against the first release of y_{t+1}, $y_{t+1|1} = y_{t+1}(t+1)$. But if $r' = 2$, the forecast is evaluated against $y_{t+1|2} = y_{t+1}(t+2)$, the second release of y_{t+1} where $y_{t+1}(t+2) = y_{t+1}$ if $r = 2$.

Given a sequence of real-time forecasts, one is interested in testing the scalar null hypothesis

$$H_0 : Ef\left(y_{t+\tau|r'}, x_t(t), \beta_0\right) = 0,$$

for a known function $f(.)$, which we will often simplify as $f_{t+\tau|r'} = f\left(y_{t+\tau|r'}, x_t(t), \beta_0\right)$, and for a given realization horizon r' for the target variable. Looking ahead toward our bootstrap, this implies that $f_{t+\tau|r'}$ depends on data from two different vintages: the vintage containing $y_{t+\tau|r'}$ and the vintage containing $x_t(t)$.

As a simple example, for a test of zero-mean prediction error $Eu_{t+\tau|r'} = 0$, we have $f_{t+\tau|r'} = u_{t+\tau|r'} = y_{t+\tau|r'} - x_t'(t)\beta_0$. A more complex example is a test of equal predictive ability between two models. For example, let $x_{i,t}'(t)\hat{\beta}_i(t)$, $i = 1, 2$, denote forecasts that we evaluate under quadratic loss. In this case, the null is $E(u_{1,t+\tau|r'}^2 - u_{2,t+\tau|r'}^2) = 0$ and the subsequent function is

$$f_{t+\tau|r'} = \left(y_{t+\tau|r'} - x_{1,t}'(t)\beta_{1,0}\right)^2 - \left(y_{t+\tau|r'} - x_{2,t}'(t)\beta_{2,0}\right)^2,$$

where we have now defined $\beta_0 = (\beta_{1,0}', \beta_{2,0}')'$ and $x_t(t) = (x_{1,t}'(t), x_{2,t}'(t))'$ so that we account for the parameters and predictors from both models, each of which contributes to $f_{t+\tau|r'}$.

To test the null hypothesis we form a test statistic based on the finite sample analogue of $f_{t+\tau|r'}$, where β_0 is replaced by $\hat{\beta}(t)$:

$$\hat{S}_P = P^{-1/2} \sum_{t=R}^{T} f\Big(y_{t+\tau|r'}, x_t(t), \hat{\beta}(t)\Big).$$

The first $P = T - R + 1$ vintages are used to produce τ-step-ahead forecasts and $P - \tau - r'$ of these vintages are also used to evaluate the forecasts. An additional $\tau + r' - 1$ vintages are used only for forecast evaluation implying a total of $P + \tau + r' - 1$ vintages.

In the context of tests of equal predictive ability between two OLS estimated linear models, Clark and McCracken (2009) show that $\hat{S}_P$ can be asymptotically normal with an asymptotic variance reminiscent of that developed in West (1996). Their method of proof, which we will follow, takes advantage of the assumption that the revision process is finite lived. This is a useful approximation because it implies that under additional mixing and moment conditions, $\hat{\beta}(t)$ is asymptotically equivalent to the estimator that uses fully revised data only,

$$\hat{\beta}_t \equiv \Big(\sum_{s=1+\tau}^{t} x_{s-\tau}x'_{s-\tau}\Big)^{-1} \sum_{s=1+\tau}^{t} x_{s-\tau}y_s.$$

Absent this assumption, we would have to consider the possibility that the probability limit of $\hat{\beta}(t)$, $\beta_0 = (E(x_{s-\tau}x'_{s-\tau}))^{-1}E(x_{s-\tau}y_s)$, would vary across forecast origins t due to the revision process itself rather than some underlying changes in the data-generating process for the fully revised data. With this in mind, in the following two sections we first extend the results in Clark and McCracken (2009), which only considered tests of equal predictive ability between two models under quadratic loss, to a wider range of functions $f_{t+\tau|r'}$. We then delineate our bootstrap algorithm and show how it allows us to replicate the null asymptotic distribution of $\hat{S}_P$ and thus be able to conduct inference using a percentile bootstrap and avoid the need to estimate the somewhat complicated asymptotic variances.

3 Assumptions

In this section, we introduce the assumptions that allow us to obtain the asymptotic distribution of $\hat{S}_P$. Throughout we let $f_{t+\tau|r',\beta} \equiv \frac{\partial}{\partial\beta'} f\left(y_{t+\tau|r'}, x_t(t), \beta_0\right)$ and define $F \equiv Ef_{t+\tau|r',\beta}$. Our assumptions are comparable to those in Clark and McCracken (2009), which adapt the results in West (1996) to environments where revisions are present.

Assumption 1 *In an open neighborhood $\mathcal{N}$ around β_0 and with probability 1, (a) $f_{t+\tau|r'}(\beta)$ is measurable and twice continuously differentiable. (b) There exists a constant $D < \infty$ such that for all t, $\sup_{\beta\in\mathcal{N}} \left|\frac{\partial^2 f_{t+\tau|r'}(\beta)}{\partial\beta\partial\beta'}\right| < m_{t+\tau}$ with a measurable function $m_{t+\tau}$ such that $E(m_{t+\tau}) < D$.*

Assumption 1 ensures that the function $f_{t+\tau|r'}$ is well approximated by a quadratic function in the neighborhood of β_0. Without data revisions, it corresponds to Assumption 1 in West (1996). As West (1996) remarks, this assumption is automatically satisfied if the function is a squared forecast error and the model is linear, provided second moments of the target variable $y_{t+\tau|r'}$ and predictors $x_t(t)$ exist. Regardless, the assumptions are general enough to permit a wider range of functions – and loss functions in the context of tests of equal predictive ability – so long as they are sufficiently well behaved in a neighborhood of β_0. For example, given sufficient moment conditions, linear models evaluated under linear-exponential (linex) loss are also permitted.

Assumption 2 *For each model $i = 1, \ldots, k$, where k is finite, the following holds. (a) The final-data estimate $\hat{\beta}_{i,t}$ satisfies $\hat{\beta}_{i,t} - \beta_{i,0} = B_i(t)H_i(t)$, where*

$$B_i(t) = \left(t^{-1} \sum_{s=1+\tau}^{t} x_{i,s-\tau} x'_{i,s-\tau} \right)^{-1} \xrightarrow{a.s.} B_i, \text{ and } H_i(t) = t^{-1} \sum_{s=1+\tau}^{t} h_{i,s} \text{ with } E(h_{i,s}) = 0,$$

where $B_i = \left(E(x_{i,s}x'_{i,s})\right)^{-1}$ and $h_{i,s} = x_{i,s-\tau}(y_s - x'_{i,s-\tau}\beta_{i,0})$. (b) The real-time data estimate $\hat{\beta}_i(t)$ satisfies $\hat{\beta}_i(t) - \beta_{i,0} = \hat{B}_i(t)\hat{H}_i(t)$, where

$$\hat{B}_i(t) = \left(t^{-1} \sum_{s=1+\tau}^{t} x_{i,s-\tau}(t) x_{i,s-\tau}(t)' \right)^{-1} \text{ and } \hat{H}_i(t) = t^{-1} \sum_{s=1+\tau}^{t} h_{i,s}(t),$$

with $h_{i,s}(t) = x_{i,s-\tau}(t)(y_s(t) - x_{i,s-\tau}(t)'\beta_{i,0})$.

The first part of Assumption 2 is a special case of Assumption 2 in West (1996), wherein multiple estimated models are allowed, but with the restriction that those parameters are estimated by OLS using fully revised data. The second part of Assumption 2 defines the real-time estimator $\hat{\beta}(t)$ as an OLS estimator based on vintage t data. Clark and McCracken (2009) rely on a similar assumption (see in particular their Assumption A1). Note that we focus exclusively on the recursive scheme when estimating model parameters. As we discuss later in the context of the bootstrap algorithm, the choice of centering constant depends explicitly on this restriction.

While our results allow for multiple models, as would be the case in a test of equal forecast accuracy, it is convenient to consolidate notation to a single parameter vector $\beta_0 = (\beta'_{1,0}, \ldots, \beta'_{k,0})'$ and let $x_t = (x'_{1,t}, \ldots, x'_{k,t})'$. Having done so we define $\hat{\beta}_t$ so that $\hat{\beta}_t - \beta_0 = B(t)H(t)$ where $B(t) = diag(B_i(t))$, $B = diag(B_i)$, $H(t) = (H'_1(t), \ldots, H'_k(t))'$, and $h_s = (h'_{1,s}, \ldots, h'_{k,s})'$. The corresponding notation for $\hat{\beta}(t)$ is analogous so that $\hat{\beta}(t) - \beta_0 = \hat{B}(t)\hat{H}(t)$.

Our next assumption is a moment and dependence assumption on the vector

$$g_{t+\tau|r'} \equiv \left(\ (f_{t+\tau|r'} - Ef_{t+\tau|r'}) \quad (f_{t+\tau|r',\beta} - F) \quad h'_{t+\tau} \quad x'_t - Ex'_t \ \right)'.$$

Assumption 3 *(a) For some $d > 1$ and $\delta > 0$, $\sup_t E \left\| g_{t+\tau|r'} \right\|^{4d+\delta} < \infty$, where $||.||$ denotes the Euclidean norm. (b) $g_{t+\tau|r'}$ is covariance stationary. (c) $\{g_{t+\tau|r'}\}$ is strong mixing with mixing coefficients of size $\frac{-3d}{d-1}$. (d) Ω is positive, where $\Omega \equiv \lim_{P,R\to\infty} Var(P^{-1/2} \sum_{t=R}^{T} \left(f_{t+\tau|r'} - Ef_{t+\tau|r'}\right) + FBP^{-1/2} \sum_{t=R}^{T} H(t))$.*

Assumption 3 modifies Assumption 3 in West (1996) to the context of real-time data. In addition, we strengthen the moment bound by an additional $\delta > 0$, which we use when proving our bootstrap results. This type of strengthening is common in the bootstrap literature, see e.g., Fitzenberger (Theorem 3.1, 1998). Similarly to Clark and McCracken (2009) (see their Assumption A2), we impose a mixing-type condition on the vector $g_{t+\tau|r'}$ which depends on the vintage horizon r' through the function $f_{t+\tau|r'}$ and its derivatives. We also require that Ω is positive as in Assumption A2 (f) of Clark and McCracken (2009). For many applications, like pairwise tests of equal accuracy between non-nested models, this restriction seems unnecessary. However, in the context of nested model comparisons, it is straightforward to show that whether Ω is positive depends on the properties of the revision process – a point we return to later.

Assumption 4 *For some $d > 1$, $r < \infty$ and $j = 1, \ldots, r$, $(y_{t|j}, x'_{t|j})'$ is uniformly L^{4d}-bounded.*[2]

Assumption 4 is the same as Assumption A2(d) of Clark and McCracken (2009). It puts restrictions on the revision process. More specifically, it restricts the total number of releases to be a finite number r. It also requires the released data to have finite moments of order slightly larger than 4. This implicitly restricts the stochastic order of magnitude of the revisions. As noted in the previous section, we rely on this assumption since it implies that $\hat{\beta}_t$ and $\hat{\beta}(t)$ are asymptotically equivalent.

Assumption 5 *$R, P \to \infty$ and $\lim_{P,R\to\infty} \frac{P}{R} \equiv \pi$, where $0 \leq \pi < \infty$.*

Assumption 5 is the same as Assumptions A4 and A4$'$ of Clark and McCracken (2009).

4 Asymptotic results

In this section we establish the asymptotic distribution of $\hat{S}_P$. To do so, we first show that $\hat{S}_P$ is asymptotically equivalent to

$$\tilde{S}_P = P^{-1/2} \sum_{t=R}^{T} f\left(y_{t+\tau|r'}, x_t(t), \hat{\beta}_t\right),$$

a test statistic based on $\hat{\beta}_t$ rather than $\hat{\beta}(t)$. Specifically, we prove the following result.

[2]Here and in what follows, we say that X_t is L^p-bounded if $(E|X_t|^p)^{1/p} < \infty$.

Lemma 4.1 *Under Assumptions 1-5, $\hat{S}_P = \tilde{S}_P + o_p(1)$.*

By eq. (4.1) of West (1996), adapted to the presence of vintage data, we can expand $\tilde{S}_P - P^{1/2}E(f_{t+\tau|r'})$ as

$$\begin{aligned} \tilde{S}_P - P^{1/2}E(f_{t+\tau|r'}) &= P^{-1/2}\sum_{t=R}^{T}(f_{t+\tau|r'} - E(f_{t+\tau|r'})) + FBP^{-1/2}\sum_{t=R}^{T}H(t) + o_p(1) \\ &\equiv S_{1P} + FBS_{2P} + o_p(1). \end{aligned} \tag{1}$$

The first term S_{1P} in (1) is equal to the scaled average of the demeaned real-time function $f_{t+\tau|r'}$ evaluated at β_0. This piece is asymptotically normally distributed by a central limit theorem (CLT) for strong mixing data, given our assumptions. The second term in (1) depends on S_{2P}, the scaled average of $H(t)$, the average of the scores h_s used in estimating β_0 based on the fully revised data. This piece captures the contribution of the parameter estimation uncertainty and can also be shown to be asymptotically normal, jointly with S_{1P}.

More specifically, following West (1996), S_{1P} and S_{2P} are jointly asymptotically normal, implying that

$$\tilde{S}_P - P^{1/2}E(f_{t+\tau|r'}) \xrightarrow{d} N(0, \Omega),$$

where

$$\Omega = \Omega_1 + FB\Omega_2 B'F' + 2FB\Omega_{12}, \tag{2}$$

with

$$\Omega_1 \equiv \lim_{R,P\to\infty} Var(S_{1P}), \; \Omega_2 \equiv \lim_{R,P\to\infty} Var(S_{2P}), \quad \text{and} \quad \Omega_{12} \equiv \lim_{R,P\to\infty} Cov(S_{1P}, S_{2P}).$$

The form of Ω is notationally the same as the one obtained in West (1996) without data revisions. One main difference is that Ω_1 is now the long-run variance of the scaled average of the real-time function $f_{t+\tau|r'}$ rather than the long-run variance of $f_{t+\tau} \equiv f(y_{t+\tau}, x_t, \beta_0)$, the function associated with fully revised data. Under Assumption 4, the parameter estimation uncertainty as measured by Ω_2 is the same as when all data used in estimation are final. However, its contribution to the overall covariance matrix Ω is different with data revisions. This is because $F \equiv E\left(f_{t+\tau|r',\beta}\right)$ is not necessarily equal to $E(f_{t+\tau,\beta}) \equiv E(\partial f(y_{t+\tau}, x_t, \beta_0)/\partial\beta')$.

The form of Ω shows that the covariance between S_{1P} and S_{2P} need not be asymptotically zero. While perhaps not immediately obvious, this has ramifications for how we design the bootstrap. In particular we take advantage of the fact that S_{2P} can be decomposed into the sum of two asymptotically uncorrelated terms, one of which is not correlated with S_{1P}. Borrowing from page 1081 of West (1996), we write S_{2P} as

$$S_{2P} \equiv P^{-1/2} \sum_{t=R}^{T} H(t) = P^{-1/2} \sum_{s=1+\tau}^{R} a_{R,0} h_s + P^{-1/2} \sum_{i=1}^{P-1} a_{R,i} h_{R+i} \equiv S_{2P.1} + S_{2P.2},$$

where the weights $a_{R,i}$ are defined as $a_{R,i} \equiv \frac{1}{R+i} + \ldots + \frac{1}{R+P-1}$ for $0 \leq i \leq P-1$.

Lemma 4.2 *Under Assumptions 1-5,*

(a) $\Omega_2 \equiv \lim_{R,P\to\infty} Var\left(S_{2P}\right) = \Omega_{2.1} + \Omega_{2.2}$, *where*

$$\begin{aligned} \Omega_{2.1} &\equiv \lim_{R,P\to\infty} Var\left(S_{2P.1}\right) = \lim_{R,P\to\infty} Var(P^{-1/2} \sum_{s=1}^{R} a_{R,0} h_s), \\ \Omega_{2.2} &\equiv \lim_{R,P\to\infty} Var\left(S_{2P.2}\right) = \lim_{R,P\to\infty} Var(P^{-1/2} \sum_{i=1}^{P-1} a_{R,i} h_{R+i}). \end{aligned}$$

(b) $\Omega_{12} \equiv \lim_{R,P\to\infty} Cov\left(S_{1P}, S_{2P}\right)$ *is equal to*

$$\Omega_{12} = \lim_{R,P\to\infty} Cov\left(S_{1P}, S_{2P.2}\right) = \lim_{R,P\to\infty} Cov\left(P^{-1/2} \sum_{t=R}^{T} f_{t+\tau|r'}, P^{-1/2} \sum_{s=1}^{P-1} a_{R,s} h_{R+s}\right).$$

Lemma 4.2 (a) shows that $S_{2P.1}$ is asymptotically uncorrelated with $S_{2P.2}$. Thus, the asymptotic variance of S_{2P} is the sum of the asymptotic variances of $S_{2P.1}$ and $S_{2P.2}$. Lemma 4.2 (b) shows that the asymptotic covariance between S_{1P} and $S_{2P.1}$ is zero. This implies that the covariance between S_{1P} and S_{2P} depends only on the covariance between S_{1P} and $S_{2P.2}$. As noted above, this proves useful when developing our bootstrap method in the next section.

5 Bootstrap results

Here we present a novel bootstrap algorithm and prove its first-order asymptotic validity when used for out-of-sample inference with real-time data. To develop intuition, we first describe our bootstrap algorithm for out-of-sample evaluation of one-step-ahead forecasts based on a simple location model where the data are subject to one revision. We then extend these results to forecast evaluation based on general linear models with forecast horizons possible greater than one and multiple (but finite) revisions.

5.1 A simple location model

Consider first the following location model:

$$y_t = \beta_0 + u_t,\ t = 1, 2, \ldots,$$

where y_t is the fully revised observation. With $r = 2$ releases and one revision, we have

$$y_s(t) = \begin{cases} y_s & 1 \leq s \leq t-1 \\ y_{t|1} & s = t, \end{cases}$$

where $y_{t|1}$ denotes the preliminary (or first release) of the value of y_t according to vintage t. This data structure is the one described in Table 1.

At each forecast origin $t = R, R+1, \ldots, T$, we forecast $y_{t+1}(t+1) = y_{t+1|1}$, next period's value of the first release of y_{t+1}. Hence, $\tau = r' = 1$. The point forecast is $\hat{\beta}(t) = t^{-1}\sum_{s=1}^{t} y_s(t)$ and the null hypothesis is

$$H_0 : E\left(f_{t+1|1}\right) = 0,$$

where $f_{t+1|1} \equiv f\left(y_{t+1|1}, \beta_0\right)$. In this example, $f_{t+1|1}$ depends only on the first-released observations $y_{t+1|1}$. The test statistics $\hat{S}_P$ and $\tilde{S}_P$ are

$$\hat{S}_P = P^{-1/2}\sum_{t=R}^{T} f\left(y_{t+1|1}, \hat{\beta}(t)\right), \text{ and } \tilde{S}_P = P^{-1/2}\sum_{t=R}^{T} f\left(y_{t+1|1}, \hat{\beta}_t\right),$$

where $\hat{\beta}_t = t^{-1}\sum_{s=1}^{t} y_s$ is the estimate of β_0 based on the fully revised data.

When specialized to this example, the asymptotic expansion of $\tilde{S}_P$ is

$$\tilde{S}_P - P^{1/2}E(f_{t+1|1}) = S_{1P} + FS_{2P} + o_p(1),$$

where $F \equiv E\left(f_{t+1|1,\beta}\right) = E(\frac{\partial}{\partial\beta} f\left(y_{t+1|1}, \beta_0\right))$,

$$\begin{aligned} S_{1P} &= P^{-1/2}\sum_{t=R}^{T}(f_{t+1|1} - Ef_{t+1|1}), \text{ and} \\ S_{2P} &= P^{-1/2}\sum_{t=R}^{T} H(t) = P^{-1/2}\sum_{s=1}^{R} a_{R,0}h_s + P^{-1/2}\sum_{i=1}^{P-1} a_{R,i}h_{R+i} \equiv S_{2P.1} + S_{2P.2}, \end{aligned}$$

where $h_s = y_s - \beta_0$ and the weights $a_{R,i}$ are as defined previously. In this simple example, S_{1P} depends on $\left\{y_{t+1|1} : t = R, \ldots, T\right\}$ whereas S_{2P} depends on $\{h_s = y_s - \beta_0 : s = 1, \ldots, T\}$.

Let $\tilde{S}_P^*$ denote a bootstrap version of the original test statistic (we provide more details below). The decision rule is to reject H_0 at level α if $|\tilde{S}_P| \geq c_{1-\alpha}^*$, where $c_{1-\alpha}^*$ is the $100(1-\alpha)^{th}$ percentile of the bootstrap distribution of $|\tilde{S}_P^*|$. To be valid, $\tilde{S}_P^*$ needs to replicate the asymptotic expansion of $\tilde{S}_P$. In particular, the bootstrap statistic needs to replicate the (zero) asymptotic mean and the asymptotic variance Ω of $\tilde{S}_P$.

Next, we propose a bootstrap algorithm that accomplishes this goal. Our algorithm relies on an application of the moving blocks bootstrap (MBB) of Künsch (1989) and Liu and Singh (1992), adapted to the out-of-sample forecasting context under data revisions. We write $\gamma_s \sim$ MBB from $\{1, \ldots, R\}$

to denote a random index that is generated from the set $\{1,\ldots,R\}$ using the MBB. Similarly, we write $\eta_s \sim$ MBB from $\{R+1,\ldots,T+1\}$ to indicate that η_s is obtained by the MBB on the set $\{R+1,\ldots,T+1\}$. We describe below precisely how to generate these indices using the MBB.

Bootstrap algorithm for a location model

1. For $s=1,\ldots,R$, let $\gamma_s \sim$ MBB from $\{1,\ldots,R\}$. For $s=R+1,\ldots,T,T+1$ generate $\eta_s \sim$ MBB from $\{R+1,\ldots,T+1\}$, independently of $\{\gamma_s\}$.

2. For each $t=R,R+1,\ldots,T$, compute

$$\hat{\beta}_t^* = t^{-1}\sum_{s=1}^{t} y_s^*,$$

where

$$y_s^* = \begin{cases} y_{\gamma_s} & \text{if } s=1,\ldots,R, \\ y_{\eta_s} & \text{if } s=R+1,\ldots,T+1. \end{cases}$$

3. For each $t=R,R+1,\ldots,T$, let

$$y_{t+1|1}^* \equiv y_{t+1}^*(t+1) = y_{\eta_{t+1}|1}.$$

and set

$$f_{t+1|1}^*\left(\hat{\beta}_t^*\right) \equiv f\left(y_{t+1|1}^*, \hat{\beta}_t^*\right) = f\left(y_{\eta_{t+1}|1}, \hat{\beta}_t^*\right).$$

4. Compute

$$\tilde{S}_P^* \equiv P^{-1/2}\sum_{t=R}^{T}\left(f_{t+1|1}^*\left(\hat{\beta}_t^*\right) - f_{t+1|1}\left(\bar{\beta}_t\right)\right),$$

where

$$\bar{\beta}_t = \frac{R}{t}\hat{\beta}_R + \frac{t-R}{t}\hat{\beta}_P,$$

with $\hat{\beta}_R \equiv R^{-1}\sum_{t=1}^{R} y_t$ and $\hat{\beta}_P \equiv P^{-1}\sum_{t=R}^{T} y_{t+1}$.

Step 1 is used to obtain the bootstrap analogs of the vintages. More specifically, we generate two sets of random indices: $\{\gamma_s : s=1,\ldots,R\}$ is used to build the first R bootstrap observations $\{y_s^* = y_{\gamma_s} : s=1,\ldots,R\}$ and $\{\eta_s : s=R+1,\ldots,T+1\}$ is used to build the remaining P observations. Since the data are assumed weakly dependent, we rely on the MBB to generate $\{\gamma_s\}$ and $\{\eta_s\}$. (These two sets are generated independently. We will explain below why this is not a problem for the validity of the bootstrap statistic.) More specifically, for a block size equal to l, and assuming that $R=k_1 l$, we generate[3]

$$I_1,\ldots,I_{k_1} \sim \text{i.i.d. Uniform on } \{1,\ldots,R-l+1\}.$$

[3] When R is not divisible by l, we let $k_1 = \lceil R/l \rceil$, the smallest integer that is greater or equal to R/l. We then obtain $R^* = k_1 l$ bootstrap observations. When $R^* \geq R$, we discard the observations in the last block so as to make the number of bootstrap observations equal to R.

These k_1 random variables indicate the beginning of each block. Then, for each $i = 1, \ldots, k_1$ and $j = 1, \ldots, l$, we set

$$\gamma_{1+(i-1)l+(j-1)} = I_i + (j-1),$$

from which we obtain

$$\{\gamma_s : s = 1, \ldots, R\} = \{\gamma_{1+(i-1)l+(j-1)} : i = 1, \ldots, k_1;\ j = 1, \ldots, l\}.$$

Similarly, we generate η_s from the set $\{R+1, \ldots, T+1\}$ using a MBB based on the same block size l.[4] These random indices are used to obtain the remaining P observations. In particular, letting $P = k_2 l$, we generate k_2 uniform draws:

$$J_1, \ldots, J_{k_2} \sim \text{i.i.d Uniform on } \{R+1, \ldots, T+1-l+1\}.$$

For each $i = 1, \ldots, k_2$ and $j = 1, \ldots, l$, we set

$$\eta_{R+1+(i-1)l+(j-1)} = J_i + (j-1),$$

from which we get $\{\eta_s : s = R+1, \ldots, T+1\} = \{\eta_{R+1+(i-1)l+(j-1)} : i = 1, \ldots, k_2;\ j = 1, \ldots, l\}$.

Table 2 provides a description of the bootstrap data structure. This table is the bootstrap analog of Table 1. In this table, for each vintage column t in $R+1, \ldots, T+1$, we set

$$y_s^*(t) = \begin{cases} y_{\gamma_s} & 1 \leq s \leq R \\ y_{\eta_s} & R+1 \leq s < t \\ y_{\eta_s|1} & s = t. \end{cases}$$

We see that except for the first column (vintage R), which sets all the observations in the bootstrap vintage R as final, all the remaining (vintages) columns replicate the triangular structure of the data.[5]

Note that the bootstrap observations indexed by η_s for $s = R+1, \ldots, T+1$ are used both in estimating β_0 (for vintages R+1 and beyond) as well as in evaluating the forecast. For this reason, these observations can be preliminary or final. This is the main reason we introduce a new random index η_s (generated independently of γ_s), which is randomly drawn from $\{R+1, \ldots, T+1\}$. Restricting the support of η_s to $\{R+1, \ldots, T+1\}$ ensures that both y_s and $y_{s|1}$ are available, implying that we can obtain their bootstrap analogs. If instead we had generated a single index η_s from the entire set $\{1, \ldots, R, R+1, \ldots, T+1\}$ (as in the bootstrap schemes of Corradi and Swanson (2007) and Calhoun

[4] Assuming the same block size is for simplicity only. We could allow for different block sizes.

[5] The fact that we do not exactly replicate vintage R's structure is not a problem because this vintage is only used for estimation of β_0 at $t = R$. Hence, this is equivalent to estimating β_0 in this vintage using only final values, i.e. we obtain $\hat{\beta}_R^* = R^{-1} \sum_{s=1}^{R} y_s^*$ rather than $\hat{\beta}^*(R) = R^{-1} \sum_{s=1}^{R} y_s^*(R)$. Since these two estimates are asymptotically equivalent under the assumption of finite revisions, our approach remains valid. When $r > 2$, a similar issue arises for the first $r-1$ vintages but since r is finite and we take limits with respect to both P and R, our approach remains valid.

Table 2: Structure of pseudo real-time data with $r = 2$

	Vintage date (t)			
Obs. s	R	$R+1$	$\cdots$	$T+1$
1	$y_1^* = y_{\gamma_1}$	$y_1^* = y_{\gamma_1}$		$y_1^* = y_{\gamma_1}$
2	$y_2^* = y_{\gamma_2}$	$y_2^* = y_{\gamma_2}$		$y_2^* = y_{\gamma_2}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$R-1$	$y_{R-1}^* = y_{\gamma_{R-1}}$	$y_{R-1}^* = y_{\gamma_{R-1}}$		$y_{R-1}^* = y_{\gamma_{R-1}}$
R	$y_R^* = y_{\gamma_R}$	$y_R^* = y_{\gamma_R}$		$y_R^* = y_{\gamma_R}$
$R+1$		$y_{R+1\|1}^* = y_{\eta_{R+1}\|1}$		$y_{R+1}^* = y_{\eta_{R+1}}$
$\vdots$				
T				$y_T^* = y_{\eta_T}$
$T+1$				$y_{T+1\|1}^* = y_{\eta_{T+1\|1}}$

(2015)), we would not be able to guarantee that a preliminary value is available for all observations. For instance, if $\eta_{R+1} = 1$, we do not observe $y_{1|1}$ and therefore cannot obtain $y_{R+1|1}^* = y_{\eta_{R+1}|1}$. The fact that we need to replicate the original vintage data structure is the crucial distinguishing feature of our paper and the main motivation for proposing a new bootstrap algorithm.

Step 2 obtains a bootstrap analogue of $\hat{\beta}_t$ using only final observations. Focusing on $\tilde{S}_P^*$ (which uses $\hat{\beta}_t^*$) rather than on the real-time bootstrap statistic $\hat{S}_P^*$ (which uses the real-time bootstrap estimate $\hat{\beta}^*(t)$) simplifies the application and the theory of the bootstrap. This approach is justified by Lemma 4.1, which shows the asymptotic equivalence of $\tilde{S}_P$ and $\hat{S}_P$.[6]

Step 3 creates the bootstrap observations used for forecast evaluation. Specifically, for $t = R, \ldots, T$, we let $y_{t+1|1}^* \equiv y_{t+1}^*(t+1) = y_{\eta_{t+1}|1}$ denote the bootstrap analog of $y_{t+1|1}$ and set

$$f_{t+1|1}^*\left(\hat{\beta}_t^*\right) \equiv f\left(y_{t+1|1}^*, \hat{\beta}_t^*\right) = f\left(y_{\eta_{t+1}|1}, \hat{\beta}_t^*\right).$$

This function is the bootstrap analog of $f_{t+1|1}\left(\hat{\beta}_t\right) \equiv f\left(y_{t+1|1}, \hat{\beta}_t\right)$.

Step 4 computes $\tilde{S}_P^*$, the bootstrap analog of $\tilde{S}_P$. This bootstrap test statistic centers $f_{t+1|1}^*\left(\hat{\beta}_t^*\right)$ around $f_{t+1|1}\left(\bar{\beta}_t\right)$, where

$$\bar{\beta}_t = \frac{R}{t}\hat{\beta}_R + \frac{t-R}{t}\hat{\beta}_P, \text{ for } t = R, R+1, \ldots, T,$$

with $\hat{\beta}_R \equiv R^{-1}\sum_{t=1}^R y_t$ and $\hat{\beta}_P \equiv P^{-1}\sum_{t=R}^T y_{t+1}$. Note that both $\hat{\beta}_R$ and $\hat{\beta}_P$ converge in probability to β_0 since $R, P \to \infty$ jointly. This implies that $\bar{\beta}_t \to_p \beta_0$ for $t = R, \ldots, T$.

A naive application of the MBB to the out-of-sample test statistic $\tilde{S}_P$ would suggest we compute

$$S_{P,naive}^* \equiv P^{-1/2}\sum_{t=R}^T\left(f_{t+1|1}^*\left(\hat{\beta}_t^*\right) - f_{t+1|1}\left(\hat{\beta}_t\right)\right).$$

[6] For vintages $t = R+1, \ldots, T+1$, an alternative estimator is $\hat{\beta}^*(t) = t^{-1}\sum_{s=1}^t y_s^*(t)$, the real-time bootstrap analog of $\hat{\beta}(t)$. Since the proofs of bootstrap validity are more involved for this estimator, we do not consider it here.

However, this naive bootstrap statistic is not asymptotically valid. This was first remarked by Corradi and Swanson (2003, 2007) in a context without data revisions. The main reason is that recursive estimation of β_0 implies that earlier observations in the sample are used more frequently than subsequent observations. This implies that $P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_t^* - \hat{\beta}_t)$ does not mimic the distribution of $P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_t - \beta_0)$ when $\hat{\beta}_t$ is recursively estimated. Consequently, $\hat{\beta}_t$ no longer approximates $E^*(\hat{\beta}_t^*)$.[7] Corradi and Swanson (2007) propose a bias correction method to recenter $P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_t^* - \hat{\beta}_t)$ appropriately. Our approach is different: we replace $\hat{\beta}_t$ by $\bar{\beta}_t$ when defining $\tilde{S}_P^*$, and we show that $P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_t^* - \bar{\beta}_t)$ mimics the limiting behavior of $P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_t - \beta_0)$ successfully. In particular, it is easy to show that $\bar{\beta}_t = E^*\left(\hat{\beta}_t^*\right)$ when we generate γ_s and η_s using the i.i.d. bootstrap. More generally, for the MBB, $\bar{\beta}_t = E^*\left(\hat{\beta}_t^*\right) + O_p\left(\frac{l}{\min(P,R)}\right)$, so $\bar{\beta}_t$ is only an approximation to $E^*\left(\hat{\beta}_t^*\right)$. Given the rate conditions in l, this approximation is sufficiently accurate to guarantee that $P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_t^* - \bar{\beta}_t)$ is centered at zero. It's worth re-iterating that this result requires that the centering constant is designed for use under the recursive scheme. If instead, the parameters were estimated using a rolling window of R observations (e.g., $\hat{\beta}\left(t\right) = R^{-1}\sum_{s=t-R+1}^{t} y_s\left(t\right)$), a distinct centering constant would be needed – an issue we do not pursue here.

Next, we explain heuristically why our bootstrap algorithm is asymptotically valid. Let $f_{t+1|1}^* \equiv f_{t+1|1}^*(\beta_0) = f\left(y_{\eta_{t+1}|1}, \beta_0\right)$. By considering two second-order mean value expansions of $f_{t+1|1}(\bar{\beta}_t)$ and $f_{t+1|1}^*(\hat{\beta}_t^*)$, both around β_0, we obtain the following stochastic expansion[8] of $\tilde{S}_P^*$:

$$\begin{aligned}
\tilde{S}_P^* &= P^{-1/2}\sum_{t=R}^{T}\left(f_{t+1|1}^*(\hat{\beta}_t^*) - f_{t+1|1}(\bar{\beta}_t)\right) \\
&= P^{-1/2}\sum_{t=R}^{T}\left(f_{t+1|1}^* - f_{t+1|1}\right) + FP^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_t^* - \bar{\beta}_t) + o_p^*(1) \\
&\equiv S_{1P}^* + FS_{2P}^* + o_p^*(1),
\end{aligned}$$

where S_{1P}^* is the bootstrap analog of S_{1P}, and S_{2P}^* is the bootstrap analog of S_{2P}; this is the bootstrap analog of the asymptotic expansion of $\tilde{S}_P$. We can further decompose S_{2P}^* as follows. Since $\hat{\beta}_t^* = t^{-1}\sum_{s=1}^{t} y_s^*$ and $\bar{\beta}_t = t^{-1}\sum_{s=1}^{t} E^*(y_s^*) + O_p(l/\min R, P)$, S_{2P}^* is asymptotically equivalent to

$$P^{-1/2}\sum_{s=1}^{R} a_{R,0}h_s^* + P^{-1/2}\sum_{i=1}^{P-1} a_{R,i}h_{R+i}^* \equiv S_{2P.1}^* + S_{2P.2}^*,$$

where $h_s^* \equiv y_s^* - E^*\left(y_s^*\right)$.

[7] Here and in the following, an asterisk appearing in E (and Var) denotes expectation (and variance) with respect to the bootstrap probability measure P^*, conditional on the original sample.

[8] See Appendix A for the definition of $o_p^*(1)$, as well as the definition of $\overset{d^*}{\to}$, which is used below.

We now show that $\tilde{S}_P^*$ is approximately centered at zero. First, since $E^*(h_s^*) = 0$ by construction, $E^*(S_{2P}^*) = 0$. Thus, the bootstrap distribution of $P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_t^* - \bar{\beta}_t)$ is asymptotically centered at zero. This solves the bias problem discussed by Corradi and Swanson (2007) (who considered $\hat{\beta}_t$ rather than $\bar{\beta}_t$ when forming their bootstrap statistic in a context without data revisions). Second, we can show that $E^*\left(S_{1P}^*\right) = O_p\left(l/R\right) = o_p\left(1\right)$ if $l/R = o\left(1\right)$. This follows from standard MBB results (see e.g. Fitzenberger, 1998). Thus, $E^*(\tilde{S}_P^*)$ is asymptotically equal to zero.

We can also show that the bootstrap variance of $\tilde{S}_P^*$ mimics the asymptotic variance Ω. To see this, note that $S_{1P}^* \overset{d^*}{\to} N\left(0, \Omega_1\right)$. This follows from Theorem 3.1 of Fitzenberger (1998) given that $f_{t+1|1}^* = f\left(y_{t+1|1}^*, \beta_0\right) = f\left(y_{\eta_{t+1}|1}, \beta_0\right)$ is obtained by applying the MBB to the first released observations $\left\{y_{t+1|1} : t = R, \ldots, T\right\}$. Thus, our bootstrap mimics Ω_1.

Next, we explain why this bootstrap also mimics Ω_2. First, note that since we resample the first R observations independently of the last P observations, the covariance between $S_{2P.1}^*$ and $S_{2P.2}^*$ is zero in the bootstrap world. Thus, the bootstrap mimics the zero asymptotic covariance between $S_{2P.1}$ and $S_{2P.2}$ (which is established in Lemma 4.2(a)). Second, note that the fact that we use the MBB to obtain γ_s and η_s implies that the bootstrap variances of $S_{2P.1}^*$ and $S_{2P.2}^*$ converge to the long-run variances $\Omega_{2.1}$ and $\Omega_{2.2}$, respectively.

Finally, note that our bootstrap also captures the covariance between S_{1P} and $S_{2P.2}$. The independence between γ_s and η_s implies that the bootstrap analogs of S_{1P}^* and $S_{2P.1}^*$ are independent, but this is not a concern since the covariance between S_{1P} and $S_{2P.1}$ is asymptotically zero by Lemma 4.2(b).

5.2 Extension to linear models

In the general framework, we forecast $y_{t+\tau|r'}$ using a linear model with predictors $x_t\left(t\right)$, where the coefficients β_0 are estimated recursively by OLS. The location model is a special case of this setup where $x_s\left(t\right) \equiv 1$ for all s, t, with the difference that the target variable is $y_{t+\tau|r'}$ (rather than $y_{t+1|1}$). The forecast function is now $f_{t+\tau|r'} \equiv f\left(y_{t+\tau|r'}, x_t\left(t\right); \beta_0\right)$, where $x_t\left(t\right)$ contains the predictors for $y_{t+\tau|r'}$ available in vintage t. When lagged dependent variables are present, $x_t\left(t\right)$ may contain a mix of preliminary and final observations.

The main difference with respect to the simple location model is in the bootstrap estimation of β_0. For each vintage $t = R, \ldots, T$ we resample the "pairs" $z_s \equiv (y_s, x_{s-\tau}')'$ used in estimating β_0 at each forecast origin. As in the location model, we estimate $\hat{\beta}_t^*$ using only finally revised bootstrap data. While we state the algorithm in the context of a single model, the extension to multiple models is straightforward.

The bootstrap algorithm for forecasts based on general linear regression models is as follows.

Bootstrap algorithm for linear models

1. Let $R-(1+\tau)+1=k_1 l$ and generate $I_1,\ldots,I_{k_1}\sim$ i.i.d. Uniform on $\{1+\tau,\ldots,R-l+1\}$. Then, for each $i=1,\ldots,k_1$ and $j=1,\ldots,l$, set $I_i+(j-1)=\gamma_{1+\tau+(i-1)l+(j-1)}$ and let

$$\{\gamma_s : s=1+\tau,\ldots,R\}=\{\gamma_{1+\tau+(i-1)l+(j-1)} : i=1,\ldots,k_1;\ j=1,\ldots,l\}.$$

Let $T+\tau-(R+1)+1=k_2 l$ and generate $J_1,\ldots,J_{k_2}\sim$ i.i.d Uniform on $\{R+\tau,\ldots,T+\tau-l+1\}$. For each $i=1,\ldots,k_2$ and $j=1,\ldots,l$, set $J_i+(j-1)=\eta_{R+1+(i-1)l+(j-1)}$, and let

$$\{\eta_s : s=R+1,\ldots,T+\tau\}=\{\eta_{R+1+(i-1)l+(j-1)} : i=1,\ldots,k_2;\ j=1,\ldots,l\}.$$

2. For $t=R,\ldots,T$, set

$$z_s^{*\prime}\equiv(y_s^*,x_{s-\tau}^{*\prime})=\begin{cases}(y_{\gamma_s},x'_{\gamma_s-\tau}) & 1+\tau\le s\le R\\ (y_{\eta_s},x'_{\eta_s-\tau}) & R+1\le s\le t,\end{cases}$$

and compute

$$\hat\beta_t^*=\left(\frac{1}{t}\sum_{s=1+\tau}^{t}x_{s-\tau}^*x_{s-\tau}^{*\prime}\right)^{-1}\left(\frac{1}{t}\sum_{s=1+\tau}^{t}x_{s-\tau}^*y_s^*\right).$$

3. For $t=R,\ldots,T$, let

$$(y^*_{t+\tau|r'},x_t^*(t)')=(y_{\eta_{t+\tau|r'}},x_{\eta_{t+\tau}-\tau}(\eta_{t+\tau}-\tau)'),$$

and compute

$$f^*_{t+\tau|r'}\left(\hat\beta_t^*\right)\equiv f\left(y^*_{t+\tau|r'},x_t^*(t)',\hat\beta_t^*\right).$$

4. Compute

$$\tilde S_P^*\equiv P^{-1/2}\sum_{t=R}^{T}\left(f^*_{t+\tau|r'}(\hat\beta_t^*)-f_{t+\tau|r'}(\bar\beta_t)\right),$$

where $\bar\beta_t=\frac{R}{t}\hat\beta_R+\frac{t-R}{t}\hat\beta_P$, with

$$\hat\beta_R=\left(\frac{1}{R}\sum_{s=1+\tau}^{R}x_{s-\tau}x'_{s-\tau}\right)^{-1}\left(\frac{1}{R}\sum_{s=1+\tau}^{R}x_{s-\tau}y_s\right)$$

and

$$\hat\beta_P=\left(\frac{1}{P}\sum_{s=R+\tau}^{T+\tau}x_{s-\tau}x'_{s-\tau}\right)^{-1}\left(\frac{1}{P}\sum_{s=R+\tau}^{T+\tau}x_{s-\tau}y_s\right).$$

Remark 1 *The presence of the predictors $x_t\,(t)$ when forecasting $y_{t+\tau|r'}$ creates some differences with respect to the simple location model's algorithm. The first difference is that we restrict the support of the MBB indices γ_s to $\{1+\tau,\ldots,R\}$ rather than $\{1,\ldots,R\}$. This is because the recursive estimates of*

β_0 depend on $(y_s, x_{s-\tau})$ for $s = 1+\tau, \ldots, t$. Thus, restricting γ_s this way ensures that we can evaluate $x^*_{s-\tau} \equiv x_{\gamma_s-\tau}$. Setting $x_s(t) = 1$ for all s implies this restriction is not needed. Similarly, we also restrict the support of η_s to the set $\{R+\tau, ..., T+\tau\}$. This ensures that $\eta_s - \tau$ is in the set $\{R, \ldots, T\}$, for which we have both final and preliminary values of the variables. This is particularly important in step 3, where we need to obtain the predictors $x^*_t(t) = x_{\eta_{t+\tau}-\tau}(\eta_{t+\tau} - \tau)$.

Lemma 5.1 *Under Assumptions 1-5 and letting $l \to \infty$ such that $l/\min\{\sqrt{R}, \sqrt{P}\} \to 0$,*

$$\tilde{S}^*_P \equiv P^{-1/2} \sum_{t=R}^{T} \left(f^*_{t+\tau|r'}(\hat{\beta}^*_t) - f_{t+\tau|r'}(\bar{\beta}_t) \right) = S^*_{1P} + FBS^*_{2P} + o^*_p(1),$$

where

$$S^*_{1P} = P^{-1/2} \sum_{t=R}^{T} \left(f^*_{t+\tau|r'} - f_{t+\tau|r'} \right),$$

and

$$S^*_{2P} = a_{R,0} P^{-1/2} \sum_{s=1+\tau}^{R} (h^*_s - \bar{h}_R) + P^{-1/2} \sum_{i=1}^{P-1} a_{R,i} (h^*_{R+i} - \bar{h}_P) \equiv S^*_{2P.1} + S^*_{2P.2},$$

where $h^*_t = x^*_{t-\tau}(y^*_t - x^{*\prime}_{t-\tau}\beta_0)$, $\bar{h}_R = R^{-1}\sum_{s=1+\tau}^{R} h_s$ and $\bar{h}_P = P^{-1}\sum_{s=R+\tau}^{T+\tau} h_s$.

Under our assumptions, $S^*_{1P} \overset{d}{\to} N(0, \Omega_1)$ by Fitzenberger (1998) (cf. Theorem 3.1). As in the simple location model, the term S^*_{2P} has two components, $S^*_{2P.1}$ and $S^*_{2P.2}$, both centered at zero asymptotically. To see this, note that $\bar{h}_R = E^*\left(R^{-1}\sum_{s=1+\tau}^{R} h^*_s\right) + O_P(l/R)$, whereas $\bar{h}_P = E^*\left(P^{-1}\sum_{i=1}^{P-1} h^*_{R+i}\right) + O_P(l/P)$. In addition, we can show that this term's bootstrap variance converges to $\Omega_2 = \Omega_{2.1} + \Omega_{2.2}$. Since the bootstrap covariance between $S^*_{2P.1}$ and S^*_{1P} is zero by construction and we show that the covariance between S^*_{1P} and $S^*_{2P.2}$ is asymptotically equal to Ω_{12}, the following result follows.

Theorem 5.1 *Suppose Assumptions 1-5 hold and $l \to \infty$ such that $l/\min\{\sqrt{R}, \sqrt{P}\} \to 0$. Then,*

$$\sup_{u \in \mathbb{R}} \left| P^*\left(\tilde{S}^*_P \le u\right) - \Pr\left(\tilde{S}^\mu_P \le u\right) \right| \to_p 0,$$

where $\tilde{S}^\mu_P = \tilde{S}_P - P^{1/2} E\left(f_{t+\tau|r'}\right)$.

The proof of Theorem 5.1 relies in part on Lemma A.4 in Appendix A (which shows the consistency of the bootstrap variance estimator of $\tilde{S}^*_P$). Theorem 5.1 implies that our bootstrap algorithm can be used to approximate the asymptotic distribution of $\tilde{S}^\mu_P$, a centered version of $\tilde{S}_P$. When the null hypothesis $H_0 : E\left(f_{t+\tau|r'}\right) = 0$ is true, $\tilde{S}^\mu_P$ coincides with the test statistic $\tilde{S}_P$, in which case Theorem 5.1 proves the consistency of the bootstrap critical values obtained from $\tilde{S}^*_P$. When the null hypothesis does not hold, Theorem 5.1 shows that the bootstrap distribution of $\tilde{S}^*_P$ is consistent for the distribution of $\tilde{S}^\mu_P$, implying that the bootstrap test based on $\tilde{S}^*_P$ has power.

5.3 Bootstrap results for nested linear models

The bootstrap algorithm in the previous section can be applied when conducting a test of equal predictability between nested models. Nevertheless, when the models are nested, the algorithm can be simplified considerably. In the following we present that simplification in the context of tests of equal accuracy under quadratic loss. By doing so we are also able to discuss how the properties of the data revisions affect whether Ω is positive.

In this application the loss differential defines the function $f_{t+\tau|r'}$ and takes the form

$$f_{t+\tau|r'} \equiv f\left(y_{t+\tau|r'}, x_t\left(t\right), \beta_0\right) = \left(y_{t+\tau|r'} - x_{1,t}'\left(t\right)\beta_{1,0}\right)^2 - \left(y_{t+\tau|r'} - x_{2,t}'\left(t\right)\beta_{2,0}\right)^2,$$

where $x_{2,t}\left(t\right) = \left(x_{1,t}\left(t\right)', x_{22,t}\left(t\right)'\right)'$ and $\beta_0 = \left(\beta_{1,0}', \beta_{2,0}'\right)'$. Under the null of equal predictive ability, model 2 includes $\dim(x_{22,s}(t)) = k_{22}$ excess parameters, i.e., $\beta_{2,0} = (\beta_{1,0}', 0')'$ and $x_{1,t}(t)'\beta_{1,0} = x_{2,t}(t)'\beta_{2,0}$.

As we have done before, we let $\hat{\beta}_t = \left(\hat{\beta}_{1,t}', \hat{\beta}_{2,t}'\right)'$ denote the estimators of β_0 based on final data and we let $\hat{\beta}\left(t\right) = \left(\hat{\beta}_1'\left(t\right), \hat{\beta}_2'\left(t\right)\right)'$ denote their real-time data analogs. $\hat{S}_P$ is the test statistic based on $\hat{\beta}\left(t\right)$, and $\tilde{S}_P$ denotes its analog based on $\hat{\beta}_t$. Lemma 4.1 immediately implies that

$$\hat{S}_P = P^{-1/2}\sum_{t=R+1}^{T}((y_{t+\tau|r'} - x_{1,t}'\left(t\right)\hat{\beta}_{1,t})^2 - (y_{t+\tau|r'} - x_{2,t}'\left(t\right)\hat{\beta}_{2,t})^2) + o_p(1) \equiv \tilde{S}_P + o_p\left(1\right).$$

More importantly, since the models are nested we know that under the null, not only is $Ef_{t+\tau|r'} = 0$, but also $f_{t+\tau|r'} = 0$ since $x_{1,t}(t)'\beta_{1,0} = x_{2,t}(t)'\beta_{2,0}$. This makes bootstrapping the distributions of $\hat{S}_P$ and $\tilde{S}_P$ easier since $\tilde{S}_P = FBP^{-1/2}\sum_{t=R}^{T}H(t) + o_p(1)$ and the uncertainty in this term is determined solely by fully revised data, and hence we no longer need to replicate the triangular structure of the different vintages.

Before delineating this bootstrap, it is useful to note that the expansion for $\tilde{S}_P$ simplifies even further under the null hypothesis. Let $F \equiv E[\frac{\partial}{\partial\beta'}f_{t+\tau|r'}\left(\beta\right)] = [F_1, F_2]$, with $F_i \equiv E[\frac{\partial}{\partial\beta_i'}f_{t+\tau|r'}\left(\beta\right)]$ for $i = 1, 2$ and recall that $B = diag(B_i)$. Since the models are nested we know that for a selection matrix $J' = (I_{k_1\times k_1}, 0_{k_1\times k_{22}})$, $H_1(t) = J'H_2(t)$ and $F_1 = -F_2J$ and hence

$$\tilde{S}_P = F_2(-JB_1J' + B_2)P^{-1/2}\sum_{t=R}^{T}H_2(t) + o_p\left(1\right). \tag{3}$$

In addition, noting that for $t = R, \ldots, T$,

$$\hat{\beta}_{2,t} - \beta_{2,0} = B_2\left(t\right)H_2\left(t\right),$$

we can further rewrite (3) as

$$\tilde{S}_P = F_2(-JB_1J'B_2^{-1} + I_{k_2})P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_{2,t} - \beta_{2,0}) + o_p\left(1\right).$$

This expansion shows that we can replicate the distribution of $\tilde{S}_P$ by replicating the distribution of $P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_{2,t}-\beta_{2,0})$. Since this term only depends on OLS estimates from the larger model evaluated with final data, we can rely on Corradi and Swanson's (2007) method to bootstrap its distribution. This combined with a consistent estimator of the factor $F_2(-JB_1J'B_2^{-1}+I_{k_2})$ provides a valid bootstrap method for computing the quantiles of $\tilde{S}_P$ under the null hypothesis.

The bootstrap algorithm is as follows.

Bootstrap algorithm for nested linear models

1. Let $T-(1+\tau)+1=kl$ and generate $I_1,\ldots,I_k\sim$ i.i.d. Uniform on $\{1+\tau,\ldots,T-l+1\}$. Then, for each $i=1,\ldots,k$ and $j=1,\ldots,l$, set $I_i+(j-1)=\gamma_{1+\tau+(i-1)l+(j-1)}$ and let

$$\{\gamma_s : s=1+\tau,\ldots,T\}=\{\gamma_{1+\tau+(i-1)l+(j-1)} : i=1,\ldots,k;\ j=1,\ldots,l\}.$$

2. For $s=1+\tau,\ldots,T$, set

$$z_s^*\equiv(y_s^*,x_{2,s-\tau}^{*\prime})'=(y_{\gamma_s},x'_{2,\gamma_s-\tau})',$$

and for $t=R,\ldots,T$, compute

$$\tilde{\beta}_{2,t}^*=\arg\min_{\beta_2}\left[t^{-1}\sum_{s=1+\tau}^{t}(y_s^*-x_{2,s-\tau}^{*}{}'\beta_2)^2-\beta_2'(T-\tau)^{-1}\sum_{s=1+\tau}^{T}2(y_s-x'_{2,s-\tau}\hat{\beta}_{2,t})x_{2,s-\tau}\right].$$

3. Compute

$$\tilde{S}_P^*=\hat{F}_2(-J\hat{B}_1J'\hat{B}_2^{-1}+I_{k_2})P^{-1/2}\sum_{t=R}^{T}(\tilde{\beta}_{2,t}^*-\hat{\beta}_{2,t}),$$

where $\hat{F}_2=2P^{-1}\sum_{t=R}^{T}(y_{t+\tau|r'}-x'_{2,t}(t)\,\hat{\beta}_{2,T})x'_{2,t}(t)$, and $\hat{B}_i=(T^{-1}\sum_{s=1}^{T}x_{i,s}x'_{i,s})^{-1}$ are consistent estimates of F_2 and B_i for $i=1,2$, respectively.

Steps 1 and 2 amount to using the block bootstrap method of Corradi and Swanson (2007) to replicate the distribution of $P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}_{2,t}-\beta_{2,0})$. Contrary to our previous bootstrap algorithms, we need only one set of MBB indices in Step 1. The main reason for using two sets of MBB indices in Step 1 of our previous methods was the need to replicate the triangular structure of the vintages data. This is no longer required because the term that depends on the function $f_{t+\tau|r'}$ is zero when the models are nested. The other key difference is that we now incorporate a bias correction term in the definition of $\tilde{\beta}_{2,t}^*$, i.e. we do not estimate $\beta_{2,0}$ using the standard OLS estimator. This correction term is one way of correcting for the bias introduced by the recursive estimation scheme in the bootstrap world and was suggested by Corradi and Swanson (2007) in a context without data revisions.

The asymptotic validity of $\tilde{S}_P^*$ follows from Theorem 1 of Corradi and Swanson (2007) and the consistency of $\hat{F}_2$, $\hat{B}_i$ for $i = 1, 2$, provided the condition $F_2(-JB_1J' + B_2) \neq 0$ holds (ensuring that Ω is positive). This last condition is non-trivial and can depend on the statistical properties of the revision process. Following Mankiw, Runkle, and Shapiro (1984), we treat revisions as consisting of news (v) and noise (w) components. A revision is said to be pure news if it is uncorrelated with any data available at the time of the provisional estimate. If the revision is correlated with the provisional estimate, then the revision contains a noise component.

Specifically, in the context of verifying whether $F_2(-JB_1J' + B_2) \neq 0$, consider the following example in which $r = 2$ and an $AR(2)$ model nests an $AR(1)$ model, so that $x_{2,t} = (1, y_t, y_{t-1})'$ and $x_{1,t} = (1, y_t)'$. The fully revised data and initial releases then take the form

$$y_t = \delta_0 + \delta_1 y_{t-1} + v_t + e_t,$$
$$y_t(t) = y_t - v_t + w_t,$$

for error terms e_t, and revision components v_t and w_t, which are mutually independent i.i.d. zero-mean Gaussian variates with variances σ_e^2, σ_v^2, and σ_w^2, respectively. If $\tau = 1$ and the initial release is used for forecast evaluation, this implies that F_2 takes the form

$$\begin{aligned} F_2 &= 2E\left[(y_{t+1|1} - x'_{2,t}(t)\,\beta_2)x'_{2,t}(t)\right] \\ &= 2E\left[(y_{t+1} - v_{t+1} + w_{t+1} - \delta_0 - \delta_1(y_t - v_t + w_t))(1, y_t - v_t + w_t, y_{t-1})\right] \\ &= 2E\left[(e_{t+1} + w_{t+1} - \delta_1(-v_t + w_t))(1, y_t - v_t + w_t, y_{t-1})\right] \\ &= (0, -2\delta_1\sigma_w^2, 0). \end{aligned}$$

We immediately observe a simple instance for which F_2, and hence $F_2(-JB_1J' + B_2)$, will be zero. F_2 is zero if δ_1 is zero, or if there is no noise component to the revision (i.e., $\sigma_w^2 = 0$). While we do not pursue it here, if there is uncertainty about whether F is zero, an alternative approach to asymptotically valid inference is developed in Corradi et al. (2024) albeit in the absence of data-revisions.

6 Monte Carlo simulations

In this section, we consider the finite sample size and power of bootstrap-based inference for tests of equal predictive ability when data are subject to revisions. In each case we use OLS to estimate two predictive models $x'_{i,t}\beta_i$ $i = 1, 2$ and evaluate accuracy under quadratic loss. The design of the experiments is similar to that in Clark and McCracken (2009) but is calibrated to align with our empirics in Section 7.

6.1 Non-nested models

We begin with size and power experiments associated with tests of equal predictive ability between two non-nested models. The final data are generated according to

$$y_t = 0.3x_{1,t-1} + (0.3 + \Delta)x_{2,t-1} + e_{y,t} + v_{y,t},$$

where $e_{y,t}$ and $v_{y,t}$ are independently generated as i.i.d. Gaussian variables with mean zero and variance equal to $\sigma^2_{e,y}$ and $\sigma^2_{v,y}$, respectively. Similarly, we let each regressor's final data be generated as $x_{i,t} = e_{x_i,t} + v_{x_i,t}$, $i = 1, 2$, where $e_{x_i,t}$ and $v_{x_i,t}$ are also mutually independent (jointly with $e_{y,t}$ and $v_{y,t}$) i.i.d. Gaussian random variables with mean zero and variances $\sigma^2_{e,x}$ and $\sigma^2_{v,x}$, respectively. For instance, we can think of y_t as headline PCE inflation and $x_{1,t-1}$ and $x_{2,t-1}$ as real GDP growth and changes to total capacity utilization, as in the empirical application considered later in Section 7. When $\Delta = 0$, the two predictors have the same predictive content for inflation, but not otherwise.

We consider the case of a single revision, where time t's preliminary estimates of y_t and $x_{i,t}$ are

$$y_t(t) = y_t - v_{y,t} + w_{y,t}, \quad \text{and} \quad x_{i,t}(t) = x_{i,t} - v_{x_i,t} + w_{x_i,t} \text{ for } i = 1, 2,$$

with $w_{y,t}$ and $w_{x_i,t}$ denoting i.i.d. Gaussian random variables with variances $\sigma^2_{w,y}$ and $\sigma^2_{w,x}$, respectively. These random variables are also mutually independent (jointly with $e_{y,t}$, $v_{y,t}$, $e_{x_i,t}$ and $v_{x_i,t}$). Following the real-time data literature, we interpret $v_{y,t}$ and $v_{x_i,t}$ as the news components of the revisions to $y_t(t)$ and $x_i(t)$, respectively, whereas $w_{y,t}$ and $w_{x_i,t}$ represent the noise components.[9]

Our goal is to test for equal predictability of $y_{t+1}(t+1) \equiv y_{t+1|1}$ using two non-nested models, each based on a real-time predictor $x_{i,t}(t)$ for $i = 1, 2$. The null hypothesis is then

$$H_0 : E(f_{t+1|1}) \equiv E\left[(y_{t+1|1} - x_{1,t}(t)\beta_{1,0})^2 - (y_{t+1|1} - x_{2,t}(t)\beta_{2,0})^2\right] = 0.$$

This is true when $\Delta = 0$ (used in the experiments describing the size properties of our test), but not otherwise (we set $\Delta = 0.7$ when evaluating power).

We consider two different data-generating processes, DGP1 and DGP2. These include both noise and news components in the revisions and are described as "DGP1, noise and news" and "DGP2, noise and news" in Table 3 below. We also consider two variations of these DGPs, without a noise component, i.e. we set $\sigma^2_{w,y} = \sigma^2_{w,x} = 0$. In this case, we write "DGP1, news only" and "DGP2, news only," respectively.

[9] To explain briefly this terminology, take as an example the equations that describe y_t and $y_t(t)$. Given that $v_{y,t}$ enters y_t and $y_t(t)$ with a positive and negative sign, respectively, the preliminary value $y_t(t)$ does not depend on the component $v_{y,t}$, implying that this term is uncorrelated with $y_t(t)$. This is the sense in which $v_{y,t}$ describes "news". Instead, the presence of $w_{y,t}$ in the preliminary value $y_t(t)$ and its absence from y_t explain why we call this component a "noise" component.

The versions of DGP1 and DGP2 that contain both noise and news differ in the way they parameterize the revisions process. More specifically,

- DGP1: $\sigma^2_{e,y} = 1.69,\ \sigma^2_{e,x} = 0.3,\ \sigma^2_{v,y} = 0.01,\ \sigma^2_{v,x} = 3,\ \sigma^2_{w,y} = 0.03,\ \sigma^2_{w,x} = 3.$
- DGP2: $\sigma^2_{e,y} = 1.69,\ \sigma^2_{e,x} = 3.2,\ \sigma^2_{v,y} = 0.01,\ \sigma^2_{v,x} = 0.1,\ \sigma^2_{w,y} = 0.03,\ \sigma^2_{w,x} = 0.3.$

The variances used in these DGPs are motivated by the empirical application considered in Section 7. Using 2019:Q4 vintage data, we calibrate the variances of the fully revised values of y using observations of annualized quarterly headline PCE inflation (PCE) from 1984:Q1 through 2019:Q4. For x_1 and x_2 we use the same vintage and sample but, for simplicity, assign them a common variance that is roughly the average of the variances of annualized quarterly real GDP growth (RGDP) and the annualized quarterly average of monthly changes to total capacity utilization (TCU). Using the same sample, the slope coefficients used to link y to x_1 and x_2 are roughly those from an OLS regression of PCE on RGDP and TCU. For DGP 2, the revisions are then calibrated based on quarterly vintages of the same three variables from 1996:Q1 through 2019:Q4. Specifically, the variances are those of the first quarterly revision. Again, for simplicity, we assign a common revision variance to x_1 and x_2 that is roughly the average of the revision variance for RGDP and TCU. For DGP1 we consider a much larger revision variance for x_1 and x_2. The larger revision variance can lead to substantial differences in actual power, as we will see when discussing Table 3 below.

At each forecast origin, we use the generated real-time data to make a one-step-ahead forecast for the target variable $y_{t+1|1}$. Forecasts take the form $x'_{i,t}(t)\hat{\beta}_i(t)$ with subsequent forecast errors $\hat{u}_{i,t+1|1} = y_{t+1|1} - x_{i,t}(t)'\hat{\beta}_i(t)$ for $i = 1, 2$. As noted above, accuracy is evaluated under quadratic loss.

Table 3 contains the results. We consider three different tests. One is the bootstrap test described in Section 5.2, which is used to obtain bootstrap critical values for $\hat{S}_P$. This test is a percentile-type test that does not require studentizing the statistic $\hat{S}_P$. It is labeled "Bootstrap" in Table 3. We also include two alternative tests. These tests are based on studentized versions of $\hat{S}_P$ and rely on critical values taken from the standard normal distribution (hence, they do not involve the bootstrap and are included as benchmark methods). One is the Diebold and Mariano (1995) test, labeled $t(\hat{\Omega}_{1P})$ in Table 3. It takes the form $t(\hat{\Omega}_{1P}) = \hat{\Omega}_{1P}^{-1/2}\hat{S}_P$, where $\hat{\Omega}_{1P}$ is a consistent estimator of Ω_1 given in (2).[10] The third test (which appears under the label $t(\hat{\Omega}_P)$ in Table 3) is the one proposed by Clark and McCracken (2009). It is given by $t(\hat{\Omega}_P) = \hat{\Omega}_P^{-1/2}\hat{S}_P$, where $\hat{\Omega}_P$ is a consistent estimator of Ω_P in (2). See Section 3.3 of Clark and McCracken (2009) for more details in obtaining $\hat{\Omega}_{1P}$ and $\hat{\Omega}_P$.

[10] Here and throughout we use the original variant of the Diebold and Mariano statistic (S_1, p. 135) but replace the rectangular kernel with the Bartlett kernel. The kernel bandwidth is $\lfloor \min(R, P)^{1/3} \rfloor$ for the non-nested comparisons and $\lfloor T^{1/5} \rfloor$ for the nested comparisons.

When estimating the relevant long-run variances we use a bandwidth of $\lfloor \min\{R^{1/3}, P^{1/3}\}\rfloor$. Results are obtained with 10,000 Monte Carlo replications and 499 bootstrap replications each. We set $R = 80$ and allow P to grow from 20 to 160. The block length is equal to $l = \lfloor \min\{R^{1/3}, P^{1/3}\}\rfloor$. This choice ensures that $l \to \infty$ such that $l/\min\{\sqrt{R}, \sqrt{P}\} \to 0$, as assumed in Theorem 5.1. It also ensures that $l > 1$ when $P = 20$ and $R = 80$. The nominal level α is 5%.

The left panel of Table 3 shows that the asymptotic-based tests are typically oversized, particularly when P is small. This is especially true for the Diebold and Mariano test, which does not account for parameter estimation uncertainty. The Clark and McCracken test, which accounts for the presence of parameter estimation error, is usually more accurately sized. However, the results in Table 3 suggest that $t(\hat{\Omega}_P)$ still tends to over-reject in finite samples. For instance, for DGP2 without noise, the rejections rates of $t(\hat{\Omega}_P)$ vary between 0.077 ($P = 20$) and 0.058 ($P = 160$). These numbers are 0.073 and 0.053 for DGP2 with noise, respectively. In contrast, the bootstrap test yields rejection rates equal to 0.052 and 0.046, for DGP2, news only, when $P = 20$ and $P = 160$, respectively. These rates are 0.049 and 0.045 for DGP2, noise and news. Hence, the bootstrap corrects the size distortions of the asymptotic-based tests. This is also true for DGP1. For DGP1, Table 3 shows that the bootstrap can be slightly conservative. For DGP2, the degree of conservativeness of the bootstrap is smaller than for DGP1. Overall, it appears that the bootstrap test does a reasonable job controlling size in finite samples and, more importantly, does so without having to compute $\hat{\Omega}_P$.

Table 3: Non-nested model size and power results with 5% nominal level

Tests	P = 20	40	80	160	P = 20	40	80	160
	size: DGP1, news only				power: DGP1, news only			
$t(\hat{\Omega}_{1P})$	0.104	0.084	0.072	0.058	0.130	0.147	0.185	0.294
$t(\hat{\Omega}_P)$	0.066	0.057	0.048	0.044	0.120	0.137	0.178	0.290
Bootstrap	0.043	0.042	0.040	0.037	0.094	0.121	0.162	0.279
	size: DGP2, news only				power: DGP2, news only			
$t(\hat{\Omega}_{1P})$	0.102	0.078	0.071	0.060	0.627	0.891	0.995	1.000
$t(\hat{\Omega}_P)$	0.077	0.064	0.063	0.058	0.622	0.893	0.995	1.000
Bootstrap	0.052	0.045	0.047	0.046	0.588	0.883	0.995	1.000
	size: DGP1, noise and news				power: DGP1, noise and news			
$t(\hat{\Omega}_{1P})$	0.112	0.102	0.105	0.113	0.269	0.420	0.667	0.911
$t(\hat{\Omega}_P)$	0.050	0.045	0.041	0.041	0.233	0.376	0.625	0.885
Bootstrap	0.036	0.034	0.034	0.037	0.197	0.341	0.605	0.878
	size: DGP2, noise and news				power: DGP2, noise and news			
$t(\hat{\Omega}_{1P})$	0.096	0.076	0.067	0.052	0.497	0.768	0.967	1.000
$t(\hat{\Omega}_P)$	0.073	0.065	0.065	0.053	0.494	0.773	0.971	1.000
Bootstrap	0.049	0.046	0.047	0.045	0.457	0.755	0.967	1.000

The right panel of Table 3 shows that the three test statistics have power converging to 1 as P

increases for all DGPs, except DGP1 with news only in the revision process (we provide an explanation below). Under the alternative, the bootstrap rejection rates are typically smaller than those of the asymptotic-based tests. This is not surprising since the latter have larger rejection rates than the bootstrap test even under the null hypothesis.

For DGP1 without noise, the actual power of the test can be low when the magnitude of $\sigma^2_{e,x}$ is small. While not immediately obvious, the root of the problem lies with how revisions affect the mean loss differential. For example, in DGPs 1 and 2 the population squared forecast errors take the form $u^2_{i,t+1|1} = (y_{t+1|1} - x_{i,t}(t)\beta_{i,0})^2$, $i = 1, 2$. After taking expectations, straightforward algebra reveals

$$E(u^2_{i,t+1|1}) = \beta^2_{j,0}\sigma^2_{e,x} + \beta^2_{j,0}\sigma^2_{v,x} + \sigma^2_{e,y} + \sigma^2_{w,y} + \beta^2_{i,0}\sigma^2_{v,x} + \beta^2_{i,0}\sigma^2_{w,x},$$

for $i \neq j$ and hence

$$E(u^2_{1,t+1|1} - u^2_{2,t+1|1}) = (\beta^2_{2,0} - \beta^2_{1,0})(\sigma^2_{e,x} - \sigma^2_{w,x}).$$

When the revisions are pure news, as in DGP1, $\sigma^2_{w,x} = 0$. Because of this the value of mean loss differential can be small when $\sigma^2_{e,x}$ is small.

6.2 Nested models

For the nested case, we consider two DGPs with both noise and news. More specifically, the final data are generated according to

$$y_t = 0.4y_{t-1} + \beta_{22}x_{t-1} + e_{y,t} + v_{y,t}$$
$$x_t = 0.4x_{t-1} + e_{x,t} + v_{x,t}$$
$$Var\begin{pmatrix} e_{y,t} \\ e_{x,t} \\ v_{y,t} \\ v_{x,t} \end{pmatrix} = \begin{pmatrix} \sigma^2_{e,y} & 0.50 & 0 & 0 \\ 0.50 & 2.67 & 0 & 0 \\ 0 & 0 & 0.01 & 0 \\ 0 & 0 & 0 & 0.1 \end{pmatrix}$$

where $\sigma^2_{e,y} = 0.60$ for DGP1 and $\sigma^2_{e,y} = 1.92$ for DGP2. We set $\beta_{22} = 0$ when evaluating size and $\beta_{22} = 0.3$ when evaluating power.

The structure of the revisions is similar to what we considered in the non-nested case, i.e. we consider a single revision, given by

$$y_t\,(t) = y_t - v_{y,t} + w_{y,t} \text{ and } x_t\,(t) = x_t - v_{x,t} + w_{x,t},$$

where $w_{y,t}$ and $w_{x,t}$ are the noise components. They are generated as i.i.d. Gaussian random variables with mean zero. As we did for the non-nested case, the variances of y and x, as well as the autoregressive coefficients, are calibrated using the 2019:Q4 vintages of PCE, RGDP, and TCU and a sample ranging from 1984:Q1 to 2019:Q4. In addition, x is an amalgamation of RGDP and TCU.

We set $Cov(e_{y,t}, e_{x,t}) = 0.50$. When $\beta_{22} = 0$, this implies that $Cov(y_t, x_t) = 0.6$, in line with the data. The revisions are then calibrated based on the same quarterly vintages from 1996:Q1 through 2019:Q4. For both DGPs we set $\sigma^2_{w,x}$ to 0.3. For $\sigma^2_{w,y}$ we set it to 3.5 in DGP1 and 0.03 in DGP2. DGP1 is a less realistic case in which the revision variance of y is much larger. As we will see later, DGP1 implies a much larger value of Ω which seems to play a role in the finite-sample efficacy of the asymptotics.

At each forecast origin, we use the generated real-time data to make a one-step-ahead forecast of the target variable $y_{t+1|1}$. In each DGP the forecast takes the form $x'_{i,t}(t)\hat{\beta}_i(t)$ for $i = 1, 2$, where $x_{1,t}(t)' = y_t(t)$ and $x_{2,t}(t)' = (y_t(t), x_t(t))$. The forecast errors take the form $\hat{u}_{i,t+1|1} = y_{t+1|1} - x_{i,t}(t)'\hat{\beta}_i(t)$ for $i = 1, 2$. For all DGPs, we test equal predictive ability under quadratic loss. In contrast to the experiments in Table 3, here we use the simpler bootstrap algorithm described in Section 5.3 to obtain critical values for the test statistic $\hat{S}_P$. We set the block length to $l = \lfloor T^{1/5} \rfloor$, which satisfies the block length requirement of Theorem 1 in Corradi and Swanson (2007). We include the same two asymptotic tests, $t(\hat{\Omega}_{1P})$ and $t(\hat{\Omega}_P)$, as benchmarks. These tests are defined as previously, and their critical values are obtained from the standard normal distribution.

Table 4: Nested model size and power results with 5% nominal level

Tests	P = 20	40	80	160	P = 20	40	80	160
	size: DGP1, noise and news				power: DGP1, noise and news			
$t(\hat{\Omega}_{1P})$	0.099	0.116	0.166	0.239	0.527	0.752	0.939	0.998
$t(\hat{\Omega}_P)$	0.096	0.087	0.073	0.063	0.926	0.982	0.998	1.000
Bootstrap	0.092	0.083	0.070	0.060	0.927	0.982	0.998	1.000
	size: DGP2, noise and news				power : DGP2, noise and news			
$t(\hat{\Omega}_{1P})$	0.040	0.021	0.009	0.005	0.199	0.251	0.354	0.571
$t(\hat{\Omega}_P)$	0.107	0.099	0.089	0.076	0.641	0.727	0.837	0.937
Bootstrap	0.104	0.096	0.088	0.076	0.641	0.726	0.836	0.937

Table 4 contains results for the nested models comparisons. For DGP1, where Ω is large, the asymptotic test based on the Diebold and Mariano statistic $t(\hat{\Omega}_{1P})$ is oversized under the null hypothesis, with rejections rates that increase from 0.099 when $P = 20$ to 0.239 when $P = 160$. In sharp contrast, in DGP2 the Diebold-Mariano version is severely undersized with rejection rates that decline from 0.040 when $P = 20$ to 0.005 when $P = 160$. The bootstrap test is comparable to the Clark and McCracken test and both lead to better size control under the null hypothesis and substantial power under the alternative.

6.3 Robustness

Each of the simulations were designed to align with our assumptions. For example, we require r to be finite and, in particular, to be small relative to both R and P. In addition, we abstract from annual benchmark revisions. In the following we provide a limited collection of simulations designed to highlight the performance of the bootstrap when these assumptions are relaxed.

6.3.1 Small number of revisions

Here we consider a modified version of DGP2 applied to tests of equal accuracy for non-nested models. The environment is largely the same as that in Section 6.1 except that we now allow values of r equaling $2, 4, 8, 16$ and 32. For example, the final, first, and intermediate releases $j \in \{1, \ldots, r-1\}$ of the dependent variable take the form

$$\begin{aligned} y_{t|r} &= 0.3x_{1,t-1} + (0.3+\Delta)x_{2,t-1} + e_{y,t} + \sum_{i=1}^{r-1} v_{y,t|i}, \\ y_{t|1} &= 0.3x_{1,t-1} + (0.3+\Delta)x_{2,t-1} + e_{y,t} + \sum_{i=1}^{r-1} w_{y,t|i}, \\ y_{t|(j+1)} &= y_{t|j} - w_{y,t|j} + v_{y,t|j}. \end{aligned}$$

Similarly, the final, first, and intermediate releases of each regressor $x_{i,t}$, $i = 1, 2$ take the form

$$\begin{aligned} x_{i,t|r} &= e_{x_i,t} + \sum_{l=1}^{r-1} v_{x_i,t|l}, \\ x_{i,t|1} &= e_{x_i,t} + \sum_{l=1}^{r-1} w_{x_i,t|l}, \\ x_{i,t|(j+1)} &= x_{i,t|j} - w_{x_i,t|j} + v_{x_i,t|j}. \end{aligned}$$

While the parameterization is similar to DGP2, we rescale the variance of the news and noise components so that their contribution to the variances of y and x_i are invariant to the choice of r.

- DGP3: $\sigma^2_{e,y} = 1.69$, $\sigma^2_{e,x} = 3.2$, $\sigma^2_{v,y} = 0.01/(r-1)$, $\sigma^2_{v,x} = 0.1/(r-1)$, $\sigma^2_{w,y} = 0.03/(r-1)$, $\sigma^2_{w,x} = 0.3/(r-1)$.

Table 5: Non-nested model size and power results with 5% nominal level: multiple revisions

r =2	4	8	16	32	r =2	4	8	16	32
size: DGP3, news only					power: DGP3, news only				
0.043	0.049	0.046	0.052	0.046	1.000	1.000	1.000	1.000	1.000
size: DGP3, noise and news					power: DGP3, noise and news				
0.048	0.046	0.047	0.045	0.044	1.000	1.000	1.000	1.000	1.000

In Table 5 we report actual size and power of our bootstrap test for a range of values of r when the sample sizes are $R = 80$ and $P = 160$. As we saw in Table 3, regardless of the choice of r, actual size and power of the test tends to be reasonable when using the percentile bootstrap.

6.3.2 Annual revisions

Here we consider a stylized example of how our bootstrap can, but need not, be robust to the presence of annual revisions. Suppose that in each quarterly vintage t there is an initial release $y_{t|1}$. However, revisions occur only once a year and when they do, all previous releases are final. Assume that all realizations y_s, $s = 1, \ldots, R-1$ are final.

We consider two autoregressive models for forecasting y_{t+1}, one based on an AR(1) model, which uses one lagged value as a predictor, and another based on a restricted version of an AR(2) model (where the twice lagged value is the predictor). Specifically, the DGP can be described as

$$y_t = x_{t-1}\beta + e_t + v_t \quad \text{and} \quad y_{t|1} = y_t - v_t + w_t,$$

where $e_t \sim \text{i.i.d.}N(0, \sigma_e^2)$, $v_t \sim \text{i.i.d.}N(0, \sigma_v^2)$, and $w_t \sim \text{i.i.d.}N(\mu_w, \sigma_w^2)$. We set $x_{t-1} = y_{t-1}$ for the AR(1) model and $x_{t-1} = y_{t-2}$ for the AR(2) model. Under the null hypothesis, $\beta = 0$, which implies $y_t \sim \text{i.i.d.}N(0, \sigma^2)$, with $\sigma^2 = \sigma_e^2 + \sigma_v^2$. We set $\beta = 0.5$ under the alternative hypothesis.

We consider a test of zero-mean prediction error based on an OLS-estimated autoregressive model that does not contain an intercept. The one-step-ahead forecast is evaluated using the fully revised value $y_{t+1|r'} = y_{t+1}$. The moment of interest takes the form $E(f_{t+1}) = E(y_{t+1} - x_t(t)\beta) = 0$ where the specific form of $x_t(t)$ depends on the lag length of the model.

For the AR(1) model, $x_t(t) = y_{t|1}$ for all t. Given our DGP, $\beta = 0$, which implies $f_{t+1} = y_{t+1}$ and the null hypothesis holds. Since the revision process is finite lived and the functional form for $f_{t+1}(\beta)$ is time invariant, the asymptotics in Clark and McCracken still apply. Intuitively, our bootstrap algorithm will also apply because it will enforce the feature that for every vintage, the most recent value will be an initial release while all previous values will be final. Put differently, while our algorithm does not replicate the entire pattern of the data when annual revisions are present, it replicates what is needed for this example.

Now suppose the model has a twice-lagged value (as in the restricted AR(2) model described above) and hence $x_t(t) = y_{t-1}(t)$. We immediately find that the functional form for f_{t+1} changes across the calendar year. In most periods it takes the form $f_{t+1}^{(1)}(\beta) = y_{t+1} - \beta y_{t-1|1}$, but during the annual revision it takes the form $f_{t+1}^{(2)}(\beta) = y_{t+1} - \beta y_{t-1}$. In both cases, under our assumed DGP, $f_{t+1} = y_{t+1}$ and the null hypothesis holds for all t. The problem is that $Ef_{t+1,\beta}^{(j)}$ need not be constant for all t. If we let $F^{(j)}$ denote $Ef_{t+1,\beta}^{(j)}$ then $F^{(1)} = -Ey_{t-1|1}$ and $F^{(2)} = -Ey_{t-1} = 0$, which are distinct

Table 6: Size and power results: AR(1) and AR(2) models

λ =1	4	12	λ =1	4	12
size: AR(1)			power: AR(1)		
0.056	0.051	0.072	0.959	0.964	0.977
size: AR(2)			power: AR(2)		
0.056	0.111	0.163	0.061	0.944	0.988

so long as $F^{(1)}$ is non-zero. Since $F^{(j)}$ varies, the asymptotics in Clark and McCracken and our bootstrap algorithm do not apply, because neither is designed to distinguish between regular vintages and vintages that contain annual revisions.

In Table 6 we provide simulation evidence on the actual size and power of the bootstrap-based test of zero-mean prediction error in the presence of an annual revision. For both models, we let

$$\sigma_e^2 = 0.3, \quad \sigma_v^2 = 0.2, \quad \sigma_w^2 = 0.2, \quad \mu_w = 0.85.$$

The initial estimation sample size is $R = 80$, while the out-of-sample size is $P = 80$. We set the bootstrap's block length to 1 since our examples have the m.d.s. property. We consider annual revision frequencies of $\lambda = 1,\ 4,\ 12$. When $\lambda = 1$, the vintages have a single regular revision structure. When $\lambda > 1$, revisions only arise with frequency λ. For instance, if data have a quarterly frequency, $\lambda = 4$ implies that each year we have one annual revision. If data are observed at the monthly frequency, one annual revision corresponds to $\lambda = 12$.

When the model is an AR(1), our percentile bootstrap provides adequately sized tests of the null regardless of annual revisions. In contrast, when the model is an AR(2), actual size of the test rises sharply, well above the nominal 5% level. As to power, in most cases the test rejects with frequencies near 95%. Even so, when the model is an AR(2) and there are no annual revisions, setting β to 0.5 does not constitute a deviation from the null and the rejection frequency aligns with the nominal size of the test.

7 Forecasting inflation

In this section we apply our bootstrap procedure to tests of equal forecast accuracy in the context of forecasting both quarterly CPI and PCE inflation. In particular, we do so with an eye toward revisiting Ang et al. (2007) who compare several forecasting methods and conclude that survey-based forecasts of inflation are generally superior under quadratic loss. While their results broadly support that thesis, they consider only current vintage data and hence do not address a more realistic environment in which data are subject to revision.

With that in mind, we compare the forecast accuracy of a small handful of linear forecasting models to survey forecasts and do so using vintage data. The surveys consist of the Blue Chip (BC) and the Survey of Professional Forecasters (SPF). For the models we keep it simple and consider only two from Ang et al. (2007): an AR(2) and an AR(2) augmented with one lag of real GDP (RGDP) growth - the latter of which includes their preferred measure of economic slack. We also consider an AR(2) that includes one lag of the change in total capacity utilization (TCU), which is the preferred indicator of economic slack used by Stock and Watson (1999) who also conduct forecasting exercises using current, rather than real-time, vintages of data.

Note that while revisions to CPI inflation are typically small, revisions to PCE inflation as well as RGDP and TCU can be substantial. Therefore, it is not immediately obvious that our results will align with those of Ang et al. (2007). Unfortunately, these three series also exhibit annual benchmark revisions and hence our assumption of a regular revision pattern across all vintages is violated. Even so, we take some comfort from the simulations in section 6.3.2 which suggest that our bootstrap can yield reasonable results for models with short lags like the ones we consider.

Vintages of the CPI and PCE price indices, RGDP growth, and TCU are obtained from the RTDSM hosted by the Federal Reserve Bank of Philadelphia. In each instance, the vintages are available monthly. We use the January, April, July, and October vintages exclusively as these are the first months for which (previous) quarter values can be constructed for all of the series. The SPF is also obtained from the Federal Reserve Bank of Philadelphia and is released within the first two weeks of February, May, August, and November. The BC forecasts are obtained from the Haver database. While the BC forecasts are updated monthly, we use those vintages that align with the SPF. Together this implies a modest timing advantage for the surveys as they are released later than the implied forecast origins for the estimated models.[11]

It's worth noting that Ang et al. (2007) only use survey forecasts of CPI inflation even when forecasting PCE inflation. The reason is that the surveys only recently began asking survey participants about PCE inflation (i.e., 2007 for SPF and 2021 for BC). To benchmark our results to theirs, we follow their approach over two samples: one with forecasted dates of 1997:Q1 to 2002:Q4 that aligns with their post-1995 sample, and a second with forecasted dates of 1997:Q1 to 2019:Q4 that provides more recent evidence.[12] We also consider a third sample with forecasted dates of 2008:Q1 to 2019:Q4

[11]There are a few instances in which this timing is irregular. If there is more than one vintage within a month we use the latest release. If there is no release within a month, but there is one early in the subsequent month, we use it.

[12]Note that, for example, a Q4 forecast origin is based on the January (February) vintage of data when using the model (survey) to form a forecast.

that allows us to evaluate SPF forecasts of PCE inflation relative to SPF forecasts of CPI inflation.[13]

Within each vintage t we apply the following data transformations. Annualized quarterly inflation ($y_t^{(1)} = y_t$) is constructed as four times the log difference of the price index associated with the last calendar month of quarters t and $t-1$. Annual inflation is constructed in the obvious way as $y_t^{(4)} = \sum_{j=0}^{3} y_{t-j}/4$. RGDP growth is annualized quarterly percent change. TCU is transformed to be four-times the average of the monthly first differences across the quarter.

For each forecast horizon $\tau = 1, 4$, the three OLS estimated forecasting models take the form

$$y_t^{(\tau)}(t) = \beta_0 + \beta_1 y_{t-\tau}(t) + \beta_2 y_{t-\tau-1}(t) + \beta_3 x_{t-\tau}(t) + u_t^{(\tau)}(t)$$

where x is either omitted or denotes $RGDP$ or TCU. We restrict attention to a Great Moderation sample and hence all OLS regressions are estimated on a sample ranging from 1984:Q3 through a given forecast origin. Throughout, all forecasts are evaluated against the initial release $y_{t+\tau|1}^{(\tau)}$.

Tables 7 and 8 provide the results of our forecasting exercises when using survey forecasts of CPI inflation. In each, the first two columns denote the ten pairwise model (survey) comparisons while the remaining columns distinguish the target variable and horizon. For each permutation of model comparison, target variable, and horizon we report three numbers. The first denotes the ratio of root mean squared errors (RMSE) such that a value less than one favors model (survey) 1. The second number denotes the percentile bootstrapped p-value (in parentheses) associated with the test of equal forecast accuracy under quadratic loss.[14] For the same test, the third number is the p-value (in brackets) implied by the asymptotic distribution of the test statistic delineated in Clark and McCracken (2009).

First, consider the results for CPI inflation in Tables 7 and 8. In both cases there is little to suggest that any of the models are more accurate than the others. Each of the RMSE ratios are close to one and none of the tests of equal ability are statistically significant using either the asymptotic or bootstrapped p-values. The largest RMSE deviations from one occur when comparing the surveys to the models. Here we find benefits to using surveys with gains as large as 7 percent at the one quarter horizon and up to roughly 15 percent at the four quarter horizon. Even so, the tests of equal predictive ability do not provide any evidence that these improvements are statistically significant using either set of p-values. Finally, there is very little that distinguishes the two sets of survey forecasts from one another. Overall, the results for CPI inflation support the conclusions in Ang et al. (2007) even if the benefits to using surveys are not statistically significant.

[13]We end our sample prior to the onset of COVID to avoid the large atypical realizations of RGDP and TCU as well as their revisions.

[14]For ease of comparison, we use the bootstrap algorithm for linear models regardless of whether the models are nested or non-nested. When a survey is used it is paired with the dependent variable to preserve dependence between the two. In all comparisons, the algorithm shares a common seed across all 999 bootstrap replications.

Table 7: Application to Forecasting Inflation (Ang et al. sample)

Model 1	Model 2	CPI		PCE	
		$\tau = 1$	$\tau = 4$	$\tau = 1$	$\tau = 4$
AR(2)	AR(2) + RGDP	0.988	0.987	1.012	0.989
		(0.875)	(0.835)	(0.333)	(0.911)
		[0.387]	[0.290]	[0.426]	[0.199]
AR(2)	AR(2) + TCU	0.997	1.021	1.021	1.015
		(0.663)	(0.483)	(0.415)	(0.329)
		[0.938]	[0.520]	[0.714]	[0.636]
AR(2) + RGDP	AR(2) + TCU	1.009	1.034	1.008	1.027
		(0.701)	(0.435)	(0.659)	(0.269)
		[0.769]	[0.232]	[0.883]	[0.357]
BC	AR(2)	0.931	0.872	1.261	1.172
		(0.593)	(0.289)	(0.161)	(0.525)
		[0.417]	[0.429]	[0.079]	[0.123]
BC	AR(2) + RGDP	0.919	0.861	1.277	1.158
		(0.551)	(0.279)	(0.121)	(0.519)
		[0.316]	[0.386]	[0.075]	[0.132]
BC	AR(2) + TCU	0.928	0.890	1.288	1.190
		(0.647)	(0.347)	(0.085)	(0.327)
		[0.319]	[0.455]	[0.065]	[0.073]
SPF	AR(2)	0.932	0.891	1.286	1.221
		(0.563)	(0.321)	(0.113)	(0.357)
		[0.323]	[0.443]	[0.039]	[0.034]
SPF	AR(2) + RGDP	0.921	0.879	1.302	1.207
		(0.543)	(0.313)	(0.079)	(0.365)
		[0.220]	[0.394]	[0.037]	[0.035]
SPF	AR(2) + TCU	0.930	0.909	1.313	1.240
		(0.691)	(0.375)	(0.061)	(0.209)
		[0.246]	[0.478]	[0.047]	[0.021]
BC	SPF	0.998	0.979	0.981	0.960
		(0.871)	(0.477)	(0.465)	(0.065)
		[0.944]	[0.495]	[0.521]	[0.056]

Notes: For each pairwise comparison, the table presents: the ratio of root mean squared errors, the p-value associated with a test of equal forecast accuracy based on our percentile bootstrap (in parentheses), and the p-value for the same test based on the asymptotic distribution associated with the test statistic delineated in Clark and McCracken (2009) (in square brackets). RMSE ratios less (greater) than one favor model 1 (2). Results are provided for an initial window size $R = 50$ and horizons $\tau = 1$ and $\tau = 4$, across 999 bootstrap replications. The forecasted dates range from 1997:Q1 to 2002:Q4.

Table 8: Application to Forecasting Inflation (pre-COVID sample)

Model 1	Model 2	CPI		PCE	
		$\tau = 1$	$\tau = 4$	$\tau = 1$	$\tau = 4$
AR(2)	AR(2) + RGDP	0.991	1.000	0.996	0.992
		(0.977)	(0.557)	(0.397)	(0.543)
		[0.294]	[0.986]	[0.694]	[0.645]
AR(2)	AR(2) + TCU	0.982	0.989	0.969	0.971
		(0.481)	(0.557)	(0.745)	(0.719)
		[0.561]	[0.733]	[0.345]	[0.410]
AR(2) + RGDP	AR(2) + TCU	0.990	0.989	0.974	0.978
		(0.481)	(0.977)	(0.981)	(0.969)
		[0.704]	[0.702]	[0.316]	[0.426]
BC	AR(2)	0.938	0.825	0.957	0.909
		(0.629)	(0.321)	(0.975)	(0.855)
		[0.142]	[0.163]	[0.519]	[0.460]
BC	AR(2) + RGDP	0.930	0.825	0.952	0.902
		(0.663)	(0.453)	(0.893)	(0.957)
		[0.130]	[0.180]	[0.503]	[0.434]
BC	AR(2) + TCU	0.921	0.816	0.927	0.883
		(0.901)	(0.417)	(0.931)	(0.947)
		[0.111]	[0.110]	[0.244]	[0.288]
SPF	AR(2)	0.944	0.834	0.970	0.926
		(0.727)	(0.335)	(0.811)	(0.985)
		[0.167]	[0.162]	[0.653]	[0.523]
SPF	AR(2) + RGDP	0.936	0.833	0.966	0.919
		(0.745)	(0.489)	(0.675)	(0.887)
		[0.151]	[0.181]	[0.627]	[0.495]
SPF	AR(2) + TCU	0.927	0.824	0.941	0.899
		(0.963)	(0.461)	(0.751)	(0.907)
		[0.131]	[0.113]	[0.329]	[0.344]
BC	SPF	0.994	0.990	0.986	0.982
		(0.161)	(0.455)	(0.171)	(0.225)
		[0.088]	[0.467]	[0.127]	[0.246]

Notes: For each pairwise comparison, the table presents: the ratio of root mean squared errors, the p-value associated with a test of equal forecast accuracy based on our percentile bootstrap (in parentheses), and the p-value for the same test based on the asymptotic distribution associated with the test statistic delineated in Clark and McCracken (2009) (in square brackets). RMSE ratios less (greater) than one favor model 1 (2). Results are provided for an initial window size $R = 50$ and horizons $\tau = 1$ and $\tau = 4$, across 999 bootstrap replications. The forecasted dates range from 1997:Q1 to 2019:Q4.

Table 9: Forecasting PCE Inflation with SPF

Model 1	Model 2	PCE	
		$\tau = 1$	$\tau = 4$
SPF-CPI	AR(2)	0.979	1.026
		(0.717)	(0.727)
		[0.761]	[0.697]
SPF-CPI	AR(2) + RGDP	0.988	1.013
		(0.923)	(0.837)
		[0.845]	[0.842]
SPF-CPI	AR(2) + TCU	0.943	0.980
		(0.583)	(0.909)
		[0.274]	[0.796]
SPF-PCE	AR(2)	0.952	0.951
		(0.363)	(0.523)
		[0.525]	[0.586]
SPF-PCE	AR(2) + RGDP	0.961	0.939
		(0.447)	(0.411)
		[0.561]	[0.493]
SPF-PCE	AR(2) + TCU	0.917	0.908
		(0.329)	(0.425)
		[0.140]	[0.289]
SPF-CPI	SPF-PCE	1.028	1.079
		(0.125)	(0.063)
		[0.047]	[0.053]

Notes: For each pairwise comparison, the table presents: the ratio of root mean squared errors, the p-value associated with a test of equal forecast accuracy based on our percentile bootstrap (in parentheses), and the p-value for the same test based on the asymptotic distribution associated with the test statistic delineated in Clark and McCracken (2009) (in square brackets). RMSE ratios less (greater) than one favor model 1 (2). Results are provided for an initial window size $R = 94$ and horizons $\tau = 1$ and $\tau = 4$, across 999 bootstrap replications. The forecasted dates range from 2008:Q1 to 2019:Q4.

Now consider the results for PCE inflation in Tables 7 and 8. These results are much more complex – though the explanation may be simple. Recall that Ang et al. (2007) use SPF forecasts of CPI inflation when forecasting either CPI or PCE inflation. In their Tables 7 and 8 (pages 1190, 1192), the SPF forecasts are nominally worse than a benchmark model when forecasting PCE inflation but nominally better when forecasting CPI inflation. In our Table 7, which uses a similar sample to theirs, this is exactly what we find. When using a comparable forecast sample, we find that the BC and SPF surveys also do worse than the models when forecasting PCE inflation. Moreover, there are several instances in which either the asymptotic or bootstrapped p-values indicate a rejection of the null of equal predictive ability at conventional levels.

The results in our Table 7, for PCE inflation, make sense. In the words of Ang et al. (2007), "Naturally, surveys [of CPI inflation] do a relatively poor job at forecasting PCE inflation, which they are not designed to forecast." Which leads us to our Table 8 and the results for PCE inflation. In all instances, the surveys are nominally more accurate than the models. The gains are as large as 5 percent at the one quarter horizon and up to 10 percent at the four quarter horizon. Nevertheless, in all instances the asymptotic and bootstrapped p-values suggest failing to reject the null of equal predictive ability.

The results for PCE inflation in Table 8 are counter intuitive and an explanation is not immediately obvious. A leading possibility is simply that we are using a different sample and that something is fundamentally different between the two. Interestingly, one distinction is easy to identify. Over Ang et al.'s (2007) 1995-2002 sample, quarterly CPI and PCE inflation exhibit a correlation of 81 percent. But over the subsequent 2002-2019 sample they exhibit a much higher correlation of 98 percent! This suggests that for much of the sample used for Table 8, forecasts of CPI inflation may be good for PCE inflation as well!

Regardless, neither Table 7 or 8 informs us about the relative accuracy of models versus surveys of PCE inflation when forecasting PCE inflation. Albeit with a shorter sample, Table 9 provides this comparison. For brevity we omit the model-to-model comparisons and focus on the comparison between models and surveys. In addition, we provide evidence on whether the SPF forecasts of PCE inflation have marginal value relative to the SPF forecasts of CPI inflation that seemed to perform pretty well in Table 8.

In Table 9 we find that at the one quarter horizon, both survey forecasts nominally outperform the models as all RMSE ratios are less than one. This is not surprising given the high correlation between CPI and PCE inflation over this sample. At the four quarter horizon, all but two of the pairwise comparisons (both of which use SPF forecasts of CPI inflation) indicate that the survey forecasts are nominally more accurate. Nevertheless, as we saw in Table 8, the differences in accuracy are not statistically significant using either choice of p-value. Perhaps most interestingly, there is reasonably clear evidence that despite the high correlation between the two measures of inflation, choosing the correct survey matters for accuracy. At both horizons, the SPF forecast of PCE inflation is not only nominally more accurate, there is evidence that it is statistically significant and especially so at the four quarter horizon.

Overall, our results reinforce those in Ang et al. (2007) but with a slightly different interpretation. Across our three samples it is clear that survey forecasts of inflation tend to be more accurate than simple models when the survey is designed to forecast the relevant measure of inflation. Even so, our results suggest a somewhat obvious modification to the Ang et al. (2007) quote, "Naturally, surveys [of CPI inflation] do a relatively poor job at forecasting PCE inflation, which they are not designed to forecast, unless CPI and PCE inflation happen to be very highly correlated."

8 Conclusions

The main contribution of this paper is to propose a new bootstrap algorithm for out-of-sample predictability tests when the data are subject to a finite number of data revisions. Our bootstrap algorithm replicates the triangular structure of the different vintages by relying on an application of the moving blocks bootstrap. The novel feature of our method is that it not only preserves the time series dependence of the data within each vintage, but it also preserves the dependence across the different vintages. We provide a set of regularity conditions under which our bootstrap method is asymptotically valid. Simulations show that the proposed bootstrap tests have comparable size properties to the data revision-robust test statistic proposed by Clark and McCracken (2009). However, the bootstrap is easier to apply as it avoids estimating directly the asymptotic variance of the test. We conclude with an application to inflation forecasting in the presence of real-time vintage data. Our results support the conclusion in Ang et al. (2007) that survey forecasts of inflation are more accurate than simple models and particularly so if the surveys are aligned with the relevant measure of inflation.

A Appendix

As usual in the bootstrap literature, we use P^* to denote the bootstrap probability measure, conditional on the original sample (defined on a given probability space $(\Omega, \mathcal{F}, P)$). For any bootstrap statistic t_T^*, we write $t_T^* = o_p^*(1)$, or $t_T^* \to^{P^*} 0$, when for any $\delta > 0$, $P^*(|t_T^*| > \delta) = o_p(1)$. We write $t_T^* = O_p^*(1)$, when for all $\delta > 0$ there exists $M_\delta < \infty$ such that $\lim_{T\to\infty} P[P^*(|t_T^*| > M_\delta) > \delta] = 0$. By Markov's inequality, this follows if $E^*|t_T^*|^q = O_p(1)$ for some $q > 0$. Finally, we write $t_T^* \to^{d^*} D$, in probability, if conditional on a sample with probability that converges to one, t_T^* weakly converges to the distribution D under P^*, i.e. $E^*(f(t_T^*)) \to^P E(f(D))$ for all bounded and uniformly continuous functions f.

For simplicity, we treat β as a scalar and focus on the case of a single model, i.e., we let $k = 1$ in Assumption 2. Following West (1996), we write "$\sup_t$" to mean "$\sup_{R\leq t\leq T}$."

A.1 Auxiliary lemmas

Here, we provide several auxiliary lemmas, followed by their proofs.

Lemma A.1 *Under Assumptions 1-5,*

(a) $\sup_t \left|\hat{B}(t) - B(t)\right| = o_p(1)$.

(b) $P^{-1/2}\sum_{t=R}^{T}\left|\hat{H}(t) - H(t)\right| = o_p(1)$.

(c) *For any* $0 \leq a < 1/2$, $\sup_t P^a|\hat{H}(t) - H(t)| = o_p(1)$.

(d) *For any* $0 \leq a < 1/2$, $\sup_t |P^a(\hat{\beta}(t) - \beta_0)| = o_p(1)$.

Lemma A.1(d) is the analog of Lemma A.3(b) of West (1996) when the estimator of β_0 is $\hat{\beta}(t)$, the OLS estimator based on real-time vintage data.

Lemma A.2 *Under Assumptions 1-5 and assuming that* $l \to \infty$ *such that* $l/\min\{\sqrt{R}, \sqrt{P}\} \to 0$,

(a) For any $0 \leq a < 1/2$, $\sup_t P^a|H^*(t) - E^*H^*(t)| = o_p^*(1)$, where $H^*(t) \equiv t^{-1}\sum_{s=1+\tau}^{t} h_s^*$, with $h_s^* \equiv x_{s-\tau}^*(y_s^* - x_{s-\tau}^{*\prime}\beta_0)$.

(b) $\sup_t |B^*(t) - B| = o_p^*(1)$, where $B^*(t) \equiv (t^{-1}\sum_{s=1+\tau}^{t} x_{s-\tau}^* x_{s-\tau}^{*\prime})^{-1}$.

(c) For any $0 \leq a < 1/2$, $\sup_t \left|P^a\left(\hat{\beta}_t^* - \beta_0\right)\right| = o_p^*(1)$.

(d) $P^{-1/2}\sum_{t=R}^{T}(f_{t+\tau|r',\beta}^* - F)BH^*(t) = o_p^*(1)$.

(e) $P^{-1/2}\sum_{t=R}^{T} F(B^*(t) - B)H^*(t) = o_p^*(1)$.

(f) $P^{-1/2}\sum_{t=R}^{T}(f^*_{t+\tau|r',\beta} - F)(B^*(t) - B)H^*(t) = o^*_p(1)$.

Parts (a) and (c) are the bootstrap analogs of Lemma A.3 (a) and (b) of West (1996). Parts (d) through (f) are the bootstrap analogs of Lemma A.4 of West (1996).

To prove Lemma A.2, we rely on the following result. It provides an asymptotic approximation to the MBB expectation of a given observation in the MBB sample.

Lemma A.3 *Let* $\{Z^*_t : t = M, \ldots, M'\}$ *denote a MBB resample of* $\{Z_t : t = N, \ldots, N'\}$ *with block size* l *such that* $l \to \infty$ *with* $l/n \to 0$*, where* $n = N' - N + 1$*. If* $E|Z_t| \le \Delta$*,* $t = N, \ldots, N'$*, for some* $\Delta < \infty$*, then* $E^*(Z^*_t) = \bar{Z}_n + O_p(l/n)$*, uniformly in* $t = M, \ldots, M'$*, where* $\bar{Z}_n \equiv n^{-1}\sum_{t=N}^{N'} Z_t$*.*

Lemma A.4 *Under Assumptions 1-5 and* $l \to \infty$ *such that* $l/\min\{\sqrt{R}, \sqrt{P}\} \to 0$*,*

(a) $Var^*(S^*_{1P}) \xrightarrow{p} \Omega_1$.

(b) $Var^*(S^*_{2P}) \xrightarrow{p} \Omega_2$.

(c) $Cov^*(S^*_{1P}, S^*_{2P}) \xrightarrow{p} \Omega_{12}$.

Proof of Lemma A.1. To prove (a), it suffices to show that $\sup_t \left|\hat{B}^{-1}(t) - B^{-1}(t)\right| = o_p(1)$. We can write

$$\hat{B}^{-1}(t) - B^{-1}(t) = t^{-1}\sum_{s=1+\tau}^{t}\left(x_{s-\tau}(t)x_{s-\tau}(t)' - x_{s-\tau}x'_{s-\tau}\right) \equiv t^{-1}V_t,$$

where

$$V_t \equiv \sum_{s=t-r+1}^{t}(x_{s-\tau}(t)x_{s-\tau}(t)' - x_{s-\tau}x'_{s-\tau}).$$

Hence, for any $\varepsilon > 0$,

$$P\left(\sup_t |t^{-1}V_t| > \varepsilon\right) \le \sum_{t=R}^{T} P\left(|t^{-1}V_t| > \varepsilon\right) \le \varepsilon^{-2}\sum_{t=R}^{T} t^{-2}E|V_t|^2.$$

The result follows by noting that $\sum_{t=R}^{T} t^{-2} \le PR^{-2} \to 0$ and $E|V_t|^2 = O(1)$ by Assumption 4. Part (b) follows similarly by writing $t^{-1}V_t \equiv \hat{H}(t) - H(t) = t^{-1}\sum_{s=t-r+1}^{t}(h_s(t) - h_s)$. Part (c) follows similarly. For part (d),

$$\sup_t |P^a(\hat{\beta}(t) - \beta_0)| \le \sup_t |P^a(\hat{\beta}(t) - \hat{\beta}_t)| + \sup_t |P^a(\hat{\beta}_t - \beta_0)|,$$

where $\sup_t |P^a(\hat{\beta}(t) - \hat{\beta}_t)| = o_p(1)$ given parts (a) and (c), and $\sup_t |P^a(\hat{\beta}_t - \beta_0)| = o_p(1)$ by West's (1996) Lemma A.3(b). ∎

Proof of Lemma A.2. Part (a). First, write

$$H^*(t) = \frac{1}{t}\sum_{s=1+\tau}^{R} h_s^* + \mathbf{1}_{\{t\geq R+1\}}\frac{1}{t}\sum_{s=R+1}^{t} h_s^*$$

where $\mathbf{1}_{\{t\geq R+1\}}$ is an indicator function equal to 1 if $t \geq R+1$. This decomposition is useful because $h_s^* = h_{\gamma_s}$ for $s \leq R$, whereas $h_s^* = h_{\eta_s}$ for $s > R$. Using it, for any $0 \leq a < 1/2$, we write

$$P^a \sup_t |H^*(t) - E^*H^*(t)| \leq \mathcal{A}_1^* + \mathcal{A}_2^*,$$

where

$$\mathcal{A}_1^* = P^a\frac{1}{\sqrt{R}}\left|\frac{1}{\sqrt{R}}\sum_{s=1+\tau}^{R}(h_s^* - E^*h_s^*)\right|, \; \mathcal{A}_2^* = P^a \sup_{R+1\leq t\leq T}\left|\frac{1}{t}\sum_{s=R+1}^{t}(h_s^* - E^*h_s^*)\right|,$$

where $\mathcal{A}_1^* = o_p^*(1)$ since $\left|R^{-1/2}\sum_{s=1+\tau}^{R}(h_s^* - E^*h_s^*)\right| = O_p^*(1)$. This follows by Chebyshev's inequality, since under Assumption 3 we can show that $Var^*\left(R^{-1/2}\sum_{s=1+\tau}^{R} h_s^*\right) = O_p(1)$ by Corollary 3.1 of Fitzenberger (1998). We are left to show that $\mathcal{A}_2^* = o_p^*(1)$. For simplicity, we assume that the number of blocks of size l needed to obtain the $t-R$ observations indexed by $s = R+1, \ldots, t$ is k (where $R+1 \leq t \leq T$), i.e. $t - R = kl$. Note that k is such that $1 \leq k \leq k_2$, since we have defined k_2 as the number of blocks of size l needed to obtain the last $T + \tau - (R+1) + 1 = P + \tau - 1$ bootstrap observations in the sample. With this notation, we can write

$$\mathcal{A}_2^* = P^a \sup_{1\leq k\leq k_2}\left|\frac{1}{kl+R}\sum_{s=R+1}^{R+kl}(h_s^* - E^*h_s^*)\right| \leq P^a R^{-1}\sup_{1\leq k\leq k_2}\left|\sum_{i=1}^{k}\mathcal{U}_i^*\right|,$$

where

$$\mathcal{U}_i^* \equiv \sum_{t=R+1+(i-1)l}^{R+(i-1)l+l}(h_t^* - E^*h_t^*) = \sum_{j=1}^{l}(h_{J_i+j-1} - E^*(h_{J_i+j-1})).$$

The last equality uses the fact that for $t = R+1+(i-1)\,l, \ldots, R+(i-1)l+l$,

$$h_t^* = h_{\eta_{R+1+(i-1)l+(j-1)}} = h_{J_i+j-1},$$

where $J_i \sim$ i.i.d. Uniform on $\{R+\tau, \ldots, T+\tau-l+1\}$. To prove that $\mathcal{A}_2^* = o_p^*(1)$, it suffices to show that $\sup_{1\leq k\leq k_2}|P^{-1/2}\sum_{i=1}^{k}\mathcal{U}_i^*| = O_p^*(1)$. To prove this, note that by the independence of $\{J_i\}$, $\{\mathcal{U}_i^* : i = 1, \ldots, k\}$ is an array of independent variables, implying that it is a martingale difference array with respect to the σ-field $\mathcal{G}_{i-1}^* = \sigma(J_1, \ldots, J_{i-1})$. Using Theorem 15.14 of Davidson (1994), for any $\epsilon > 0$,

$$P^*\left(\sup_{1\leq k\leq k_2}\left|P^{-1/2}\sum_{i=1}^{k}\mathcal{U}_i^*\right| > \epsilon\right) \leq \frac{E^*\left|P^{-1/2}\sum_{i=1}^{k_2}\mathcal{U}_i^*\right|^2}{\epsilon^2},$$

where

$$E^*\left(P^{-1/2}\sum_{i=1}^{k_2}\mathcal{U}_i^*\right)^2 = E^*\left(P^{-1/2}\sum_{s=R+1}^{T+\tau}(h_s^*-E^*h_s^*)\right)^2 = Var^*\left(P^{-1/2}\sum_{t=R+1}^{T+\tau}h_s^*\right),$$

which is $O_p(1)$ by Corollary 3.1 of Fitzenberger (1998). The result follows by noting that $P^aR^{-1}P^{1/2} = o(1)$ under Assumption 5 and $a < 1/2$.

Part (b). It suffices to show that $\sup_t |B^*(t)^{-1} - B^{-1}| = o_p^*(1)$, which follows if $\sup_t |B^*(t)^{-1} - E^*B^*(t)^{-1}| = o_p^*(1)$ and $\sup_t |E^*B^*(t)^{-1} - B^{-1}| = o_p^*(1)$. Let $a_s^* = x_{s-\tau}^*x_{s-\tau}^{*\prime}$, and $a_s = x_{s-\tau}x_{s-\tau}'$. By the triangle inequality,

$$\sup_t\left|B^*(t)^{-1} - E^*B^*(t)^{-1}\right| \leq \left|R^{-1}\sum_{s=1+\tau}^{R}(a_s^*-E^*a_s^*)\right| + \sup_{R+1\leq t\leq T}\left|t^{-1}\sum_{s=R+1}^{t}(a_s^*-E^*a_s^*)\right|.$$

We can show that the two terms on the right-hand-side (RHS) of the inequality are $o_p^*(1)$ by relying on an argument similar to that used in the proof of part (a). Thus, we only need to show that $\sup_t |E^*B^*(t)^{-1} - B^{-1}| = o_p(1)$. Noting that $E^*B^*(t)^{-1} = t^{-1}\sum_{s=1+\tau}^{t}E^*a_s^*$ and $B^{-1} = Ea_s$ (which is constant under our stationarity assumption on x_t), we can write

$$\begin{aligned}\sup_t\left|E^*B^*(t)^{-1} - B^{-1}\right| &= \sup_t\left|t^{-1}\sum_{s=1+\tau}^{t}E^*(a_s^*-Ea_s) + \frac{t-\tau}{t}E(a_s) - E(a_s)\right| \\ &\leq \sup_t\left|t^{-1}\sum_{s=1+\tau}^{t}E^*(a_s^*-Ea_s)\right| + O\left(R^{-1}\right),\end{aligned}$$

where the last inequality uses the fact that τ is finite. Noting that $a_s^* = a_{\gamma_s}$ for $s = 1+\tau,\ldots,R$ and $a_s^* = a_{\eta_s}$ for $s \geq R+1$, we can apply the triangular inequality to obtain

$$\sup_t\left|t^{-1}\sum_{s=1+\tau}^{t}E^*(a_s^*-Ea_s)\right| \leq \left|\underbrace{R^{-1}\sum_{s=1+\tau}^{R}E^*(a_s^*-Ea_s)}_{\equiv\xi_1}\right| + \underbrace{\sup_t\left|t^{-1}\sum_{s=R+1}^{t}E^*(a_s^*-Ea_s)\right|}_{\equiv\xi_2}.$$

The first term is the absolute value of ξ_1, which can be rewritten as

$$\xi_1 \equiv \frac{R-\tau}{R}(R-\tau)^{-1}\sum_{s=1+\tau}^{R}E^*(a_s^*-Ea_s),$$

where $(R-\tau)^{-1}\sum_{s=1+\tau}^{R}E^*(a_s^*-Ea_s)$ is the MBB sample average of $a_s^*-Ea_s$. By well-known properties of the MBB (see e.g. Fitzenberger, 1998), we can show that this is equal to $(R-\tau)^{-1}\sum_{s=1+\tau}^{R}(a_s - Ea_s)+O_p\left(\frac{l}{R-\tau}\right)$. Hence, $\xi_1 = R^{-1}\sum_{s=1+\tau}^{R}(a_s-Ea_s)+O_p(l/R)$. We can show that $\xi_1 = O_p\left(R^{-1/2}\right)+o_p(1) = o_p(1)$ under Assumption 3 and the fact that $l/R = o(1)$. That $R^{-1}\sum_{s=1+\tau}^{R}(a_s - Ea_s)$ is $O_p\left(R^{-1/2}\right)$ follows by applying a maximal inequality for mixingales (see e.g. Lemma A.1 of Goncalves and Vogelsang (2011)). This result follows if $a_s - Ea_s \equiv x_{s-\tau}x'_{s-\tau} - Ex_{s-\tau}x'_{s-\tau}$ is an L_2-mixingale of size -1 (which is ensured by Assumption 3). We study ξ_2 next. This term relies on the MBB indices η_s. Using again the simplified assumption that $t - R = kl$, we can write

$$\sum_{s=R+1}^{t} E^*(a_s^* - Ea_s) = \sum_{i=1}^{k}\sum_{j=1}^{l} E^*\left(a_{J_i+(j-1)} - Ea_s\right) \text{ (where we note that } Ea_s \text{ is a constant)}$$

$$= l\sum_{i=1}^{k} E^*\left(\underbrace{l^{-1}\sum_{j=1}^{l}\left(a_{J_i+(j-1)} - Ea_s\right)}_{\equiv \mathcal{U}_i^*}\right) = klE^*(\mathcal{U}_1^*),$$

where the last equality holds because $\mathcal{U}_i^*$ is i.i.d. across i. Hence, we get that

$$\xi_2 \equiv \sup_{R+1\le t\le T}\left|t^{-1}\sum_{s=R+1}^{t} E^*(a_s^* - Ea_s)\right| \le R^{-1}\sup_{1\le k\le k_2}|klE^*(\mathcal{U}_1^*)| \le R^{-1}\underbrace{(k_2 l)}_{P-\tau+1}|E^*(\mathcal{U}_1^*)|.$$

The result follows by Assumption 5 (which implies that $R^{-1}(P-\tau+1) = O(1)$) and the fact that $E^*(\mathcal{U}_1^*) = O_p\left(P^{-1/2}\right)$. The latter follows because we can write

$$E^*(\mathcal{U}_1^*) = k_2^{-1}\sum_{i=1}^{k_2} E^*\left(\mathcal{U}_i^*\right) = E^*\left\{(k_2 l)^{-1}\sum_{i=1}^{k_2}\sum_{j=1}^{l}(a_{J_i+(j-1)} - Ea_s)\right\} = \frac{P}{P-\tau+1}E^*\left(P^{-1}\sum_{t=R+\tau}^{T+\tau}(a_t^* - Ea_t)\right),$$

which is the MBB bootstrap expectation of the sample average of $\{a_t^* - Ea_t : t = R+\tau, \ldots, T+\tau\}$ (note that $Ea_t = Ea_s$ since this is a constant under the stationarity assumption on x_t). Thus, by the properties of the MBB bootstrap expectation, we can write $E^*(\mathcal{U}_1^*)$ as

$$E^*(\mathcal{U}_1^*) = P^{-1}\sum_{t=R+\tau}^{T+\tau}(a_t - Ea_t) + O_p(l/P) = O_p\left(P^{-1/2}\right) + o_p(1) = o_p(1),$$

since the first term after the first equality is $O_p\left(P^{-1/2}\right)$, under Assumption 3 (as explained above). This concludes the proof that $\xi_2 = o_p(1)$ and the proof of part b).

Part (c). By definition, $\hat{\beta}_t^* = \left(t^{-1}\sum_{s=1+\tau}^{t} x_{s-\tau}^* x_{s-\tau}^{*\prime}\right)^{-1} t^{-1}\sum_{s=1+\tau}^{t} x_{s-\tau}^* y_s^*$. Letting $y_s^* = x_{s-\tau}^{*\prime}\beta_0 + (y_s^* - x_{s-\tau}^{*\prime}\beta_0)$ and recalling the definitions of $B^*(t)$ and $H^*(t)$ yields $\hat{\beta}_t^* - \beta_0 = B^*(t)H^*(t)$. Hence,

$$\begin{aligned} P^a\sup_t|\hat{\beta}_t^* - \beta_0| =& P^a\sup_t|B^*(t)H^*(t)| \\ \le& \sup_t\left|B^*(t) - B\right|P^a\sup_t\left|H^*(t) - E^*H^*(t)\right| + BP^a\sup_t|E^*H^*(t)| \\ &+ \sup_t\left|B^*(t) - B\right|P^a\sup_t\left|E^*H^*(t)\right| + BP^a\sup_t|H^*(t) - E^*H^*(t)|. \end{aligned}$$

Since $\sup_t\left|B^*(t) - B\right| = o_p^*(1)$ by part (b), and for $0 \le a < 1/2$, $P^a\sup_t\left|H^*(t) - E^*H^*(t)\right| = o_p^*(1)$ by part (a), the result follows by showing that $P^a\sup_t|E^*H^*(t)| = o_p(1)$. This result follows by the exact same arguments as in the proof of part (b). In particular, we can decompose $\sup_t |E^* H^*(t)| \leq \chi_1 + \chi_2$ where $\chi_1 = O_p\left(R^{-1/2}\right) + O_p\left(l/R\right)$ and $\chi_2 = O_p\left(P^{-1/2}\right) + O_p\left(l/P\right)$ (the proof of these results is the same as that used to study ξ_1 and ξ_2 in part b)). It follows that

$$P^a \sup_t |E^* H^*(t)| = O_p(P^a R^{-1/2}) + O_p\left(P^a l/R\right) + O_p(P^a P^{-1/2}) + O_p\left(P^a l/P\right) = o_p\left(1\right),$$

since P and R are of the same order magnitude (by Assumption 5), and $a < 1/2$. This implies that the first and third terms are $o_p\left(1\right)$. The second and fourth terms are also $o_p\left(1\right)$ under our assumptions. For instance, for the last term, because $l \to \infty$ such that $l/\sqrt{P} = o\left(1\right)$, we can write $O_p\left(P^a l/P\right) = O_p(P^a P^{-1/2} l P^{-1/2}) = o_p\left(1\right)$ since $l/P^{1/2} = o\left(1\right)$ and $P^{a-1/2} = o\left(1\right)$.

Part (d). Adding and subtracting appropriately, we can center $f^*_{t+\tau|r'}$ and $H^*(t)$ around their bootstrap means, i.e., $P^{-1/2}\sum_{t=R}^{T}(f^*_{t+\tau|r',\beta} - F)BH^*(t) = \sum_{i=1}^{4}\mathcal{F}_i$ where

$$\mathcal{F}_1 = P^{-1/2}\sum_{t=R}^{T}(f^*_{t+\tau|r',\beta} - E^* f^*_{t+\tau|r',\beta})B(H^*(t) - E^* H^*(t)), \quad \mathcal{F}_2 = P^{-1/2}\sum_{t=R}^{T}(E^* f^*_{t+\tau|r',\beta} - F)BE^* H^*(t)$$

$$\mathcal{F}_3 = P^{-1/2}\sum_{t=R}^{T}(E^* f^*_{t+\tau|r',\beta} - F)B(H^*(t) - E^* H^*(t)), \text{ and } \mathcal{F}_4 = P^{-1/2}\sum_{t=R}^{T}(f^*_{t+\tau|r',\beta} - E^* f^*_{t+\tau|r',\beta})BE^* H^*(t).$$

To prove part (d), it suffices to show that each $\mathcal{F}_i$ vanishes asymptotically. We start by showing $\mathcal{F}_2 = o_p(1)$. Note that $\mathcal{F}_2$ is bounded by

$$\mathcal{F}_2 \leq P^{-1/2}\sum_{t=R}^{T}\Big|E^* f_{t+\tau|r',\beta} - F\Big|B\sup_t\Big|E^* H^*(t)\Big|,$$

implying that $\mathcal{F}_2 = o_p(1)$ if $P^{-1/2}\sum_{t=R}^{T}\Big|E^* f_{t+\tau|r',\beta} - F\Big| = O_p(1)$ and $\sup_t \Big|E^* H^*(t)\Big| = o_p(1)$, where $\sup_t \Big|E^* H^*(t)\Big| = o_p(1)$ follows from the result of part (c). To show $P^{-1/2}\sum_{t=R}^{T}\Big|E^* f_{t+\tau|r',\beta} - F\Big| = O_p(1)$, it suffices to show that $P^{-1}\sum_{t=R}^{T}P^{1/2}E\Big|E^* f_{t+\tau|r',\beta} - F\Big| = O(1)$. This condition requires $P^{1/2}E\Big|E^* f_{t+\tau|r',\beta} - F\Big| = O(1)$ to hold for $t = R, \ldots, T$. Using Jensen's inequality, we can write

$$P^{1/2}E\Big|E^* f^*_{t+\tau|r',\beta} - F\Big| \leq P^{1/2}\Big(E\left(E^* f^*_{t+\tau|r',\beta} - F\right)^2\Big)^{1/2},$$

where for $t = R, \ldots, T$, $E\left(E^* f^*_{t+\tau|r',\beta} - F\right)^2 \leq O(1/P)$. Hence, $P^{1/2}E\Big|E^* f^*_{t+\tau|r',\beta} - F\Big| \leq O(1)$, completing the proof that $\mathcal{F}_2 = o^*_p(1)$.

For $\mathcal{F}_3$, we write

$$\mathcal{F}_3 \leq P^{-1}\sum_{t=R}^{T}P^{1/2}\Big|E^* f^*_{t+\tau|r',\beta} - F\Big|B\sup_t\Big|H^*(t) - E^* H^*(t)\Big|,$$

where $\sup_t \Big|H^*(t) - E^* H^*(t)\Big| = o^*_p(1)$ by Lemma A.2 (a) and $P^{1/2}\Big|E^* f_{t+\tau|r',\beta} - F\Big| = O_p(1)$. Hence, $\mathcal{F}_3 = o^*_p(1)$.

For $\mathcal{F}_4$, it suffices to show $E^*\mathcal{F}_4 = 0$ and $Var^*(\mathcal{F}_4) = o_p(1)$. Note that $E^*\mathcal{F}_4 = 0$ by design. Hence, we only need to show that $Var^*(\mathcal{F}_4) = o_p(1)$. For the sake of brevity, we let $\tau = 1$, and we define $a^*_{t+\tau} \equiv f^*_{t+\tau|r',\beta} - E^* f^*_{t+\tau|r',\beta}$, $c_t \equiv BE^*H^*(t)$. For $i = 1,\ldots,k_2$, we let $n_i = R+1+(i-1)l$, so $n_1 = R+1$ and $n_{k_2} = T+1-l+1$, where $R+1$ is the index that links to the first element in the first generated block, and $T+1-l+1$ is the index that links to the first element in the last generated block. Exploiting the independence between bootstrap blocks, we can write

$$Var^*(\mathcal{F}_4) = E^*\left(P^{-1/2}\sum_{t=R}^{T} a^*_{t+1}c_t\right)^2 = k_2^{-1}\sum_{i=1}^{k_2} l^{-1}E^*\left(\sum_{j=1}^{l} a^*_{n_i+(j-1)}c_{n_i+(j-1)-1}\right)^2.$$

Now, we let $d^*_{n_i+(j-1)} = a^*_{n_i+(j-1)}c_{n_i+(j-1)-1}$ and apply the standard MBB variance formula to get

$$Var^*(\mathcal{F}_4) = k^{-1}\sum_{i=1}^{k_2}\left(l^{-1}\sum_{j=1}^{l} E^* d^{*2}_{n_i+(j-1)} + 2l^{-1}\sum_{m=1}^{l-1}\sum_{j=1}^{l-m} E^*(d^*_{n_i+(j-1)}d^*_{n_i+(j-1+m)})\right),$$

where $l^{-1}\sum_{j=1}^{l} E^* d^{*2}_{n_i+(j-1)} \le O_p(P^{-1/2})$ and $\sum_{j=1}^{l-m} E^*(d^*_{n_i+(j-1)}d^*_{n_i+(j-1+m)}) \le O_p(l/P^{1/2})$ for $i = 1,\ldots,k_2$. For brevity, we only show $E^* d^{*2}_{n_i+(j-1)} = O_p(P^{-1/2})$ for $i = 1,\ldots,k_2$ and $j = 1,\ldots,l$. The proof that $\sum_{j=1}^{l-m} E^*(d^*_{n_i+(j-1)}d^*_{n_i+(j-1+m)}) = o_p(1)$ follows from a similar argument. Using the definition $d^*_{n_i+(j-1)} = a^*_{n_i+(j-1)}c_{n_i+(j-1)-1}$, we can write

$$\begin{aligned}
E^* d^{*2}_{n_i+(j-1)} &= Var^*(a^*_{n_i+(j-1)})c^2_{n_i+(j-1)-1} \\
&\le \left|Var^*(a^*_{n_i+(j-1)}) - \Gamma_{aa}(0)\right|c^2_{n_i+(j-1)-1} + |\Gamma_{aa}(0)|c^2_{n_i+(j-1)-1} \\
&\le \left|Var^*(a^*_{n_i+(j-1)}) - \Gamma_{aa}(0)\right|\left(B\sup_t |E^*H^*(t)|\right)^2 + |\Gamma_{aa}(0)|\left(B\sup_t |E^*H^*(t)|\right)^2
\end{aligned}$$

where $\Gamma_{aa}(0) \equiv Var(f_{t|r',\beta})$ and B are constants, $\sup_t |E^*H^*(t)| \le O_p(P^{-1/2})$ by result of part (c). We are left to show that $E\left|Var^*(a^*_{n_i+(j-1)}) - \Gamma_{aa}(0)\right| = O(1)$. Using Jensen's inequality,

$$\begin{aligned}
E\left|Var^*(a^*_{n_i+(j-1)}) - \Gamma_{aa}(0)\right| &\le \left[E\left(Var^*(a^*_{n_i+(j-1)}) - \Gamma_{aa}(0)\right)^2\right]^{1/2} \\
&\le \left[Var\left(Var^*(a^*_{n_i+(j-1)})\right) + \left(EVar^*(a^*_{n_i+(j-1)}) - \Gamma_{aa}(0)\right)^2\right]^{1/2}
\end{aligned}$$

where for $i = 1,\ldots,k_2$,

$$Var^*(a^*_{n_i+(j-1)}) = \frac{1}{P+1-l}\sum_{t=R+1}^{T+1-l+1}\left(f_{t+(j-1)|r',\beta} - \mathcal{C}_{f,j-1}\right)^2 \text{ with } \mathcal{C}_{f,j-1} = \frac{1}{P+1-l}\sum_{s=R+1}^{T+1-l+1} f_{s+(j-1)|r',\beta}.$$

Using the uniform fourth moment bound on a_t, $Var\left(Var^*(a^*_{n_i+(j-1)})\right) \to 0$. Note that

$$EVar^*(a^*_{n_i+(j-1)}) - \Gamma_{aa}(0) = -E(\mathcal{C}_{f,j-1})^2 \le O(1)$$

which completes this part of the proof. To show $\mathcal{F}_1 = o_p^*(1)$, we use the definition of $H^*(t)$ to decompose $\mathcal{F}_1$ as $\mathcal{F}_1 = \mathcal{F}_{1.1} + \mathcal{F}_{1.2}$ where

$$\mathcal{F}_{1.1} = P^{-1/2} \sum_{t=R}^{T} \left(f^*_{t+\tau|r',\beta} - E^* f^*_{t+\tau|r',\beta} \right) B \left(t^{-1} \sum_{s=1+\tau}^{R} (h_s^* - E^* h_s^*) \right),$$

$$\mathcal{F}_{1.2} = P^{-1/2} \sum_{t=R+1}^{T} \left(f^*_{t+\tau|r',\beta} - E^* f^*_{t+\tau|r',\beta} \right) B \left(t^{-1} \sum_{s=R+1}^{t} (h_s^* - E^* h_s^*) \right).$$

Then the result follows if $\mathcal{F}_{1.1}$ and $\mathcal{F}_{1.2}$ are both $o_p^*(1)$. For $\mathcal{F}_{1.1}$, we first rewrite it as

$$\mathcal{F}_{1.1} = \left(\frac{R}{P} \right)^{1/2} \left(\sum_{t=R}^{T} t^{-1} \left(f^*_{t+\tau|r',\beta} - E^* f^*_{t+\tau|r',\beta} \right) \right) B \left(R^{-1/2} \sum_{s=1+\tau}^{R} (h_s^* - E^* h_s^*) \right).$$

The result follows if $\sum_{t=R}^{T} t^{-1} \left(f^*_{t+\tau|r',\beta} - E^* f^*_{t+\tau|r',\beta} \right) = o_p^*(1)$, i.e., the middle term on the right-hand side of the above equation vanishes asymptotically. This is because $\frac{R}{P} \to \pi < \infty$, and by Corollary 3.1 in Fitzenberger (1998) $R^{-1/2} \sum_{s=1+\tau}^{R} (h_s^* - E^* h_s^*) = O_p^*(1)$. To show the middle term vanishes asymptotically, it suffices to show the bootstrap variance of the middle term is $o_p(1)$. This is because the bootstrap mean of the middle term is exactly zero, and the result will follow by Chebyshev's inequality. Using independence of the bootstrap blocks, we can write

$$Var^* \left(\sum_{t=R}^{T} t^{-1} \left(f^*_{t+\tau|r',\beta} - E^* f^*_{t+\tau|r',\beta} \right) \right) = \sum_{i=0}^{k_2-1} E^* \left(\sum_{j=0}^{l-1} \frac{f^*_{R+\tau+il+j|r',\beta} - E^* f^*_{R+\tau+il+j|r',\beta}}{R+\tau+il+j} \right)^2.$$

Then we argue the right-hand side of the above equality is $o_p(1)$. For simplicity, we let $m_i = R+\tau+il$. By Cauchy-Schwarz inequality, we can write,

$$\begin{aligned} \sum_{i=0}^{k_2-1} E^* \left(\sum_{j=0}^{l-1} \frac{f^*_{m_i+j|r',\beta} - E^* f^*_{m_i+j|r',\beta}}{m_i+j} \right)^2 &\leq \sum_{i=0}^{k_2-1} \left(\sum_{j=0}^{l-1} \left(\frac{1}{m_i+j} \right)^2 \right) \left(\sum_{j=0}^{l-1} E^* \left(f^*_{m_i+j|r',\beta} - E^* f^*_{m_i+j|r',\beta} \right)^2 \right) \\ &\leq \underbrace{\sum_{i=0}^{k_2-1} \left(\frac{l}{R^2} \right) \sum_{j=0}^{l-1} Var^* \left(f^*_{m_i+j|r',\beta} \right)}_{\equiv \mathcal{M}}. \end{aligned}$$

Note that we can also write the right-hand side of the above inequality as

$$\mathcal{M} = \left(\frac{l}{R^2} \right) \left(\sum_{t=R}^{T} Var^* \left(f^*_{t+\tau|r',\beta} - F \right) \right).$$

By Markov's inequality, the result follows since $T/R < \infty$, $l/R \to 0$, and $EVar^* \left(f^*_{t+\tau|r',\beta} \right) \leq O(1)$ uniformly in t. In particular, we can show $E\mathcal{M} \to 0$ by noting that

$$EVar^* \left(f^*_{t+\tau|r',\beta} - F \right) \leq E \left(E^* \left(f^*_{t+\tau|r',\beta} - F \right)^2 \right) + E \left(E^* f^*_{t+\tau|r',\beta} - F \right)^2.$$

Let $\gamma_j \equiv Cov\left(f_{s|r',\beta} f_{s+j|r',\beta}\right) = Cov\left(f_{s|r',\beta} f_{s-j|r',\beta}\right) = \gamma_{-j}$. By the definition of bootstrap expectation and the covariance stationary assumption, $E\left(E^*\left(f^*_{t+\tau|r',\beta} - F\right)^2\right) \leq |\gamma_0|$, and

$$E\left(E^* f^*_{t+\tau|r',\beta} - F\right)^2 \leq \frac{1}{P} E\left(\sqrt{P-l+1} E^*\left(f^*_{t+\tau|r',\beta} - F\right)\right)^2 \leq \frac{1}{P}\left(\sum_{j=-\infty}^{\infty} |\gamma_j|\right).$$

For these reasons, $EVar^*\left(f_{t+\tau|r',\beta} - F\right) \leq |\gamma_0| + \frac{1}{P}\sum_{j=-\infty}^{\infty} |\gamma_j|$. This implies that

$$E\mathcal{M} \leq \frac{l}{R}\frac{P}{R}|\gamma_0| + \frac{l}{R}\frac{1}{R}\sum_{j=-\infty}^{\infty} |\gamma_j|.$$

Since $\frac{l}{R} \to 0$, $\frac{P}{R} \to \pi < \infty$, and $\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty$, $E\mathcal{M} \to 0$. The term $\mathcal{F}_{1.2}$ can be handled similarly, so we omit the details here.

Part (e). Note that

$$\begin{aligned} P^{-1/2}\sum_{t=R}^{T} F(B^*(t) - B)H^*(t) \leq & \sup_t |B^*(t) - B| F P^{-1/2}\sum_{t=R}^{T} |H^*(t) - E^*H^*(t)| \\ & + \sup_t |B^*(t) - B| F P^{-1/2}\sum_{t=R}^{T} |E^*H^*(t)|, \end{aligned}$$

where $\sup_t |B(t)^* - B| = o_p^*(1)$ by part (b). The result follows by showing that (i) $P^{-1/2}\sum_{t=R}^{T} |H^*(t) - E^*H^*(t)| = O_p^*(1)$ and (ii) $P^{-1/2}\sum_{t=R}^{T} |E^*H^*(t)| = O_p(1)$. To prove (i), it suffices to show $E^*|H^*(t) - E^*H^*(t)| \leq O_p(R^{-1/2})$ uniformly in t. For $t = R, \ldots, T$, observe that $E^*|H^*(t) - E^*H^*(t)| \leq \left[E^*\big(H^*(t) - E^*H^*(t)\big)^2\right]^{1/2}$, where

$$\begin{aligned} E^*\big(H^*(t) - E^*H^*(t)\big)^2 =& E^*\left(t^{-1}\sum_{s=1+\tau}^{R}(h_s^* - E^*h_s^*) + \frac{\mathbf{1}_{\{R+1\leq t\leq T\}}}{t}\sum_{s=R+1}^{t}(h_s^* - E^*h_s^*)\right)^2 \\ =& \frac{R}{t^2}E^*\left(R^{-1/2}\sum_{s=1+\tau}^{R}(h_s^* - E^*h_s^*)\right)^2 \\ & + \frac{\mathbf{1}_{\{R+1\leq t\leq T\}}(t-R)}{t^2}E^*\left((t-R)^{-1/2}\sum_{s=R+1}^{t}(h_s^* - E^*h_s^*)\right)^2 \end{aligned}$$

where $\mathbf{1}_{\{R+1\leq t\leq T\}}$ is an indicator function that equals 1 if $R+1 \leq t \leq T$. Note that $\frac{R}{t^2} \leq R^{-1}$, $(t-R)t^{-2} \leq PR^{-2} = O(R^{-1})$, $Var^*(R^{-1/2}\sum_{s=1+\tau}^{R} h_s^*) = O_p(1)$, $Var^*((t-R)^{-1/2}\sum_{s=R+1}^{t} h_s^*) = O_p(1)$, and $Cov^*(R^{-1/2}\sum_{s=1+\tau}^{R} h_s^*, (t-R)^{-1/2}\sum_{s=R+1}^{t} h_s^*) = 0$. To complete the proof, we show (ii). This follows from noting that

$$P^{-1/2}\sum_{t=R}^{T} |E^*H^*(t)| \leq P^{1/2}\sup_t |E^*H^*(t)| = O_p(1),$$

since we already showed that $\sup_t |E^*H^*(t)| \leq \chi_1 + \chi_2 = O_p\left(R^{-1/2}\right) + O_p(l/R) + O_p(P^{-1/2}) + O_p(l/P)$ in part (c).

Part (f). Adding and subtracting appropriately,

$$P^{-1/2}\sum_{t=R}^{T}(f^*_{t+\tau|r',\beta} - F)\Big(B^*(t) - B\Big)H^*(t) \leq \sup_t \Big|B^*(t) - B\Big| P^{-1/2}\sum_{t=R}^{T}\Big|f^*_{t+\tau|r',\beta} - F\Big|\, |H^*(t) - E^*H^*(t)|$$

$$+ \sup_t \Big|B^*(t) - B\Big| P^{-1/2}\sum_{t=R}^{T}\Big|f^*_{t+\tau|r',\beta} - F\Big|\, |E^*H^*(t)|.$$

Given part b), it suffices to show (i) $P^{-1/2}\sum_{t=R}^{T}\Big|f^*_{t+\tau|r',\beta} - F\Big||H^*(t) - E^*H^*(t)| = O^*_p(1)$ and (ii) $P^{-1/2}\sum_{t=R}^{T}\Big|f^*_{t+\tau|r',\beta} - F\Big|\,|E^*H^*(t)| = O^*_p(1)$. We can prove (i) by applying the Cauchy-Schwarz inequality and using the fact that $E^*\big(H^*(t) - E^*H^*(t)\big)^2 = O_p\left(\min(R,P)^{-1/2}\right) + O_p\left(l/\min(R,P)\right)$, as in the proof of part (e). Part (ii) follows by noting that $E^*\big(f^*_{t+\tau|r',\beta} - F\big)^2 = O_p(1)$ (by Lemma A.3) and the fact that $\sup_t |E^*H^*(t)| = O_p\left(\min(R,P)^{-1/2}\right) + O_p\left(l/\min(R,P)\right)$. ■

Proof of Lemma A.3. Let $M' - M + 1 = kl$ and generate $I_1, \ldots, I_k \sim$ i.i.d. Uniform on $\{N, \ldots, N' - l + 1\}$. Set

$$Z^*_{M+(i-1)l+(j-1)} = Z_{I_i+(j-1)}, \text{ for } i = 1, \ldots, k,\ j = 1, \ldots, l,$$

and note that this yields a bootstrap sample

$$\left\{Z^*_M = Z_{I_1}, Z^*_{M+1} = Z_{I_1+1}, \ldots, Z^*_{M+l-1} = Z_{I_1+l-1}, Z^*_{M+l} = Z_{I_2}, \ldots, Z^*_{M'} = Z_{I_k+l-1}\right\}.$$

Since I_i is i.i.d. Uniform on $\{N, \ldots, N' - l + 1\}$, it follows that for $i = 1, \ldots, k$, $j = 1, \ldots, l$,

$$E^*(Z^*_{M+(i-1)l+(j-1)}) = E^*(Z_{I_i+(j-1)}) = (N' - N - l + 2)^{-1}\sum_{t=N}^{N'-l+1} Z_{t+(j-1)},$$

which varies with $j = 1, \ldots, l$, but not with $i = 1, \ldots, k$. Let $n = N' - N + 1$ and $\bar{Z}_n \equiv n^{-1}\sum_{t=N}^{N'} Z_t$. For each $j = 1, \ldots, l$, we can write

$$(N' - N - l + 2)^{-1}\sum_{t=N}^{N'-l+1} Z_{t+(j-1)} = \frac{n}{n - l + 1}\left(n^{-1}\sum_{t=N}^{N'} Z_t\right) + (n - l + 1)^{-1} A_j,$$

where $A_j \equiv \sum_{t=N}^{N'-l+1} Z_{t+(j-1)} - \sum_{t=N}^{N'} Z_t$. Each term A_j can be written as a sum involving $l - 1$ observations in $\{Z_t\}$. For instance, for $j = 1$,

$$(n - l + 1)^{-1} A_1 = \frac{l - 1}{n - l + 1}\underbrace{(l-1)^{-1}(Z_{N'-l+2} + \ldots + Z_{N'})}_{a_1 = O_p(1)} = O_p(l/n) \text{ if } l/n = o(1),$$

where $a_1 = O_p(1)$ because $E|Z_t| \leq \Delta$ for all t. For $j = 2$,

$$(n - l + 1)^{-1} A_2 = \frac{l - 1}{n - l + 1}\underbrace{(l-1)^{-1}(Z_{N+1} + Z_{N'-l+3} + \ldots + Z_{N'})}_{a_2 = O_p(1)} = O_p(l/n) \text{ if } l/n = o(1).$$

Since, for any j, we can show that $(n-l+1)^{-1}E|A_j| = O(l/n)$ uniformly in j, the result follows. ■

Proof of Lemma A.4. **Part (a).** Recall that $S_{1P}^{*} = P^{-1/2}\sum_{t=R+\tau}^{T+\tau}(f_{t|r'}^{*} - f_{t|r'})$. We start by showing $S_{1P}^{*} = \tilde{S}_{1P}^{*} + o_p^{*}(1)$ where $\tilde{S}_{1P}^{*} \equiv P^{-1/2}\sum_{t=R+1}^{T+\tau}(f_{t|r'}^{*} - E^{*}f_{t|r'}^{*})$. Using $\tilde{S}_{1P}^{*}$ helps to set the first summand of S_{1P}^{*} from $f_{R+\tau|r'}^{*}$ to $f_{R+1|r'}^{*}$ where $f_{R+1|r'}^{*} = f_{\eta_{R+1}|r'}$ is the first element of the first random block based on $\{\eta_{R+1},\ldots,\eta_{R+1}+(l-1)\}$, and $\tilde{S}_{1P}^{*}$ is centered around bootstrap mean. Adding and subtracting appropriately,

$$S_{1P}^{*} = \tilde{S}_{1P}^{*} + P^{-1/2}\Delta^{*} + P^{-1/2}\Delta$$

where $\Delta^{*} = \sum_{t=R+\tau}^{T+\tau} f_{t|r'}^{*} - \sum_{s=R+1}^{T+\tau} f_{s|r'}^{*}$, $\Delta = \sum_{s=R+1}^{T+\tau} Ef_{s|r'}^{*} - \sum_{t=R+\tau}^{T+\tau} f_{t|r'}$. Note that Δ^{*} is at most $O_p^{*}(\tau)$, and it is exactly zero when $\tau = 1$. Hence $P^{-1/2}\Delta^{*}$ vanishes asymptotically. Let $\mathcal{C}_f = P^{-1}\sum_{t=R+\tau}^{T+\tau} f_{t|r'}$ then

$$\Delta = \sum_{s=R+1}^{T+\tau}\left(E^{*}f_{s|r'} - \mathcal{C}_f + \mathcal{C}_f\right) - P\mathcal{C}_f = \sum_{s=R+1}^{T+\tau}\left(E^{*}f_{s|r'} - \mathcal{C}_f\right) + (\tau-1)\mathcal{C}_f$$

where $E^{*}f_{s|r'} - \mathcal{C}_f \leq O_p(l/P)$ for $t = R+1,\ldots,T+\tau$ by Lemma A.3, and $\mathcal{C}_f \leq O_p(1)$. This implies $P^{-1/2}\Delta \leq P^{-1/2}(P+\tau-1)O_p(l/P) + P^{-1/2}(\tau-1)O_p(1)$. Hence, $P^{-1/2}\Delta$ vanishes asymptotically under the block length condition $l/\sqrt{P} \to 0$. Using these results, we can write $S_{1P}^{*} = \tilde{S}_{1P}^{*} + o_p^{*}(1)$. This implies

$$\lim_{R,P\to\infty} Var^{*}\left(S_{1P}^{*}\right) = \lim_{R,P\to\infty} Var^{*}\left(\tilde{S}_{1P}^{*}\right)$$

where

$$Var^{*}\left(\tilde{S}_{1P}^{*}\right) \xrightarrow{p} \lim_{R,P\to\infty} Var\left(P^{-1/2}\sum_{t=R+\tau}^{T+\tau} f_{t|r'}\right) = \Omega_1$$

by Corollary 3.1 of Fitzenberger (1998).

Part (b). From Lemma 5.1, we know that

$$S_{2P}^{*} = \underbrace{a_{R,0}P^{-1/2}\sum_{s=1+\tau}^{R}(h_s^{*} - \bar{h}_R)}_{S_{2P.1}^{*}} + \underbrace{P^{-1/2}\sum_{i=1}^{P-1} a_{R,i}(h_{R+i}^{*} - \bar{h}_P)}_{S_{2P.2}^{*}}.$$

We first recenter S_{2P}^{*} by adding and subtracting the appropriate bootstrap mean of $S_{2P.1}^{*}$ and $S_{2P.2}^{*}$

$$S_{2P.1}^{*} = a_{R,0}P^{-1/2}\sum_{s=1+\tau}^{R}(h_s^{*} - E^{*}h_s^{*}) + a_{R,0}P^{-1/2}\sum_{s=1+\tau}^{R}(E^{*}h_s^{*} - \bar{h}_R)$$
$$S_{2P.2}^{*} = P^{-1/2}\sum_{i=1}^{P-1} a_{R,i}(h_{R+i}^{*} - E^{*}h_{R+i}^{*}) + P^{-1/2}\sum_{i=1}^{P-1} a_{R,i}(E^{*}h_{R+i}^{*} + \bar{h}_P)$$

where $a_{R,0}P^{-1/2}\sum_{s=1+\tau}^{R}(E^*h_s^* - \bar{h}_R) = o_p(1)$ since $a_{R,0} < \infty$ and $E^*h_s^* - \bar{h}_R = O_p(l/R)$ for $s = 1+\tau,\ldots,R$, and $P^{-1/2}\sum_{i=1}^{P-1} a_{R,i}(E^*h_{R+i}^* - \bar{h}_P) = o_p(1)$ since $P^{-1}\sum_{i=1}^{P-1} a_{R,i} \to 1 - \pi^{-1}\ln(1+\pi)$ (see West (1996), Lemma 4.1) and $E^*h_{R+i}^* - \bar{h}_P = O_p(l/R)$ for $i = 1,\ldots,P-1$. Now, we can write

$$S_{2P}^* = \underbrace{a_{R,0}P^{-1/2}\sum_{s=1+\tau}^{R}(h_s^* - E^*h_s^*)}_{\tilde{S}_{2P.1}^*} + \underbrace{P^{-1/2}\sum_{i=1}^{P-1} a_{R,i}(h_{R+i}^* - E^*h_{R+i}^*)}_{\tilde{S}_{2P.2}^*} + o_p^*(1),$$

and

$$\lim_{R,P\to\infty} Var^*\left(S_{2P}^*\right) = \lim_{R,P\to\infty}\left(Var^*\left(\tilde{S}_{2P.1}^*\right) + Var^*\left(\tilde{S}_{2P.2}^*\right) + 2Cov^*\left(\tilde{S}_{2P.1}, \tilde{S}_{2P.2}^*\right)\right).$$

Note that

$$\lim_{R,P\to\infty} Var^*(\tilde{S}_{2P.1}^*) = \lim_{R,P\to\infty} a_{R,0}^2 P^{-1} R\, Var^*\left(R^{-1/2}\sum_{s=1+\tau}^{R}(h_s^* - E^*h_s^*)\right)$$

where $a_{R,0}^2 P^{-1}R \to \pi^{-1}\ln^2(1+\pi)$ by West (1996), page 1082 (A-1a); by using Fitzenberger's (1998) Corollary 3.1,

$$Var^*\left(R^{-1/2}\sum_{t=1+\tau}^{R}(h_t^* - E^*h_s^*)\right) \xrightarrow{p} Var\left(R^{-1/2}\sum_{t=1+\tau}^{R} h_t\right) \to \sum_{j=-\infty}^{\infty}\Gamma_{hh}(j)$$

where $\Gamma_{hh}(j) \equiv E(h_t h_{t+j}) = E(h_t h_{t-j}) \equiv \Gamma_{hh}(-j)$. Hence $Var^*(\tilde{S}_{2P.1}^*) \xrightarrow{p} \Omega_{2.1}$.

For $Var^*(\tilde{S}_{2P.2}^*)$, we first rewrite $\tilde{S}_{2P.2}^*$ as

$$\tilde{S}_{2P.2}^* = P^{-1/2}\sum_{i=1}^{P-1} a_{R,i}(h_{R+i}^* - E^*h_{R+i}^*) = P^{-1/2}\sum_{t=R+1}^{T+\tau} c_t(h_t^* - E^*h_t^*).$$

where

$$c_{R+i} = \begin{cases} a_{R,i} \text{ for } 1 \le i \le P-1 \\ 0 \text{ if } P \le i \le P-1+\tau. \end{cases}$$

By exploiting the independence between blocks, we write

$$Var^*\left(P^{-1/2}\sum_{t=R+1}^{T+\tau} c_t(h_t^* - E^*h_t^*)\right) = \underbrace{P^{-1}\sum_{i=1}^{k_2} Var^*\left(\sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1} c_t(h_t^* - E^*h_t^*)\right)}_{\mathcal{V}}$$

Now, we show $\mathcal{V} \xrightarrow{p} \Omega_{2.2}$. Adding and subtracting appropriately, we write $\mathcal{V} = \mathcal{V}_1 + \mathcal{V}_2 + \mathcal{V}_3$ where

$$\mathcal{V}_1 = P^{-1}\sum_{i=1}^{k_2}\left(\sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1} c_t^2\Gamma_{hh}(0) + 2\sum_{j=1}^{l-1}\sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1-j} c_t c_{t+j}\Gamma_{hh}(j)\right)$$

$$\mathcal{V}_2 = P^{-1}\sum_{i=1}^{k_2}\sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1} c_t^2\Big(Var^*(h_t^*) - \Gamma_{hh}(0)\Big)$$

$$\mathcal{V}_3 = 2\sum_{j=1}^{l-1} P^{-1}\sum_{i=1}^{k_2}\sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1-j} c_t c_{t+j}\Big(Cov^*(h_t^*, h_{t+j}^*) - \Gamma_{hh}(j)\Big)$$

where $\mathcal{V}_2$ and $\mathcal{V}_3$ goes to zero in probability. We show $\mathcal{V}_2 = o_p(1)$; similar arguments apply to proving $\mathcal{V}_3 = o_p(1)$. For $i = 1, \ldots, k_2$, we let $m_i = R + 1 + (i-1)l$. Using this notation, we can bound $\mathcal{V}_2$ as follows

$$\mathcal{V}_2 = P^{-1}\sum_{i=1}^{k_2}\sum_{j=1}^{l} c^2_{m_i+(j-1)}\left(Var^*(h^*_{m_i+(j-1)}) - \Gamma_{hh}(0)\right) \leq P^{-1}\sum_{i=1}^{k_2}\sum_{j=1}^{l} c^2_{m_i+(j-1)}\left|Var^*(h^*_{m_i+(j-1)}) - \Gamma_{hh}(0)\right|.$$

A sufficient condition for $\mathcal{V}_2 = o_p(1)$ is $\left(P^{-1}\sum_{t=R+1}^{T+\tau} c_t^2\right) E\left|Var^*(h^*_{m_i+(j-1)}) - \Gamma_{hh}(0)\right| \to 0$ where

$$P^{-1}\sum_{i=1}^{k_2}\sum_{j=1}^{l} c^2_{m_i+(j-1)} = P^{-1}\sum_{t=R+1}^{T+\tau} c_t^2 \to 2[1-\pi^{-1}\ln(1+\pi)] - \pi^{-1}\ln(1+\pi)$$

by equation (A1-1b) in Lemma A.5 of West (1996). Thus, we only need to show $E\left|Var^*(h^*_{m_i+(j-1)}) - \Gamma_{hh}(0)\right| \to 0$. Using Jensen's inequality,

$$\begin{aligned} E\left|Var^*(h^*_{m_i+(j-1)}) - \Gamma_{hh}(0)\right| &\leq \left[E\left(Var^*(h^*_{m_i+(j-1)}) - \Gamma_{hh}(0)\right)^2\right]^{1/2} \\ &\leq \left[Var\left(Var^*(h^*_{m_i+(j-1)})\right) + \left(EVar^*(h^*_{m_i+(j-1)}) - \Gamma_{hh}(0)\right)^2\right]^{1/2} \end{aligned}$$

where for $i = 1, \ldots, k_2$,

$$Var^*(h_{m_i+(j-1)}) = \frac{1}{P+\tau-l}\sum_{t=R+1}^{T+\tau-l+1}\left(h_{t+(j-1)} - \mathcal{C}_{h,j}\right)^2 \text{ with } \mathcal{C}_{h,j} = \frac{1}{P+\tau-l}\sum_{s=R+1}^{T+\tau-l+1} h_{s+(j-1)}.$$

Using the uniform fourth moment bound on h_t, $Var\left(Var^*(h^*_{m_i+(j-1)})\right) \to 0$. Note that

$$EVar^*(h^*_{m_i+(j-1)}) - \Gamma_{hh}(0) = -E(\mathcal{C}_{h,j})^2 \leq O\left(\frac{1}{P+\tau-l}\right) \to 0$$

which completes the proof of $\mathcal{V}_2 = o_p(1)$. Similar proofs can be found in the proof of Lemma A3 of Corradi and Swanson (2003) equation (38)-(39) or equation A.9 of Corradi and Swanson (2007).

Since $\mathcal{V}_2$ and $\mathcal{V}_3$ both converge to zero in probability, we only need to focus on $\mathcal{V}_1$. Adding and subtracting, we can write $\mathcal{V}_1 = \mathcal{V}_{1.1} + \mathcal{V}_{1.2}$ where

$$\begin{aligned} \mathcal{V}_{1.1} =& P^{-1}\sum_{i=1}^{k_2}\left(\sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1} c_t^2\Gamma_{hh}(0) + 2\sum_{j=1}^{l-1}\sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1} c_t^2\Gamma_{hh}(j)\right) \\ \mathcal{V}_{1.2} =& P^{-1}\sum_{i=1}^{k_2}\left(2\sum_{j=1}^{l-1}\left(\sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1-j} c_t c_{t+j} - \sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1} c_t^2\right)\Gamma_{hh}(j)\right), \end{aligned}$$

where $\mathcal{V}_{1.2}$ converges to zero. This step can be shown by using an argument similar to the proof of West (1996), equation (A-1b). In particular, for $i = 1, ..., k_2$, we can write

$$2l^{-1}\sum_{j=1}^{l-1}\left(\sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1-j} c_t c_{t+j} - \sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1} c_t^2\right)\Gamma_{hh}(j) \to 0.$$

Now, we are only left with $\mathcal{V}_{1.1}$ and

$$\mathcal{V}_{1.1} = \left(P^{-1} \sum_{t=R+1}^{T+\tau} c_t^2\right) \left(\sum_{-l+1}^{l-1} \Gamma_{hh}(j)\right),$$

where

$$P^{-1} \sum_{t=R+1}^{T+\tau} c_t^2 = P^{-1} \sum_{i=1}^{P-1} a_{R,i}^2 \to 2[1 - \pi^{-1} \ln(1+\pi)] - \pi^{-1} \ln(1+\pi).$$

Hence,

$$\lim_{R,P\to\infty} Var^* \left(S_{2P.2}^*\right) = \lim_{R,P\to\infty} \mathcal{V} + o_p(1) \xrightarrow{p} \lim_{R,P\to\infty} \mathcal{V}_{1.1} = \Omega_{2.2}.$$

Part (c). Using the results of part (a) and (b) of this lemma, we can write

$$\lim_{R,P\to\infty} Cov^*(S_{1P}, S_{2P}) = \lim_{R,P\to\infty} Cov^*\Big(\tilde{S}_{1P}, (\tilde{S}_{2P.1}^* + \tilde{S}_{2P.2}^*)\Big).$$

Exploiting the independence between $\{\gamma_{1+\tau}, \ldots, \gamma_R\}$ and $\{\eta_{R+1}, \ldots, \eta_{T+\tau}\}$, we can write

$$Cov^*\Big(\tilde{S}_{1P}, (\tilde{S}_{2P.1}^* + \tilde{S}_{2P.2}^*)\Big) = Cov^* \left(\tilde{S}_{1P}^*, \tilde{S}_{2P.2}^*\right)$$

where $Cov^* \left(\tilde{S}_{1P}^*, \tilde{S}_{2P.1}^*\right) = 0$. Using the notation c_{R+i}, see definition in Part (b), we can write

$$Cov^*(\tilde{S}_{1P}^*, \tilde{S}_{2P.2}^*) = P^{-1} Cov^* \left(\sum_{t=R+1}^{T+\tau} (f_{t|r'}^* - E^* f_{t|r'}^*), \sum_{t=R+1}^{T+\tau} c_t(h_t^* - E^* h_t^*)\right).$$

Exploiting the independence between blocks, we write

$$\begin{aligned} &P^{-1} Cov^* \left(\sum_{t=R+1}^{T+\tau} (f_{t|r'}^* - E^* f_{t|r'}^*), \sum_{t=R+1}^{T+\tau} c_t(h_t^* - E^* h_t^*)\right) \\ &= \underbrace{P^{-1} \sum_{i=1}^{k_2} Cov^* \left(\sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1} (f_{t|r'}^* - E^* f_{t|r'}^*), \sum_{t=R+1+(i-1)l}^{R+1+(i-1)l+l-1} c_t(h_t^* - E^* h_t^*)\right)}_{\mathcal{W}} \end{aligned}$$

For notation simplicity, we let $b_i = R + 1 + (i-1)l$ for $i = 1, \ldots, k_2$ then $\mathcal{W} = \mathcal{W}_1 + \mathcal{W}_2 + \mathcal{W}_3$ where

$$\mathcal{W}_1 = P^{-1} \sum_{i=1}^{k_2} \sum_{m=1}^{l-1} \sum_{j=1}^{l} Cov^* \left(f_{b_i+(j-1)|r'}^*,\ c_{b_i+(j-1)+m} h_{b_i+(j-1)+m}^*\right),$$

$$\mathcal{W}_2 = P^{-1} \sum_{i=1}^{k_2} \sum_{j=1}^{l} Cov^* \left(f_{b_i+(j-1)|r'}^*,\ c_{b_i+(j-1)} h_{b_i+(j-1)}^*\right),$$

$$\mathcal{W}_3 = P^{-1} \sum_{i=1}^{k_2} \sum_{m=1}^{l-1} \sum_{j=1}^{l} Cov^* \left(f_{b_i+(j-1)+m|r'}^*,\ c_{b_i+(j-1)} h_{b_i+(j-1)}^*\right).$$

Adding and subtracting appropriately, we have

$$\mathcal{W}_1 = \mathcal{W}_{1.1} + (\mathcal{W}_1 - \mathcal{W}_{1.1}) \text{ where } \mathcal{W}_{1.1} = P^{-1}\sum_{i=1}^{k_2}\sum_{m=1}^{l-1}\sum_{j=1}^{l} c_{b_i+(j-1)+m}\Gamma_{fh}(-m),$$

$$\mathcal{W}_2 = \mathcal{W}_{2.1} + (\mathcal{W}_2 - \mathcal{W}_{2.1}) \text{ where } \mathcal{W}_{2.1} = P^{-1}\sum_{i=1}^{k_2}\sum_{j=1}^{l} c_{b_i+(j-1)}\Gamma_{fh}(0)$$

$$\mathcal{W}_3 = \mathcal{W}_{3.1} + (\mathcal{W}_3 - \mathcal{W}_{3.1}) \text{ where } \mathcal{W}_{3.1} = P^{-1}\sum_{i=1}^{k_2}\sum_{m=1}^{l-1}\sum_{j=1}^{l} c_{b_i+(j-1)}\Gamma_{fh}(m)$$

where $\Gamma_{fh}(m) = E(f_{t|r'}h_{t+m}) = E(f_{t|r'}h_{t-m}) = \Gamma_{fh}(-m)$. Note that $|\mathcal{W}_1 - \mathcal{W}_{1.1}| = o_p(1)$, $|\mathcal{W}_2 - \mathcal{W}_{2.1}| = o_p(1)$ and $|\mathcal{W}_3 - \mathcal{W}_{3.1}| = o_p(1)$ by using the same arguments that proves the results of part (b). Further adding and subtracting on $\mathcal{W}_{1.1}$ and $\mathcal{W}_{3.1}$, we can write

$$\mathcal{W}_{1.1} = \mathcal{W}_{1.1.1} + (\mathcal{W}_{1.1} - \mathcal{W}_{1.1.1}) \text{ where } \mathcal{W}_{1.1.1} = \sum_{m=1}^{l-1} P^{-1}\sum_{i=1}^{k_2}\sum_{j}^{l} c_{b_i+(j-1)}\Gamma_{fh}(-m),$$

$$\mathcal{W}_{3.1} = \mathcal{W}_{3.1.1} + (\mathcal{W}_{3.1} - \mathcal{W}_{3.1.1}) \text{ where } \mathcal{W}_{3.1.1} = \sum_{m=1}^{l-1} P^{-1}\sum_{i=1}^{k_2}\sum_{j=1}^{l} c_{b_i+(j-1)}\Gamma_{fh}(m)$$

Note that $|\mathcal{W}_{1.1} - \mathcal{W}_{1.1.1}| = o(1)$ and $|\mathcal{W}_{3.1} - \mathcal{W}_{3.1.1}| = o(1)$ by an argument similar to that used in the proof of West's (1996) Lemma A.6. Hence,

$$Cov^*\left(\tilde{S}^*_{1P}, \tilde{S}^*_{2P.2}\right) \xrightarrow{p} P^{-1}\left(\sum_{i=1}^{P-1} a_{R,i}\right)\sum_{m=-l+1}^{l-1}\Gamma_{fh}(m) \to \Omega_{12}$$

where $P^{-1}\left(\sum_{i=1}^{P-1} a_{R,i}\right) \to 1 - \pi^{-1}\ln(1+\pi)$, see remark in part (b) or by Lemma A6 of West (1996). ■

A.2 Proofs of results in the paper

Proof of Lemma 4.1. Given Lemma A.1, by two mean value expansions of $f_{t+\tau|r'}(\hat{\beta}_t)$ and $f_{t+\tau|r'}(\hat{\beta}(t))$ around β_0, we can write

$$\hat{S}_P - \tilde{S}_P = FP^{-1/2}\sum_{t=R}^{T}(\hat{\beta}(t) - \hat{\beta}_t) + o_p(1).$$

The result follows by showing that $P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}(t) - \hat{\beta}_t) = o_p(1)$. Using the definitions of $\hat{\beta}_t$ and $\hat{\beta}(t)$, we can write

$$P^{-1/2}\sum_{t=R}^{T}(\hat{\beta}(t) - \hat{\beta}_t) = \sum_{i=1}^{3}\mathcal{C}_i,$$

where $\mathcal{C}_1 = P^{-1/2}\sum_{t=R}^{T}\left(\hat{B}(t) - B(t)\right)H(t)$, $\mathcal{C}_2 = P^{-1/2}\sum_{t=R}^{T}B(t)\left(\hat{H}(t) - H(t)\right)$, and $\mathcal{C}_3 = P^{-1/2}\sum_{t=R}^{T}\left(\hat{B}(t) - B(t)\right)\left(\hat{H}(t) - H(t)\right)$. Next, we show $\mathcal{C}_i = o_p(1)$ for $i = 1, 2, 3$. Starting with $\mathcal{C}_1$,

$$\mathcal{C}_1 \leq \sup_t \left| \hat{B}(t) - B(t) \right| P^{-1/2} \sum_{t=R}^{T} |H(t)|,$$

where $\sup_t \left| \hat{B}(t) - B(t) \right| = o_p(1)$ by Lemma A.1 (a) and $P^{-1/2} \sum_{t=R}^{T} |H(t)| = O_p(1)$ by the proof of West's (1996) Lemma A.4 (c). Next, adding and subtracting appropriately,

$$\mathcal{C}_2 = P^{-1/2} \sum_{t=R}^{T} B \left(\hat{H}(t) - H(t) \right) + P^{-1/2} \sum_{t=R}^{T} (B(t) - B) \left(\hat{H}(t) - H(t) \right).$$

It follows that

$$\mathcal{C}_2 \leq BP^{-1/2} \sum_{t=R}^{T} \left(\hat{H}(t) - H(t) \right) + \sup_t |B(t) - B| \, P^{-1/2} \sum_{t=R}^{T} |\hat{H}(t) - H(t)|$$

where $P^{-1/2} \sum_{t=R}^{T} |\hat{H}(t) - H(t)| = o_p(1)$ by Lemma A.1 (b), and $\sup_t |B(t) - B| = o_p(1)$ by Assumption 2 (a). We can show that $\mathcal{C}_3 = o_p(1)$ by a similar argument, completing the proof. ■

Proof of Lemma 4.2. Part (a) follows from eq. (A-1.c) in Lemma A5 in West (1996), whereas part (b) follows from West's (1996) Lemma A2 (a). ■

Proof of Lemma 5.1. This result is obtained by taking the difference of two second-order mean value expansions. The first expansion expands $f^*_{t+\tau|r'}(\hat{\beta}^*_t)$ around β_0, whereas the second expansion expands $f_{t+\tau|r'}(\bar{\beta}_t)$ around β_0, where $\bar{\beta}_t \equiv t^{-1} R \hat{\beta}_R + t^{-1}(t-R)\hat{\beta}_P$, with $\hat{\beta}_R \equiv (R^{-1} \sum_{s=1+\tau}^{R} x_{s-\tau} x'_{s-\tau})^{-1} R^{-1} \sum_{s=1+\tau}^{R} x_{s-\tau} y_s$ and $\hat{\beta}_P \equiv (P^{-1} \sum_{s=R+\tau}^{T+\tau} x_{s-\tau} x'_{s-\tau})^{-1} P^{-1} \sum_{s=R+\tau}^{T+\tau} x_{s-\tau} y_s$.

More specifically, we have that

$$P^{-1/2} \sum_{t=R}^{T} f^*_{t+\tau|r'} \left(\hat{\beta}^*_t \right) = P^{-1/2} \sum_{t=R}^{T} f^*_{t+\tau|r'} + \xi^*_1 + \xi^*_2$$

where

$$\xi^*_1 \equiv P^{-1/2} \sum_{t=R}^{T} f^*_{t+\tau|r',\beta} (\hat{\beta}^*_t - \beta_0) \quad \text{and} \quad \xi^*_2 \equiv 0.5 P^{-1/2} \sum_{t=R}^{T} \frac{\partial^2}{\partial \beta^2} f^*_{t+\tau|r'}(\tilde{\beta}^*_t)(\hat{\beta}^*_t - \beta_0)^2,$$

where $\tilde{\beta}^*_t$ lies between $\hat{\beta}^*_t$ and β_0, and we recall that $f^*_{t+\tau|r'} \equiv f^*_{t+\tau|r'}(\beta_0)$ and $f^*_{t+\tau|r',\beta} \equiv f^*_{t+\tau|r',\beta}(\beta_0)$. To show $\xi^*_2 = o^*_p(1)$, note that

$$|\xi^*_2| \leq 0.5 \left(\sup_t \left| P^{1/4} \left(\hat{\beta}^*_t - \beta_0 \right) \right| \right)^2 P^{-1} \sum_{t=R}^{T} \left| \frac{\partial^2}{\partial \beta^2} f^*_{t+\tau|r'} \left(\tilde{\beta}^*_t \right) \right|.$$

The result follows by Lemma A.2 (c) and the fact that we can show that $P^{-1} \sum_{t=R}^{T} | \frac{\partial^2}{\partial \beta^2} f^*_{t+\tau|r'} \left(\tilde{\beta}^*_t \right) | = O^*_p(1)$, as we argue next. By Assumption 1, and the fact that $\tilde{\beta}^*_t \xrightarrow{P^*} \beta_0$, we can bound $| \frac{\partial^2}{\partial \beta^2} f^*_{t+\tau|r'} \left(\tilde{\beta}^*_t \right) |$ by $\sup_{\beta \in N} | \frac{\partial^2}{\partial \beta^2} f_{\eta_{t+\tau}|r'}(\beta) | \leq m_{\eta_{t+\tau}} \equiv m^*_{t+\tau}$. The result follows by Markov's inequality,

$$P^* \left(P^{-1} \sum_{t=R}^{T} \left| \frac{\partial^2}{\partial \beta^2} f^*_{t+\tau|r'}(\tilde{\beta}^*_t) \right| > \delta \right) \leq P^* \left(P^{-1} \sum_{t=R}^{T} m^*_{t+\tau} > \delta \right) \leq \delta^{-1} P^{-1} \sum_{t=R}^{T} E^*(m^*_{t+\tau}),$$

since $P^{-1}\sum_{t=R}^{T}E^{*}(m_{t+\tau}^{*}) = O_p(1)$ by the properties of the MBB expectation. For ξ_1^*, adding and subtracting appropriately yields

$$\xi_1^* = P^{-1/2}\sum_{t=R}^{T} f_{t+\tau|r',\beta}^{*}\left(\hat{\beta}_t^* - \beta_0\right) = \sum_{i=1}^{4}\xi_{1.i}^{*},$$

where

$$\xi_{1.1}^* = FBP^{-1/2}\sum_{t=R}^{T} H^*(t), \quad \xi_{1.2}^* = P^{-1/2}\sum_{t=R}^{T}(f_{t+\tau|r',\beta}^{*} - F)BH^*(t)$$
$$\xi_{1.3}^* = P^{-1/2}\sum_{t=R}^{T} F(B^*(t) - B)H^*(t), \text{ and } \xi_{1.4}^* = P^{-1/2}\sum_{t=R}^{T}(f_{t+\tau|r',\beta}^{*} - F)(B^*(t) - B)H^*(t).$$

By Lemma A.2 (d),(e) and (f), $\xi_{1,i}^* = o_p^*(1)$ for $i = 2, 3, 4$, respectively. Hence,

$$P^{-1/2}\sum_{t=R}^{T} f_{t+\tau|r'}^{*}(\hat{\beta}_t^*) = P^{-1/2}\sum_{t=R}^{T} f_{t+\tau|r'}^{*} + FBP^{-1/2}\sum_{t=R}^{T} H^*(t) + o_p^*(1). \tag{4}$$

Similarly, an expansion of $f_{t+\tau|r'}(\bar{\beta}_t)$ around β_0 yields

$$P^{-1/2}\sum_{t=R}^{T} f_{t+\tau|r'}(\bar{\beta}_t) = P^{-1/2}\sum_{t=R}^{T} f_{t+\tau|r'} + \bar{\xi}_1 + \bar{\xi}_2,$$

where

$$\bar{\xi}_1 = P^{-1/2}\sum_{t=R}^{T} f_{t+\tau|r',\beta}(\bar{\beta}_t - \beta_0) \text{ and } \bar{\xi}_2 = 0.5P^{-1/2}\sum_{t=R}^{T}\frac{\partial^2}{\partial\beta^2} f_{t+\tau|r'}(\ddot{\beta}_t)(\bar{\beta}_t - \beta_0)^2.$$

where $\ddot{\beta}_t$ lies between $\bar{\beta}_t$ and β_0, and $f_{t+\tau|r',\beta} \equiv f_{t+\tau|r',\beta}(\beta_0)$. We can show that $\bar{\xi}_2 = o_p(1)$ using a similar argument to that used to show that $\xi_2^* = o_p^*(1)$. In particular, it suffices to show that $\sup_t \left|P^{1/4}\left(\bar{\beta}_t - \beta_0\right)\right| = o_p(1)$ and $P^{-1}\sum_{t=R}^{T}\left|\frac{\partial^2}{\partial\beta^2} f_{t+\tau|r'}\left(\ddot{\beta}_t\right)\right| = O_p(1)$. For $\bar{\xi}_1$, note that by definition we can write $\bar{\beta}_t - \beta_0 = \frac{R}{t}(\hat{\beta}_R - \beta_0) + \frac{t-R}{t}(\hat{\beta}_P - \beta_0)$, where

$$\hat{\beta}_R - \beta_0 = B(R)H(R),$$

with $B(R) \equiv \left(R^{-1}\sum_{s=1+\tau}^{R} x_s x_s'\right)^{-1}$, $H(R) = R^{-1}\sum_{s=1+\tau}^{R} h_s$, using our previous definitions of $B(t)$ and $H(t)$. Similarly, given the definition of $\hat{\beta}_P$, we can write

$$\hat{\beta}_P - \beta_0 = \underbrace{\left(P^{-1}\sum_{s=R+\tau}^{T+\tau} x_s x_s'\right)^{-1}}_{\equiv B_P}\underbrace{P^{-1}\sum_{s=R+\tau}^{T+\tau} h_s}_{\equiv H_P = \bar{h}_P}.$$

With this notation, we have that

$$\bar{\xi}_1 = P^{-1/2}\sum_{t=R}^{T} f_{t+\tau|r',\beta}\left(\frac{R}{t}B(R)H(R) + \frac{t-R}{t}B_P H_P\right).$$

Adding and subtracting appropriately, we rewrite $\bar{\xi}_1$ as $\bar{\xi}_1 = \sum_{i=1}^{8} \bar{\xi}_{1.i}$ where

$$\bar{\xi}_{1.1} = FBP^{-1/2}\sum_{t=R}^{T}\frac{R}{t}H(R) \ , \ \bar{\xi}_{1.2} = FBP^{-1/2}\sum_{t=R}^{T}\frac{t-R}{t}H_P$$

$$\bar{\xi}_{1.3} = FP^{-1/2}\sum_{t=R}^{T}\frac{R}{t}\big(B(R)-B\big)H(R) \ , \ \bar{\xi}_{1.4} = FP^{-1/2}\sum_{t=R}^{T}\frac{t-R}{t}\big(B_P-B\big)H_P$$

$$\bar{\xi}_{1.5} = P^{-1/2}\sum_{t=R}^{T}\frac{R}{t}\big(f_{t+\tau|r',\beta}-F\big)BH(R) \ , \ \bar{\xi}_{1.6} = P^{-1/2}\sum_{t=R}^{T}\frac{t-R}{t}\big(f_{t+\tau|r',\beta}-F\big)BH_P$$

$$\bar{\xi}_{1.7} = P^{-1/2}\sum_{t=R}^{T}\frac{R}{t}\big(f_{t+\tau|r',\beta}-F\big)\big(B(R)-B\big)H(R) \ , \ \bar{\xi}_{1.8} = P^{-1/2}\sum_{t=R}^{T}\frac{t-R}{t}\big(f_{t+\tau|r',\beta}-F\big)\big(B_P-B\big)H_P.$$

We can show that $\bar{\xi}_{1.3}$, $\bar{\xi}_{1.5}$ and $\bar{\xi}_{1.7}$ are $o_p(1)$ by applying arguments similar to those in West (1996) (cf. his Lemma A.4). Similar proofs show that $\bar{\xi}_{1.4}$, $\bar{\xi}_{1.6}$ and $\bar{\xi}_{1.8}$ are also $o_p(1)$. Hence, we obtain that $\bar{\xi}_1 = \bar{\xi}_{1.1} + \bar{\xi}_{1.2} + o_p(1)$ and

$$P^{-1/2}\sum_{t=R}^{T} f_{t+\tau|r'}(\bar{\beta}_t) = P^{-1/2}\sum_{t=R}^{T} f_{t+\tau|r'} + \bar{\xi}_{1.1} + \bar{\xi}_{1.2} + o_p(1). \tag{5}$$

Subtracting (4) from (5) yields

$$\tilde{S}_P^* \equiv P^{-1/2}\sum_{t=R}^{T}\Big(f^*_{t+\tau|r'}(\hat{\beta}^*_t) - f_{t+\tau|r'}(\bar{\beta}_t)\Big) = P^{-1/2}\sum_{t=R}^{T}\Big(f^*_{t+\tau|r'} - f_{t+\tau|r'}\Big) + \big(\xi^*_{1.1} - \bar{\xi}_{1.1} - \bar{\xi}_{1.2}\big) + o_p(1),$$

where

$$P^{-1/2}\sum_{t=R}^{T}\Big(f^*_{t+\tau|r'} - f_{t+\tau|r'}\Big) = S^*_{1P},$$

and $\xi^*_{1.1} - \bar{\xi}_{1.1} - \bar{\xi}_{1.2} = FBS^*_{2P}$, as we show next. In particular, using arguments similar to those of Lemma A.5 of West (1996), we have that

$$\xi^*_{1.1} \equiv FBP^{-1/2}\sum_{t=R}^{T} H^*(t) = FBa_{R,0}P^{-1/2}\sum_{s=1+\tau}^{R} h^*_s + FBP^{-1/2}\sum_{i=1}^{P-1} a_{R,i}h^*_{R+i},$$

where $a_{R,i} \equiv (R+i)^{-1} + \ldots + (R+P-1)^{-1}$ for $0 \le i \le P-1$. Moreover, we can rewrite $\bar{\xi}_{1.1}$ as follows,

$$\bar{\xi}_{1.1} = FBP^{-1/2}\underbrace{\left(\sum_{t=R}^{T} t^{-1}\right)}_{\equiv a_{R,0}} R\underbrace{H(R)}_{\equiv \bar{h}_R} = \frac{R}{R-\tau}FBP^{-1/2}a_{R,0}(R-\tau)\,\bar{h}_R = FBa_{R,0}P^{-1/2}\sum_{s=1+\tau}^{R}\bar{h}_R + o_p(1),$$

where the second equality holds by $H(R) \equiv \bar{h}_R$ and the fact that $a_{R,0} = \sum_{t=R}^{T} t^{-1}$, and the third equality follows from $R/(R-\tau) \to 1$. Lastly, we can rewrite

$$\bar{\xi}_{1.2} = FBP^{-1/2}\left(\sum_{t=R}^{T}\frac{t-R}{t}\right)H_P = FBP^{-1/2}\underbrace{\left(\frac{1}{R+1} + \ldots + \frac{P-1}{T}\right)}_{=\sum_{i=1}^{P-1} a_{R,i}}\bar{h}_P = FBP^{-1/2}\left(\sum_{i=1}^{P-1} a_{R,i}\right)\bar{h}_P.$$

Hence,

$$\xi_{1.1}^* - \bar{\xi}_{1.1} - \bar{\xi}_{1.2} = FBa_{R,0}P^{-1/2}\sum_{s=1+\tau}^{R}(h_s^* - \bar{h}_R) + FBP^{-1/2}\sum_{i=1}^{P-1}a_{R,i}(h_{R+i}^* - \bar{h}_P) \equiv FBS_{2P}^*.$$

■

Theorem 5.1. Theorem 5.1 follows from Polya's theorem (e.g., Serfling, 1980 Chapter 1.5.3 page 18) if $\Omega^{-1/2}\tilde{S}_P^{\mu} \xrightarrow{d} N(0,1)$ and $\Omega^{-1/2}\tilde{S}_P^* \xrightarrow{d^*} N(0,1)$. Using the expansion in Lemma 4.1, we get $\tilde{S}_P^{\mu} \xrightarrow{d} N(0,\Omega)$ by Theorem 4.1 of West (1996). We are left to show that $\tilde{S}_P^* \xrightarrow{d^*} N(0,\Omega)$. Recall that $\tilde{S}_{1P}^* = P^{-1/2}\sum_{t=R+1}^{T+\tau}(f_{t|r'}^* - E^*f_{t|r'}^*)$, $\tilde{S}_{2P.1}^* = a_{R,0}P^{-1/2}\sum_{s=1+\tau}^{R}(h_s^* - E^*h_s^*)$, and $\tilde{S}_{2P.2}^* = P^{-1/2}\sum_{t=R+1}^{T+\tau}c_t(h_t^* - E^*h_t^*)$ with

$$c_{R+i} = \begin{cases} a_{R,i} \text{ for } 1 \le i \le P-1 \\ 0 \text{ if } P \le i \le P-1+\tau. \end{cases}$$

Using Lemma 5.1 and the results in Lemma A.4, we can write,

$$\begin{aligned}\tilde{S}_P^* &= \tilde{S}_{1P}^* + FB\left(\tilde{S}_{2P.1}^* + \tilde{S}_{2P.2}^*\right) + o_p^*(1) \\ &= \underbrace{(\tilde{S}_{1P}^* + FB\tilde{S}_{2P.2}^*)}_{\equiv \mathcal{K}_1^*} + \underbrace{FB\tilde{S}_{2P.1}^*}_{\equiv \mathcal{K}_2^*} + o_p^*(1).\end{aligned}$$

This alternative representation of $\tilde{S}_P^*$ enable us to separate the whole bootstrap statistic into two independent, zero bootstrap mean, bootstrap statistics $\mathcal{K}_1^*$ and $\mathcal{K}_2^*$. The independence of these two terms can be seen from bootstrap algorithm in section 5. Recall that $\tilde{S}_{2P.1}^*$ is based on random indexes $I_1,\ldots,I_{k_1}$ whereas $\tilde{S}_{1P}^*$ and $\tilde{S}_{2P.2}^*$ are both based on random indexes $J_1,\ldots,J_{k_2}$, which are independent from $I_1,\ldots,I_{k_1}$. Next we show $\mathcal{K}_1^*$ and $\mathcal{K}_2^*$ both converge to normal distribution with zero mean and appropriate limiting variances. Without loss of generality, we let $\tau = 1$ for the rest of the proof.

For $\mathcal{K}_2^*$, we write

$$\mathcal{K}_2^* = FBa_{R,0}\left(\frac{P}{R-1}\right)^{-1/2}(R-1)^{-1/2}\sum_{s=2}^{R}(h_s^* - E^*h_s^*)$$

where $\left(Var\left((R-1)^{-1/2}\sum_{s=2}^{R}h_s\right)\right)^{-1/2}(R-1)^{-1/2}\sum_{s=2}^{R}(h_s^* - E^*h_s^*) \xrightarrow{d^*} N(0,1)$ by Corollary 3.1 of Fitzenberger (1998). Using the result of $Var^*(\tilde{S}_{2P.1}^*)$ in Lemma A.4 (b), we can write

$$Var^*\left(\mathcal{K}_2^*\right) \xrightarrow{p} F^2B^2\Omega_{2.1}.$$

where the definition of $\Omega_{2.1}$ can be found in Lemma 4.2. Hence, $\mathcal{K}_2^*$ converge to normal distribution with zero mean and variance $F^2B^2\Omega_{2.1}$.

For $\mathcal{K}_1^*$, we first simplify the notation by letting $m_i = R + 1 + (i-1)l$ then write $\mathcal{K}_1^*$ as sum of independent block sums, i.e., $\mathcal{K}_1^* = P^{1/2}k_2^{-1}\sum_{i=1}^{k_2}\tilde{U}_{m_i}^*$ where

$$\tilde{U}_{m_i}^* = l^{-1}\sum_{j=1}^{l}\left(\tilde{f}_{m_i+(j-1)|r'}^* + FBc_{m_i+(j-1)}\tilde{h}_{m_i+(j-1)}^*\right)$$

with $\tilde{f}_{m_i+(j-1)|r'}^* = f_{m_i+(j-1)|r'}^* - E^*f_{m_i+(j-1)|r'}^*$ and $\tilde{h}_{m_i+(j-1)}^* = h_{m_i+(j-1)}^* - E^*h_{m_i+(j-1)}^*$. Using results in Lemma A.4, we can write

$$Var^*(\mathcal{K}_1^*) \xrightarrow{p} \underbrace{\Omega_1 + F^2B^2\Omega_{2,2} + 2FB\Omega_{12}}_{\equiv\Sigma}$$

where the definition of Ω_1, $\Omega_{2,2}$, and Ω_{12} can be found in equation (1) and Lemma 4.2 respectively. Then $\Sigma^{-1/2}\mathcal{K}_1^* = \sum_{i=1}^{k}\tilde{Z}_{m_i}$ where $\tilde{Z}_{m_i} = \Sigma^{-1/2}P^{-1/2}l\tilde{U}_{m_i}^*$. Noting that $\{\tilde{Z}_{m_1}^*, \ldots, \tilde{Z}_{m_{k_2}}^*\}$ is a zero mean, independent heterogeneous sequence which satisfies a CLT for independent heterogeneous sequence (see, 23.6 Lindeberg's Theorem and 23.11 Liapunov's Theorem in Davidson 1994). We verify Liapunov's condition using arguments similar to those in Goncalves and White (2002) (see equation (A.5) on p. 1384). ■

References

[1] Ang, A., Bekaert, G., and M. Wei (2007), "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?," *Journal of Monetary Economics*, 54, 1163-1212.

[2] Aruoba, S.B. (2008), "Data Revisions Are Not Well-Behaved," *Journal of Money, Credit, and Banking*, 40, 319-340.

[3] Calhoun, G. (2015), "A Simple Block Bootstrap for Asymptotically Normal Out-of-Sample Test Statistics," Iowa State University working paper.

[4] Chauvet, M., and J. Piger (2008), "A Comparison of the Real-Time Performance of Business Cycle Dating Methods," *Journal of Business and Economic Statistics*, 26, 42-49.

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

[6] Clark, T.E., and M.W. McCracken (2009), "Tests of Equal Predictive Ability with Real-Time Data," *Journal of Business and Economic Statistics* 27, 441-454.

[7] Clements, M.P., and A.B. Galvão (2013), "Real-time Forecasting of Inflation and Output Growth with Autoregressive Models in the Presence of Data Revisions," *Journal of Applied Econometrics*, 28, 458-477.

[8] Clements, M.P., and A.B. Galvão (2019), "Data Revisions and Real-Time Forecasting," in *Oxford Research Encyclopedia of Economics and Finance*, Oxford University Press.

[9] Corradi, V., Fosten, J., and Gutknecht, D. (2024), "Predictive Ability Tests with Possibly Overlapping Models," *Journal of Econometrics*, forthcoming.

[10] Corradi, V., and N. Swanson (2003), "The Block Bootstrap for Parameter Estimation Error in Recursive Estimation Schemes, with Applications to Predictive Inference Evaluation", Working paper.

[11] Corradi, V., and N. Swanson (2007), "Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes", *International Economic Review*, 48:67–109, 2007.

[12] Corradi, V., Swanson, N. R., and Olivetti, C. (2001), "Predictive Ability with Cointegrated Variables," *Journal of Econometrics*, 104(2), 315-358.

[13] Croushore, D. (2006), "Forecasting with Real–Time Macroeconomic Data," in *Handbook of Economic Forecasting* , eds. G. Elliott, C. Granger, and A. Timmermann, Amsterdam: North-Holland, pp. 961-982.

[14] Croushore, D. (2011), "Frontiers of Real–Time Data Analysis," *Journal of Economic Literature*, 49, 72-100.

[15] Croushore, D. and T. Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, 105, 111-130.

[16] Davidson, J. (1994), *Stochastic Limit Theory*. Oxford: Oxford University Press.

[17] Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business Economic Statistics*, 20(1), 134-144.

[18] Faust, J., Rogers, J.H., and Wright, J.H. (2005), "News and Noise in G-7 GDP Announcements," *Journal of Money, Credit, and Banking*, 37, 403-420.

[19] Faust, J., and Wright, J.H. (2009), "Comparing Greenbook and Reduced Form Forecasts using a Large Real-time Dataset " *Journal of Business and Economic Statistics*, 27, 468-479.

[20] Fitzenberger, B. (1998), "The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions, " *Journal of Econometrics*, 82, 235-287.

[21] Garratt, A., Mitchell, J., Vahey, S.P., and Wakerly, E.C.. (2011), "Real-time Inflation Forecast Densities from Ensemble Phillips Curves," *The North American Journal of Economics and Finance*, 14, 15-27.

[22] Giacomini, R., and B. Rossi (2010), "Forecast Comparisons in Unstable Environments," *Journal of Applied Econometrics* 25, 595-620.

[23] Gonçalves, S. and H. White (2002), "The Bootstrap of the Mean for Dependent Heterogeneous Arrays", *Econometric Theory*, 18, 1367-1384.

[24] Gonçalves, S. and T. Vogelsang (2011), "Block Bootstrap Puzzles in HAC Robust Testing: the Sophistication of the Naive Bootstrap", *Econometric Theory*, 27, 745-791.

[25] Koenig, E.F., Dolmas, S., and J. Piger (2003), "The Use and Abuse of Real-Time Data in Economic Forecasting," *The Review of Economics and Statistics*, 85, 618-628.

[26] Künsch, H. R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations", *Annals of Statistics*, 17, 1217-1241.

[27] Liu, R. Y. and K. Singh (1992), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence". In: LePage, R., Billiard, L. (Eds.), *Exploring the Limits of the Bootstrap*, Wiley, New York, 224-248.

[28] Mankiw, N.G., Runkle, D.E., and Shapiro, M.D. (1984), "Are Preliminary Announcements of the Money Stock Rational Forecasts?" *Journal of Monetary Economics*, 14, 15-27.

[29] Odendahl, F., Rossi, B., and T. Sekhposyan (2022), "Evaluating Forecast Performance with State Dependence," *Journal of Econometrics*, forthcoming.

[30] Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley.

[31] Stark, T., and Croushore, D. (2002), "Forecasting with a Real-Time Data Set for Macroeconomists," *Journal of Macroeconomics*, 24, 507-531.

[32] Stock, J.H. and M. Watson (1999), "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293-335.

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