**School of Economics Working Paper 2016-05**

University of East Anglia

SCHOOL OF ECONOMICS

# Forecast Evaluation with Factor-Augmented Models

**by Jack Fosten***

***School of Economics, University of East Anglia**

**Abstract**

This paper extends Diebold-Mariano-West (DMW) forecast accuracy tests and corresponding bootstrap procedures 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 or recursive fashion, posing several new challenges which we address in this paper. The first main contribution is to show the convergence rates of factors estimated in different estimation windows, and provide conditions under which the asymptotic distribution of the DMW test statistic is not affected by factor estimation error. The next main contribution is to propose a new method for constructing bootstrap critical values for the DMW test with estimated factors when parameter estimation error is non-negligible. This is a non-trivial problem as the factors and factor-augmented model coefficients resulting from standard Principal Components are not sign-identified, resulting in arbitrary \sign-changing" across different estimation windows. We show that it is not possible to apply existing bootstrap methods in the case where the factors are not sign identified. We propose a novel new normalization for different windows of factor estimates, which adjusts them in such a way that their sign is matched across all estimation windows. We establish the first-order validity of a simple-to-implement block bootstrap procedure based on resampling the adjusted factor estimates 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, Di_usion Index, Factor Model, Predictive Ability.

**School of Economics**
**University of East Anglia**
**Norwich Research Park**
**Norwich NR4 7TJ**
**United Kingdom**
**www.uea.ac.uk/economics**

# Forecast Evaluation with Factor-Augmented Models

Jack Fosten*
University of East Anglia, UK.

September 27, 2016

**Abstract**

This paper extends Diebold-Mariano-West (DMW) forecast accuracy tests and corresponding bootstrap procedures 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 or recursive fashion, posing several new challenges which we address in this paper. The first main contribution is to show the convergence rates of factors estimated in different estimation windows, and provide conditions under which the asymptotic distribution of the DMW test statistic is not affected by factor estimation error. The next main contribution is to propose a new method for constructing bootstrap critical values for the DMW test with estimated factors when parameter estimation error is non-negligible. This is a non-trivial problem as the factors and factor-augmented model coefficients resulting from standard Principal Components are not sign-identified, resulting in arbitrary "sign-changing" across different estimation windows. We show that it is not possible to apply existing bootstrap methods in the case where the factors are not sign identified. We propose a novel new normalization for different windows of factor estimates, which adjusts them in such a way that their sign is matched across all estimation windows. We establish the first-order validity of a simple-to-implement block bootstrap procedure based on resampling the adjusted factor estimates 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

*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 Valentina 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.

of our knowledge, there has been no research formally addressing the effect of factor estimation on 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 solve in this paper.

The first main issue is that an increasing amount of empirical studies test the accuracy of factor-augmented models by treating 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 the existing results of Stock and Watson (2002a), Bai (2003) and Bai and Ng (2002, 2006) for estimating factors over the full sample, to the case of out-of-sample 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. Recently, Gonçalves et al. (2015) have 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 is that existing bootstrap approaches to the DMW test, such as Corradi and Swanson (2006), do not deal explicitly with the case where model regressors are generated. In the context of non-generated regressors, Corradi and Swanson (2006) show that it is possible to mimic the contribution of parameter estimation error using bootstrap analogues of the rolling or recursive parameter estimates, obtained by performing the pseudo out-of-sample method to variables which have been resampled over the full sample. However, applying this procedure becomes very problematic in the case where the generated regressors are the factors estimated by PCA, due to the lack of sign-identification of the factors and factor-augmented model coefficients. We show that sequences of factor-augmented model parameter estimates arising from the rolling or recursive estimation schemes are subject to "sign-changing," which is unavoidable in empirical applications. The implication is that it is not possible for the bootstrap parameter estimates to match the sign of the original out-of-sample parameter estimates. Without solving this issue, it is not possible to obtain valid bootstrap critical values for the test.

The second main contribution of the paper therefore addresses the bootstrap by first proposing a method to solve the sign-changing issue. We suggest a novel new normalization which is applied to the factors in every estimation window after the first. Specifically, the method adjusts the standard PCA estimates in each estimation window by using the estimate of the factor loadings from the first window as a normalization for a subset of the loadings in all subsequent windows. This has the effect of 'matching' the signs in every 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.

The final result of the paper establishes the first-order validity of block bootstrap critical values, calculated in the same way as Corradi and Swanson (2006), where the new normalization is used for the factors throughout the pseudo out-of-sample experiment. Our paper provides contributions not only to the literature of bootstrapping DMW tests, 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 and not the out-of-sample context.

The results of this paper will allow empirical researchers to perform robust inference in a wide range of different forecast comparisons involving 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 typically 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 provides a simple-to-implement procedure which covers the majority of commonly-used forecast comparison types, 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 organised as follows. Section 2 introduces the factor-augmented model and the construction of the DMW test statistic. Section 3 outlines the assumptions required and the asymptotic properties of the factor estimates and the DMW test statistic. Section 4 details the problems which occur when attempting to bootstrap this test statistic, provides a new normalization for rolling or recursive factor estimates, and outlines how to construct first-order valid bootstrap critical values. Section 5 provides simulation evidence to evaluate the performance of this bootstrap procedure. Finally, Section 6 provides an empirical illustration of forecasting U.S. CPI inflation and Section 7 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:

$$y_{t+h} = F_t'\beta + \epsilon_{1,t+h} \tag{1}$$

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:

$$X_t = \Lambda F_t + u_t \tag{2}$$

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. Stock and Watson (2002a,b) 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):

$$H_0 : \mathrm{E}\left[g\left(\epsilon_{1,t+h}\right) - g\left(\epsilon_{2,t+h}\right)\right] = 0 \tag{3}$$

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 $g\left(\epsilon_{1,t+h}\right)$ and $g\left(\epsilon_{2,t+h}\right)$, given some loss function $g(\cdot)$. This loss function may be differentiable as in West (1996), such as mean squared forecast error (MSFE) where $g\left(\epsilon_{t+h}\right) = \epsilon_{t+h}^2$, or non-differentiable as in McCracken (2000), such as mean absolute error (MAE) where $g\left(\epsilon_{t+h}\right) = \left|\epsilon_{t+h}\right|$. 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 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.

[1]For example, often $W_t$ contains lags of the dependent variable, in which case the model would be called a factor-augmented autoregression.

(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 $Z_t$:

$$y_{t+h} = Z_t'\gamma + \epsilon_{2,t+h} \tag{4}$$

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 $X_t$.[2] 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 main estimation schemes, as described in West (1996), are rolling estimation which uses a fixed window length $R$ of data from $t - R + 1$ to $t$ for each horizon $t = R, ..., T$, and recursive estimation which uses all observations from 1 to $t$. 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). Since rolling and recursive factor estimation are both used in the empirical literature (see the meta-analysis of Eickmeier and Ziegler (2008) for a discussion), we write both the rolling and recursive PCA estimation procedures as follows:

[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.

$$\text{Rolling} \qquad \left(\widehat{F}^{(t)}, \widehat{\Lambda}^{(t)}\right) = \arg\min_{\Lambda,F} \frac{1}{NR} \sum_{i=1}^{N} \sum_{j=t-R+1}^{t} \left(X_{ij} - \lambda_i' F_j\right)^2 \tag{5}$$

or:

$$\text{Recursive} \qquad \left(\widehat{F}^{(t)}, \widehat{\Lambda}^{(t)}\right) = \arg\min_{\Lambda,F} \frac{1}{Nt} \sum_{i=1}^{N} \sum_{j=1}^{t} \left(X_{ij} - \lambda_i' F_j\right)^2 \tag{6}$$

Note that we superscript the factors by $t$ to denote which estimation window has been used in their estimation. It is for simplicity that we do not repeatedly differentiate the notation between rolling and recursive estimates, and simply use $\widehat{F}^{(t)}$ rather than, say, $\widehat{F}^{(t,rol)}$ and $\widehat{F}^{(t,rec)}$. In what follows we will tend to focus on the case of rolling estimation of the factors, as the recursive case is very similar and in any case is explicitly dealt with in the paper of Gonçalves et al. (2015).

We identify the factors in the rolling case using the assumptions that $F^{(t)\prime}F^{(t)}/R = I_r$, where $I_r$ is the $r \times r$ identity matrix, and that $\Lambda'\Lambda/N$ is diagonal.[3] These 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 $\widehat{F}^{(t)}$, 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 $X^{(t)}X^{(t)\prime}/RN$. The factor estimates are correspondingly normalized:

**Normalization N1:** $\widehat{F}^{(t)\prime}\widehat{F}^{(t)}/R = I_r$

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

Having estimated the factors by either rolling or recursive estimation, for each $R \leq t \leq T$ we can estimate the factor-augmented model coefficients in Model 1 as follows:

$$\text{Rolling} \qquad \widehat{\beta}_t = \arg\min_{\beta} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left(y_{j+h} - \widehat{F}_j^{(t)\prime}\beta\right)^2 \tag{7}$$

or:

$$\text{Recursive} \qquad \widehat{\beta}_t = \arg\min_{\beta} \frac{1}{t} \sum_{j=1}^{t-h} \left(y_{j+h} - \widehat{F}_j^{(t)\prime}\beta\right)^2 \tag{8}$$

and the parameters in the benchmark Model 2, $\gamma$, are estimated in the same way. These parameter estimates are used to make model-based forecasts which give rise to the forecast errors $\widehat{\epsilon}_{1,t+h} = y_{t+h} - \widehat{F}_t^{(t)\prime}\widehat{\beta}_t$ and $\widehat{\epsilon}_{2,t+h} = y_{t+h} - Z_t'\widehat{\gamma}_t$ for each $R \leq t \leq T$, which are then used to calculate the

[3]The recursive scheme uses $F^{(t)\prime}F^{(t)}/t = I_r$.

[4]As noted by Bai and Ng (2008), estimating $\widehat{F}^{(t)}$ from $X^{(t)}X^{(t)\prime}$ using I1 will give the same common component as estimating $\widehat{\Lambda}^{(t)}$ from $X^{(t)\prime}X^{(t)}$ and normalizing such that $\Lambda'\Lambda/N = I_k$ and $F^{(t)\prime}F^{(t)}/R$ being symmetric.

(non-scaled)[5] Diebold-Mariano-West test statistic:

$$\widehat{S}_P = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g\left(\widehat{\epsilon}_{1,t+h}\right) - g\left(\widehat{\epsilon}_{2,t+h}\right) \right) \tag{9}$$

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 $\widehat{S}_P$ on estimated factors as well as estimated model coefficients. The estimation of the factors across different windows is new to the literature on out-of-sample testing and is what we formally analyse in the next section.

# 3 Asymptotic Theory

We now detail the assumptions required to show the properties of the factor estimates in different windows, and the asymptotic distribution of the Diebold-Mariano-West statistic, $\widehat{S}_P$. 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 $\|M\| = \left(tr\left(M'M\right)\right)^{1/2}$. 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)** $(F_t', Z_t', \epsilon_{1,t+h}, \epsilon_{2,t+h}, u_{1t,\dots,}u_{Nt})$ is strong mixing with mixing coefficients of size $-3d/\left(d-1\right)$ for some $d > 1$;

**(b)** $(F_t', Z_t', \epsilon_{1,t+h}, \epsilon_{2,t+h}, u_{1t,\dots,}u_{Nt})$ is strictly stationary.

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

**(a)** $\mathrm{E}\|F_t\|^{4d} \leq C$, and $\frac{1}{R}\sum_{j=t-R+1}^{t} F_j F_j' \xrightarrow{p} \Sigma_F > 0$ 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 $\mathrm{E}\|\lambda_i\|^4 \leq C$. 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)** $\mathrm{E}(u_{it}) = 0$, $\mathrm{E}|u_{it}|^{8d} \leq C$;

**(b)** $E\left(\frac{1}{N}\sum_{i=1}^{N} u_{is}u_{it}\right) = \gamma_{st}$, $|\gamma_{ss}| \leq C$ 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 $\frac{1}{P}\sum_{t=R}^{T}\sum_{k=t-R+1}^{t} |\gamma_{tk}| \leq C$;

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

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

[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 $\widehat{S}_P$ unscaled and its variance will be calculated along the lines of West (1996).

**(e)** $E\left(u_{ik}u_{jh}\right)=\tau_{ij,kh}$ and $\sup_t \frac{1}{NR}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=t-R+1}^{t}\sum_{h=t-R+1}^{t}\left|\tau_{ij,kh}\right|\leq C$.

**Assumption 4:** (Dependence between idiosyncratic errors and loadings, regressions errors and variables)
**(a)** For all $s$, $E\left\|\sup_t \frac{1}{\sqrt{NR}}\sum_{k=t-R+1}^{t}\sum_{i=1}^{N}F_k\left(u_{is}u_{ik}-\mathrm{E}\left(u_{is}u_{ik}\right)\right)\right\|^2\leq C$
**(b)** For all $s$, and $h\geq 0$, $E\left|\sup_{t\geq R}\frac{1}{\sqrt{NR}}\sum_{k=t-R+1}^{t-h}\sum_{i=1}^{N}\left(u_{is}u_{ik}-\mathrm{E}\left(u_{is}u_{ik}\right)\right)\epsilon_{1,k+h}\right|^2\leq C$
**(c)** $E\left\|\frac{1}{\sqrt{RN}}\sup_t\sum_{j=t-R+1}^{t-h}\Lambda' u_j\epsilon_{1,j+h}\right\|^2\leq C$, and $E\left(\lambda_i u_{it}\epsilon_{1,t+h}\right)=0$ for all $(i,t)$
**(d)** $E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\sup_t\frac{1}{\sqrt{R}}\sum_{j=t-R+1}^{t}F_j u_{ij}\right\|\right)^2\leq C$, and $E\left(F_t u_{it}\right)=0$ for all $(i,t)$
**(e)** $E\left\|\frac{1}{\sqrt{RN}}\sup_t\sum_{j=t-R+1}^{t}F_j u_j'\Lambda\right\|^2\leq C$, and $E\left(\lambda_i u_{it}F_t\right)=0$ for all $(i,t)$
**(f)** $E\left(\frac{1}{R}\sup_t\sum_{j=t-R+1}^{t}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i u_{ij}\right\|^2\right)\leq C$, and $E\left(\lambda_i u_{ij}\right)=0$ for all $(i,t)$

**Assumption 5:** (Forecast model moments)
**(a)** $\mathrm{E}|\epsilon_{1,t+h}|^{4d}\leq C$ and $\mathrm{E}|\epsilon_{2,t+h}|^{4d}\leq C$ where $d>1$;
**(b)** $\mathrm{E}\|Z_t\|^{4d}\leq C$ and for all $t$, $\frac{1}{R}\sum_{j=t-R+1}^{t}Z_jZ_j'\xrightarrow{p}\Sigma_Z>0$;
**(c)** $\mathrm{E}(F_t\epsilon_{1,t+h})=0$ and $\mathrm{E}(Z_t\epsilon_{2,t+h})=0$.

**Assumption 6:** (Functional form of loss function)
**(a1)** For the $(2r\times 1)$ vector $\theta=[F_t',\beta']'$, in an open neighbourhood $N_1$ around $\theta$, and with probability one, the function $g\left(\epsilon_{1,t+h}\right)=g\left(y_{t+h}-F_t'\beta\right)$ 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 $g\left(\epsilon_{2,t+h}\right)=g\left(y_{t+h}-Z_t'\gamma\right)$ is measurable and twice continuously differentiable with respect to $\gamma$;
**(b1)** For all $t$, $\sup_{\theta\in N_1}\left\|\nabla_\theta^2 g\left(\epsilon_{1,t+h}\right)\right\|\leq m_{1t}$, for a measurable $m_{1t}$ with $\mathrm{E}(m_{1t})\leq C$;
**(b2)** Also for all $t$, $\sup_{\gamma\in N_2}\left\|\nabla_\gamma^2 g\left(\epsilon_{2,t+h}\right)\right\|\leq m_{2t}$, for a measurable $m_{2t}$ with $\mathrm{E}(m_{2t})\leq C$.

**Assumption 7:** (Test statistic and score assumptions)
**(a)** $\mathrm{E}\left\|\left[\nabla_\beta g\left(\epsilon_{1,t+h}\right),\nabla_\gamma g\left(\epsilon_{2,t+h}\right),g\left(\epsilon_{1,t+h}\right),g\left(\epsilon_{2,t+h}\right),F_t\epsilon_{1,t+h},Z_t\epsilon_{2,t+h}\right]'\right\|^{4d}\leq C$, where $d>1$;
**(b)** Denote $L_{t+h}=g\left(\epsilon_{1,t+h}\right)-g\left(\epsilon_{2,t+h}\right)$ and $E\left(L_{t+h}\right)$ its expectation. Furthermore let $V_\epsilon=\sum_{j=-\infty}^{\infty}E\left(L_{t+h}-\mathrm{E}\left(L_{t+h}\right)\right)\left(L_{t+h-j}-\mathrm{E}\left(L_{t+h-j}\right)\right)$. Then $V_\epsilon>0$.
**(c)** $\mathrm{E}\|\nabla_F g\left(\epsilon_{1,t+h}\right)\|^{4d}\leq C$ and $D_F=E\left(\nabla_F g\left(\epsilon_{1,t+h}\right)\right)$

**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.[6]

[6]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. To reiterate the statement above, assumptions for the recursive estimation procedure are not repeated here but are easily derived from those above 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). 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.[7]

In order to show the asymptotic distribution of the test statistic $\widehat{S}_P$, we first rely on several key Lemmas on the properties of the factors in different estimation windows. In the case of the rolling estimation scheme, the key Lemma is:

**Lemma A:** *Under Normalization N1 and Assumptions 1-8:*
**(i)**

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

where the rotation matrix, $H_{NR}^{(t)}$, is time-dependent and has the same form as Bai (2003):

$$H_{NR}^{(t)} = \widehat{V}^{(t)-1} \frac{\widehat{F}^{(t)\prime} F^{(t)}}{R} \frac{\Lambda' \Lambda}{N} \tag{10}$$

This Lemma shows is the equivalent of Bai and Ng (2002) Theorem 1, extended to the case of rolling estimation of the factors, and shows that the rolling window average of the squared difference between the rolling factor estimates $\widehat{F}_j^{(t)}$ and the rotated factors $H_{NR}^{(t)} F_j$ vanishes at a rate $\min\{N, R\}$, uniformly over the rolling windows $t = R, ..., T$. It is possible to show that the

[7] For cases where the rolling window length is kept constant, even asymptotically, see the conditional predictive ability approach of Giacomini and White (2006).

same Lemma holds over recursive windows of the factors. However, as these results are already shown by Gonçalves et al. (2015) and almost identical to the rolling case, we do not repeat these here. The following result shows the asymptotic distribution of the test statistic $\widehat{S}_P$:

**Theorem 1:** *Under Normalization N1, Assumptions 1-8 and under $H_0$:*

$$\widehat{S}_P \xrightarrow{d} N(0, \Omega)$$

*where:*

$$\begin{aligned}\Omega = &V_\epsilon + \theta_1 D_\beta V_F D'_\beta + 2\theta_2 D_\beta C_{\epsilon,F} \\ &+ \theta_1 D_\gamma V_Z D'_\gamma - 2\theta_2 D_\gamma C_{\epsilon,Z} - 2\theta_1 D_\beta C_{F,Z} D'_\gamma\end{aligned} \tag{11}$$

*and $\theta_1$ and $\theta_2$ depend on the use of rolling or recursive estimation as follows:*

| | $\theta_1$ | $\theta_2$ |
|---|---|---|
| *Recursive:* | $2\left(1 - \frac{1}{\pi} ln(1+\pi)\right)$ | $1 - \frac{1}{\pi} ln(1+\pi)$ |
| *Rolling* ($\pi \geq 1$)*:* | $1 - \frac{1}{3\pi}$ | $1 - \frac{1}{2\pi}$ |
| *Rolling* ($\pi < 1$)*:* | $\pi - \frac{\pi^2}{3}$ | $\frac{\pi}{2}$ |

*Also:*

$$\begin{aligned}V_\epsilon = \sum_{j=-\infty}^{\infty} \mathrm{E}&\left[\left(g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})\right) - \mathrm{E}\left(g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})\right)\right. \\ &\left.\times \left(g(\epsilon_{1,t+h+j}) - g(\epsilon_{2,t+h+j})\right) - \mathrm{E}\left(g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})\right)\right]\end{aligned}$$

$$V_F = \left(H^\dagger \Sigma_F H^{\dagger\prime}\right)^{-1} H^\dagger \left(\sum_{j=-\infty}^{\infty} \mathrm{E}\left[F_t \epsilon_{1,t+h} \epsilon_{1,t+h+j} F'_{t+j}\right]\right) H^{\dagger\prime} \left(H^\dagger \Sigma_F H^{\dagger\prime}\right)^{-1},$$

$$V_Z = \Sigma_Z^{-1} \left(\sum_{j=-\infty}^{\infty} \mathrm{E}\left[Z_t \epsilon_{2,t+h} \epsilon_{2,t+h+j} Z'_{t+j}\right]\right) \Sigma_Z^{-1},$$

$$C_{\epsilon,F} = \left(\sum_{j=-\infty}^{\infty} \mathrm{E}\left[\left(\left(g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})\right) \epsilon_{1,t+h+j} F'_{t+j}\right)\right]\right) H^{\dagger\prime} \left(H^\dagger \Sigma_F H^{\dagger\prime}\right)^{-1},$$

$$C_{\epsilon,Z} = \left(\sum_{j=-\infty}^{\infty} \mathrm{E}\left[\left(g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})\right) \epsilon_{2,t+h+j} Z'_{t+j}\right]\right) \Sigma_Z^{-1},$$

$$C_{F,Z} = \left(H^\dagger \Sigma_F H^{\dagger\prime}\right)^{-1} H^\dagger \left(\sum_{j=-\infty}^{\infty} \mathrm{E}\left[F_t \epsilon_{1,t+h} \epsilon_{2,t+h+j} Z'_{t+j}\right]\right) \Sigma_Z^{-1},$$

*where the rotation matrix $H^\dagger = V^{-1} Q \Sigma_\Lambda$, where $V$, $Q$ and $\Sigma_\Lambda$ are the probability limits of $\widehat{V}^{(t)}$,*

$\widehat{F}^{(t)\prime}F^{(t)}/R$ *and* $\Lambda'\Lambda/N$*, and we define* $D_\beta = E\left(\nabla_\beta g\left(\epsilon_{1,t+h}\right)\right)$ *and* $D_\gamma = E\left(\nabla_\gamma g\left(\epsilon_{2,t+h}\right)\right)$.

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).

# 4 Bootstrap Inference

## 4.1 Introduction

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).

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. We now detail several challenges which result when using generated regressors, particularly factors estimated by PCA, which are subject to a lack of sign-identification across different estimation windows.

## 4.2 Which Factors to Resample?

In order to apply the bootstrap procedure of Corradi and Swanson (2006) in the case where PEE is non-negligible, we cannot simply resample the forecast errors which enter the DMW test statistic, we need to be able to resample the estimated factors and other model variables in order to mimic the estimation of the model parameters. In other words, even though Assumption 8(a) maintains the Bai and Ng (2006) assumption that $\sqrt{T}/N \to 0$, which ensures that factor estimation error is negligible, we still need to be able to resample the factors over the full sample in order for the bootstrap to mimic the contribution of PEE.[8]

However, given that the pseudo out-of-sample approach used in this paper yields different factor estimated for different estimation windows, it is not obvious how to proceed with this bootstrap

[8]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.

methodology, as there is a multiplicity of overlapping windows of generated factors. For rolling and recursive PCA as in Equations (5) and (6), the overlapping estimates can be displayed in the following matrix:

$$\widehat{F}^{rol} = \begin{bmatrix} \widehat{F}_1^{(R)} & \widehat{F}_2^{(R)} & \cdots & \widehat{F}_R^{(R)} \\ \widehat{F}_2^{(R+1)} & \widehat{F}_3^{(R+1)} & \cdots & \widehat{F}_{R+1}^{(R+1)} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{F}_P^{(T)} & \widehat{F}_{P+1}^{(T)} & \cdots & \widehat{F}_T^{(T)} \end{bmatrix} \tag{12}$$

$$\widehat{F}^{rec} = \begin{bmatrix} \widehat{F}_1^{(R)} & \widehat{F}_2^{(R)} & \cdots & \widehat{F}_R^{(R)} & - & - & - \\ \widehat{F}_1^{(R+1)} & \widehat{F}_2^{(R+1)} & \cdots & \widehat{F}_R^{(R+1)} & \widehat{F}_{R+1}^{(R+1)} & - & - \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & - \\ \widehat{F}_1^{(T)} & \widehat{F}_2^{(T)} & \cdots & \widehat{F}_R^{(T)} & \widehat{F}_{R+1}^{(T)} & \cdots & \widehat{F}_T^{(T)} \end{bmatrix} \tag{13}$$

Neither the rolling nor recursive estimates give rise to a single factor estimate which we are able to resample over the full sample from 1 to $T$. The only exception is, of course, the full sample factor estimate which coincides with the final window from recursive estimation. Therefore the most obvious candidate to use for bootstrap resampling is the set of full sample factor estimates $\widehat{F}_1^{(T)}, ..., \widehat{F}_T^{(T)}$, as used in Corradi and Swanson (2014).[9] In spite of the finite sample differences in the estimated factors across different windows, for example $\widehat{F}_2^{(R)}$ and $\widehat{F}_2^{(R+1)}$, in principle a bootstrap procedure based on resampling the full sample factors should be valid because, as mentioned above, we assume that factor estimation error is asymptotically negligible.

However, there is one final problem we must address before being able to derive valid critical values. In addition to the factors differing across estimation windows due to finite sample differences, they may also differ due to a change in sign, resulting from the lack of sign identification of PCA. This causes a problem in the Corradi and Swanson (2006) bootstrap methodology, as outlined in the next section.

### 4.3 Lack of Sign-Identification and Failure of Existing Bootstrap Methods

Suppose that we proceed to use the bootstrap procedure of Corradi and Swanson (2006) and decide to resample from the full-sample factor estimates $\widehat{F}^{(T)} = [\widehat{F}_1^{(T)}, ..., \widehat{F}_T^{(T)}]$. That is, we obtain the block bootstrap resamples $\{y^*_{t+h}, \widehat{\mathcal{F}}^*_t, Z^*_t\}_{t=1}^T$ by resampling the variables $\{y_{t+h}, \widehat{F}_t^{(T)}, Z_t\}_{t=1}^T$ over full sample from $t = 1, ..., T$. The precise details of the Corradi and Swanson (2006) procedure are described later. Then it is suggested to obtain bootstrap estimates $\widehat{\beta}^*_t$ and $\widehat{\gamma}^*_t$ of the parameters in Equations 1 and 4 by applying the pseudo out-of-sample method to the resampled data, but using a recentering term which accounts for the fact that these bootstrap estimates are based on

[9] Another potential candidate would be to join together a finite number of different rolling windows of factor estimates to produce one full-sample factor estimate. This was explored in previous versions of the paper, which are available on request.

resamples taken from the full sample and not rolling or recursive windows of the data.

Finally, Corradi and Swanson (2006) show that the bootstrap version of the DMW test statistic, which uses the bootstrap errors $\widehat{\epsilon}^*_{1,t+h} = y^*_{t+h} - \widehat{\mathcal{F}}^{*\prime}_t \widehat{\beta}^*_t$ and $\widehat{\epsilon}^*_{2,t+h} = y^*_{t+h} - Z^{*\prime}_t \widehat{\gamma}^*_t$, must be recentered around the test statistic over the full sample as follows:

$$\widehat{S}^*_P = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \left( g\left(\widehat{\epsilon}^*_{1,t+h}\right) - g\left(\widehat{\epsilon}^*_{2,t+h}\right) \right) - \frac{1}{T} \sum_{j=1}^{T} \left( g\left(\widehat{\epsilon}_{1,j+h,t}\right) - g\left(\widehat{\epsilon}_{2,j+h,t}\right) \right) \right) \tag{14}$$

where the error term in the recentering, $\widehat{\epsilon}_{1,j+h,t}$, is constructed using the rolling OLS estimators $\widehat{\beta}_t$ from the original pseudo out-of-sample procedure and the full sample factors, $\widehat{F}^{(T)}_j$, which were used for resampling:

$$\widehat{\epsilon}_{1,j+h,t} = y_{j+h} - \widehat{F}^{(T)\prime}_j \widehat{\beta}_t \tag{15}$$

The use of $\widehat{\beta}_t$ in the recentering term results from taking an expansion of $\widehat{\beta}^*_t$ in the bootstrap test statistic around $\widehat{\beta}_t$, which is required to show that $\widehat{\beta}^*_t$ behaves like $\widehat{\beta}_t$ asymptotically.

However, with the error terms in Equation (15) in mind, we encounter a serious problem with this bootstrap resampling procedure, resulting from the lack of sign identification briefly mentioned in the set-up section. It is well known that the PCA estimates of the factors and therefore factor augmented model coefficients are only identified up to a rotation matrix *and* a change in sign. In proving Theorem 1, neither the rotation nor sign of the estimated factors in a particular window was of any consequence as it cancelled out with the equivalent rotation and sign of the factor-augmented model coefficients which had been estimated in the same window of data. However, in Equation (15), since the full sample factors $\widehat{F}^{(T)}$ and $\widehat{\beta}_t$ have been estimated using different windows of data, we cannot be sure that the sign of the factors will cancel out with the factor augmented model coefficients.

To see this, we show the following Proposition:

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

$$\sup_t \frac{1}{R} \sum_{j=t-R+1}^{t} \left\| \widehat{F}^{(t)}_j - H^{\dagger}_t F_j \right\|^2 = o_p(1) \tag{16}$$

*and therefore:*

$$\sup_t \left\| \widehat{\beta}_t - H^{\dagger\prime -1}_t \beta \right\| = o_p(1) \tag{17}$$

*where* $H^{\dagger}_t = S^{(t)} H^{\dagger}$, $S^{(t)} = diag(\pm 1, ..., \pm 1)$ *is any sign matrix and* $H^{\dagger}$ *is described in Theorem 1.*

Proposition 1a shows that across different rolling windows, (the same result holds for recursive windows), the factors and factor-augmented model coefficients are consistent up to the same rotation matrix, but have a column sign $S^{(t)}$ which is potentially different across windows.[10]

[10]Assumptions leading to the case where $H^{\dagger} = I_r$ 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 $H^{\dagger}_t = S^{(t)}$.

In the context of the recentering term based on the errors in Equation (15), the implications of Proposition 1a are that, with some abuse of notation, if the factors in $\widehat{F}_j^{(T)}$, which we use to resample, are consistent for $+H^{\dagger}F_j$, whereas some rolling OLS estimator $\widehat{\beta}_t$ is consistent for $-H^{\dagger\prime-1}\beta$, then the limit of $\widehat{\epsilon}_{1,j+h,t}$ will be $(y_{j+h} + F_j'\beta)$ and not $(y_{j+h} - F_j'\beta)$. This feature leads to an incorrect recentering after the errors enter the function $g(.)$ and it will not be possible to show that the bootstrap resampling scheme results in valid bootstrap critical values for the DMW test. This is explained further in the next section, where a graphical depiction is given of how sign-changing affects the recentering term in Equations (14) and (15).

We now propose a new normalization for the estimated factors which eliminates the problem of the sign-changing.

### 4.4 Normalizing Rolling and Recursive PCA Estimates: a New Approach

To overcome the problem of sign-changing in the rolling and recursive 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 structure onto the factor loadings. To our knowledge there have been no papers looking at the normalization of factors across different estimation windows.

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:

$$\Lambda = \begin{bmatrix} \underset{r\times r}{\Lambda_1} \\ \underset{(N-r)\times r}{\Lambda_2} \end{bmatrix}$$

Now take the estimate of $\Lambda_1$ from the first window, $\widehat{\Lambda}_1^{(R)}$, and use this $r \times r$ matrix as a normalization in the remaining windows. For each $R \leq t \leq T$ this normalization is written:

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

The reason for using $\widehat{\Lambda}_1^{(R)}$ from the first window is to keep the exercise truly out-of-sample, in that $\widehat{\Lambda}_1^{(R)}$ is available to the researcher in each estimation window after $t = R$. One could instead use the approach of Bai and Ng (2013) who propose to assign particular values of 1's and 0's to certain factor loadings in $\Lambda_1$. We prefer to remain agnostic on the nature of the factor loadings in this paper. In partitioning the $\Lambda$ matrix 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.[11]

[11] It is not advisable to simply choose $\Lambda_1$ corresponding to the first $r$ variables listed in a given dataset. In the

To implement Normalization N2, in the first window ($t = R$) the factors and loadings are normalized using the standard PCA normalization N1. In all subsequent windows ($t > R$) the $r \times r$ sub-matrix $\Lambda_1$ is normalized to be equal to its standard PCA estimate from the first 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, which we call $\widetilde{\Lambda}^{(t)}$ and $\widetilde{F}^{(t)}$:

$$\widetilde{\Lambda}^{(t)} = \widehat{\Lambda}^{(t)} \left(\widehat{\Lambda}_1^{(t)}\right)^{-1} \widehat{\Lambda}_1^{(R)}, \qquad \widetilde{F}^{(t)} = \widehat{F}^{(t)} \widehat{\Lambda}_1^{(t)\prime} \left(\widehat{\Lambda}_1^{(R)\prime}\right)^{-1} \tag{18}$$

This, in turn, gives rise to the adjusted OLS estimates of the factor-augmented model coefficients:

$$\widetilde{\beta}_t = \left(\widetilde{F}^{(t)\prime} \widetilde{F}^{(t)}\right)^{-1} \widetilde{F}^{(t)\prime} y^{(t)} \tag{19}$$

The following Proposition provides results for the convergence of these adjusted factors and factor-augmented model coefficients, $\widetilde{F}^{(t)}$ and $\widetilde{\beta}_t$:

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

$$\sup_t \frac{1}{R} \sum_{j=t-R+1}^{t} \left\| \widetilde{F}_j^{(t)} - H_R^{\dagger} F_j \right\|^2 = o_p(1) \tag{20}$$

*and therefore:*

$$\sup_t \left\| \widetilde{\beta}_t - H_R^{\dagger\prime -1} \beta \right\| = o_p(1) \tag{21}$$

*where $H_R^{\dagger}$ is the limiting rotation matrix from the first window, and is 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 (as before, an identical result holds for recursive estimation). The implication of Proposition 1b, relative to Proposition 1a, is that by normalizing the estimates using 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 $H_t^{\dagger}$ for $t = R$, so our method has the effect of matching the sign of each rolling window to that of the first window.

Note that the change in estimation of the factors under Normalization N2 does not alter the asymptotic distribution derived in Theorem 1. This is because the test statistic $\widehat{S}_P$ is numerically identical to the equivalent test statistic using the adjusted factor estimates. Namely, with the forecast errors where $\widetilde{\epsilon}_{1,t+h} = y_{t+h} - \widetilde{F}_t^{(t)\prime} \widetilde{\beta}_t$ and $\widetilde{\epsilon}_{2,t+h} = y_{t+h} - Z_t' \widetilde{\gamma}_t$, we can re-define the test statistic as:

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.

$$\widetilde{S}_P = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g\left(\widetilde{\epsilon}_{1,t+h}\right) - g\left(\widetilde{\epsilon}_{2,t+h}\right) \right) \tag{22}$$

and the same result as Theorem 1 holds:

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

$$\widetilde{S}_P \xrightarrow{d} N\left(0, \Omega\right)$$

*where* $\Omega$ *is exactly as stated in Theorem 1.*

However, returning to the recentering term based on the errors in Equation (15), Proposition 1b suggests that no problems now arise in the bootstrap procedure based on resampling the adjusted full-sample factor estimates $\widetilde{F}^{(T)}$, since for each $t$, we are assured that the limiting rotation and sign of $\widetilde{F}^{(T)}$ matches that of $\widetilde{\beta}_t$ in each and every window, and so the recentering terms will result in valid bootstrap critical values. The new recentering error term can be written:

$$\widetilde{\epsilon}_{1,j+h,t} = y_{j+h} - \widetilde{F}_j^{(T)\prime} \widetilde{\beta}_t \tag{23}$$

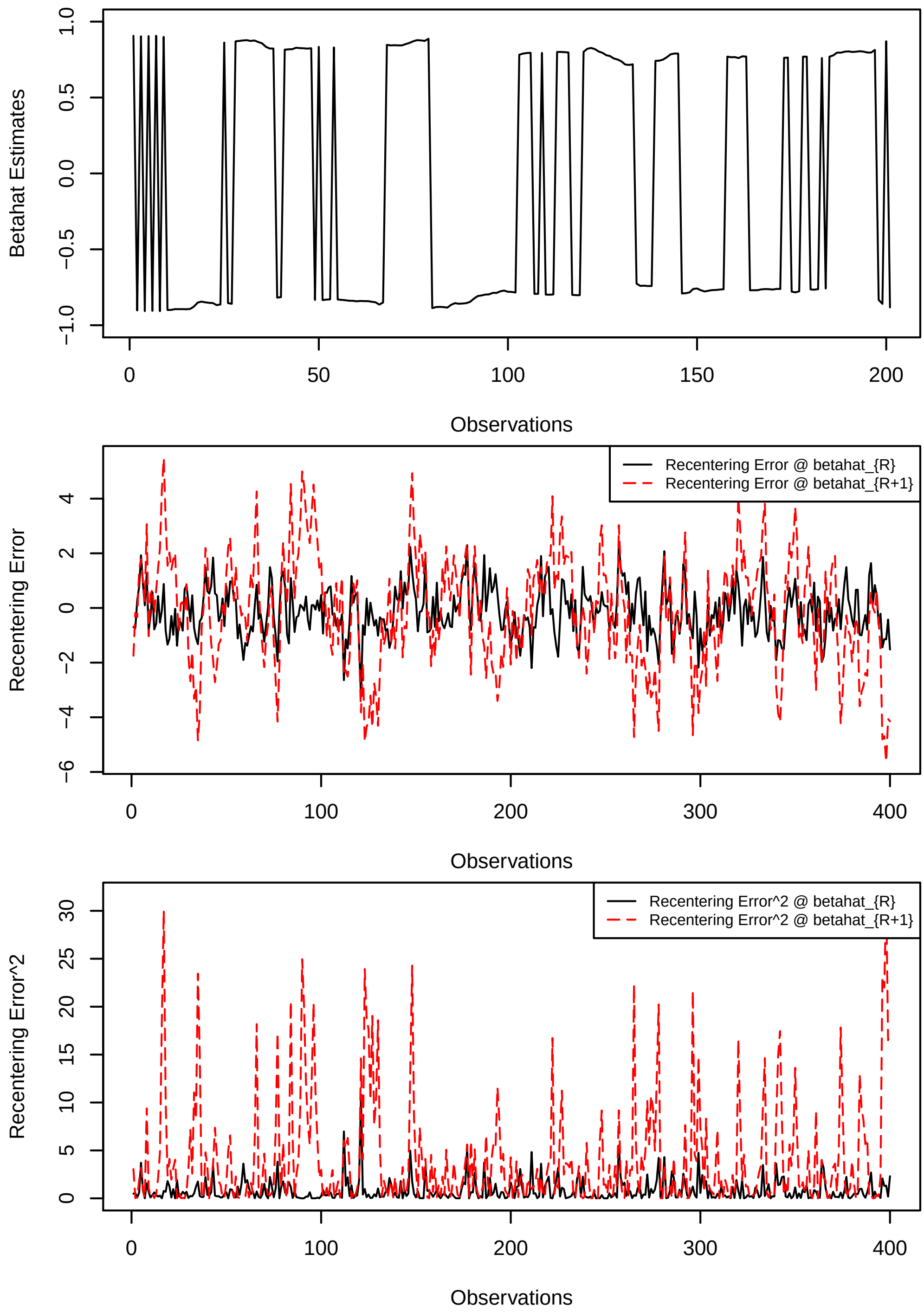
To gain some intuition about the concept of sign-changing and how they affect the recentering terms, Figures 1 and 2 use simulated data to display some graphical evidence concerning the rolling OLS estimators based on the non-adjusted and adjusted PCA factor estimates. The top panels of Figures 1 and 2 display the evolution of $\widehat{\beta}_t$ and $\widetilde{\beta}_t$ across rolling windows in a controlled environment, which clearly displays the results of Propositions 1a and 1b regarding sign-changing and sign-robustness. The second panel of Figure 1 plots the recentering error term $\widehat{\epsilon}_{1,j+h,t}$ evaluated at the OLS estimators from the first two windows, $t = R$ and $t = R + 1$. This shows that the sign-change between $\widehat{\beta}_R$ and $\widehat{\beta}_{R+1}$ causes very large differences in the recentering error term, which in the bottom panel of Figure 1 causes large bias in the recentering term $g(\widehat{\epsilon}_{1,j+h,t})$. The equivalent panels in Figure 2 show that this does not happen when using the sign-robust factor estimates and factor augmented model coefficient estimates.

### 4.5 Bootstrap Critical Values

We follow the block bootstrap procedure of Corradi and Swanson (2006) to obtain bootstrap critical values. We obtain bootstrap samples, $(y^*_{t+h}, \widetilde{\mathcal{F}}^*_t, Z_t)_{t=1}^T$, for all variables in Models 1 and 2, $(y_{t+h}, \widetilde{F}_t^{(T)}, Z_t)_{t=1}^T$ 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 $j = 1, ..., b$ and by forming $b$ blocks of $y_{t+h}$, $\widetilde{\mathcal{F}}_t$ and $Z_t$ such that $[y^*_{1+h}, ..., y^*_{T+h}] = [y_{I_1+1+h}, ..., y_{I_1+l+h}, ..., y_{I_b+1+h}, ..., y_{I_b+l+h}]$ , $[\widetilde{\mathcal{F}}^*_1, ..., \widetilde{\mathcal{F}}^*_T] = [\widetilde{F}^{(T)}_{I_1+1}, ..., \widetilde{F}^{(T)}_{I_1+l}, ..., \widetilde{F}^{(T)}_{I_b+1}, ..., \widetilde{F}^{(T)}_{I_b+l}]$ and $[Z^*_1, ..., Z^*_T] = [Z_{I_1+1}, ..., Z_{I_1+l}, ..., Z_{I_b+1}, ..., Z_{I_b+l}]$.

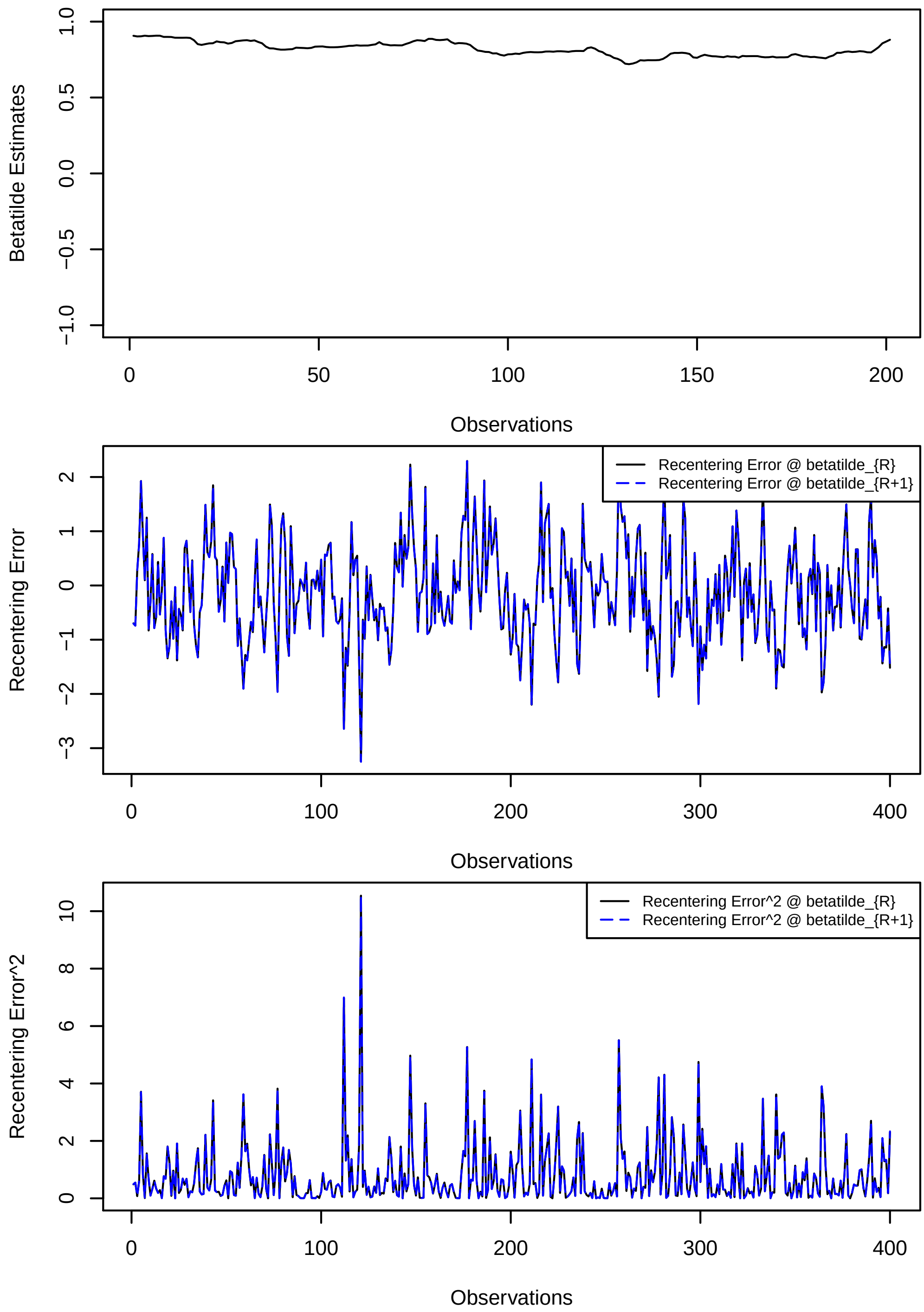
We proceed by applying the rolling out-of-sample methodology on this resampled data. Bootstrap estimation of the parameters $\beta$ and $\gamma$ proceeds by recentering the OLS criterion function

**Figure 1:** Properties of $\widehat{\beta}_t$ from non-adjusted PCA factors and corresponding recentering term.




**Notes:** The three graphs depict (i) the rolling estimates of $\widehat{\beta}_t$ plotted over time for $t = R, ..., T$, (ii) the recentering error from Equation (15), $\widehat{\epsilon}_{1,j+h,t}$, plotted over all periods $j = 1, ..., T$, evaluated at $t = R$ and $t = R + 1$, and (iii) the recentering term, $g(\widehat{\epsilon}_{1,j+h,t})$, for the same errors using squared error loss $g(\epsilon) = \epsilon^2$. The data was simulated with $T = 400$ and $N = 200$ with $r = 1$ factor for $X_t$ and such that $y_t = 1 + F_t + \epsilon_t$, full details of which are available upon request.

**Figure 2:** Properties of $\widetilde{\beta}_t$ from adjusted PCA factors and corresponding recentering term.




**Notes:** The three graphs depict (i) the rolling estimates of $\widetilde{\beta}_t$ plotted over time for $t = R, ..., T$, (ii) the recentering error from Equation (23), $\widetilde{\epsilon}_{1,j+h,t}$, plotted over all periods $j = 1, ..., T$, evaluated at $t = R$ and $t = R + 1$, and (iii) the recentering term, $g(\widetilde{\epsilon}_{1,j+h,t})$, for the same errors using squared error loss $g(\epsilon) = \epsilon^2$. The data was simulated with $T = 400$ and $N = 200$ with $r = 1$ factor for $X_t$ and such that $y_t = 1 + F_t + \epsilon_t$, full details of which are available upon request.

around the score over the full sample as in Corradi and Swanson (2006):

$$\text{Rolling:} \qquad \widetilde{\beta}_t^* = \arg\min_{\beta} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( \left( y_{j+h}^* - \widetilde{\mathcal{F}}_j^{*\prime} \beta \right)^2 + 2\beta' \left( \frac{1}{T} \sum_{j'=1}^{T} \widetilde{F}_{j'}^{(T)} \widetilde{\epsilon}_{1,j'+h,t} \right) \right) \tag{24}$$

$$= \left( \frac{1}{R} \sum_{j=t-R+1}^{t-h} \widetilde{\mathcal{F}}_j^* \widetilde{\mathcal{F}}_j^{*\prime} \right)^{-1} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( \widetilde{\mathcal{F}}_j^* y_{j+h}^* - \left( \frac{1}{T} \sum_{j'=1}^{T} \widetilde{F}_{j'}^{(T)} \widetilde{\epsilon}_{1,j'+h,t} \right) \right)$$

$$\text{Recursive:} \qquad \widetilde{\beta}_t^* = \arg\min_{\beta} \frac{1}{t} \sum_{j=1}^{t-h} \left( \left( y_{j+h}^* - \widetilde{\mathcal{F}}_j^{*\prime} \beta \right)^2 + 2\beta' \left( \frac{1}{T} \sum_{j'=1}^{T} \widetilde{F}_{j'}^{(T)} \widetilde{\epsilon}_{1,j'+h,t} \right) \right) \tag{25}$$

$$= \left( \frac{1}{t} \sum_{j=1}^{t-h} \widetilde{\mathcal{F}}_j^* \widetilde{\mathcal{F}}_j^{*\prime} \right)^{-1} \frac{1}{t} \sum_{j=1}^{t-h} \left( \widetilde{\mathcal{F}}_j^* y_{j+h}^* - \left( \frac{1}{T} \sum_{j'=1}^{T} \widetilde{F}_{j'}^{(T)} \widetilde{\epsilon}_{1,j'+h,t} \right) \right)$$

and similarly for Model 2:

$$\text{Rolling:} \qquad \widetilde{\gamma}_t^* = \arg\min_{\gamma} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( \left( y_{j+h}^* - Z_j^{*\prime} \gamma \right)^2 + 2\gamma' \left( \frac{1}{T} \sum_{j'=1}^{T} Z_{j'} \widetilde{\epsilon}_{2,j'+h,t} \right) \right) \tag{26}$$

$$= \left( \frac{1}{R} \sum_{j=t-R+1}^{t-h} Z_j^* Z_j^{*\prime} \right)^{-1} \frac{1}{R} \sum_{j=t-R+1}^{t-h} \left( Z_j^* y_{j+h}^* - \left( \frac{1}{T} \sum_{j'=1}^{T} Z_{j'} \widetilde{\epsilon}_{2,j'+h,t} \right) \right)$$

$$\text{Recursive:} \qquad \widetilde{\gamma}_t^* = \arg\min_{\gamma} \frac{1}{t} \sum_{j=1}^{t-h} \left( \left( y_{j+h}^* - Z_j^{*\prime} \gamma \right)^2 + 2\gamma' \left( \frac{1}{T} \sum_{j'=1}^{T} Z_{j'} \widetilde{\epsilon}_{2,j'+h,t} \right) \right) \tag{27}$$

$$= \left( \frac{1}{t} \sum_{j=1}^{t-h} Z_j^* Z_j^{*\prime} \right)^{-1} \frac{1}{t} \sum_{j=1}^{t-h} \left( Z_j^* y_{j+h}^* - \left( \frac{1}{T} \sum_{j'=1}^{T} Z_{j'} \widetilde{\epsilon}_{2,j'+h,t} \right) \right)$$

for all $R \leq t \leq T$, where the recentering error terms as described in Equation (23) are $\widetilde{\epsilon}_{1,j'+h,t} = y_{j'+h} - \widetilde{F}_{j'}^{(T)\prime} \widetilde{\beta}_t$ and $\widetilde{\epsilon}_{2,j'+h,t} = y_{j'+h} - Z_{j'}' \widetilde{\gamma}_t$, which are the error terms of Models 1 and 2 evaluated at the respective rolling or recursive OLS estimators $\widetilde{\beta}_t$ and $\widetilde{\gamma}_t$ over the full sample $j' = 1, ..., T$. 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 $\widetilde{\beta}_t$ rather than $\widehat{\beta}_t$ is crucial for ensuring that the recentering term is well-behaved.

The bootstrap estimators $\widetilde{\beta}_t^*$ and $\widetilde{\gamma}_t^*$ are used in constructing the bootstrap forecast errors:

$$\widetilde{\epsilon}_{1,t+h}^* = y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime} \widetilde{\beta}_t^* \tag{28}$$

and:

$$\widetilde{\epsilon}_{2,t+h}^* = y_{t+h}^* - Z_t^{*\prime} \widetilde{\gamma}_t^* \tag{29}$$

Finally, the bootstrap counterpart of the test statistic $\widetilde{S}_P$ is then given by:

$$\widetilde{S}_P^* = \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \left( g\left( \widetilde{\epsilon}_{1,t+h}^* \right) - g\left( \widetilde{\epsilon}_{2,t+h}^* \right) \right) - \frac{1}{T} \sum_{j=1}^{T} \left( g\left( \widetilde{\epsilon}_{1,j+h,t} \right) - g\left( \widetilde{\epsilon}_{2,j+h,t} \right) \right) \right) \tag{30}$$

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 1-8, and assuming that $l, b \to \infty$ such that $l/T^{1/4} \to 0$. Then, under $H_0$:*

$$P\left( \omega : \sup_{s \in \mathbb{R}} \left| \Pr{}^*\left( \widetilde{S}_P^* \leq s \right) - \Pr\left( \widetilde{S}_P \leq s \right) \right| > \epsilon \right) \to 0$$

Using this result we can generate $B$ bootstrap replications of the test statistic $\widetilde{S}_P^*$ 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 $\widetilde{S}_P$ to these percentiles of the empirical distribution of $\widetilde{S}_P^*$.

# 5 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, under the rolling and recursive estimation schemes. The loss functions we consider for $g(.)$ are mean squared forecast error (MSFE) and mean absolute error (MAE):

$$MSFE\left(\epsilon_{t+h}\right) = \epsilon_{t+h}^2$$
$$MAE\left(\epsilon_{t+h}\right) = \left|\epsilon_{t+h}\right|$$

We consider model estimation using OLS, in which case parameter estimation error is negligible for the MSFE test, as shown by West (1996), so $\Omega = V_\epsilon$ in that case.[12] On the other hand, when $g(.)$ 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.

[12]Note that, as explained in Corradi and Swanson (2006), with negligible PEE we do not need to use the recentered estimator $\widehat{\beta}_t^*$ to construct the bootstrap forecast errors.

The MSFE and MAE tests are based on the forecasts generated from the two linear regressions:

$$y_{j+1} = \mu_1 + \beta \widetilde{F}_j^{(t)} + \epsilon_{1,j+1} \qquad\qquad y_{j+1} = \mu_2 + \gamma Z_j + \epsilon_{2,j+1} \tag{31}$$

Model 1 is the feasible factor-augmented model where $\widetilde{F}^{(t)}$ is the rolling adjusted PCA factor estimate described in Equation (18). 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):

$$\begin{aligned} X_{it} &= \lambda_i F_t + \sqrt{\theta} u_{it} \\ F_t &= \rho_F F_{t-1} + e_t \\ u_{it} &= \rho_u u_{i,t-1} + v_{it} \end{aligned}$$

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 $v_{it}$ are drawn independently from $N(0, 1-\rho_F^2)$ and $N(0, 1-\rho_u^2)$ 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 McCracken (2000):

$$\begin{aligned} y_{t+1} &= Z_t + (1+c)\, F_t + \epsilon_{t+1} \qquad (32) \\ Z_t &= \rho_Z Z_{t-1} + w_t \\ \epsilon_t &= \rho_\epsilon \epsilon_{t-1} + \eta_t \end{aligned}$$

In a similar way to above, $w_t$ and $\eta_t$ are drawn independently from $N(0, 1-\rho_Z^2)$ and $N(0, 1-\rho_\epsilon^2)$ 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 (32), so the null of equal MSFE and MAE are both satisfied. In this case, both Models 1 and 2 in Equation (31) 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. A comprehensive Monte Carlo treatment of predictive ability tests is given by Busetti and Marcucci (2013), but to our knowledge there are no Monte Carlo studies of the performance of these tests involving estimated factors.

The sample sizes we consider are $T = 240$ and $T = 480$ and we assess the performance based on different splits between $R$ and $P$ corresponding to values of $\pi = \{0.5, 1, 2\}$, recalling from Assumption 8b that $\pi = \lim_{T\to\infty}(P/R)$. We use a panel dimension of $N = 200$ as a medium between small- and large-scale factor model studies seen in the literature. In any case, we expect finite sample performance to vary most in the direction of $T$ and $\pi$, and not $N$. We select the persistence parameters $\rho_F$, $\rho_u$, $\rho_Z$ and $\rho_\epsilon$ 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 DMW test using the non-adjusted standard normal critical values by way of comparison as these are often reported in the empirical literature. The results indicate the the bootstrap test has good size properties for both the rolling and recursive estimation schemes 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, and to a lesser extent under MSFE loss. The results for the different block lengths are very similar, though the results for $l = 6$ are slightly better than those for $l = 3$. There are only very small differences in the performance between rolling and recursive estimation.

Since the bootstrap test with $l = 6$ is almost correctly sized, we only present the power results for this test in Table 2.[13] These results show that the bootstrap test has good power properties, with larger power for the larger values of $\pi$. The improvement of power with $\pi$ is expected given the Monte Carlo analysis of other studies such as Hansen and Timmermann (2015). 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.

# 6 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 series. 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:

$$\pi_{t+h}^{h} = \phi_1 + \beta_1 (L) F_t + \mu_1 (L) \pi_t + \epsilon_{t+h} \tag{33}$$

[13] Results for the power of the bootstrap test with $l = 3$ and the non size-adjusted power of the standard normal version of the test are omitted and available upon request.

**Table 1:** Empirical size for MSFE and MAE tests for equal out-of-sample predictive ability conducted at the nominal 10% level.

| | | Rolling | | | | | |
|---|---|---|---|---|---|---|---|
| | | $T = 240$ | | | $T = 480$ | | |
| | | $\pi = 0.5$ | $\pi = 1$ | $\pi = 2$ | $\pi = 0.5$ | $\pi = 1$ | $\pi = 2$ |
| | Standard Normal | 0.18 | 0.18 | 0.15 | 0.15 | 0.17 | 0.18 |
| *MSFE* | Bootstrap ($l = 3$) | 0.14 | 0.16 | 0.13 | 0.10 | 0.12 | 0.13 |
| | Bootstrap ($l = 6$) | 0.13 | 0.14 | 0.11 | 0.10 | 0.12 | 0.13 |
| | Standard Normal | 0.17 | 0.18 | 0.14 | 0.14 | 0.15 | 0.17 |
| *MAE* | Bootstrap ($l = 3$) | 0.14 | 0.15 | 0.10 | 0.11 | 0.11 | 0.13 |
| | Bootstrap ($l = 6$) | 0.11 | 0.14 | 0.09 | 0.08 | 0.12 | 0.11 |

| | | Recursive | | | | | |
|---|---|---|---|---|---|---|---|
| | | $T = 240$ | | | $T = 480$ | | |
| | | $\pi = 0.5$ | $\pi = 1$ | $\pi = 2$ | $\pi = 0.5$ | $\pi = 1$ | $\pi = 2$ |
| | Standard Normal | 0.17 | 0.17 | 0.15 | 0.15 | 0.17 | 0.18 |
| *MSFE* | Bootstrap ($l = 3$) | 0.14 | 0.15 | 0.13 | 0.10 | 0.12 | 0.14 |
| | Bootstrap ($l = 6$) | 0.12 | 0.13 | 0.10 | 0.10 | 0.13 | 0.13 |
| | Standard Normal | 0.17 | 0.19 | 0.14 | 0.14 | 0.15 | 0.17 |
| *MAE* | Bootstrap ($l = 3$) | 0.14 | 0.16 | 0.11 | 0.11 | 0.11 | 0.14 |
| | Bootstrap ($l = 6$) | 0.12 | 0.14 | 0.10 | 0.09 | 0.12 | 0.11 |

**Notes:** 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:

$$\pi_{t+h}^{h} = \phi_2 + \beta_2 (L) u_t + \mu_2 (L) \pi_t + \epsilon_{t+h} \tag{34}$$

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 $\pi_t^1 \equiv \pi_t = 1200 \ln (P_t/P_{t-1})$. 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)[14] 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

[14]We thank these authors for kindly providing us with their data.

**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.

| | Rolling | | | | | | |
|---|---|---|---|---|---|---|---|
| | | $T = 240$ | | | $T = 480$ | | |
| | $c$ | $\pi = 0.5$ | $\pi = 1$ | $\pi = 2$ | $\pi = 0.5$ | $\pi = 1$ | $\pi = 2$ |
| | 0.1 | 0.21 | 0.28 | 0.25 | 0.21 | 0.30 | 0.36 |
| | 0.2 | 0.35 | 0.41 | 0.41 | 0.43 | 0.56 | 0.63 |
| $MSFE$ | 0.3 | 0.46 | 0.58 | 0.61 | 0.64 | 0.77 | 0.86 |
| | 0.4 | 0.64 | 0.69 | 0.79 | 0.78 | 0.92 | 0.94 |
| | 0.5 | 0.75 | 0.85 | 0.88 | 0.90 | 0.97 | 0.99 |
| | 0.1 | 0.19 | 0.25 | 0.20 | 0.19 | 0.27 | 0.30 |
| | 0.2 | 0.31 | 0.39 | 0.35 | 0.44 | 0.48 | 0.58 |
| $MAE$ | 0.3 | 0.44 | 0.57 | 0.54 | 0.64 | 0.71 | 0.79 |
| | 0.4 | 0.58 | 0.65 | 0.73 | 0.77 | 0.88 | 0.92 |
| | 0.5 | 0.71 | 0.83 | 0.86 | 0.88 | 0.96 | 0.98 |

| | Recursive | | | | | | |
|---|---|---|---|---|---|---|---|
| | | $T = 240$ | | | $T = 480$ | | |
| | $c$ | $\pi = 0.5$ | $\pi = 1$ | $\pi = 2$ | $\pi = 0.5$ | $\pi = 1$ | $\pi = 2$ |
| | 0.1 | 0.21 | 0.27 | 0.26 | 0.22 | 0.32 | 0.36 |
| | 0.2 | 0.36 | 0.41 | 0.42 | 0.43 | 0.57 | 0.63 |
| $MSFE$ | 0.3 | 0.47 | 0.59 | 0.61 | 0.64 | 0.78 | 0.87 |
| | 0.4 | 0.64 | 0.67 | 0.80 | 0.77 | 0.92 | 0.95 |
| | 0.5 | 0.76 | 0.86 | 0.88 | 0.91 | 0.98 | 0.99 |
| | 0.1 | 0.20 | 0.26 | 0.24 | 0.21 | 0.28 | 0.32 |
| | 0.2 | 0.33 | 0.41 | 0.38 | 0.43 | 0.49 | 0.60 |
| $MAE$ | 0.3 | 0.46 | 0.57 | 0.58 | 0.64 | 0.73 | 0.83 |
| | 0.4 | 0.59 | 0.67 | 0.75 | 0.77 | 0.90 | 0.93 |
| | 0.5 | 0.71 | 0.84 | 0.88 | 0.88 | 0.97 | 0.98 |

**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.

identified as the "Great Moderation." Given that the present method is valid only with stable 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. We use rolling estimation, splitting the sample equally for estimation and prediction so that the rolling window length is $R = 150$ and $P \equiv T - R + 1 = 151$. 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 (34) and (33) 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.

| | | MAE Loss | | | MSFE Loss | | |
|---|---|---|---|---|---|---|---|
| | | $h=1$ | $h=3$ | $h=12$ | $h=1$ | $h=3$ | $h=12$ |
| | Relative Loss | 0.9982 | 0.9872 | 0.8562 | 1.0126 | 1.0104 | 0.7177 |
| | Statistic | -0.0546 | -0.2844 | -2.1842 | 1.9123 | 0.9983 | -8.0507 |
| $l=3$ | 5% | -0.3556 | -0.2401 | 0.1694 | -4.1110 | -6.3238 | -3.5696 |
| | 10% | -0.2123 | 0.0312 | 0.3232 | -3.1921 | -4.6923 | -2.6402 |
| | 50% | 0.2178 | 0.6339 | 0.7197 | -0.5480 | 0.0008 | 0.0999 |
| | $p$-value | 0.3860 | 0.0852 | 0.0000 | 0.3609 | 0.7569 | 0.0000 |
| $l=6$ | 5% | -0.4035 | -0.5694 | -0.2086 | -4.0514 | -6.3368 | -4.2747 |
| | 10% | -0.2460 | -0.0754 | 0.0879 | -3.0901 | -4.2660 | -3.0541 |
| | 50% | 0.2060 | 0.6368 | 0.6203 | -0.6212 | 0.3409 | 0.3047 |
| | $p$-value | 0.4411 | 0.1604 | 0.0000 | 0.2957 | 0.7870 | 0.0100 |
| $l=12$ | 5% | -0.4448 | -0.4820 | -0.5996 | -3.6582 | -4.8561 | -5.1587 |
| | 10% | -0.2808 | -0.1860 | -0.3670 | -3.0259 | -3.1668 | -3.6389 |
| | 50% | 0.1863 | 0.5801 | 0.6846 | -0.5504 | 0.6797 | 1.0798 |
| | $p$-value | 0.4010 | 0.1654 | 0.0000 | 0.1704 | 0.9023 | 0.0100 |
| | Standard Normal | 0.4314 | 0.3521 | 0.1143 | 0.7170 | 0.6112 | 0.1171 |

**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 $\widetilde{S}_P$ 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 a great deal 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 of 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.

# 7 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) and associated bootstrap procedures 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 contribution of this paper shows the properties of factors estimated by Principal Components under a rolling or recursive estimation scheme which is used to construct the forecast errors for the DMW test statistic. We provide convergence rates for the factor estimates which extends the existing results of Bai and Ng (2002) and others for the case of full-sample factor estimation, to the out-of-sample estimation context. We then show the conditions under which factor estimation error does not have any effect on the asymptotic distribution of the DMW test statistic.

The second main contribution of the paper is to provide a method to obtain bootstrap critical values for the DMW test with estimated factors. This is not a simple undertaking, as we show that the existing bootstrap method of Corradi and Swanson (2006) does not carry over to the case involving estimates factors, due to the "sign-changing" of the estimated factors and factor-augmented model parameters across different windows. 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 the block bootstrap method 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.

# 8 Appendix

The proofs of Theorem 1, Proposition 1 and Theorem 2 are provided here. Several additional technical Lemmas are also required but the detailed proofs are consigned to an Online Appendix. As mentioned in the text, these results are shown foe the case of rolling estimation. The proofs of the recursive case follow identical lines and are not included here. The full proofs for the recursive case are outlined in Gonçalves et al. (2015).

## 8.1 Proof of Theorem 1

The following technical Lemmas are required in proving Theorem 1:

**Lemma A:** *Under Normalization N1 and Assumptions 1-8:*
**(i)**

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

**(ii)**

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

**(iii)**

$$\sup_t \frac{1}{R} \sum_{j=t-R+1}^{t} \left( \widehat{F}_j^{(t)} - H_{NR}^{(t)} F_j \right) F_j' = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)$$

**(iv)**

$$\sup_t \left( \widehat{\Lambda}^{(t)} - \Lambda H_{NR}^{(t)-1} \right) = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{\sqrt{R}} \right\} \right)$$

**(v)**

$$\sup_t \frac{1}{R} \sum_{j=t-R+1}^{t} \left( \widehat{F}_j^{(t)} - H_{NR}^{(t)} F_j \right) \epsilon_{1.j+h} = O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)$$

**(vi)**

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

**(vii)**

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

*where $H_{NR}^{(t)}$ is described in Equation (10).*

The proofs of these Lemmas are provided in a separate Online Appendix. Lemma A extends existing results on full-sample factor estimation error in the standard Principal Components estimates $\widehat{F}^{(t)}$ 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 $\min\left\{\sqrt{R},N\right\}$. Finally, Lemma A(v) extends Bai and Ng (2006) Lemma A.1 (iv) to the case of rolling estimation. Lemmas A(vi) and (vii) extend the results of Bai and Ng (2006) Lemma A.1 to the rolling estimation case.

Taking the test statistic $\widehat{S}_P$ in Equation (9):

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

The part which is new to this paper is the first term in $g\left(\widehat{\epsilon}_{1,t+h}\right)$ which involves the adjusted rolling PCA estimates $\widehat{F}_j^{(t)}$ and corresponding regression estimates $\widehat{\beta}_t$ (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 $H_{NR}^{(t)}F_t$ and $H_{NR}^{(t)\prime-1}\beta$ yields:

$$\begin{aligned}\frac{1}{\sqrt{P}}\sum_{t=R}^{T}g\left(\widehat{\epsilon}_{1,t+h}\right)=&\frac{1}{\sqrt{P}}\sum_{t=R}^{T}g\left(\epsilon_{1,t+h}\right)+\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\nabla_F g\left(\epsilon_{1,t+h}\right)\left(\widehat{F}_t^{(t)}-H_{NR}^{(t)}F_t\right)\\&+\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\nabla_\beta g\left(\epsilon_{1,t+h}\right)\left(\widehat{\beta}_t-H_{NR}^{(t)\prime-1}\beta\right)+o_p\left(1\right)\end{aligned}$$

The second-order terms are $o_p\left(1\right)$ 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, Lemma A(vii) shows that:

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

therefore we can write:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T} g\left(\widehat{\epsilon}_{1,t+h}\right) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T} g\left(\epsilon_{1,t+h}\right) + \frac{1}{\sqrt{P}}\sum_{t=R}^{T} \nabla_{\beta} g\left(\epsilon_{1,t+h}\right)\left(\widehat{\beta}_t - H_{NR}^{(t)\prime-1}\beta\right) + o_p\left(1\right)$$

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

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

since it can be seen from Lemma A and Proposition 1a that $\frac{1}{R}\sum_{j=t-R+1}^{t-h}\widehat{F}_j^{(t)}\widehat{F}_j^{(t)\prime} = H^{\dagger}\Sigma_F H^{\dagger\prime} + o_p\left(1\right)$ uniformly in $t$, noting that the sign matrices $S^{(t)}$ cancel in each product. Therefore since stationarity and strong mixing of $\epsilon_{1,t+h}$ in Assumption 1 along with measurability of $g\left(.\right)$ by Assumption 6, $\nabla_{\beta} g\left(\epsilon_{1,t+h}\right)$ 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:

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\nabla_{\beta} g\left(\epsilon_{1,t+h}\right)\left(\widehat{\beta}_t - H_{NR}^{(t)\prime-1}\beta\right)\\
&= D_{\beta}\left(H^{\dagger}\Sigma_F H^{\dagger\prime}\right)^{-1} H^{\dagger}\frac{1}{\sqrt{P}R}\sum_{t=R}^{T}\sum_{j=t-R+1}^{t-h} F_j\epsilon_{1,j+h}\\
&+ D_{\beta}\left(H^{\dagger}\Sigma_F H^{\dagger\prime}\right)^{-1}\frac{1}{\sqrt{P}R}\sum_{t=R}^{T}\sum_{j=t-R+1}^{t-h}\left(\widehat{F}_j^{(t)} - H_{NR}^{(t)}F_j\right)\epsilon_{1,j+h} + o_p\left(1\right)
\end{aligned}$$

where $D_{\beta}$ is described in Theorem 1.[15] This gives a similar expression to that in West (1996) but for the factor estimation error term. For this term we use Lemma A(vi), which shows that:

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

Therefore:

$$\begin{aligned}
\frac{1}{\sqrt{P}}\sum_{t=R}^{T} g\left(\widehat{\epsilon}_{1,t+h}\right) &= \frac{1}{\sqrt{P}}\sum_{t=R}^{T} g\left(\epsilon_{1,t+h}\right)\\
&+ D_{\beta}\left(H^{\dagger}\Sigma_F H^{\dagger\prime}\right)^{-1} H^{\dagger}\frac{1}{\sqrt{P}R}\sum_{t=R}^{T}\sum_{j=t-R+1}^{t-h} F_j\epsilon_{1,j+h} + o_p\left(1\right)
\end{aligned}$$

[15] Note that, again, the sign matrix $S^{(R)}$ cancels out between $\nabla_{\beta} g\left(\epsilon_{1,t+h}\right)$ and $H_R^{\dagger}F_j$.

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

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R}^{T} g\left(\widehat{\epsilon}_{2,t+h}\right) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T} g\left(\epsilon_{2,t+h}\right) + \frac{1}{\sqrt{P}}\sum_{t=R}^{T} \nabla_{\gamma} g\left(\epsilon_{2,t+h}\right)\left(\widehat{\gamma}_t - \gamma\right) + o_p\left(1\right) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T} g\left(\epsilon_{2,t+h}\right) + D_{\gamma}\Sigma_Z^{-1}\frac{1}{\sqrt{P}R}\sum_{t=R}^{T}\sum_{j=t-R+1}^{t-h} Z_j\epsilon_{2,j+h} + o_p\left(1\right)
\end{aligned}$$

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

$$\begin{aligned}
\widehat{S}_P &= \frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(g\left(\widehat{\epsilon}_{1,t+h}\right) - g\left(\widehat{\epsilon}_{2,t+h}\right)\right) \qquad (35) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(g\left(\epsilon_{1,t+h}\right) - g\left(\epsilon_{2,t+h}\right)\right) + D_{\beta}\left(H^{\dagger}\Sigma_F H^{\dagger\prime}\right)^{-1} H^{\dagger}\frac{1}{\sqrt{P}R}\sum_{t=R}^{T}\sum_{j=t-R+1}^{t-h} F_j\epsilon_{1,j+h} \\
&- D_{\gamma}\Sigma_Z^{-1}\frac{1}{\sqrt{P}R}\sum_{t=R}^{T}\sum_{j=t-R+1}^{t-h} Z_j\epsilon_{2,j+h} + O_p\left(\max\left\{\frac{\sqrt{P}}{N},\frac{\sqrt{P}}{R}\right\}\right) + o_p\left(1\right)
\end{aligned}$$

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 $H_0 : \mathrm{E}\left(g\left(\epsilon_{1,t+h}\right) - g\left(\epsilon_{2,t+h}\right)\right) = 0$:

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

where:

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

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 $\pi < 1$, $\theta_1 = \left(\pi - \frac{\pi^2}{3}\right)$ and $\theta_2 = \frac{\pi}{2}$. For the recursive scheme, $\theta_1 = 2\left(1 - \frac{1}{\pi}\ln\left(1+\pi\right)\right)$ and $\theta_2 = 1 - \frac{1}{\pi}\ln\left(1+\pi\right)$.

The variance-covariance matrices are fully described in the statement of Theorem 1 in the text. This shows what was required.

## 8.2 Proof of Proposition 1a and 1b

**Proof of Proposition 1a:**

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

$$
\begin{aligned}
&\frac{1}{R}\sum_{j=t-R+1}^{t}\left\|\widehat{F}_j^{(t)}-H_t^{\dagger}F_j\right\|^2\\
&=\frac{1}{R}\sum_{j=t-R+1}^{t}\left\|\left(\widehat{F}_j^{(t)}-H_{NR}^{(t)}F_j\right)+\left(H_{NR}^{(t)}-H_t^{\dagger}\right)F_j\right\|^2\\
&\leq\frac{2}{R}\sum_{j=t-R+1}^{t}\left\|\left(\widehat{F}_j^{(t)}-H_{NR}^{(t)}F_j\right)\right\|^2+\frac{2}{R}\sum_{j=t-R+1}^{t}\left\|\left(H_{NR}^{(t)}-H_t^{\dagger}\right)F_j\right\|^2\\
&\leq\frac{2}{R}\sum_{j=t-R+1}^{t}\left\|\left(\widehat{F}_j^{(t)}-H_{NR}^{(t)}F_j\right)\right\|^2+\left\|H_{NR}^{(t)}-H_t^{\dagger}\right\|^2\frac{2}{R}\sum_{j=t-R+1}^{t}\left\|F_j\right\|^2 \qquad (36)
\end{aligned}
$$

where:

$$H_{NR}^{(t)}=\widehat{V}^{(t)-1}\frac{\widehat{F}^{(t)\prime}F^{(t)}}{R}\frac{\Lambda'\Lambda}{N}$$

is the rotation matrix of Equation (10) in the text, $H_t^{\dagger}=S^{(t)}H^{\dagger}$ and $H^{\dagger}=V^{-1}Q\Sigma_{\Lambda}$ and $S^{(t)}=\operatorname{diag}(\pm1,...,\pm1)$ is any sign matrix. Now, it follows from Equation (36) that:

$$
\begin{aligned}
\sup_t\frac{1}{R}\sum_{j=t-R+1}^{t}\left\|\widehat{F}_j^{(t)}-H_t^{\dagger}F_j\right\|^2\leq&\sup_t\frac{2}{R}\sum_{j=t-R+1}^{t}\left\|\left(\widehat{F}_j^{(t)}-H_{NR}^{(t)}F_j\right)\right\|^2\\
&+\sup_t\left\|H_{NR}^{(t)}-H_t^{\dagger}\right\|^2\sup_t\frac{2}{R}\sum_{j=t-R+1}^{t}\left\|F_j\right\|^2
\end{aligned}
$$

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

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

we can combine these two expressions to get:

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

Since $\sup_t\frac{1}{R}\sum_{j=t-R+1}^{t}\left\|F_j\right\|^2$ is $O_p(1)$ by Assumption 2 we are left with the part in $\left\|H_{NR}^{(t)}-H_t^{\dagger}\right\|^2$. 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 $\Lambda'\Lambda/N\xrightarrow{p}\Sigma_{\Lambda}$ by Assumption 2b. For the first part of the rotation matrix in $\widehat{V}^{(t)-1}$, we note that by using the rolling eigen-identity:

$$\frac{X^{(t)}X^{(t)\prime}}{RN}\widehat{F}^{(t)}=\widehat{F}^{(t)}\widehat{V}^{(t)}$$

we can get:

$$\widehat{V}^{(t)} = \frac{1}{R}\widehat{F}^{(t)\prime}\frac{X^{(t)}X^{(t)\prime}}{RN}\widehat{F}^{(t)}$$

using the fact that $\widehat{F}^{(t)\prime}\widehat{F}^{(t)}/R = I_r$ under Normalization N1. Now showing the limit of $\widehat{V}^{(t)}$ follows an identical proof to the full-sample estimation result in Stock and Watson (2002a) as we assume in Assumption 2 that $F^{(t)\prime}F^{(t)}/R \xrightarrow{p} \Sigma_F$ uniformly in $t$ and since the factors are normalized to have unit length in every rolling window. Therefore as in Stock and Watson (2002a) we have that:

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

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

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

Finally for part of the matrix in $\widehat{F}^{(t)\prime}F^{(t)}/R$, we note that this term is the analogue to the full sample version in Bai (2003) Proposition 1, and we have that:

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

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

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

uniformly in $t$, which in turn yields $\sup_t \left\|H_{NR}^{(t)} - H_t^\dagger\right\|^2 = O_p\left(\max\left\{\frac{1}{N}, \frac{1}{R}\right\}\right)$ since we define $H_t^\dagger = S^{(t)}H^\dagger$ with $H^\dagger = V^{-1}Q\Sigma_\Lambda$.

Combining all of these results we have:

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

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:

$$\sup_t \left\|\widehat{\beta}_t - H_t^{\dagger\prime-1}\beta\right\| = o_p(1)$$

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 $\widetilde{F}^{(t)}$ to the standard PCA estimates $\widehat{F}^{(t)}$ in order to apply the results from Lemma A and Proposition 1a. Recall from Equation (18) that the adjusted PCA estimates under Normalization N2 for rolling window $R \leq t \leq T$ are:

$$\widetilde{F}^{(t)} = \widehat{F}^{(t)} \widehat{\Lambda}_1^{(t)\prime} \left( \widehat{\Lambda}_1^{(R)\prime} \right)^{-1}$$

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

$$\widetilde{F}_j^{(t)} = \left( \widehat{\Lambda}_1^{(R)} \right)^{-1} \widehat{\Lambda}_1^{(t)} \widehat{F}_j^{(t)}$$

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:

$$\begin{aligned}
\widetilde{F}_j^{(t)} - H_{NR}^{(R)} F_j &= \left( \widehat{\Lambda}_1^{(R)} \right)^{-1} \widehat{\Lambda}_1^{(t)} \widehat{F}_j^{(t)} - H_{NR}^{(R)} F_j \\
&= \left( \widehat{\Lambda}_1^{(R)} \right)^{-1} \widehat{\Lambda}_1^{(t)} \left( \widehat{F}_j^{(t)} - H_{NR}^{(t)} F_j \right) \\
&+ \left[ \left( \widehat{\Lambda}_1^{(R)} \right)^{-1} \widehat{\Lambda}_1^{(t)} H_{NR}^{(t)} - H_{NR}^{(R)} \right] F_j
\end{aligned}$$

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

$$\begin{aligned}
\frac{1}{R} \sum_{j=t-R+1}^{t} \left( \widetilde{F}_j^{(t)} - H_{NR}^{(R)} F_j \right) &= \left[ \left( \widehat{\Lambda}_1^{(R)} \right)^{-1} \widehat{\Lambda}_1^{(t)} \right] \frac{1}{R} \sum_{j=t-R+1}^{t} \left( \widehat{F}_j^{(t)} - H_{NR}^{(t)} F_j \right) \\
&+ \left[ \left( \widehat{\Lambda}_1^{(R)} \right)^{-1} \widehat{\Lambda}_1^{(t)} H_{NR}^{(t)} - H_{NR}^{(R)} \right] \frac{1}{R} \sum_{j=t-R+1}^{t} F_j
\end{aligned} \tag{37}$$

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:

$$\widehat{\Lambda}_1^{(t)} = \Lambda_1 H_{NR}^{(t)-1} + O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{\sqrt{R}} \right\} \right)$$

This result therefore implies for the terms in square brackets on the RHS of Equation (37) we have

$$\begin{aligned}
\left( \widehat{\Lambda}_1^{(R)} \right)^{-1} \widehat{\Lambda}_1^{(t)} &= H_{NR}^{(R)} \Lambda_1^{-1} \Lambda_1 H_{NR}^{(t)-1} + O_p \left( \max \left\{ \frac{1}{N}, \frac{1}{\sqrt{R}} \right\} \right) \\
&= H_{NR}^{(R)} H_{NR}^{(t)-1} + o_p(1) \\
&= O_p(1)
\end{aligned} \tag{38}$$

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:

$$\begin{aligned}\left(\widehat{\Lambda}_1^{(R)}\right)^{-1}\widehat{\Lambda}_1^{(t)}H_{NR}^{(t)}-H_{NR}^{(R)}&=H_{NR}^{(R)}\Lambda_1^{-1}\Lambda_1 H_{NR}^{(t)-1}H_{NR}^{(t)}-H_{NR}^{(R)}+O_p\left(\max\left\{\frac{1}{N},\frac{1}{\sqrt{R}}\right\}\right)\\&=O_p\left(\max\left\{\frac{1}{N},\frac{1}{\sqrt{R}}\right\}\right)\end{aligned}\tag{39}$$

as the rotation matrix $H_{NR}^{(t)}$ cancels out with its own inverse, as does $\Lambda_1$, and we are left with $H_{NR}^{(R)}-H_{NR}^{(R)}$. Finally, from Equation (37) we have:

$$\begin{aligned}\sup_t\frac{1}{R}\sum_{j=t-R+1}^{t}\left\|\widetilde{F}_j^{(t)}-H_{NR}^{(R)}F_j\right\|^2&\le 2\left\|\sup_t\left[\left(\widehat{\Lambda}_1^{(R)}\right)^{-1}\widehat{\Lambda}_1^{(t)}\right]\right\|^2\sup_t\frac{1}{R}\sum_{j=t-R+1}^{t}\left\|\widehat{F}_j^{(t)}-H_{NR}^{(t)}F_j\right\|^2\\&\quad+2\left\|\sup_t\left[\left(\widehat{\Lambda}_1^{(R)}\right)^{-1}\widehat{\Lambda}_1^{(t)}H_{NR}^{(t)}-H_{NR}^{(R)}\right]\right\|^2\sup_t\frac{1}{R}\sum_{j=t-R+1}^{t}\|F_j\|^2\\&=O_p(1)\times O_p\left(\max\left\{\frac{1}{N},\frac{1}{R}\right\}\right)+O_p\left(\max\left\{\frac{1}{N^2},\frac{1}{R}\right\}\right)\times O_p(1)\\&=O_p\left(\max\left\{\frac{1}{N},\frac{1}{R}\right\}\right)\end{aligned}$$

by the results in Equation (38), Lemma A(i), Equation (39) and Assumption 2a. Finally, exactly as in the proof of Proposition 1a we can now replace the rotation matrix $H_{NR}^{(R)}$ with $H_R^{\dagger}$, where in this case we only have the rotation matrix from the first rolling window. It therefore follows that:

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

by the same logic as in Equation (36). 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 $\widetilde{F}_j^{(t)}$, and we have:

$$\sup_t\left\|\widetilde{\beta}_t-H_R^{\dagger\prime-1}\beta\right\|=o_p(1)$$

using the results in West (1996) Lemma A3.

## 8.3 Proof of Theorem 1'

In proving Theorem 1', the modification of Theorem 1 to the adjusted PCA estimates, we additionally require the following results which come as a corollary of Proposition 1b.

**Corollary A** *Under Normalization N2, and given Proposition 1b and Lemma A:*
**(i)**

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

**(ii)**

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

*where* $H_R^{\dagger}$ *is the rotation matrix described in Proposition 1b.*

The proof of Corollary A can be found in the separate Appendix. This shows the same result as in Lemma A(vi) and (vii) holds for the adjusted PCA estimates $\widetilde{F}_t^{(t)}$, which are consistent for the rotation matrix $H_R^{\dagger}$ as shown in Proposition 1b.

The proof of Theorem 1' follows immediately from the same steps used to prove Theorem 1, but using the results in Corollary A. Specifically, in the same way as Equation (35), we can write the test statistic $\widetilde{S}_P$ from Equation (22) as:

$$\begin{aligned}\widetilde{S}_P &= \frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(g\left(\widetilde{\epsilon}_{1,t+h}\right)-g\left(\widetilde{\epsilon}_{2,t+h}\right)\right) \qquad (40)\\ &= \frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(g\left(\epsilon_{1,t+h}\right)-g\left(\epsilon_{2,t+h}\right)\right)+\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\nabla_F g\left(\epsilon_{1,t+h}\right)\left(\widetilde{F}_t^{(t)}-H_R^{\dagger}F_t\right)\\ &+ D_\beta\left(H^{\dagger}\Sigma_F H^{\dagger\prime}\right)^{-1}H^{\dagger}\frac{1}{\sqrt{P}R}\sum_{t=R}^{T}\sum_{j=t-R+1}^{t-h}F_j\epsilon_{1,j+h}\\ &+ D_\beta\left(H^{\dagger}\Sigma_F H^{\dagger\prime}\right)^{-1}\frac{1}{\sqrt{P}R}\sum_{t=R}^{T}\sum_{j=t-R+1}^{t-h}\left(\widetilde{F}_j^{(t)}-H_R^{\dagger}F_j\right)\epsilon_{1,j+h}\\ &- D_\gamma\Sigma_Z^{-1}\frac{1}{\sqrt{P}R}\sum_{t=R}^{T}\sum_{j=t-R+1}^{t-h}Z_j\epsilon_{2,j+h}+o_p\left(1\right)\end{aligned}$$

Now the second and fourth terms of the last expression are $O_p\left(\max\left\{\frac{\sqrt{P}}{N},\frac{\sqrt{P}}{R}\right\}\right)$ by Corollary A, and the remaining terms are the same as in Equation (35), meaning that the same distribution obtains for $\widetilde{S}_P$ as for $\widehat{S}_P$, as required.

## 8.4 Proof of Theorem 2

Taking a Taylor series expansion of $\widetilde{S}_P^*$ around $\widetilde{\beta}_t$ and $\widetilde{\gamma}_t$ yields:

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

$$= \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \left( g\left( y_{t+h}^{*} - \widetilde{\mathcal{F}}_{t}^{*\prime} \widetilde{\beta}_{t} \right) - g\left( y_{t+h}^{*} - Z_{t}^{*\prime} \widetilde{\gamma}_{t} \right) \right) - \frac{1}{T} \sum_{j=1}^{T} \left( g\left( \widetilde{\epsilon}_{1,j+h,t} \right) - g\left( \widetilde{\epsilon}_{2,j+h,t} \right) \right) \right)$$

$$+ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\beta} g\left( y_{t+h}^{*} - \widetilde{\mathcal{F}}_{j}^{*\prime} \overline{\beta}_{t} \right) \left( \widetilde{\beta}_{t}^{*} - \widetilde{\beta}_{t} \right) + \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\gamma} g\left( y_{t+h}^{*} - Z_{t}^{*\prime} \overline{\gamma}_{t} \right) \left( \widetilde{\gamma}_{t}^{*} - \widetilde{\gamma}_{t} \right) \tag{41}$$

where $\overline{\beta}_t \in \left(\widetilde{\beta}_t^*, \widetilde{\beta}_t\right)$ and $\overline{\gamma}_t \in \left(\widetilde{\gamma}_t^*, \widetilde{\gamma}_t\right)$ and $\widetilde{\epsilon}_{1,j+h,t} = y_{j+h} - \widetilde{F}_j^{(T)\prime} \widetilde{\beta}_t$ and $\widetilde{\epsilon}_{2,j+h,t} = y_{j+h} - Z_j' \widetilde{\gamma}_t$ as described in Equation (23) in the text.

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

### 8.4.1 Proof of Bootstrap Mean

We begin by showing that the three terms on the RHS of Equation (41) have mean zero. For the first part of Equation (41), we need to show that:

$$\begin{aligned} & \mathrm{E}^* \left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g\left( y_{t+h}^{*} - \widetilde{\mathcal{F}}_{t}^{*\prime} \widetilde{\beta}_{t} \right) - g\left( y_{t+h}^{*} - Z_{t}^{*\prime} \widetilde{\gamma}_{t} \right) \right) \right] \\ &= \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \frac{1}{T} \sum_{j=1}^{T} \left( g\left( \widetilde{\epsilon}_{1,j+h,t} \right) - g\left( \widetilde{\epsilon}_{2,j+h,t} \right) \right) \right) + o_p(1) \end{aligned}$$

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

$$\begin{aligned} & \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( g\left( y_{t+h}^{*} - \widetilde{\mathcal{F}}_{t}^{*\prime} \widetilde{\beta}_{t} \right) - g\left( y_{t+h}^{*} - Z_{t}^{*\prime} \widetilde{\gamma}_{t} \right) \right) \\ & - \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \frac{1}{T} \sum_{j=1}^{T} \left( g\left( \widetilde{\epsilon}_{1,j+h,t} \right) - g\left( \widetilde{\epsilon}_{2,j+h,t} \right) \right) \right) = o_{p^*}(1) \end{aligned} \tag{42}$$

Consider the first part of this expectation:

$$\begin{aligned} & \mathrm{E}^* \left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} g\left( y_{t+h}^{*} - \widetilde{\mathcal{F}}_{t}^{*\prime} \widetilde{\beta}_{t} \right) \right] \\ &= \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \frac{1}{T-l+1} g\left( y_{1+h} - \widetilde{F}_{1}^{(T)\prime} \widetilde{\beta}_{t} \right) + ... + \frac{1}{T-l+1} g\left( y_{T+h} - \widetilde{F}_{T}^{(T)\prime} \widetilde{\beta}_{t} \right) \right) + O_p\left( \frac{l}{T} \right) \\ &= \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left( \frac{1}{T} \sum_{j=1}^{T} g\left( y_{j+h} - \widetilde{F}_{j}^{(T)\prime} \widetilde{\beta}_{t} \right) \right) + o_p(1) \end{aligned}$$

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

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

And recalling the definitions of $\widetilde{\epsilon}_{1,j+h,t}$ and $\widetilde{\epsilon}_{2,j+h,t}$ we have:

$$\begin{aligned}\mathrm{E}^*&\left[\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\left(g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}\widetilde{\beta}_t\right) - g\left(y_{t+h}^* - Z_t^{*\prime}\widetilde{\gamma}_t\right)\right)\right]\right. \\ &\quad -\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\frac{1}{T}\sum_{j=1}^{T}\left(g\left(\widetilde{\epsilon}_{1,j+h,t}\right) - g\left(\widetilde{\epsilon}_{2,j+h,t}\right)\right)\right) = o_p(1)\end{aligned}$$

which shows (42) as required. For the expectation of the second part of Equation (41) we need to show that:

$$\mathrm{E}^*\left[\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\nabla_\beta g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}\overline{\beta}_t\right)\left(\widetilde{\beta}_t^* - \widetilde{\beta}_t\right)\right] = o_p(1)$$

so that by the bootstrap LLN:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\nabla_\beta g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}\overline{\beta}_t\right)\left(\widetilde{\beta}_t^* - \widetilde{\beta}_t\right) = o_{p^*}(1) \tag{43}$$

Consider the bootstrap estimator from the recentered OLS objective function described in (24).

$$\begin{aligned}\widetilde{\beta}_t^* &= \arg\min_\beta \frac{1}{R}\sum_{j=t-R+1}^{t-h}\left(\left(y_{j+h}^* - \widetilde{\mathcal{F}}_j^{*\prime}\beta\right)^2 + 2\beta^\prime\left(\frac{1}{T}\sum_{j^\prime=1}^{T}\widetilde{F}_{j^\prime}^{(T)}\widetilde{\epsilon}_{1,j^\prime+h,t}\right)\right) \\ &= \left(\frac{1}{R}\sum_{j=t-R+1}^{t-h}\widetilde{\mathcal{F}}_j^*\widetilde{\mathcal{F}}_j^{*\prime}\right)^{-1}\frac{1}{R}\sum_{j=t-R+1}^{t-h}\left(\widetilde{\mathcal{F}}_j^* y_{j+h}^* - \left(\frac{1}{T}\sum_{j^\prime=1}^{T}\widetilde{F}_{j^\prime}^{(T)}\widetilde{\epsilon}_{1,j^\prime+h,t}\right)\right)\end{aligned}$$

Now for all $R \le t \le T$, defining the following error term $\widehat{\epsilon}_{1,j+h,t}^* = y_{j+h}^* - \widetilde{\mathcal{F}}_j^{*\prime}\widetilde{\beta}_t$ for $j = t-R+1,...,t-h$ we substitute $y_{j+h}^*$ into the above expression to get:

$$\widetilde{\beta}_t^* - \widetilde{\beta}_t = \left(\frac{1}{R}\sum_{j=t-R+1}^{t-h}\widetilde{\mathcal{F}}_j^*\widetilde{\mathcal{F}}_j^{*\prime}\right)^{-1}\frac{1}{R}\sum_{j=t-R+1}^{t-h}\left(\widetilde{\mathcal{F}}_j^*\widehat{\epsilon}_{1,j+h,t}^* - \left(\frac{1}{T}\sum_{j^\prime=1}^{T}\widetilde{F}_{j^\prime}^{(T)}\widetilde{\epsilon}_{1,j^\prime+h,t}\right)\right)$$

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 $\widetilde{\epsilon}_{1,j+h,t} = y_{j+h} - \widetilde{F}_j^{(T)\prime}\widetilde{\beta}_t$ , we first show that

$$\mathrm{E}^*\left[\frac{1}{R}\sum_{j=t-R+1}^{t-h}\widetilde{\mathcal{F}}_j^*\widehat{\epsilon}_{1,j+h,t}^*\right] = \frac{1}{T}\sum_{j=1}^{T}\widetilde{F}_j^{(T)}\widetilde{\epsilon}_{1,j+h,t} + o_p(1) \tag{44}$$

since:

$$
\begin{aligned}
&\mathrm{E}^*\left[\frac{1}{R}\sum_{j=t-R+1}^{t-h}\widetilde{\mathcal{F}}_j^*\widehat{\epsilon}_{1,j+h,t}^*\right] \\
&=\mathrm{E}^*\left[\frac{1}{R}\sum_{j=t-R+1}^{t-h}\widetilde{\mathcal{F}}_j^*\left(y_{j+h}^*-\widetilde{\mathcal{F}}_j^{*\prime}\widetilde{\beta}_t\right)\right] \\
&=\left(\frac{1}{T-l+1}\widetilde{F}_1^{(T)}\left(y_{1+h}-\widetilde{F}_1^{(T)\prime}\widetilde{\beta}_t\right)+\ldots+\frac{1}{T-l+1}\widetilde{F}_T^{(T)}\left(y_{T+h}-\widetilde{F}_T^{(T)\prime}\widetilde{\beta}_t\right)\right)+O_p\left(\frac{l}{T}\right) \\
&=\frac{1}{T}\sum_{j=1}^{T}\widetilde{F}_j^{(T)}\widetilde{\epsilon}_{1,j+h,t}+o_p(1)
\end{aligned}
$$

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

$$
\begin{aligned}
\mathrm{E}^*\left[\frac{1}{R}\sum_{j=t-R+1}^{t-h}\widetilde{\mathcal{F}}_j^*\widetilde{\mathcal{F}}_j^{*\prime}\right] &=\frac{1}{T}\sum_{j=1}^{T}\widetilde{F}_j^{(T)}\widetilde{F}_j^{(T)\prime}+o_p(1) \\
&=H^{\dagger}\Sigma_F H^{\dagger\prime}+o_p(1)
\end{aligned}
$$

uniformly over $t$, which follows from the recursive analogue to Proposition 1b as $\widetilde{F}_j^{(T)}$ is consistent for $H_R^{\dagger}F_j$ and because the sign matrix $S^{(R)}$ cancels in the product. Therefore it follows that:

$$
\begin{aligned}
&\mathrm{E}^*\left[\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\nabla_{\beta}g\left(y_{t+h}^*-\widetilde{\mathcal{F}}_t^{*\prime}\overline{\beta}_t^*\right)\left(\widetilde{\beta}_t^*-\widetilde{\beta}_t\right)\right] \\
&=D_{\beta}\left(H^{\dagger}\Sigma_F H^{\dagger\prime}\right)^{-1}\mathrm{E}^*\left[\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\frac{1}{R}\sum_{j=t-R+1}^{t-h}\left(\widetilde{\mathcal{F}}_j^*\widehat{\epsilon}_{1,j+h,t}^*-\frac{1}{T}\sum_{j'=1}^{T}\widetilde{F}_{j'}^{(T)}\widetilde{\epsilon}_{1,j'+h,t}\right)\right]+o_p(1) \qquad (45) \\
&=o_p(1)
\end{aligned}
$$

because, as in Corradi and Swanson (2006), it follows from (44) that the bias term on the RHS of (45) 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 (43) as required.

Finally, for the expectation of the last part of Equation (41) we need to show that:

$$
\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\nabla_{\gamma}g\left(y_{t+h}^*-Z_t^{*\prime}\overline{\gamma}_t\right)\left(\widetilde{\gamma}_t^*-\widetilde{\gamma}_t\right)=o_{p^*}(1) \qquad (46)
$$

Since we assume for simplicity that $Z_t$ does not contain 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 (42), (43) and (46) shows that the statistic $\widetilde{S}_P^*$ has mean zero, as required.

### 8.4.2 Proof of Bootstrap Variance

As Equation (41) 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:

$$\overline{g}\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}\widetilde{\beta}_t\right) = g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}\widetilde{\beta}_t\right) - \frac{1}{T}\sum_{j=1}^{T} g\left(y_{j+h} - \widetilde{F}_j^{(T)\prime}\widetilde{\beta}_t\right)$$

$$\nabla_F \overline{g}\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}\widetilde{\beta}_t\right) = \nabla_F g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}\widetilde{\beta}_t\right) - \frac{1}{T}\sum_{j=1}^{T} \nabla_F g\left(y_{j+h} - \widetilde{F}_j^{(T)\prime}\widetilde{\beta}_t\right)$$

For the first term of Equation (41), we must show that:

$$\mathrm{Var}^*\left[\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}\widetilde{\beta}_t\right) - g\left(y_{t+h}^* - Z_t^{*\prime}\widetilde{\gamma}_t\right)\right)\right] = V_\epsilon + o_p(1) \tag{47}$$

We start by taking a Taylor series expansion around the true $\gamma$ and true (rotated) $H_R^{\dagger\prime-1}\beta$:

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}\widetilde{\beta}_t\right) - g\left(y_{t+h}^* - Z_t^{*\prime}\widetilde{\gamma}_t\right)\right) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}H_R^{\dagger\prime-1}\beta\right) - g\left(y_{t+h}^* - Z_t^{*\prime}\gamma\right)\right) \\
&+ \frac{1}{\sqrt{P}}\sum_{t=R}^{T}\nabla_\beta g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}\overline{\beta}_t\right)\left(\widetilde{\beta}_t - H_R^{\dagger\prime-1}\beta\right) + \frac{1}{\sqrt{P}}\sum_{t=R}^{T}\nabla_\gamma g\left(y_{t+h}^* - Z_t^{*\prime}\overline{\gamma}_t\right)\left(\widetilde{\gamma}_t - \gamma\right) \qquad (48) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}H_R^{\dagger\prime-1}\beta\right) - g\left(y_{t+h}^* - Z_t^{*\prime}\gamma\right)\right) + o_{p^*}(1)
\end{aligned}$$

since the last two terms on the RHS of (48) 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 $R = b_1 l$

$$\mathrm{Var}^*\left[\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(g\left(y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime}H_R^{\dagger\prime-1}\beta\right) - g\left(y_{t+h}^* - Z_t^{*\prime}\gamma\right)\right)\right]$$

$$
\begin{aligned}
&= \mathrm{Var}^* \left[ \frac{1}{\sqrt{P}} \sum_{j=b_1+1}^{b} \sum_{i=1}^{l} \left( g\left( y_{I_j+i+h} - \widetilde{F}_{I_j+i}^{(T)\prime} H_R^{\dagger\prime -1} \beta \right) - g\left( y_{I_j+i+h} - Z_{I_j+i}^{\prime} \gamma \right) \right) \right] \\
&= \mathrm{E}^* \left[ \frac{1}{l} \sum_{i=1}^{l} \sum_{i'=1}^{l} \left( g\left( y_{I_j+i+h} - \widetilde{F}_{I_j+i}^{(T)\prime} H_R^{\dagger\prime -1} \beta \right) - g\left( y_{I_j+i+h} - Z_{I_j+i}^{\prime} \gamma \right) \right) \right. \\
&\left. \times \left( g\left( y_{I_j+i'+h} - \widetilde{F}_{I_j+i'}^{(T)\prime} H_R^{\dagger\prime -1} \beta \right) - g\left( y_{I_j+i'+h} - Z_{I_j+i'}^{\prime} \gamma \right) \right) \right] \\
&= \frac{1}{T} \sum_{t=l}^{T-l} \sum_{j=-l}^{l} \left( \left( \overline{g}\left( y_{t+h} - \widetilde{F}_t^{(T)\prime} H_R^{\dagger\prime -1} \beta \right) - \overline{g}\left( y_{t+h} - Z_t^{\prime} \gamma \right) \right) \right. \qquad (49) \\
&\left. \times \left( \overline{g}\left( y_{t+h+j} - \widetilde{F}_{t+j}^{(T)\prime} H_R^{\dagger\prime -1} \beta \right) - \overline{g}\left( y_{t+h+j} - Z_{t+j}^{\prime} \gamma \right) \right) \right) + O_p\left( \frac{l^2}{T} \right)
\end{aligned}
$$

Now, unlike in Corradi and Swanson (2006) we also need to relate estimated factors back to true (rotated) factors. Therefore we can use the recursive analogue to Proposition 1b, which follows from the recursive version of Lemma A (i), shown explicitly by Gonçalves et al. (2015), which states that:

$$
\frac{1}{T} \sum_{t=1}^{T} \left\| \widetilde{F}_t^{(T)} - H_R^{\dagger} F_t \right\|^2 = O_p\left( \max\left\{ \frac{1}{N}, \frac{1}{R} \right\} \right)
$$

Therefore we expand $\overline{g}\left( y_{t+h} - \widetilde{F}_t^{(T)\prime} H_R^{\dagger\prime -1} \beta \right)$ around $H_R^{\dagger} F_t$ for all $t$, which gives:

$$
\overline{g}\left( y_{t+h} - \widetilde{F}_t^{(T)\prime} H_R^{\dagger\prime -1} \beta \right) = \overline{g}\left( y_{t+h} - F_t^{\prime} \beta \right) + \nabla_F \overline{g}\left( y_{t+h} - \overline{F}_t^{\prime} H_R^{\dagger\prime -1} \beta \right) \left( \widetilde{F}_t^{(T)} - H_R^{\dagger} F_t \right) \qquad (50)
$$

for some $\overline{F}_t \in \left( \widetilde{F}_t^{(T)}, H_R^{\dagger} F_t \right)$. It is crucial here that $\widetilde{F}_t^{(T)}$ has the same rotation as $\widetilde{\beta}_t$ as in Condition 3, otherwise the last line would not hold as the first term on the RHS may be $\overline{g}\left( y_{t+h} + F_t^{\prime} \beta \right)$ instead of $\overline{g}\left( y_{t+h} - F_t^{\prime} \beta \right)$. Therefore the variance expression in Equation (49) can be written as the variance of the true forecast error loss differential $\overline{g}\left( \epsilon_{1,t+h} \right) - \overline{g}\left( \epsilon_{2,t+h} \right)$ plus cross-products and the square of the factor estimation error term:

$$
\begin{aligned}
&= \frac{1}{T} \sum_{t=l}^{T-l} \sum_{j=-l}^{l} \left( \left( \overline{g}\left( \epsilon_{1,t+h} \right) - \overline{g}\left( \epsilon_{2,t+h} \right) + \nabla_F \overline{g}\left( y_{t+h} - \overline{F}_t^{\prime} H_R^{\dagger\prime -1} \beta \right) \left( \widetilde{F}_t^{(T)} - H_R^{\dagger} F_t \right) \right) \right. \\
&\left. \times \left( \overline{g}\left( \epsilon_{1,t+h+j} \right) - \overline{g}\left( \epsilon_{2,t+h+j} \right) + \nabla_F \overline{g}\left( y_{t+h+j} - \overline{F}_{t+j}^{\prime} H_R^{\dagger\prime -1} \beta \right) \left( \widetilde{F}_{t+j}^{(T)} - H_R^{\dagger} F_{t+j} \right) \right) \right) + O_p\left( \frac{l^2}{T} \right) \\
&= \frac{1}{T} \sum_{t=l}^{T-l} \sum_{j=-l}^{l} \left[ \left( \overline{g}\left( \epsilon_{1,t+h} \right) - \overline{g}\left( \epsilon_{2,t+h} \right) \right) \left( \overline{g}\left( \epsilon_{1,t+h+j} \right) - \overline{g}\left( \epsilon_{2,t+h+j} \right) \right) \right] + o_p(1)
\end{aligned}
$$

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 1b. For example take the first factor estimation error term:

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=l}^{T-l}\sum_{j=-l}^{l}\left[\left(\overline{g}\left(\epsilon_{1,t+h}\right)-\overline{g}\left(\epsilon_{2,t+h}\right)\right)\times\nabla_F\overline{g}\left(y_{t+h+j}-\overline{F}'_{t+j}H_R^{\dagger\prime-1}\beta\right)\left(\widetilde{F}_{t+j}^{(T)}-H_R^{\dagger}F_{t+j}\right)\right]\\
&\leq(2l+1)\sup_{-l\leq j\leq l}\left|\frac{1}{T}\sum_{t=l}^{T-l}\left(\overline{g}\left(\epsilon_{1,t+h}\right)-\overline{g}\left(\epsilon_{2,t+h}\right)\right)\times\nabla_F\overline{g}\left(y_{t+h+j}-\overline{F}'_{t+j}H_R^{\dagger\prime-1}\beta\right)\left(\widetilde{F}_{t+j}^{(T)}-H_R^{\dagger}F_{t+j}\right)\right|\\
&=(2l+1)\times O_p(1)\,O_p\left(\max\left\{\frac{1}{\sqrt{N}},\frac{1}{\sqrt{R}}\right\}\right)
\end{aligned}
$$

Since for any $j$ we have:

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=l}^{T-l}\left(\overline{g}\left(\epsilon_{1,t+h}\right)-\overline{g}\left(\epsilon_{2,t+h}\right)\right)\times\nabla_F\overline{g}\left(y_{t+h+j}-\overline{F}'_{t+j}H_R^{\dagger\prime-1}\beta\right)\left(\overline{F}_{t+j}^{(R)}-H_R^{\dagger}F_{t+j}\right)\\
&\leq\left(\frac{1}{T}\sum_{t=l}^{T-l}\left\|\left(\overline{g}\left(\epsilon_{1,t+h}\right)-\overline{g}\left(\epsilon_{2,t+h}\right)\right)\nabla_F\overline{g}\left(y_{t+h+j}-\overline{F}'_{t+j}H_R^{\dagger\prime-1}\beta\right)\right\|^2\right)^{1/2}\\
&\times\left(\frac{1}{T}\sum_{t=l}^{T-l}\left\|\widetilde{F}_{t+j}^{(T)}-H_R^{\dagger}F_{t+j}\right\|^2\right)^{1/2}
\end{aligned}
$$

And since $\frac{1}{T}\sum_{t=l}^{T-l}\left\|\widetilde{F}_{t+j}^{(T)}-H_R^{\dagger}F_{t+j}\right\|^2=O_p\left(\max\left\{\frac{1}{N},\frac{1}{R}\right\}\right)$, 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:

$$
\begin{aligned}
&\mathrm{Var}^*\left[\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(g\left(y_{t+h}^*-\widetilde{\mathcal{F}}_t^{*\prime}\widetilde{\beta}_t\right)-g\left(y_{t+h}^*-Z_t^{*\prime}\widetilde{\gamma}_t\right)\right)\right]\\
&=\frac{1}{T}\sum_{t=l}^{T-l}\sum_{j=-l}^{l}\left[\left(\overline{g}\left(\epsilon_{1,t+h}\right)-\overline{g}\left(\epsilon_{2,t+h}\right)\right)\left(\overline{g}\left(\epsilon_{1,t+h+j}\right)-\overline{g}\left(\epsilon_{2,t+h+j}\right)\right)\right]\\
&+O_p(l)\,O_p\left(\max\left\{\frac{1}{\sqrt{N}},\frac{1}{\sqrt{R}}\right\}\right)+o_p(1)
\end{aligned}
$$

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 that $l/\sqrt{R}\to 0$. Finally defining the population autocovariance $\gamma_j^{\epsilon}=E\left[\left(\overline{g}\left(\epsilon_{1,t+h}\right)-\overline{g}\left(\epsilon_{2,t+h}\right)\right)\left(\overline{g}\left(\epsilon_{1,t+h+j}\right)-\overline{g}\left(\epsilon_{2,t+h+j}\right)\right)\right]$ it follows from West (1996) that the last term equals:

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

$$= \sum_{j=-l}^{l} \gamma_j^{\epsilon} + o_p(1)$$
$$\xrightarrow{p} V_{\epsilon}$$

which shows (47) 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:

$$\mathrm{Var}^* \left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\beta} g\left( y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime} \overline{\beta}_t \right) \left( \widetilde{\beta}_t^* - \widetilde{\beta}_t \right) \right] = \left( 1 - \frac{1}{3\pi} \right) D_{\beta} V_F D_{\beta}' + o_p(1) \tag{51}$$

$$\mathrm{Var}^* \left[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\gamma} g\left( y_{t+h}^* - Z_t^{*\prime} \overline{\gamma}_t \right) \left( \widetilde{\gamma}_t^* - \widetilde{\gamma}_t \right) \right] = \left( 1 - \frac{1}{3\pi} \right) D_{\gamma} V_Z D_{\gamma}' + o_p(1) \tag{52}$$

$$\begin{aligned} &\mathrm{Cov}^* \Bigg[ \left( g\left( y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime} \widetilde{\beta}_t \right) - g\left( y_{t+h}^* - Z_t^{*\prime} \widetilde{\gamma}_t \right) \right) - \frac{1}{T} \sum_{j=1}^{T} \left( g\left( \widetilde{\epsilon}_{1,j+h,t} \right) - g\left( \widetilde{\epsilon}_{2,j+h,t} \right) \right) \\ &, \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\beta} g\left( y_{t+h}^* - \widetilde{\mathcal{F}}_j^{*\prime} \overline{\beta}_t \right) \left( \widetilde{\beta}_t^* - \widetilde{\beta}_t \right) \Bigg] = 2 \left( 1 - \frac{1}{2\pi} \right) D_{\beta} C_{\epsilon,F} + o_p(1) \end{aligned} \tag{53}$$

$$\begin{aligned} &\mathrm{Cov}^* \Bigg[ \left( g\left( y_{t+h}^* - \widetilde{\mathcal{F}}_t^{*\prime} \widetilde{\beta}_t \right) - g\left( y_{t+h}^* - Z_t^{*\prime} \widetilde{\gamma}_t \right) \right) - \frac{1}{T} \sum_{j=1}^{T} \left( g\left( \widetilde{\epsilon}_{1,j+h,t} \right) - g\left( \widetilde{\epsilon}_{2,j+h,t} \right) \right) \\ &, \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\gamma} g\left( y_{t+h}^* - Z_t^{*\prime} \overline{\gamma}_t \right) \left( \widetilde{\gamma}_t^* - \widetilde{\gamma}_t \right) \Bigg] = 2 \left( 1 - \frac{1}{2\pi} \right) D_{\gamma} C_{\epsilon,Z} + o_p(1) \end{aligned} \tag{54}$$

$$\begin{aligned} &\mathrm{Cov}^* \Bigg[ \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\beta} g\left( y_{t+h}^* - \widetilde{\mathcal{F}}_j^{*\prime} \overline{\beta}_t \right) \left( \widetilde{\beta}_t^* - \widetilde{\beta}_t \right) \\ &, \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \nabla_{\gamma} g\left( y_{t+h}^* - Z_t^{*\prime} \overline{\gamma}_t \right) \left( \widetilde{\gamma}_t^* - \widetilde{\gamma}_t \right) \Bigg] = 2 \left( 1 - \frac{1}{3\pi} \right) D_{\beta} C_{F,Z} D_{\gamma}' + o_p(1) \end{aligned} \tag{55}$$

Therefore combining the results in Equations (47) and (51)-(55) it follows that

$$\mathrm{Var}^* \left[ \widetilde{S}_P^* \right] = \Omega + o_p(1)$$

as required. This completes the proof of Theorem 2.

# References


Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

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

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*(4), 1133–1150.

Bai, J. and S. Ng (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics 3*(2), 89–168.

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of Econometrics 176*(1), 18–29.

Banbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-Casting and the Real-Time Data Flow. In G. Elliott and A. Timmermann (Eds.), *Hanbook of Economic Forecasting: Volume 2A*, pp. 195–236. Amsterdam: North-Holland.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *Quarterly Journal of Economics 120*(1), 387–422.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169–194.

Breitung, J. and S. Eickmeier (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics 163*(1), 71–84.

Busetti, F. and J. Marcucci (2013). Comparing forecast accuracy: a monte carlo investigation. *International Journal of Forecasting 29*(1), 13–27.

Castle, J. L., M. P. Clements, and D. F. Hendry (2013). Forecasting by factors, by variables, by both or neither? *Journal of Econometrics 177*(2), 305–319.

Cheng, X. and B. E. Hansen (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics 186*(2), 280–293.

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

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

Corradi, V. and N. R. Swanson (2006). Predictive density and conditional confidence interval accuracy tests. *Journal of Econometrics 135*(1-2), 187–228.

Corradi, V. and N. R. Swanson (2014). Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics 182*(1), 100–118.

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

Eickmeier, S. and C. Ziegler (2008). How Successful are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach. *Journal of Forecasting 27*(3), 237–265.

Giacomini, R., D. N. Politis, and H. White (2013). A warp-speed method for conducti monte carlo experiments involving bootstrap estimators. *Econometric theory 29*(3), 567–589.

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

Gonçalves, S., M. W. McCracken, and B. Perron (2015). Tests of Equal Accuracy for Nested Models with Estimated Factors. *Federal Reserve Bank of St. Louis Working Paper 2015-025A*.

Gonçalves, S. and B. Perron (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics 182*(1), 156–173.

Grover, S. and M. W. McCracken (2014). Factor-based prediction of industry-wide bank stress. *Federal Reserve Bank of St.Louis Review 96*(2), 173.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Hansen, P. R. and A. Timmermann (2015). Equivalence Between Out-of-Sample Forecast Comparisons and Wald Statistics. *Econometrica 83*(6), 2485.

Kim, H. H. and N. R. Swanson (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics 178*(1), 352–367.

Ludvigson, S. C. and S. Ng (2007). The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics 83*(1), 171–222.

McCracken, M. W. (2000). Robust out-of-sample inference. *Journal of Econometrics 99*(2), 195–223.

Stock, J. and M. Watson (2009). Forecasting in dynamic factor models subject to structural instablility. In J. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: a Festschrift in Honour of David F. Hendry*. Oxford: Oxford University Press.

Stock, J. and M. Watson (2011). Dynamic factor models. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford handbook of economic forecasting*, pp. 35–60. New York: Oxford University Press.

Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics 44*(2), 293–335.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*(2), 147–162.

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