Forecast Evaluation with Factor-Augmented Models
by Jack Fosten*

*SCHOOL OF ECONOMICS, UNIVERSITY OF EAST ANGLIA

Abstract
This paper provides an extension of Diebold-Mariano-West (DMW) forecast accuracy tests to allow for factor-augmented models to be compared with non-nested benchmarks. The out-of-sample approach to forecast evaluation requires that both the factors and the forecasting model parameters are estimated in a rolling fashion, which poses several new challenges which we address in this paper. Firstly, we show the convergence rates of factors estimated in different rolling windows, and then give conditions under which the asymptotic distribution of the DMW test statistic is not affected by factor estimation error. Secondly, we draw attention to the issue of “sign-changing” across rolling windows of factor estimates and factor-augmented model coefficients, caused by the lack of sign identification when using Principal Components Analysis to estimate the factors. We show that arbitrary sign-changing does not affect the distribution of the DMW test statistic, but it does prohibit the construction of valid bootstrap critical values using existing procedures. We solve this problem by proposing a novel new normalization for rolling factor estimates, which has the effect of matching the sign of factors estimated in different rolling windows. We establish the first-order validity of a simple-to-implement block bootstrap procedure and illustrate its properties using Monte Carlo simulations and an empirical application to forecasting U.S. CPI inflation.

JEL classification codes
C12, C22, C38, C53

Keywords
Bootstrap, Diffusion Index, Factor Model, Predictive Ability.
Forecast Evaluation with Factor-Augmented Models

Jack Fosten∗
University of East Anglia, UK.

January 28, 2016

Abstract

This paper provides an extension of Diebold-Mariano-West (DMW) forecast accuracy tests to allow for factor-augmented models to be compared with non-nested benchmarks. The out-of-sample approach to forecast evaluation requires that both the factors and the forecasting model parameters are estimated in a rolling fashion, which poses several new challenges which we address in this paper. Firstly, we show the convergence rates of factors estimated in different rolling windows, and then give conditions under which the asymptotic distribution of the DMW test statistic is not affected by factor estimation error. Secondly, we draw attention to the issue of “sign-changing” across rolling windows of factor estimates and factor-augmented model coefficients, caused by the lack of sign identification when using Principal Components Analysis to estimate the factors. We show that arbitrary sign-changing does not affect the distribution of the DMW test statistic, but it does prohibit the construction of valid bootstrap critical values using existing procedures. We solve this problem by proposing a novel new normalization for rolling factor estimates, which has the effect of matching the sign of factors estimated in different rolling windows. We establish the first-order validity of a simple-to-implement block bootstrap procedure and illustrate its properties using Monte Carlo simulations and an empirical application to forecasting U.S. CPI inflation.

JEL Classification: C12, C22, C38, C53

Keywords: Bootstrap, Diffusion Index, Factor Model, Predictive Ability.

1 Introduction

This paper considers a number of unresolved challenges which arise when comparing the out-of-sample accuracy of factor-augmented models to a wide variety of competing models using the test procedure of Diebold and Mariano (1995) and West (1996). Since the factor-augmented or “diffusion index” model was proposed by Stock and Watson (2002a,b), there has been a growing amount of literature on the estimation and properties of factor models; comprehensively surveyed by Bai and Ng (2008) and Stock and Watson (2011). There has also been substantial empirical interest in comparing the forecast performance of factor-augmented models to competing models. Recent examples include Castle et al. (2013) and Kim and Swanson (2014). However, to the best of our knowledge, there has been no research formally addressing the effect of factor estimation on

∗Correspondence to: School of Economics, University of East Anglia, Norwich, NR4 7TJ, UK. E-mail address: j.fosten@uea.ac.uk. I would like to thank Valentia Corradi, Federico Martellosio, Vasco Gabriel, Jean-Yves Pitarikis, Gianni Amisano, Domenico Giannone, Michael Clements and Peter Phillips for their discussion and comments on this version and earlier versions of the paper. Also I would like to thank participants at the 4th (2013) Carlo Giannini PhD Workshop in Econometrics at EIEF, Rome, the 2014 Southampton-Surrey Econometrics Event at the University of Southampton, and the 10th (2015) RES PhD Meetings at University College London.
Diebold-Mariano-West (DMW) type tests of equal out-of-sample predictive ability between non-nested models. The fundamental difference in this set-up compared to the standard framework of West (1996) is that out-of-sample forecast comparisons require the estimation of both the factor model and the factor-augmented forecasting model in a rolling or recursive estimation scheme. This brings about several new problems which we identify and solve in this paper.

The first main issue is that an increasing amount of empirical studies of the accuracy of factor-augmented models treat the estimated factors as if they were observed variables even though, as remarked by Grover and McCracken (2014): “neither the results in Diebold and Mariano (1995) nor those in West (1996) are directly applicable to situations where generated regressors are used for prediction.” The literature on predictive ability testing for non-nested models has remained largely silent on the effect factor estimation has on DMW type test statistics. Further results are therefore required which address the properties of factors estimated in different windows, particularly when used in constructing test statistics along the lines of West (1996).

Our first contribution is to provide results on the convergence of factors estimated by Principal Components Analysis (PCA) in different rolling windows. This extends existing results such as such as Stock and Watson (2002a), Bai (2003) and Bai and Ng (2002, 2006), where factors are estimated only once over the full sample, to the case of rolling estimation. We then give conditions under which factor estimation error does not contribute to the asymptotic distribution of the DMW test statistic. This implies that the distribution of West (1996), which accounts for cases where parameter estimation error is non-negligible, is unaffected by factor estimation and the usual critical values can be obtained. Independently to this paper, Gonçalves et al. (2015) have recently provided similar results for tests involving nested model comparisons and, as such, their results are strongly complementary to our results for non-nested comparisons.

The second main issue we address, which tends to be overlooked in the literature, is the issue of sign-identification of PCA estimates of the factors. It is well-known that PCA is unable to identify the sign of the ‘true’ factors or the factor-augmented model parameters. This is of no consequence when estimating the factors only once over the full sample as in the studies of Stock and Watson (2002a,b) and subsequent works. However, when using rolling or recursive estimation, we show that sequences of rolling factor-augmented model parameter estimates are subject to “sign-changing,” which is unavoidable in empirical applications. We show that sign-changing has consequences, not for the asymptotic distribution of the DMW test, but in constructing valid bootstrap critical values for this test. We show that existing bootstrap methods for DMW tests, such as Corradi and Swanson (2006), cannot be directly applied unless the issue of sign-changing is eliminated.

We therefore propose a method to solve the sign-changing issue using a novel new normalization for rolling factor estimates. This method adjusts the standard PCA estimates by using a preliminary estimate of the factor loadings from the first rolling window as a normalization for a subset of the loadings in all subsequent windows. This has the effect of ‘matching’ the signs in every rolling window to that from the first window, which eliminates sign-changing in an environment with stable factor loadings. Our approach bears similarities to the work on factor model identification.
such as Bai and Ng (2013), who propose a similar normalization to the factor estimates, though not in the context of out-of-sample forecast evaluation.

Our final contribution is to propose a simple-to-implement block bootstrap resampling scheme, applicable to either rolling or recursively estimated factors, and establish the first-order validity of corresponding bootstrap critical values for the DMW test statistic. Our approach extends the existing method of Corradi and Swanson (2006) which is not able to deal with the case of generated regressors. This problem is non-trivial for two reasons. Firstly, rolling estimation of the factors gives rise to a multiplicity of overlapping windows of generated regressors, and it is therefore not obvious which factor estimates to resample. Secondly, we show that standard PCA estimates cannot yield valid block bootstrap critical values due to the sign-changing issue. We show that only the sign-adjusted PCA factor estimates, using the normalization proposed in this paper, should be used in the construction of valid bootstrap critical values. Our methodology not only contributes to the literature of bootstrapping DMW tests, by allowing for generated factors in the approach of Corradi and Swanson (2006), but also to the literature of bootstrapping factor estimates, such as Gonçalves and Perron (2014) and Corradi and Swanson (2014), which are only applicable to full-sample factor estimation.

The results of this paper will open up a wider range of possible types of forecast comparison to empirical forecasters interested in factor-augmented models. The main benefit of our bootstrap approach is in cases where parameter estimation error is non-negligible as standard error calculations can become convoluted, as shown by McCracken (2000). Previous empirical studies avoid this issue by assuming away parameter estimation error, normally by using ordinary least squares (OLS) for model estimation and the mean squared forecast error (MSFE) loss function for evaluation. Our paper allows full generality of forecast comparisons, for example when OLS is used for estimation but evaluation is performed using loss functions such as mean absolute error (MAE) or direction-of-change. The method can be used to evaluate forecasts from the FAVAR model of Bernanke et al. (2005), which has become a popular tool in macroeconometrics and forecasting. It can also be used to compare different types of factor-augmented models, such as in the paper of Boivin and Ng (2006) who compare models using factors from real variables versus factors from nominal or financial variables.

The rest of the paper is organized as follows. Section 2 introduces the factor-augmented model and the construction of the Diebold-Mariano-West test statistic. Section 3 describes in detail the “sign-changing” problem and its consequences, and proposes a new normalization to overcome this. Section 4 outlines the assumptions required and the asymptotic properties of the test statistic. Section 5 discusses the required conditions for resampling rolling and recursive factor estimates shows the first-order validity of block bootstrap critical values. Section 6 provides simulation evidence to evaluate this procedure. Finally, Section 7 provides an empirical illustration forecasting U.S. CPI inflation and Section 8 concludes the paper.
2 Forecast Evaluation Set-up

2.1 Models and Forecast Comparison

In this paper we are interested in comparing the forecast accuracy of the factor-augmented model with a competing benchmark forecasting procedure. The factor-augmented, or “diffusion index” model of Stock and Watson (2002a,b) comprises of two equations: a forecast model for predicting a target variable \( y_{t+h} \) at horizon \( h > 0 \) and a factor-model which approximates a high-dimensional set of \( N \) predictors \( X_t \). The forecasting equation is:

\[
y_{t+h} = F_t' \beta + \epsilon_{1,t+h}
\]  

(1)

where \( F_t \) is an \( r \times 1 \) vector of unobserved factors. Equation (1) could also be specified to include other non-factor variables, such as a set of ‘must-have’ regressors \( W_t \) as in Bai and Ng (2006).\(^1\) We omit these additional regressors to simplify notation. The factor model for \( X_t \) has the following representation:

\[
X_t = \Lambda F_t + u_t
\]  

(2)

where \( \Lambda \) is an \( N \times r \) matrix of factor loadings and \( u_t \) is an \( N \times 1 \) vector of errors which are idiosyncratic to each variable. By combining Equations (1) and (2), there is significant data reduction in predicting \( y_{t+h} \) using \( X_t \) when the number of factors is much smaller than the dimension of \( X_t \), in other words \( r << N \). Stock and Watson (2002a,b) and subsequently Bai and Ng (2006) suggest to estimate the unknown factors by Principal Components Analysis (PCA), which makes the factor-augmented regression feasible.

In order to test the accuracy of the factor-augmented model relative to some other benchmark, we formulate the null hypothesis of equal unconditional predictive ability as in Diebold and Mariano (1995) and West (1996):

\[
H_0 : E[g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})] = 0
\]  

(3)

where \( \epsilon_{1,t+h} \) is the forecast error of the factor-augmented model (Model 1) and \( \epsilon_{2,t+h} \) is the forecast error of the benchmark procedure. This null hypothesis tests equality between the expected forecast error losses \( g(\epsilon_{1,t+h}) \) and \( g(\epsilon_{2,t+h}) \), given some loss function \( g(\cdot) \). This loss function may be differentiable as in West (1996), such as mean squared forecast error (MSFE) where \( g(\epsilon_{t+h}) = \epsilon_{t+h}^2 \), or non-differentiable as in McCracken (2000), such as mean absolute error (MAE) where \( g(\epsilon_{t+h}) = |\epsilon_{t+h}| \). The alternative hypothesis \( H_A \) can be two-sided or one-sided in favour of a particular model.

We can consider various different types of competitor forecasting procedures giving rise to \( \epsilon_{2,t+h} \). One case is when benchmark forecast errors \( \epsilon_{2,t+h} \) are non model-based; a case which was the original purpose of the study of Diebold and Mariano (1995). This situation may arise when comparing forecasts from the factor-augmented model to an external set of forecasts such as those

\(^1\)For example, often \( W_t \) contains lags of the dependent variable, in which case the model would be called a factor-augmented autoregression.
from a Survey of Professional Forecasters. It is increasingly common for studies to use professional forecasters as a benchmark for comparison with factor model-based forecasts; see Banbura et al. (2013) for a recent example. Professional forecast data is available for the United States from institutions such as the Federal Reserve Bank of Philadelphia.

Alternatively, the benchmark forecast could also be model-based as in the case of West (1996). For example, we could specify a competitor model (Model 2) which uses a different vector of explanatory variables $Z_t$:

$$\begin{align*}
y_{t+h} &= Z_t' \gamma + \epsilon_{2,t+h} \\
\end{align*}$$

(4)

In this paper, we maintain the framework of West (1996) which requires that $Z_t$ is not nested within $F_t$. This allows for a wide variety of different forecast comparisons. For example $Z_t$ could be any set of non-factor indicators which can be either within $X_t$ or external to $X_t$. This covers studies such as Stock and Watson (1999, 2009) which compares factor forecasts of CPI inflation to Phillips-curve forecasts using unemployment series. Alternatively $Z_t$ could also contain other (non-nested) factors. For example, if $Z_t$ are factors from a financial dataset and $F_t$ are macroeconomic factors then this framework can be used. This includes papers such as Ludvigson and Ng (2007) who study different factor specifications for the risk-return relation. Other examples are comparisons of real against nominal factors from the same database, such as in Boivin and Ng (2006).

By ruling out nested model comparisons, as studied by Clark and McCracken (2001, 2005), we cannot use this approach to compare a 1-factor against a 2-factor model, or compare an autoregression with a factor-augmented autoregression. The study of nested model comparisons involving factor-augmented models has been addressed by the recent paper of Gonçalves et al. (2015).

### 2.2 Pseudo Out-of-sample Methodology and Test Statistic

To test the null hypothesis in Equation (3) when forecast errors are derived from model-based forecasts, it is very common to use a pseudo out-of-sample forecasting exercise. We have a sample of $T + h$ observations on the observed variables $(y_{t+h}, X_t, Z_t)_{t=1}^T$. The pseudo out-of-sample procedure involves splitting the sample into an ‘in-sample’ section and an ‘out of sample’ section. We can then use the in-sample data to make $P$ forecasts of $y_{t+h}$ at horizons $t = R, ..., T$ based on estimates obtained from the models described in Equations (1) and (4). The total sample size is therefore split into $T = R + P - 1$.

The paper of West (1996), and subsequent works, have suggested three main schemes to estimate the models: rolling, recursive and fixed. The rolling estimation scheme estimates the models holding the estimation window fixed at length $R$ and uses observations $j = t - R + 1, ..., t$ for each $R \leq t \leq T$. The recursive scheme, on the other hand, starts with the beginning of the sample and increases the estimation window one period at a time, using observations $j = 1, ..., t$ for each $R \leq t \leq T$. Since the fixed scheme is not used by practitioners we do not include this in our analysis. Throughout

---

2 Note that even if $Z_t$ is a set of variables from $X_t$, nestedness is prevented by the presence of the idiosyncratic errors in the factor model, $u_t$. In this way, variables in $X_t$ might be strongly correlated with $F_t$ but are never an exact linear combination of $F_t$, which is the case of nestedness.
this paper, many of the results will focus on the rolling scheme where extension to the recursive scheme is trivial. Details for use of the recursive estimation in this context can also be found in Gonçalves et al. (2015).

The main difference in this paper to the framework of West (1996) is that, not only do we need to estimate the parameters $\beta$ and $\gamma$ to obtain forecasts, we also need to estimate the factors which enter Equation (1). In order for this to be truly out-of-sample, empirical studies use rolling or recursive Principal Components Analysis. This means that at forecast origin $t = T$, the rolling factor estimates can be described as the $R \times r$ matrix $\hat{F}(t) = [\hat{F}_{t-R+1}, ..., \hat{F}_t]'$, whereas the recursive estimates are the $t \times r$ matrix $\hat{F}(t) = [\hat{F}_1, ..., \hat{F}_t]'$. The superscript $t$ in both cases denotes that data up to period $t$ is used in estimating the factors. This is an important distinction to make relative to papers such as Bai and Ng (2006) which only estimate factors once over the full sample.

For the rolling estimation case we have the following PCA optimization for each $R \leq t \leq T$:

$$
\left( \hat{F}(t), \hat{\Lambda}(t) \right) = \arg \min_{\Lambda, F} \frac{1}{NR} \sum_{i=1}^{N} \sum_{j=t-R+1}^{t} (X_{ij} - \Lambda_i F_j)^2
$$

subject to the normalization $F(t)'F(t)/R = I_r$ and that $\Lambda'\Lambda/N$ is diagonal. The recursive scheme is dealt with analogously by modifying the window length.\(^3\) The normalization conditions are standard in the literature and provide $r^2$ restrictions required to uniquely fix the factors and loadings, see Bai and Ng (2008). The solution is to set $\hat{F}(t)$, the estimated factors in the rolling window from $t-R+1$ to $t$, as the $r$ normalized eigenvectors corresponding to the $r$ largest eigenvalues of the rolling $R \times R$ matrix $X(t)X(t)'/RN$. The factor estimates are correspondingly normalized:

**Normalization N1:** $\hat{F}(t)'\hat{F}(t)/R = I_r$

This normalization implies that the factor loadings are estimated by $\hat{\Lambda}(t) = X(t)'\hat{F}(t)/R$. Since $\hat{\Lambda}(t)'\hat{\Lambda}(t)/N = V(t)$, where $V(t)$ is the diagonal matrix of the $r$ largest eigenvalues of $X(t)X(t)'/RN$, the second requirement that $\Lambda'\Lambda/N$ is diagonal, is satisfied.\(^4\)

Finally, for each $R \leq t \leq T$ we can estimate $\hat{\beta}_t = \arg \min_{\beta} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( y_{j+h} - \hat{F}_j'(t)\beta \right)^2$ for the factor-augmented model and $\hat{\gamma}_t = \arg \min_{\gamma} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( y_{j+h} - Z_j'\gamma \right)^2$ for the benchmark model. These parameter estimates are used to make model-based forecasts which give rise to the forecast errors $\hat{e}_{1,t+h} = y_{t+h} - \hat{F}_t'(t)\hat{\beta}_t$ and $\hat{e}_{2,t+h} = y_{t+h} - Z_t'\hat{\gamma}_t$ for each $R \leq t \leq T$, which are then used to calculate the (non-scaled)\(^5\) Diebold-Mariano-West test statistic:

\(^3\)Intuitively, this is the same as the rolling procedure but with the window length $t$ instead of $R$ and where data from $1$ up to $t$ is used instead of $t - R + 1$ up to $t$, for each $R \leq t \leq T$.

\(^4\)As noted by Bai and Ng (2008), estimating $\hat{F}(t)$ from $X(t)X(t)'$ using $I_1$ will give the same common component as estimating $\hat{\Lambda}(t)$ from $X(t)'X(t)$ and normalizing such that $\Lambda'\Lambda/N = I_k$ and $F(t)'F(t)/R$ being symmetric.

\(^5\)We show later that, as in West (1996), the variance of this test statistic will depend on whether the recursive or rolling scheme is used, we leave $\hat{S}_P$ unscaled and its variance will be calculated along the lines of West (1996).
\[
\hat{S}_P = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} (g(\hat{\epsilon}_{1,t+h}) - g(\hat{\epsilon}_{2,t+h})) 
\] (6)

This test statistic is used to test the null hypothesis of equal predictive ability in Equation (3). The key difference of this set-up to that of West (1996) is the dependence of the test statistic \( \hat{S}_P \) on estimated factors as well as estimated model coefficients. The rolling estimation of the factors is new to the literature on out-of-sample testing and is what we formally analyse in this paper. In the next section, we draw particular attention to the issue of sign-identification of the factors, and show under which circumstances this is problematic.

3 Rolling Estimation of Factor Models

3.1 Convergence Rates and “Sign Changing”

The out-of-sample approach to forecast evaluation with factor-augmented models, described in the previous section, uses rolling or recursive Principal Components Analysis (PCA) to estimate the factors in different estimation windows. This yields a sequence of generated regressors which has not yet been addressed in the literature on forecast accuracy testing in the context of West (1996). Furthermore, the convergence properties of rolling PCA estimates of the factors have not been provided in the literature,\(^6\) whereas papers such as Stock and Watson (2002a,b), Bai (2003) and Bai and Ng (2006) provide convergence rates for the case where estimation takes place only once over the full sample.

Using rolling PCA as in Equation (5), which re-estimates the factor model in each rolling window, there are multiple different factor estimates in overlapping rolling windows which can be collected into the following matrix:

\[
\hat{F} = \begin{bmatrix}
\hat{F}^{(R)}_1 & \hat{F}^{(R)}_2 & \cdots & \hat{F}^{(R)}_P \\
\hat{F}^{(R+1)}_1 & \hat{F}^{(R+1)}_2 & \cdots & \hat{F}^{(R+1)}_P \\
\vdots & \vdots & \ddots & \vdots \\
\hat{F}^{(T)}_P & \hat{F}^{(T)}_{P+1} & \cdots & \hat{F}^{(T)}_T
\end{bmatrix} \] (7)

Each row of this matrix describes the factors estimated from a single rolling window of the data matrix \( X \), as described in the previous section. We can see from matrix (7) that there is only 1 estimate for the factors at observation date 1, \( \hat{F}^{(R)}_1 \), whereas there are 2 estimates at observation date 2, \( \hat{F}^{(R+1)}_1 \) and \( \hat{F}^{(R+1)}_2 \), and so on.

In extending results on the convergence of factor estimates to the case of rolling estimation, an issue arises to do with the sign identification of the rolling factors, which has not been addressed by the existing literature. It is a well-known result that the standard PCA estimates of the factors

\(^6\)With the exception of the paper of Gonçalves et al. (2015), mentioned above, who provide results required for analysing test statistics for nested model comparisons.
only identify the true factors only up to a change in column sign, and a full-rank \( r \times r \) rotation matrix:

\[
H_{NR}^{(t)} = \hat{V}(t)^{-1} \hat{F}(t)^t \Lambda' \Lambda R \frac{N}{N} \tag{8}
\]

This rotation matrix, unlike that of previous studies using full-sample estimation, is time dependent in that it depends on the rolling window \( R \leq t \leq T \). Since the middle part of the rotation matrix contains the rolling estimate \( \hat{F}(t) \), which comprises of eigenvectors of unit length under Normalization N1, in a similar way to Bai (2003), the matrix \( \hat{F}(t)^t F(t) / R \) only has a unique probability limit \( Q \) up to a change in column sign, as detailed in the Appendix. On the other hand, the matrices \( \hat{V}(t)^{-1} \) and \( \Lambda' \Lambda / N \) have unique probability limits \( V^{-1} \) and \( \Sigma_{\Lambda} \). This gives rise to the following Proposition detailing the convergence of the rolling factor estimates and model coefficients up to a sign-changing rotation matrix:

**Proposition 1a:** “Sign-Changing” Under Normalization N1 and Assumptions 1-8, below:

\[
\sup_t \frac{1}{R} \sum_{j=t-R+1}^{t} \left\| \hat{F}_j - H_t^{\dagger} F_j \right\|^2 = o_p(1) \tag{9}
\]

and therefore:

\[
\sup_t \left\| \hat{\beta}_t - H_t^{\dagger \dagger -1} \beta \right\| = o_p(1) \tag{10}
\]

where \( H_t^{\dagger} = S^{(t)} H^{\dagger} \), \( S^{(t)} = \text{diag}(\pm 1, \ldots, \pm 1) \) is any sign matrix and \( H^{\dagger} = V^{-1} Q \Sigma_{\Lambda} \).

Proposition 1a shows the convergence of the rolling factors and factor-augmented model coefficients up to a rotation matrix which is the same across different rolling windows except for a change in column sign \( S^{(t)} \). The first part of the Proposition presents the analogue of the full-sample factor consistency results of Stock and Watson (2002a,b), Bai (2003) and Bai and Ng (2006) extended to the rolling case. The crucial difference here is that the rotation matrix varies across \( t \) due to the sign-changing issue. In terms of the matrix of factor estimates mentioned above, this implies that factor estimates in different rolling windows such as \( \hat{F}_2^{(R)} \) and \( \hat{F}_2^{(R+1)} \), having been estimated using different windows of data, may differ not just by finite-sample heterogeneity caused by the different data used in estimation, but also by a change in sign.

For the factor-augmented model coefficients, the second part of Proposition 1a similarly shows that the probability limit of the rolling OLS estimator is time-dependent due to the sign matrix \( S^{(t)} \) present in \( H_t^{\dagger} \), even though the true \( \beta \) is stable. The fact that rolling estimators of the true \( \beta \) do not have the same probability limit is something which is not standard in the framework of West (1996) and subsequent papers on predictive ability testing. In the next section we show that, while this is not expected to affect the asymptotics of the test statistic \( \hat{S}_P \), the sign-changing is of concern when we wish to bootstrap this test statistic in the presence of parameter estimation error.

---

7 Assumptions leading to the case \( H^{\dagger} = I_r \) are considered by Bai and Ng (2013), which is carried over to the recursive rotation matrix in Gonçalves et al. (2015) where they simply have that \( H_{t+1} = S^{(t)} \).
3.2 Consequences of “Sign-Changing”

There are a number of reasons why the issue of sign-changing in factor estimates has been largely overlooked in the literature to date. Firstly, starting with the work of Stock and Watson (2002a,b), studies which formally address the properties of PCA factor estimation focus on the case where estimation occurs only once using the full sample of data. In that case, there is no possibility of “sign-changing” as there are not multiple sets of factor estimates to compare. This is why papers such as Bai and Ng (2013) are able to fix the column sign of each of the factors to be +1 without loss of generality.

The second reason why sign-changing has received little attention in the literature is that it is well-known that the rotation of factors does not have any effect on the prediction made from the factor-augmented model. This is because, in products such as the prediction $\hat{\mathbf{F}}(t)'\hat{\mathbf{\beta}}_t$, any rotation of the factors cancels out with the corresponding factor-augmented model coefficients which are consistent up to the inverse of the same rotation matrix, as in Proposition 1a. In turn, this implies that the forecast errors $\hat{\epsilon}_{1,t+h} = y_{t+h} - \hat{\mathbf{F}}(t)'\hat{\mathbf{\beta}}_t$ are unaffected by changing sign in the rolling factor estimates. Therefore, the Diebold-Mariano-West statistic $\hat{S}_P$, shown in Equation (6), is also invariant to any changes in sign in the rolling factor estimates and its asymptotic properties will be the same regardless of the sign given to the factors.

However, we show in a later section that sign-changing in the factor estimates is a problem when it comes to performing bootstrap inference for the DMW test. The reasons for this are explained in full detail in Section 5. Intuitively, the issue is as follows. In following a bootstrap resampling procedure such as Corradi and Swanson (2006), we must resample factor estimates over the full sample from 1 to $T$. The factor estimates from which we resample must be consistent for the ‘true’ factors over the full sample. These resampled factors are then used to generate a bootstrap equivalent to the rolling OLS estimator, denoted $\hat{\mathbf{\beta}}_t^*$ for each $R \leq t \leq T$. However, since $\hat{\mathbf{\beta}}_t^*$ is calculated using factor estimates consistent over the full sample, its asymptotic behaviour is with respect to only a single rotation of the true $\mathbf{\beta}$. On the other hand, the asymptotic behaviour of the original $\hat{\mathbf{\beta}}_t$ is with respect to sign-changing rotations of the true $\mathbf{\beta}$, as shown in Proposition 1a. Therefore, in cases where parameter estimation error (PEE) is non-negligible, it is impossible to show that the bootstrapped $\hat{\mathbf{\beta}}_t^*$ behaves like the rolling OLS estimator $\hat{\mathbf{\beta}}_t$, as these estimators do not have the same rotation matrix. As shown in Section 5, valid bootstrap critical cannot be obtained until the issue of sign-changing has been resolved.

In the next section, we therefore propose a novel new normalization of the factors which removes the issue of sign-changing. This will be of importance in obtaining valid bootstrap critical values for the Diebold-Mariano-West test.

3.3 Normalizing Rolling PCA Estimates: a New Approach

To overcome the problem of sign-changing in the rolling estimation of factor-augmented models, we propose a novel new way to estimate the factors in the out-of-sample framework, which entails using a new normalization rather than Normalization N1 used in standard PCA. Our method uses
the rolling structure of the Principal Components estimates and bears similarities with an identification framework proposed in Bai and Ng (2013), without needing to impose strict identification conditions. To our knowledge there have been no papers looking at the normalization of factors for rolling estimation.

Firstly, partition the $N \times r$ factor loading matrix $\Lambda$ into two sub-matrices $\Lambda_1$ and $\Lambda_2$ of dimension $r \times r$ and $(N-r) \times r$ respectively:

$$\Lambda = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix}$$

The idea is to place normalizing restrictions directly onto the matrix $\Lambda_1$ rather than onto the covariance of the factors, as was the case in Normalization N1. An interesting issue is how to place these restrictions without putting too much structure onto this subset of factor loadings. We propose an approach which uses a preliminary estimate of $\Lambda_1$ from the first rolling window, $\hat{\Lambda}_1^{(R)}$, as a normalization in the remaining rolling windows. For each $R \leq t \leq T$ we propose:

**Normalization N2:** For $\Lambda_1$ of full rank, in each $R \leq t \leq T$ normalize $\Lambda_1 = \hat{\Lambda}_1^{(R)}$ where $\hat{\Lambda}_1^{(R)}$ is the standard PCA estimate of $\Lambda_1$ from the first rolling window.

To elaborate on the partition between $\Lambda_1$ and $\Lambda_2$, since the ordering of the variables in the dataset is irrelevant for factor estimation, the choice of the first $r$ variables is arbitrary. The full rank of $\Lambda_1$ is required for its invertability, and can simply be ensured by re-positioning the variables. In empirical studies this may be done by selecting variables from different groups such as a sample across real, nominal and financial variables.\(^8\)

To implement Normalization N2, in the first rolling window ($t = R$) the factors and loadings are normalized using the standard PCA normalization N1. In all subsequent rolling windows ($t > R$) the $r \times r$ sub-matrix $\Lambda_1$ is normalized to be equal to its standard PCA estimate from the first rolling window. In a similar way to Bai and Ng (2013), this normalization is simple to implement based on a straightforward adjustment to the standard PCA estimates. On the other hand, unlike in Bai and Ng (2013), it does not require us to put strong identifying assumptions onto $\Lambda_1$, and it can also be viewed as truly “out-of-sample” as we have chosen to normalize based on an estimate from the first rolling window, which is available to the researcher in all subsequent rolling windows.

The adjusted estimates under Normalization N2, which we call $\tilde{\Lambda}^{(t)}$ and $\tilde{F}^{(t)}$, are calculated using a simple adjustment of the standard PCA estimates:

$$\tilde{\Lambda}^{(t)} = \hat{\Lambda}^{(t)} \left( \hat{\Lambda}_1^{(R)} \right)^{-1} \hat{\Lambda}_1^{(R)}$$

$$\tilde{F}^{(t)} = \hat{F}^{(t)} \hat{\Lambda}_1^{(t)^t} \left( \hat{\Lambda}_1^{(R)^t} \right)^{-1}$$

This, in turn, gives rise to the adjusted rolling OLS estimates of the factor-augmented model coefficients $\tilde{\beta}_t := \left( \tilde{F}^{(t)^t} \tilde{F}^{(t)} \right)^{-1} \tilde{F}^{(t)^t} y^{(t)}$. The following Proposition provides results for the convergence

\(^8\)It is not advisable to simply choose $\Lambda_1$ corresponding to the first $r$ variables listed in a given dataset. In the case of the Stock and Watson (2002a,b) dataset, for instance, the first variables are all disaggregates of industrial production which all load onto the factors in a similar way and may give a $\Lambda_1$ with some eigenvalues close to zero.
of these adjusted factors and factor-augmented model coefficients, \( \tilde{F}(t) \) and \( \tilde{\beta}_t \):

**Proposition 1b: “Sign-Robustness”** Under Normalization N2 and Assumptions 1-8, below:

\[
\sup_t \frac{1}{R} \sum_{j=t-R+1}^{t} \left\| \tilde{F}^{(t)}_j - H^t_R F_j \right\|^2 = o_p(1) \quad (12)
\]

and therefore:

\[
\sup_t \left\| \tilde{\beta}_t - H^t_R^{-1} \beta \right\| = o_p(1) \quad (13)
\]

where \( H^t_R \) is the limiting rotation matrix from the first rolling window, and is therefore not dependent on \( t \).

The two parts of Proposition 1b show the convergence of the adjusted factors and factor-augmented model coefficients over rolling windows. The critical difference to Proposition 1a is that Normalization N2 ensures that the adjusted factors and factor-augmented model coefficients are consistent about a rotation matrix which does not depend on \( t \), which means that the sign-changing issue is eliminated. In fact, the limiting rotation matrix is equal to that of the first rolling window under standard PCA, in other words \( H^t_R \) for \( t = R \), so our method has the effect of matching the sign of each rolling window to that of the first window. This was due to the choice of normalization we used in N2. Intuitively, in the expression for \( \tilde{F}(t) \) in Equation (11), the standard PCA estimates \( \hat{F}(t) \) and \( \hat{\Lambda}'(t) \) have probability limits \( F(t) H^t S(t)' \) and \( S(t)'^{-1} H^{-1} \Lambda_1' \), as shown by Proposition 1a. Therefore \( S(t) \) is eliminated in the limit since the product of the sign matrix with its own inverse is the identity matrix, and the only sign matrix which remains in the expression is \( S(R) \), the sign from the first rolling window.

To illustrate how sign-changing is unavoidable in empirical studies when using standard PCA, we refer to the two graphs displayed in Figures 1 and 2 below. These use the updated Stock and Watson (2002a,b) dataset. The red line in Figure 1 shows two different rolling window estimates of the first factor corresponding to the first and last half of the data. The black line shows the same factor estimated using the full sample. The change in sign in the red line clearly demonstrates the first part of Proposition 1a of sign-changing in the factors, whereas the blue line does not have this property. In Figure 2, the red line depicts a sequence of rolling estimates of the factor-augmented model coefficient in a simple regression of U.S. CPI inflation. The sign fluctuation as a result of the second part of Proposition 1a is clear from this figure, and is not present in the blue line for the adjusted estimators. In principle one could suggest ad hoc rule-based algorithms to fix the sign of eigenvectors in each rolling window, but these may fail in small samples. Our method, instead, is always robust to sign-changing.

---

9 For a description of the data see Section 7, below.
Figure 1: Graph plotting the rolling factor estimates $\hat{F}(t)$ and $\tilde{F}(t)$ in the first and last rolling windows against the full-sample estimate of the first Stock and Watson factor.

Notes: The vertical dashed line denotes the point $R = T/2$ which separates the two separate rolling window estimates forming the red and blue lines. The black line denotes the factor estimated over the full sample of data.

Figure 2: Graph plotting the values of rolling OLS estimates $\hat{\beta}_t$ and the adjusted $\tilde{\beta}_t$ for the first factor in 1-step ahead factor-augmented regressions of U.S. CPI inflation.

Notes: These results are based on an AR(6) specification with 1 factor with $R = 240$ and $P = 241$. The average of $\hat{\beta}_t$ over the $P$ rolling windows is 0.002 whereas the average of $\tilde{\beta}_t$ is -0.221.
4 Asymptotic Theory

The Diebold-Mariano-West statistic $\hat{S}_p$ displayed in Equation (6) was derived based on the standard Principal Components estimates of the factors under Normalization N1. In the previous section we also showed how to adjust these estimates under Normalization N2 to make them robust to sign changing. We can write a new test statistic $\tilde{S}_p$ which has been constructed using the adjusted factor estimates $\tilde{F}^{(t)}$ rather than $\hat{F}^{(t)}$:

$$\tilde{S}_p = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} (g(\tilde{\epsilon}_{1,t+h}) - g(\tilde{\epsilon}_{2,t+h}))$$

where analogously to above, $\tilde{\epsilon}_{1,t+h} = y_{t+h} - \tilde{F}^{(t)}\tilde{\beta}_t$. If $Z_t$ contains estimated factors, $\tilde{\epsilon}_{2,t+h}$ also uses the adjusted estimates under Normalization N2. However, as mentioned in Section 3.2, we note that the empirical magnitudes of $S_p$ and $\tilde{S}_p$ are the same:

$$\hat{S}_p = \tilde{S}_p$$

since the rotation of the factors always cancels with the equivalent rotation in the factor-augmented model coefficients. These statistics will therefore behave equivalently asymptotically. Since the later bootstrap section focuses on the use of the sign-robust estimates, we will proceed to use the tilde notation, keeping in mind that this does not alter the results in this section.

We now detail the assumptions required to show the asymptotic distribution of $\tilde{S}_p$. For simplicity, we assume that the benchmark model variables $Z_t$ are a set of non-factor variables though this can be relaxed at the cost of further notation. In the assumptions, $C$ denotes a generic constant. For a matrix $M$, $M > 0$ means that $M$ is positive definite, and $\|M\| = (\text{tr}(M'M))^{1/2}$. We adopt the notation of West (1996) that “sup$_t$” is shorthand for “sup$_{R \leq t \leq T}$”. Assumptions 1 to 8 detail what is required:

Rolling Estimation Assumptions

**Assumption 1:** (Model Variables, Forecast Errors and Idiosyncratic Error Processes)

(a) $(F'_t, Z'_t, \epsilon_{1,t+h}, \epsilon_{2,t+h}, u_{1t}, ..., u_{Nt})$ is strong mixing with mixing coefficients of size $-3d/(d-1)$ for some $d > 1$;

(b) $(F'_t, Z'_t, \epsilon_{1,t+h}, \epsilon_{2,t+h}, u_{1t}, ..., u_{Nt})$ is strictly stationary.

**Assumption 2:** (Factors and Loadings)

(a) $E\|F_t\|^4 \leq C$, and $\frac{1}{R} \sum_{j=-R+1}^{t} F_j F'_j \xrightarrow{p} \Sigma_F > 0$ uniformly in $t$ as $R \to \infty$;

(b) The loadings $\lambda_i$ for $i = 1, ..., N$ are either deterministic such that $\|\lambda_i\| \leq C$ or stochastic such that $E\|\lambda_i\|^4 \leq C$. In any case $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\Lambda > 0$;

(c) The eigenvalues of the $r \times r$ matrix $\Sigma_\Lambda \Sigma_F$ are unique.

**Assumption 3:** (Idiosyncratic Error Dependence)

(a) $E(u_{it}) = 0$, $E|u_{it}|^8 \leq C$;
(b) $E\left(\frac{1}{N}\sum_{i=1}^{N} u_{ts} u_{it}\right) = \gamma_{st}, |\gamma_{ss}| \leq C$ for all $s$, and $\sup_t \frac{1}{R} \sum_{j=t-R+1}^{t} \sum_{k=t-R+1}^{t} |\gamma_{jk}| \leq C$ and $\frac{1}{N} \sum_{t=R}^{T} \sum_{k=t-R+1}^{t} |\gamma_{tk}| \leq C$;

(c) For all $(t, s)$, $E\left(\left|N^{-1/2} \sum_{i=1}^{N} u_{it} u_{is} - E (u_{it} u_{is})\right|^{4}\right) \leq C$;

(d) $E(\tau_{ij,t}) = \gamma_{ij,t}$ with $|\gamma_{ij,t}| \leq |\tau_{ij}|$ for all $t$ and $\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |\tau_{ij}| \leq C$;

(e) $E(\tau_{ijk,h}) = \gamma_{ijk,h}$ and $\sup_t \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=t-R+1}^{t} \sum_{h=t-R+1}^{t} |\tau_{ijk,h}| \leq C$.

**Assumption 4:** (Dependence between idiosyncratic errors and loadings, regressions errors and variables)

(a) For all $s$, $E\left(\sup_t \frac{1}{\sqrt{NR}} \sum_{k=t-R+1}^{t} \sum_{i=1}^{N} F_k (u_{is} u_{ik} - E (u_{is} u_{ik}))\right)^2 \leq C$

(b) For all $s$, and $h \geq 0$, $E\left(\sup_t \frac{1}{\sqrt{NR}} \sum_{j=t-R+1}^{t-h} \sum_{i=1}^{N} u_{ij} (u_{ij,t+h} - E (u_{ij} u_{ij,t+h}))\right)^2 \leq C$

(c) $E\left(\sup_t \frac{1}{\sqrt{NR}} \sum_{j=t-R+1}^{t} \sum_{i=1}^{N} u_{ij} (u_{ij,t+h} - E (u_{ij} u_{ij,t+h}))\right)^2 \leq C$, and $E (\sum_{i=1}^{N} E (u_{i(t^+)} e_{1,t+h}^t) = 0$ for all $(i, t)$(d) $E\left(\sup_t \frac{1}{\sqrt{NR}} \sum_{j=t-R+1}^{t} \sum_{i=1}^{N} u_{ij} (u_{ij,t+h} - E (u_{ij} u_{ij,t+h}))\right)^2 \leq C$, and $E (F_t u_{it}) = 0$ for all $(i, t)$(e) $E\left(\sup_t \frac{1}{\sqrt{NR}} \sum_{j=t-R+1}^{t} \sum_{i=1}^{N} u_{ij} (u_{ij,t+h} - E (u_{ij} u_{ij,t+h}))\right)^2 \leq C$, and $E (\sum_{i=1}^{N} E (u_{i(t^+)} e_{2,t+h}^t) = 0$ for all $(i, t)$(f) $E\left(\sup_t \frac{1}{\sqrt{NR}} \sum_{j=t-R+1}^{t} \sum_{i=1}^{N} u_{ij} (u_{ij,t+h} - E (u_{ij} u_{ij,t+h}))\right)^2 \leq C$, and $E (\sum_{i=1}^{N} E (u_{i(t^+)} e_{2,t+h}^t) = 0.$

**Assumption 5:** (Forecast model moments)

(a) $E\left|\epsilon_{1,t+h}\right|^{4d} \leq C$ and $E\left|\epsilon_{2,t+h}\right|^{4d} \leq C$ where $d > 1$;

(b) $E\left\|Z_t\right\|^{4d} \leq C$ and for all $t, \frac{1}{R} \sum_{j=t-R+1}^{t} \sum_{i=1}^{N} 1 \leq C$;

(c) $E (F_t e_{1,t+h}^t) = 0$ and $E (Z_t e_{2,t+h}^t) = 0$.

**Assumption 6:** (Functional form of loss function)

(a1) For the $(2r \times 1)$ vector $\theta = [F_t, \beta]^t$, in an open neighbourhood $N_1$ around $\theta$, and with probability one, the function $g(\epsilon_{1,t+h}) = g(y_{t+h} - F_t^t \beta)$ is measurable and twice continuously differentiable with respect to $\theta$;

(a2) Also in an open neighbourhood $N_2$ around $\gamma$, and with probability one, the function $g(\epsilon_{2,t+h}) = g(y_{t+h} - Z_t^t \gamma)$ is measurable and twice continuously differentiable with respect to $\gamma$;

(b1) For all $t, sup_{\theta \in N_1} \left\|\nabla_{\theta} g(\epsilon_{1,t+h})\right\| \leq m_{1t},$ for a measurable $m_{1t}$ with $E(m_{1t}) \leq C$;

(b2) Also for all $t, sup_{\gamma \in N_2} \left\|\nabla_{\gamma} g(\epsilon_{1,t+h})\right\| \leq m_{2t},$ for a measurable $m_{2t}$ with $E(m_{2t}) \leq C$.

**Assumption 7:** (Test statistic and score assumptions)

(a) $E\left|\left|\nabla_{\beta} g(\epsilon_{1,t+h}), \nabla_{\gamma} g(\epsilon_{2,t+h}), \nabla_{\beta} g(\epsilon_{1,t+h}), \nabla_{\beta} g(\epsilon_{2,t+h}), F_t \epsilon_{1,t+h}, Z_t \epsilon_{2,t+h}\right|\right|^{4d} \leq C,$ where $d > 1$;

(b) Denote $L_{t+h} = g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})$ and $E (L_{t+h})$ its expectation. Furthermore let $V_t = \sum_{j=-\infty}^{\infty} E (L_{t+h} - E (L_{t+h})) (L_{t+h-j} - E (L_{t+h-j}))$. Then $V_t > 0$.

(c) $E\left|\left|\nabla_{\Gamma} g(\epsilon_{1,t+h})\right|\right|^{4d} \leq C$ and $E (D_t F_t = E (\nabla_{\Gamma} g(\epsilon_{1,t+h}))$.

**Assumption 8:** (Asymptotic Rates)

(a) $T, N \to \infty$ such that $\sqrt{T}/N \to 0$;

(b) $P, R \to \infty$ as $T \to \infty$ and $\lim_{T \to \infty} (P/R) = \pi$ with $0 \leq \pi < \infty$.

**Recursive Estimation Assumptions**
If the recursive estimation scheme is used, the required assumptions coincide almost identically with that of rolling estimation and, as mentioned above, are ensured by appropriately replacing partial sums from $t - R + 1$ to $t$ with sums from 1 to $t$. These are therefore not repeated here.\footnote{Assumptions required for recursive estimation are detailed in Gonçalves et al. (2015).}

Assumptions 1-8 are closely related to the assumptions of West (1996) for predictive ability testing and Bai and Ng (2006) for factor-augmented models. However, since rolling estimation is used for the factors, we must modify assumptions on the dependence of the factors, loadings and idiosyncratic errors. These modifications are similar in spirit to that of Corradi and Swanson (2014), but differ because their paper takes rolling averages of full-sample factor estimates whereas we take rolling averages of rolling factor estimates. Assumptions 1 through 5 essentially mirror the assumptions in Bai and Ng (2006) for factor estimation. Assumptions 3(b)-3(c) and Assumption 4 modify these, in a similar way to Corradi and Swanson (2014), to the rolling estimation case. Assumptions 1 and 5 include the additional assumptions required for the benchmark model, and also guarantee all variables are stationary mixing processes, as required by West (1996) and more recently Cheng and Hansen (2015) for the factor-augmented case. Assumptions for the recursive estimation procedure are not repeated here but, in a similar way to above, can be insured by changing rolling averages from observations $t - R + 1, ..., t$ to being recursive averages over 1, ..., $t$, as in Gonçalves et al. (2015).

Assumptions 6 and 7 are concerned with the moments, differentiability and measurability of the loss function $g(\cdot)$ and are the same as those in West (1996). As mentioned before, these conditions can be weakened along the lines of McCracken (2000) to allow for non-differentiable loss functions. For the sake of brevity, we do not re-write these additional assumptions here. Finally, Assumption 8 places assumptions on the relative rate of increase of $N$, $T$, $R$ and $P$, the first of which is standard in factor model studies, the second of which is standard in the out-of-sample predictive ability testing literature.\footnote{For cases where the rolling window length is kept constant, even asymptotically, see the conditional predictive ability approach of Giacomini and White (2006).}

The following result shows the asymptotic distribution of the test statistic $\tilde{S}_P$:

**Theorem 1:** Under Normalization $N2$, Assumptions 1-8 and under $H_0$:

$$\tilde{S}_P \overset{d}{\rightarrow} N(0, \Omega)$$

where:

$$\begin{align*}
\Omega &= V_\epsilon + \theta_1 D_\beta V_F D_\beta' + 2\theta_2 D_\beta C_{\epsilon, F} \\
&\quad + \theta_1 D_\gamma V_Z D_\gamma' - 2\theta_2 D_\gamma C_{\epsilon, Z} - 2\theta_1 D_\beta C_{F;Z} D_\gamma'
\end{align*}$$

where for the rolling estimation scheme with $\pi \geq 1$, $\theta_1 = \left(1 - \frac{1}{3\pi}\right)$ and $\theta_2 = \left(1 - \frac{1}{2\pi}\right)$. For the rolling estimation scheme with $\pi < 1$, $\theta_1 = \left(\pi - \frac{\pi^2}{4}\right)$ and $\theta_2 = \frac{\pi}{2}$. For the recursive scheme,
\[ \theta_1 = 2 \left( 1 - \frac{1}{\pi} \ln (1 + \pi) \right) \text{ and } \theta_2 = 1 - \frac{1}{\pi} \ln (1 + \pi). \] Also:

\[
V_e = \sum_{j=-\infty}^{\infty} E \left[ (g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})) - E (g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})) \right] \times (g(\epsilon_{1,t+h+j}) - g(\epsilon_{2,t+h+j})) - E (g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})) \right]
\]

\[
V_F = \left( H^\dagger \Sigma_F H^\dagger \right)^{-1} H^\dagger \left( \sum_{j=-\infty}^{\infty} E \left[ F_t \epsilon_{1,t+h} \epsilon_{1,t+h+j} F_t \right] \right) H^\dagger \left( H^\dagger \Sigma_F H^\dagger \right)^{-1},
\]

\[
V_Z = \Sigma_Z^{-1} \left( \sum_{j=-\infty}^{\infty} E \left[ Z_t \epsilon_{2,t+h} \epsilon_{2,t+h+j} Z_t \right] \right) \Sigma_Z^{-1},
\]

\[
C_{e,F} = \left( \sum_{j=-\infty}^{\infty} E \left[ ((g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})) \epsilon_{1,t+h+j} F_t \right] \right) H^\dagger \left( H^\dagger \Sigma_F H^\dagger \right)^{-1},
\]

\[
C_{e,Z} = \left( \sum_{j=-\infty}^{\infty} E \left[ (g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})) \epsilon_{2,t+h+j} Z_t \right] \right) \Sigma_Z^{-1},
\]

\[
C_{F,Z} = \left( H^\dagger \Sigma_F H^\dagger \right)^{-1} H^\dagger \left( \sum_{j=-\infty}^{\infty} E \left[ F_t \epsilon_{1,t+h} \epsilon_{2,t+h+j} Z_t \right] \right) \Sigma_Z^{-1},
\]

where the rotation matrix \( H^\dagger \) is described in Proposition 1b, and we define \( D_\beta = E (\nabla \beta g(\epsilon_{1,t+h})) \) and \( D_\gamma = E (\nabla \gamma g(\epsilon_{2,t+h})). \)

This result shows that exactly the same distribution of West (1996) obtains even in the presence of additional factor estimation, under the assumptions outlined earlier. This result is useful as it implies that the same critical values can be used for this test, regardless of the presence of factor estimation error. The variance-covariance matrix \( \Omega \) can be consistently estimated by Newey-West type estimators, as outlined in Comment 6 of West (1996).

The next section discusses methods to construct first-order valid bootstrap critical values for the test statistic \( \hat{S}_p \).

5 Bootstrap Inference

5.1 Bootstrapping Factors Estimated in the Out-of-Sample Method

In this section we provide details of the earlier claim that only the sign-adjusted factor estimates and OLS coefficients, \( \hat{F}^{(t)} \) and \( \hat{\beta}_t \), can be used if we wish to construct valid bootstrap critical values for the test statistic \( \hat{S}_p \). The bootstrap is useful in the forecast accuracy testing framework as Theorem 1 shows that the variance-covariance matrix \( \Omega \) may be made of many parts when parameter estimation error (PEE) is non-negligible. In finite samples, the aggregation of estimates
of these parts may give unreliable standard errors and it may be advisable to use bootstrap inference. However, the bootstrapping of rolling or recursively estimated factors has not been addressed in the literature as far as we are aware. On the other hand, bootstrapping factors estimated over the full sample has been considered by Corradi and Swanson (2014) and Gonçalves and Perron (2014). There are several new challenges to overcome when bootstrapping such factor estimates for the purpose of out-of-sample predictive ability testing.

A block bootstrap procedure for inference on Diebold-Mariano-West type tests when PEE is non-negligible was suggested by Corradi and Swanson (2006). Their approach is to resample model variables over the full sample from 1 to \( T \) and then proceed to perform a pseudo out-of-sample procedure on the resampled data. They show that quantiles of the empirical distribution of a recentered bootstrap test statistic gives rise to valid critical values for DMW-type tests. However, this approach is only valid in the context of non-generated regressors. For example one could directly apply the Corradi and Swanson (2006) approach for Model 2 in Equation (4) in cases where the predictors \( Z_t \) are non-generated regressors; in which case \((y_{t+h}^*, Z_t^*)_T^T\) would be generated as \( T \) resampled observations from the data for \( y_{t+h} \) and \( Z_t \).

In the context of factor estimation presented in this paper, it is not obvious how to proceed with a bootstrap methodology, as we have a multiplicity of overlapping windows of factors generated in the pseudo out-of-sample procedure. Since we assume that factor estimation error is non-negligible, ensured by Assumption 8(a) that \( \sqrt{T/N} \to 0 \) like in Bai and Ng (2006), then we are able to resample the factor estimates directly as in Corradi and Swanson (2014).\(^{12}\) Moreover, since we have presented various different factor estimates throughout the paper, corresponding to factors estimated under the rolling and recursive schemes, and also sign-changing and sign-robust factor estimates, it is useful to have some general conditions required for resampling factor estimates in order to provide valid bootstrap inference in this situation.

Consider first some generic matrix of factor estimates \( F \) which we are to resample in a bootstrap procedure. We first assume the following top-level conditions for \( F \), and then show which type of factor estimates considered in this paper fulfil these assumptions.

**Condition 1:** \( F \) is a matrix of factor estimates of dimension \( T \times r \).

**Condition 2:** The estimates in \( F \) are consistent for the true factors \( F \) over the full sample such that:

\[
\frac{1}{T} \sum_{t=1}^{T} \| F_t - HF_t \|^2 = o_p(1)
\]

where \( H \) is some full-rank \( r \times r \) rotation matrix.

**Condition 3:** For every \( R \leq t \leq T \), the rotation matrix of the resampled factors, \( H \), must match the inverse of the rotation matrix of the estimated factor-augmented model coefficient \( \beta \) from the

\(^{12}\)This is in contrast to Gonçalves and Perron (2014) where they assume that \( \sqrt{T/N} \to c \) and therefore need to mimic factor estimation error as well.

17
out-of-sample procedure: \( \hat{\beta}_t \) in relation to the standard PCA factor estimates, or \( \tilde{\beta}_t \) in the case of the adjusted PCA estimates.

Condition 1 requires that the factor estimates we resample are of dimension \( T \times r \), which is not as trivial a requirement as it seems. As shown in matrix (7), when we use rolling estimation for the factors, there is no single factor estimate over the full sample from 1 to \( T \), but a sequence of overlapping rolling windows from 1 to \( R \), 2 to \( R+1 \) and so on. On the other hand, under recursive estimation, the final estimation window does yield a factor estimate from 1 to \( T \) and so we might resample directly from the full-sample factors in that case.

For the case of rolling estimation, we suggest that \( T \) is assumed to be strictly a multiple of \( R \), for example \( T = 2R \) or \( T = 3R \) and so on. In many empirical studies, sample splits such as \( P \approx R \) or \( P > R \) are used, and so this technical requirement is not of much consequence. This can be ensured with a simple modification to Assumption 8:

**Assumption 8’**: (Asymptotic Rates for Rolling)
(a) Same as Assumption 8(a);
(b’) \( P,R \to \infty \) as \( T \to \infty \) and \( T = KR \) with \( K \) a finite integer such that \( 2 \leq K < \infty \) and therefore \( \lim_{T \to \infty} (P/R) = (K - 1) \).

With this assumption, it is then possible to join together \( K \) rolling windows of length \( R \) to obtain a matrix of dimension \( T \). This gives rise to four suggestions for \( \mathcal{F} \) which satisfy Condition 1. These correspond to the standard PCA estimates under Normalization N1 for both the rolling and recursive schemes, and the adjusted PCA estimates under Normalization N2 for both the rolling and recursive schemes:

\[
\begin{align*}
\hat{\mathcal{F}}^{rol} &= \left[ \left( \hat{F}_1^{(R)}, \ldots, \hat{F}_R^{(R)} \right), \left( \hat{F}_{R+1}^{(2R)}, \ldots, \hat{F}_{2R}^{(2R)} \right), \ldots, \left( \hat{F}_P^{(T)}, \ldots, \hat{F}_T^{(T)} \right) \right]' \quad (16a) \\
\hat{\mathcal{F}}^{rec} &= \left[ \left( \hat{F}_1^{(T)}, \ldots, \hat{F}_T^{(T)} \right) \right]' \quad (16b) \\
\tilde{\mathcal{F}}^{rol} &= \left[ \left( \tilde{F}_1^{(R)}, \ldots, \tilde{F}_R^{(R)} \right), \left( \tilde{F}_{R+1}^{(2R)}, \ldots, \tilde{F}_{2R}^{(2R)} \right), \ldots, \left( \tilde{F}_P^{(T)}, \ldots, \tilde{F}_T^{(T)} \right) \right]' \quad (16c) \\
\tilde{\mathcal{F}}^{rec} &= \left[ \left( \tilde{F}_1^{(T)}, \ldots, \tilde{F}_T^{(T)} \right) \right]' \quad (16d)
\end{align*}
\]

It is clear that the recursive versions, \( \hat{\mathcal{F}}^{rec} \) and \( \tilde{\mathcal{F}}^{rec} \) in Equations (16b) and (16d), are formed only from the final (full-sample) window of the recursive estimation scheme, whereas the rolling versions in Equations (16a) and (16c) are formed of a finite number of \( K \) different rolling windows.

Condition 2 ensures that the resampled estimates are consistent for the true factors over the full sample. This is required as the bootstrap expectation of some resampled \( \mathcal{F}_t^* \) equals the average \( \frac{1}{T} \sum_{t=1}^{T} \mathcal{F}_t \), which in turn can be related back to the true rotated factors under Condition 2, up to asymptotically negligible terms. Turning to the four suggestions above, it is clear that the

\[\text{See, for example, Stock and Watson (2002a) or Kim and Swanson (2014).}\]
recursive versions $\hat{F}_{rec}$ and $\tilde{F}_{rec}$ satisfy Condition 2 as they are nothing other than full-sample factor estimates, which have been shown to be consistent by papers such as Bai and Ng (2006) and recently Gonçalves et al. (2015). Similarly, since Proposition 1b shows that each window of the adjusted PCA estimates is consistent up to the same rotation matrix, $\tilde{F}_{rol}$ will also satisfy Condition 1 as it comprises $K$ windows of factor estimates which have the same rotation matrix. However, as Proposition 1a shows that different rolling windows of the standard PCA factor estimates are subject to sign-changing, $\hat{F}_{rol}$ cannot satisfy Condition 2. This is formalized in the following:

**Proposition 2a:** Under the adjusted PCA Normalization N2 and Rolling Assumptions 1-7 and 8’ for $\tilde{F}_{rol}$ and Recursive Assumptions 1-8 for $\hat{F}_{rec}$:

$$\frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{F}_{t}^{rol} - H_{R}^\dagger R_{t} \right\|^2 = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)$$

and:

$$\frac{1}{T} \sum_{t=1}^{T} \left\| \hat{F}_{t}^{rec} - H_{R}^\dagger R_{t} \right\|^2 = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)$$

Under the standard PCA Normalization N1 and Recursive Assumptions 1-8:

$$\frac{1}{T} \sum_{t=1}^{T} \left\| \hat{F}_{t}^{rec} - H_{R}^\dagger R_{t} \right\|^2 = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)$$

But Condition 2 does not hold for $\hat{F}_{rol}$.

The last line of Proposition 2a demonstrates that, due to the sign changing problem, it is not possible to use $\hat{F}_{rol}$ for bootstrap resampling. This can be seen by returning to Figure 1 where the red line represents the estimator $\hat{F}_{rol}$ with $T = 2R$. This gives graphical intuition that resampling the factors in this way will not be valid due to the sign-changing problem.

Finally, Condition 3 above requires that the factors we resample must have the same rotation as the inverse of the factor-augmented model coefficients used in the pseudo out-of-sample procedure. This condition is required as Corradi and Swanson (2006) show that a recentering term is required to ensure the bootstrap statistic has mean zero. Their bootstrap counterpart of the test statistic $\tilde{S}_{P}$, described in more detail later, has the following form:

$$\tilde{S}_{P}^* = \frac{1}{\sqrt{T}} \sum_{t=R}^{T} \left( g(\tilde{\epsilon}_{1,t+h}) - g(\tilde{\epsilon}_{2,t+h}) \right) - \frac{1}{T} \sum_{j=1}^{T} \left( g(\tilde{\epsilon}_{1,j+h,t}) - g(\tilde{\epsilon}_{2,j+h,t}) \right)$$

where the second part of this expression is the recentering term which is an average over the full sample.\textsuperscript{14} The forecast errors for the factor-augmented Model 1 in this recentering term would be $\tilde{\epsilon}_{1,j+h,t} = y_{j+h} - F_{j}'\tilde{\beta}_{t}$ if the sign-changing $\tilde{\beta}_{t}$ were to be used, or $\tilde{\epsilon}_{1,j+h,t} = y_{j+h} - F_{j}'\hat{\beta}_{t}$ if the

\textsuperscript{14}The bootstrap error terms $\tilde{\epsilon}_{1,t+h}$ and $\tilde{\epsilon}_{2,t+h}$ are described in Equations (19) and (20).
adjusted \( \tilde{\beta}_t \) were to be used. However, it is clear that we need Condition 3, because if \( F \) does not share the same rotation matrix as the estimator for \( \beta \) used for the recentering term, then the rotation matrices will not cancel in the product. This would result in the recentering term inside the bracket being \( O_p(1) \) and, when rescaled by \( \sqrt{T} \), would diverge to infinity. The final Proposition gives the final guidance on how to proceed to resample the estimated factors described above:

**Proposition 2b:** Condition 3 only holds when either \( \tilde{\mathcal{F}}_t^{\text{rec}} \) or \( \tilde{\mathcal{F}}_t^{\text{rol}} \) is used for resampling and \( \tilde{\beta}_t \) is used for the recentering term.

Using Proposition 1b and Proposition 2a, we know that the adjusted estimator \( \tilde{\beta}_t \) is consistent in each rolling window up to \( H_R^{t-1} \) and \( \tilde{\mathcal{F}}_t^{\text{rec}} \) and \( \tilde{\mathcal{F}}_t^{\text{rol}} \) are consistent up to the rotation \( H_R^t \) due to the use of Normalization N2, therefore the first part of Proposition 2b is immediate. However, it is not possible to use \( \tilde{\beta}_t \) for the recentering term as Proposition 1a shows that it has a time-varying rotation matrix and therefore cannot have the same rotation as the resampled factors for each \( R \leq t \leq T \). Finally, Proposition 2b rules out the use of the recursive version \( \tilde{\mathcal{F}}_t^{\text{rec}} \) as this has rotation matrix \( H_R^t \) which does not match \( H_R^t \) due to the column sign.

The next section formally details the procedure for calculating bootstrap critical values.

### 5.2 Bootstrap Critical Values

Under the results in the previous section, we can proceed to resample the full-sample factor estimates \( \tilde{\mathcal{F}}^{\text{rec}} \) or \( \tilde{\mathcal{F}}^{\text{rol}} \) described in Equations (16c) and (16d). The choice of either of these is not relevant asymptotically as we have assumed that factor estimation error is negligible. We therefore drop the superscript “\( \text{rec} \)” or “\( \text{rol} \)” and denote the matrix of estimates \( \tilde{\mathcal{F}} \) for either \( \tilde{\mathcal{F}}^{\text{rec}} \) or \( \tilde{\mathcal{F}}^{\text{rol}} \). We follow the block bootstrap procedure of Corradi and Swanson (2006) to obtain bootstrap critical values. We can resample \( (y_{t+h}, \tilde{\mathcal{F}}_t, Z_t)_{t=1}^T \) using \( b \) blocks of length \( l \) such that \( bl = T \). This is done by drawing an index \( I_j \) from the discrete iid random uniform distribution on the interval \([0, T-l]\) with equal probability where \( j = 1, \ldots, b \) and by forming \( b \) blocks of \( y_{t+h}, \tilde{\mathcal{F}}_t \) and \( Z_t \) such that \[
[y_{t+h}^1, \ldots, y_{T+h}^b, \tilde{\mathcal{F}}_{t+h}, \ldots, \tilde{\mathcal{F}}_{T+h}] = [y_{t+1}, \ldots, y_{T+h}, \ldots, y_{b+1}, \ldots, y_{b+l+h}] , \quad \left[ \tilde{\mathcal{F}}_{t+1}, \ldots, \tilde{\mathcal{F}}_T \right] = [\tilde{\mathcal{F}}_{t+1}, \ldots, \tilde{\mathcal{F}}_{t+h}, \ldots, \tilde{\mathcal{F}}_{T+h}] \quad \text{and} \quad [Z_{t+1}^1, \ldots, Z_{T+h}^b] = [Z_{t+1}, \ldots, Z_{T+h}, \ldots, Z_{b+1}, \ldots, Z_{b+l+h}].
\]

We proceed by applying the rolling out-of-sample methodology on this resampled data. Bootstrap estimation of the parameters \( \beta \) and \( \gamma \) proceeds by recentering the OLS criterion function around the score over the full sample as in Corradi and Swanson (2006):

\[
\tilde{\beta}_t^* = \arg\min_{\beta} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( y_{j+h}^* - \tilde{\mathcal{F}}_{j}^* \beta \right)^2 + 2\beta' \left( \frac{1}{T} \sum_{j'=1}^{T} \tilde{\mathcal{F}}_{j'}^* \tilde{c}_{1,j'+h,t} \right)
\]

\[
= \left( \frac{1}{R} \sum_{j=t-R+1}^{t-h} \tilde{\mathcal{F}}_{j}^* \tilde{F}_{j}^* \right)^{-1} \left( \frac{1}{R} \sum_{j=t-R+1}^{t-h} \tilde{\mathcal{F}}_{j}^* y_{j+h}^* - \left( \frac{1}{T} \sum_{j'=1}^{T} \tilde{\mathcal{F}}_{j'}^* \tilde{c}_{1,j'+h,t} \right) \right)
\]
and:
\[ \tilde{\gamma}_t^* = \arg \min_{\gamma} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( \left( y_{j+h}^* - Z_j^* \tilde{\gamma}_j \right)^2 + 2\gamma \left( \sum_{j=1}^{T} Z_j \tilde{\epsilon}_{2,j+h,t} \right) \right) \]
\[ = \left( \frac{1}{R} \sum_{j=t-R+1}^{t-h} Z_j^* Z_j^* \right)^{-1} \left( \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( Z_j^* y_{j+h}^* - \left( \frac{1}{T} \sum_{j'=1}^{T} Z_j \tilde{\epsilon}_{2,j+h,t} \right) \right) \right) \] (18)

for all \( R \leq t \leq T \), where \( \tilde{\epsilon}_{1,j+h,t} = y_{j+h} - \tilde{\mathcal{F}}_t \tilde{\beta}_t \) and \( \tilde{\epsilon}_{2,j+h,t} = y_{j+h} - Z_j \tilde{\gamma}_t \) are the error terms of Models 1 and 2 evaluated at the rolling OLS estimators \( \tilde{\beta}_t \) and \( \tilde{\gamma}_t \) over the full sample \( j' = 1, \ldots, T \). The recentering corrects for the fact that resampling takes place over the full sample, and ensures that the bootstrap first-order conditions are equal to zero. As mentioned above, the use of \( \tilde{\beta}_t \) rather than \( \hat{\beta}_t \) is crucial for ensuring that the recentering term is well-behaved.

The bootstrap estimators \( \tilde{\beta}_t^* \) and \( \tilde{\gamma}_t^* \) are used in constructing the bootstrap forecast errors:
\[ \tilde{\epsilon}_{1,t+h} = y_{t+h} - \tilde{\mathcal{F}}_t \tilde{\beta}_t^* \] (19)
and:
\[ \tilde{\epsilon}_{2,t+h} = y_{t+h} - Z_t \tilde{\gamma}_t^* \] (20)

Finally, the bootstrap counterpart of the test statistic \( \tilde{S}_P \) is then given by:
\[ \tilde{S}_P^* = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g(\tilde{\epsilon}_{1,t+h}) - g(\tilde{\epsilon}_{2,t+h}) \right) - \frac{1}{T} \sum_{j=1}^{T} \left( g(\tilde{\epsilon}_{1,j+h,t}) - g(\tilde{\epsilon}_{2,j+h,t}) \right) \] (21)

This test statistic is also recentered around the full sample as in Corradi and Swanson (2006). The main difference in our paper is that the choice of estimator for \( \beta \) used in the recentering term is non-trivial.

The first-order validity of this bootstrap resampling procedure is presented in the following Theorem:

**Theorem 2:** Under Normalization N2, Assumptions 1-7, Assumption 8', and assuming that \( l, b \to \infty \) such that \( l/T^{1/4} \to 0 \). Then, under \( H_0 \):
\[ P \left( \omega : \sup_{s \in \mathbb{R}} \left| \Pr^* \left( \tilde{S}_P^* \leq s \right) - \Pr \left( \tilde{S}_P \leq s \right) \right| > \epsilon \right) \to 0 \]

Using this result we can generate \( B \) bootstrap replications of the test statistic \( \tilde{S}_P^* \) and calculate the \( \alpha \) and \( (1 - \alpha) \) percentiles of its empirical distribution. Rejection or non-rejection of the null hypothesis will be based on comparing the test statistic \( \tilde{S}_P \) to these percentiles of the empirical distribution of \( \tilde{S}_P^* \).
6 Monte Carlo Simulation

In this section we analyse the finite sample properties of the bootstrap critical values in testing the null of equal out-of-sample predictive ability. We take the case of rolling estimation, and as such we use the matrix \( \tilde{F}^{rol} \) to resample the factors, as described above in Equation (16c). The loss functions we consider for \( g(\cdot) \) are mean squared forecast error (MSFE) and mean absolute error (MAE):

\[
\text{MSFE}(\epsilon_{t+h}) = \epsilon_{t+h}^2 \\
\text{MAE}(\epsilon_{t+h}) = |\epsilon_{t+h}|
\]

We consider model estimation using OLS, in which case parameter estimation error is negligible for the MSFE test, as shown by West (1996), so \( \Omega = V \) in that case.\(^{15}\) On the other hand, when \( g(\cdot) \) is MAE, McCracken (2000) shows that, under a slight modification to Assumption 6, parameter estimation error contributes to \( \Omega \) in the same way as in West (1996). Moreover, its functional form is complex and cumbersome to compute.

Due to the computational burden of this problem, we will apply the warp-speed bootstrap method of Giacomini et al. (2013) which only uses one bootstrap draw per Monte Carlo draw. The MSFE and MAE tests are based on the forecasts generated from the two linear regressions:

\[
y_{j+1} = \mu_1 + \beta \tilde{F}^{(t)}_{j} + \epsilon_{1,j+1} \quad y_{j+1} = \mu_2 + \gamma Z_j + \epsilon_{2,j+1}
\]  

(22)

for each \( R \leq t \leq T \) and where \( j = t - R + 1, ... , t \) as rolling estimation is used. Model 1 is the feasible factor-augmented model where \( \tilde{F}^{(t)} \) is the rolling adjusted PCA factor estimate described in Section 3. We assume \( Z_t \) to be a non-factor regressor. There is one constant and one regressor in each model. We choose a data generating process (DGP) for the factor model similar to that in Stock and Watson (2002a):

\[
X_{it} = \lambda_i F_t + \sqrt{\theta} u_{it} \\
F_t = \rho F_{t-1} + e_t \\
u_{it} = \rho_u u_{i,t-1} + v_{it}
\]

For simplicity we specify a single factor corresponding to \( r = 1 \). The loadings are drawn as independent \( N(1,1) \) variates, and the processes \( e_t \) and \( u_{it} \) are drawn independently from \( N(0,1-\rho^2_F) \) and \( N(0,1-\rho^2_u) \) so that \( F_t \) and \( u_{it} \) have unit unconditional variance. We set \( \theta = r \) which fixes the signal to noise ratio in the factor model to 1, following previous studies.

For the forecast variable DGP we use the following representation which is similar to that of

\(^{15}\)Note that, as explained in Corradi and Swanson (2006), with negligible PEE we do not need to use the recentered estimator \( \hat{\beta}^*_t \) to construct the bootstrap forecast errors.
McCracken (2000):

\[ y_{t+1} = Z_t + (1 + c) F_t + \epsilon_{t+1} \]  
\[ Z_t = \rho Z_{t-1} + w_t \]  
\[ \epsilon_t = \rho \epsilon_{t-1} + \eta_t \]  

Equation (23)

In a similar way to above, \( w_t \) and \( \eta_t \) are drawn independently from \( N(0, 1 - \rho_F^2) \) and \( N(0, 1 - \rho_Z^2) \) so that all variables have unit unconditional variance. The initial conditions \( F_0, u_0, Z_0 \) and \( \epsilon_0 \) are all drawn from their stationary distributions.

The key parameter \( c \) is used to move between the null and alternative. When \( c = 0 \) the variables have equal weight in the DGP for \( y_{t+1} \) in Equation (23), so the null of equal MSFE and MAE are both satisfied. In this case, both Models 1 and 2 in Equation (22) have \( R^2 \) equal to 0.33. When \( c > 0 \) the alternative holds and the factor-augmented model has lower MSFE and MAE. We will vary \( c \) between 0 and 0.5 thereby allowing the assessment of size and power of the test. Therefore when \( c = 0.5 \) the \( R^2 \) of models 1 and 2 are roughly 0.53 and 0.24. This approach is similar in nature to that of McCracken (2000) though we do not consider as extreme deviations from the null. Most other studies using Monte Carlo simulations focus on nested model comparisons, see Busetti and Marcucci (2013) for a comprehensive study.

The sample sizes we consider for the rolling window length are \( R = \{120, 240\} \) and for the evaluation period we let the parameter \( K \) in Assumption 8b’ to be \( K = \{2, 3\} \). This corresponds to \( \pi = \{1, 2\} \) and \( T \) ranges between 240 and 720. We use a panel dimension of \( N = 200 \) as a medium between small- and large-scale factor model studies seen in the literature. We select the persistence parameters \( \rho_F, \rho_u, \rho_Z \) and \( \rho_e \) to be equal to 0.5 for all of the AR(1) processes.

Due to the computational burden of this problem we use the warp-speed bootstrap of Giacomini et al. (2013) with \( M = 999 \) Monte Carlo replications and \( B = 1 \) bootstrap draw per replication. Since we do not have any optimality results governing optimal block length, we compare the results for the values \( l = \{3, 6\} \).

Table 1 displays the empirical size for the bootstrap tests of equal out-of-sample MAE and MSFE for a nominal size equal to 0.1. We also provide the results for the basic Diebold-Mariano-West test using the non-adjusted standard normal critical values by way of comparison as these are often reported in the literature. The results indicate the the bootstrap test has good size properties for all configurations we consider, particularly for the test of equal MAE. On the other hand, the standard normal Diebold-Mariano test is found to be oversized in all cases. Using bootstrap critical values abates this problem considerably under MAE loss, though there is still some moderate oversizing in the test for equal MSFE when smaller sample sizes are considered. The results for the different block lengths are very similar, though the results for \( l = 6 \) are marginally better than those for \( l = 3 \).

Since the bootstrap test with \( l = 6 \) is almost correctly sized, we only present the power results for this configuration.
Table 1: Empirical size for MSFE and MAE tests for equal out-of-sample predictive ability conducted at the nominal 10% level.

<table>
<thead>
<tr>
<th>MAE Loss</th>
<th>$R = 120$</th>
<th>$R = 240$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\pi = 1$</td>
<td>$\pi = 2$</td>
</tr>
<tr>
<td>Standard Normal</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Bootstrap $(l = 3)$</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>Bootstrap $(l = 6)$</td>
<td>0.09</td>
<td>0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MSFE Loss</th>
<th>$R = 120$</th>
<th>$R = 240$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\pi = 1$</td>
<td>$\pi = 2$</td>
</tr>
<tr>
<td>Standard Normal</td>
<td>0.17</td>
<td>0.15</td>
</tr>
<tr>
<td>Bootstrap $(l = 3)$</td>
<td>0.15</td>
<td>0.11</td>
</tr>
<tr>
<td>Bootstrap $(l = 6)$</td>
<td>0.13</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Notes: Based on $M = 999$ Monte Carlo replications. Warp-speed bootstrap uses $B = 1$ bootstrap draw per Monte Carlo replication. See description in text.

for this test in Table 2. These results show that the bootstrap test has good power properties, with larger power for the larger values of $\pi$. For the most extreme deviation from the null we consider, $c = 0.5$, the test approaches unit power for the larger sample sizes. It is worth re-iterating that we have chosen fairly subtle deviations from the null in these specifications. If we had instead chosen the most extreme deviation from the null to be of similar magnitude to that of, say, McCracken (2000) we would expect the results to display power even closer to unity.

The next section presents an empirical illustration of this test before the paper concludes.

7 Empirical Illustration

Stock and Watson (1999) pose the empirical question of whether high-dimensional factor based forecasts can improve on simple Phillips curve models in forecasting inflation using unemployment. This problem has subsequently been analysed by a large number of papers, which are surveyed by Stock and Watson (2009), and more recently by Hansen et al. (2011). In this empirical illustration we offer a re-examination of this question. The factor-augmented model is written:

$$\pi_{t+h} = \phi_1 + \beta_1 (L) F_t + \mu_1 (L) \pi_t + \epsilon_{t+h}$$  (24)
Table 2: Power of the bootstrap test \((l = 6)\) for MSFE and MAE tests for equal out-of-sample predictive ability conducted at the nominal 10% level.

<table>
<thead>
<tr>
<th></th>
<th>(R = 120)</th>
<th>(R = 240)</th>
<th>(R = 120)</th>
<th>(R = 240)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(c = 1)</td>
<td>(c = 2)</td>
<td>(c = 1)</td>
<td>(c = 2)</td>
</tr>
<tr>
<td>(0.1)</td>
<td>0.18</td>
<td>0.25</td>
<td>0.21</td>
<td>0.34</td>
</tr>
<tr>
<td>(0.2)</td>
<td>0.38</td>
<td>0.48</td>
<td>0.42</td>
<td>0.58</td>
</tr>
<tr>
<td>(0.3)</td>
<td>0.52</td>
<td>0.72</td>
<td>0.62</td>
<td>0.83</td>
</tr>
<tr>
<td>(0.4)</td>
<td>0.70</td>
<td>0.85</td>
<td>0.83</td>
<td>0.94</td>
</tr>
<tr>
<td>(0.5)</td>
<td>0.83</td>
<td>0.94</td>
<td>0.92</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Notes: These results present the power of the bootstrap test with higher values of \(c\) denoting higher divergence from the null. Based on \(M = 999\) Monte Carlo replications. Warp-speed bootstrap uses \(B = 1\) bootstrap draw per Monte Carlo replication. See description in text.

and the Phillips curve benchmark is:

\[
\pi_{t+h}^h = \phi_2 + \beta_2 (L) u_t + \mu_2 (L) \pi_t + \epsilon_{t+h} \tag{25}
\]

where \(u_t\) is the rate of unemployment and, following Stock and Watson (1999), the inflation variable is the annualized cumulative growth of the consumer price index (CPI) \(\pi_t^h = (1200/h) \ln (P_t/P_{t-h})\) with the autoregressive terms \(\pi_t^1 \equiv \pi_t = 1200 \ln (P_t/P_{t-1}).^{17}\) These models are clearly non-nested when \(\beta_1 (L) \neq 0\) and \(\beta_2 (L) \neq 0\).

For evaluating the forecasts from these models, as in the Monte Carlo section we compare the results of the MSFE loss function to that of MAE. The data we use is that of Stock and Watson (2002a,b) updated by Kim and Swanson (2014)\(^{18}\) which contains 144 macroeconomic and financial variables. Evidence of Stock and Watson (2009) and Breitung and Eickmeier (2011) suggests the presence of large structural breaks in the factor loadings around 1984, corresponding to the date identified as the “Great Moderation.” Given that the present method is valid only with stable

\(^{17}\)We also ran results where we had an \(I(1)\) specification for inflation. These results are available on request.

\(^{18}\)We thank these authors for kindly providing us with their data.
loadings, after transformation to stationarity and lagging the explanatory variables $h$ times for the direct forecasting scheme, we retain a sample size $T = 300$ over the time period 1984:06 to 2009:05. With a relatively small time series sample we choose $K = 2$ which implies that $R = 150$ and $P = T - R + 1 = 151$ and gives an equal split of observations for estimation and prediction. We will consider the 1-, 3- and 12-month forecast horizons.

Since we still find evidence of instability in factor loadings after the first factor, even after splitting the sample, we use a 1-factor forecasting model. We select the number of autoregressive lags and lags of the explanatory variables corresponding to $\mu_1(L)$, $\mu_2(L)$, $\beta_1(L)$ and $\beta_2(L)$ in Equations (25) and (24) using the BIC. This takes place over the first rolling window so that the number of variables in the model is held fixed over the evaluation period.

Table 3: Statistical comparison of forecasts of the U.S. CPI inflation rate from the factor-augmented model against a Phillips curve benchmark. Test statistics, bootstrap critical values and $p$-values for the test of equal predictive ability.

<table>
<thead>
<tr>
<th></th>
<th>MAE Loss</th>
<th>MSFE Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$h = 1$</td>
<td>$h = 3$</td>
</tr>
<tr>
<td>Relative Loss</td>
<td>0.9982</td>
<td>0.9782</td>
</tr>
<tr>
<td>Statistic</td>
<td>-0.0546</td>
<td>-0.2844</td>
</tr>
<tr>
<td>$l = 3$</td>
<td>5%</td>
<td>-0.2523</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>-0.1821</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>0.2581</td>
</tr>
<tr>
<td>$p$-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2587</td>
<td>0.0652</td>
</tr>
<tr>
<td>$l = 6$</td>
<td>5%</td>
<td>-0.3756</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>-0.2095</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>0.2025</td>
</tr>
<tr>
<td>$p$-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3609</td>
<td>0.1103</td>
</tr>
<tr>
<td>$l = 12$</td>
<td>5%</td>
<td>-0.4335</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>-0.2461</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>0.1453</td>
</tr>
<tr>
<td>$p$-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4511</td>
<td>0.1754</td>
</tr>
<tr>
<td>Standard Normal</td>
<td>0.4314</td>
<td>0.3521</td>
</tr>
</tbody>
</table>

Notes: The row entitled relative loss is the ratio of forecast error loss from the factor-augmented model to the Phillips curve benchmark. The row entitled Statistic presents the actual test statistic $S_P$ for forecast horizons $h = 1$, $h = 3$ and $h = 12$. For block lengths $l = 3, 6, 12$, the 5th, 10th and 50th percentiles of the bootstrap empirical distribution are presented for $B = 399$, with 2-sided symmetric bootstrap $p$-values. For comparison, the final row presents the standard Normal $p$-value of Diebold and Mariano (1995), with standard error estimated using the rectangular kernel with $h - 1$ lags.

For the block bootstrap implementation we use $B = 399$ bootstrap draws and different values for
the block length $l = \{3, 6, 12\}$ meaning a number of blocks equal to $b = \{100, 50, 25\}$ respectively. Table 3 documents the results, displaying the two-sided symmetric bootstrap $p$-values testing the null hypothesis of equal predictive ability. It can be seen that the results do not depend on this choice of $l$.

The results across the two different loss functions are qualitatively similar. At the shorter horizons $h = 1$ and $h = 3$, the relative error losses for both MAE and MSFE are very close to one, with less than a 2% difference between the factor augmented model and the Phillips curve model. The evidence from the bootstrap critical values, and indeed the standard normal Diebold-Mariano critical values indicates no evidence to reject the null of equal predictive ability.

However, at the 12-month horizon we see larger differences between the predictive ability of the two models. The factor augmented model has around 15% better predictive ability under MAE loss and almost 30% better under MSFE loss. The bootstrap critical values indicate enough evidence to reject the null, whereas a test based on standard normal critical values finds no such evidence. This shows that empirical papers basing their decisions only on the standard normal critical values may in some instances make different conclusions to when using bootstrap critical values. We conclude that, in the post-1984 period, factors have had superior predictive ability at predicting 12-month ahead cumulative inflation growth relative to the Phillips curve benchmark, but not at shorter horizons.

8 Conclusion

This paper provides solutions to several problems posed by extending out-of-sample predictive ability tests of Diebold and Mariano (1995) and West (1996) to allow for factor-augmented models. This is an important problem as the rising interest in factor-augmented models has not yet been matched with a formal treatment of forecast accuracy tests involving estimated factors. The first part of this paper provides new results on the properties of estimated factors under a rolling estimation scheme, used to construct the forecast errors for the DMW test statistic. We provide convergence rates for the rolling factor estimates which extend existing results in the literature for full-sample factor estimation. We then show the conditions under which factor estimation error does not have any effect on the asymptotic distribution of the DMW test statistic.

We also address an important issue which has been overlooked in the literature, to do with “sign-changing” in the estimated factors and factor-augmented model parameters across different rolling windows. This results from Principal Components estimation which does not identify the sign of the true factors. We show that, while this has no impact on the asymptotic distribution of the DMW test statistic, it does prohibit the use of existing methods for calculating bootstrap critical values for this test. We propose a novel new normalization to the factors which corrects for sign-changing and allows us to establish the first-order validity of bootstrap critical values using a block bootstrap method similar to that of Corradi and Swanson (2006).

The paper concludes with simulation evidence and an empirical application to demonstrate
the use of the bootstrap procedure. We compare forecasts of the U.S. CPI inflation rate from a factor-augmented model to a Phillips curve benchmark, discovering that inference based on the non-adjusted standard normal critical values gives different findings compared to the valid bootstrap critical values. The results in this paper can be built upon in future work, both in empirical applications of factor-augmented models, and methodological extensions of DMW-type tests to allow for estimated factors.

9 Appendix

The proofs of Proposition 1a and 1b, Propositions 2a and 2b and Theorems 1 and 2 from the text are provided here. Additional Lemmas are also required but the detailed proofs are consigned to an Online Appendix.

9.1 Proof of Proposition 1a and 1b

The proof of Propositions 1a and 1b rely on the following Lemmas:

Lemma A: Under Normalization N1 and Assumptions 1-8:

(i)

\[
\sup_{t} \frac{1}{R} \sum_{j=t-R+1}^{t} \left\| \hat{F}_j^{(t)} - H_{NR}^{(t)} F_j \right\|^2 = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)
\]

(ii)

\[
\sup_{t} \frac{1}{R} \sum_{j=t-R+1}^{t} \left( \hat{F}_j^{(t)} - H_{NR}^{(t)} F_j \right) u_{ij} = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)
\]

(iii)

\[
\sup_{t} \frac{1}{R} \sum_{j=t-R+1}^{t} \left( \hat{F}_j^{(t)} - H_{NR}^{(t)} F_j \right) F_j = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)
\]

(iv)

\[
\sup_{t} \left( \hat{\Lambda}^{(t)} - \Lambda H_{NR}^{(t)-1} \right) = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{\sqrt{R}} \right\} \right)
\]

(v)

\[
\sup_{t} \frac{1}{R} \sum_{j=t-R+1}^{t} \left( \hat{F}_j^{(t)} - H_{NR}^{(t)} F_j \right) \epsilon_{j+h} = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)
\]

where \( H_{NR}^{(t)} \) is described in Equation (8).
Lemma A extends existing results on full-sample factor estimation error in the standard Principal Components estimates $\hat{F}(t)$ to the case of rolling estimation. Respectively, Lemmas A(i)-(iii) are the rolling analogues of Bai and Ng (2002) Theorem 1 and Bai (2003) Lemma B.1 and Lemma B.2. Lemma A(iv) comes as a result of parts (i)-(iii) and is the analogue of Bai (2003) Theorem 2 which shows the rolling factor loadings are consistent at a rate $\min\{\sqrt{R}, N\}$. Finally, Lemma A(v) extends Bai and Ng (2006) Lemma A.1 (iv) to the case of rolling estimation. The proofs of these are fully provided in a separate Appendix.

**Proof of Proposition 1a:**
We start by reformulating the expression in Proposition 1a in terms of the rotation matrix $H^{(t)}_{NR}$:

$$\frac{1}{R} \sum_{j=t-R+1}^{t} \| \hat{F}_{j}^{(t)} - H_{NR}^{(t)} F_j \|^2$$

$$= \frac{1}{R} \sum_{j=t-R+1}^{t} \left\| \left( \hat{F}_{j}^{(t)} - H_{NR}^{(t)} F_j \right) + \left( H_{NR}^{(t)} - H_{i}^{(t)} \right) F_j \right\|^2$$

$$\leq \frac{2}{R} \sum_{j=t-R+1}^{t} \left\| \left( \hat{F}_{j}^{(t)} - H_{NR}^{(t)} F_j \right) \right\|^2 + \frac{2}{R} \sum_{j=t-R+1}^{t} \left\| \left( H_{NR}^{(t)} - H_{i}^{(t)} \right) F_j \right\|^2$$

$$\leq \frac{2}{R} \sum_{j=t-R+1}^{t} \left\| \left( \hat{F}_{j}^{(t)} - H_{NR}^{(t)} F_j \right) \right\|^2 + \left\| H_{NR}^{(t)} - H_{i}^{(t)} \right\|^2 \frac{2}{R} \sum_{j=t-R+1}^{t} \| F_j \|^2 \quad \text{(26)}$$

where:

$$H_{NR}^{(t)} = \hat{V}^{(t)} - \frac{\hat{F}^{(t)} F(t)}{R} - \frac{\Lambda^{(t)}}{N}$$

is the rotation matrix of Equation (8) in the text, $H_{i}^{(t)} = S^{(t)} H^{(t)}$ and $H^{(t)} = V^{-1} Q \Sigma \Lambda$ and $S^{(t)} = \text{diag}(\pm 1, \ldots, \pm 1)$ is any sign matrix. Now, it follows from Equation (26) that:

$$\sup_{t} \frac{1}{R} \sum_{j=t-R+1}^{t} \left\| \hat{F}_{j}^{(t)} - H_{i}^{(t)} F_j \right\|^2 \leq \sup_{t} \frac{2}{R} \sum_{j=t-R+1}^{t} \left\| \left( \hat{F}_{j}^{(t)} - H_{NR}^{(t)} F_j \right) \right\|^2$$

$$+ \sup_{t} \left\| H_{NR}^{(t)} - H_{i}^{(t)} \right\|^2 \sup_{t} \frac{2}{R} \sum_{j=t-R+1}^{t} \| F_j \|^2$$

and using Lemma A(i), which states that:

$$\sup_{t} \frac{1}{R} \sum_{j=t-R+1}^{t} \left\| \hat{F}_{j}^{(t)} - H_{NR}^{(t)} F_j \right\|^2 = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)$$

we can combine these two expressions to get:

$$\sup_{t} \frac{1}{R} \sum_{j=t-R+1}^{t} \left\| \hat{F}_{j}^{(t)} - H_{i}^{(t)} F_j \right\|^2 \leq \sup_{t} \left\| H_{NR}^{(t)} - H_{i}^{(t)} \right\|^2 \sup_{t} \frac{2}{R} \sum_{j=t-R+1}^{t} \| F_j \|^2 + O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)$$
Since $\sup_t \frac{1}{N} \sum_{j=t-R+1}^t \|F_j\|^2$ is $O_p(1)$ by Assumption 2 we are left with the part in $\left\|H_{NR}^{(t)} - H_t^{(1)}\right\|^2$. To deal with this part we can directly follow the proofs in Bai (2003), with Lemma A showing that required results on factor estimation error hold uniformly in $t$. Firstly, we know that $\Lambda' \Lambda/N \xrightarrow{P} \Sigma$ by Assumption 2b. For the first part of the rotation matrix in $\hat{V}^{(t)}-1$, we note that this term is the analogue to the full sample version in Bai (2003) Proposition 1, and therefore as in Stock and Watson (2002a) we have that:

$$\hat{V}^{(t)} = V + o_p(1)$$

uniformly over $t$ where $V$ is the diagonal matrix of the eigenvalues of $\Sigma \Lambda$, using Lemma A. Finally, since $\hat{V}^{(t)}-1 - V^{-1} = \hat{V}^{(t)}-1 \left( V - \hat{V}(t) \right) V^{-1}$ it follows that:

$$\sup_t \left\|\hat{V}^{(t)}-1 - V^{-1}\right\| = o_p(1)$$

in the same way as Gonçalves et al. (2015) Lemma A6. Finally for part of the matrix in $\hat{F}^{(t)^\prime}/F^{(t)}/R$, we note that this term is the analogue to the full sample version in Bai (2003) Proposition 1, and we have that:

$$\frac{\hat{F}^{(t)^\prime}/F^{(t)}}{R} = Q^{(t)} + o_p(1)$$

uniformly in $t$, where $Q^{(t)} = \Sigma^{-1/2} T^{(t)} V^{1/2}$ and $T^{(t)}$ is the eigenvector matrix of $\Sigma^{1/2} \Sigma F \Sigma^{1/2}$ but whose sign is determined by the column sign of $\hat{F}^{(t)}$. Therefore we write $Q^{(t)} = S^{(t)} Q$, where $Q^{(t)} = \Sigma^{-1/2} T V^{1/2}$. Combining these three parts, we have that:

$$H_{NR}^{(t)} = S^{(t)} V^{-1} Q \Sigma + o_p(1)$$

uniformly in $t$, which in turn yields $\sup_t \left\|H_{NR}^{(t)} - H_t^{(1)}\right\|^2 = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)$ since we define $H_t^{(1)} = S^{(t)} H^{1} \text{ with } H^{1} = V^{-1} Q \Sigma$. Combining all of these results we have:

$$\sup_t \frac{1}{R} \sum_{j=t-R+1}^t \left\|\hat{F}_j^{(t)} - H_t^{(1)} F_j\right\|^2 = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)$$

30
as required for the first part of Proposition 1a.

The second part in \( \widehat{\beta}_t \) follows directly using Lemmas A(i),(iii) and (v), since in the same way as Bai and Ng (2006), \( \widehat{\beta}_t \) is consistent up to the inverse of the rotation matrix for \( \widehat{F}^{(t)} \), therefore:

\[
\sup_t \| \widehat{\beta}_t - H_t^{(t-1)} \beta \| = o_p(1)
\]

using the results of West (1996) Lemma A3. This shows what was required.

**Proof of Proposition 1b:**

This Proposition relies on relating the adjusted PCA estimates \( \tilde{F}^{(t)} \) to the standard PCA estimates \( \hat{F}^{(t)} \) in order to apply the results from Lemma A and Proposition 1a. Recall from Section 3 that the adjusted PCA estimates under Normalization N2 for rolling window \( R \leq t \leq T \) are:

\[
\tilde{F}^{(t)} = \hat{F}^{(t)} \tilde{\Lambda}_1^{(R)} \left( \tilde{\Lambda}_1^{(R)^T} \right)^{-1}
\]

Writing this as an \( r \times 1 \) vector of factor estimates for a given date \( j \) yields:

\[
\tilde{F}^{(t)}_j = \left( \tilde{\Lambda}_1^{(R)} \right)^{-1} \tilde{\Lambda}_1^{(t)} \hat{F}^{(t)}_j
\]

where \( j = t - R + 1, ..., t \) for each \( R \leq t \leq T \). Subtracting from both sides \( H_{NR}^{(R)} F_j \), the true factors rotated about the PCA rotation matrix from the first rolling window, and manipulating the expression:

\[
\tilde{F}^{(t)}_j - H_{NR}^{(R)} F_j = \left( \tilde{\Lambda}_1^{(R)} \right)^{-1} \tilde{\Lambda}_1^{(t)} \hat{F}^{(t)}_j - H_{NR}^{(R)} F_j
\]

Therefore writing this as an average over the \( R \) observations \( j = t - R + 1, ..., t \) we have:

\[
\frac{1}{R} \sum_{j=t-R+1}^{t} \left( \tilde{F}^{(t)}_j - H_{NR}^{(R)} F_j \right) = \left[ \left( \tilde{\Lambda}_1^{(R)} \right)^{-1} \tilde{\Lambda}_1^{(t)} \right] \frac{1}{R} \sum_{j=t-R+1}^{t} \left( \hat{F}^{(t)}_j - H_{NR}^{(R)} F_j \right)
\]

\[
+ \left[ \left( \tilde{\Lambda}_1^{(R)} \right)^{-1} \tilde{\Lambda}_1^{(t)} H_{NR}^{(t)} - H_{NR}^{(R)} \right] \frac{1}{R} \sum_{j=t-R+1}^{t} F_j
\]

Now we use the result in Lemma A(iv) on rolling factor loading consistency uniformly over \( t \). Since this result holds over all the rows of \( \Lambda \) including those in \( \Lambda_1 \) it follows uniformly in \( t \) that:

\[
\tilde{\Lambda}_1^{(t)} = \Lambda_1 H_{NR}^{(t)} + o_p \left( \max \left\{ \frac{1}{N}, \frac{1}{\sqrt{R}} \right\} \right)
\]

31
This result therefore implies for the terms in square brackets on the RHS of Equation (27) we have

\begin{align*}
\left(\hat{\Lambda}_1^{(R)}\right)^{-1} \hat{\Lambda}_1^{-1} &= H_{NR}^{(R)} \Lambda_1^{-1} H_{NR}^{(t)-1} + O_p \left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{R}}\right\}\right) \\
&= H_{NR}^{(R)} H_{NR}^{(t)-1} + o_p (1) \\
&= O_p (1)
\end{align*}

(28)

uniformly in \( t \), since we assume that \( \Lambda_1 \) is invertible, and since \( H_{NR}^{(t)} \) is \( O_p (1) \) uniformly in \( t \) and is of full rank. For the second square-bracketed term we can again use Lemma A(iv) to show uniformly in \( t \) that:

\begin{align*}
\left(\hat{\Lambda}_1^{(R)}\right)^{-1} \hat{\Lambda}_1 \Lambda_1^{-1} H_{NR}^{(t)-1} H_{NR}^{(R)} &= H_{NR}^{(R)} \Lambda_1^{-1} \Lambda_1 H_{NR}^{(t)-1} - H_{NR}^{(R)} + O_p \left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{R}}\right\}\right) \\
&= O_p \left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{R}}\right\}\right)
\end{align*}

(29)

as the rotation matrix \( H_{NR}^{(R)} \) cancels out with its own inverse, as does \( \Lambda_1 \), and we are left with \( H_{NR}^{(R)} - H_{NR}^{(R)} \). Finally, from Equation (27) we have:

\begin{align*}
\sup_t \frac{1}{R} \sum_{j=t-R+1}^t \|\tilde{F}_j - H_{NR}^{(R)} F_j\|_2^2 &\leq 2 \sup_t \left\| \left(\hat{\Lambda}_1^{(R)}\right)^{-1} \right\|_2^2 \sup_t \frac{1}{R} \sum_{j=t-R+1}^t \|\tilde{F}_j - H_{NR}^{(R)} F_j\|_2^2 \\
&+ 2 \sup_t \left\| \left(\hat{\Lambda}_1^{(R)}\right)^{-1} \hat{\Lambda}_1 \right\|_2^2 \sup_t \frac{1}{R} \sum_{j=t-R+1}^t \|F_j\|_2^2 \\
&= O_p (1) \times O_p \left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{R}}\right\}\right) + O_p \left(\max\left\{\frac{1}{N^2}, \frac{1}{R}\right\}\right) \times O_p (1) \\
&= O_p \left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{R}}\right\}\right)
\end{align*}

(26)

by the results in Equation (28), Lemma A(i), Equation (29) and Assumption 2a. Finally, exactly as in the proof of Proposition 1a we can now replace the rotation matrix \( H_{NR}^{(R)} \) with \( H_{NR}^{(R)} \), where in this case we only have the rotation matrix from the first rolling window. It therefore follows that:

\begin{align*}
\sup_t \frac{1}{R} \sum_{j=t-R+1}^t \|\tilde{F}_j - H_{NR}^{(R)} F_j\|_2^2 &= \sup_t \left\| \tilde{\beta}_t - H_{R}^{(t)} \tilde{\beta}_t\right\|_2^2 = O_p \left(\max\left\{\frac{1}{N}, \frac{1}{\sqrt{R}}\right\}\right)
\end{align*}

by the same logic as in Equation (26). This shows what was required for the first part. Having shown this result, as in Proposition 1b it follows that the rolling OLS estimator is consistent up to the inverse of the same rotation matrix as \( \tilde{F}_j^{(R)} \), and we have:

\begin{align*}
\sup_t \left\| \tilde{\beta}_t - H_{R}^{(t)} \tilde{\beta}_t\right\|_2 &= o_p (1)
\end{align*}


32
9.2 Proof of Theorem 1

In proving Theorem 1 we additionally require the following Lemma and results which come as a corollary of Proposition 1.

**Lemma B** Under Normalization N1 and Assumptions 1-8:

(i) \[
\frac{1}{\sqrt{PR}} \sum_{t=R}^{T} \sum_{j=t-R+1}^{t-h} \left( \tilde{F}_j^{(t)} - H_{NR}^{(t)} F_j \right) \epsilon_{1,j+h} = O_p \left( \max \left\{ \frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R} \right\} \right)
\]

(ii) \[
\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla g(\epsilon_{1,t+h}) \left( \tilde{F}_t^{(t)} - H_{NR}^{(t)} F_t \right) = O_p \left( \max \left\{ \frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R} \right\} \right)
\]

where \( H_{NR}^{(t)} \) is the rotation matrix described in Proposition 1a.

**Corollary A** Under Normalization N2, and given Proposition 1b and Lemma B:

(i) \[
\frac{1}{\sqrt{PR}} \sum_{t=R}^{T} \sum_{j=t-R+1}^{t-h} \left( \tilde{F}_j^{(t)} - H_{R}^{(t)} R F_j \right) \epsilon_{1,j+h} = O_p \left( \max \left\{ \frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R} \right\} \right)
\]

(ii) \[
\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla g(\epsilon_{1,t+h}) \left( \tilde{F}_t^{(t)} - H_{R}^{(t)} R F_t \right) = O_p \left( \max \left\{ \frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R} \right\} \right)
\]

where \( H_{R}^{(t)} \) is the rotation matrix described in Proposition 1b.

Lemma B provides results on factor estimation error of the standard PCA factor estimates under Normalization N1 when used in factor-augmented regressions. These extend the results of Bai and Ng (2006) Lemma A.1 to the rolling estimation case, and are also confined to the separate Appendix. Corollary A shows the same result as in Lemma B, but for the adjusted PCA estimates \( \tilde{F}_i^{(t)} \), which are consistent for the rotation matrix \( H_{R}^{(t)} \) as shown in Proposition 1b. The proofs of these can also be found in the separate Appendix.

Taking the test statistic \( \tilde{S}_P \) in Equation (14):

\[
\tilde{S}_P = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g(\tilde{c}_{1,t+h}) - g(\tilde{c}_{2,t+h}) \right)
\]

The part which is new to this paper is the first term in \( g(\tilde{c}_{1,t+h}) \) which involves the adjusted rolling PCA estimates \( \tilde{F}_j^{(t)} \) and corresponding regression estimates \( \tilde{\beta}_t \) (since we assume for simplicity in
these proofs that Model 2 does not contain factors.) Taking a second-order Taylor series expansion of this first part around the (rotated) probability limits $H_R^1 F_t$ and $H_R^{t-1} \beta$ yields:

$$
\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \frac{d}{d(\epsilon_{1,t+h})} g(\epsilon_{1,t+h}) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \frac{d}{d(\epsilon_{1,t+h})} g(\epsilon_{1,t+h}) + \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_F g(\epsilon_{1,t+h}) \left( \tilde{F}_t^{(t)} - H_R^1 F_t \right) + \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\beta} g(\epsilon_{1,t+h}) \left( \tilde{\beta}_t - H_R^{t-1} \beta \right) + o_p(1)
$$

The second-order terms are $o_p(1)$ as in West (1996) proof of Equation 4.1 part (b) since Assumptions 1, 5c, 6 and 7 in this paper ensure that Assumptions 1, 2 and 3 of West (1996) hold. Furthermore, Corollary A(ii) shows that:

$$
\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_F g(\epsilon_{1,t+h}) \left( \tilde{F}_t^{(t)} - H_R^1 F_t \right) = O_p \left( \max \left\{ \frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R} \right\} \right)
$$

therefore we can write:

$$
\frac{1}{\sqrt{P}} \sum_{t=R}^{T} g(\epsilon_{1,t+h}) \left( \tilde{F}_t^{(t)} - H_R^1 F_t \right) = O_p \left( \max \left\{ \frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R} \right\} \right)
$$

since we assume in Assumption 8(a) that $\sqrt{T}/N \to 0$. Now using a similar argument to Bai and Ng (2006) proof of Theorem 1, with $\tilde{\beta}_t$ estimated by OLS we can write:

$$
\tilde{\beta}_t - H_R^{t-1} \beta = \left( H^1 \Sigma_F H^\dagger \right)^{-1} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \tilde{F}_j^{(t)} \epsilon_{1,j+h} + o_p(1)
$$

since Proposition 1b implies that $\frac{1}{R} \sum_{j=t-R+1}^{t-h} \tilde{F}_j^{(t)} \tilde{F}_j^{(t)'} = H^1 \Sigma_F H^\dagger + o_p(1)$ uniformly in $t$, noting that the sign matrices $S^{(R)}$ cancel in the product. Therefore since stationarity and strong mixing of $\epsilon_{1,t+h}$ in Assumption 1 along with measurability of $g(.)$ by Assumption 6, $\nabla_{\beta} g(\epsilon_{1,t+h})$ and $F_t \epsilon_{1,t+h}$ are also stationary and strong mixing with moments bounded as in Assumption 7a, so Assumptions 2 and 3 of West (1996) hold for the factor augmented-model. Now in a similar way to West (1996) proof of Equation 4.1 part (a) we therefore have:

$$
\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\beta} g(\epsilon_{1,t+h}) \left( \tilde{\beta}_t - H_R^{t-1} \beta \right)
$$

$$
= D_\beta \left( H^1 \Sigma_F H^\dagger \right)^{-1} H^\dagger \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \sum_{j=t-R+1}^{t-h} F_j \epsilon_{1,j+h}
$$

$$
+ D_\beta \left( H^1 \Sigma_F H^\dagger \right)^{-1} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \sum_{j=t-R+1}^{t-h} \left( \tilde{F}_j^{(t)} - H_R^1 F_j \right) \epsilon_{1,j+h} + o_p(1)
$$
where $D_3$ is described in Theorem 1. Now we have a similar expression to that in West (1996) but for the factor estimation error term. For this term we can use Corollary A(i), which shows that:

$$\frac{1}{\sqrt{PR}} \sum_{t=R}^{T} \sum_{j=t-h}^{t} \left( \tilde{F}_j^{(t)} - H_R^i F_j \right) \epsilon_{1,j+h} = O_p \left( \max \left\{ \frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R} \right\} \right)$$

Therefore:

$$\frac{1}{\sqrt{PR}} \sum_{t=R}^{T} g \left( \tilde{\epsilon}_{1,t+h} \right) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} g \left( \epsilon_{1,t+h} \right) + D_\beta \left( H^i \Sigma F H^i \right)^{-1} H^i \frac{1}{\sqrt{PR}} \sum_{t=R}^{T} \sum_{j=t-h}^{t} F_j \epsilon_{1,j+h} + o_p (1)$$

The second part of $\tilde{S}_P$ does not involve any factor estimation error and therefore is simply a direct application of West (1996) to the linear forecasting model case.

$$\frac{1}{\sqrt{PR}} \sum_{t=R}^{T} g \left( \tilde{\epsilon}_{2,t+h} \right) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} g \left( \epsilon_{2,t+h} \right) + D_\gamma \Sigma^{-1} \frac{1}{\sqrt{PR}} \sum_{t=R}^{T} \sum_{j=t-h}^{t} Z_j \epsilon_{2,j+h} + o_p (1)$$

Therefore $\tilde{S}_P$ can be written fully as:

$\tilde{S}_P = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g \left( \tilde{\epsilon}_{1,t+h} \right) - g \left( \tilde{\epsilon}_{2,t+h} \right) \right)$

$$= \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g \left( \epsilon_{1,t+h} \right) - g \left( \epsilon_{2,t+h} \right) \right) + D_\beta \left( H^i \Sigma F H^i \right)^{-1} H^i \frac{1}{\sqrt{PR}} \sum_{t=R}^{T} \sum_{j=t-h}^{t} F_j \epsilon_{1,j+h}$$

$$- D_\gamma \Sigma^{-1} \frac{1}{\sqrt{PR}} \sum_{t=R}^{T} \sum_{j=t-h}^{t} Z_j \epsilon_{2,j+h} + O_p \left( \max \left\{ \frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R} \right\} \right) + o_p (1)$$

Having established this, asymptotic normality completes the proof of Theorem 1 in the same way as Theorem 4.1 of West (1996). Under the hypothesis $H_0 : \text{E} \left( g \left( \epsilon_{1,t+h} \right) - g \left( \epsilon_{2,t+h} \right) \right) = 0$:

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g \left( \tilde{\epsilon}_{1,t+h} \right) - g \left( \tilde{\epsilon}_{2,t+h} \right) \right) \xrightarrow{d} N \left( 0, \Omega \right)$$

\(^{19}\)Note that, again, the sign matrix $S^{(R)}$ cancels out between $\nabla g \left( \epsilon_{1,t+h} \right)$ and $H_R^i F_j$. 

35
where:
\[
\Omega = V_t + \theta_1 D_\beta V_t D'_\beta + 2 \theta_2 D_\beta C_t F_t \\
+ \theta_1 D_\gamma V_t D'_\gamma - 2 \theta_2 D_\gamma C_t, Z - 2 \theta_1 D_\beta C_t, Z D'_\gamma
\]

where for the rolling estimation scheme with \(\pi \geq 1\), \(\theta_1 = (1 - \frac{1}{2\pi})\) and \(\theta_2 = (1 - \frac{1}{2\pi})\). For the rolling estimation scheme with \(\pi < 1\), \(\theta_1 = (\pi - \frac{\pi^2}{3})\) and \(\theta_2 = \frac{\pi}{2}\). For the recursive scheme, \(\theta_1 = 2(1 - \frac{1}{\pi}\ln(1 + \pi))\) and \(\theta_2 = 1 - \frac{1}{\pi}\ln(1 + \pi)\).

The variance-covariance matrices are fully described in the statement of Theorem 1 in the text.

This shows what was required.

9.3 Proof of Proposition 2a and 2b

Proof of Proposition 2a

We begin by showing the result for \(\tilde{F}_{t}^{rol}\) described in Equation (16c). We can relate the full sample sum over \(\tilde{F}_{t}^{rol}\) to a sum over the \(K\) adjusted rolling PCA components \(\tilde{F}_{\gamma}(R), \tilde{F}_{\gamma}(2R), \ldots, \tilde{F}_{\gamma}(T)\) in the following way:

\[
\sum_{t=1}^{T} \tilde{F}_{t}^{rol} = \tilde{F}_{1}^{(R)} + \ldots + \tilde{F}_{R}^{(R)} + \tilde{F}_{R+1}^{(2R)} + \ldots + \tilde{F}_{P}^{(2R)} + \ldots + \tilde{F}_{T}^{(T)}
\]

Now using the fact that \(T = KR\) by Assumption 8’ we can relate the average of \(\tilde{F}_{t}^{rol}\) over \(T\) observations to exactly \(K\) averages over \(R\) observations:

\[
\frac{1}{T} \sum_{t=1}^{T} \tilde{F}_{t}^{rol} = \frac{1}{T} \left[ \sum_{j=1}^{R} \tilde{F}_{j}^{(R)} + \sum_{j=R+1}^{2R} \tilde{F}_{j}^{(2R)} + \ldots + \sum_{j=P}^{T} \tilde{F}_{j}^{(T)} \right]
\]

As shown in Proposition 1b, the adjusted PCA estimates under Normalization N2 have the same asymptotic rotation matrix \(H_{\gamma}^{\dagger}\). Therefore we can subtract from both sides the whole time series of true factors rotated by \(H_{\gamma}^{\dagger}\):

\[
\frac{1}{T} \sum_{t=1}^{T} \left( \tilde{F}_{t}^{rol} - H_{\gamma}^{\dagger} F_t \right) = \frac{1}{K} \left[ \frac{1}{R} \sum_{j=1}^{R} \tilde{F}_{j}^{(R)} + \frac{1}{R} \sum_{j=R+1}^{2R} \tilde{F}_{j}^{(2R)} + \ldots + \frac{1}{R} \sum_{j=P}^{T} \tilde{F}_{j}^{(T)} \right] - \frac{1}{T} \sum_{t=1}^{T} H_{\gamma}^{\dagger} F_t
\]

(31)
Finally, since $K$ is finite the following inequality holds:

$$
\frac{1}{T} \sum_{t=1}^{T} \left\| \hat{F}_{rol} - H^*_RF_t \right\|^2 \leq \frac{1}{R} \sum_{j=1}^{R} \left\| \hat{F}^{(R)}_j - H^*_RF_j \right\|^2 + \frac{1}{R} \sum_{j=R+1}^{2R} \left( \hat{F}^{(2R)}_j - H^*_RF_j \right) + \frac{1}{R} \sum_{j=T-R+1}^{T} \left\| \hat{F}^{(T)}_j - H^*_RF_j \right\|^2
$$

$$= O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)
$$

were the $O_p(\max \left\{ \frac{1}{N}, \frac{1}{R} \right\})$ comes from Proposition 1b which shows convergence uniformly in $t$, which implies convergence over this finite number of $K$ rolling windows. This shows what was required for $\hat{F}_{rol}$.

To show that the same result does not hold for $\hat{F}_{rol}$, described in Equation (16a), which uses the sign-changing estimates under Normalization N1, note that the analogue of Equation (31) for $\hat{F}_{rol}$ becomes:

$$
\frac{1}{T} \sum_{t=1}^{T} \left( \hat{F}_{rol} - H^*_RF_t \right)
$$

$$= \frac{1}{K} \left[ \frac{1}{R} \sum_{j=1}^{R} \left( \hat{F}^{(R)}_j - H^*_RF_j \right) + \frac{1}{R} \sum_{j=R+1}^{2R} \left( \hat{F}^{(2R)}_j - H^*_RF_j \right) + \frac{1}{R} \sum_{j=T-R+1}^{T} \left( \hat{F}^{(T)}_j - H^*_RF_j \right) \right]
$$

$$= \frac{1}{K} \left[ \frac{1}{R} \sum_{j=1}^{R} \left( \hat{F}^{(R)}_j - H^*_RF_j \right) + \frac{1}{R} \sum_{j=R+1}^{2R} \left( \hat{F}^{(2R)}_j - H^*_RF_j \right) + \frac{1}{R} \sum_{j=R+1}^{2R} \left( H^*_R - H^*_R \right) F_j \right]
$$

$$\leq 2 \left[ \left\| H^*_R - H^*_R \right\|^2 \frac{1}{R} \sum_{j=R+1}^{2R} \left\| F_j \right\|^2 + \left\| H^*_T - H^*_T \right\|^2 \frac{1}{R} \sum_{j=T-R+1}^{T} \left\| F_j \right\|^2 \right] + O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)
$$

and since the rotation matrices $H^*_R, H^*_2R, ..., H^*_T$ are not the same asymptotically due to the sign-
changing issue it follows that \( \| H_{2R}^\dagger - H_R^\dagger \|^2 = O_p(1) \) and so on, which means that the entity \( \tilde{F}_{t}^{\text{rot}} \) is not consistent for the true \( F_t \) over the full sample.

Since the recursive versions \( \tilde{F}_t^{\text{rec}} \) and \( \hat{F}_t^{\text{rec}} \), described in Equations (16b) and (16d), are both constructed using a single full-sample factor estimate, consistency follows Bai and Ng (2006) for full-sample estimation, or using the uniform results of Gonçalves et al. (2015), which shows the consistency of PCA estimation uniformly over all recursive windows, and therefore applies to the last window. We therefore do not repeat these here.

**Proof of Proposition 2b**

The statement in Proposition 2b follows directly from the results of Propositions 1a, 1b and 2a. Specifically, since the sign-changing estimator \( \tilde{\beta}_t \) is shown in Proposition 1a to have a time-varying rotation matrix \( H_t^{\dagger \dagger -1} \), it therefore cannot have the same rotation for each \( R \leq t \leq T \) as \( \hat{F}_t^{\text{rec}}, \tilde{F}_t^{\text{rot}} \) or \( \hat{F}_t^{\text{rec}} \), which all satisfy Condition 2, and respectively have rotations \( H_t^\dagger, H_t^\dagger R \) and \( H_t^\dagger R \).

On the other hand, in Proposition 1b, the adjusted estimator \( \tilde{\beta}_t \) is shown to have rotation \( H_t^{\dagger \dagger -1} \) for all \( t \), and therefore Condition 3 holds for both the tilde estimates \( \tilde{F}_t^{\text{rot}} \) and \( \tilde{F}_t^{\text{rec}} \), but not \( \hat{F}_t^{\text{rec}} \), and therefore rules out the use of \( \tilde{F}_t^{\text{rec}} \) for resampling.

### 9.4 Proof of Theorem 2

Taking a Taylor series expansion of \( \tilde{S}_p^* \) around \( \tilde{\beta}_t \) and \( \tilde{\gamma}_t \) yields:

\[
\tilde{S}_p^* = \frac{1}{\sqrt{p}} \sum_{t=R}^{T} \left( g \left( \tilde{c}^*_t, y_{t+h} \right) - g \left( \tilde{c}^*_2, y_{t+h} \right) \right) - \frac{1}{T} \sum_{j=1}^{T} \left( g \left( \tilde{c}^*_1, y_{t+h} \right) - g \left( \tilde{c}^*_2, y_{t+h} \right) \right)
\]

\[
= \frac{1}{\sqrt{p}} \sum_{t=R}^{T} \left( g \left( y_{t+h} - \tilde{F}_t^{\text{rot}} \tilde{\beta}_t \right) - g \left( y_{t+h} - \tilde{F}_t^{\text{rot}} \tilde{\gamma}_t \right) \right) - \frac{1}{T} \sum_{j=1}^{T} \left( g \left( \tilde{c}^*_1, y_{t+h} \right) - g \left( \tilde{c}^*_2, y_{t+h} \right) \right)
\]

\[
+ \frac{1}{\sqrt{p}} \sum_{t=R}^{T} \nabla g \left( y_{t+h} - \tilde{F}_t^{\text{rot}} \tilde{\beta}_t \right) \left( \tilde{\beta}_t - \tilde{\beta}_t \right) + \frac{1}{\sqrt{p}} \sum_{t=R}^{T} \nabla g \left( y_{t+h} - \tilde{F}_t^{\text{rot}} \tilde{\gamma}_t \right) \left( \tilde{\gamma}_t - \tilde{\gamma}_t \right)
\]

(32)

where \( \tilde{\beta}_t \in \left( \tilde{\beta}^*_t, \tilde{\beta}_t \right) \) and \( \tilde{\gamma}_t \in \left( \tilde{\gamma}^*_t, \tilde{\gamma}_t \right) \) and \( \tilde{c}_{1,j+h,t} = y_{j+h} - \tilde{F}_j^{\text{rot}} \beta_t \) and \( \tilde{c}_{2,j+h,t} = y_{j+h} - \tilde{F}_j^{\text{rot}} \gamma_t \) as described in the text.

The proof of Theorem 2 is in two parts. We firstly need to show that \( \tilde{S}_p^* \) mean zero and then we need to show that \( \tilde{S}_p^* \) has variance \( \Omega \).

#### 9.4.1 Proof of Bootstrap Mean

We begin by showing that the three terms on the RHS of Equation (32) have mean zero. For the first part of Equation (32), we need to show that:

\[
E^* \left[ \frac{1}{\sqrt{p}} \sum_{t=R}^{T} \left( g \left( y_{t+h} - \tilde{F}_t^{\text{rot}} \tilde{\beta}_t \right) - g \left( y_{t+h} - \tilde{F}_t^{\text{rot}} \tilde{\gamma}_t \right) \right) \right]
\]
\[
\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \frac{1}{T} \sum_{j=1}^{T} \left( g(\tilde{\epsilon}_{1,j+h,t}^*) - g(\tilde{\epsilon}_{2,j+h,t}^*) \right) \right) + o_p(1)
\]

so that it follows by the bootstrap law of large numbers (LLN) that:

\[
\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g(y_{t+h}^* - \tilde{F}_t^* \tilde{\beta}_t^*) - g(y_{t+h}^* - Z_t^* \tilde{\gamma}_t) \right) \\
- \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \frac{1}{T} \sum_{j=1}^{T} \left( g(\tilde{\epsilon}_{1,j+h,t}^*) - g(\tilde{\epsilon}_{2,j+h,t}^*) \right) \right) = o_p(1) \tag{33}
\]

Consider the first part of this expectation:

\[
E^* \left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} g(y_{t+h}^* - \tilde{F}_t^* \tilde{\beta}_t^*) \right] \\
= \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \frac{1}{T-l+1} g(y_{1+h}^* - \tilde{F}_1^* \tilde{\beta}_1) + \ldots + \frac{1}{T-l+1} g(y_{T+h}^* - \tilde{F}_T^* \tilde{\beta}_T) \right) + O_p \left( \frac{l}{T} \right) \\
= \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \frac{1}{T} \sum_{j=1}^{T} g(y_{j+h}^* - \tilde{F}_j^* \tilde{\beta}_j^*) \right) + o_p(1)
\]

since \(l/T \to 0\). Similarly for the second part it follows that:

\[
E^* \left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} g(y_{t+h}^* - Z_t^* \tilde{\gamma}_t) \right] = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \frac{1}{T} \sum_{j=1}^{T} g(y_{j+h}^* - Z_j^* \tilde{\gamma}_j) \right) + o_p(1)
\]

And recalling the definitions of \(\tilde{\epsilon}_{1,j+h,t}^*\) and \(\tilde{\epsilon}_{2,j+h,t}^*\) we have:

\[
E^* \left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g(y_{t+h}^* - \tilde{F}_t^* \tilde{\beta}_t^*) - g(y_{t+h}^* - Z_t^* \tilde{\gamma}_t) \right) \right] \\
- \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \frac{1}{T} \sum_{j=1}^{T} \left( g(\tilde{\epsilon}_{1,j+h,t}^*) - g(\tilde{\epsilon}_{2,j+h,t}^*) \right) \right) = o_p(1)
\]

which shows (33) as required. For the expectation of the second part of Equation (32) we need to show that:

\[
E^* \left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla g \left( y_{t+h}^* - \tilde{F}_t^* \tilde{\beta}_t \right) \left( \tilde{\beta}_t^* - \tilde{\beta}_t \right) \right] = o_p(1)
\]

so that by the bootstrap LLN:

\[
\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla g \left( y_{t+h}^* - \tilde{F}_t^* \tilde{\beta}_t \right) \left( \tilde{\beta}_t^* - \tilde{\beta}_t \right) = o_p(1) \tag{34}
\]
Consider the bootstrap estimator from the recentered OLS objective function described in (17).

\[
\tilde{\beta}_t^* = \arg\min_{\beta} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( y_{j+h}^* - \tilde{F}_j^* \beta \right)^2 + 2\beta' \left( \frac{1}{T} \sum_{j'=1}^{T} \tilde{F}_{j'} \tilde{e}_{1,j'+h,t} \right)
\]

\[
= \left( \frac{1}{R} \sum_{j=t-R+1}^{t-h} \tilde{F}_j^* \tilde{F}_j^* \right)^{-1} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( \tilde{F}_j^* y_{j+h}^* - \left( \frac{1}{T} \sum_{j'=1}^{T} \tilde{F}_{j'} \tilde{e}_{1,j'+h,t} \right) \right)
\]

Now for all \( R \leq t \leq T \), defining the following error term \( \tilde{e}_{1,j+h,t}^* = y_{j+h}^* - \tilde{F}_j^* \tilde{\beta}_t \) for \( j = t-R+1, \ldots, t-h \) we substitute \( y_{j+h}^* \) into the above expression to get:

\[
\tilde{\beta}_t^* - \tilde{\beta}_t = \left( \frac{1}{R} \sum_{j=t-R+1}^{t-h} \tilde{F}_j^* \tilde{F}_j^* \right)^{-1} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( \tilde{F}_j^* \tilde{e}_{1,j+h,t}^* - \left( \frac{1}{T} \sum_{j'=1}^{T} \tilde{F}_{j'} \tilde{e}_{1,j'+h,t} \right) \right)
\]

This gives an expression equivalent to that in Corradi and Swanson (2006) Proof of Proposition 2 but for the linear case. Now recalling the definition that \( \tilde{e}_{1,j+h,t} = y_{j+h} - \tilde{F}_j \tilde{\beta}_t \), we first show that

\[
E^* \left[ \frac{1}{R} \sum_{j=t-R+1}^{t-h} \tilde{F}_j^* \tilde{e}_{1,j+h,t} \right] = \frac{1}{T} \sum_{j=1}^{T} \tilde{F}_j \tilde{e}_{1,j+h,t} + o_p(1) \tag{35}
\]

since:

\[
E^* \left[ \frac{1}{R} \sum_{j=t-R+1}^{t-h} \tilde{F}_j^* \tilde{e}_{1,j+h,t}^* \right] = E^* \left[ \frac{1}{R} \sum_{j=t-R+1}^{t-h} \tilde{F}_j^* \left( y_{j+h}^* - \tilde{F}_j \tilde{\beta}_t \right) \right]
\]

\[
= \left( \frac{1}{T-l+1} \tilde{F}_1 \right) (y_{l+h} - \tilde{F}_1 \tilde{\beta}_t) + \ldots + \left( \frac{1}{T-l+1} \tilde{F}_T \right) (y_{T+h} - \tilde{F}_T \tilde{\beta}_t) + O_p \left( \frac{l}{T} \right)
\]

\[
= \frac{1}{T} \sum_{j=1}^{T} \tilde{F}_j \tilde{e}_{1,j+h,t} + o_p(1)
\]

as \( l/T \to 0 \). Also note that:

\[
E^* \left[ \frac{1}{R} \sum_{j=t-R+1}^{t-h} \tilde{F}_j^* \tilde{F}_j^* \right] = \frac{1}{T} \sum_{j=1}^{T} \tilde{F}_j \tilde{F}_j' + o_p(1)
\]

\[
= H_\tilde{F} \Sigma F H_\tilde{F}' + o_p(1)
\]

uniformly over \( t \), which follows from Proposition 2a as \( \tilde{F}_j \) is consistent for \( H_\tilde{F} \tilde{F}_j \) and because the
sign matrix $S^{(R)}$ cancels in the product. Therefore it follows that:

$$
E^{*}\left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\beta} g \left( y_{t+h} - \tilde{F}_{t} \beta_{t} \right) \left( \tilde{\beta}_{t}^{*} - \tilde{\beta}_{t} \right) \right] = D_{\beta} \left( H^\dagger \Sigma_{P} H^\dagger \right)^{-1} E^{*}\left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( \tilde{F}_{j} \tilde{\gamma}_{1,j+h,t} - \frac{1}{T} \sum_{j'=1}^{T} \tilde{F}_{j'} \tilde{\gamma}_{1,j'+h,t} \right) \right] + o_{P}(1) \quad (36)
$$

$$
= o_{P}(1)
$$

because, as in Corradi and Swanson (2006), it follows from (35) that the bias term on the RHS of (36) when rescaled by $\sqrt{P}$ is of order $O_{p} \left( l/\sqrt{T} \right)$ since $P = O(T)$ and therefore this bias is asymptotically negligible since we assume that $l/\sqrt{T} \to 0$. This shows (34) as required.

Finally, for the expectation of the last part of Equation (32) we need to show that:

$$
\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\gamma} g \left( y_{t+h} - Z_{t} \gamma_{t} \right) \left( \tilde{\gamma}_{t}^{*} - \tilde{\gamma}_{t} \right) = o_{P^{*}}(1) \quad (37)
$$

Since we assume for simplicity that $Z_{t}$ does not contain estimated factors, this is the same as the proof in Corradi and Swanson (2006) for the linear case, so we do not repeat this proof here. If $Z_{t}$ contains estimated factors then we can treat it in the same way as the previous proof.

Combining the results in Equations (33), (34) and (37) shows that the statistic $\tilde{S}_{P}^{*}$ has mean zero, as required.

### 9.4.2 Proof of Bootstrap Variance

As Equation (32) is made up of three terms, the bootstrap variance contains three variances and three distinct covariances. We need to show that each of these are consistent for the three variance and three covariance parts of the matrix $\Omega$ described in Theorem 1. We will only show the first of these 6 terms, and the rest follow very similar lines, or are direct applications of the results in Corradi and Swanson (2006).

Firstly define the following variables de-meaned by bootstrap expectation:

$$
\bar{g} \left( y_{t+h} - \tilde{F}_{t} \beta_{t} \right) = g \left( y_{t+h} - \tilde{F}_{t} \beta_{t} \right) - \frac{1}{T} \sum_{j=1}^{T} g \left( y_{j+h} - \tilde{F}_{j} \beta_{t} \right)
$$

$$
\nabla_{F} \bar{g} \left( y_{t+h} - \tilde{F}_{t} \beta_{t} \right) = \nabla_{F} g \left( y_{t+h} - \tilde{F}_{t} \beta_{t} \right) - \frac{1}{T} \sum_{j=1}^{T} \nabla_{F} g \left( y_{j+h} - \tilde{F}_{j} \beta_{t} \right)
$$

41
We start by taking a Taylor series expansion around the true \( \gamma \) and true (rotated) \( H_{R}^{t-1} \beta \):

\[
\frac{1}{\sqrt{P}} \sum_{t=1}^{T} \left( g \left( y_{t+h}^{*} - \tilde{F}_{t}^{\text{rot}} \tilde{\beta}_{t} \right) - g \left( y_{t+h}^{*} - Z_{t}^{\text{rot}} \gamma_{t} \right) \right) = V_{e} + o_{p}(1) \tag{38}
\]

We start by taking a Taylor series expansion around the true \( \gamma \) and true (rotated) \( H_{R}^{t-1} \beta \):

\[
\frac{1}{\sqrt{P}} \sum_{t=1}^{T} \left( g \left( y_{t+h}^{*} - \tilde{F}_{t}^{\text{rot}} \tilde{\beta}_{t} \right) - g \left( y_{t+h}^{*} - Z_{t}^{\text{rot}} \gamma_{t} \right) \right)
= \frac{1}{\sqrt{P}} \sum_{t=1}^{T} \left( g \left( y_{t+h}^{*} - \tilde{F}_{t}^{\text{rot}} H_{R}^{t-1} \beta \right) - g \left( y_{t+h}^{*} - Z_{t}^{\text{rot}} \gamma \right) \right)
+ \frac{1}{\sqrt{P}} \sum_{t=1}^{T} \nabla_{\beta} g \left( y_{t+h}^{*} - \tilde{F}_{t}^{\text{rot}} \tilde{\beta}_{t} \right) \left( \tilde{\beta}_{t} - H_{R}^{t-1} \beta \right)
+ \frac{1}{\sqrt{P}} \sum_{t=1}^{T} \nabla_{\gamma} g \left( y_{t+h}^{*} - Z_{t}^{\text{rot}} \gamma_{t} \right) \left( \gamma_{t} - \gamma \right) \tag{39}
\]

since the last two terms on the RHS of (39) are \( o_{p}(1) \) as shown by Corradi and Swanson (2006).

We now analyse the bootstrap variance of the last line. Without loss of generality take \( R = b_{1} \ell \)

\[
\text{Var}^{*} \left[ \frac{1}{\sqrt{P}} \sum_{t=1}^{T} \left( g \left( y_{t+h}^{*} - \tilde{F}_{t}^{\text{rot}} H_{R}^{t-1} \beta \right) - g \left( y_{t+h}^{*} - Z_{t}^{\text{rot}} \gamma \right) \right) \right]
= \text{Var}^{*} \left[ \frac{1}{\sqrt{P}} \sum_{j=b_{1}+1}^{b} \sum_{i=1}^{l} \left( g \left( y_{j+i+h}^{*} - \tilde{F}_{j+i}^{\text{rot}} H_{R}^{t-1} \beta \right) - g \left( y_{j+i+h}^{*} - Z_{j+i}^{\text{rot}} \gamma \right) \right) \right]
= \mathbb{E}^{*} \left[ \frac{1}{T} \sum_{j=1}^{l} \sum_{i=1}^{l} \left( g \left( y_{j+i+h}^{*} - \tilde{F}_{j+i}^{\text{rot}} H_{R}^{t-1} \beta \right) - g \left( y_{j+i+h}^{*} - Z_{j+i}^{\text{rot}} \gamma \right) \right) \right.
\times \left( g \left( y_{j+i+h}^{*} - \tilde{F}_{j+i}^{\text{rot}} H_{R}^{t-1} \beta \right) - g \left( y_{j+i+h}^{*} - Z_{j+i}^{\text{rot}} \gamma \right) \right) \right]
= \frac{1}{T} \sum_{t=1}^{T-1} \sum_{j=1}^{l} \left( \mathbb{E} \left( y_{t+h}^{*} - \tilde{F}_{t}^{\text{rot}} H_{R}^{t-1} \beta \right) - \mathbb{E} \left( y_{t+h}^{*} - Z_{t}^{\text{rot}} \gamma \right) \right)
\times \left( \mathbb{E} \left( y_{t+h}^{*} - \tilde{F}_{t}^{\text{rot}} H_{R}^{t-1} \beta \right) - \mathbb{E} \left( y_{t+h}^{*} - Z_{t}^{\text{rot}} \gamma \right) \right) + o_{p} \left( \frac{T^{2}}{T} \right) \tag{40}
\]

Now, unlike in Corradi and Swanson (2006) we also need to relate estimated factors back to true (rotated) factors. Therefore from Proposition 2 which states that:

\[
\frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{F}_{t} - H_{R}^{t} F_{t} \right\|^{2} = o_{p} \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)
\]

which applies to both \( \tilde{F}_{t}^{\text{rod}} \) and \( \tilde{F}_{t}^{\text{ec}} \), we expand \( \mathbb{E} \left( y_{t+h}^{*} - \tilde{F}_{t}^{\text{rot}} H_{R}^{t-1} \beta \right) \) around \( H_{R}^{t} F_{t} \) for all \( t \), which
gives:

$$\bar{g} \left( y_{t+h} - \hat{F}_t H_{R}^{t-1} \beta \right) = \bar{g} \left( y_{t+h} - F_t' \beta \right) + \nabla F \bar{g} \left( y_{t+h} - F_t' H_{R}^{t-1} \beta \right) \left( \hat{F}_t - H_{R} \hat{F}_t \right) \quad \text{(41)}$$

for some $F_t \in \left( \hat{F}_t, H_{R} \hat{F}_t \right)$. It is crucial here that $\tilde{\beta}$ has the same rotation as $\beta$ as in Condition 3, otherwise the last line would not hold. Therefore the variance expression in Equation (40) can be written as the variance of the true forecast error loss differential $\bar{g} (\epsilon_{1,t+h}) - \bar{g} (\epsilon_{2,t+h})$ plus cross-products and the square of the factor estimation error term:

$$= \frac{1}{T} \sum_{t=1}^{T-l} \sum_{j=-l}^{l} \left( \left( \bar{g} (\epsilon_{1,t+h}) - \bar{g} (\epsilon_{2,t+h}) + \nabla F \bar{g} \left( y_{t+h+j} - F_{t+j}' H_{R}^{t-1} \beta \right) \left( \tilde{F}_{t+j} - H_{R} \tilde{F}_{t+j} \right) \right) \right) \times \left( \bar{g} (\epsilon_{1,t+h+j}) - \bar{g} (\epsilon_{2,t+h+j}) + \nabla F \bar{g} \left( y_{t+h+j} - F_{t+j}' H_{R}^{t-1} \beta \right) \left( \tilde{F}_{t+j} - H_{R} \tilde{F}_{t+j} \right) \right) + O_p \left( \frac{l^2}{T} \right) \quad \text{(40)}$$

$$= \frac{1}{T} \sum_{t=1}^{T-l} \sum_{j=-l}^{l} \left[ \left( \bar{g} (\epsilon_{1,t+h}) - \bar{g} (\epsilon_{2,t+h}) \right) \left( \bar{g} (\epsilon_{1,t+h}) - \bar{g} (\epsilon_{2,t+h}) \right) + o_p (1) \right]$$

For the last line to hold we use a similar argument to that used implicitly in Corradi and Swanson (2014) proof of Theorem 2. Since the remaining terms are squares and cross-products of factor estimation error with mean-zero variables, they are all $o_p (1)$ by Proposition 2. For example take the first factor estimation error term:

$$= \frac{1}{T} \sum_{t=1}^{T-l} \sum_{j=-l}^{l} \left[ \left( \bar{g} (\epsilon_{1,t+h}) - \bar{g} (\epsilon_{2,t+h}) \right) \times \nabla F \bar{g} \left( y_{t+h+j} - F_{t+j}' H_{R}^{t-1} \beta \right) \left( \tilde{F}_{t+j} - H_{R} \tilde{F}_{t+j} \right) \right] \quad \text{(41)}$$

$$\leq (2l + 1) \sup_{-l \leq j \leq l} \left| \frac{1}{T} \sum_{t=1}^{T-l} \left( \bar{g} (\epsilon_{1,t+h}) - \bar{g} (\epsilon_{2,t+h}) \right) \times \nabla F \bar{g} \left( y_{t+h+j} - F_{t+j}' H_{R}^{t-1} \beta \right) \left( \tilde{F}_{t+j} - H_{R} \tilde{F}_{t+j} \right) \right|$$

$$= (2l + 1) \times O_p (1) O_p \left( \max \left\{ \frac{1}{\sqrt{N}}, \frac{1}{\sqrt{R}} \right\} \right)$$

Since for any $j$ we have:

$$\frac{1}{T} \sum_{t=1}^{T-l} \left( \bar{g} (\epsilon_{1,t+h}) - \bar{g} (\epsilon_{2,t+h}) \right) \times \nabla F \bar{g} \left( y_{t+h+j} - F_{t+j}' H_{R}^{t-1} \beta \right) \left( \tilde{F}_{t+j} - H_{R} \tilde{F}_{t+j} \right)$$

$$\leq \left( \frac{1}{T} \sum_{t=1}^{T-l} \left\| \left( \bar{g} (\epsilon_{1,t+h}) - \bar{g} (\epsilon_{2,t+h}) \right) \nabla F \bar{g} \left( y_{t+h+j} - F_{t+j}' H_{R}^{t-1} \beta \right) \right\|^2 \right)^{1/2}$$

$$\times \left( \frac{1}{T} \sum_{t=1}^{T-l} \left\| \tilde{F}_{t+j} - H_{R} \tilde{F}_{t+j} \right\|^2 \right)^{1/2}$$

And by Proposition 2 which states that $\frac{1}{T} \sum_{t=1}^{T-l} \left\| \tilde{F}_{t+j} - H_{R} \tilde{F}_{t+j} \right\|^2 = O_p \left( \max \left\{ \frac{1}{\sqrt{N}}, \frac{1}{\sqrt{R}} \right\} \right)$, the second term is $O_p \left( \max \left\{ \frac{1}{\sqrt{N}}, \frac{1}{\sqrt{R}} \right\} \right)$ and clearly the first term is $O_p (1)$. Therefore, since all the other factor
estimation error terms follow a similar logic (and the squared factor estimation error term is of yet smaller order), we will finally be able to show that:

\[
\text{Var}^* \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( g \left( y_{t+h} - \tilde{F}_t^* \tilde{\beta}_t \right) - g \left( y_{t+h} - Z_t^* \gamma_t \right) \right) \right] = \frac{1}{T} \sum_{t=1}^{T} \left[ (\bar{g}(\epsilon_{t+h}) - \bar{g}(\epsilon_{2,t+h})) (\bar{g}(\epsilon_{1,t+h+j}) - \bar{g}(\epsilon_{2,t+h+j})) \right]
\]

\[+ O_p(l) O_p \left( \max \left\{ \frac{1}{\sqrt{N}}, \frac{1}{\sqrt{R}} \right\} \right) + o_p(1)
\]

And the final two error terms are both negligible since we have that \( R = o(T), \sqrt{T/N} \to 0 \) and \( l/T^{1/4} \to 0 \) which all imply that both \( l/\sqrt{N} \to 0 \) and \( l/\sqrt{R} \to 0 \). Finally defining the population autocovariance \( \gamma_j = E \left[ (\bar{g}(\epsilon_{1,t+h}) - \bar{g}(\epsilon_{2,t+h})) (\bar{g}(\epsilon_{1,t+h+j}) - \bar{g}(\epsilon_{2,t+h+j})) \right] \) it follows from West (1996) that the last term equals:

\[= \sum_{j=-l}^{l} \gamma_j + \left[ \frac{1}{T} \sum_{t=1}^{T-1} \sum_{j=-l}^{l} (\bar{g}(\epsilon_{t+h}) - \bar{g}(\epsilon_{2,t+h})) (\bar{g}(\epsilon_{1,t+h+j}) - \bar{g}(\epsilon_{2,t+h+j})) - \sum_{j=-l}^{l} \gamma_j \right] + o_p(1)
\]

\[= \sum_{j=-l}^{l} \gamma_j + o_p(1)
\]

\[\Rightarrow V_e
\]

which shows (38) as required.

We do not repeat this proof for the other 5 variances and covariances as these are similar in logic. Specifically, we can show that:

\[
\text{Var}^* \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla \beta g \left( y_{t+h} - \tilde{F}_t^* \tilde{\beta}_t \right) \left( \tilde{\beta}_t^* - \tilde{\beta}_t \right) \right] = \left( 1 - \frac{1}{3\pi} \right) D_\beta V_F D_\beta + o_p(1) \tag{42}
\]

\[
\text{Var}^* \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nabla \gamma g \left( y_{t+h} - Z_t^* \gamma_t \right) \left( \gamma_t^* - \gamma_t \right) \right] = \left( 1 - \frac{1}{3\pi} \right) D_\gamma V_Z D_\gamma + o_p(1) \tag{43}
\]

\[
\text{Cov}^* \left[ \left( g \left( y_{t+h} - \tilde{F}_t^* \tilde{\beta}_t \right) - g \left( y_{t+h} - Z_t^* \gamma_t \right) \right) - \frac{1}{T} \sum_{j=1}^{T} (g(\tilde{\epsilon}_{1,j+h,t}) - g(\tilde{\epsilon}_{2,j+h,t})) \right] = 2 \left( 1 - \frac{1}{2\pi} \right) D_\beta C_{e,F} + o_p(1) \tag{44}
\]

\[
\text{Cov}^* \left[ \left( g \left( y_{t+h} - \tilde{F}_t^* \tilde{\beta}_t \right) - g \left( y_{t+h} - Z_t^* \gamma_t \right) \right) - \frac{1}{T} \sum_{j=1}^{T} (g(\tilde{\epsilon}_{1,j+h,t}) - g(\tilde{\epsilon}_{2,j+h,t})) \right] = 2 \left( 1 - \frac{1}{2\pi} \right) D_\gamma C_{e,Z} + o_p(1) \tag{45}
\]
\[ \text{Cov}^* \left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla g \left( y_{t+h}^* - \tilde{F}_j^* \tilde{\beta}_t \right) \left( \tilde{\beta}_t^* - \tilde{\beta}_t \right) \right], \] 
\[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla g \left( y_{t+h}^* - Z_t^* \gamma_t \right) \left( \tilde{\gamma}_t^* - \tilde{\gamma}_t \right) \] 
\[ = 2 \left( 1 - \frac{1}{3\pi} \right) D_{\beta} C_{F,Z} D_{\gamma}^\prime + o_p(1) \]  

(46)

Therefore combining the results in Equations (38) and (42)-(46) it follows that

\[ \text{Var}^* \left[ \tilde{S}_P^* \right] = \Omega + o_p(1) \]

as required. This completes the proof of Theorem 2.

References


Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.


