Factor Augmented Forecasting Subject to Structural Breaks in the Factor Structure*

Xu Han[1] and Ze-Yu Zhong[2]

[1] *City University of Hong Kong*

[2] *Monash University*

November 2024

Abstract

This paper investigates the impact of structural breaks in the factor structure on factor-augmented forecasting. We decompose the break in the factor loading matrix into rotational and shift components. To effectively utilise the pre-break data and maintain robustness against shift breaks, we propose a novel factor estimator that minimises the L2 distance between pre- and post-break loading matrices through the rotation of factor estimates. We call this estimator the "rotated factors" and analyse its asymptotic properties, along with two competing factor estimators, in the presence of different types of breaks. To leverage the respective advantages of each factor estimator in an automatic data driven way, we introduce a method that averages over sets of factor estimates using a leave-h-out cross-validation criterion. Simulations demonstrate that combining different factor estimates through the proposed cross-validation averaging approach leads to improved forecasting performance compared to existing methods. Furthermore, we evaluate the effectiveness of our methods in an empirical application with US macroeconomic data and emphasise the importance of incorporating structural breaks into factor-augmented forecasting models.

*We thank the participants at the 2024 IAAE Xiamen, 2024 IAAE Thessaloniki, and 2024 International Symposium on Forecasting for their comments. We acknowledge that this research was supported in part by the Monash eResearch Centre and eSolutions Research Support Services through the use of the MonARCH HPC Cluster.

1 Introduction

Factor-augmented regressions, pioneered by Stock and Watson (2002a, 2012) have emerged as the prevailing benchmark for macroeconomic forecasting. These models leverage unobserved factors that summarise information from a large set of predictors, resulting in significant empirical success in forecasting. However, because the existing literature on factor-augmented forecasting generally assumes structural stability, the presence of structural breaks in macroeconomic data poses a significant challenge. These breaks in macroeconomic data can introduce disruptions in the factor structure of dynamic factor models, thereby undermining the reliability and predictive power of the estimated factors.

In forecasting models that rely solely on observed predictors, addressing changes in regression coefficients is usually sufficient (Pesaran et al., 2006, 2013). However, in factor-augmented forecasting, equations are affected by structural breaks in both the regression coefficients and the factor estimators. Previous research has investigated the impacts of small and large breaks in the factor loading matrix on the factor estimators. When the break size is small, the full-sample Principal Components (PC) estimator remains robust; thus the break can be ignored during the estimation process, and the estimated factors remain consistent up to a rotational basis, (Stock and Watson, 2002a; Bates et al., 2013). Conversely, large breaks can increase the dimension of the factor space, leading to breaks in both the factor moments and the coefficients in the forecasting equation (e.g. Han and Inoue, 2015; Duan et al., 2022). Indeed, large breaks can contaminate the factor space, resulting in the PC estimator instead recovering some alternative "pseudo" representation, which absorbs the effects of breaks. It is for this reason that the full sample principal components are also known as the "pseudo-factor" method. In such cases, a split-sample method that estimates the factors using post-break data becomes a natural choice for forecasting, (Baltagi et al., 2021).

However, this existing literature on factor-augmented forecasting is incomplete, because it only considers the magnitude of the breaks, without differentiating the respective impacts of different types of breaks that occur in the factor structure. In this paper, we propose a model where the post-break loading matrix is represented as a sum of two components: a shift component that is uncorrelated with the pre-break loading matrix, and a rotational component that rotates the pre-break loading matrix. Motivated by the uncorrelatedness between the shift and the pre-break loadings, we propose a new "rotated" factor estimator. Specifically, the factors are first estimated using pre- and post-break data separately, and then the factors are rotated by minimising the L^2 distance between the pre- and post-break loading matrices. This rotation ensures that the pre- and post-break factor estimates align asymptotically, allowing them to be combined effectively to utilise pre-break data. Forecasting performance can thus be improved by mitigating the potentially significant bias-variance trade-off of a traditional split-sample approach.

Our paper makes the following theoretical contributions. First, we analyse the impacts of shift and rotational breaks on the asymptotic properties of three types of factor estimators: the (full-sample) pseudo-factors, split-sample factors, and our newly introduced rotated factors. We obtain the convergence rates of these factor estimators for different magnitudes of breaks under a local asymptotic framework. Notably, in cases where there is a small or no rotational change, we find that the rotated factor estimator can achieve the regular convergence rate obtained by Bai (2003) in factor models without breaks, even with a large shift break. Consequently, our rotated factor estimator allows for much larger shift breaks compared to the pseudo-factor estimator analysed by Bates et al. (2013).

Second, we derive the precise out-of-sample forecasting bias-variance trade-offs of the different factor estimators, and are thus able to compare their performance under different sizes of breaks. We find that the proposed rotated factors are weakly dominant for small rotational breaks, while split-sample factors are the best for large rotational breaks. For very large shift breaks or moderate rotational breaks, no single factor estimator is universally superior. As an additional byproduct of this analysis, we find that under certain conditions, the bias terms induced by the rotational and shift breaks may cancel out other to some extent, which offers an additional explanation for the successful forecast outcomes obtained with pseudo-factor estimators in empirical applications, in comparison to the small breaks framework of Bates et al. (2013).

Third, given the practical difficulty in estimating the sizes of rotational and shift breaks, we propose a cross-validation criterion to average over all possible sets of factors and obtain data-driven weights. We demonstrate that while the factor estimates are affected by the presence of structural breaks, these bias terms can be shown to be asymptotically normally distributed within the context of model averaging criteria, allowing them to be disregarded. This establishes the validity of our cross-validation criterion, and extends the results of Cheng and Hansen (2015) by incorporating structural breaks into the factor-loading matrix.

We conduct simulations to examine the impact of varying break sizes on the different sets of factor estimators, confirming the theoretical properties outlined earlier. Additionally, we assess the effectiveness of the proposed cross-validation averaging estimator in automatically assigning appropriate weights to the different factor estimates. In an empirical study, we apply the proposed methods to the FRED-QD macroeconomic dataset of McCracken and Ng (2020), with focusing on breaks associated with the Great Moderation, (considered by Stock and Watson, 2009; Breitung and Eickmeier, 2011; Baltagi et al., 2021, and others) and the Global Financial Crisis, (Cheng et al., 2016; Bai et al., 2020). By analysing this real-world dataset, we evaluate the performance of the proposed averaging estimators in comparison to existing approaches. We find that simply allowing for a break in the forecasting equation as suggested by the literature generally performs very poorly, and that estimating the factors in a way that is robust to structural breaks as we have with our proposed rotated factors offers much better performance; the application of a model averaging step then works at automatically leveraging the respective advantages of each factor estimator. Together, these findings show that the proposed estimators exhibit favourable outcomes, and the importance of incorporating structural breaks into factor-augmented forecasting models.

Our work is closely related to the existing literature on forecasting using factors estimated from factor models with structural breaks. Corradi and Swanson (2014) and Massacci (2019) introduce tests to assess whether the forecasting equation and/or the factor structure exhibit any breaks; they respectively report mixed and improved empirical out of sample forecasting performance from incorporating breaks. Stock and Watson (2009) find substantial gains for in-sample fit by accounting for the Great Moderation as a structural break. Fu et al. (2023) propose a framework that allows for time-varying factor loadings in a factor-augmented vector-autoregression (FAVAR) setting. Banerjee et al. (2008) and Bates et al. (2013) demonstrate through simulation evidence that forecast accuracy deteriorates when there is time-varying instability in the factor structure. Massacci and Kapetanios (2024) explore the effects of structural breaks in factor-augmented forecasting using the Common Correlated Effects approach (CCE) of Pesaran (2006). Hansen (2007) and Wan et al. (2010) lower the prediction loss of estimators via frequentist model averaging, an approach extended to structural breaks by Hansen (2009) and Zhang and Zhang (2023), and factor-augmented forecasting by Cheng and Hansen (2015).

Our work is different from these studies in several key aspects. First, we differentiate between the impacts of rotational and shift breaks in a local asymptotic framework that allows for both small and large magnitudes. Second, we develop the rotated factor estimator and its asymptotic properties, which is designed to be used directly and thus does not require allowing breaks in the forecasting equation. Third, we propose a model averaging approach that is robust in the presence of structural breaks based on cross-validation, which addresses the practical difficulty of knowing the magnitudes and types of breaks present in the data.

The paper is structured as follows. In Section 2, we introduce three candidate factor estimators and discuss the implementation of the cross-validation criterion for model averaging. Section 3 outlines the assumptions made in our analysis, and establishes the asymptotic properties of the factor estimators, detailed comparisons of their forecasts in terms of their out-of-sample mean squared forecast error, and the validity of the proposed cross-validation criterion. Section 4 presents our simulation experiments. Section 5 presents an empirical application of our methods. For notations, we use $\|A\| = [\text{trace}(A^\top A)]^{1/2}$ to denote the Euclidean norm of matrix A, $\lfloor . \rfloor$ to denote the floor operator, M to denote a generic finite constant, and $\xrightarrow{p}$ and $\xrightarrow{d}$ to denote convergence in probability and distribution, respectively. All proofs are relegated to the Appendix.

2 Model and Estimation

2.1 Model Setup

Suppose we have observations (y_t, x_{it}) for $t = 1, \dots, T$ and $i = 1, \dots, N$, and the goal is to produce a direct forecast for y_{T+h} using the factor-augmented regression model

$$y_{t+h} = f_t^\top \beta(L) + z_t^\top \delta + \eta_{t+h}, \tag{2.1}$$

where $h \geq 1$ is the forecast horizon, and $\beta(L)$ is a lag polynomial of order p for some $0 \leq p \leq p_{max}$. The term z_t collects all other regressors thought to improve forecasting performance; typically this includes a constant term, and y_t itself, whose lags enter in a lag polynomial $\gamma(L)$ of order q for some $0 \leq q \leq q_{max}$ collected in δ. We focus on the case where all polynomials have their roots lie outside the unit circle.

We restrict our attention to the case of a structural break in the factor structure, as there exists a breadth of literature in handling breaks in the forecasting equation itself (e.g. Pesaran et al., 2013; Corradi and Swanson, 2014). To this end, the r-dimensional factors f_t are unobserved but related to the panel of time series subject to one time break in the factor structure

$$x_{it} = \begin{cases} \lambda_{1i}^\top f_t + e_{it}, & t = 1, \dots, \lfloor \pi T \rfloor, \\ \lambda_{2i}^\top f_t + e_{it}, & t = \lfloor \pi T \rfloor + 1, \dots, T, \end{cases} \tag{2.2}$$

where $\pi \in (0, 1)$ is the break fraction, partitioning the data into $T_1 = \lfloor \pi T \rfloor$ and $T_2 = T - \lfloor \pi T \rfloor$ sized partitions, each loading onto a set of pre- and post-break loadings λ_{1i} and λ_{2i} respectively. In matrix notation, we have

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} F_1\Lambda_1^\top + e_{(1)} \\ F_2\Lambda_2^\top + e_{(2)} \end{bmatrix}, \tag{2.3}$$

where X is $T \times N$, $F = (f_1, \ldots, f_T)^\top$ is $T \times r$, $\Lambda_1 = (\lambda_{1,1}, \ldots, \lambda_{1,N})^\top$ and $\Lambda_2 = (\lambda_{2,1}, \ldots, \lambda_{2N})^\top$ are $N \times r$, and $e_{(1)}, e_{(2)}$ are the corresponding error matrices. Due to the large dimensionality of the loading matrices, the literature has documented different types of breaks that can occur in them (see Han and Inoue, 2015; Baltagi et al., 2017; Bai et al., 2024; Koo et al., 2023, and others). We show that different break types affect factor estimation, and hence factor-augmented forecasting, in different ways. Following Koo et al. (2023), we decompose the break as

$$\Lambda_2 = \Lambda_1 Z + W \tag{2.4}$$

where Z denotes a rotational change common to the cross-section, and W denotes a leftover idiosyncratic shift component that is uncorrelated with Λ_1. These two break types have become associated with breaks in the factors and breaks in the loadings, (see Wang and Liu, 2021; Pelger and Xiong, 2022; Koo et al., 2023). The case of no structural break corresponds to the case of $Z = I_r$ and $W = \mathbf{0}$.

To study the impacts of breaks of differing magnitudes, we consider parameterising Z as close to I_r, and W as close to $\mathbf{0}$

$$Z = I_r + \frac{R}{N^{1-\nu}}, \tag{2.5}$$

$$W = \frac{D}{N^{(1-\alpha)/2}}, \tag{2.6}$$

where $\frac{D^\top \Lambda_1}{N} = O_p\left(\frac{1}{\sqrt{N}}\right)$, and $\nu, \alpha \in [0, 1]$, and control the size of the rotation and shift breaks, respectively. Our formulation allows us to consider the cases of small, moderate and large rotational and shift breaks, corresponding to the case of $\nu < 0.5, \nu = 1/2, \nu < 1/2$ and respectively for α. This characterisation is related to existing frameworks employed by the literature to analyse weak *loadings* (see Bailey et al., 2021; Bai and Ng, 2023); we use it here to analyse possibly small *breaks*. The formulation in Equation (2.5) implies the following rates

$$\|I_r - Z\| = O_p\left(\frac{N^\nu}{N}\right), \tag{2.7}$$

$$\frac{\Lambda_1^\top W}{N} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \tag{2.8}$$

where the latter is implied by the Central Limit Theorem for Λ_1 and W that are uncorrelated in population, but possibly not exactly orthogonal in finite sample. Our characterisation of the shift break is compatible with the interpretation that a fraction of series have a break in their loadings. If w_i is non-zero for $i = 1, \ldots, N_1$ with $N_1 \propto N^\alpha$ and $\frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} \lambda_{1i} w_i^\top = O_p(1)$, then this implies that $\frac{\Lambda_1^\top W}{N} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)$, the same rate as Equation (2.5).

Remark. *In general, both the number of factors r and the break fraction π can be consistently estimated and can be conditioned on without affecting the main asymptotic results. Additionally, the case of a change in the number of factors can also be accommodated with a "rectangular" $r_1 \times r_1$ rotation Z where $r_2 < r_1$. For notational simplicity, we therefore treat both r and π as known and focus on the case where the number of factors remains constant. Should a practitioner wish to, we also show that with some suitable and tedious adjustments, a finite set of candidate breaks and number of factors can be averaged over in our model averaging step.*

2.2 Effects of Structural Breaks on Factor Estimates

2.2.1 Factor Space

We study the effects of a structural break on the factor estimates. It is well known that the principal components estimator as estimated over the whole sample is inherently robust to small degrees of structural changes, (see Stock and Watson, 1998; Bates et al., 2013; Baltagi et al., 2017). Our parameterisation of the structural break naturally allows us to derive the specific rates induced by the respective bias terms. To illustrate this, note that the parameterisation in Equation (2.4) implies the following equivalent representation

$$\begin{aligned} X &= \begin{bmatrix} F_1 & 0 \\ F_2 Z^\top & F_2 \end{bmatrix} \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + e \\ &= \begin{bmatrix} G_r & G_p \end{bmatrix} \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + e \\ &= G \Xi^\top + e. \end{aligned} \tag{2.9}$$

Equation (2.9) shows that if the break is ignored, the principal components estimator estimates the *pseudo*-factors G, where the first r columns G_r are subject to the effects of the rotational break (if any), and is augmented by extra r columns in the form of G_p due to the shift type break (if any). The extra r columns G_p are what is known as the augmentation effect, and is an extra bias term which depends on α. Hence, the first set of factor estimates we consider are simply $\sqrt{T}$ multiplied by the first r eigenvectors of $XX^\top/(TN)$. We denote these as $\tilde{F}_P$, as these are now understood to be the *pseudo*-factors, which are a noisy estimate of G_r. Thus, the pseudo-factors are subject to bias terms originating from both rotational and shift type breaks.

As noted by Baltagi et al. (2021), a structural break in the factors can also be accommodated by using the subsample factors $\tilde{F}_1$ and $\tilde{F}_2$, which are $\sqrt{T_1}$ times the first r eigenvectors of $X_1X_1^\top/(T_1N)$ and $\sqrt{T_2}$ times the first r eigenvectors of $X_2X_2^\top/(T_2N)$, respectively. The subsample factors recover the true factors F_1 and F_2 up to two different rotational bases; the split-sample factors $\tilde{F}_S = \left[\tilde{F}_1^\top, \tilde{F}_2^\top\right]^\top$ therefore require adding a structural break in the forecasting equation. Algebraically, this is identical to simply using the post-break data, and thus can be viewed as way to cover *all* possible structural breaks, at the potentially large cost of increased variance due to less data used in estimation.

Perhaps unsurprisingly, we show that such split-sample approaches do not work well empirically. Thus, we propose a way of combining the subsample factors directly, and thus alleviate the need for a break in the forecasting equation. To this end, we follow Koo et al. (2023) and define a set of "rotated" factors, which rotates the estimated post-break factors onto the same rotational basis as the pre-break factors, and is additionally able to purge out the effects of shift breaks. Specifically, we define the rotated factors as $\tilde{F}_R = \left[\tilde{F}_1^\top, \tilde{Z}\tilde{F}_2^\top\right]^\top$ where

$$\tilde{Z} = \left(\tilde{\Lambda}_1^\top \tilde{\Lambda}_1\right)^{-1} \tilde{\Lambda}_1^\top \tilde{\Lambda}_2 \tag{2.10}$$

is an estimate of the true rotational break using the OLS estimates of the pre- and post-break loadings $\tilde{\Lambda}_1 = \frac{1}{T_1} X_1^\top \tilde{F}_1$ and $\tilde{\Lambda}_2 = \frac{1}{T_2} X_2^\top \tilde{F}_2$, respectively. In essence, the set of rotated factors aims to be a much more robust to shift type breaks, and importantly can be used directly without the need for a break in the forecasting equation. The use of a rotational operation, however, means that the rotated factors are still subject to any rotational breaks. The maximum order of each break tolerated by the pseudo-, split-sample and rotated factors are summarised in Table 1, and are a preview of the theoretical results in Section 3.1.

Table 1: Maximum order of breaks allowed to achieve the regular convergence rate if $N \propto T$.

	Rotation	Shift	Notes
Pseudo Factors $\tilde{F}_P$	$\nu \leq 0.5$	$\alpha \leq 0.5$	Uses whole sample of data
Split Sample Factors $\tilde{F}_S$	$\nu = 1$	$\alpha = 1$	Requires break in forecast equation
Rotated Factors $\tilde{F}_R$	$\nu \leq 0.5$	$\alpha = 1$	Robust to shift breaks

2.3 Bias-variance Trade-offs

The theoretical results for the factor space allow us to analyse the precise bias-variance trade-offs for the mean squared forecast error (MSFE) across all sets of factor estimates. To begin, rewrite Equation (2.1) as

$$\begin{aligned} y_{t+h} &= c_t^\top \theta + \eta_{t+h} \\ &= \mu_t + \eta_{t+h}, \end{aligned} \tag{2.11}$$

where $c_t = \left[f_t^\top, (1, y_t)^\top\right]$ collects the regressors, $\theta = \left(\beta(L)^\top, \gamma(L)^\top\right)^\top$ collects all the lag polynomials, and μ_t denotes the conditional mean. The h-step ahead forecast is produced in using the following "two-step" approach:

1. Use x_{it} for $t = 1 : T$ to estimate $\tilde{F}_P = [\tilde{f}_{P,1}, \ldots, \tilde{f}_{P,T}]^\top$, $\tilde{F}_S = \left[\tilde{f}_{S,1}, \ldots, \tilde{f}_{S,T}\right]^\top$, and $\tilde{F}_R = \left[\tilde{f}_{R,1}, \ldots, \tilde{f}_{R,T}\right]^\top$.

2. Estimate Equation (2.11) using $\tilde{C}_P$, $\tilde{C}_S$, and $\tilde{C}_R$, the matrix counterparts of $\tilde{c}_{P,t}$, $\tilde{c}_{S,t}$ and $\tilde{c}_{R,t}$, which replace the f_t in c_t with $\tilde{f}_{P,t}$, $\tilde{f}_{S,t}$, and $\tilde{f}_{R,t}$ with data up to $T - h$, to produce $\widehat{\theta}_P$, $\widehat{\theta}_S$, and $\widehat{\theta}_R$.

3. Compute the pseudo-, split-sample, and rotated factor forecasts, respectively, as $\tilde{c}_{P,T}^\top \widehat{\theta}_P$, $\tilde{c}_{S,T}^\top \widehat{\theta}_S$, and $\tilde{c}_{R,T}^\top \widehat{\theta}_R$.

We highlight three main findings from our analysis of the bias-variance trade-offs of these forecasts; the detailed expressions for these are complex and therefore relegated to Section 3.2.

Pseudo-factors and rotated factors are asymptotically equivalent for small shift breaks, regardless of the size of rotational breaks. When the shift break is small where $\alpha < 0.5$, we find that the pseudo- and rotated factor methods recover the same factor space G_r, and therefore produce asymptotically identical forecasts.

Rotated factors weakly dominate pseudo-factors for small rotational breaks, regardless of the size of shift breaks. Although both the pseudo- and rotated factors estimate G_r, the rotated factors have the effects of shift breaks "purged out" and are therefore much more robust to them. Thus, for small rotational breaks (i.e. G_r is close to F), the rotated factors weakly dominate the pseudo-factors in terms of MSFE, regardless of the size of the shift break.

Split-sample factors dominate for large rotational breaks $\nu > 0.5$. Naturally, the fact that pseudo- and rotated factors both estimate G_r means that they are always subject to the effects of rotational breaks. Thus, when the rotational break large, both are dominated by the split-sample factors.

In practice, estimating the sizes of the shift and rotational breaks is challenging, making it difficult to determine which set of factors is best. This motivates us to develop the theoretical justification for the use of traditional frequentist criteria as a data-driven way to automatically select and/or average over the set of factor estimates.

2.4 Model Averaging and Cross-validation

2.4.1 Model Averaging Framework

Although it is possible to test for evidence of breaks in the factor structure as well as disentangle which type of break has occurred (e.g. Koo et al., 2023), it is generally difficult to estimate the corresponding size of the breaks ν and α. Additionally, forecasting strategies based on the results of hypothesis tests amount to essentially an all-or-nothing approach and are noted to not work well empirically (e.g. Hansen, 2009). Thus, in a final step we propose averaging over the possible factor estimates. Doing so naturally allows us to additionally average over an unknown lag structure in the forecasting equation, similar to Cheng and Hansen (2015). Suppose that there are $\mathcal{M}$ approximating models, each specifying a different lag structure or subset of the largest set of regressors $c_t(\mathcal{M}) = \left(1, y_t, \ldots, y_{t-p_{max}}, f_t^\top, \ldots, f_{t-q_{max}}^\top\right)^\top$. Doing so allows us to re-write Equation (2.1) in scalar and matrix forms, respectively:

$$y_{t+h} = c_t(\mathcal{M})^\top \theta + \eta_{t+h}, \tag{2.12}$$

$$Y = C(\mathcal{M})\theta + \eta. \tag{2.13}$$

Remark. *Equation* (2.1) *assumes that* y_t *is generated from* f_t*, which are the true factors subject to strict stationarity, and implicitly assumes that the rotational break is not part of the factors. Conversely, some literature interprets the rotational change as part of the factors themselves changing (e.g. Massacci, 2021; Wang and Liu, 2021; Pelger and Xiong, 2022; Koo et al., 2023), implying that* y_t *is generated from* g_t*. In this case, estimators of* g_t*, including the pseudo-factors and rotated factors, would be effective; the rotational break would not be relevant. Our model averaging approach is based on in-sample model fit, and therefore capable of automatically handling this ambiguity.*

To accommodate the possibility of a possible structural break in the factor structure, we consider three different possible sets of factor estimates: the first r pseudo-factors $\tilde{F}_P$, the split-sample factors $\tilde{F}_S$, and the rotated factors $\tilde{F}_R$. Combining the three different factor estimates with the $\mathcal{M}$ different possible lag structures yields $3 \times \mathcal{M}$ possible models in total. Without loss of generality, we define each mth set of regressors as

$$\tilde{c}_t(m) = \begin{cases} \tilde{c}_{P,t}(m) & m = 1, \ldots, \mathcal{M}, \\ \tilde{c}_{S,t}(m) & m = \mathcal{M}+1, \ldots, 2\mathcal{M}, \\ \tilde{c}_{R,t}(m) & m = 2\mathcal{M}+1, \ldots 3\mathcal{M}, \end{cases} \tag{2.14}$$

i.e. $\tilde{c}_t(m)$ contains the $\mathcal{M}$ possible lag structures for the pseudo-factors, split-sample factors, and rotated factors. The choice of lag structures to consider is not critical; a simple choice we use are sequentially nested subsets of $c_t(m)$. Defining $\tilde{C}(m)$ as the matrix counterpart of $\tilde{c}_t(m)$, the least squares estimate of $\theta(m)$ is then $\widehat{\theta}(m) = \left(\tilde{C}(m)^\top \tilde{C}(m)\right)^{-1} \tilde{C}(m)^\top Y$ with residual $\tilde{\eta}_{t+h} = y_{t+h} - \tilde{c}_t(m)^\top \widehat{\theta}(m)$. The least squares conditional forecast of y_{T+h} by the mth approximating model is

$$\widehat{y}_{t+h|T}(m) = \tilde{c}_t(m)^\top \widehat{\theta}(m). \tag{2.15}$$

Forecast combinations across all $3\mathcal{M}$ models can then be constructed by a weighted average

$$\widehat{y}_{t+h|T}(w) = \sum_{m=1}^{3\mathcal{M}} w(m) \widehat{y}_{t+h|T}(m), \tag{2.16}$$

where $w(m), m = 1, \ldots, 3\mathcal{M}$ are forecast weights such that all weights are in the unit simplex. Correspondingly, the forecast combination residual is $\widehat{\eta}_{t+h}(w) = \sum_{m=1}^{3\mathcal{M}} w(m) \widehat{\eta}_{t+h}(m)$.

2.4.2 Cross-validation Criterion

We propose the use of a post-break cross-validation for model selection and averaging in the presence of a possible structural break. In the case of no structural break, the whole sample cross-validation criterion remains valid for $h > 1$ multi-step-ahead forecasts in the case of serial correlation in η_{t+h} unlike the Mallows Criterion (Cheng and Hansen, 2015). The presence of a structural break in the regressors thus necessitates the use of post-break cross-validation residuals. To construct this criterion, define the leave-h-out prediction residual $\tilde{\eta}_{t+h,h}(m) = y_{t+h} - \tilde{c}_t(m)^\top \tilde{\theta}_{t,h}(m)$ where $\tilde{\theta}_{t,h}(m)$ is the least squares fit from a regression of y_{t+h} on $\tilde{c}_t(m)$ with the observations $\{y_{j+h}, \tilde{c}_j(m) : j = T-h+1, \ldots, T+h-1\}$ omitted. Note that this set of leave-h-out residuals uses the factors estimated from the whole sample. When $h = 1$ the leave-one-out prediction residual has the simple formula

$$\tilde{\eta}_{t+h,h}(m) = \widehat{\eta}_{t+h}(m) \left(1 - \tilde{c}_t(m)^\top \left(\tilde{C}(m)^\top \tilde{C}(m)\right)^{-1} \tilde{c}_t(m)\right)^{-1}.$$

More generally for $h > 1$, the leave-h-out residual has the formula

$$\tilde{\eta}_{t+h,h} = \widehat{\eta}_{t+h}(m) + \tilde{c}_t(m)^\top \left(\sum_{|j-t|\geq h} \tilde{c}_j(m)\tilde{c}_j(m)^\top\right)^{-1} \times \left(\sum_{|j-t|\geq h} \tilde{c}_j(m)\widehat{\eta}_{j+t}(m)\right).$$

The cross-validation criterion for forecast selection is

$$CV_{h,T}(m) = \frac{1}{\lfloor(1-\pi)T\rfloor} \sum_{t=\lfloor\pi T+1\rfloor}^{T} \tilde{\eta}_{t+h,h}(m)^2, \tag{2.17}$$

and the corresponding cross-validation selected model is $\widehat{m} = argmin_{1\leq m\leq 3\mathcal{M}} CV_{h,T}(m)$; the selected forecast is $\widehat{y}_{T+h|T}(\widehat{m})$. Let the leave-$h$-out prediction residuals for forecast combination be $\tilde{\eta}_{t+h,h}(w) = \sum_{m=1}^{3\mathcal{M}} w(m)\tilde{\eta}_{t+h,h}(m)$. The corresponding cross-validation criterion is then

$$\begin{aligned} CV_{h,T}(w) &= \frac{1}{\lfloor(1-\pi)T\rfloor} \sum_{t=\lfloor\pi T+1\rfloor}^{T} \tilde{\eta}_{t+h,h}(w)^2 \\ &= \frac{1}{\lfloor(1-\pi)T\rfloor} \sum_{t=\lfloor\pi T+1\rfloor}^{T} \left(\sum_{m=1}^{3\mathcal{M}} w(m)\tilde{\eta}_{t+h,h}(m)\right)^2. \end{aligned} \tag{2.18}$$

The cross-validation weight vector is the minimiser of the criterion:

$$\widehat{w} = \underset{w\in\mathscr{H}^{3\mathcal{M}}}{\arg\min}\, CV_{h,T}(w), \tag{2.19}$$

which is quadratic in w and can therefore be solved via quadratic programming routines. The cross-validation selected combination forecast is $\widehat{y}_{T+h|T}(\widehat{w})$, which we call the leave-h-out cross-validation averaging (CVA_h) forecast.

3 Asymptotic Theory

We provide the detailed asymptotic theory for the behaviour of the factor estimates, the bias-variance trade-offs of their subsequent forecasts, and the validity of the cross-validation procedure.

3.1 Effects on Factor Estimates

We first provide the precise theoretical justification for the effects of structural breaks in the factor structure on the proposed factor estimates. To do so, we make the following assumptions.

Assumption 1. *$E\|f_t\|^4 < \infty$, $E(f_t f_t^\top) = \Sigma_F$ and $\frac{1}{T}\sum_{t=1}^T f_t f_t^\top \xrightarrow{p} \Sigma_F$ for some positive definite Σ_F.*

Assumption 2. *There exists a positive constant $M < \infty$ such that*

a) $E\|\lambda_{1,i}\|^4 \leq M$, $\left\|\Lambda_1^\top \Lambda_1/N\right\| - \Sigma_{\Lambda_1} \xrightarrow{p} 0$ for some $\Sigma_{\Lambda_1} > 0$.

b) $Z = I_r + \frac{R}{N^{1-\nu}}$, where $\|R\| \leq M$ and $\nu \in [0,1]$.

c) $W = \frac{D}{N^{(1-\alpha)/2}}$ where $\frac{D^\top D}{N} \xrightarrow{p} \Sigma_D > 0$, $D^\top \Lambda_1 = O_p\left(\frac{1}{\sqrt{N}}\right)$ and $\alpha \in [0,1]$.

Assumption 3. *There exists a positive constant $M < \infty$ such that for all N and T:*

a) $E(e_{it}) = 0, E|e_{it}|^8 \leq M$.

b) $E(e_s^\top e_t/N) = E(N^{-1}\sum_{i=1}^N e_{is}e_{it}) = \gamma_N(s,t)$, $|\gamma_N(s,s)| \leq M$ for all s, and $T^{-1}\sum_{t=1}^T \sum_{s=1}^T |\gamma_N(s,t)| \leq M$.

c) $E(e_{it}e_{jt}) = \tau_{ij,t}$, with $|\tau_{ij,t}| < \tau_{ij}$ for some τ_{ij} and for all t. In addition, $N^{-1}\sum_{i=1}^N \sum_{j=1}^N |\tau_{ij}| \leq M$.

d) $E(e_{it}e_{js}) = \tau_{ij,ts}$, and $(NT)^{-1}\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T |\tau_{ij,ts}| \leq M$.

e) For every (t,s), $E\left|N^{-1/2}\sum_{i=1}^N [e_{is}e_{it} - E(e_{is}e_{it})]\right|^4 \leq M$.

Assumption 4. *For $m = 1,2$, the variables $\{\lambda_{m,i}\}$, $\{f_t\}$, and $\{e_{it}\}$ are mutually independent groups.*

Assumption 5. *There exists an $M < \infty$ such that for all T and N, and for every $t \leq T$ and $i \leq N$ such that:*

a) $\sum_{s=1}^{T} |\gamma_N(s,t)| \leq M$;

b) $\sum_{k=1}^{N} |\tau_{ki}| \leq M$.

Assumption 6. *There exists an $M < \infty$ such that for all N, T, and $m = 1, 2$:*

a) $E\left\| \frac{1}{NT} \sum_{s=1}^{T} \sum_{k=1}^{N} f_s [e_{ks} e_{kt} - E(e_{ks} e_{kt})] \cdot \iota_{ms} \right\|^2 \leq M$ *for each* t.

b) $E\left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{k=1}^{N} f_t \lambda_{m,k}^{\top} e_{kt} \cdot \iota_{mt} \right\|^2 \leq M$.

c) $E\left\| \frac{1}{\sqrt{N^{\alpha} T}} \sum_{t=1}^{T} \sum_{k=1}^{N} f_t w_k^{\top} e_{kt} \cdot \iota_{mt} \right\|^2 \leq M$.

d) For each t $E\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_{1,i} e_{it} \right\|^4 \leq M$.

e) For each t $E\left\| \frac{1}{\sqrt{N^{\alpha}}} \sum_{i=1}^{N} w_i e_{it} \right\|^4 \leq M$.

Assumption 7. *The eigenvalues of $(\Sigma_{\Lambda_1} \Sigma_F)$ and $(\Sigma_{\Lambda_2} \Sigma_F)$ are distinct.*

Assumption 8. *The break fraction π is bounded away from 0 and 1, and*

a) $\left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{k=1}^{N} f_t \lambda_{m,k}^{\top} e_{kt} \iota_{mt} \right\|^2 = O_p(1)$, $\left\| \frac{1}{\sqrt{NT}} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} \sum_{k=1}^{N} f_t \lambda_{m,k}^{\top} e_{kt} \iota_{mt} \right\|^2 = O_p(1)$ *for* $m = 1, 2$, *and*

b) $\left\| \frac{\sqrt{T}}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} (f_t f_t^{\top} - \Sigma_F) \right\| = O_p(1)$, *and* $\left\| \frac{\sqrt{T}}{T - \lfloor \pi T \rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} (f_t f_t^{\top} - \Sigma_F) \right\| = O_p(1)$.

Assumptions 1 to 7 are either straight from, or slight modifications of, those in Bai (2003). Assumption 1 is the same as Assumption A in Bai (2003), except that we require the second moment of f_t to be time invariant. This additional "strict" stationarity assumption is common as an identification condition (e.g. Han and Inoue, 2015; Baltagi et al., 2017, and others). Assumption 2 (a) is the same as Assumption B in Bai (2003), and allows for the loadings to be random. Assumptions 2 (b) and 2 (c) characterise the sizes of the rotational and shift breaks, respectively. Assumption 3 allows for weak serial and cross-sectional correlation and defines the *approximate* factor model, corresponding to Assumption C of Bai (2003). Assumption 4 is standard in the factor modelling literature, and is the subsample version of Assumption D of Bai and Ng (2006). Assumption 5 is a strengthened version of Assumption 3, but still allows for heterogeneity in time and cross-sectional dimensions, corresponding to Assumption E in Bai (2003). Assumption 6 corresponds to Assumptions F1-F2 in Bai (2003). Although we require Assumption 6, which are moment conditions in Bai (2003), asymptotic normality of $N^{-1/2}\sum_{i=1}^{N}\lambda_i e_{it}$ are not required for the purposes of estimation. Also, Assumption 6 (c) is slightly stronger that Assumption F3 of Bai (2003), which only requires the existence of the second moments. Assumption 7 corresponds to Assumption G in Bai (2003). Assumption 8 requires that the sample sizes before and after the potential break date go to infinity. It is a weaker version of Assumption 8 in Han and Inoue (2015), who assumes that the terms are bounded uniformly in a range of potential π.

Remark. *Similar to Koo et al. (2023) we require the break fraction π and the number of factors r pre- and post-break to be known. This is not restrictive, as several consistent estimates of π exist (e.g. Baltagi et al., 2017; Bai et al., 2020, 2024). Conditional on some consistent estimate $\widehat{\pi}$, the subsample factors $\tilde{F}_1$ and $\tilde{F}_2$ are able to achieve the usual $O_p\left(\delta_{NT}^{-2}\right)$ consistency rate, and r can be estimated consistently by either applying consistent estimators of r such as the information criterion of Bai and Ng (2002) in either subsample (see Baltagi et al., 2017), or using an information criterion robust to breaks over the whole sample (see Su and Wang, 2017). With some adjustments, our theoretical results also hold as long as the number of factors specified by the practitioner does not exceed r, similar to Cheng and Hansen (2015). For notational clarity, we proceed as if r is known. If practitioners wish to consider different candidate r and π, these can simply be averaged over in our model averaging step following some suitable adjustments to the theory.*[1]

Pseudo-factors

To analyse the asymptotic properties of $\tilde{F}_P$, we separate the analysis in the two cases of $\alpha < 1$ and $\alpha = 1$. In the case of $\alpha < 1$, the analysis of $\tilde{F}_P$ proceeds by treating the first r factors G_r as "strong" factors, and the additional G_p columns induced by the shift break as additional noise. Hence, $\tilde{F}_R$ is estimating $G_r H_G$ where the normalisation basis is defined as

$$H_G = \frac{\Lambda_1^\top \Lambda_1}{N}\frac{G_r^\top \tilde{F}_P}{T}V_{NT,r}^{-1}, \tag{3.1}$$

where $V_{NT,r}$ is a diagonal matrix of the first r eigenvalues of $(NT)^{-1}XX^\top$ in descending order.

However, when $\alpha = 1$ the shift break is too large to ignore, and hence H_G is unsuitable in the sense that it does not have a well-defined limit. In this case, we can recognise that the factor structure now consists of $2r$ "strong" factors $G = \begin{bmatrix} G_r & G_p \end{bmatrix}$ which load onto the pseudo- loadings Ξ. Hence, $\tilde{F}_P$ which are the first r eigenvectors can be analysed as a subset of $\tilde{G}$, the first $2r$ principal components, and we are

[1] See Appendices A.5 and A.7

able to specify a normalisation basis with a valid probability limit[2] as

$$H_{\Xi,r} = \frac{\Xi^\top \Xi}{N} \frac{G^\top \tilde{F}_P}{T} V_{NT,r}^{-1}. \tag{3.2}$$

Split Sample Factors

The results for using the split-sample factors $\tilde{F}_S$ follow from Bai and Ng (2002). Define the following subsample rotational bases as

$$H_1 = \frac{\Lambda_1^\top \Lambda_1}{N} \frac{F_1^\top \tilde{F}_1}{T_1} V_{NT,1}^{-1}, \quad H_2 = \frac{\Lambda_2^\top \Lambda_2}{N} \frac{F_2^\top \tilde{F}_2}{T_2} V_{NT,2}^{-1}, \tag{3.3}$$

where $V_{NT,1}$ and $V_{NT,2}$ are diagonal matrices consisting of the first r eigenvalues of $X_1 X_1^\top/(NT_1)$ and $X_2 X_2^\top/(NT_2)$, respectively. However, in general, $H_1 \neq H_2$, and this necessitates introducing allowing for a break in the forecasting equation. This is algebraically equivalent to using the post-break data to yield an unbiased estimator, at the potentially large cost of increased variance.

Rotated Factors

The rotated factors $\tilde{F}_R$ are designed to overcome the shortcoming of the split-sample factors, and produce a set of factors on the same normalisation basis that are robust to structural breaks.

Proposition 1. *Under Assumptions 1 to 8 and as $N, T \to \infty$,*

$$\tilde{Z} = H_1^\top Z H_2^{-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).$$

The proof of Proposition 1 in provided in Appendix A.4. Because $\tilde{F}_2$ estimates $F_2 H_2$, and $\tilde{F}_1$ estimates $F_1 H_1$, Proposition 1 shows that the post-break factors $\tilde{F}_2$ can be rotated onto the same basis as $\tilde{F}_1$ by simply post-multiplying it by $\tilde{Z}^\top$. Because the shift break W is uncorrelated with Λ_1, this operation "purges out" any shift breaks. Note, however, that this rotation operation absorbs the effect of any rotational break Z, and is therefore not robust to this type of break.

With the above specification of the various normalisation bases, the consistency rates of the pseudo, split-sample, and rotated factors can be summarised in the following theorem.

Theorem 1. *Under Assumptions 1 to 8, as $N, T \to \infty$,*

[2] See Lemma 2.

a) The pseudo-factors $\tilde{F}_P$ satisfy:

$$T^{-1}\left\|\tilde{F}_P - G_r H_G\right\|^2 = O_p\left(\frac{1}{\delta^2_{NT}}\right) + O_p\left(\frac{N^{2\alpha}}{N^2}\right), \qquad \textit{for} \quad \alpha < 1,$$

$$T^{-1}\left\|\tilde{F}_P - F H_G\right\|^2 = O_p\left(\frac{1}{\delta^2_{NT}}\right) + O_p\left(\frac{N^{2\alpha}}{N^2}\right) + O_p\left(\frac{N^{2\nu}}{N^2}\right), \qquad \textit{for} \quad \alpha < 1, \ \textit{and}$$

$$T^{-1}\left\|\tilde{F}_P - G H_{\Xi,r}\right\|^2 = O_p\left(\frac{1}{\delta^2_{NT}}\right), \qquad \textit{for} \quad \alpha = 1,$$

b) The split-sample factors $\tilde{F}_S = [\tilde{F}_1^\top, \tilde{F}_2^\top]^\top$ for $\iota = 1, 2$ satisfy:

$$T^{-1}\left\|\tilde{F}_\iota - F_\iota H_\iota\right\|^2 = O_p\left(\frac{1}{\delta^2_{NT}}\right),$$

c) The rotated factors $\tilde{F}_R = \left[\tilde{F}_1^\top, \tilde{Z}\tilde{F}_2^\top\right]^\top$ satisfy:

$$T^{-1}\left\|\tilde{F}_R - G_r H_1\right\|^2 = O_p\left(\frac{1}{\delta^2_{NT}}\right) + O_p\left(\frac{N^{\alpha}}{N^2}\right), \ \textit{and}$$

$$T^{-1}\left\|\tilde{F}_R - F H_1\right\|^2 = O_p\left(\frac{1}{\delta^2_{NT}}\right) + O_p\left(\frac{N^{\alpha}}{N^2}\right) + O_p\left(\frac{N^{2\nu}}{N^2}\right).$$

Theorem 1 (a) provides the convergence rates for the pseudo-factors $\tilde{F}_P$. For $\alpha < 1$, the consistency result is stated in terms of G_r which absorbs the effects of the rotational break into the factor space, and F, the original factor space. For $\alpha = 1$, Theorem 1 (a) is stated in terms of the consistency to $GH_{\Xi,r}$, and therefore formalises how $\tilde{F}_P$ estimates a linear combination of G_r and G_p when the shift break is very large. Theorem 1 (b) are simply the subsample versions of Theorem 1 of Bai and Ng (2002), and show that $\tilde{F}_1$ and $\tilde{F}_2$ are estimating F_1H_1 and F_2H_2 respectively. Because the normalisation bases H_1 and H_2 generally differ, this necessitates a break in the forecasting equation. Theorem 1 (c) presents the mean square consistency results for the rotated factors $\tilde{F}_R$, and is similarly presented in terms of both G_r and F. It shows that $\tilde{F}_R$ can tolerate much larger values of α compared to the pseudo-factors $\tilde{F}_P$; the second part similarly follows by adding and subtracting the true factors F. In either case, because $\alpha \in [0, 1]$, the additional $O_p\left(N^{\alpha-2}\right)$ term arising from the shift break is no larger than the usual $O_p\left(\delta_{NT}^{-2}\right)$ rate.

3.2 Forecasting Bias-variance Trade-offs

3.2.1 Model and Expansion Results

We next provide the precise theoretical treatment of the bias-variance trade-offs for out-of-sample forecasting of the different factor estimators. Equation (2.1) can be rewritten in matrix form as

$$
\begin{aligned}
Y &= F\beta(L) + z\delta + \eta \\
&= C\theta + \eta \\
&= \mu + \eta.
\end{aligned} \tag{3.4}
$$

For simplicity, we treat the lag structure $\beta(L)$ and $\gamma(L)$ as known for this section - an extension to an unknown lag structure can be handled by our model averaging framework in Section 3.3.

To analyse the effects of the structural break on the forecasting equation, we make the following additional assumptions. Let $\mathcal{F}_t = \sigma\left(y_t, f_t, X_{1t}, X_{2t}, \ldots, f_{t-1}, y_{t-1}, x_{1,t-1}, x_{2,t-1}, \ldots\right)$ denote the information set at time t.

Assumption 9.

a) $\mathbb{E}\left(\eta_{t+h}|\mathcal{F}_t\right) = 0$.

b) $\left(c_t^\top, \eta_{t+h}, e_{1t}, \ldots, e_{Nt}\right)$ *is piece-wise strictly stationary and ergodic before and after the break.*

c) $\mathbb{E}\|c_t\|^4 \leq M$, $\mathbb{E}\eta_t^4 \leq M$, *and* $\frac{1}{T}\sum_{t=1}^{T}\left(c_t c_t^\top\right) \xrightarrow{p} \Sigma_{CC} > 0$.

d) $\frac{1}{\sqrt{T}}\sum_{t=1-h}^{T-h} c_t\eta_{t+h} \xrightarrow{d} N(0, \Omega_{CC,\eta})$, *where* $\Omega_{CC,\eta} = \operatorname{plim} \frac{1}{T}\sum_{t=1-h}^{T-h}\eta_{t+h}^2 c_t c_t^\top > 0$.

Assumption 9 places additional assumptions on the forecasting error term η_t, and follows from Assumption R of Cheng and Hansen (2015). Assumption 9 (a) implies that η_{t+h} is conditionally unpredictable at time t, but does not imply that η_{t+h} is serially uncorrelated when $h > 1$. This is consistent with the fact that η_{t+h} is typically a moving average process of order $h-1$. Assumption 9 (b) assumes that the data is piece-wise stationary and ergodic pre- and post-break. Assumptions 9 (c) and 9 (d) are standard moment conditions and central limit theorems, the latter of which is satisfied under standard weak dependence conditions.

With some suitable adjustments on the parameterisation of Z and W, we can accommodate the case of disappearing factors, and therefore allow for r to be different before and after the break. Similar to Cheng and Hansen (2015) and the existing literature, approximating models may use less than r factors, but cannot use more than r factors. This allows practitioners to average over a set of candidate number of factors.

Using the rates derived in Theorem 1, we show the pseudo-, split-sample, and rotated factor methods have the following expressions for their out-of-sample squared biases.

Proposition 2. *Under Assumptions 1 to 9, as $N, T \to \infty$ and under the condition that $N \propto T$, then:*

$$\text{bias}(\tilde{c}_{P,T}^{\top}\widehat{\theta}_P) = \left[\left((I-Z) - \left(\frac{\Lambda_1^{\top}\Lambda_1}{N}\right)^{-1}\left(\frac{\tilde{F}_P^{\top}G_r}{T}\right)^{-1}\frac{\tilde{F}_P^{\top}G_p}{T}\frac{W^{\top}W}{N}\right)\left(f_T - \frac{G_p^{\top}\tilde{C}_P}{T}\left(\frac{\tilde{C}_P^{\top}\tilde{C}_P}{T}\right)^{-1}\tilde{c}_{P,T}\right)\right.$$
$$\left. - \left(\frac{\Lambda_1^{\top}\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^{\top}e_T}{N}\right]^{\top}\beta(L) + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right), \tag{3.5}$$

$$\text{bias}(\tilde{c}_{S,T}^{\top}\widehat{\theta}_S)^2 = \frac{-e_T^{\top}\Lambda_2}{N}\left(\frac{\Lambda_2^{\top}\Lambda_2}{N}\right)^{-1}\beta(L) + O_p\left(\frac{1}{\delta_{NT}^2}\right), \tag{3.6}$$

$$\text{bias}(\tilde{c}_{R,T}^{\top}\widehat{\theta}_R)^2 = \left(H_1^{\top}(I-Z)\left(f_T - \frac{\tilde{F}_2^{\top}\tilde{C}_{R,2}}{T}\left(\frac{\tilde{C}_R^{\top}\tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\right) - H_1^{\top}Z\left(\frac{\Lambda_2^{\top}\Lambda_2}{N}\right)^{-1}\frac{\Lambda_2^{\top}e_T}{N}\right)^{\top}H_1^{-1}\beta(L)$$
$$+ O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right). \tag{3.7}$$

Equations (3.5) to (3.7) in Proposition 2 express the squared bias in terms of the rotational break $(I - Z)$, shift break $\left(W^{\top}W/N\right)$, and inherent estimation uncertainty in the factors e_T. The additional variance and cross terms which additionally make up the total squared loss are of order $O_p(1)$ and negligible, respectively, with their specific forms are relegated to Appendix B of the Appendix. Therefore, by analysing these bias terms in detail for $(\alpha, \nu) \in \{[0, 0.5), 0.5, (0.5, 1]\}$, corresponding to small, moderate, and large breaks, we have the following comparisons between the different forecasts.

Theorem 2. *Under Assumptions 1 to 9, as $N, T \to \infty$ and under the condition that $N \propto T$, then:*

a) For small shift breaks $\alpha < 0.5$, $\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - \tilde{c}_{R,T}^{\top}\widehat{\theta}_R = o_p(N^{-1/2})$,

b) For small rotational breaks $\nu < 0.5$,

$$E\left(\text{plim}_{N,T\to\infty} N\left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2\right) < E\left(\text{plim}_{N,T\to\infty} N\left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - c_T^{\top}\theta\right\|^2\right), \qquad \textit{for} \quad \alpha = 0.5, \textit{ and}$$
$$\left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2 / \left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - c_T^{\top}\theta\right\|^2 \xrightarrow{p} 0, \qquad \textit{for} \quad \alpha > 0.5,$$

c) For moderate rotational breaks $\nu = 0.5$,

$$\left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - c_T^{\top}\theta\right\|^2 \asymp_p \left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2 \asymp_p \left\|\tilde{c}_{S,T}^{\top}\widehat{\theta}_S - c_T^{\top}\theta\right\|^2, \qquad \textit{for} \quad \alpha = 0.5, \textit{ and}$$

$$\left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2 \asymp_p \left\|\tilde{c}_{S,T}^{\top}\widehat{\theta}_S - c_T^{\top}\theta\right\|^2,$$
$$\left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - c_T^{\top}\theta\right\|^2 / max\left[\left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2, \left\|\tilde{c}_{S,T}^{\top}\widehat{\theta}_S - c_T^{\top}\theta\right\|^2\right] \xrightarrow{p} \infty, \qquad for \quad \alpha > 0.5,$$

d) For large rotational breaks $\nu > 0.5$

$$\left\|\tilde{c}_{S,T}^{\top}\widehat{\theta}_S - c_T^{\top}\theta\right\|^2 / min\left[\left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2, \left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - c_T^{\top}\theta\right\|^2\right] \xrightarrow{p} 0.$$

Theorem 2 provides the detailed comparisons between the different forecasts produced by each set of factor estimates for varying sizes of shift and rotational breaks, which we summarise into four cases. Theorem 2 (a) implies that $\left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - \tilde{c}_{R,T}^{\top}\widehat{\theta}_R\right\|^2 / max\left[\left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - c_T^{\top}\theta\right\|^2, \left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2\right] \xrightarrow{p} 0$, and shows the asymptotic equivalence between the pseudo-factors and the rotated factors for $\alpha < 0.5$. This result holds regardless of the size of the rotational break, and follows because both the pseudo- and rotated factors $\tilde{F}_P$ and $\tilde{F}_R$ are estimating G_r, the first r pseudo-factors. Theorem 2 (b) shows how the rotated factors weakly dominate the pseudo-factors when the rotational break is small. Theorem 2 (c) shows that rotated and split-sample factors have MSFEs that are of the same asymptotic order for moderate rotational breaks; additionally if the shift break is also moderate, then the MSFE of the pseudo-factors is also of the same asymptotic order. This represents the region where the magnitude induced by the break terms is on the same order as the loss in efficiency from using the split-sample factors, and therefore the specific ranking of each method depends on the data-generating process. Theorem 2 (d) shows that the both the pseudo- and rotated factors cannot handle large rotational breaks, and are therefore dominated by the split-sample factors. For clarity, the results of Theorem 2 are summarised in Table 2.

	$\nu < 0.5$	$\nu = 0.5$	$\nu > 0.5$
$\alpha < 0.5$			
$\alpha = 0.5$	R		
$0.5 < \alpha < 1$			S
$\alpha = 1$			

Table 2: Summary of Theorem 2. Yellow region represents rotated factors are the best, orange represents the split-sample factors are the best, white represents no dominating method. Red box represents region where rotated factors dominate the pseudo-factors, blue box represents where the rotated factors are equivalent to pseudo-factors.

Remark. *The pseudo-factors $\tilde{F}_P$ are subject to a possible additional "bias cancellation" effect for large shift and rotational breaks. This is due to the cross term between the shift and rotational breaks which appears in the expression when calculating the MSFE, which may be negative depending on the specific* *data-generating process. This cross term is asymptotically relevant only when $\alpha = \nu$; thus in finite sample this effect can appear when α and ν are similar and greater than or equal to* $1/2$.

3.3 Forecast Model Selection and Averaging

Next, we provide the theoretical justification of the proposed cross-validation selection and averaging procedure, which holds even in the context of a structural break, and for $h > 1$ or if the errors are possibly conditionally heteroskedastic. First, it helps to understand that a h-step ahead forecast is actually a specific leave-h-out estimator. Following Hansen (2010) and Cheng and Hansen (2015), the h-step ahead forecast is $\widehat{y}_{T+h|T}(m) = \tilde{c}_T(m)^\top \widehat{\theta}(m)$, where $\widehat{\theta}(m)$ is the least squares estimate with data sample $\{y_{t+h}, \tilde{c}_t(m) : t = 1-h, \dots, T-h\}$. Compared to a leave-$h$-out[3] estimator $\tilde{\theta}_{T,h}(m)$ with the last h observations $\{y_{j+h}, \tilde{c}_j(m) : j = T-h+1, \dots, T+h-1\}$ omitted, the sample used in estimation is identical. Hence, $\widehat{\theta}(m) = \tilde{\theta}_{T,h}(m)$, and the h-step ahead forecast can be written as $\widehat{y}_{T+h|T}(m) = \tilde{c}_T(m)^\top \tilde{\theta}_{T,h}(m)$. The forecast error is also equivalent to the leave-h-out prediction residual and is $y_{T+h} - \widehat{y}_{T+h|T}(m) = y_{T+h} - \tilde{c}_T(m)^\top \tilde{\theta}_{T,h}(m) = \tilde{\eta}_{T+h,h}(w)$. The MSFE of the point forecast equals

$$MSFE_T(w) = \mathbb{E}\left(y_{T+h} - \widehat{y}_{T+h|T}(w)\right)^2 = \mathbb{E}\tilde{\eta}^2_{T+h,h}(w)^2. \tag{3.8}$$

This equivalence between the MSFE and the expected post-break squared leave-h-out prediction residual, allows us to view the cross-validation criterion as a natural estimator of the expectation $\mathbb{E}\tilde{\eta}^2_{T+h,h}(w)^2$.

Let the leave-h-out fitted values for the mth model be $\tilde{\mu}_{t,h}(m) = \tilde{c}_t(m)^\top \tilde{\theta}_{t,h}(m)$ and for the weighted model as $\tilde{\mu}_{t,h}(w) = \sum_{m=1}^{3\mathcal{M}} w(m)\tilde{c}_t(m)^\top \tilde{\theta}_{t,h}(m)$. The leave-$h$-out prediction residuals are $\tilde{\eta}_{t+h,h}(w) = y_{t+h} - \tilde{\mu}_{t,h}(w)$, or equivalently using vector notation, $\tilde{\eta}_h(w) = \eta + \mu - \tilde{\mu}_h(w)$. Therefore, we have

$$\begin{aligned} CV_{h,T}(w) &= \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} \tilde{\eta}_{t,h}(w)^\top \tilde{\eta}_{t,h}(w) \\ &= \tilde{L}_{T_2}(w) + \frac{1}{T}\eta_{(2)}^\top \eta_{(2)} + \frac{2}{\sqrt{T}}\tilde{r}_{1T}(w) \end{aligned} \tag{3.9}$$

where $\eta_{(2)}$ represents the vector of post-break errors,

$$\tilde{L}_{T_2}(w) = \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} (\mu_t - \tilde{\mu}_{t,h}(w))^2$$

[3] We follow the terminology used by Hansen (2010) and Cheng and Hansen (2015); in reality, as noted by Hansen (2010), this is actually a leave-$(2h-1)$-out cross-validation estimator, where the $h-1$ observations within observation t are removed.

$$= \frac{1}{T_2} \left(\mu_{(2)} - \tilde{\mu}_{(2),h}(w)\right)^\top \left(\mu_{(2)} - \tilde{\mu}_{(2),h}(w)\right) \tag{3.10}$$

is the post-break in-sample squared error from the leave h out estimator, and

$$\begin{aligned}
\tilde{r}_{1T}(w) &= \frac{1}{\sqrt{T_2}} \left(\mu_2 - \tilde{\mu}_{2,h}(w)\right)^\top \eta_{(2)} \\
&= \sum_{m=1}^{M} w(m) \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1-h}^{T-h} \left(\mu_t - \tilde{c}_t(m)^\top \tilde{\theta}_{t,h}(m)\right) \eta_{t+h} \\
&= \sum_{m=1}^{M} w(m) \tilde{r}_{1T}(m). \qquad (3.11)
\end{aligned}$$

Thus, provided that $\tilde{r}_{1T}(m)$ can be ignored, the post-break cross-validation criterion is a natural estimate of the post-break MSFE. Similar to Cheng and Hansen (2015), our strategy is to show that $\tilde{r}_{1T}(m)$ is asymptotically normally distributed with zero mean, and hence can be ignored when calculating the cross-validation criterion. Define $\theta(m) = \left(C_H(m)^\top C_H(m)\right)^{-1} C_H(m)^\top y$ as the projection coefficient from the regression of y_{t+h} onto $c_{Ht}(m)$, where $C_H(m) = C(m)H(m)$ and $H(m)$ is a rotation matrix which suitably transforms the columns of $C(m)$.[4] This allows us to establish the asymptotic negligibility of $\tilde{r}_{1T}(m)$, and therefore legitimacy of the post-break cross-validation criterion.

Proposition 3. *Under Assumptions 1 to 9,*

$$\begin{aligned}
\tilde{r}_{1T}(m) &\xrightarrow{d} S_1(m) \sim N(0, \sigma^2 Q(m)), \\
\tilde{r}_{1T}(w) &\xrightarrow{d} \xi_1(w) = \sum_{m=1}^{3\mathcal{M}} w(m) S_1(m),
\end{aligned}$$

where $Q(m) = \operatorname{plim}_{T\to\infty} \frac{1}{(1-\pi)^2} \frac{1}{T} (\mu_{(2)} - C_{2,H}(m)\theta(m))^\top (\mu_{(2)} - C_{2,H}(m)\theta(m))$, and $C_{2,H}(m)$ are the post break rows of $C_H(m)$.

Theorem 3. *Under Assumptions 1 to 9, we have for any $h \geq 1$, fixed $\mathcal{M}$ and w, and $N, T \to \infty$,*

$$CV_{h,T}(w) = \tilde{L}_{T_2}(w) + \frac{1}{T_2} \eta_{(2)}^\top \eta_{(2)} + \frac{2}{\sqrt{T_2}} \tilde{r}_{1T}(w),$$

where $\tilde{r}_{1T}(w) \xrightarrow{d} \xi_1(w)$ and $E\xi_1(w) = 0$.

[4] The exact form of $H(m)$ follows similarly to the definition of H in Lemma A.1 of Bai and Ng (2006), with suitable adjustments so that the appropriate subsets of lags of y_t and f_t are allowed. Specifically, $H(m)$ is a block diagonal matrix where the top upper left block associated with the lags of y_t are identity, and the bottom right block associated with the factors are a suitable choice of rotational basis with a valid limit, i.e. H_G or $H_{\Xi,r}$ for the pseudo-factors depending on whether $\alpha < 1$ or $\alpha = 1$, and H_1 for the split-sample or rotated factors.

Theorem 3 shows that $CV_{h,T}(w)$ is an asymptotically unbiased estimate of $\tilde{L}_{T_2}(w)$, the in-sample squared loss from the leave-h-out estimator, plus σ^2. This holds for any weight vector, for any set of estimated factors considered, for any forecast horizon, and allows for conditional heteroskedasticity. Theorem 3 mirrors and extends Theorem 2 of Cheng and Hansen (2015) to allow for the case of a structural break in the factor structure.

Remark. *It is important to note that the true weight vector w need not be unique, which may cause some convergence issues in estimation, but does not affect forecasting performance. This is a direct consequence of Theorem 2 (a), which states that the pseudo- and rotated factors produce asymptotically equivalent forecasts when the shift break is small. Consequently, any numerical instability in the weight vector merely reflects the similar quality of the different factor estimates and poses no issues for forecasting.*

4 Monte Carlo Study

4.1 Model Specification

We investigate the finite sample MSFE of the proposed factor estimators by themselves as well as in conjunction with cross-validation selection and averaging. The data-generating process follows that of Bai and Ng (2009) and Cheng and Hansen (2015), but we focus on linear models and add a structural break in the factor structure.

We generate $\lambda_{1i} \sim N(0, I_r)$, which can be stacked to form Λ_1. We then generate the rotational and shift break components respectively as

$$Z = I_r + \frac{R}{N^{1-\nu}}; \quad R \sim i.i.d.MVN(0_r, I_r), \quad W = \frac{1.5 \times D}{N^{(1-\alpha)/2}}; \quad D \sim i.i.d.MVN(0, I_r). \tag{4.1}$$

This allows us to generate the post-break loadings $\Lambda_2 = \Lambda_1 Z + W$. The approximate factor model with a structural break is then

$$x_{it} = \begin{cases} \lambda_{1,i}^\top f_t + \sqrt{\theta} e_{it}, & t = 1, \ldots, \lfloor \pi T \rfloor \\ \lambda_{2,i}^\top f_t + \sqrt{\theta} e_{it}, & t = \lfloor \pi T \rfloor + 1, \ldots, T, \end{cases} \tag{4.2}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The parameter θ is set to 6 in order to calibrate the signal to noise ratio to be 50%.

The factors and errors are generated as follows:

$$f_{k,t} = \rho f_{k,t-1} + u_{it}, u_{it} \sim i.i.d.N(0, 1-\rho^2), \tag{4.3}$$

$$e_{it} = \alpha e_{i,t-1} + v_{it}, \tag{4.4}$$

where $\rho \in \{0, 0.7\}$ captures the serial correlation in the factors, and ϵ_{it}, v_{it} are mutually independent with $v_t = (v_{1,t}, \dots, v_{N,t})^\top$ being i.i.d. $N(0, \Omega)$ for $t = 1, \dots, T$. For $t = 1$, $e_{\cdot t} = (e_{1,1}, \dots, e_{N,1})^\top$ is $N\left(0, \frac{1}{1-\alpha^2}\Omega\right)$ to initialise the errors at their stationary distributions. The scalar α captures the serial correlation in the errors, and as in Bates et al. (2013) and Baltagi et al. (2017), $\Omega_{ij} = \beta^{|i-j|}$ captures the cross-sectional correlation in the errors. We consider $\alpha = \beta = 0.3$ to allow for mild serial and cross-sectional correlation. The true break fraction is set to 0.5 and treated as known.[5]

The regression equation for the forecast is

$$y_{t+h} = \beta_1 f_t + \beta_2 f_{t-1} + \beta_3 f_{t-2} + \eta_{t+h} \tag{4.5}$$

$$\eta_{t+h} = \sum_{j=1}^{h-1} \kappa^j \varepsilon_{t+h-j} \tag{4.6}$$

where $\beta_1 = 0.5, \beta_2 = 0.2, \beta_3 = 0.1$, and $\varepsilon_t \sim N(0, 1)$ i.i.d. over t and is independent of v_{is} and u_{is} for all t and s. For multi-step forecasting, the moving average parameter κ controls the serial dependence in the error term, which we set to $\kappa \in \{0.1, 0.5, 0.9\}$. The sample size is $N = 100$ and $T = (200, 500)$, and 1,000 simulation repetitions are conducted.

We treat the number of factors r as known. The first r factors are then estimated as the 1) the first r pseudo-factors $\tilde{F}_P$, 2) the split/post-break factors $\tilde{F}_S$, and 3) the rotated factors $\tilde{F}_R$. For each set of possible factors, the set of candidate regressors for the model averaging and model selection is

$$\mathcal{C}_t = (1, \tilde{f}_t^\top, \dots, \tilde{f}_{t-q_{max}}^\top, y_t, \dots, y_{t-p_{max}}). \tag{4.7}$$

Feasible models are constructed using all possible combinations of lags of q and p. We consider $q_{max} = p_{max} = 4$, and this yields a total of $3 \times (4 \times 4)$ models in total.

We compare the MSFE of various model averaging and model selection methods. The model averaging methods include leave-h-out cross-validation averaging (CVA_h), Mallows model averaging (MMA),

[5] Additional results for $\pi \in \{0.3, 0.7\}$ are similar and omitted.

Bayesian model averaging (BMA),[6] and simple averaging with equal weights. The model selection methods include the proposed post-break leave-h-out cross-validation and Mallows selection of Cheng and Hansen (2015).

4.2 Results

Across most parameter values and all forecast horizons, the proposed post-break leave-h-out cross-validation averaged forecasts yield the smallest RMSE and hence the best forecasting performance. For compactness, we report the results of each factor estimator and the model averaging estimators in Section 4.2, with poorly performing models omitted. Specifically, other data adaptive weighted forecasts such as BIC-weighted and Mallows weighted forecasts offered similar, but slightly worse performance compared to post-break cross validation weighted forecasts, whereas all model selection methods and equal weighted forecasts were dominated. The RMSE are normalised by the RMSE of the infeasible forecast using the true unobserved factors.

In the case of small to moderate rotational break where $\nu \leq 0.5$, both the pseudo-factors and rotated factors show deteriorated performance as α increases, although the rotated factors are significantly more robust. For the case of a large rotational $\nu = 1$, the pseudo-factors perform better as α increases, whereas the rotated factors' performance stays constant. This seemingly counter-intuitive result is because the MSFE of the pseudo-factors are subject to both the shift and rotational breaks, and therefore, the product of their cross terms as well. For the case of a large rotational break, this means that large bias from large ν can be diluted by the effect increasing α, depending on the signs of the bias terms. In contrast, the rotated factors are designed to be a more precise estimator of G_r, and hence increasing α does not improve performance. Neither of these estimators perform well as ν increases, and remain dominated by the split-sample factors and by extension, weighted forecasts, which can leverage their effect.

4.3 Value of Rotated factors

A main contribution of our paper is the set of rotated factors $\tilde{F}_R$, which are robust to shift type breaks without the need for including a break in the forecasting equation. To illustrate this, we conduct an additional simulation study and compare the model averaged forecasts constructed from two model sets: one with the full set of possible factor estimates, and one with a model set that excludes the rotated factors $\tilde{F}_R$, on a specification of $\nu = 0$ and varying levels of α. Section 4.3 displays the results of this exercise.

[6] Our Bayesian model averaging weights are set as $w(m) = exp(-BIC(m)/2)/\sum_{m=1}^{3\mathcal{M}} exp(-BIC(m)/2)$, where $BIC(m)$ is the BIC for the mth model. This approximates the case of equal model priors and diffuse model priors on parameters.

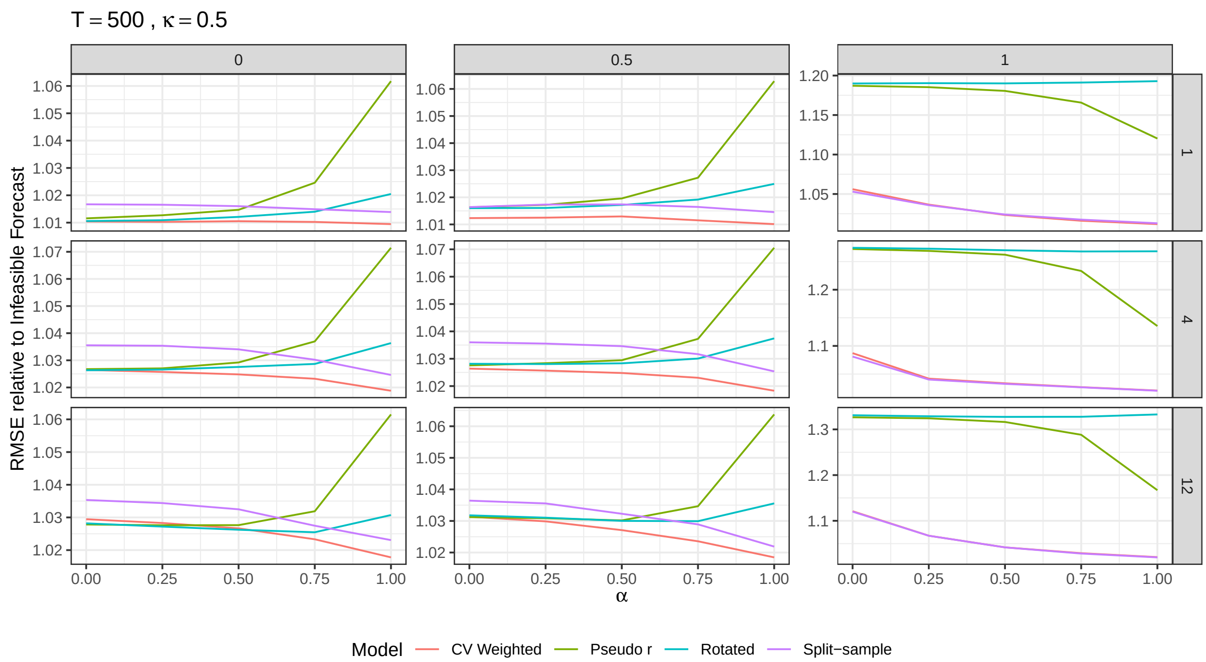

Figure 1: Relative MSFE for each factor estimator and proposed post-break cross-validation weighted forecasts, faceted by h (rows) and ν (columns), $\kappa = 0.5$ for moderate serial correlation in errors, $q_{max} = p_{max} = 4$.

The results clearly show that, across all forecasting horizons and regardless of the model averaging method used, excluding the rotated factors $\tilde{F}_R$ results in poorer forecasting performance. Thus, this demonstrates the value of using the rotated factors, as they offer a parsimonious way of adding possible robustness to shift type breaks in the factor structure. Additionally, although not theoretically validated, we find that the Bayesian model averaged forecasts can tend to produce good forecasts, particularly when the size of the breaks is not too large.

5 Empirical Study

5.1 Data

We apply the proposed sets of factor estimates that deal with a possible structural break in the factor structure in combination with the proposed cross-validation selection and averaging methods to forecast U.S. macroeconomic series. We compare their performance with model averaging approaches that do not consider possible structural breaks and use the principal components over the whole subsample - i.e. the pseudo-factors; this approach corresponds to the frequentist averaging approach of Cheng and Hansen (2015), which only averages over the number of factors and lag structure.

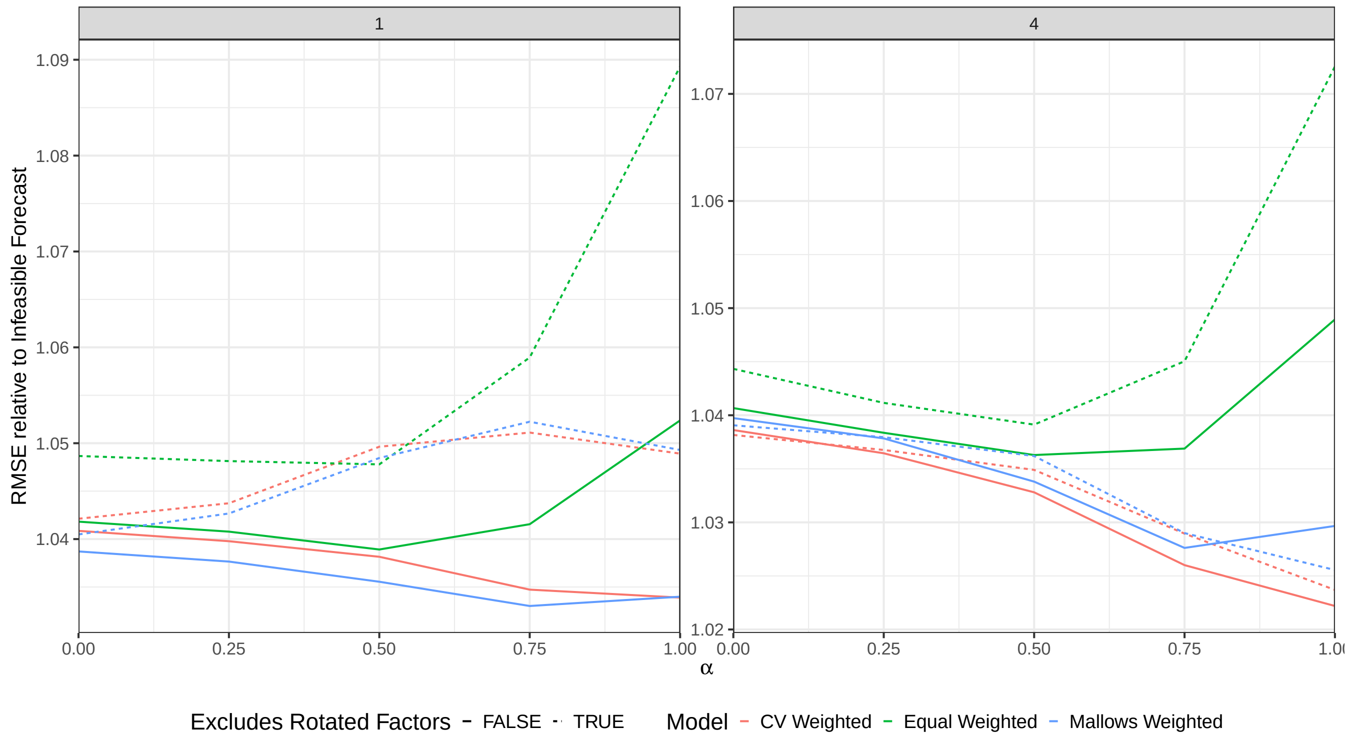

Figure 2: Relative MSFE, faceted by h, $\kappa = 0.1$ for mild serial correlation, $q_{max} = p_{max} = 4$. Solid line represents model averaged forecasts where the model set includes the rotated factors $\tilde{F}_R$, dashed line represents model averaged forecasts where the model set excludes $\tilde{F}_R$.

We consider the FRED-MD database of McCracken and Ng (2016), which consists of 125[7] U.S. macroeconomic time series, 93 of which are non-top-level aggregates used to estimate the factors.[8] As our method only deals with one possible structural break, we focus on the subsample(s) of 1984 March - 2020 February and 2008 December - 2024 September, which are dictated by evidence of breaks occurring in 1984 February, 2008 November, and 2020 March associated with the Great Moderation (Stock and Watson, 2009; Breitung and Eickmeier, 2011), Global Financial Crisis (Cheng et al., 2016), and COVID19 Pandemic (Ng, 2021), respectively.

5.2 Methodology

As in Stock and Watson (2012), all forecasting models contain a fixed set of 4 lagged dependent variables, and the models differ by the number of included factors, as well as the method of estimating the factors. Due to the need for subsample principal components estimation and the potentially short panel in the second subsample, the number of factors included in each model ranges from $r = 0$ to 10, rather than $r = 0$

[7]This number is obtained by following McCracken and Ng (2016), and suitably removing series which cannot be included after the suggested transformation(s).

[8]We follow the classification provided by Han (2024), corresponding to include = 99.

to $r = 50$ as in Stock and Watson (2012).

Given this set of models, we construct forecasts using both selection and model averaging approaches. The averaging methods include the proposed post-break leave-h-out cross-validation, Mallows model averaging following Cheng and Hansen (2015), Bayesian model averaging, and equal weights. The selection methods include the proposed post-break leave-h-out cross-validation and Mallows model selection similarly following Cheng and Hansen (2015). The out-of-sample MSFE is calculated through an expanding window exercise. Due to the short span of the post-break data, we withhold only the final 12 observations (corresponding to one years) as the test set, and we report the relative root mean squared error relative to the dynamic factor model with 5 factors (DFM-5). This model is chosen because Stock and Watson (2012) demonstrated that DFM-5 improves upon AR(4) in more than 75% of series, while the shrinkage methods offering little to no improvement, and hence serves as a good benchmark.

We follow the methodology of Stock and Watson (2002b) in estimating the forecasting equation. Specifically, we do not purge the effects of the lagged regressors on the x_{it} series in a preliminary step as suggested by Stock and Watson (2012). Our experience shows that this omission results in forecasts of very similar quality, and is unnecessary.

5.3 Results

Tables 3 and 4 reports the percentiles of the distributions of the one-, two- and three- month ahead expanding window out-of-sample RMSEs (relative to the DFM-5 benchmark) over the 125 series for the proposed forecasting methods, for the Global Financial Crisis and COVID19 subsamples. These results are remarkable, given that Indeed, Stock and Watson (2012) report that many advanced shrinkage methods fail to improve upon the DFM-5 benchmark, while Cheng and Hansen (2015) show that frequentist averaging over the number of factors yields only modest gains for a minority of series. Thus, it is particularly noteworthy that our proposed methods, which account for potential breaks in the factor structure provide further - albeit modest - gains.

For both subsamples, we find that the proposed post-break cross-validation weighted forecasts generally exhibit the least deterioration and show potential for substantial improvement over the benchmark for at least half the series. Of all the methods, it is often within the top three methods by ranking, and still remains competitive when it is not. Of particular note is how the cross-validated forecasts remain competitive even for COVID19 subsample, a scenario with an extreme low number of post-break observations, and where data-adaptive methods using the full sample such as the Mallows criterion tends to dominate.

Examining the performance of each factor estimator helps reveals the source of these gains. Generally, of the three factor estimators, the rotated factors performs the best, followed closely by the pseudo factors, and occasionally the split-sample factors. Indeed, among the three factor estimators, using the rotated factors alone can yield very competitive forecasting performance - this effect is particularly evident on the COVID19 subsample. In contrast, the split-sample factors typically offer the worst performance, representing a significant bias-variance trade-off, though can occasionally offer competitive forecasts for a minority of series. This demonstrates that the gains from the model averaged forecasts generally come from including the rotated factors; the combination of these two strategies is then able to generally dominate the benchmark for at least half of the series, and highlights the importance of modelling structural breaks. Similar to Stock and Watson (2012) and Cheng and Hansen (2015), we find that most methods fail to improve upon the DFM-5 benchmark for at least three quarters of the series in the dataset.

Tables 5 and 6 break down the results of Tables 3 and 4 by category at the median RMSE relative to the DFM-5 benchmark. Generally, we find that the rotated factors can offer better forecasting performance for variables in the Output and Income, Labor Market, Money and Credit, Interest Rates, and on occasion in the Prices and Stock Market categories. These patterns are not entirely consistent across all samples and forecasting horizons, with the split-sample strategy sometimes performing well in categories that were particularly chaotic, such as Housing for the Global Financial Crisis subsample, and the pseudo factors sometimes providing better performance in the COVID19 subsample, where the post-break sample is very short. This reinforces the need for a data-adaptive method to automatically select or weight forecasts. Unsurprisingly, the use of weighted forecasts, in particular the post-break cross-validation weighted forecasts, can result in more reliable forecasting performance, compared to other weighting schemes such as equal weights and Mallows weights. This, together with evidence that they can improve performance for a minority of series as detailed in Tables 3 and 4 further exemplifies the need for data-adaptive weighting procedures.

Table 3: Distributions of relative RMSEs by forecasting method, relative to DFM-5, h = 1, 2, 3, FREDMD Global Financial Crisis Subsample (1984 March - 2020 February, 2008 November Break), outlier adjusted, include = 99.

Percentile	h = 1			h = 2			h = 3		
Model	0.250	0.500	0.750	0.250	0.500	0.750	0.250	0.500	0.750

CV Select	0.992	1.002	1.014	0.985***	0.999	1.009	0.987	0.999	1.007
CV Weighted	0.982*	0.996**	1.006	0.984**	0.996**	1.005***	0.982*	0.996***	1.005
Equal Weighted	0.983**	0.995*	1.004**	0.983*	0.995*	1.003*	0.983**	0.994**	1.002*
Mallows Select	0.986***	1.003	1.018	0.992	1.001	1.015	0.986	0.997	1.009
Mallows Weighted	0.988	0.998	1.005***	0.988	0.998	1.005***	0.984***	0.997	1.005
Pseudo r	0.995	1.000	1.003*	0.996	1.000	1.003*	0.994	1.000	1.002*
Rotated	0.986***	0.996**	1.005***	0.985***	0.996**	1.009	0.985	0.993*	1.003***
Split Break	0.994	1.010	1.032	0.989	1.005	1.027	0.991	1.008	1.028

Note:

Entries are percentiles of distributions of relative RMSEs over the 125 variables being forecast, by series, at the specified forecast horizon. RMSEs are relative to the DFM-5 forecast and calculated as an expanding pseudo out of sample exercise. All forecasts are direct. Cross validation implemented using post break residuals. No. of asterisks denote ranking. Pseudo r factors is obtained by averaging over the number of factors using post-break CV, and hence similar to Cheng and Hansen (2015)'s approach, rotated and split-sample factors are also similarly averaged.

Table 4: Distributions of relative RMSEs by forecasting method, relative to DFM-5, h = 1, 2, 3, FREDMD COVID-19 Subsample (2008 December - 2024 September, 2020 March Break), outlier adjusted, include = 99.

Percentile	h = 1			h = 2			h = 3		
Model	0.250	0.500	0.750	0.250	0.500	0.750	0.250	0.500	0.750
CV Select	0.929	0.984***	1.019	0.948	0.997	1.013	0.962	1.000	1.025
CV Weighted	0.912*	0.970**	1.011**	0.946	0.988	1.008**	0.935	0.990	1.010***
Equal Weighted	0.924***	0.985	1.022	0.937***	0.976*	1.012	0.933***	0.982**	1.011
Mallows Select	0.953	0.985	1.024	0.917*	0.982***	1.011***	0.930**	0.979*	1.001*
Mallows Weighted	0.955	0.993	1.023	0.938	0.981**	1.007*	0.963	0.994	1.019
Pseudo r	0.974	0.997	1.004*	0.981	1.000	1.011***	0.962	0.994	1.003**
Rotated	0.920**	0.969*	1.015***	0.924**	0.983	1.021	0.923*	0.984***	1.016
Split Break	0.989	1.049	1.116	0.965	1.009	1.080	0.990	1.047	1.115

Note:

Entries are percentiles of distributions of relative RMSEs over the 125 variables being forecast, by series, at the specified forecast horizon. RMSEs are relative to the DFM-5 forecast and calculated as an expanding pseudo out of sample exercise. All forecasts are direct. Cross validation implemented using post break residuals. No. of asterisks denote ranking.

Table 5: Median RMSE by forecasting method and category of series, relative to DFM-5, expanding window forecast estimates, FREDMD Global Financial Crisis Subsample (1984 March - 2020 February, 2008 November Break), outlier adjusted, include = 99.

Group	CV Select	CV Weighted	Equal Weighted	Mallows Select	Mallows Weighted	Pseudo r	Rotated	Split Break
h = 1								
Output and Income	0.998***	1.001	0.993*	1.003	0.996**	1.002	0.998***	1.004
Labor Market	1.000***	1.002	1.000***	1.008	0.999**	1.000***	0.993*	1.010
Housing	1.000	0.999***	0.997**	1.016	0.999***	1.000	1.004	0.996*
Consumption, Orders, and Inventories	1.005	1.001	0.995***	0.998	0.992*	1.001	0.992*	1.024
Money and Credit	1.003	1.002	0.998	0.986**	1.001	0.992***	0.985*	1.032
Interest and Exchange Rates	1.009	0.990*	0.990*	0.995***	0.997	1.001	0.996	1.012
Prices	1.008	0.991**	0.989*	1.003	0.997	1.000	0.994***	1.010
Stock Market	1.008	1.015	1.015	1.005**	1.038	1.005**	1.002*	1.082
h = 2								
Output and Income	0.988**	0.987*	0.988**	0.998	0.996	1.000	0.990	0.992
Labor Market	1.000**	1.001	0.999*	1.009	1.001	1.000**	1.008	1.008
Housing	1.004	0.998	0.983**	1.008	0.981*	1.001	1.008	0.985***
Consumption, Orders, and Inventories	0.987*	0.991**	1.002	0.996***	0.999	1.000	0.996***	1.013
Money and Credit	0.998	0.998	0.996**	0.996**	0.996**	0.998	0.995*	1.010
Interest and Exchange Rates	1.007	0.998**	0.996*	1.004	1.001	1.000***	1.000***	1.040
Prices	0.990*	0.991	0.990*	1.002	0.994	1.000	0.990*	0.996
Stock Market	0.993	0.993	0.992**	0.992**	0.996	0.996	0.990*	1.029
h = 3								
Output and Income	1.000	0.983*	0.988***	1.005	0.983*	1.000	0.989	0.997
Labor Market	1.001	0.995**	0.996***	0.997	0.998	1.000	0.992*	1.008
Housing	0.990	0.977*	0.982***	0.998	0.986	1.002	0.995	0.979**

Consumption, Orders, and Inventories	0.987**	0.993	0.989	0.978*	0.989	0.999	0.988***	1.009
Money and Credit	0.997	0.996***	0.994*	0.996***	0.995**	0.997	0.997	1.002
Interest and Exchange Rates	1.006***	1.008	1.004**	1.013	1.006***	1.000*	1.011	1.024
Prices	0.994*	0.998	0.995**	0.996***	1.001	0.998	0.996***	1.010
Stock Market	1.000	0.996***	0.997	0.976**	1.005	0.997	0.972*	1.045

Note:

Entries are median RMSEs, relative to DFM-5, for the row category of variables. No. of asterisks denote ranking.

Table 6: Median RMSE by forecasting method and category of series, relative to DFM-5, expanding window forecast estimates, FREDMD COVID-19 Subsample (2008 December - 2024 September, 2020 March Break), outlier adjusted, include = 99.

Group	CV Select	CV Weighted	Equal Weighted	Mallows Select	Mallows Weighted	Pseudo r	Rotated	Split Break
h = 1								
Output and Income	0.968**	0.903*	0.974	0.987	1.000	1.001	0.970***	1.123
Labor Market	0.905**	0.910***	0.940	0.982	0.967	0.993	0.895*	1.041
Housing	0.977***	0.974**	0.988	0.969*	0.990	0.984	1.007	1.053
Consumption, Orders, and Inventories	1.024	0.988**	1.005	0.979*	1.002	0.988**	1.005	1.104
Money and Credit	0.971*	0.985***	1.000	0.988	0.990	0.988	0.974**	1.088
Interest and Exchange Rates	1.000	0.988	0.973**	0.980***	0.981	1.000	0.967*	1.011
Prices	1.002**	1.009	1.021	1.014	1.008***	0.994*	1.027	1.068
Stock Market	0.981	0.970***	0.980	0.965**	0.971	0.994	0.960*	0.972
h = 2								
Output and Income	0.993	0.991***	0.987**	1.023	0.985*	1.011	1.012	0.991***
Labor Market	0.987	0.950***	0.947**	0.954	0.956	0.980	0.941*	1.007
Housing	0.998	0.986	0.963**	0.928*	0.972***	1.000	0.988	1.054
Consumption, Orders, and Inventories	0.994*	1.002	1.009	1.005	0.994*	1.001***	1.018	1.041
Money and Credit	0.926	0.897**	0.959	0.901***	0.926	1.000	0.895*	1.000
Interest and Exchange Rates	1.000	0.999	0.975*	0.983**	0.992	1.000	0.988***	1.019
Prices	0.998	0.996***	0.993**	1.000	1.004	0.996***	0.989*	1.005
Stock Market	0.968	0.963	0.945	0.892*	0.956	0.957	0.898**	0.935***
h = 3								
Output and Income	0.972**	0.975	0.963*	0.991	0.979	0.972**	0.977	1.004
Labor Market	1.005	0.929*	0.930**	0.968***	0.977	0.976	0.995	1.053
Housing	1.000	0.998	0.985***	0.912*	0.999	0.986	0.936**	1.120

Consumption, Orders, and Inventories	0.996	0.998	0.994	0.984**	0.993***	0.995	0.977*	1.106
Money and Credit	0.981	0.965**	0.966***	0.985	0.978	0.985	0.923*	1.003
Interest and Exchange Rates	1.000	0.998	0.987***	0.970**	0.993	1.000	0.967*	1.038
Prices	1.029	1.018	1.021	0.992*	1.036	1.003**	1.016***	1.116
Stock Market	0.933**	0.932*	0.968	0.973	1.005	1.003	0.953***	1.052

Note:

Entries are median RMSEs, relative to DFM-5, for the row category of variables. No. of asterisks denote ranking.

5.4 Robustness Check using Stock and Watson (2012) Data

As a robustness check, we also conduct a forecasting exercise for one-, two-, and four- quarter ahead forecasts using the quarterly dataset of Stock and Watson (2012), which is also used by Cheng and Hansen (2015). This dataset consists of 143 macroeconomic series, of which only 108[9] disaggregated series are used to estimate the factors. Due to data availability, we focus on a structural break in 1984 Q1, corresponding to the Great Moderation, which is documented as a break by Stock and Watson (2009); **?**); Breitung and Eickmeier (2011); Baltagi et al. (2021), among others. Our results are available in Appendix D.1, and broadly similar to the results using FRED-MD.

6 Conclusion

This paper proposes and derives the theoretical properties of three different factor estimates in the presence of structural breaks in the factor structure: the whole sample principal components, split-sample factors, and our novel set of rotated factors, which are the subsample factors normalised onto the same basis. We show that these factor estimates are respectively robust to small breaks, all large breaks at the cost of more parameters, and large shift type types. In practice, it is difficult to know or estimate the sizes of each type of break, and to this end we propose and prove the validity of the use of post-break leave-h-out cross-validation selection and weighting for data driven selection and weighting. Monte Carlo simulations support the theoretical results. An application with U.S. macroeconomic data demonstrates the potential gains from leveraging knowledge of structural break in the dataset and highlights the poor performance of traditional approaches, which directly allows for breaks in the forecasting equation. The theoretical results proposed in the paper have notable applications beyond factor-augmented forecasting. One possible extension is to generalise the single variable forecast in the factor-augmented vector autoregression (FAVAR) model of Bernanke et al. (2005). Indeed, it is well documented that the dynamic impulse responses of monetary policy have changed since before and after the Great Moderation (see Boivin and Giannoni, 2006), and the tools developed in this paper may be helpful in investigating further.

[9] Stock and Watson (2012) mistakenly say there are 109 series.

A Factor Model Proofs

A.1 Preliminary

We first state some preliminary results used throughout the proofs.

$$\frac{F^\top e}{\sqrt{NT}} = O_p(1) \tag{A.1}$$
$$\frac{\Lambda_1^\top e}{\sqrt{NT}} = O_p(1) \tag{A.2}$$
$$\frac{W^\top e}{\sqrt{N^\alpha T}} = O_p(1) \tag{A.3}$$
$$\frac{ee^\top}{NT} = O_p\left(\frac{1}{\delta_{NT}}\right) \tag{A.4}$$
$$\frac{F^\top e\Lambda}{NT} = O_p\left(\frac{1}{\delta_{NT}^2}\right) \tag{A.5}$$
$$\frac{F^\top eW}{N^\alpha T} = O_p\left(\frac{1}{\delta_{NT}^2}\right) \tag{A.6}$$

which are implied by Assumption 6 (a), Assumption 6 (d), Assumption 6 (e), Assumption 3 (e), Assumption 6 (b), and Assumption 6 (c), respectively.

A.2 Pseudo-factors $\tilde{F}_P$

To begin, we make the following expansion

$$\begin{aligned}\tilde{F}_P V_{NT,r} &= \frac{1}{TN} XX^\top \tilde{F}_P \\ \tilde{F}_P &= \frac{1}{TN}\left(G_r\Lambda_1^\top + G_pW^\top + e\right)\left(G_r\Lambda_1^\top + G_pW^\top + e\right)^\top V_{NT,r}^{-1}.\end{aligned} \tag{A.7}$$

Proof of Theorem 1 (a). Expanding out Equation (A.7), we have

$$\begin{aligned}\tilde{F}_P = \frac{1}{TN}\Big(&G_r\Lambda_1^\top\Lambda_1G_r^\top\tilde{F}_P + G_r\Lambda_1^\top e^\top\tilde{F}_P + e\Lambda_1^\top G_r^\top\tilde{F}_P + ee^\top\tilde{F}_P \\ &+eWG_p^\top\tilde{F}_P + G_pW^\top WG_p^\top\tilde{F}_P + G_r\Lambda_1^\top WG_p^\top\tilde{F}_P + G_pW^\top\Lambda_1G_r^\top\tilde{F}_P\Big)V_{NT,r}^{-1}.\end{aligned} \tag{A.8}$$

Substituting in H_G and rearranging yields

$$\begin{aligned}\tilde{F}_P - G_rH_G = \frac{1}{TN}\Big(&G_r\Lambda_1^\top e^\top\tilde{F}_P + e\Lambda_1G_r^\top\tilde{F}_P + ee^\top\tilde{F}_P + eWG_p^\top\tilde{F}_P \\ &+G_pW^\top e^\top\tilde{F}_P + G_pW^\top WG_p^\top\tilde{F}_P + G_r\Lambda_1^\top WG_p^\top\tilde{F}_P + G_pW^\top\Lambda_1G_r^\top\tilde{F}_P\Big)V_{NT,r}^{-1}.\end{aligned} \tag{A.9}$$

Next, multiply both sides by $\frac{1}{\sqrt{T}}$ to get

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\left(\tilde{F}_P - G_r H_G\right)\\
=&\frac{1}{\sqrt{T}}\frac{1}{TN}\Big(G_r\Lambda_1^\top e^\top \tilde{F}_P + e\Lambda_1 G_r^\top \tilde{F}_P + ee^\top \tilde{F}_P + eWG_p^\top \tilde{F}_P\\
&\qquad +G_pW^\top e^\top \tilde{F}_P + G_pW^\top W G_p^\top \tilde{F}_P + G_r\Lambda_1^\top W G_p^\top \tilde{F}_P + G_pW^\top \Lambda_1 G_r^\top \tilde{F}_P\Big) V_{NT,r}^{-1}\\
=&\left(a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8\right) V_{NT,r}^{-1}.
\end{aligned}
$$

Noting that $V_{NT,r}^{-1} = O_p(1)$, we have by, analysing the asymptotic order of the terms of the RHS

$$
\begin{aligned}
a_1 &= \frac{G_r}{\sqrt{T}}\frac{\Lambda_1^\top e^\top}{\sqrt{TN}}\frac{\tilde{F}_P}{\sqrt{T}}\frac{1}{\sqrt{N}} = O_p\left(\frac{1}{\sqrt{N}}\right),\\
a_2 &= \frac{e\Lambda_1}{\sqrt{TN}}\frac{G_r^\top \tilde{F}_P}{T}\frac{1}{\sqrt{N}} = O_p\left(\frac{1}{\sqrt{N}}\right),\\
a_3 &= \frac{ee^\top}{NT}\frac{\tilde{F}_P}{\sqrt{T}} = O_p\left(\frac{1}{\delta_{NT}}\right),\\
a_4 &= \frac{eW}{\sqrt{N^\alpha T}}\frac{G_p^\top \tilde{F}_P}{T}\frac{\sqrt{N^\alpha}}{N} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right),\\
a_5 &= \frac{G_p}{\sqrt{T}}\frac{W^\top e^\top}{\sqrt{N^\alpha T}}\frac{\tilde{F}_P}{\sqrt{T}}\frac{\sqrt{N^\alpha}}{N} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right),\\
a_6 &= \frac{G_p}{\sqrt{T}}\frac{W^\top W}{N^\alpha}\frac{N^\alpha}{N}\frac{G_p^\top \tilde{F}_P}{T} = O_p\left(\frac{N^\alpha}{N}\right),\\
a_7 &= \frac{G_r}{\sqrt{T}}\frac{\Lambda_1^\top W}{N}\frac{G_p^\top \tilde{F}_P}{T} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \quad \text{and}\\
a_8 &= \frac{G_p}{\sqrt{T}}\frac{W^\top \Lambda_1}{N}\frac{Gp^\top \tilde{F}_P}{T} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}
$$

Note that the terms a_7 and a_8 are not zero due to W and Λ_1 not being exactly orthogonal, but are still asymptotically negligible. Thus, term a_6 characterises the dominating bias term. Collecting the dominating terms yields

$$
\frac{1}{\sqrt{T}}\left(\tilde{F}_P - G_r H_G\right) = O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{N^\alpha}{N}\right).
$$

Squaring both sides yields the main result.

This mean square consistency result can be used to derive a sharper bound for some of the terms in

$\frac{1}{\sqrt{T}}\left(\tilde{F}_P - G_r H_G\right)$. Specifically,

$$\begin{aligned}
a_1 &= \frac{G_r \Lambda_1^\top e^\top}{\sqrt{TN}} \frac{\tilde{F}_P}{\sqrt{T}} \frac{1}{\sqrt{TN}} \\
&= \frac{G_r}{\sqrt{T}} \frac{\Lambda_1^\top e^\top G_r H_G}{\sqrt{TN}} \frac{1}{\sqrt{TN}} + \frac{G_r}{\sqrt{T}} \frac{\Lambda_1^\top e^\top (\tilde{F}_P - G_r H_G)}{\sqrt{TN}} \frac{1}{\sqrt{TN}} \\
&= O_p\left(\frac{1}{\sqrt{TN}}\right) + O_p\left(\frac{1}{\sqrt{N}\delta_{NT}} + \frac{N^\alpha}{N\sqrt{N}}\right), \\
a_3 &= \frac{ee^\top}{NT} \frac{\tilde{F}_P}{\sqrt{T}} = O_p\left(\frac{1}{\sqrt{TN}}\right) + O_p\left(\frac{1}{\sqrt{N}\delta_{NT}} + \frac{N^\alpha}{N\sqrt{N}}\right),
\end{aligned}$$

where the detailed derivation for a_3 follows by

$$\begin{aligned}
\frac{1}{NT\sqrt{T}}\left\|ee^\top \tilde{F}_P\right\| &= \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{TN}\sum_{t=1}^{T} e_s^\top e_t \tilde{F}_{P,t}\right\|^2\right)^{1/2}, \\
\frac{1}{TN}\sum_{t=1}^{T} e_s^\top e_t \tilde{F}_{P,t} &= \frac{1}{TN}\sum_{t=1}^{T}\left[e_s^\top e_t - E(e_s^\top e_t)\right]\tilde{F}_{P,t}^\top + \frac{1}{TN}\sum_{t=1}^{T} E(e_s^\top e_t)\tilde{F}_{P,t}^\top, \\
\frac{1}{TN}\sum_{t=1}^{T}\left[e_s^\top e_t - E(e_s^\top e_t)\tilde{F}_{P,t}\right] &= \frac{1}{TN}\sum_{t=1}^{T}\left[e_s^\top e_t - E(e_s^\top e_t)\right] G_{r,t}^\top H_G \\
&\quad + \frac{1}{TN}\sum_{t=1}^{T}\left[e_s^\top e_t - E(e_s^\top e_t)\right]\left(\tilde{F}_{P,t}^\top - G_{r,t}^\top H_G\right) \\
&= O_p\left(\frac{1}{\sqrt{TN}}\right) + O_p\left(\frac{1}{\sqrt{N}\delta_{NT}} + \frac{N^\alpha}{N\sqrt{N}}\right), \quad \text{and} \qquad \text{(A.10)} \\
\frac{1}{TN}\sum_{t=1}^{T} E\left[e_s^\top e_t\right]\tilde{F}_{P,t}^\top &= \frac{1}{T}\sum_{t=1}^{T} E(e_s^\top e_t/N) G_{r,t}^\top H_G + \frac{1}{T}\sum_{t=1}^{T} E(e_s^\top e_t/N)\left(\tilde{F}_{P,t}^\top - G_{r,t} H_G\right) \\
&= O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{T}\delta_{NT}} + \frac{N^\alpha}{\sqrt{T}N}\right). \qquad \text{(A.11)}
\end{aligned}$$

The remaining terms a_4, a_5, a_6, a_7, and a_8 all contain W, and therefore cannot be sharpened. ■

Proof of Theorem 1 (a).

$$\begin{aligned}
\frac{1}{\sqrt{T}}\left(\tilde{F}_P - GH_{\Xi,r}\right) &= \frac{1}{TN\sqrt{T}}\left(G_r\Lambda_1^\top e^\top \tilde{F}_P + e\Lambda_1^\top G_r^\top \tilde{F}_P + ee^\top \tilde{F}_P + eWG_p^\top \tilde{F}_P\right) V_{NT,r}^{-1} \\
&= (a_9 + a_{10} + a_{11} + a_{12}) V_{NT,r}^{-1}. \\
a_9 &= \frac{G_r}{\sqrt{T}} \frac{\Lambda_1^\top e^\top}{\sqrt{NT}} \frac{\tilde{F}_P}{\sqrt{T}} = O_p\left(\frac{1}{\sqrt{N}}\right), \\
a_{10} &= \frac{e\Lambda}{\sqrt{TN}} \frac{G_r^\top \tilde{F}_P}{T} = O_p\left(\frac{1}{\sqrt{N}}\right), \\
a_{11} &= \frac{ee^\top}{\sqrt{TN}} \frac{\tilde{F}_P}{\sqrt{T}} = O_p\left(\frac{1}{\delta_{NT}}\right),
\end{aligned}$$

$$a_{12} = \frac{eW}{\sqrt{TN}} \frac{G_r^\top \tilde{F}_P}{T} = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Collecting the dominating terms and squaring both sides of the equation proves the result. ■

Theorem 1 (a) can then be used to prove the following lemmas for the pseudo-factors $\tilde{F}_P$.

Lemma 1. *Under Assumptions 1 to 8 and as $N, T \to \infty$ and $\alpha < 1$,*

a) $\frac{1}{T}\left(\tilde{F}_P - G_r H_G\right)^\top G_r = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right)$, *if $\alpha < 1$,*

b) $\frac{1}{T}\left(\tilde{F}_P - G H_{\Xi,r}\right)^\top G_r = O_p\left(\frac{1}{\delta_{NT}^2}\right)$, *if $\alpha = 1$,*

c) $\frac{1}{T}\left(\tilde{F}_P - G_r H_G\right)^\top e_i = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N\sqrt{T}}\right)$, *if $\alpha < 1$,*

d) $\frac{1}{T}\left(\tilde{F}_P - G H_{\Xi,r}\right)^\top e_i = O_p\left(\frac{1}{\delta_{NT}^2}\right)$, *if $\alpha = 1$.*

Proof of Lemma 1 (a).

$$\begin{aligned}
\frac{1}{T}\left(\tilde{F}_P - G_r H_G\right)^\top G_r = \frac{1}{T^2 N} V_{NT,r}^{-1} \Big(& \tilde{F}_P^\top G_r \Lambda_1^\top W G_p^\top G_r + \tilde{F}_P^\top G_p W^\top \Lambda_1 G_r^\top G_r \\
&+ \tilde{F}_P^\top G_p W^\top W G_p^\top G_r + \tilde{F}_P^\top e W G_p^\top G_r \\
&+ \tilde{F}_P^\top G_p W^\top e^\top G_r + \tilde{F}_P^\top e e^\top G_r + \tilde{F}_P^\top G_r \Lambda_1^\top e^\top G_r + \tilde{F}_P^\top e \Lambda_1 G_r^\top G_r\Big) \\
= V_{NT,r}^{-1} &\left(a_{13} + a_{14} + a_{15} + a_{16} + a_{17} + a_{18} + a_{19} + a_{20}\right).
\end{aligned}$$

Analysing each term, we have

$$\begin{aligned}
a_{13} &= \frac{\tilde{F}_P^\top G_r}{T} \frac{\Lambda_1^\top W}{N} \frac{G_p^\top G_r}{T} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \\
a_{14} &= \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top \Lambda}{N} \frac{G_r^\top G_r}{T} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \\
a_{15} &= \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N^\alpha} \frac{N^\alpha}{N} \frac{G_p^\top G_r}{T} = O_p\left(\frac{N^\alpha}{N}\right), \\
a_{16} &= \frac{\tilde{F}_P}{\sqrt{T}} \frac{eW}{\sqrt{N^\alpha T}} \frac{G_p^\top G_r}{T} \frac{\sqrt{N^\alpha}}{N} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \\
a_{17} &= \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top e^\top G_r}{N^\alpha T} \frac{N^\alpha}{N} = \frac{N^\alpha}{N} O_p\left(\frac{1}{\delta_{NT}^2}\right), \\
a_{18} &= \frac{\left(\tilde{F}_P - G_r H_G\right)^\top}{\sqrt{T}} \frac{ee^\top}{NT} \frac{G_r}{\sqrt{T}} + \frac{H_G^\top G_r^\top e e^\top G_r}{TNT} \\
&= \left(O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{N^\alpha}{N}\right)\right) O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{1}{T}\right)
\end{aligned}$$

$$
\begin{aligned}
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + \frac{N^\alpha}{N} O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{1}{T}\right), \\
a_{19} &= \frac{\tilde{F}_P^\top G_r}{T} \frac{\Lambda_1^\top e^\top G_r}{TN} = O_p\left(\frac{1}{\delta_{NT}^2}\right), \quad \text{and} \\
a_{20} &= \frac{\left(\tilde{F}_P - G_r H_G\right)^\top}{\sqrt{T}} \frac{e\Lambda}{\sqrt{TN}} \frac{G_r^\top G_r}{T} \frac{1}{\sqrt{N}} + \frac{H_G^\top G_r^\top e\Lambda}{TN} \frac{G_r^\top G_r}{T} \\
&= \left(O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{N^\alpha}{N}\right)\right) \frac{1}{\sqrt{N}} + O_p\left(\frac{1}{\delta_{NT}^2}\right) \\
&= O_p\left(\frac{1}{\sqrt{N}\delta_{NT}}\right) + O_p\left(\frac{N^\alpha}{N^{3/2}}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}
$$

Collecting the dominating terms proves the lemmas. ■

Proof of Lemma 1 (b).

$$
\begin{aligned}
\frac{1}{T}\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top G_r &= V_{NT,r}^{-1} \frac{1}{T^2 N}\left(\tilde{F}_P^\top G_p W^\top e^\top G_r + \tilde{F}_P^\top e e^\top G_r + \tilde{F}_P^\top G_r \Lambda_1^\top e^\top G_r + \tilde{F}_P^\top e \Lambda_1 G_r^\top G_r\right) \\
&= V_{NT,r}^{-1}\left(a_{21} + a_{22} + a_{23} + a_{24}\right), \\
a_{21} &= \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top e^\top G_r}{NT} = O_p\left(\frac{1}{\delta_{NT}^2}\right), \\
a_{22} &= \frac{\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top}{\sqrt{T}} \frac{ee^\top}{NT} \frac{G_r}{\sqrt{T}} + \frac{\left(GH_{\Xi,r}\right)^\top}{\sqrt{T}} \frac{ee^\top}{NT} \frac{G_r}{\sqrt{T}} \\
&= O_p\left(\frac{1}{\delta_{NT}}\right) O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{1}{T}\right), \\
a_{23} &= \frac{\tilde{F}_P^\top G_r}{T} \frac{\Lambda_1^\top e^\top G_r}{NT} = O_p\left(\frac{1}{\delta_{NT}^2}\right), \quad \text{and} \\
a_{24} &= \frac{\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top}{\sqrt{T}} \frac{e\Lambda}{\sqrt{TN}} \frac{G_r^\top G_r}{T} + \frac{H_{\Xi,r}^\top G_r^\top e\Lambda}{NT} \frac{G_r^\top G_r}{T} \\
&= O_p\left(\frac{1}{\delta_{NT}}\right) \frac{1}{\sqrt{N}} + O_p\left(\frac{1}{\delta_{NT}^2}\right). \\
\therefore \frac{1}{T}\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top G_r &= O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}
$$

■

Proof of Lemma 1 (c).

$$
\begin{aligned}
\frac{1}{T}\left(\tilde{F}_P - G_r H_G\right)^\top e_i = \frac{1}{T^2 N} V_{NT,r}^{-1} \Big(&\tilde{F}_P^\top G_r \Lambda_1^\top W G_p^\top e_i + \tilde{F}_P^\top G_p W^\top \Lambda_1 G_r^\top e_i \\
&+ \tilde{F}_P^\top G_p W^\top W G_p^\top e_i + \tilde{F}_P^\top e W G_p^\top e_i
\end{aligned}
$$

$$+\tilde{F}_P^\top G_p W^\top e^\top e_i + \tilde{F}_P^\top e e^\top e_i + \tilde{F}_P^\top G_r \Lambda_1^\top e^\top e_i + \tilde{F}_P^\top e \Lambda_1 G_r^\top e_i\Big)$$

$$= V_{NT,r}^{-1}\left(a_{25} + a_{26} + a_{27} + a_{28} + a_{29} + a_{30} + a_{31} + a_{32}\right)$$

These terms have the following asymptotic order:

$$\begin{aligned}
a_{25} &= \frac{\tilde{F}_P^\top G_r}{T}\frac{\Lambda_1^\top W}{N}\frac{G_p^\top e_i}{\sqrt{T}}\frac{1}{\sqrt{T}} = O_p\left(\frac{\sqrt{N^\alpha}}{N\sqrt{T}}\right),\\
a_{26} &= \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top \Lambda}{N}\frac{G_r^\top e_i}{\sqrt{T}}\frac{1}{\sqrt{T}} = O_p\left(\frac{\sqrt{N^\alpha}}{N\sqrt{T}}\right),\\
a_{27} &= \frac{\tilde{F}_P^\top G_r}{T}\frac{W^\top W}{N^\alpha}\frac{G_p^\top e_i}{\sqrt{T}}\frac{1}{\sqrt{T}}\frac{N^\alpha}{N} = O_p\left(\frac{N^\alpha}{N}\right)\frac{1}{\sqrt{T}},\\
a_{28} &= \frac{\tilde{F}_P^\top}{\sqrt{T}}\frac{eW}{\sqrt{N^\alpha T}}\frac{G_p^\top e_i}{\sqrt{T}}\frac{\sqrt{N^\alpha}}{N\sqrt{T}} = O_p\left(\frac{\sqrt{N^\alpha}}{N\sqrt{T}}\right),\\
a_{29} &= \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top e^\top e_i}{TN^\alpha}\frac{N^\alpha}{N}\\
&= \frac{N^\alpha}{N}\left(O_p\left(\frac{1}{N^\alpha}\right) + O_p\left(\frac{1}{\sqrt{TN^\alpha}}\right)\right),\\
&= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N\sqrt{T}}\right),\\
a_{30} &= \frac{\tilde{F}_P^\top}{\sqrt{T}}\frac{ee^\top}{TN}\frac{e_i}{\sqrt{T}}\\
&= O_p\left(\frac{1}{\sqrt{TN}}\right) + O_p\left(\frac{1}{\sqrt{N}\delta_{NT}} + \frac{N^\alpha}{N\sqrt{N}}\right)
\end{aligned}$$

$$\begin{aligned}
a_{31} &= \frac{\tilde{F}_P^\top G_r}{T}\frac{\Lambda_1^\top e^\top e_i}{TN} = O_p\left(\frac{1}{\delta_{NT}^2}\right), \quad \text{and}\\
a_{32} &= \frac{\left(\tilde{F}_P - G_r H_G\right)^\top}{\sqrt{T}}\frac{e\Lambda}{\sqrt{TN}}\frac{G_r^\top e_i}{\sqrt{T}}\frac{1}{\sqrt{TN}} + \frac{H_G^\top G_r e\Lambda}{TN}\frac{G_r^\top e_i}{\sqrt{T}}\frac{1}{\sqrt{T}}\\
&= \left(O_p\left(\frac{N^\alpha}{N}\right) + O_p\left(\frac{1}{\delta_{NT}}\right)\right)\frac{1}{\sqrt{TN}} + O_p\left(\frac{1}{\delta_{NT}^2}\right)\frac{1}{\sqrt{T}}\\
&= O_p\left(\frac{N^\alpha}{N\sqrt{TN}}\right) + O_p\left(\frac{1}{\delta_{NT}\sqrt{TN}}\right) + O_p\left(\frac{1}{\delta_{NT}^2\sqrt{T}}\right).
\end{aligned}$$

The first part of the theorem follows by collecting the dominating terms.

The second part of the theorem follows by adding and subtracting FH_G

$$\tilde{F}_P - G_r H_G = \tilde{F}_P - FH_G + (F - G_r)H_G$$

$$
\begin{aligned}
&= \tilde{F}_P - FH_G + \begin{bmatrix} 0 \\ F_2(I_r - Z^\top) \end{bmatrix} H_G \\
&= \tilde{F}_P - FH_G + G_p(I_r - Z^\top)H_G \\
\tilde{F}_P - FH_G &= \tilde{F}_P - G_rH_G - G_p(I_r - Z^\top)H_G,
\end{aligned}
$$

where the result follows after taking the squared norms of both sides and diving by T. ■

Proof of Lemma 1 (d).

$$
\begin{aligned}
\frac{1}{T}\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top e_i &= V_{NT}^{-1}\frac{1}{T^2N}\left(\tilde{F}_P^\top G_pW^\top e^\top e_i + \tilde{F}_P^\top ee^\top e_i + \tilde{F}_P^\top e\Lambda_1 G_r^\top e_i + \tilde{F}_P^\top G_r\Lambda_1^\top e^\top e_i\right) \\
&= V_{NT,r}^{-1}\left(a_{33} + a_{34} + a_{35} + a_{36}\right), \\
a_{33} &= \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top e^\top e_i}{TN} = O_p\left(\frac{1}{\delta_{NT}^2}\right), \\
a_{34} &= \frac{\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top}{\sqrt{T}}\frac{ee^\top}{TN}\frac{e_i}{\sqrt{T}} + \frac{H_{\Xi,r}^\top G^\top e}{\sqrt{TN}}\frac{e^\top e_i}{T\sqrt{N}}\frac{1}{\sqrt{T}} \\
&= O_p\left(\frac{1}{\delta_{NT}}\right)O_p\left(\frac{1}{\delta_{NT}}\right) + \frac{1}{\sqrt{T}}O_p\left(\frac{1}{\delta_{NT}}\right), \\
a_{35} &= \frac{\tilde{F}_P}{\sqrt{T}}\frac{e\Lambda}{\sqrt{TN}}\frac{G_r^\top e_i}{\sqrt{T}}\frac{1}{\sqrt{TN}} = \frac{1}{\sqrt{TN}}O_p(1), \quad \text{and} \\
a_{36} &= \frac{\tilde{F}_P^\top G_p}{T}\frac{\Lambda_1^\top e^\top e_i}{TN} = O_p\left(\frac{1}{\delta_{NT}^2}\right), \\
\therefore \frac{1}{T}\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top e_i &= O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}
$$

■

Additionally, Equation (A.7) allows us to study the expansion of each $\tilde{g}_{r,t}$. Begin by considering $\tilde{f}_{P,t} - H_G^\top g_{r,t}$

$$
\begin{aligned}
\tilde{f}_{P,t} - H_G^\top g_{r,t} =& \frac{1}{NT}V_{NT}^{-1}\Big(\tilde{F}_P^\top e\Lambda_1 g_{r,t} + \tilde{F}_P^\top G_r\Lambda_1^\top e_t + \tilde{F}_P^\top ee_t + \tilde{F}_P^\top G_pW^\top e_t \\
&+ \tilde{F}_P^\top eWg_{p,t} + \tilde{F}_P^\top G_pW^\top Wg_{p,t} + \tilde{F}_P^\top G_pW^\top\Lambda_1 g_{r,t} + \tilde{F}_P^\top G_r\Lambda_1^\top Wg_{p,t}\Big) \\
=& V_{NT}^{-1}\left(a_{37} + a_{38} + a_{39} + a_{40} + a_{41} + a_{42} + a_{43} + a_{44}\right).
\end{aligned}
$$

Analysing each term, we have

$$
\begin{aligned}
a_{37} &= \frac{\tilde{F}_P^\top}{\sqrt{T}} \frac{e\Lambda_1}{\sqrt{TN}} g_{rt} \frac{1}{\sqrt{N}} \\
&= \frac{\left(\tilde{F}_P - G_r H_G\right)^\top}{\sqrt{T}} \frac{e\Lambda_1}{\sqrt{TN}} g_{rt} \frac{1}{\sqrt{N}} + \frac{H_G^\top G_r^\top e\Lambda_1}{TN} g_{rt} \\
&= \left(O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{N^\alpha}{N}\right)\right) \frac{1}{\sqrt{N}} + O_p\left(\frac{1}{\delta_{NT}^2}\right) \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N\sqrt{N}}\right), \\
a_{38} &= \frac{\tilde{F}_P^\top G_r}{T} \frac{\Lambda_1^\top e_t}{N} \\
&= \frac{\tilde{F}_P^\top G_r}{T} \frac{1}{N} \sum_{i=1}^N \lambda_{1i} e_{it} \\
&= O_p\left(\frac{1}{\sqrt{N}}\right), \\
a_{39} &= \frac{\tilde{F}_P^\top e e_t}{NT} \\
&= \frac{1}{TN} \sum_{s=1}^T e_t^\top e_s \widehat{F}_{P,s}^\top \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N\sqrt{N}}\right) \\
a_{40} &= \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top e_t}{N} \\
&= \frac{\tilde{F}_P^\top G_p}{T} \frac{\sqrt{N^\alpha}}{N} \frac{1}{\sqrt{N^\alpha}} \sum_{i=1}^N w_i e_{it} \\
&= O_p\left(\frac{\sqrt{N^\alpha}}{N}\right),
\end{aligned}
$$

$$
\begin{aligned}
a_{41} &= \frac{\left(\tilde{F}_P^\top - G_r H_G\right)^\top}{T} \frac{eW}{\sqrt{N^\alpha T}} \frac{\sqrt{N^\alpha}}{N} g_{p,t} + \frac{H_G^\top G_r^\top eW}{\sqrt{N^\alpha T}} \frac{\sqrt{N^\alpha}}{N\sqrt{T}} g_{p,t} \\
&= \left(O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{N^\alpha}{N}\right)\right) O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N\sqrt{T}}\right), \\
a_{42} &= \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N^\alpha} \frac{N^\alpha}{N} g_{p,t} \\
&= O_p\left(\frac{N^\alpha}{N}\right), \\
a_{43} &= \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top \Lambda_1}{N} g_{r,t} \\
&= O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \quad \text{and}
\end{aligned}
$$

$$
\begin{aligned}
a_{44} &= \frac{\tilde{F}_P^\top G_r}{T}\frac{\Lambda_1^\top W}{N} g_{p,t} \\
&= O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
\tilde{f}_{P,t} - H_G^\top g_{r,t} &= V_{NT}^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\frac{\Lambda_1^\top e_T}{N} + \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N} g_{p,t}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) \\
&= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{N^\alpha}{N}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right). \qquad \text{(A.12)}
\end{aligned}
$$

Finally, note that $g_{r,T} = Zf_T$, and therefore implies

$$
\tilde{f}_{P,T} - H_G^\top f_T = V_{NT}^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\frac{\Lambda_1^\top e_T}{N} + \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N} g_{p,T}\right) - H_G^\top (I - Z)F_T + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right). \qquad \text{(A.13)}
$$

Lemma 2. *Under Assumptions 1 to 8,*

$$
\text{plim}\, H_{\Xi,r} = Q_{G,r}^+,
$$

where $Q_{G,r} \equiv \Upsilon_G \Sigma_\Xi^{-1/2}$, V_G *is a diagonal matrix consisting of the first* $2r$ *largest eigenvalues of* $\Sigma_\Xi^{1/2}\Sigma_G\Sigma_\Xi^{1/2}$ *in descending order,* $\Sigma_G = \text{plim}\, \frac{1}{T}G^\top G$, *and* $+$ *denotes the pseudo inverse.*

Proof of Lemma 2. To see this, first note that the case of $\alpha = 1$ implies that $\frac{1}{N}\Xi^\top\Xi$ converges to Σ_Ξ which is positive definite. This allows us to use Proposition 1 of Bai (2003) to state the following probability limit for the $2r$ pseudo-factors

$$
\frac{\tilde{G}^\top G}{T} \xrightarrow{p} Q_G \equiv V_G^{1/2}\Upsilon_G^\top \Sigma_\Xi^{-1/2}, \qquad \text{(A.14)}
$$

where $\tilde{G}$ are $\sqrt{T}$ times the first $2r$ eigenvectors of $XX^\top/NT$, V_G is a diagonal matrix consisting of the first $2r$ largest eigenvalues of $\Sigma_\Xi^{1/2}\Sigma_G\Sigma_\Xi^{1/2}$ in descending order, and $\Sigma_G = \text{plim}\,\frac{1}{T}G^\top G$. A slight modification of the result in Bai (2003) via the continuous mapping theorem yields

$$
\begin{aligned}
\text{plim}\, \frac{\tilde{F}_P^\top G}{T} &= \text{plim}\begin{bmatrix} I_r & 0_r \end{bmatrix}\frac{\tilde{G}^\top G}{T} \\
&= \begin{bmatrix} I_r & 0_r \end{bmatrix} Q_G
\end{aligned}
$$

$$= \begin{bmatrix} V_r^{1/2} & 0_r \end{bmatrix} \Upsilon_G \Sigma_\Xi^{-1/2} \equiv Q_{G,r}, \tag{A.15}$$

which is an $r \times 2r$ matrix. The limit of $H_{\Xi,r}$ is therefore

$$\begin{aligned} H_{0,\Xi,r} &= \operatorname{plim} \frac{\Xi^\top \Xi}{N} \frac{G^\top \tilde{F}_P}{T} \left(\begin{bmatrix} I_r & 0_r \end{bmatrix} V_{NT,r} \begin{bmatrix} I_r \\ 0_r \end{bmatrix} \right)^{-1} \\ &= \Sigma_\Xi Q_{G,r}^\top V_r^{-1} \\ &= \Sigma_\Xi^{1/2} \Upsilon_G \begin{bmatrix} V_r^{-1/2} \\ 0_r \end{bmatrix} = Q_{G,r}^+, \end{aligned} \tag{A.16}$$

where $Q_{G,r}^+$ is the *pseudo* inverse of $Q_{G,r}$, and is a $2r \times r$ matrix.[10] ■

A.3 Split-sample Factors $\tilde{F}_S$

Proof of Theorem 1 (b). This is simply the subsample version of Theorem 1 of Bai and Ng (2002).

Note that Theorem 1 (b) can be equivalently stated as

$$\frac{1}{\sqrt{T}} \left\| \left(\tilde{F}_\iota - F_\iota H_\iota\right) - \frac{e_{(\iota)}\Lambda_\iota}{N} \frac{F_\iota^\top \tilde{F}_\iota}{T_\iota} V_{NT,\iota}^{-1} \right\| = O_p\left(\frac{1}{\delta_{NT}^2}\right), \tag{A.17}$$

due to $\frac{e_{(\iota)}\Lambda_\iota}{N} \frac{F_\iota^\top \tilde{F}_\iota}{T_\iota} V_{NT,\iota}^{-1}$ being the largest term. ■

Theorem 1 (b) also implies the following lemmas.

Lemma 3. *Under Assumptions 1 to 8 and as $N, T \to \infty$, for $\iota = 1, 2$,*

a) $\frac{1}{T} \left(\tilde{F}_\iota - F_\iota H_\iota\right)^\top F_\iota = O_p\left(\frac{1}{\delta_{NT}^2}\right)$,

b) $\frac{1}{T} \left(\tilde{F}_\iota - F_\iota H_\iota\right)^\top e_{i,(\iota)} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$,

where $e_{i,(1)} = (e_{i1}, \ldots, e_{i,\lfloor \pi T \rfloor})^\top$ *and* $e_{i,(2)} = (e_{i,\lfloor \pi T \rfloor + 1}, \ldots, e_{i,T})^\top$.

Proof of Lemma 3. These are simply the subsample versions of Lemmas B.1 and B.2 of Bai (2003). ■

Additionally, by eigen-identity, we have the following expansion:

$$\tilde{F}_2 - F_2 H_2 = \frac{1}{T_2 N} \left(F_2 \Lambda_2^\top e_{(2)} \tilde{F}_2 + e_{(2)} \Lambda_2 F_2^\top \tilde{F}_2 + e_{(2)} e_{(2)}^\top \tilde{F}_2 \right) V_{NT,2}^{-1}, \quad \text{and} \tag{A.18}$$

[10] The pseudo inverse identity follows from the fact that $(AB)^+ = B^+ A^+$ if A has linearly independent columns and B has linearly independent rows.

$$\tilde{f}_{2,t} - H_2^\top f_t = V_{NT,2}^{-1}\left(\frac{\tilde{F}_2^\top e_{(2)}^\top \Lambda_2}{T_2 N} f_t + \frac{\tilde{F}_2^\top F_2}{T_2}\frac{\Lambda_2^\top e_t}{N} + \frac{\tilde{F}_2^\top e_{(2)} e_t}{T_2 N}\right), \tag{A.19}$$

where following Bai (2003) it can be shown that the 1st and 3rd terms are $O_p\left(\frac{1}{\delta_{NT}^2}\right)$, and the second term is the $O_p\left(\frac{1}{\sqrt{N}}\right)$ dominating term.

A.4 Rotated Factors $\tilde{F}_R$

Proof of Proposition 1. Let $e_{(1)} = [e1, \ldots, e_{T_1}]^\top$ and $e_{(2)} = [e_{(T_1+1),\ldots,e_T}]^\top$ denote the partitioned errors.

$$\begin{aligned}
\tilde{Z} =& (\tilde{\Lambda}_1^\top \tilde{\Lambda}_1)^{-1}\tilde{\Lambda}_1^\top \tilde{\Lambda}_2 \\
=& \frac{1}{NT_1T_2} V_{NT,1}^{-1}(\tilde{F}_1^\top X_1)^\top (\tilde{F}_2^\top X_2) \\
=& V_{NT,1}^{-1}\frac{1}{NT_1T_2}\left(\tilde{F}_1^\top F_1 \Lambda_1^\top + \tilde{F}_1^\top e_1\right)\left(\tilde{F}_2^\top F_2 Z^\top \Lambda_1^\top + \tilde{F}_2^\top F_2 W^\top + \tilde{F}_2^\top e_{(2)}\right)^\top \\
=& V_{NT,1}^{-1}\frac{1}{NT_1T_2}\left(\tilde{F}_1^\top F_1 \Lambda_1^\top e_{(2)}^\top \tilde{F}_2 + \tilde{F}_1^\top F_1 \Lambda_1^\top W F_2^\top \tilde{F}_2 + \tilde{F}_1^\top F_1 \Lambda_1^\top \Lambda_1 Z F_2^\top \tilde{F}_2\right. \\
&\left. + \tilde{F}_1^\top e_{(1)} e_{(2)}^\top \tilde{F}_2 + \tilde{F}_1^\top e_{(1)} W F_2^\top \tilde{F}_2 + \tilde{F}_1^\top e_{(1)} \Lambda_1 Z F_2^\top \tilde{F}_2\right) \\
=& V_{NT,1}^{-1}(Z.I + Z.II + Z.III + Z.IV + Z.V + Z.VI)
\end{aligned}$$

We shall see that $Z.III$ characterises the convergence behaviour, and the remaining terms are all asymptotically negligible.

$$\begin{aligned}
Z.I &= \frac{\tilde{F}_1^\top F_1 \Lambda_1^\top e_{(2)}^\top \tilde{F}_2}{NT_1T_2} \\
&= \frac{\tilde{F}_1^\top F_1 \Lambda_1^\top e_{(2)}^\top (\tilde{F}_2 - F_2 H_2)}{T_1 N T_2} + \frac{\tilde{F}_1^\top F_1 \Lambda_1^\top e_{(2)}^\top F_2 H_2}{T_1 N T_2} \\
&\leq \left\|\frac{\tilde{F}_1^\top F_1}{T_1}\right\|\left\|\frac{\Lambda_1^\top e_{(2)}^\top}{N\sqrt{T_2}}\right\|\left\|\frac{\tilde{F}_2 - F_2 H_2}{\sqrt{T_2}}\right\| + \left\|\frac{\tilde{F}_1^\top F_1}{T_1}\right\|\left\|\frac{\Lambda_1^\top e_{(2)}^\top F_2}{NT_2}\right\|\|H_2\| \\
&= O_p(1) O_p\left(\frac{1}{\sqrt{N}}\right) O_p\left(\frac{1}{\delta_{NT}}\right) + O_p(1) O_p\left(\frac{1}{\delta_{NT}^2}\right) O_p(1) \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right). \\
Z.II &= \frac{\tilde{F}_1^\top F_1 \Lambda_1^\top W F_2^\top \tilde{F}_2}{NT_1T_2} \\
&= \frac{\tilde{F}_1^\top F_1}{T_1}\frac{\Lambda_1^\top W}{N}\frac{F_2^\top \tilde{F}_2}{T_2} \\
&= O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}$$

$$
\begin{aligned}
Z.IV &= \frac{\tilde{F}_1^\top e_{(1)} e_{(2)}^\top \tilde{F}_2}{NT_1T_2} \\
&= \frac{(\tilde{F}_1 - F_1H_1)^\top}{T_1} \frac{e_{(1)}e_{(2)}^\top}{N} \frac{(\tilde{F}_2 - F_2H_2)}{T_2} + \frac{(F_1H_1)^\top}{T_1} \frac{e_{(1)}e_{(2)}^\top}{N} \frac{(\tilde{F}_2 - F_2H_2)}{T_2} + \\
&\quad \frac{(\tilde{F}_1 - F_1H_1)^\top}{T_1} \frac{e_{(1)}e_{(2)}^\top}{N} \frac{(F_2H_2)}{T_2} + \frac{(F_1H_1)^\top}{T_1} \frac{e_{(1)}e_{(2)}^\top}{N} \frac{(F_2H_2)}{T_2} \\
&= Z.IV.a + Z.IV.b + Z.IV.c + Z.IV.d. \\
\|Z.IV.a\| &\leq \left\| \frac{(\tilde{F}_1 - F_1H_1)}{\sqrt{T_1}} \right\| \left\| \frac{e_{(1)}e_{(2)}^\top}{\sqrt{T_1}\sqrt{T_2}N} \right\| \left\| \frac{(\tilde{F}_2 - F_2H_2)}{\sqrt{T_2}} \right\| \\
&= O_p\left(\frac{1}{\delta_{NT}}\right) O_p\left(\frac{1}{\delta_{NT}}\right) O_p\left(\frac{1}{\delta_{NT}}\right) = O_p\left(\frac{1}{\delta_{NT}^3}\right). \\
\|Z.IV.b\| &\leq \left\| \frac{(F_1H_1)}{\sqrt{T_1}} \right\| \left\| \frac{e_{(1)}e_{(2)}^\top}{\sqrt{T_1}\sqrt{T_2}N} \right\| \left\| \frac{(\tilde{F}_2 - F_2H_2)}{\sqrt{T_2}} \right\| \\
&= O_p(1)\, O_p\left(\frac{1}{\delta_{NT}}\right) O_p\left(\frac{1}{\delta_{NT}}\right) = O_p\left(\frac{1}{\delta_{NT}^2}\right). \\
\|Z.IV.c\| &\leq \left\| \frac{(\tilde{F}_1 - F_1H_1)}{\sqrt{T_1}} \right\| \left\| \frac{e_{(1)}e_{(2)}^\top}{\sqrt{T_1}\sqrt{T_2}N} \right\| \left\| \frac{(F_2H_2)}{\sqrt{T_2}} \right\| \\
&= O_p\left(\frac{1}{\delta_{NT}}\right) O_p\left(\frac{1}{\delta_{NT}}\right) O_p(1) = O_p\left(\frac{1}{\delta_{NT}^2}\right). \\
\|Z.IV.d\| &\leq \|H_1\| \left\| \frac{F_1^\top e_{(2)}^\top}{T_1\sqrt{N}} \right\| \left\| \frac{e^\top F_2}{T_2\sqrt{N}} \right\| \|H_2\| \\
&= O_p(1)\, O_p\left(\frac{1}{\sqrt{T}}\right) O_p\left(\frac{1}{\sqrt{T}}\right) O_p(1) = O_p\left(\frac{1}{\delta_{NT}^2}\right) \\
\therefore Z.IV &= O_p\left(\frac{1}{\delta_{NT}^2}\right). \\
Z.V &= \frac{\tilde{F}_1^\top}{T_1} \frac{e_{(1)}W}{N} \frac{F_2^\top \tilde{F}_2}{T_2} \\
&\leq \left\| \frac{(\tilde{F}_1 - F_1H_1)}{\sqrt{T_1}} \right\| \left\| \frac{e_{(1)}W}{N^\alpha\sqrt{T_1}} \right\| \frac{N^\alpha}{N} \left\| \frac{F_2^\top \tilde{F}_2}{T_2} \right\| + \|H_2\| \left\| \frac{F_1^\top e_{(1)}W}{T_1N^\alpha} \right\| \frac{N^\alpha}{N} \left\| \frac{F_2^\top \tilde{F}_2}{T_2} \right\| \\
&= O_p\left(\frac{1}{\delta_{NT}}\right) O_p\left(\frac{1}{\sqrt{N^\alpha}}\right) \frac{N^\alpha}{N} O_p(1) + O_p(1)\, O_p\left(\frac{1}{\sqrt{N^\alpha T}}\right) \frac{N^\alpha}{N} O_p(1) \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}
$$

$$
\begin{aligned}
Z.VI &= \frac{\tilde{F}_1^\top}{T_1} \frac{e_{(1)}\Lambda_1 Z}{N} \frac{F_2^\top \tilde{F}_2}{T_2} \\
&\leq \left\| \frac{(\tilde{F}_1 - F_1H_1)}{\sqrt{T_1}} \right\| \left\| \frac{e_{(1)}\Lambda_1}{N\sqrt{T_1}} \right\| \|Z\| \left\| \frac{F_2^\top \tilde{F}_2}{T_2} \right\| + \|H\| \left\| \frac{F_1^\top e_{(1)}\Lambda_1}{T_1N} \right\| \|Z\| \left\| \frac{F_2^\top \tilde{F}_2}{T_2} \right\| \\
&= O_p\left(\frac{1}{\delta_{NT}}\right) O_p\left(\frac{1}{\sqrt{N}}\right) O_p(1)\, O_p(1) + O_p(1)\, O_p\left(\frac{1}{\delta_{NT}^2}\right) O_p(1)\, O_p(1)
\end{aligned}
$$

$$= O_p\left(\frac{1}{\delta_{NT}^2}\right).$$

Finally, note that $H_2 = \frac{\Lambda_2^\top \Lambda_2}{N}\frac{F_2^\top \tilde{F}_2}{T_2} V_{NT,2}^{-1}$ and

$$\begin{aligned}
F_2 H_2 + \tilde{F}_2 - F_2 H_2 &= \tilde{F}_2 \\
\frac{1}{T_2}\tilde{F}_2^\top F_2 H_2 + \frac{1}{T_2}\tilde{F}_2^\top(\tilde{F}_2 - F_2 H_2) &= I_r \\
\frac{1}{T_2}\tilde{F}_2^\top F_2 H_2 + O_p\left(\frac{1}{\delta_{NT}^2}\right) &= I_r \\
\frac{1}{T_2}\tilde{F}_2^\top F_2 &= H_2^{-1} + O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}$$

Therefore

$$\tilde{Z} = H_1^\top Z H_2^{-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)$$

as required. ■

Proof of Theorem 1 (c). From the consistency of $\tilde{Z}$, it follows that

$$\begin{aligned}
\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1 &= \tilde{F}_2(\tilde{Z}^\top - H_2^{-1} Z^\top H_1) + (\tilde{F}_2 H_2^{-1} - F_2) Z^\top H_1, \\
\tilde{F}_2\tilde{Z}^\top - F_2 H_1 &= \tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1 + F_2(Z^\top - I_r) H_1. \qquad \text{(A.20)}
\end{aligned}$$

Taking the squared norms of both sides and dividing by T yields the result. ■

Theorem 1 (c) additionally can be used to derive the following lemmas.[11]

Lemma 4. *Under Assumptions 1 to 8 and as $N, T \to \infty$:*

a) $\frac{1}{T}\left(\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1\right)^\top F_2 = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)$,

b) $\frac{1}{T}\left(\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1\right)^\top e_{i,(2)} = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{\sqrt{T}N}\right)$, *and*

c) $\frac{1}{T}\left(\tilde{F}_2\tilde{Z}^\top - F_2 H_1\right)^\top \tilde{F}_2 = -H_1^\top(I - Z)H_2^{-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)$.

Proof of Lemma 4 (a).

$$\frac{1}{T}\left(\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1\right)^\top F_2 = \frac{1}{T}\left(\tilde{F}_2(\tilde{Z}^\top - H_2^{-1} Z^\top H_1) + (\tilde{F}_2 - F_2 H_2) H_2^{-1} Z^\top H_1\right)^\top F_2$$

[11] Similarly, lemmas for $\frac{1}{T}(\tilde{F}_2\tilde{Z}^\top - F_2 H_1)^\top F_2$ (in terms of the true factors F) should be unnecessary.

$$
\begin{aligned}
&= \frac{1}{T}(\tilde{Z}^\top - H_2^{-1}Z^\top H_1)^\top \tilde{F}_2^\top F_2 + \frac{1}{T}H_1^\top Z H_2^{-\top}(\tilde{F}_2 - F_2H_2)^\top F_2 \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}
$$

■

Proof of Lemma 4 (b).

$$
\begin{aligned}
\frac{1}{T}\left(\tilde{F}_2\tilde{Z}^\top - F_2Z^\top H_1\right)^\top e_{i(2)} &= \frac{1}{T}\left(\tilde{F}_2(\tilde{Z}^\top - H_2^{-1}Z^\top H_1) + (\tilde{F}_2 - F_2H_2)H_2^{-1}Z^\top H_1\right)^\top e_{i(2)} \\
&= \frac{1}{T}(\tilde{Z}^\top - H_2^{-1}Z^\top H_1)^\top \tilde{F}_2^\top e_{i(2)} + \frac{1}{T}H_1^\top Z H_2^{-\top}(\tilde{F}_2 - F_2H_2)^\top e_{i(2)} \\
&= \left(\tilde{Z}^\top - H_2^{-1}Z^\top H_1\right)^\top \left(\frac{(\tilde{F}_2 - F_2H_2)^\top e_i}{T} + \frac{F_2^\top e_{i(2)}}{\sqrt{T}}\frac{1}{\sqrt{T}}\right) \\
&\quad + \frac{1}{T}H_1^\top Z H_2^{-\top}(\tilde{F}_2 - F_2H_2)^\top e_{i(2)} \\
&= \left(O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\right)\left(O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{\sqrt{T}N}\right).
\end{aligned}
$$

■

Proof of Lemma 4 (c).

Beginning with $\frac{(\tilde{F}_2\tilde{Z}^\top - F_2Z^\top H_1)^\top \tilde{F}_2}{T}$, we have

$$
\begin{aligned}
\frac{(\tilde{F}_2\tilde{Z}^\top - F_2Z^\top H_1)^\top \tilde{F}_2}{T} &= \frac{(\tilde{F}_2\tilde{Z}^\top - F_2Z^\top H_1)^\top F_2H_2}{T} + \frac{(\tilde{F}_2\tilde{Z}^\top - F_2Z^\top H_1)^\top}{\sqrt{T}}\frac{(\tilde{F}_2 - F_2H_2)}{\sqrt{T}} \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N\delta_{NT}}\right) \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}
$$

Adding and subtracting terms implies

$$
\begin{aligned}
\frac{(\tilde{F}_2\tilde{Z}^\top - F_2H_1)^\top \tilde{F}_2}{T} &= \frac{(\tilde{F}_2\tilde{Z}^\top - F_2Z^\top H_1)^\top \tilde{F}_2}{T} - \frac{H_1^\top(I-Z)F_2^\top \tilde{F}_2}{T} \\
&= -H_1^\top(I-Z)H_2^{-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \\
\therefore \frac{1}{T}(\tilde{F}_R - FH_1)^\top \tilde{F}_R &= -H_1^\top(I-Z)H_2^{-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}
$$

■

A.5 Case of $\tilde{r} < r$

We detail how our method still holds if $\tilde{r} < r$, and hence allows for averaging over an unknown number of factors, as long as this is below the true r. The proof consists of defining appropriate rotational bases H_G, H_Ξ, H_1, H_2 which comply with the existing theory, and ensuring that they have a valid probability limit.

Suppose that the practitioner wishes to use the factor estimates with $r^* < r$ as a possible averaging model. Define the $\tilde{F}_{P,r^*}, \tilde{F}_{1,r^*}$ and $\tilde{F}_{2,r^*}$ as the respective counterparts of $\tilde{F}_P, \tilde{F}_1$ and $\tilde{F}_2$ but using r^*. We specify counterparts of their rotational bases H_G, H_Ξ, H_1 and H_2 as

$$\underset{r\times r^*}{H_{G,r^*}} = \frac{\Lambda_1^\top \Lambda_1}{N} \frac{G_r^\top \tilde{F}_{r^*}}{T} V_{NT,r^*}^{-1}, \tag{A.21}$$

$$\underset{2r\times r^*}{H_{\Xi,r^*}} = \frac{\Xi^\top \Xi}{N} \frac{G^\top \tilde{F}_{r^*}}{T} V_{NT,r^*}^{-1}, \tag{A.22}$$

$$\underset{r\times r^*}{H_{1,r^*}} = \frac{\Lambda_1^\top \Lambda_1}{N} \frac{F_1^\top \tilde{F}_{1,r^*}}{T} V_{NT,1,r^*}^{-1}, \tag{A.23}$$

$$\underset{r\times r^*}{H_{2,r^*}} = \frac{\Lambda_2^\top \Lambda_2}{N} \frac{F_2^\top \tilde{F}_{2,r^*}}{T} V_{NT,2,r^*}^{-1}, \tag{A.24}$$

where $V_{NT,r^*}, V_{NT,1,r^*}$ and $V_{NT,2,r^*}$ are diagonal matrices consisting of the first r^* eigenvalues of $XX^\top/(NT)$, $X_1X_1^\top/(NT_1)$, and $X_2X_2^\top/(NT_2)$, respectively. First, note that all of these rotational bases are $O_p(1)$ because

$$\|H_{G,r^*}\| \leq \left\|\frac{\tilde{F}_{P,r^*}^\top \tilde{F}_{P,r^*}}{T}\right\|^{1/2} \left\|\frac{G_r^\top G_r}{T}\right\|^{1/2} \left\|\frac{\Lambda_1^\top \Lambda_1}{N}\right\| \left\|V_{NT,r^*}^{-1}\right\| = O_p(1),$$

$$\|H_{\Xi,r^*}\| \leq \left\|\frac{\tilde{F}_{P,r^*}^\top \tilde{F}_{P,r^*}}{T}\right\|^{1/2} \left\|\frac{G^\top G}{T}\right\|^{1/2} \left\|\frac{\Xi^\top \Xi}{N}\right\| \left\|V_{NT,r^*}^{-1}\right\| = O_p(1),$$

$$\|H_{1,r^*}\| \leq \left\|\frac{\tilde{F}_{1,r^*}^\top \tilde{F}_{1,r^*}}{T}\right\|^{1/2} \left\|\frac{F_1^\top F_1}{T}\right\|^{1/2} \left\|\frac{\Lambda_1^\top \Lambda_1}{N}\right\| \left\|V_{NT,1,r^*}^{-1}\right\| = O_p(1), \quad \text{and}$$

$$\|H_{2,r^*}\| \leq \left\|\frac{\tilde{F}_{2,r^*}^\top \tilde{F}_{2,r^*}}{T}\right\|^{1/2} \left\|\frac{F_2^\top F_2}{T}\right\|^{1/2} \left\|\frac{\Lambda_2^\top \Lambda_2}{N}\right\| \left\|V_{NT,2,r^*}^{-1}\right\| = O_p(1).$$

Therefore, Theorem 1 (a) and Lemma 1 which are the mean square consistency results for the pseudo-factors $\tilde{F}_P$ are all unaffected and still hold.

Next, we establish that these rotational bases have well defined probability limits. Similar to the case of $\alpha = 1$, we have

$$\begin{aligned}
\text{plim} \frac{\tilde{F}_{P,r^*}^\top G}{T} &= \text{plim} \begin{bmatrix} I_{r^*} & 0_{2r-r^*} \end{bmatrix} \frac{\tilde{G}^\top G}{T} \\
&= \begin{bmatrix} I_{r^*} & 0_{2r-r^*} \end{bmatrix} Q_G \\
&= \begin{bmatrix} V_{r^*}^{1/2} & 0_{2r-r^*} \end{bmatrix} \Upsilon_G \Sigma_\Xi^{-1/2} \equiv Q_{G,r^*},
\end{aligned}$$

which is a $r^* \times 2r$ matrix. The limit of H_{Ξ,r^*} is therefore

$$\begin{aligned}
H_{0,\Xi,r^*} &= \text{plim}\, H_{\Xi,r^*} \\
&= \text{plim} \frac{\Xi^\top \Xi}{N} \frac{G^\top \tilde{F}_{P,r^*}}{T} \left(\begin{bmatrix} I_{r^*} & 0_{2r-r^*} \end{bmatrix} V_{NT,r^*} \begin{bmatrix} I_{r^*} \\ 0_{2r-r^*} \end{bmatrix} \right)^{-1} \\
&= \Sigma_\Xi Q_{G,r^*}^\top \begin{bmatrix} V_{r^*}^{1/2} \\ 0_{2r-r^*} \end{bmatrix} V_{r^*}^{-1} \\
&= \Sigma_\Xi \Sigma_\Xi^{-1/2} \Upsilon_G \begin{bmatrix} V_{r^*}^{1/2} \\ 0_{2r-r^*} \end{bmatrix} V_{r^*}^{-1} \\
&= \Sigma_\Xi^{1/2} \Upsilon_G \begin{bmatrix} V_{r^*}^{-1/2} \\ 0_{2r-r^*} \end{bmatrix} = Q_{G,r^*}^+,
\end{aligned}$$

where Q_{G,r^*}^+ is the *pseudo* inverse of Q_{G,r^*}, and is a $2r \times r^*$ matrix. By defining $Q_1 = \text{plim} \frac{\tilde{F}_1^\top F_1}{T}$ and $Q_2 = \text{plim} \frac{\tilde{F}_2^\top F_2}{T}$, we can derive the limits of H_{1,r^*} and H_{2,r^*} as Q_{1,r^*} and $Q_{2,r*}$ in a similar way.

By Theorem 1 of Bai and Ng (2002), we have

$$\begin{aligned}
\left\| \tilde{F}_{1,r^*} - F_1 H_{1,r^*} \right\|^2 &= O_p\left(\frac{1}{\delta_{NT}^2} \right) \\
\left\| \tilde{F}_{2,r^*} - F_2 H_{2,r^*} \right\|^2 &= O_p\left(\frac{1}{\delta_{NT}^2} \right)
\end{aligned}$$

which shows that Theorem 1 (b) containing the mean square consistency of the split-sample factors $\tilde{F}_S$ are unaffected. Lemma 3 corresponds to Lemmas B.1 and B.2 of Bai (2003) by applying Theorem 1 (b), and therefore also holds.

Similarly, the proof of Proposition 1 still holds by simply replacing all cases of $\tilde{F}_1 - F_1 H_1$ and $\tilde{F}_2 - F_2 H_2$ with $\tilde{F}_1 - F_1 H_{1,r^*}$ and $\tilde{F}_2 - F_2 H_{2,r^*}$, respectively. The final step of Proposition 1 requires establishing that

$\frac{\tilde{F}_2^\top F_2}{T} = H_{2,r^*}^+$, where the result is now stated in terms of a pseudo inverse due to H_2 being a rectangular $r \times r^*$ matrix. This can holds because

$$
\begin{aligned}
F_2 H_{2,r^*} + \tilde{F}_{2,r^*} - F_2 H_{2,r^*} &= \tilde{F}_{2,r^*} \\
\frac{1}{T_2}\tilde{F}_{2,r^*}^\top F_2 H_{2,r^*} + \frac{1}{T_2}\tilde{F}_{2,r^*}^\top \left(\tilde{F}_{2,r^*} - F_2 H_{2,r^*}\right) &= I_{r^*} \\
\frac{1}{T_2}\tilde{F}_{2,r^*} F_2 &= H_2^+ + O_p\left(\frac{1}{\delta_{NT}^2}\right),
\end{aligned}
$$

where $\frac{1}{T_2}\tilde{F}_{2,r^*}^\top \left(\tilde{F}_{2,r^*} - F_2 H_{2,r^*}\right) = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ is implied by $\left\|\tilde{F}_{2,r^*} - F_2 H_{2,r^*}\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

A.6 Changing r

We detail how our decomposition can be extended to the allow for disappearing factors, and hence a change in the number of factors. Note that the case of an emerging factor can always be parameterised in by reversing the pre- and post-break samples, and it thus suffices to focus on the case of a disappearing factor.

Existing work tends to parameterise a disappearing factor by allowing for a singular Z, (e.g. Han and Inoue, 2015; Baltagi et al., 2017; Bai et al., 2024). However, these approaches work by using the *pseudo*-factors - the case of split-sample estimation is more difficult. The main issue is to ensure that H_2 has valid limiting behaviour - once this is done, the proofs for the split-sample factors and rotated factors can follow on without major adjustments.

Without loss of generality, suppose that the $r - r_2$th factors disappear. To avoid Λ_2 not being of full column rank, we instead parameterise Λ_2 as an $N \times (r - r_2)$ matrix:

$$
\begin{aligned}
\Lambda_2 &= (\Lambda_1 + W)\begin{bmatrix} I_{r-r_2} \\ 0 \end{bmatrix} \\
&= \Lambda_1 Z_0 + W_0. \qquad \text{(A.25)}
\end{aligned}
$$

This allows us to write

$$
\begin{aligned}
X_2 &= F_2 \Lambda_2^\top + e_{(2)} \\
&= F_2 \begin{bmatrix} I_{r-r_2} \\ 0 \end{bmatrix} \left((\Lambda_1 Z + W) \begin{bmatrix} I_{r-r_2} \\ 0 \end{bmatrix} \right)^\top + e_{(2)}
\end{aligned}
$$

$$= F_{2,r-r_2} \left(\Lambda_1 Z_0 + W_0 \right)^\top + e_{(2)}, \tag{A.26}$$

which expresses the post-break data as a factor structure with $r - r_2$ factors. We can therefore apply the usual framework of Bai (2003) and use

$$H_{2,r-r_2} = \frac{\Lambda_2^\top \Lambda_2}{N} \frac{F_{2,r-r_2}^\top \tilde{F}_2}{T_2} V_{NT,2,r-r_2}^{-1} \tag{A.27}$$

where we can use the first $r - r_2$ post-break factors denoted by $\tilde{F}_{2,r-r_2}$. All of the above quantities exhibit full rank, and hence $H_{2,r-r_2}$ is an $(r - r_2) \times (r - r_2)$ square matrix.

Lemma 5. *Under Assumptions 1 to 8, as $N, T \to \infty$*

a) $\frac{1}{T} \left\| \tilde{F}_{2,r-r_2} - F_{2,r-r_2} H_{2,r-r_2} \right\|^2 = O_p \left(\frac{1}{\delta_{NT}^2} \right)$,

b) $\frac{1}{T} \left(\tilde{F}_{2,r-r_2} - F_{2,r-r_2} H_{2,r-r_2} \right)^\top F_{2,r-r_2} = O_p \left(\frac{1}{\delta_{NT}^2} \right)$

c) $\frac{1}{T} \left(\tilde{F}_{2,r-r_2} - F_{2,r-r_2} H_{2,r-r_2} \right)^\top e_{i,(2)} = O_p \left(\frac{1}{\delta_{NT}^2} \right)$

Proof of Lemma 5. These correspond to Theorem of Bai and Ng (2002) and Lemmas B.1 and B.2 of Bai (2003). ■

Lemma 5 can also be used to prove analogous results for the rotated factors, where $\tilde{Z}$ is now an $r \times (r - r_2)$ matrix.

Lemma 6. *Under Assumptions 1 to 8, as $N, T \to \infty$*

a) $\tilde{Z} = H_1^\top Z_0 H_{2,r-1}^{-\top} + O_p \left(\frac{1}{\delta_{NT}^2} \right)$,

b) $\frac{1}{T} \left\| \tilde{F}_{2,r-r_2} \tilde{Z}^\top - F_2 Z_0^\top H_1 \right\|^2 = O_p \left(\frac{1}{\delta_{NT}^2} \right) + O_p \left(\frac{N^{2\alpha}}{N^2} \right)$,
$\frac{1}{T} \left\| \tilde{F}_{2,r-r_2} \tilde{Z}^\top - F_1 H_1 \right\|^2 = O_p \left(\frac{1}{\delta_{NT}^2} \right) + O_p \left(\frac{N^{2\nu}}{N^2} \right) + O_p \left(\frac{\sqrt{N^\alpha}}{N} \right)$, *and*

c) $\frac{1}{T} \left(\tilde{F}_{2,r-r_2} - F_{2,r-r_2} Z_0^\top H_1 \right)^\top F_{2,r-r_2} = O_p \left(\frac{1}{\delta_{NT}^2} \right) + O_p \left(\frac{\sqrt{N^\alpha}}{N} \right)$,
$\frac{1}{T} \left(\tilde{F}_{2,r-r_2} - F_{2,r-r_2} Z_0^\top H_1 \right)^\top e_{i(2)} = O_p \left(\frac{1}{\delta_{NT}^2} \right) + O_p \left(\frac{\sqrt{N^\alpha}}{N\sqrt{T}} \right)$.

Proof of Lemma 6. Lemmas 6 (a) to 6 (c) are analogous to Proposition 1, Theorem 1 (c), and Lemma 4, and are all proved in a similar way. ■

A.7 Mis-specified Break Fraction

We show how our method can adapt to a possible mis-specified break fraction π^*, enabling a practitioner to average over a finite number of candidate breaks.

Consistent Estimation of the Break Fraction

We first detail the rates and conditions regarding estimation of π. The least-squares estimator of Bai et al. (2020) is consistent for the *break index* $k = \lfloor \pi T \rfloor$ for $\alpha > 0$. Therefore, for any $\alpha > 0$ the break fraction can be treated as known, regardless of ν.

Rotational breaks are more difficult to deal with. When $\nu < 0.5$ and $\alpha = 0$, the impact of the rotational break is small enough to not impact the forecasting coefficients. Therefore, even though these breaks cannot be consistently estimated, they are safe to ignore. When $\nu > 0.5$, this constitutes a large enough break in the coefficients that can be consistently estimated. To see this, the results of Bai (1997) show that the break fraction can still be consistently estimated as long as the break is large enough. In our context, this would correspond to $N^{2-2\nu}$, implying an error of $o(N)$ for the break index.

The case of $\nu = 0.5$ and $\alpha = 0$ represents a rare case where the break fraction cannot be consistently estimated, and also coincides to the case where no one estimation method for the factors clearly dominates any of the others.

Therefore, it is only in the rare cases of $\nu < 0.5, \alpha = 0$, and $\nu = 0.5, \alpha = 0$ where the break fraction cannot be estimated - and only the latter case could be of interest to a practitioner. We work around this by showing that the split-sample factors $\tilde{F}_S$ and rotated factors $\tilde{F}_R$ still exhibit proper limiting behaviour when the break is possibly mis-specified. This allows the practitioner to additionally select and/or average over a finite number of "candidate" break fractions for forecasting. The use of model averaging using cross-validation can be justified by showing analogous results, and requires the careful specification of a rotational matrix that has clearly defined limits.

Note that the pseudo-factors $\tilde{F}_P$ do not use any partitioning of the data, and thus the following results are only necessary for analysing the split-sample factors $\tilde{F}_S$ and rotated factors $\tilde{F}_R$. Let X_1^* and X_2^* denote the $T_1^* = \lfloor \pi^* T \rfloor \times N$ and $\lfloor (1-\pi^*)T \times N \rfloor$ partitions defined by π^*, $\tilde{F}_1^*$ and $\tilde{F}_2^*$ the respective estimates of the factors using principal components, and $\tilde{\Lambda}_1^*$ and $\tilde{\Lambda}_2^*$ the respective factor loadings as estimated by least squares.

Case 1: Break Fraction is under-estimated $\pi^* < \pi$.

In this case, write the X matrix as

$$X = \left[\begin{array}{cc} F_{11}^* & 0 \\ \hline F_{12}^* & 0 \\ F_2 Z^\top & F_2 \end{array}\right] \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + e,$$

where F_{11}^* is $\lfloor \pi^* T \rfloor \times r = T_1^* \times r$, and F_{12}^* is $\lfloor (1-\pi^*)T \rfloor \times r$.

Therefore, using π^* to partition X implies the following equivalent representation theorem:

$$\begin{aligned} X &= \begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix} \\ &= \begin{bmatrix} F_{11}^* & 0 \\ G_r^* & G_p^* \end{bmatrix} \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + e \\ &= \begin{bmatrix} F_{11}^* & 0 \\ G^* & \end{bmatrix} \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + e, \end{aligned}$$

where G_r^* and G^* are both T_2^* in length. Thus, the case of using a mis-specified $\pi^* < \pi$ can be analysed as the case of a factor structure with no break F_{11}^*, and *pseudo*-factors G_r^* or $G^* = \begin{bmatrix} G_r^* & G_p^* \end{bmatrix}$ after the break.

We specify the following rotational bases

$$\begin{aligned} H_1^* &= \frac{\Lambda_1^\top \Lambda_1}{N} \frac{F_{11}^{*\top} \tilde{F}_1^*}{T_1^*} V_{NT,1}^{*-1}, \\ H_{2,r}^* &= \frac{\Lambda_1^\top \Lambda_1}{N} \frac{G_r^{*\top} \tilde{F}_2^*}{T_2^*} V_{NT,2}^{*-1}, \quad \text{and} \\ H_{2,\Xi}^* &= \frac{\Xi^\top \Xi}{N} \frac{G^{*\top} \tilde{F}_2^*}{T_2^*} V_{NT,2}^{*-1}, \end{aligned}$$

where $V_{NT,1}^*$ and $V_{NT,2}^*$ are diagonal matrices consisting of the first r eigenvalues of $X_1^* X_1^{*\top}/(NT_1^*)$ and $X_2^* X_2^{*\top}/(NT_2^*)$.

This allows us to state the following consistency result for the split-sample factors $\tilde{F}_S^*$.

Lemma 7. *Under Assumptions 1 to 8, as $N, T \to \infty$*

a) $\frac{1}{T}\left\| \tilde{F}_1^* - F_{11}^* H_1^* \right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$,

b) $\frac{1}{T}\left\| \tilde{F}_2^* - G_r^* H_{2,r}^* \right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^{2\alpha}}{N^2}\right)$, *for $\alpha < 1$,*

$\frac{1}{T}\left\|\tilde{F}_2^* - G^* H_{2,\Xi}^*\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *for* $\alpha = 1$.

c) $\frac{1}{T}\left(\tilde{F}_1^* - F_{11}^* H_{1,r}^*\right)^\top F_{11}^* = O_p\left(\frac{1}{\delta_{NT}^2}\right)$,

$\frac{1}{T}\left(\tilde{F}_2^* - G_r^* H_{2,r}^*\right)^\top G_r^* = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right)$ *if* $\alpha < 1$,

$\frac{1}{T}\left(\tilde{F}_2^* - G^* H_{2,\Xi}^*\right)^\top G^* = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *if* $\alpha = 1$,

d) $\frac{1}{T}\left(\tilde{F}_1^* - F_{11}^* H_{1,r}^*\right)^\top e_{i(1)} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$,

$\frac{1}{T}\left(\tilde{F}_2^* - G_r^* H_{2,r}^*\right)^\top e_{i(2)} = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N\sqrt{T}}\right)$ *if* $\alpha < 1$,

$\frac{1}{T}\left(\tilde{F}_2^* - G^* H_{2,\Xi}^*\right)^\top e_{i(2)} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *if* $\alpha = 1$.

Lemma 7 (a) follows from Theorem 1 of Bai and Ng (2002). Lemma 7 (b) follows by applying the results of Theorem 1 (a) to the post-break factors. Lemmas 7 (c) and 7 (d) are the counterparts to Lemma 3. Lemma 7 also allows us to state the following lemmas.

Next, we focus on the rotated factors. Lemma 7 also allows us to state the following results for the rotated factors $\tilde{F}_R^* = \left[\tilde{F}_1^{*\top}, Z^* \tilde{F}_2^{*\top}\right]^\top$, where $\tilde{Z}^* = \left(\tilde{\Lambda}_1^{*\top}\tilde{\Lambda}_1^*\right)^{-1}\tilde{\Lambda}_1^{*\top}\tilde{\Lambda}_2^*$.

Lemma 8. *Under Assumptions 1 to 8, as* $N, T \to \infty$

$$
a)\ \tilde{Z}^* = \begin{cases} H_1^{*\top} H_{2,r}^{*-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right), & \alpha < 1, \\ H_1^{*\top}\frac{G_r^{*\top}\tilde{F}_2^*}{T_2^*} + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right), & \alpha = 1; \end{cases}
$$

b) $\frac{1}{T}\left\|\tilde{F}_2^*\tilde{Z}^{*\top} - G_r^* H_{1,r}^*\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^{2\alpha}}{N^2}\right)$ *if* $\alpha < 1$,

$\frac{1}{T}\left\|\tilde{F}_2^*\tilde{Z}^{*\top} - \frac{G^* H_{2,\Xi}\tilde{F}_2^{*\top}}{T_2} G_r^* H_1^*\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N^2}\right)$ *if* $\alpha = 1$.

Lemma 8 (a) shows that, in the case of a mis-specified break fraction, the estimated rotation $\tilde{Z}^*$ can still be used as a way to join the pre- and post-break factors together. Lemma 8 (a) shows the corresponding mean square consistency results for the rotated factors, which can be used to formulate their limiting behaviour. Because $\tilde{F}_2^*$ is estimating a set of *pseudo*-factors, both sets of results need to be stated for $\alpha < 1$ and $\alpha = 1$ separately.

Proof of Lemma 8. We first prove the consistency of $\tilde{Z}^*$. Expanding out $\tilde{Z}^*$ we have

$$
\begin{aligned}
\tilde{Z}^* &= \left(\tilde{\Lambda}_1^{*\top}\tilde{\Lambda}_1^*\right)^{-1}\tilde{\Lambda}_1^{*\top}\tilde{\Lambda}_2^* \\
&= \frac{1}{NT_1^*T_2^*} V_{NT,1}^{*-1}\left(\tilde{F}_1^{*\top}X_1^*\right)\left(\tilde{F}_2^{*\top}X_2^*\right)^\top
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{NT_1^*T_2^*} V_{NT,1}^{*-1} \left(\tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top + \tilde{F}_1^{*\top} e_{(1)} \right) \left(\tilde{F}_2^* G^* \Xi^\top + \tilde{F}_2^* e_{(2)} \right)^\top \\
&= \frac{1}{NT_1^*T_2^*} V_{NT,1}^{*-1} \left(\tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top \Xi G^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top e_{(2)}^\top \tilde{F}_2^* + \tilde{F}_1^{*\top} e_{(1)} \Xi^* G^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} e_{(1)} e_{(2)}^\top \tilde{F}_2^* \right) \\
&= \frac{1}{NT_1^*T_2^*} V_{NT,1}^{*-1} \left(\tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top \Lambda_1 G_r^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top W G_p^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top e_{(2)}^\top \tilde{F}_2^* \right. \\
&\qquad \left. + \tilde{F}_1^{*\top} e_{(1)} \Lambda_1 G_r^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} e_{(1)} W G_p^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} e_{(1)} e_{(2)}^\top \tilde{F}_2^* \right) \\
&= (Z.i + Z.ii + Z.iii + Z.iv + Z.v + Z.vi),
\end{aligned}
$$

where the first term is the main dominating term, and $Z.ii, Z.iii, Z.iv$ and $Z.v$ are asymptotically negligible because

$$
\begin{aligned}
Z.ii &= V_{NT,1}^{*-1} \frac{\tilde{F}_1^{*\top} F_{11}^*}{T_1^*} \frac{\Lambda_1^\top W}{N} \frac{G_p^{*\top} \tilde{F}_2}{T_2^*} \\
&= O_p\left(\frac{\sqrt{N^\alpha}}{N} \right), \\
Z.iii &= V_{NT,1}^{*-1} \frac{\tilde{F}_1^{*\top} F_{11}^*}{T_1^*} \frac{\Lambda_1^\top e_{(2)}^\top G_2^* H_{2,\Xi}^*}{NT_2^*} + \frac{\tilde{F}_1^{*\top}}{\sqrt{T_1^*}} \frac{F_{11}^* \Lambda_1^\top e_{(2)}^\top}{N\sqrt{T_2^*}} \frac{\tilde{F}_2^* - G_2^* H_{2,\Xi}^*}{\sqrt{T_2^*}} \\
&= O_p\left(\frac{1}{\delta_{NT}^2} \right) + O_p\left(\frac{1}{\delta_{NT}^2} \right) = O_p\left(\frac{1}{\delta_{NT}^2} \right), \\
Z.iv &= V_{NT,1}^{*-1} \frac{\left(\tilde{F}_1^* - F_{11}^* H_1^* \right)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} \Lambda_1}{N\sqrt{T_1^*}} \frac{G_r^{*\top} \tilde{F}_2^*}{T_2^*} + \frac{(F_{11}^* H_1^*)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} \Lambda_1}{N\sqrt{T_1^*}} \frac{G_r^{*\top} \tilde{F}_2^*}{T_2^*} \\
&= O_p\left(\frac{1}{\delta_{NT}^2} \right) + O_p\left(\frac{1}{\delta_{NT}^2} \right) = O_p\left(\frac{1}{\delta_{NT}^2} \right), \\
Z.v &= V_{NT,1}^{*-1} \frac{\left(\tilde{F}_1^* - F_{11}^* H_1^* \right)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} W}{\sqrt{N^\alpha T_1^*}} \frac{G_p^{*\top} \tilde{F}_2^*}{T_2^*} \frac{\sqrt{N^\alpha}}{N} + \frac{(F_{11}^* H_1^*)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} W G_p^{*\top}}{\sqrt{T_2^* N^\alpha}} \frac{\tilde{F}_2^*}{\sqrt{T_2^*}} \frac{\sqrt{N^\alpha}}{N} \frac{1}{\sqrt{T_1^*}} \\
&= O_p\left(\frac{1}{\delta_{NT}} \right) \frac{\sqrt{N^\alpha}}{N} + O_p\left(\frac{1}{\sqrt{T}} \right) \frac{\sqrt{N^\alpha}}{N} = O_p\left(\frac{1}{\delta_{NT}^2} \right).
\end{aligned}
$$

The term $Z.vi$ can be further decomposed as

$$
\begin{aligned}
Z.vi &= V_{NT,1}^{*-1} \frac{\left(\tilde{F}_1^* - F_{11}^* H_1^* \right)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} e_{(2)}^\top}{N\sqrt{T_1^* T_2^*}} \frac{\left(\tilde{F}_2^* - G^* H_{2,\Xi}^* \right)}{\sqrt{T_2^*}} + V_{NT,1}^{*-1} \frac{(F_{11}^* H_1^*)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} e_{(2)}^\top}{N\sqrt{T_1^* T_2^*}} \frac{\left(\tilde{F}_2^* - G^* H_{2,\Xi}^* \right)}{\sqrt{T_2^*}} \\
&\quad + V_{NT,1}^{*-1} \frac{\left(\tilde{F}_1^* - F_{11}^* H_1^* \right)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} e_{(2)}^\top}{N\sqrt{T_1^* T_2^*}} \frac{G^* H_{2,\Xi}^*}{\sqrt{T_2^*}} + V_{NT,1}^{*-1} \frac{(F_{11}^* H_1^*)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} e_{(2)}^\top}{N\sqrt{T_1^* T_2^*}} \frac{G^* H_{2,\Xi}^*}{\sqrt{T_2^*}} \\
&= O_p\left(\frac{1}{\delta_{NT}^3} \right) + O_p\left(\frac{1}{\delta_{NT}^2} \right) + O_p\left(\frac{1}{\delta_{NT}^2} \right) + O_p\left(\frac{1}{\delta_{NT}^2} \right), \\
&= O_p\left(\frac{1}{\delta_{NT}^2} \right).
\end{aligned}
$$

To analyse the leading term $Z.i$, note that for the case of $\alpha < 1$, $H_{2,r}^*$ is an $r \times r$ invertible matrix, and therefore

$$
\begin{aligned}
G_r H_{2,r}^* + \tilde{F}_2^* - G_r H_{2,r}^* &= \tilde{F}_2^* \\
\frac{1}{T_2^*} \tilde{F}_2^{*\top} G_r H_{2,r}^* &= I_r - \frac{1}{T_2^*} \tilde{F}_2^{*\top} \left(\tilde{F}_2^* - G_r H_{2,r}^* \right) \\
\frac{\tilde{F}_2^{*\top} G_r}{T_2^*} &= H_{2,r}^{*-1} + O_p\left(\frac{N^\alpha}{N} \right) + O_p\left(\frac{1}{\delta_{NT}^2} \right),
\end{aligned}
$$

where the last line uses Lemma 7 (c). Using the definition of H_1^*, it follows that

$$
\begin{aligned}
\tilde{Z}^* &= H_1^{*\top} \frac{G_r^\top \tilde{F}_2^*}{T_2^*} + O_p\left(\frac{\sqrt{N^\alpha}}{N} \right) + O_p\left(\frac{1}{\delta_{NT}^2} \right) \\
&= H_1^{*\top} H_{2,r}^{*-\top} + O_p\left(\frac{N^\alpha}{N} \right) + O_p\left(\frac{\sqrt{N^\alpha}}{N} \right) + O_p\left(\frac{1}{\delta_{NT}^2} \right),
\end{aligned}
$$

where the first and last lines can be used to establish the results for $\tilde{F}_2^* \tilde{Z}^{*\top}$ for $\alpha = 1$ and $\alpha < 1$, respectively. For $\alpha < 1$, we have

$$
\begin{aligned}
\frac{1}{\sqrt{T}} \left(\tilde{F}_2^* \tilde{Z}^{*\top} - G_r^* H_1^* \right) &= \frac{1}{\sqrt{T}} \tilde{F}_2^* \left(\tilde{Z}^{*\top} - H_{2,r}^{*-1} H_1^* \right) + \frac{1}{\sqrt{T}} \left(\tilde{F}_2^* H_{2,r}^{*-1} - G_r^* \right) H_1, \\
&= O_p\left(\frac{N^\alpha}{N} \right) + O_p\left(\frac{\sqrt{N^\alpha}}{N} \right) + O_p\left(\frac{1}{\delta_{NT}^2} \right) + O_p\left(\frac{1}{\delta_{NT}} \right) + O_p\left(\frac{N^\alpha}{N} \right).
\end{aligned}
$$

For the case of $\alpha = 1$, $\tilde{F}_2^*$ is consistent for $G^* H_{2,\Xi}^*$, where the rotation matrix is $2r \times r$ and therefore does not have an inverse. Our consistency result is, therefore,

$$
\begin{aligned}
\frac{1}{\sqrt{T}} \left(\tilde{F}_2^* \tilde{Z}^{*\top} - \frac{G H_{2,\Xi} \tilde{F}_2^{*\top}}{T_2} G_r^* H_1^* \right) &= \frac{1}{\sqrt{T}} G H_{2,\Xi} \left(\tilde{Z}^{*\top} - \frac{\tilde{F}_2^{*\top} G_r H_1}{T} \right) + \frac{1}{\sqrt{T}} \left(\tilde{F}_2^* - G H_{2,\Xi} \right) \tilde{Z}^{*\top} + O_p\left(\frac{1}{\delta_{NT}^2} \right) \\
&= O_p\left(\frac{\sqrt{N^\alpha}}{N} \right) + O_p\left(\frac{1}{\delta_{NT}^2} \right).
\end{aligned}
$$

In both cases, collecting the dominating terms and squaring both sides yields the result. ■

Case 2: Break Fraction is over-estimated $\pi^* > \pi$.

In this case, consider the following partition for X:

$$X = \left[\begin{array}{cc} F_1 & 0 \\ F_{21}^* Z^\top & F_{21}^* \\ \hline F_{22}^* Z^\top & F_{22}^* \end{array}\right] \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + e.$$

This implies the following equivalent representation theorem:

$$\begin{aligned} X &= \begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix} \\ &= \begin{bmatrix} G_r^* & G_p^* \\ F_{22}^* Z^\top & F_{22}^* \end{bmatrix} \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + e \\ &= \begin{bmatrix} G_1^* \\ G_2^* \end{bmatrix} \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + e. \end{aligned}$$

Note that in this parameterisation, $G_2^* \Xi^\top = F_{22}^* (\Lambda_1 Z + W)^\top = F_{22}^* \Lambda_2$. This allows us to specify the following rotational bases

$$\begin{aligned} H_{1,r}^* &= \frac{\Lambda_1^\top \Lambda_1}{N} \frac{G_r^{*\top} \tilde{F}_1^*}{T_1} V_{NT,1}^{*-1}, \\ H_{1,\Xi}^* &= \frac{\Xi^\top \Xi}{N} \frac{G_1^{*\top} \tilde{F}_1^*}{T_1^*} V_{NT,1}^{*-1}, \\ H_2^* &= \frac{\Lambda_2^\top \Lambda_2}{N} \frac{F_{22}^{*\top} \tilde{F}_2^*}{T_2^*} V_{NT,2}^{*-1}. \end{aligned}$$

where $V_{NT,1}^*$ and $V_{NT,2}^*$ are diagonal matrices consisting of the first r eigenvalues of $X_1^* X_1^{*\top}/(NT_1^*)$ and $X_2^* X_2^{*\top}/(NT_2^*)$.

Lemma 9. *Under Assumptions 1 to 8 and as $N, T \to \infty$*

a) $\frac{1}{T}\left\|\tilde{F}_1^* - G_r^* H_{1,r}^*\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^{2\alpha}}{N^2}\right),$
$\frac{1}{T}\left\|\tilde{F}_1^* - G^* H_{1,\Xi}^*\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right),$
$\frac{1}{T}\left\|\tilde{F}_2^* - F_{22}^* H_2^*\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right),$

b) $\frac{1}{T}\left(\tilde{F}_1^* - G_r^* H_{1,r}^*\right)^\top G_r^* = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right)$ *for* $\alpha < 1,$
$\frac{1}{T}\left(\tilde{F}_1^* - G^* H_{1,\Xi}^*\right)^\top G_r^* = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *for* $\alpha = 1,$

$\frac{1}{T}\left(\tilde{F}_2^* - F_{22}^* H_2^*\right)^\top F_{22}^* = O_p\left(\frac{1}{\delta_{NT}^2}\right),$

c) $\frac{1}{T}\left(\tilde{F}_1^* - G_r^* H_{1,r}^*\right)^\top e_{i(1)} = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N\sqrt{T}}\right)$ *for* $\alpha < 1$,

$\frac{1}{T}\left(\tilde{F}_1^* - G^* H_{1,\Xi}^*\right)^\top e_{i(2)} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *for* $\alpha = 1$,

$\frac{1}{T}\left(\tilde{F}_2^* - F_{22}^* H_2^*\right)^\top e_{i(2)} = O_p\left(\frac{1}{\delta_{NT}^2}\right).$

Proof of Lemma 9. Lemma 9 (a) corresponds to Theorem 1 (a) and Theorem 1 (b) and can be proven similarly. Lemmas 9 (b) and 9 (c) correspond to Lemmas 1 and 3 and can be proven similarly. ■

Lemma 9 similarly allows us to present the following results for the rotated factors.

Lemma 10. *Under Assumptions 1 to 8 and as* $N, T \to \infty$

a) $$\tilde{Z}^* = \begin{cases} H_{1,r}^{*\top} Z H_{2,r}^{*-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right), & \alpha < 1, \\ H_{1,\Xi}^{*\top} H_{2,\Xi}^{*-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right), & \alpha = 1. \end{cases}$$

b) $\frac{1}{T}\left\|\tilde{F}_2^* \tilde{Z}^{*\top} - F_{22}^* Z^\top H_{1,r}^*\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^{2\alpha}}{N^2}\right)$ *for* $\alpha < 1$,

$\frac{1}{T}\left\|\tilde{F}_2^* \tilde{Z}^{*\top} - G_2^* H_{1,\Xi}^*\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *for* $\alpha = 1$.

c) $\frac{1}{T}\left(\tilde{F}_2^* \tilde{Z}^{*\top} - F_{22}^* H_2^*\right)^\top F_{22}^* = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right)$ *for* $\alpha < 1$,

$\frac{1}{T}\left(\tilde{F}_2^* \tilde{Z}^{*\top} - G_2^* H_2^*\right)^\top G_2^* = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *for* $\alpha = 1$.

d) $\frac{1}{T}\left(\tilde{F}_2^* \tilde{Z}^{*\top} - F_{22}^* H_2^*\right)^\top e_{i(2)} = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N\sqrt{T}}\right)$ *for* $\alpha < 1$,

$\frac{1}{T}\left(\tilde{F}_2^* \tilde{Z}^{*\top} - G_2^* H_2^*\right)^\top e_{i(2)} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *for* $\alpha = 1$.

The rotated factors work by rotating the post-break factors onto the same rotational basis as the pre-break factors. In the case of an over-estimated break fraction $\pi^* > \pi$, this causes the estimated pre-break factors to exhibit a pseudo-factor representation, and similar to the case of analysing the pseudo-factors, care needs to be taken in specifying a rotational basis with proper limiting behaviour. To achieve this, Lemma 10 is stated separately for the cases of $\alpha < 1$ and $\alpha = 1$.

Proof of Lemma 10. We first prove the consistency result for $\tilde{Z}^*$.

$$\begin{aligned} \tilde{Z}^* &= \left(\tilde{\Lambda}_1^{*\top} \tilde{\Lambda}_1^*\right)^{-1} \tilde{\Lambda}_1^{*\top} \tilde{\Lambda}_2^* \\ &= \frac{1}{NT_1^* T_2^*} V_{NT,1}^{*-1} \left(\tilde{F}_1^{*\top} X_1^*\right) \left(\tilde{F}_2^{*\top} X_2^*\right)^\top \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{NT_1^* T_2^*} V_{NT,1}^{*-1} \left(\tilde{F}_1^* G_1^* \Xi^\top + \tilde{F}_1^* e_{(1)} \right) \left(\tilde{F}_2^{*\top} G_2^* \Xi + \tilde{F}_2^{*\top} e_{(2)} \right)^\top \\
&= \frac{1}{NT_1^* T_2^*} V_{NT,1}^{*-1} \left(\tilde{F}_1^* G_1^* \Xi^\top \Xi G_2^{*\top} \tilde{F}_2^* + \tilde{F}_1^* G_1^* \Xi^\top e_{(2)}^\top \tilde{F}_2^* + \tilde{F}_1^* e_{(1)}^\top \Xi G_2^{*\top} \tilde{F}_2^* + \tilde{F}_1^* e_{(1)}^\top e_{(e)} \tilde{F}_2^* \right) \\
&= Z.vi + Z.vii + Z.viii + Z.ix.
\end{aligned}
$$

The the last three terms are asymptotically negligible because

$$
\begin{aligned}
Z.vii &= V_{NT,1}^{*-1} \frac{\tilde{F}_1^{*\top} G^*}{T_1^*} \frac{\Xi^\top e_{(2)}^\top \tilde{F}_2^*}{NT_2^*}, \\
&= V_{NT,1}^{*-1} \left(\frac{\tilde{F}_1^{*\top} G_r^*}{T_1^*} \frac{\Lambda_1^\top e_{(2)}^\top F_{22}^* H_2^*}{NT_2^*} + \frac{\tilde{F}_1^{*\top} G_r^*}{T_1^*} \frac{\Lambda_1^\top e_{(2)}^\top}{N\sqrt{T_2^*}} \frac{(\tilde{F}_2^* - F_{22}^* H_2^*)}{\sqrt{T_2^*}} \right. \\
&\qquad \left. + \frac{\tilde{F}_1^{*\top} G_p^*}{T_1^*} \frac{W^\top e_{(2)}^\top F_{22}^* H_2^*}{N^\alpha T_2^*} \frac{N^\alpha}{N} + \frac{\tilde{F}_1^{*\top} G_p^*}{T_1^*} \frac{W^\top e_{(2)}^\top}{\sqrt{N^\alpha T_2^*}} \frac{(\tilde{F}_2^* - F_{22}^* H_2^*)}{\sqrt{T_2^*}} \frac{\sqrt{N^\alpha}}{N} \right), \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) \frac{N^\alpha}{N} + O_p\left(\frac{1}{\delta_{NT}^2}\right) \frac{\sqrt{N^\alpha}}{N} \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right), \\
Z.ix &= V_{NT,1}^{*-1} \frac{\tilde{F}_1^{*\top}}{\sqrt{T_1^*}} \frac{e_{(1)} \Xi}{N\sqrt{T_1^*}} \frac{G_2^{*\top} \tilde{F}_2^*}{T_2^*} \\
&= V_{NT,1}^{*-1} \left(\frac{\left(\tilde{F}_1 - G_1^* H_{1,\Xi}^* \right)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} \Lambda_1}{N\sqrt{T_1^*}} \frac{G_r^{*\top} \tilde{F}_2^*}{T_2^*} + \frac{\left(G_1^* H_{1,\Xi}^* \right)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} \Lambda_1}{N\sqrt{T_1^*}} \frac{G_r^{*\top} \tilde{F}_2^*}{T_2^*} \right. \\
&\qquad \left. + \frac{\left(\tilde{F}_1 - G_1^* H_{1,\Xi}^* \right)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} W}{N^\alpha \sqrt{T_1^*}} \frac{G_r^{*\top} \tilde{F}_2^*}{T_2^*} \frac{N^\alpha}{N} + \frac{\left(G_1^* H_{1,\Xi}^* \right)^\top}{\sqrt{T_1^*}} \frac{e_{(1)} W}{\sqrt{N^\alpha T_1^*}} \frac{G_r^{*\top} \tilde{F}_2^*}{T_2^*} \frac{\sqrt{N^\alpha}}{N} \right) \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) \frac{N^\alpha}{N} + O_p\left(\frac{1}{\delta_{NT}^2}\right) \frac{\sqrt{N^\alpha}}{N} \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right), \\
Z.x &= O_p\left(\frac{1}{\delta_{NT}^2}\right),
\end{aligned}
$$

where the negligibility for $Z.x$ can be proven in a similar way.

The remaining $Z.vi$ is the leading term, whose behaviour depends on α. When $\alpha < 1$, we have

$$
\begin{aligned}
Z.vi = V_{NT,1}^{*-1} &\left(\frac{\tilde{F}_1^{*\top} G_r^*}{T_1^*} \frac{\Lambda_1^\top \Lambda_1 Z}{N} \frac{F_{22}^{*\top} \tilde{F}_2^*}{T_2^*} + \frac{\tilde{F}_1^{*\top} G_p^*}{T_1^*} \frac{W^\top \Lambda_1 Z}{N} \frac{F_{22}^{*\top} \tilde{F}_2^*}{T_2^*} \right. \\
&\quad \left. + \frac{\tilde{F}_1^{*\top} G_r^*}{T_1^*} \frac{\Lambda_1^\top W}{N} \frac{F_{22}^{*\top} \tilde{F}_2^*}{T_2^*} + \frac{\tilde{F}_1^{*\top} G_p^*}{T_1^*} \frac{W^\top W}{N} \frac{F_{22}^{*\top} \tilde{F}_2^*}{T_2^*} \right)
\end{aligned}
$$

$$= H_{1,r}^{*\top} Z H_2^{*-\top} + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) + O_p\left(\frac{N^\alpha}{N}\right),$$

where the last line uses $\frac{1}{T} F_{22}^{*\top} \tilde{F}_2^* = H_2^{*-\top}$, because

$$\begin{aligned} F_{22}^* H_2^* + \tilde{F}_2^* - F_{22}^* H_2^* &= \tilde{F}_2^* \\ \frac{1}{T_2^*} \tilde{F}_2^{*\top} F_{22}^* H_2^* + \frac{1}{T_2^*} \tilde{F}_2^{*\top} \left(\tilde{F}_2^* - F_{22}^* H_2^*\right) &= I_r \\ \frac{1}{T_2^*} F_{22}^{*\top} \tilde{F}_2^* &= H_2^{*-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right). \end{aligned}$$

For the case of $\alpha = 1$, the leading term $Z.vi.I$ can instead be characterised using $H_{1,\Xi}^*$

$$\begin{aligned} Z.vi.I &= H_{1,\Xi}^{*\top} \frac{G_2^{*\top} \tilde{F}_2^*}{T_2^*} \\ &= H_{1,\Xi}^{*\top} \begin{bmatrix} Z \\ I_r \end{bmatrix} \frac{F_{22}^{*\top} \tilde{F}_2}{T_2^*} \\ &= H_{1,\Xi}^{*\top} \begin{bmatrix} Z \\ I_r \end{bmatrix} H_2^{*-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right), \end{aligned}$$

which uses the fact that $G_2^* = \begin{bmatrix} F_{22}^* Z^\top & F_{22}^* \end{bmatrix} = \begin{bmatrix} Z^\top & I_r \end{bmatrix} F_{22}^*$. Collecting the dominating terms for the two cases yields the consistency result for $\tilde{Z}^*$.

For the $\alpha < 1$, the mean square consistency for $\tilde{F}_2^* \tilde{Z}^{*\top}$ follows as

$$\begin{aligned} \frac{1}{\sqrt{T}} \left(\tilde{F}_2^* \tilde{Z}^{*\top} - F_{22}^* Z^\top H_{1,r}^{*\top}\right) &= \frac{1}{\sqrt{T}} \tilde{F}_2^* \left(\tilde{Z}^{*\top} - H_2^{*-1} Z^\top H_{1,r}\right) + \frac{1}{\sqrt{T}} \left(\tilde{F}_2^* - F_{22}^* H_2^*\right) H_2^{*-1} Z^\top H_{1,r} \\ &= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right) + O_p\left(\frac{1}{\delta_{NT}}\right). \end{aligned}$$

For $\alpha = 1$ we have

$$\begin{aligned} \frac{1}{\sqrt{T}} \left(\tilde{F}_2^* \tilde{Z}^{*\top} - F_{22}^* \begin{bmatrix} Z^\top & I_r \end{bmatrix} H_{1,\Xi}^*\right) &= \frac{1}{\sqrt{T}} \tilde{F}_2^* \left(\tilde{Z}^{*\top} - H_2^{*-1} \begin{bmatrix} Z^\top & I_r \end{bmatrix} H_{1,\Xi}\right) \\ &\quad + \frac{1}{\sqrt{T}} \left(\tilde{F}_2^* - F_{22}^* H_2^*\right) H_2^{*-1} \begin{bmatrix} Z^\top & I_r \end{bmatrix} H_{1,\Xi} \\ \frac{1}{\sqrt{T}} \left(\tilde{F}_2^* \tilde{Z}^{*\top} - G_2^* H_{1,\Xi}^*\right) &= O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{1}{\delta_{NT}}\right). \end{aligned}$$

For both cases, taking the squared norm of both sides yields the result.

Lemmas 10 (c) and 10 (d) can be proven in a similar way to the pseudo-factors for the cases $\alpha < 1$ and $\alpha = 1$. ■

B Bias Variance Trade-off Proofs

B.1 Out-of-sample Asymptotic Expansions

B.1.1 Pseudo-factors

In the case where the pseudo-factors $\tilde{F}_P$ are used, the regressor matrix is $\tilde{C}_P = [\tilde{c}_{P,t-1}, \dots, \tilde{c}_{P,T-h}]^\top$. Because $\tilde{F}_P$ is an estimate of $G_r H_G$, we define $c_{G_r,t} = [g_{r,t}^\top, (1, y_t)^\top]^\top$, its matrix counterpart $C_{G_r} = [c_{G_r,t-h}, \dots, c_{G_r,T-h}]^\top$, and the corresponding rotation matrix $H_P = diag(H_G, I)$, which rotates the columns of the factors but leaves the observed regressors unchanged.

The least squares estimate of the forecast coefficient and resulting forecast $\tilde{\mu}_{P,T}$ is then

$$
\begin{aligned}
\widehat{\theta}_P &= \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top Y, \\
\tilde{\mu}_{P,T} &= \tilde{c}_{P,T}^\top \widehat{\theta}_P.
\end{aligned}
$$

The out-of-sample forecast error is then

$$
c_T^\top \theta - \tilde{c}_{P,T}^\top \widehat{\theta}_P = c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top C\theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top \eta. \tag{B.1}
$$

The squared norm of this is

$$
\begin{aligned}
\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \widehat{\theta} \right\|^2 =& \left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top C\theta \right\|^2 + \left\| \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top \eta \right\|^2 \\
&+ 2\eta^\top \tilde{C}_P^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_{P,T} \tilde{C}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top C\theta \\
=& bias^2 + var + cross, .
\end{aligned} \tag{B.2}
$$

Our strategy is to treat the cross term as an asymptotically normally distributed, and therefore mean zero random variable to be ignored. Note that this is OK, because the asymptotic order of the cross term must be less than the bias term, which is dominating. Usually the cross term in cancelled out, in the case of analysing the in sample residuals. The variance term is $O_p(1)$, and is mainly of interest in comparing the split-sample factors.

The bias term can be expanded as

$$
\begin{aligned}
& c_T^\top \theta - \tilde{C}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top C\theta \\
=& (c_T^\top H_P - \tilde{c}_{P,T}^\top) H_P^{-1}\theta - \tilde{c}_{P,t}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top C + \tilde{c}_{P,T}^\top H_P^{-1}\theta \\
=& [f_T^\top H_G - \tilde{f}_{P,T}^\top, 0] \begin{bmatrix} H_G^{-1}\beta(L) \\ \delta \end{bmatrix} + \tilde{c}_{P,T}^\top \left[I - \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top C H_P\right] H_P^{-1}\theta \\
=& \left(f_T^\top H_G - \tilde{f}_{P,T}^\top\right) H_G^{-1}\beta + \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top \left(\tilde{C}_P - C H_P\right) H_P^{-1}\theta \\
=& -\left(\tilde{f}_{P,T}^\top - f_T^\top H_G\right) H_G^{-1}\beta + \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top \left(\tilde{C}_P - C H_P\right) H_P^{-1}\theta.
\end{aligned}
$$

By the expansion in the proof of Lemma 1 (a) (replacing G_r with $\tilde{C}_P$), we have

$$
\frac{(\tilde{F}_P - G_r H_G, 0)^\top \tilde{C}_P}{T} = \begin{bmatrix} V_{NT}^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \frac{G_p^\top \tilde{C}_P}{T} \\ 0 \end{bmatrix} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
$$

Consequently, $\frac{(\tilde{F}_P - F H_G, 0)^\top \tilde{C}_P}{T}$ follows by adding and subtracting

$$
\begin{aligned}
\frac{(\tilde{F}_P - F H_G, 0)^\top \tilde{C}_P}{T} &= \begin{bmatrix} \frac{(G_r H_G - F H_G)^\top \tilde{C}_P}{T} + V_{NT}^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \frac{G_p^\top \tilde{C}_P}{T} \\ 0 \end{bmatrix} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) \\
&= \begin{bmatrix} \frac{-H_G^\top (I-Z) G_p^\top \tilde{C}_P}{T} + V_{NT}^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \frac{G_p^\top \tilde{C}_P}{T} \\ 0 \end{bmatrix} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right),
\end{aligned}
$$

which expresses this substitution in terms of the rotational break and shift break.

Substituting these expansions, we have

$$
\begin{aligned}
& c_T^\top \theta - \tilde{c}_T^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top C\theta \\
=& \left(-\left(V_{NT}^{-1} \frac{\tilde{F}_P^\top G_r}{T} \frac{\Lambda_1^\top e_T}{N} + V_{NT}^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} f_T - H_G^\top (I-Z) f_T\right)\right)^\top H_G^{-1}\beta(L) \\
& + \left(\begin{bmatrix} \frac{-H_G^\top (I-Z) G_p^\top \tilde{C}_P}{T} + V_{NT}^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \frac{G_p^\top \tilde{C}_P}{T} \\ 0 \end{bmatrix} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\right)^\top H_P^{-1}\theta \\
=& \left(-\left(V_{NT}^{-1} \frac{\tilde{F}_P^\top G_r}{T} \frac{\Lambda_1^\top e_T}{N} + V_{NT}^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} f_T - H_G^\top (I-Z) f_T\right) - H_G^\top (I-Z) \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T}\right. \\
& \left. + V_{NT}^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \frac{G_P^\top \tilde{F}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\right)^\top H_G^{-1}\beta
\end{aligned}
$$

$$
\begin{aligned}
&= \Bigg(H_G^\top (I-Z) \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \\
&\qquad - V_{NT}^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \Bigg)^\top H_G^{-1} \beta(L) \\
&\quad - \frac{e_T^\top \Lambda_1}{N} \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \beta(L) + O_p\left(\frac{1}{\delta_{NT}^2} \right) + O_p\left(\frac{\sqrt{N^\alpha}}{N} \right), \\
&= \Bigg[\left((I-Z) - \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T} \right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \right) \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \\
&\qquad - \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \frac{\Lambda_1^\top e_T}{N} \Bigg]^\top \beta(L) + O_p\left(\frac{1}{\delta_{NT}^2} \right) + O_p\left(\frac{\sqrt{N^\alpha}}{N} \right),
\end{aligned}
$$

where the last two lines use the definition of H_G^{-1}. This expresses the bias in terms of the rotational break $(I-Z)$, shift break $(W^\top W)$, and inherent estimation error in the factors.

Hence, the squared bias can be expressed as

$$
\begin{aligned}
&\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{c}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C\theta \right\|^2 \\
&= \Bigg\| \Bigg[\left((I-Z) - \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T} \right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \right) \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \\
&\qquad - \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \frac{\Lambda_1^\top e_T}{N} \Bigg]^\top \beta(L) \Bigg\|^2 \\
&\quad + O_p\left(N^{\alpha/2+\nu-2} \right) + O_p\left(N^{3\alpha/2} \right) + O_p\left(N^{\alpha/2-3/2} \right).
\end{aligned} \tag{B.3}
$$

where the remainder terms follow from the cross terms between the rotational break, shift break, inherent bias, and the above remainder terms.

Remark. *The two bias terms may cancel each other out if α and ν are equal, and the two bias terms $(I-Z)$ and $-\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1}\frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N}$ have opposite signs. In finite sample, depending on the DGP, we can expect this bias cancellation to occur for similar values of α and ν when both are ≥ 0.5 (i.e. when the bias terms are large enough to affect forecasting performance).*

The variance term can be written as

$$
\frac{1}{T} \left[\frac{\eta^\top \tilde{C}_P}{\sqrt{T}} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \tilde{c}_{P,T}^\top \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \frac{\tilde{C}_P^\top \eta}{\sqrt{T}} \right]
$$

$$
\begin{aligned}
=&\frac{1}{T}\left[\frac{\eta^\top \tilde{C}_P}{\sqrt{T}}\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1}\tilde{c}_{P,T}\tilde{c}_{P,T}^\top\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1}\left[\frac{(C_{G_r}H_P)^\top\eta}{\sqrt{T}}+\frac{(\tilde{C}_P-C_{G_r}H_P)^\top\eta}{\sqrt{T}}\right]\right]\\
=&\frac{1}{T}\left[\frac{\eta^\top \tilde{C}_P}{\sqrt{T}}\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1}\tilde{c}_{P,T}\tilde{c}_{P,T}\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1}\frac{(C_{G_r}H_P)^\top\eta}{\sqrt{T}}+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)+O_p\left(\frac{N^\alpha}{N}\right)\right]\\
=&\frac{1}{T}\left[\frac{\eta^\top C_{G_r}H_P}{\sqrt{T}}\left(\frac{(H_PC_{G_r})^\top C_{G_r}H_P}{T}\right)^{-1}\tilde{c}_{P,T}\tilde{c}_{P,T}^\top\left(\frac{(H_PC_{G_r})^\top C_{G_r}H_P}{T}\right)^{-1}\frac{(C_{G_r}H_P)^\top\eta}{\sqrt{T}}\right.\\
&\left.+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)+O_p\left(\frac{N^\alpha}{N}\right)\right]\\
=&\frac{1}{T}\left[\frac{\eta^\top C_{G_r}}{\sqrt{T}}\left(\frac{(H_PC_{G_r})^\top C_{G_r}H_P}{T}\right)^{-1}H_P^\top C_{G_r,T}\tilde{c}_{P,T}^\top\left(\frac{(H_PC_{G_r})^\top C_{G_r}H_P}{T}\right)^{-1}\frac{(C_{G_r}H_P)^\top\eta}{\sqrt{T}}\right.\\
&+\frac{\eta^\top C_{G_r}}{\sqrt{T}}\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1}H_P^{-\top}\left(\tilde{c}_{P,T}-H_P^\top C_{G_r,T}\right)\tilde{c}_{P,T}^\top H_P^{-1}\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1}\frac{C_{G_r}^\top\eta}{\sqrt{T}}\\
&\left.+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)+O_p\left(\frac{N^\alpha}{N}\right)\right],\\
=&\frac{1}{T}\left[\frac{\eta^\top C_{G_r}}{\sqrt{T}}\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1}c_{G_{r,T}}^\top c_{G_{r,T}}\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1}\frac{C_{G_r}^\top\eta}{\sqrt{T}},\right.\\
&\left.+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)+O_p\left(\frac{N^\alpha}{N}\right)\right].
\end{aligned}
$$

B.1.2 Split-sample Factors

When the split-sample factors are used, this is algebraically equivalent to using only the post-break data. That is, the post-break observations of Y, denoted as Y_2, are fitted using the regressor matrix $\tilde{C}_2 = [\tilde{c}_{2,t-1},\ldots,\tilde{c}_{2,T-h}]^\top$, where $\tilde{c}_{2,t} = \left[\tilde{f}_{2,t}^\top,(1,y_t)^\top\right]^\top$. Because $\tilde{F}_2$ is an estimate of F_2H_2, we define $c_{2,t} = \left[f_t^\top,(1,y_t)^\top\right]^\top$, its matrix counterpart $C_2 = [c_{2,t-h},\ldots,c_{2,T-h}]^\top$, and its corresponding rotation matrix $H_S = diag(H_2, I)$, which rotates the columns of the factor but leaves the observed regressors unchanged.

The least squares estimate of the forecast coefficient and resulting forecast $\tilde{\mu}_{S,T}$ is then

$$
\begin{aligned}
\widehat{\theta}_S &= \left(\tilde{C}_2^\top\tilde{C}_2\right)^{-1}\tilde{C}_2^\top Y,\\
\tilde{\mu}_{S,T} &= \tilde{c}_{S,T}^\top\widehat{\theta}_S.
\end{aligned}
$$

The out-of-sample forecast error is then

$$
c_T^\top\theta-\tilde{c}_{S,T}^\top\widehat{\theta}_S = c_T^\top\theta-\tilde{c}_{S,T}^\top\left(\tilde{C}_2^\top\tilde{C}_2\right)^{-1}\tilde{C}_2^\top C_2\theta-\tilde{c}_{S,T}^\top\left(\tilde{C}_2^\top\tilde{C}_2\right)^{-1}\tilde{C}_2^\top\eta_{(2)}. \tag{B.4}
$$

The out-of-sample forecast fit is then

$$\begin{aligned}
&\left\| c_T^\top \theta - \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{C}_2^\top Y_2 \right\|^2 \\
=&\left\| c_T^\top \theta - \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{C}_2^\top C_2 \theta \right\|^2 + \left\| \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{C}_2^\top \eta_{(2)} \right\|^2 \\
&+ 2\eta_{(2)}^\top \tilde{C}_2 \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{c}_{2,T} \left(c_T^\top \theta - \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{C}_2^\top C_2 \theta \right) \\
=& bias^2 + var + cross.
\end{aligned} \tag{B.5}$$

To analyse the bias term, $\tilde{c}_{2,t} - H_S^\top c_t = \begin{bmatrix} \tilde{f}_{2,t} - H_2^\top f_t \\ 0 \end{bmatrix}$, where the first r rows follow the expansion in Lemma 3 (b). Therefore, for the bias term, we have

$$\begin{aligned}
& c_T^\top \theta - \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{C}_2^\top C_2 \theta \\
=& \left(H_S^\top c_T - H_S^\top \frac{C_2^\top \tilde{C}_2}{T_2} \left(\frac{\tilde{C}_2^\top \tilde{C}_2}{T_2} \right)^{-1} \tilde{c}_{2,T} \right)^\top H_S^{-1} \theta \\
=& \begin{bmatrix} f_T^\top H_2 - \tilde{f}_{2,t}^\top & 0 \end{bmatrix} \begin{bmatrix} H_2^{-1} \beta(L) \\ \delta \end{bmatrix} + \left(\tilde{c}_{2,T}^\top \left(\frac{\tilde{C}_2^\top \tilde{C}_2}{T_2} \right)^{-1} \frac{\tilde{C}_2^\top \left(\tilde{C}_2 - C_2 H_2 \right)}{T_2} \right) H_P^{-1} \theta \\
=& \left(-V_{NT,2}^{-1} \frac{\tilde{F}_2^\top F_2}{T_2} \frac{\Lambda_2^\top e_T}{N} \right)^\top H_2^{-1} \beta + O_p\left(\frac{1}{\delta_{NT}^2} \right) \\
=& \frac{-e_T^\top \Lambda_2}{N} \left(\frac{\Lambda_2^\top \Lambda_2}{N} \right)^{-1} \beta(L) + O_p\left(\frac{1}{\delta_{NT}^2} \right),
\end{aligned}$$

where the last line uses the definition of H_2. This implies that the squared bias is

$$\left\| c_T^\top \theta - \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{C}_2^\top C_2 \theta \right\|^2 = \left\| \frac{-e_T^\top \Lambda_2}{N} \left(\frac{\Lambda_2^\top \Lambda_2}{N} \right)^{-1} \beta(L) \right\|^2 + O_p\left(\frac{1}{\delta_{NT}^2} \right).$$

For the variance, we have

$$\begin{aligned}
& \eta_{(2)}^\top \tilde{C}_2 \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{c}_{2,T} \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{C}_2^\top \eta_{(2)} \\
=& \frac{1}{T_2} \left[\frac{\eta_{(2)}^\top \tilde{C}_2}{\sqrt{T_2}} \left(\frac{\tilde{C}_2^\top \tilde{C}_2}{T_2} \right)^{-1} \tilde{c}_{2,T} \tilde{c}_{2,T}^\top \left(\frac{\tilde{C}_2^\top \tilde{C}_2}{T_2} \right)^{-1} \frac{\tilde{C}_2^\top \eta_{(2)}}{\sqrt{T_2}} \right] \\
=& \frac{1}{T_2} \left[\frac{\eta_{(2)}^\top C_2}{\sqrt{T_2}} \left(\frac{C_2^\top C_2}{T_2} \right)^{-1} c_T c_T^\top \left(\frac{C_2^\top C_2}{T_2} \right)^{-1} \frac{C_2^\top \eta_{(2)}}{\sqrt{T_2}} + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2} \right) \right].
\end{aligned} \tag{B.6}$$

B.1.3 Rotated Factors

When the rotated factors $\tilde{F}_R$ are used, the regressor matrix is $\tilde{C}_R = [\tilde{c}_{R,t-1}, \ldots, \tilde{c}_{R,T-h}]^\top$. Because the rotated factors $\tilde{F}_R$ are an estimate of $G_r H_1$, we use $c_{G_r,t} = \left[g_{r,t}^\top, (1, y_t)^\top\right]^\top$ and its matrix counterpart C_{G_r} as with the pseudo-factors, and the corresponding rotation matrix $H_R = diag(H_1, I)$, which rotated the columns of the factors but leaves the observed regressors unchanged.

The least squares estimate of the forecast coefficient and resulting forecast $\tilde{\mu}_{R,T}$ is then

$$\begin{aligned}\widehat{\theta}_R &= \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R^\top Y,\\ \tilde{\mu}_{R,T} &= \tilde{c}_{R,T}^\top \widehat{\theta}_R.\end{aligned}$$

The out-of-sample forecast error is then

$$c_T^\top \theta - \tilde{c}_{R,T}^\top \widehat{\theta}_R = c_T^\top \theta - \tilde{c}_{R,T}^\top \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R^\top C\theta - \tilde{c}_{R,T}^\top \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R^\top \eta. \tag{B.7}$$

The squared norm of this is

$$\begin{aligned}&\left\| c_T^\top \theta - \tilde{c}_{R,T}^\top \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R^\top Y \right\|^2\\ =&\left\| c_T^\top \beta - (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R^\top C\theta \right\|^2 + \left\| (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R^\top \eta \right\|^2\\ &\quad + 2\eta^\top \tilde{C}_R \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{c}_{R,T} \left(c_T^\top \theta - (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R C\theta\right)\\ =&bias^2 + var + cross.\end{aligned}$$

For the bias term, we have

$$\begin{aligned}&c_T^\top \theta - (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R^\top C\theta\\ =&\left(H_R^\top c_T - \tilde{c}_{R,T} + \left(\tilde{C}_R - CH_R\right)^\top \tilde{C}_R \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{c}_{R,T}\right)^\top H_R^{-1}\theta\\ =&\left[\tilde{f}_{R,T}^\top - f_T^\top H_1, 0\right] \begin{bmatrix} H_1^{-1}\beta(L) \\ \delta \end{bmatrix} + \left[\left(\tilde{C}_R - CH_R\right)^\top \tilde{C}_R \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{c}_{R,T}\right]^\top H_R^{-1}\theta.\end{aligned}$$

This requires expressions for $\tilde{Z}\tilde{f}_{2,T} - H_1^\top Z f_T$ and $\frac{1}{T}(\tilde{C}_R - CH_R)^\top \tilde{C}_R$. Using the consistency of $\tilde{Z}$ and the expansion for $\tilde{f}_{2,T} - H_2^\top f_T$, it follows that

$$\begin{aligned}\tilde{Z}\tilde{f}_{2,T} - H_1^\top Z f_T &= H_1^\top Z H_2^{-\top} V_{NT,2}^{-1} \frac{\tilde{F}_2^\top F_2}{T} \frac{\Lambda_2^\top e_T}{N} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) \\ &= H_1^\top Z \left(\frac{\Lambda_2^\top \Lambda_2}{N}\right)^{-1} \frac{\Lambda_2^\top e_T}{N} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \end{aligned} \tag{B.8}$$

where the second line follows by the definition of $H_2^{-\top}$.

Next, we analyse $\frac{1}{T}\left(\tilde{C}_R - CH_1\right)^\top \tilde{C}_R$.

$$\begin{aligned} &\frac{1}{T}\left(\tilde{C}_R - CH_R\right)^\top \tilde{C}_R \\ &= \begin{bmatrix} \frac{1}{T}\left(\tilde{F}_R - FH_1\right)^\top \tilde{C}_R \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{T}\left(\tilde{F}_1 - F_1 H_1\right)^\top \tilde{C}_{R,1} + \frac{1}{T}\left(\tilde{F}_2\tilde{Z}^\top - F_2 H_1\right)^\top \tilde{C}_{R,2} \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{T}\left(\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1\right)^\top \tilde{C}_{R,2} - \frac{1}{T} H_1^\top (I-Z) F_2^\top \tilde{C}_{R,2} \\ 0 \end{bmatrix} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) \\ &= \begin{bmatrix} -H_1^\top (I-Z) \frac{F_2^\top \tilde{C}_{R,2}}{T} \\ 0 \end{bmatrix} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right). \end{aligned}$$

Therefore, the expression for the bias can be expressed as

$$\begin{aligned} &c_T^\top \theta - (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R^\top C\theta \\ &= \left(H_1^\top (I-Z) f_T - H_1^\top Z \left(\frac{\Lambda_2^\top \Lambda_2}{N}\right)^{-1} \frac{\Lambda_2^\top e_T}{N} - H_1^\top (I-Z) \frac{\tilde{F}_2^\top \tilde{C}_{R,2}}{T} \left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1} \tilde{c}_{R,T}\right)^\top H_1^{-1}\beta(L) \\ &\quad + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) \\ &= \left(H_1^\top (I-Z)\left(f_T - \frac{\tilde{F}_2^\top \tilde{C}_{R,2}}{T}\left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1} \tilde{c}_{R,T}\right) - H_1^\top Z \left(\frac{\Lambda_2^\top \Lambda_2}{N}\right)^{-1} \frac{\Lambda_2^\top e_T}{N}\right)^\top H_1^{-1}\beta(L) \\ &\quad + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right). \end{aligned}$$

Note that for $\alpha < 1$

$$
\begin{aligned}
Z\left(\frac{\Lambda_2^\top \Lambda_2}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N} &= Z\left(\left(\frac{Z^\top \Lambda_1^\top \Lambda_1 Z}{N}\right)^{-1} + O_p\left(\frac{N^\alpha}{N}\right)\right)\left(\frac{Z^\top \Lambda_1^\top e_T}{N} + \frac{W^\top e_T}{N}\right) \\
&= \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N} + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}
$$

The squared bias is therefore

$$
\begin{aligned}
&\left\| c_T^\top \theta - (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R^\top C\theta \right\|^2 \\
=&\left\| \left((I-Z)\left(f_T - \frac{\tilde{F}_2^\top \tilde{C}_{R,2}}{T}\left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\right) - \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N}\right)^\top \beta + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\right\|^2 \\
=&\left\| \left((I-Z)\left(f_T - \frac{\tilde{F}_2^\top \tilde{C}_{R,2}}{T}\left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\right) - \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N}\right)^\top \beta\right\|^2 \\
&+ O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N^2}\right) + O_p\left(\frac{N^{\alpha/2+\nu}}{N^2}\right) + O_p\left(\frac{N^\nu}{\delta_{NT}^2 N}\right).
\end{aligned}
$$

Thus, the rotated factors are much more robust to shift type breaks.

The variance term can be written as

$$
\begin{aligned}
&\frac{1}{T}\left[\frac{\eta^\top \tilde{C}_R}{\sqrt{T}}\left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\tilde{c}_{R,T}^\top\left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1}\frac{\tilde{C}_R^\top \eta}{\sqrt{T}}\right] \\
=&\frac{1}{T}\left[\frac{\eta^\top \tilde{C}_R}{\sqrt{T}}\left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\tilde{c}_{R,T}^\top\left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1}\left[\frac{(C_{G_r}H_R)^\top \eta}{\sqrt{T}} + \frac{\left(\tilde{C}_R - C_{G_r}H_R\right)^\top \eta}{\sqrt{T}}\right]\right] \\
=&\frac{1}{T}\left[\frac{\eta^\top \tilde{C}_R}{\sqrt{T}}\left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\tilde{c}_{R,T}^\top\left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1}\frac{(C_{G_r}H_R)^\top \eta}{\sqrt{T}} + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\right] \\
=&\frac{1}{T}\left[\frac{\eta^\top (C_{G_r}H_R)}{\sqrt{T}}\left(\frac{(C_{G_r}H_R)^\top (C_{G_r}H_R)}{T}\right)^{-1}\tilde{c}_{R,T}\tilde{c}_{R,T}^\top\left(\frac{(C_{G_r}H_R)^\top (C_{G_r}H_R)}{T}\right)^{-1}\frac{(C_{G_r}H_R)^\top \eta}{\sqrt{T}}\right. \\
&\qquad\left. + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\right] \\
=&\frac{1}{T}\left[\frac{\eta^\top (C_{G_r}H_R)}{\sqrt{T}}\left(\frac{(C_{G_r}H_R)^\top (C_{G_r}H_R)}{T}\right)^{-1}H_R^\top c_{G_r,T}\tilde{c}_{R,T}^\top\left(\frac{(C_{G_r}H_R)^\top (C_{G_r}H_R)}{T}\right)^{-1}\frac{(C_{G_r}H_R)^\top \eta}{\sqrt{T}}\right. \\
&\qquad + \frac{\eta^\top (C_{G_r}H_R)}{\sqrt{T}}\left(\frac{(C_{G_r}H_R)^\top (C_{G_r}H_R)}{T}\right)^{-1}\left(\tilde{c}_{R,T} - H_1^\top c_{G_r,T}\right) \\
&\qquad \times \tilde{c}_{R,T}^\top\left(\frac{(C_{G_r}H_R)^\top (C_{G_r}H_R)}{T}\right)^{-1}\frac{(C_{G_r}H_R)^\top \eta}{\sqrt{T}}
\end{aligned}
$$

$$
+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)+O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\Bigg]
$$

$$
=\frac{1}{T}\left[\frac{\eta^\top C_{G_r}}{\sqrt{T}}\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1}c_{G_r,T}c_{G_r,T}^\top\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1}\frac{C_{G_r}^\top\eta}{\sqrt{T}}\right.
$$

$$
\left.+O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)+O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\right].
$$

B.2 Small Shift Break $\alpha < 0.5$

Proof of Theorem 2 (a) - Asymptotic Equivalence of Psuedo and Rotated Forecasts.

We show that the pseudo-factors and rotated factors produce asymptotically identical forecasts for $\alpha < 1/2$. Taking the difference between the rotated and pseudo-factor forecasts, we have

$$
\begin{aligned}
&\tilde{c}_{R,T}^\top\widehat{\theta}_R-\tilde{c}_{P,T}^\top\widehat{\theta}_P\\
=&\tilde{c}_{R,T}^\top\left(\tilde{C}_R^\top\tilde{C}_P\right)^{-1}\tilde{C}_R^\top C\theta-\tilde{c}_{P,T}^\top\left(\tilde{C}_P^\top\tilde{C}_P\right)^{-1}\tilde{C}_P^\top C\theta\\
=&\tilde{c}_{R,T}^\top\left(C_R^\top C_R\right)^{-1}C_R^\top C\theta-\tilde{c}_{P,T}^\top\left(C_P^\top C_P\right)^{-1}C_P^\top C\theta+\tilde{c}_{R,T}^\top\left[\left(\tilde{C}_R^\top\tilde{C}_R\right)^{-1}\tilde{C}_R^\top C-\left(C_R^\top C_R\right)^{-1}C_R^\top C\right]\theta\\
&-\tilde{c}_{P,T}^\top\left[\left(\tilde{C}_P^\top\tilde{C}_P\right)^{-1}\tilde{C}_P^\top C-\left(C_P^\top C_P\right)^{-1}C_P^\top C\right]\theta\\
=&\tilde{c}_{R,T}^\top\left(C_R^\top C_R\right)^{-1}C_R^\top C\theta-\tilde{c}_{P,T}^\top\left(C_P^\top C_P\right)^{-1}C_P^\top C\theta+O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{N^\alpha}{N}\right)+O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\\
=&\left(\tilde{c}_{R,T}^\top H_R^{-1}-\tilde{c}_{P,T}H_P^{-1}\right)\left(C_{G_r}^\top C_{G_r}\right)^{-1}C_{G_r}^\top C\theta+O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{N^\alpha}{N}\right)+O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}
\tag{B.9}
$$

Next, we check the term $\left(\tilde{c}_{R,T}^\top H_R^{-1}-\tilde{c}_{P,T}H_P^{-1}\right)$.

Based on the expansions of $\tilde{f}_{R,T}$ and $\tilde{f}_{P,T}$, the first r entries of the row vector $\left(\tilde{c}_{R,T}^\top H_R^{-1}-\tilde{c}_{P,T}H_P^{-1}\right)$ are

$$
\begin{aligned}
&\left(\tilde{f}_{R,T}^\top H_1^{-1}-f_T^\top Z\right)-\left(\tilde{f}_{P,T}^\top H_1^{-1}-f_T^\top Z\right)\\
=&\tilde{f}_{2,T}^\top\left(O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\right)+\frac{e_T^\top\Lambda_2}{N}\left(\frac{\Lambda_2^\top\Lambda_2}{N}\right)^{-1}Z^\top-\frac{e_T^\top\Lambda_1}{N}\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\\
&-f_T^\top\frac{W^\top W}{N}\frac{G_p^\top\tilde{F}_P}{T}\left(\frac{G_r^\top\tilde{F}_P}{T}\right)^{-1}\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}+O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{\sqrt{N^\alpha}}{N}\right)\\
=&-f_T^\top\frac{W^\top W}{N}\frac{G_p^\top\tilde{F}_P}{T}\left(\frac{G_r^\top\tilde{F}_P}{T}\right)^{-1}\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}+O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}
\tag{B.10}
$$

Thus, combined with the fact that the remaining rows of $\left(\tilde{c}_{R,T}^\top H_R^{-1}-\tilde{c}_{P,T}H_P^{-1}\right)$ are 0, we therefore have for $\alpha < 1/2$:

$$\tilde{c}_{R,T}^{\top}\left(\tilde{C}_R^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_R^{\top}C\theta - \tilde{c}_{P,T}^{\top}\left(\tilde{C}_P^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_P^{\top}C\theta = o_p\left(N^{-1/2}\right). \tag{B.11}$$

Next, we focus on the difference between the variance terms of the pseudo- and rotated factors. Similarly, the variance terms of the out-of-sample prediction errors can be written as

$$\begin{aligned}
&\tilde{c}_{R,T}^{\top}\left(\tilde{C}_R^{\top}\tilde{C}_R\right)^{-1}\tilde{C}_R^{\top}\eta - \tilde{c}_{P,T}^{\top}\left(\tilde{C}_P^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_P^{\top}\eta \\
=&\tilde{c}_{R,T}^{\top}\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta - \tilde{c}_{P,T}^{\top}\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta + \tilde{c}_{R,T}^{\top}\left[\left(\tilde{C}_R^{\top}\tilde{C}_R\right)^{-1}\tilde{C}_R^{\top}\eta - \left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta\right] \\
&- \tilde{c}_{P,T}^{\top}\left[\left(\tilde{C}_P^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_P^{\top}\eta - \left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta\right] \\
=&\tilde{c}_{R,T}^{\top}\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta - \tilde{c}_{P,T}^{\top}\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^{\alpha}}{N}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right) \\
=&\left(\tilde{c}_{R,T}^{\top}H_R^{-1} - \tilde{c}_{P,T}^{\top}H_P^{-1}\right)\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^{\alpha}}{N}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right) \\
=&o_p\left(N^{-1/2}\right),
\end{aligned} \tag{B.12}$$

for $\alpha < 1/2$, where we use that face that both $\tilde{c}_{R,T}^{\top}\left(\tilde{C}_R^{\top}\tilde{C}_R\right)^{-1}\tilde{C}_R^{\top}\eta$ and $\tilde{c}_{P,T}^{\top}\left(\tilde{C}_P^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_P^{\top}\eta$ are negligible because

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\tilde{c}_{R,T}^{\top}\left(\frac{\tilde{C}_R^{\top}\tilde{C}_R}{T}\right)^{-1}\frac{\tilde{C}_R^{\top}\eta}{\sqrt{T}} \\
=&\frac{1}{\sqrt{T}}\tilde{c}_{R,T}^{\top}\left(\frac{\tilde{C}_R^{\top}\tilde{C}_R}{T}\right)^{-1}\left[\frac{C_{G_r}^{\top}\eta}{\sqrt{T}} + \frac{\left(\tilde{C}_R - C_{G_r}H_R\right)^{\top}\eta}{\sqrt{T}}\right] \\
=&\frac{1}{\sqrt{T}}\tilde{c}_{R,T}^{\top}\left(\frac{\tilde{C}_R^{\top}\tilde{C}_R}{T}\right)^{-1}\frac{C_{G_r}^{\top}\eta}{\sqrt{T}} + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right) \\
=&\frac{1}{\sqrt{T}}\left[H_R^{\top}c_{G_r,t} + (\tilde{c}_{R,T} - H_R^{\top}c_{G_r,t})\right]^{\top}H_R^{-\top}\left(\frac{C_{G_r}^{\top}C_{G_r}}{T}\right)^{-1}\frac{C_{G_r}^{\top}\eta}{\sqrt{T}} + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right) \\
=&\frac{1}{\sqrt{T}}c_{G_r,t}\left(\frac{C_{G_r}^{\top}C_{G_r}}{T}\right)^{-1}\frac{C_{G_r}^{\top}\eta}{\sqrt{T}} + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right) \\
&= O_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}$$

where the third line uses Lemma 4 (b), the fourth uses Theorem 1 (c), and the fifth line uses Equation (B.8), and

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\tilde{c}_{P,T}^{\top}\left(\frac{\tilde{C}_P^{\top}\tilde{C}_P}{T}\right)^{-1}\frac{\tilde{C}_P^{\top}\eta}{\sqrt{T}}\\
=&\frac{1}{\sqrt{T}}\tilde{c}_{P,T}^{\top}\left(\frac{\tilde{C}_P^{\top}\tilde{C}_P}{T}\right)^{-1}\left[\frac{C_{G_r}^{\top}\eta}{\sqrt{T}}+\frac{\left(\tilde{C}_P-C_{G_r}H_P\right)^{\top}\eta}{\sqrt{T}}\right]\\
=&\frac{1}{\sqrt{T}}\tilde{c}_{R,T}^{\top}\left(\frac{\tilde{C}_R^{\top}\tilde{C}_R}{T}\right)^{-1}\frac{C_{G_r}^{\top}\eta}{\sqrt{T}}+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)+O_p\left(\frac{N^{\alpha}}{N}\right)\\
=&\frac{1}{\sqrt{T}}\left[H_P^{\top}c_{G_r,t}+(\tilde{c}_{P,T}-H_R^{\top}c_{G_r,t})\right]^{\top}H_R^{-\top}\left(\frac{C_{G_r}^{\top}C_{G_r}}{T}\right)^{-1}\frac{C_{G_r}^{\top}\eta}{\sqrt{T}}+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)+O_p\left(\frac{N^{\alpha}}{N}\right)\\
=&\frac{1}{\sqrt{T}}c_{G_r,t}\left(\frac{C_{G_r}^{\top}C_{G_r}}{T}\right)^{-1}\frac{C_{G_r}^{\top}\eta}{\sqrt{T}}+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)+O_p\left(\frac{N^{\alpha}}{N}\right)\\
&=O_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}
$$

where the third line uses Lemma 1 (c), the fourth uses Theorem 1 (a), and the fifth uses Equation (A.12).

Combining the bias and variance terms, we have

$$
\tilde{c}_{R,T}^{\top}\widehat{\theta}_R-\tilde{c}_{P,T}^{\top}\widehat{\theta}_P=o_p\left(N^{-1/2}\right). \tag{B.13}
$$

Note that the non-shift bias terms for both the pseudo- and rotated methods are larger than $O_p\left(N^{-1/2}\right)$. Therefore, the difference $\tilde{c}_{R,T}^{\top}\widehat{\theta}_R-\tilde{c}_{P,T}^{\top}\widehat{\theta}_P$ is asymptotically negligible relative to the estimation errors $\tilde{c}_{R,T}^{\top}\widehat{\theta}_R-c_T^{\top}\theta$ and $\tilde{c}_{P,T}^{\top}\widehat{\theta}_P-c_T^{\top}\theta$. This shows the asymptotic equivalence. ■

B.3 Small Rotational Break $\nu\in[0,0.5)$

Proof of Theorem 2 (b). We organise the proof in the cases of $\alpha\in[0,0.5)$, $\alpha=0.5$, $\alpha\in(0.5,1)$ and $\alpha=1$.

B.3.1 $\nu\in[0,0.5)$ and $\alpha\in[0,0.5)$

Both the pseudo- and rotated methods have the same leading term of order $O_p\left(\frac{1}{N}\right)$ in their expansions for the squared bias term, i.e. respectively,

$$
\begin{aligned}
\left\|c_T^{\top}\theta-\tilde{C}_{P,T}^{\top}\left(\tilde{C}_P^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_P^{\top}C\theta\right\|^2&=\left\|\left[-\left(\frac{\Lambda_1^{\top}\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^{\top}e_T}{N}\right]^{\top}\beta(L)\right\|^2+o_p\left(\frac{1}{N}\right),\\
\left\|c_T^{\top}\theta-\tilde{c}_{R,T}^{\top}\left(\tilde{C}_R^{\top}\tilde{C}_R\right)^{-1}\tilde{C}_R^{\top}C\theta\right\|^2&=\left\|\left[-\left(\frac{\Lambda_1^{\top}\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^{\top}e_T}{N}\right]^{\top}\beta(L)\right\|^2+o_p\left(\frac{1}{N}\right).
\end{aligned}
$$

The split-sample method has the following expansion for the squared bias term:

$$\begin{aligned}
\left\|c_T^\top \theta - \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2\right)^{-1} \tilde{C}_2^\top C_2 \theta\right\|^2 &= \left\|\frac{-e_T^\top \Lambda_2}{N}\left(\frac{\Lambda_2^\top \Lambda_2}{N}\right)^{-1} \beta(L)\right\|^2 + O_p\left(\frac{1}{\delta_{NT}^2}\right) \\
&= \left\|\frac{-e_T^\top \Lambda_1}{N}\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \beta(L)\right\|^2 + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N\sqrt{N}}\right) \\
&= \left\|\frac{-e_T^\top \Lambda_1}{N}\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \beta(L)\right\|^2 + o_p\left(\frac{1}{N}\right).
\end{aligned}$$

Thus, the leading term for the squared bias terms of all methods are identical.

For the variance terms, recall that the leading terms for the variance of the pseudo-factors and rotated factors are the same for $\alpha < 1$. For $\nu < 1$, we have

$$\begin{aligned}
&\frac{1}{T}\left[\frac{\eta^\top C_{G_r}}{\sqrt{T}}\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1} c_{G_r,T} c_{G_r,T}^\top \left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1} \frac{C_{G_r}^\top \eta}{\sqrt{T}} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right)\right] \\
=&\frac{1}{T}\left[\frac{\eta^\top C_{G_r}}{\sqrt{T}}\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1} c_{G_r,T} c_{G_r,T}^\top \left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1} \left(\frac{C^\top \eta + (C_{G_r} - C)^\top \eta}{\sqrt{T}}\right)\right. \\
&\left. + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right) + O_p\left(\frac{N^\nu}{N}\right)\right] \\
=&\frac{1}{T}\left[\frac{\eta^\top C}{\sqrt{T}}\left(\frac{C^\top C}{T}\right)^{-1} c_{G_r,T} c_T^\top \left(\frac{C^\top C}{T}\right)^{-1} \frac{C^\top \eta}{\sqrt{T}}\right. \\
&+ \frac{\eta^\top C}{\sqrt{T}}\left(\frac{C^\top C}{T}\right)^{-1} c_{G_r,T}\left(c_{G_r,T} - c_T\right)^\top \left(\frac{C^\top C}{T}\right)^{-1} \frac{C^\top \eta}{\sqrt{T}} \\
&\left. + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right) + O_p\left(\frac{N^\nu}{N}\right)\right] \\
=&\frac{1}{T}\left[\frac{\eta^\top C}{\sqrt{T}}\left(\frac{C^\top C}{T}\right)^{-1} c_T c_T^\top \left(\frac{C^\top C}{T}\right)^{-1} \frac{C^\top \eta}{\sqrt{T}} + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right) + O_p\left(\frac{N^\nu}{N}\right)\right]
\end{aligned}$$

Therefore, comparing the leading terms in the expansions for the pseudo-factor method and the split-sample method, we obtain

$$\begin{aligned}
&E\left[\frac{1}{T_2} E\left[\frac{\eta_{(2)}^\top C_2}{\sqrt{T_2}}\left(\frac{C_2^\top C_2}{T_2}\right)^{-1} c_T \middle| C\right]^2 - \frac{1}{T} E\left[\frac{\eta^\top C}{\sqrt{T}}\left(\frac{C^\top C}{T}\right)^{-1} c_T \middle| C\right]^2\right] \\
=&E\left[\frac{1}{T} tr\left[\frac{1}{1-\pi} Var\left(\frac{C_2^\top \eta_{(2)}}{\sqrt{T_2}} \middle| C\right) \Sigma_{CC}^{-1} - Var\left(\frac{C^\top \eta}{\sqrt{T}} \middle| C\right) \Sigma_{CC}^{-1}\right]\right] \\
=&\frac{1}{T} tr\left(\Omega_{CC,\eta} \Sigma_C^{-1}\right) \frac{\pi}{1-\pi} + o_p\left(\frac{1}{T}\right) > 0, \qquad \text{(B.14)}
\end{aligned}$$

where $\Omega_{CC,\eta} = Var(\frac{1}{\sqrt{T}}C^\top \eta|C)$ and $\Sigma_{CC} = E(T^{-1}C^\top C)$, and we use the fact that $\left(\frac{C_2^\top C_2}{T_2}\right)^{-1} c_t c_t^\top \left(\frac{C_2^\top C_2}{T_2}\right)^{-1}$ and $\left(\frac{C^\top C}{T}\right)^{-1} c_t c_t^\top \left(\frac{C^\top C}{T}\right)^{-1} \to \Sigma_{CC}^{-1}$ under uniform integrability. Hence, the split-sample method suffers from a larger variance compared to the pseudo- and rotated factor methods. Combined with the result that its squared bias is of the same asymptotic order, this implies that the split-sample factors are inferior, so the split-sample method is therefore dominated by the other two methods in terms of MSFE for $\alpha < 1/2$ and $\nu < 1/2$.

B.3.2 $\nu \in [0, 0.5)$ and $\alpha = 0.5$

The expansion for the rotated factors remains the same, but the shift break implies an additional term for the pseudo-factor method

$$
\begin{aligned}
&\left\| c_T^\top \theta - \tilde{C}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top C\theta \right\|^2 \\
=&\left\| \left[-\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T} \right) - \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N} \right]^\top \beta(L) \right\|^2 \\
&\quad + o_p\left(\frac{1}{N}\right) \\
=&\left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta(L) \right\|^2 \\
&\quad + \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N} \right]^\top \beta(L) \right\|^2 + o_p\left(\frac{1}{N}\right)
\end{aligned}
\tag{B.15}
$$

where the last line follows from the fact that the cross term can be shown to be negligible. Specifically,

$$
\begin{aligned}
&(\beta(L))^\top \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T} \right)^\top \\
&\quad \times \frac{W^\top W}{N} \frac{G_p^\top \tilde{F}_P}{T} \left(\frac{G_r^\top \tilde{F}_P}{T}\right)^{-1} \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \beta(L) \\
=&(\beta(L))^\top \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N} f_T^\top \frac{W^\top W}{N} \frac{G_p^\top \tilde{F}_P}{T} \left(\frac{G_r^\top \tilde{F}_P}{T}\right)^{-1} \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \beta(L) \\
&- (\beta(L))^\top \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N} \left(\frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T} \right)^\top \\
&\quad \times \frac{W^\top W}{N} \frac{G_p^\top \tilde{F}_P}{T} \left(\frac{G_r^\top \tilde{F}_P}{T}\right)^{-1} \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \beta(L)
\end{aligned}
$$

$$
\begin{aligned}
=&(\beta(L))^\top \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N} f_T^\top \frac{W^\top W}{N} \frac{G_p^\top \tilde{F}_P}{T} \left(\frac{G_r^\top \tilde{F}_P}{T}\right)^{-1} \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \beta(L) \\
&- (\beta(L))^\top \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N} \begin{bmatrix} f_T^\top Z \\ z_T \end{bmatrix} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \frac{\tilde{C}_P^\top G_p}{T} \\
&\times \frac{W^\top W}{N} \frac{G_p^\top \tilde{F}_P}{T} \left(\frac{G_r^\top \tilde{F}_P}{T}\right)^{-1} \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \beta(L) \\
&- (\beta(L))^\top \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N} \begin{bmatrix} \frac{e_T^\top \Lambda_1}{N} \frac{G_r^\top \tilde{F}_P}{T} V_{NT}^{-1} \\ 0 \end{bmatrix} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \frac{\tilde{C}_P^\top G_p}{T} \\
&\times \frac{W^\top W}{N} \frac{G_p^\top \tilde{F}_P}{T} \left(\frac{G_r^\top \tilde{F}_P}{T}\right)^{-1} \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \beta(L) \\
&- (\beta(L))^\top \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N} \begin{bmatrix} f_T^\top \frac{W^\top W}{N} \frac{G_p^\top \tilde{F}_P}{T} V_{NT}^{-1} \\ 0 \end{bmatrix} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \frac{\tilde{C}_P^\top G_p}{T} \\
&\times \frac{W^\top W}{N} \frac{G_p^\top \tilde{F}_P}{T} \left(\frac{G_r^\top \tilde{F}_P}{T}\right)^{-1} \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \beta + O_p\left(\frac{1}{\delta_{NT}^2}\right) \\
=&P.I + P.II + P.III + P.IV,
\end{aligned}
$$

where the last line follows from the fact that $\tilde{c}_{P,T} = c_{P,T} + (\tilde{c}_{P,T} - c_{G_r,T})$, where the latter is a vector with $\tilde{f}_{P,T} - H_G Z^\top f_T$ in its first r columns, and 0 in its remaining columns. The terms $P.III$ and $P.IV$ are both $O_p\left(N^{-3/2}\right)$ and therefore already negligible. The terms I and II are both $O_p\left(\frac{1}{\sqrt{N}}\right)$ and therefore not negligible. However, if we assume that $\frac{\Lambda_1^\top e_T}{\sqrt{N}} \perp\!\!\!\perp f_T$ and $\frac{\Lambda_1^\top e_T}{\sqrt{N}} \perp\!\!\!\perp y_T$, then it follows that $E(\text{plim}_{N,T\to\infty} P.I) = E(\text{plim}_{N,T\to\infty} P.II) = 0$. Therefore, for the pseudo-factors we have

$$
\begin{aligned}
&E\left(\text{plim}_{N,T\to\infty} \left\| c_T^\top \theta - \tilde{C}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top C\theta \right\|^2\right) \\
=&E\left(\text{plim}_{N,T\to\infty} \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T}\right)\right]^\top \beta(L) \right\|^2 \right. \\
&\left. + \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N}\right]^\top \beta(L) \right\|^2\right) + o_p\left(\frac{1}{N}\right).
\end{aligned}
$$

In this scenario, the rotated method has a smaller squared bias term than the pseudo-factor method.

The split-sample method is inferior to the rotated method following the same argument. The ranking between the split-sample method and the pseudo method depends on a bias-variance trade-off which are of identical asymptotic order. Specifically, the variance of the split-sample method exceeds that of the pseudo-factor method by $T^{-1}tr(\Omega\Sigma_F^{-1})\frac{\pi}{1-\pi} \asymp_p N^{-1}$, whereas the pseudo-factor method suffers from an additional squared bias

$$\left\|\left[\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1}\frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N}\left(f_T - \frac{G_p^\top \tilde{F}_P}{T}\left(\frac{\tilde{F}_P^\top\tilde{F}_P}{T}\right)^{-1}\tilde{c}_{P,T}\right)\right]^\top\beta(L)\right\|^2 \asymp_p N^{-1}.$$

The specific ranking, therefore, depends on the specific DGP.

B.3.3 $\nu \in [0, 0.5)$ and $\alpha \in (0.5, 1)$

When $\alpha \in (1/2, 1)$, the expansions for the rotated factors remains the same. However, the shift bias term $\left\|\left[\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1}\frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N}\left(f_T - \frac{G_p^\top \tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top\tilde{C}_P}{T}\right)^{-1}\tilde{c}_{P,T}\right)\right]^\top\beta(L)\right\|^2$ for the pseudo-factors becomes the leading term, and therefore

$$\begin{aligned}
&\|\mu_{T+h} - \widehat{\mu}_{P,T+h}\|^2 / \|\mu_{T+h} - \widehat{\mu}_{R,T+h}\|^2 \to \infty, \\
&\|\mu_{T+h} - \widehat{\mu}_{P,T+h}\|^2 / \|\mu_{T+h} - \widehat{\mu}_{S,T+h}\|^2 \to \infty,
\end{aligned}$$

as $N, T \to \infty$. The split-sample method still remains inferior to the rotated method following the same argument.

B.3.4 $\nu \in [0, 0.5)$ and $\alpha = 1$

If $\alpha = 1$, then $\left\|\left[\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1}\frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N}\left(f_T - \frac{G_p^\top \tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top\tilde{C}_P}{T}\right)^{-1}\tilde{C}_{P,T}\right)\right]^\top\beta\right\|^2 \asymp_p 1$, so the pseudo-factor method is the least effective. The bias of the rotated factor method is

$$\begin{aligned}
&\left\|c_T^\top\theta - (\tilde{c}_{R,T})^\top\left(\tilde{C}_R^\top\tilde{C}_R\right)^{-1}\tilde{C}_R^\top C\theta\right\|^2 \\
=&\left\|\left((I-Z)\left(f_T - \frac{\tilde{F}_2^\top\tilde{C}_{R,2}}{T}\left(\frac{\tilde{C}_R^\top\tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\right) - \left(\frac{\Lambda_2^\top\Lambda_2}{N}\right)^{-1}\frac{\Lambda_2^\top e_T}{N}\right)^\top\beta(L)\right\|^2 \\
&\quad + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{N^{\alpha/2+\nu}}{N^2}\right) + O_p\left(\frac{N^\nu}{\delta_{NT}^2 N}\right) \\
=&\left\|\frac{-e_T^\top\Lambda_2}{N}\left(\frac{\Lambda_2^\top\Lambda_2}{N}\right)^{-1}\beta(L)\right\|^2 + o_p\left(\frac{1}{N}\right),
\end{aligned}$$

which is comparable to the squared bias of the split-sample method

$$\left\| c_T^\top \theta - \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{C}_2^\top C_2 \theta \right\|^2 = \left\| \frac{-e_T^\top \Lambda_2}{N} \left(\frac{\Lambda_2^\top \Lambda_2}{N} \right)^{-1} \beta(L) \right\|^2 + o_p\left(\frac{1}{N}\right).$$

Furthermore, the variance terms of the rotated and split-sample methods are both $\asymp_p N^{-1}$. Thus, the specific ranking of the rotated and split-sample methods depends on the DGP.

■

B.4 Moderate Rotational Break $\nu = 0.5$

Proof of Theorem 2 (c) - moderate rotational breaks $\nu = 0.5$. We organise the proof in the cases of $\alpha \in [0, 0.5)$, $\alpha = 0.5$, $\alpha \in (0.5, 1)$ and $\alpha = 1$.

B.4.1 $\nu = 0.5$ and $\alpha \in [0, 0.5)$

For the pseudo-factor method, the expectation of the probability limit of the squared bias term is

$$\begin{aligned}
&E\left(\operatorname{plim}_{N,T\to\infty} \left\| c_T^\top \theta - \tilde{C}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C \theta \right\|^2 \right) \\
=&E\left(\operatorname{plim}_{N,T\to\infty} \left\| \left[(I - Z) \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) - \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \frac{\Lambda_1^\top e_T}{N} \right]^\top \beta(L) \right\|^2 \right. \\
&\quad \left. + o_p\left(\frac{1}{N}\right) \right) \\
=&E\left(\operatorname{plim}_{N,T\to\infty} \left\| \left[(I - Z) \left(f_T - \frac{G_p^\top G_r}{T} \left(\frac{C_{G_r}^\top C_{G_r}}{T} \right)^{-1} c_T \right) \right]^\top \beta(L) \right\|^2 \right. \\
&\quad \left. + \operatorname{plim}_{N,T\to\infty} \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \frac{\Lambda_1^\top e_T}{N} \right]^\top \beta \right\|^2 \right) \\
&\quad + o_p\left(\frac{1}{N}\right),
\end{aligned} \tag{B.16}$$

which follows because

$$-2E\left[\operatorname{plim}_{N,T\to\infty} (\beta(L))^\top (I - Z) \left(I - \frac{G_p^\top G_r}{T} \left(\frac{G_r^\top G_r}{T} \right)^{-1} \right) f_T \frac{e_T^\top \Lambda_1}{N} \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \beta(L) \right] = 0.$$

For the rotated factor method, we have that the expected value of its probability limit is

$$
\begin{aligned}
&E\left(\operatorname{plim}_{N,T\to\infty}\left\|c_T^\top\theta-(\tilde{c}_{R,T})^\top\left(\tilde{C}_R^\top\tilde{C}_R\right)^{-1}\tilde{C}_R^\top C\theta\right\|^2\right)\\
=&E\left(\operatorname{plim}_{N,T\to\infty}\left\|\left[(I-Z)\left(f_T-\frac{\tilde{F}_2^\top\tilde{C}_R}{T}\left(\frac{\tilde{C}_R^\top\tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\right)-\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N}\right]^\top\beta\right\|^2+o_p\left(\frac{1}{N}\right)\right)\\
=&E\left(\operatorname{plim}_{N,T\to\infty}\left\|\left[(I-Z)\left(f_T-\frac{H_2^\top F_2^\top C_{G,r}}{T}\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1}c_{G_r,T}\right)\right]^\top\beta(L)\right\|^2\right.\\
&\left.+\left\|\left[\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N}\right]^\top\beta(L)\right\|^2\right)+o_p\left(\frac{1}{N}\right),
\end{aligned}
\tag{B.17}
$$

which similarly uses the fact that the cross term has a zero expected value in its probability limit

$$
E\left[\operatorname{plim}_{N,T\to\infty}(\beta(L))^\top(I-Z)\left(f_T-\frac{H_2^\top F_2^\top C_{G,r}}{T}\left(\frac{G_r^\top G_r}{T}\right)^{-1}c_{G_r,T}\right)\frac{e_T^\top\Lambda_1}{N}\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\beta(L)\right]=0.
$$

Thus, both the pseudo- and rotated factor methods have an extra $O_p\left(\frac{1}{N}\right)$ term in their squared biases. Recall that the variance term of the split-sample method exceeds that of the pseudo- and rotated factor methods by a $O_p\left(\frac{1}{N}\right)$ term, and that the forecasts of the pseudo- and rotated factor methods are asymptotically identical due to Theorem 2 (a). Therefore, the ranking between the pseudo-, rotated, and split-sample methods depends on the bias-variance trade-off determined by the bias terms magnitude of relative to $T^{-1}tr(\Omega\Sigma_{CC}^{-1})\pi/(1-\pi)$, which depends on the specific DGP.

B.4.2 $\nu = 0.5$ and $\alpha = 0.5$

If $\alpha = 0.5$, then the expansion for the rotated factors remains the same as Equation (B.17), but the pseudo-factors have an extra term due to the shift break

$$
\begin{aligned}
&\left\|c_T^\top\theta-\tilde{c}_{P,T}^\top\left(\tilde{C}_P^\top\tilde{C}_P\right)^{-1}\tilde{C}_P^\top C\theta\right\|^2\\
=&\left\|\left[\left((I-Z)-\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1}\frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N}\right)\left(f_T-\frac{G_p^\top\tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top\tilde{C}_P}{T}\right)^{-1}\tilde{c}_{P,T}\right)\right.\right.\\
&\left.\left.-\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N}\right]^\top\beta(L)\right\|^2+o_p\left(\frac{1}{N}\right)
\end{aligned}
$$

$$
\begin{aligned}
=&\left\| \left[(I-Z)\left(f_T - \frac{G_p^\top \tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T}\right)\right]^\top \beta(L) \right\|^2 \\
&+ \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1} \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N}\left(f_T - \frac{G_p^\top \tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T}\right)\right]^\top \beta(L) \right\|^2 \\
&+ \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N}\right]^\top \beta(L) \right\|^2 \\
&- 2(\beta(L))^\top (I-Z)\left\| f_T - \frac{G_p^\top \tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T}\right\|^2 \frac{W^\top W}{N}\frac{G_p^\top \tilde{F}_P}{T}\left(\frac{G_r^\top \tilde{F}_P}{T}\right)^{-1}\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1}\beta(L) \\
&- 2(\beta(L))^\top (I-Z)\left(f_T - \frac{G_p^\top \tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T}\right)\frac{e_T^\top \Lambda_1}{N}\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1}\beta(L) \\
&- 2(\beta(L))^\top \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1} \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N}\left(f_T - \frac{G_p^\top \tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T}\right) \\
&\quad \times \frac{e_T^\top \Lambda_1}{N}\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1}\beta(L) \\
&+ o_p\left(\frac{1}{N}\right),
\end{aligned} \tag{B.18}
$$

where the last two cross terms have an expected probability limit of zero, using arguments similarly employed in Appendix B.4.1. In this case, whether the rotated or the pseudo-factor method depends on the sign of

$$
\begin{aligned}
&\left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1} \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N}\left(f_T - \frac{G_p^\top \tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T}\right)\right]^\top \beta(L) \right\|^2 \\
&- 2(\beta(L))^\top (I-Z)\left\| f_T - \frac{G_p^\top \tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T}\right\|^2 \frac{W^\top W}{N}\frac{G_p^\top \tilde{F}_P}{T}\left(\frac{G_r^\top \tilde{F}_P}{T}\right)^{-1}\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1}\beta(L).
\end{aligned}
$$

If this term is positive (negative), then the rotated method has a smaller (larger) squared bias than the pseudo method. The comparison between the pseudo-, rotated, and split-sample methods follow a similar bias-variance argument. That is, their relative rankings depend on the specific DGP.

B.4.3 $\nu = 0.5$ and $\alpha \in (0.5, 1)$

If $\alpha \in (0.5, 1]$, then the bias term caused by the shift break becomes the leading term for the pseudo-factors. Since $\|\mu_{T+h} - \widehat{\mu}_{S,T+h}\|^2 \asymp_p N^{-1}$ and $\|\mu_{T+h} - \widehat{\mu}_{R,T+h}\|^2 \asymp_p N^{-1}$, we have

$$\|\mu_{T+h} - \widehat{\mu}_{P,T+h}\|^2 / \|\mu_{T+h} - \widehat{\mu}_{R,T+h}\|^2 \to \infty,$$
$$\|\mu_{T+h} - \widehat{\mu}_{P,T+h}\|^2 / \|\mu_{T+h} - \widehat{\mu}_{S,T+h}\|^2 \to \infty,$$

as $N, T \to \infty$. The ranking between the rotated and split-sample factors depends on a similar bias-variance trade-off.

B.4.4 $\nu = 0.5$ and $\alpha = 1$

If $\alpha = 1$, then $\left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T}\right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta(L) \right\|^2 \asymp_p 1$, so the pseudo-factor method is the least effect. The bias of the rotated factor method is

$$\begin{aligned}
&\left\| c_T^\top \theta - c_T^\top \theta - (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top \tilde{C}_R\right)^{-1} \tilde{C}_R^\top C\theta \right\|^2 \\
=&\left\| \left((I - Z) \left(f_T - \frac{\tilde{F}_2^\top \tilde{C}_{R,2}}{T} \left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T}\right)^{-1} \tilde{c}_{R,T} \right) - \left(\frac{\Lambda_2^\top \Lambda_2}{N}\right)^{-1} \frac{\Lambda_2^\top e_T}{N} \right)^\top \beta(L) \right\|^2 + o_p\left(\frac{1}{N}\right) \\
=& \asymp_p N^{-1}.
\end{aligned}$$

The split-sample method has the same asymptotic order for its squared bias term, and both the rotated and split-sample methods have variance terms that are $\asymp_p N^{-1}$. Therefore, the ranking between the rotated and split-sample factors depends on the DGP. ■

B.5 Large Rotational Break $\nu \in (0.5, 1]$

Proof of Theorem 2 (d) - large rotational breaks $\nu > 0.5$. We organise the proof by the cases of $\nu < 1$ and $\nu = 1$, and within those two cases by increasing values of α.

B.5.1 $\nu \in (0.5, 1)$ and $\alpha < \nu$

The squared bias of the pseudo-factor method is

$$\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C \theta \right\|^2 = \left\| \left[(I - Z) \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta(L) \right\|^2 + o_p \left(N^{2\nu - 2} \right)$$
$$\asymp_p N^{2\nu-2}. \tag{B.19}$$

The squared bias of the rotated factor method is

$$\left\| c_T^\top \theta - (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top \tilde{C}_R \right)^{-1} \tilde{C}_R^\top C \theta \right\|^2 = \left\| \left[(I - Z) \left(f_T - \frac{\tilde{F}_2^\top \tilde{C}_{R,2}}{T} \left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T} \right)^{-1} \tilde{c}_{R,T} \right) \right]^\top \beta(L) \right\|^2 + o_p \left(N^{2\nu - 2} \right)$$
$$\asymp_p N^{2\nu-2}. \tag{B.20}$$

Both of these converge to zero at a slower rate than $\left\| c_T^\top \theta - \tilde{c}_{S,T}^\top \widehat{\theta}_S \right\|^2 \asymp_p N^{-1}$. Hence, the bias term induced by the rotational break will dominate the lower variance, and the split-sample method is superior to both the pseudo- and rotated factor methods.

Note that because the leading term associated with the rotational breaks are different for the pseudo- and rotated methods, their specific ranking will depend on the DGP.

B.5.2 $\nu \in (0.5, 1)$ and $\alpha = \nu$

In this case, the expansion of the rotated method remains the same as Equation (B.20), but the expansion for the squared bias of the pseudo method has an additional term due to the shift break. Specifically

$$\begin{aligned} &\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C \theta \right\|^2 \\ &= \left\| \left[\left((I - Z) - \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T} \right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \right) \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta(L) \right\|^2 \\ &\quad + o_p \left(N^{2\nu-2} \right), \end{aligned} \tag{B.21}$$

so the specific ranking between the rotated and pseudo-factor method depends the sign of the cross term. By the same argument, the split-sample factors still remain better than the others.

B.5.3 $\nu \in (0.5, 1)$ and $\alpha \in (\nu, 1)$

The expansion for the rotated factors remains the same as Equation (B.20). However, for the pseudo-factors, the bias term induced by the shift break is now the leading term, i.e.

$$\begin{aligned}
&\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C \theta \right\|^2 \\
=& \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T} \right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta(L) \right\|^2 + o_p\left(N^{2\alpha - 2} \right) \\
=& \asymp_p N^{2\alpha-2}.
\end{aligned} \tag{B.22}$$

Recall that $\left\| \mu_{T+h} - \widehat{\mu}_{S,T+h} \right\|^2 \asymp_p N^{-1}$, so we have

$$\begin{aligned}
&\left\| \mu_{T+h} - \widehat{\mu}_{P,T+h} \right\|^2 / \left\| \mu_{T+h} - \widehat{\mu}_{S,T+h} \right\|^2 \to \infty, \\
&\left\| \mu_{T+h} - \widehat{\mu}_{R,T+h} \right\|^2 / \left\| \mu_{T+h} - \widehat{\mu}_{S,T+h} \right\|^2 \to \infty.
\end{aligned}$$

B.5.4 $\nu \in (0.5, 1)$ and $\alpha = 1$

The expansion of the rotated factor estimator remains the same as Equation (B.20). The pseudo-factor method is the least effective estimator, because

$$\begin{aligned}
&\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C \theta \right\|^2 \\
=& \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T} \right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta(L) \right\|^2 \asymp_p 1,
\end{aligned}$$

and $\left\| \mu_{T+h} - \widehat{\mu}_{S,T+h} \right\| \asymp_p N^{-1}$.

B.5.5 $\nu = 1$ and $\alpha < 1$

The squared bias terms for the pseudo- and rotated factors are respectively

$$\begin{aligned}
&\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C \theta \right\|^2 \\
=& \left\| \left[(I - Z) \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta(L) \right\|^2 + o_p(1) \\
\asymp_p& \ 1,
\end{aligned}$$

$$
\begin{aligned}
&\left\| c_T^\top\theta - c_T^\top\theta - (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top\tilde{C}_R\right)^{-1}\tilde{C}_R^\top C\theta\right\|^2 \\
&= \left\| \left[(I-Z)\left(f_T - \frac{\tilde{F}_2^\top\tilde{C}_{R,2}}{T}\left(\frac{\tilde{C}_R^\top\tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\right)\right]^\top \beta(L)\right\|^2 + o_p(1) \\
&\asymp_p 1.
\end{aligned}
$$

Again, these two leading terms are algebraically different, though of the same order. Both are dominated by the split-sample method.

B.5.6 $\nu = 1$ and $\alpha = 1$

The squared bias terms for the pseudo- and rotated factors are, respectively,

$$
\begin{aligned}
&\left\| c_T^\top\theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top\tilde{C}_P\right)^{-1}\tilde{C}_P^\top C\theta\right\|^2 \\
&= \left\| \left[\left((I-Z) - \left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1}\frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N}\right)\left(f_T - \frac{G_p^\top\tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top\tilde{C}_P}{T}\right)^{-1}\tilde{c}_{P,T}\right)\right]^\top \beta(L)\right\|^2 \\
&\quad + o_p(1) \\
&\asymp_p 1, \\
&\left\| c_T^\top\theta - (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top\tilde{C}_R\right)^{-1}\tilde{C}_R^\top C\theta\right\|^2 \\
&= \left\| \left((I-Z)\left(f_T - \frac{\tilde{F}_2^\top\tilde{C}_{R,2}}{T}\left(\frac{\tilde{C}_R^\top\tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\right) - \left(\frac{\Lambda_2^\top\Lambda_2}{N}\right)^{-1}\frac{\Lambda_2^\top e_T}{N}\right)^\top \beta(L)\right\|^2 + o_p(1) \\
&\asymp_p 1.
\end{aligned}
$$

both of which dominate their variance terms. The ranking between them depends on the realisation of the cross term. ■

C Forecasting Proofs

We first prove the following lemma, which establishes that the cross-validation estimate $\tilde{\theta}(m)_{t,h}$ is uniformly close to $\widehat{\theta}(m)$.

Lemma 11. *If u_t is piece-wise stationary and ergodic such that its pre- and post-break second moments satisfy $E\|u_{1t}\|^2 < \infty$, $E\|u_{2t}\|^2 < \infty$, and $g(u)$ is continuously differentiable at $\mu = E(u_t)$, then for the full* *sample estimator* $\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} u_t$ *and leave* h *out estimator* $\tilde{\mu}_{t,h} = (T+1-2h)^{-1}\sum_{|j-t|\geq h} u_j$,

$$\max_{1\leq t\leq T}\left\|\sqrt{T}\left(g(\widehat{\mu}) - g(\tilde{\mu}_{t,h})\right)\right\| = o_p(1)$$

Lemma 11 establishes that Lemma 1 of Cheng and Hansen (2015) still holds for data that is subject to structural break but is still piece-wise stationary.

Proof of Lemma 11. Suppose that μ_t is piece-wise stationary and ergodic, such that

$$u_t = \begin{cases} u_{1t}, & t = 1, \ldots, \pi T, \\ u_{2t}, & t = \pi T + 1, \ldots, T, \end{cases}$$

$$E(u_{1t}) = \mu_1 < \infty, \quad E\left|u_{1t}\right|^2 < \infty, \quad \text{and}$$
$$E(u_{2t}) = \mu_2 < \infty, \quad E\left|u_{2t}\right|^2 < \infty.$$

We have

$$\begin{aligned}\max_{1\leq t\leq T}\|u_t\| &= \max\left(\max_{1\leq t\leq \pi T}\|u_{1t}\|, \max_{\pi T+1\leq t\leq T}\|u_{2t}\|\right)\\ &= \max\left(o_p\left(\sqrt{T}\right), o_p\left(\sqrt{T}\right)\right)\\ &= o_p\left(\sqrt{T}\right)\end{aligned}$$

Second, since

$$\widehat{\mu} - \tilde{\mu}_{t,h} = \frac{1-2h}{T(T+1-2h)}\sum_{t=1}^{T} u_t + \frac{1}{T+1-2h}\sum_{|j-t|\geq h} u_j$$

then

$$\begin{aligned}\max_{1\leq t\leq T}\|\widehat{\mu} - \tilde{\mu}_{t,h}\| &\leq O_p\left(\frac{1}{T}\right) + \frac{2h}{T+1-2h}\max_{1\leq t\leq T}\|u_t\|\\ &= o_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

An application of the Delta method then yields

$$\max_{1\leq t\leq T}\left\|\sqrt{T}\left(g(\widehat{\mu}) - g(\tilde{\mu}_{t,h})\right)\right\| = o_p(1).$$

■

Proof of Proposition 3 and Theorem 3. The term $\tilde{r}_{1T}(m)$ can be decomposed further by directly replacing $\tilde{C}(m)$ with $C_H(m)$:

$$\begin{aligned}
\frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t\left[\mu_t-\tilde{\mu}_t(w)\right] &= \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t\left[\left(\mu_t-\left(\tilde{c}_t(m)-c_t(m)\right)^\top\tilde{\theta}_{t,h}(m)\right)-c_{Ht}(m)^\top\tilde{\theta}_{t,h}(m)\right]\\
&= \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t\left(\mu_t-c_{Ht}(m)^\top\theta(m)\right)+\frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t(c_{Ht}(m)-\tilde{c}_t(m))^\top\tilde{\theta}_{t,h}\\
&+\frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t c_t^\top\left(\widehat{\theta}(m)-\tilde{\theta}_{t,h}(m)\right)-\frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t c_t^\top\left(\widehat{\theta}(m)-\theta(m)\right)\\
&= \tilde{r}_{1T}^0(m)+\tilde{r}_{2T}(m)+\tilde{r}_{3T}(m)+\tilde{r}_{4T}(m).
\end{aligned}$$

The term $\tilde{r}_{1T}^0(m)$ and therefore $\tilde{r}_{1T}(w)$ are asymptotically normally distributed with zero mean. To see this, Assumption 9 implies that for each m,

$$\begin{aligned}
\frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t\left(\mu_t-c_t(m)^\top\theta(m)\right) &= \frac{1}{T_2}(\mu_{(2)}-C_{2,H}(m)\theta(m))^\top\eta_{(2)}\\
&= \frac{1}{(1-\pi)\sqrt{T}}\frac{1}{\sqrt{T}}(\mu_{(2)}-C_{2,H}(m)\theta(m))^\top\eta_{(2)}\\
&\xrightarrow{d} S_1(m)\sim N(0,\sigma^2 Q(m))
\end{aligned}$$

where $Q(m)=\operatorname{plim}_{T\to\infty}\frac{1}{(1-\pi)^2}\frac{1}{T}(\mu_{(2)}-C_{2,H}(m)\theta(m))^\top(\mu_{(2)}-C_{2,H}(m)\theta(m))$. Additionally,

$$\tilde{r}_{1T}^0(w)\xrightarrow{d}\xi_1(w)=\sum_{m=1}^{3\mathcal{M}} w(m)S_1(m) \tag{C.1}$$

is a weighted sum of mean zero normal variables, and thus $E\xi_1(w)=0$.

It remains to show that terms $\tilde{r}_{2T}(w),\tilde{r}_{3T}(w)$ and $\tilde{r}_{4T}(w)$ are $o_p\left(\frac{1}{\sqrt{T}}\right)$.

For term $\tilde{r}_{4T}(m)$,

$$\begin{aligned}
\widehat{\theta}(m)-\theta(m) &= \left(\tilde{C}(m)^\top\tilde{C}(m)\right)^{-1}\tilde{C}(m)^\top y-\left(C_H(m)^\top C_H(m)\right)^{-1}C_H(m)^\top Y\\
&= \left[\left(\frac{\tilde{C}(m)^\top\tilde{C}(m)}{T}\right)^{-1}-\left(\frac{C_H(m)^\top C_H(m)T}{T}\right)^{-1}\right]\frac{\tilde{C}(m)^\top Y}{T}\\
&+\left(\frac{C_H(m)^\top C_H(m)}{T}\right)^{-1}\frac{\left(\tilde{C}(m)-C_H(m)\right)^\top Y}{T}.
\end{aligned}$$

The first term is bounded by

$$
\begin{aligned}
&\frac{\tilde{C}(m)^\top \tilde{C}(m)}{T} - \frac{C_H(m)^\top C_H(m)}{T} \\
&= \frac{\left(\tilde{C}(m) - C_H(m)\right)^\top \left(\tilde{C} - C_H(m)\right)}{T} + \frac{C_H(m)^\top \left(\tilde{C}(m) - C_H(m)\right)}{T} + \frac{\left(\tilde{C}(m) - C_H(m)\right)^\top C_H(m)}{T} \\
&= \begin{cases}
O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right), & m = 1, \ldots, \mathcal{M}, \quad \text{Pseudo-factors}, \alpha < 1 \\
O_p\left(\frac{1}{\delta_{NT}^2}\right), & m = 1, \ldots, \mathcal{M}, \quad \text{Pseudo-factors}, \alpha = 1 \\
O_p\left(\frac{1}{\delta_{NT}^2}\right), & m = \mathcal{M}+1, \ldots, 2\mathcal{M}, \quad \text{Split-sample Factors} \\
O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), & m = 2\mathcal{M}, \ldots, 3\mathcal{M}, \quad \text{Rotated Factors},
\end{cases}
\end{aligned}
$$

by Lemma 1 (a), Lemma 1 (b), Lemma 4 (a) and Lemma 3.

The second term is bounded by

$$
\begin{aligned}
&\frac{\left(\tilde{C}(m) - C_H(m)\right)^\top Y}{T} \\
&= \begin{cases}
\left[\frac{\left(\tilde{F}_P - G_r H_G\right)^\top Y}{T}, \quad 0_u\right], & m = 1, \ldots, \mathcal{M}, \quad \text{Pseudo-factors}, \alpha < 1, \\
\left[\frac{\left(\tilde{F}_P - G H_\Xi\right)^\top Y}{T}, \quad 0_u\right], & m = 1, \ldots, \mathcal{M}, \quad \text{Pseudo-factors}, \alpha = 1, \\
\left[\frac{\left(\tilde{F}_1 - F_1 H_1\right)^\top Y_1}{T} + \frac{\left(\tilde{F}_2 - F_2 H_2\right)^\top Y_2}{T}, \quad 0_u\right], & m = \mathcal{M}+1, \ldots, 2\mathcal{M}, \quad \text{Split-sample Factors}, \\
\left[\frac{\left(\tilde{F}_1 - F_1 H_1\right)^\top Y_1}{T} + \frac{\left(\tilde{F}_2 \tilde{Z}^\top - F_2 Z^\top H_1\right)^\top Y_2}{T}, \quad 0_u\right], & m = 2\mathcal{M}, \ldots, 3\mathcal{M}, \quad \text{Rotated Factors},
\end{cases} \\
&= \begin{cases}
O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N}\right), & m = 1, \ldots, \mathcal{M}, \quad \text{Pseudo-factors}, \alpha < 1, \\
O_p\left(\frac{1}{\delta_{NT}^2}\right), & m = 1, \ldots, \mathcal{M}, \quad \text{Pseudo-factors}, \alpha = 1, \\
O_p\left(\frac{1}{\delta_{NT}^2}\right), & m = \mathcal{M}+1, \ldots, 2\mathcal{M}, \quad \text{Split-sample Factors}, \\
O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), & m = 2\mathcal{M}, \ldots, 3\mathcal{M}, \quad \text{Rotated Factors}.
\end{cases}
\end{aligned}
$$

Therefore, term $\tilde{r}_{4T}(w) = \sum_{m=1}^{3\mathcal{M}} \tilde{r}_{4T}(m) = o_p\left(\frac{1}{\sqrt{T}}\right)$.

Term $\tilde{r}_{3T}(m)$ can be bounded by

$$
\left\| \frac{1}{T_2} \sum_{t=T_1+1-h}^{T-h} 2\eta_t c_t^\top \left(\widehat{\theta}(m) - \tilde{\theta}_{t,h}(m)\right) \right\| \leq \frac{2}{T_2} \sum_{t=T_1+1-h}^{T-h} \left\| \eta_t c_t^\top \right\| \underset{t > \lfloor \pi T \rfloor}{max} \left\| \tilde{\theta}_{t,h}(m) - \widehat{\theta}(m) \right\|.
$$

Thus, $\tilde{r}_{3T}(w) = \sum_{m=1}^{3\mathcal{M}} \tilde{r}_{3T}(m) = o_p\left(\frac{1}{\sqrt{T}}\right)$.

For term $\tilde{r}_{2T}(m)$, we have

$$\begin{aligned}\frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t\left(c_t(m)-\tilde{c}_t(m)\right)^\top \tilde{\theta}_{t,h}(m) &= \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t\left(c_t(m)-\tilde{c}_t(m)\right)^\top \left(\tilde{\theta}_{t,h}(m)-\widehat{\theta}(m)\right) \\ &+ \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t\left(c_t(m)-\tilde{c}_t(m)\right)^\top \widehat{\theta}(m).\end{aligned}$$

The first term is negligible because

$$\begin{aligned}&\left\|\frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t\left(c_t(m)-\tilde{c}_t(m)\right)^\top \left(\tilde{\theta}_{t,h}(m)-\widehat{\theta}(m)\right)\right\| \\ &\leq \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} \|2\eta_t(c_t(m)-\tilde{c}_t(m))\| \max_{t>\lfloor \pi T\rfloor}\left\|\tilde{\theta}_{t,h}(m)-\widehat{\theta}(m)\right\| \\ &= 2\left(\frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h}\eta_t^2 \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h}\|c_t(m)-\tilde{c}_t(m)\|^2\right)^{1/2} max\left\|\tilde{\theta}_{t,h}(m)-\widehat{\theta}(m)\right\| \\ &= \begin{cases}\left(O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{N^{2\alpha}}{N^2}\right)\right)^{1/2} o_p\left(\frac{1}{\sqrt{T}}\right), & m=1,\ldots,\mathcal{M}, \text{Pseudo-factors}, \alpha<1, \\ \left(O_p\left(\frac{1}{\delta_{NT}^2}\right)\right)^{1/2} o_p\left(\frac{1}{\sqrt{T}}\right), & m=1,\ldots,\mathcal{M}, \text{Pseudo-factors}, \alpha=1, \\ \left(O_p\left(\frac{1}{\delta_{NT}^2}\right)\right)^{1/2} o_p\left(\frac{1}{\sqrt{T}}\right), & m=\mathcal{M}+1,\ldots,2\mathcal{M}, \text{Split-sample Factors}, \\ \left(O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{N^{\alpha}}{N^2}\right)\right)^{1/2} o_p\left(\frac{1}{\sqrt{T}}\right), & m=2\mathcal{M}+1,\ldots,3\mathcal{M}, \text{Rotated Factors}\end{cases} \\ &= o_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

The second term is negligible because

$$\begin{aligned}&\frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t\left(c_t(m)-\tilde{c}_t(m)\right)^\top \widehat{\theta}(m) \\ &= \begin{cases}\frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t(f_t^\top Z^\top H_G - \tilde{f}_{P,t}^\top), & m=1,\ldots,\mathcal{M}, \text{Pseudo-factors}, \alpha<1, \\ \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t(g_t^\top H_\Xi - \tilde{f}_{P,t}^\top), & m=1,\ldots,\mathcal{M}, \text{Pseudo-factors}, \alpha=1, \\ \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t(f_t^\top H_2 - \tilde{f}_{S,t}^\top), & m=\mathcal{M}+1,\ldots,2\mathcal{M}, \text{Split-sample Factors}, \\ \frac{1}{T_2}\sum_{t=T_1+1-h}^{T-h} 2\eta_t(f_t^\top Z^\top H_1 - \tilde{f}_{R,t}^\top), & m=2\mathcal{M}+1,\ldots,3\mathcal{M}, \text{Rotated Factors},\end{cases}\end{aligned}$$

$$
= \begin{cases} O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^\alpha}{N\sqrt{T}}\right), & m = 1, \ldots, \mathcal{M}, \text{Pseudo-factors}, \alpha < 1, \\ O_p\left(\frac{1}{\delta_{NT}^2}\right), & m = 1, \ldots, \mathcal{M}, \text{Pseudo-factors}, \alpha = 1, \\ O_p\left(\frac{1}{\delta_{NT}^2}\right), & m = \mathcal{M}+1, \ldots, 2\mathcal{M}, \text{Split-sample Factors}, \\ O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{\sqrt{T}N}\right), & m = 2\mathcal{M}+1, \ldots, 3\mathcal{M}, \text{Rotated Factors}, \end{cases}
$$

$$
= o_p\left(\frac{1}{\sqrt{T}}\right).
$$

Therefore, $\tilde{r}_{2T}(w) = \sum_{m=1}^{3\mathcal{M}} \tilde{r}_{2T}(m) = o_p\left(\frac{1}{\sqrt{T}}\right)$. This proves Proposition 3. The result in Theorem 3 follows immediately. ■

D Empirical Results

D.1 Empirical Robustness Checks

Table 7 and Table 8 are the analogous tables for the Stock and Watson (2012) dataset, using the Great Moderation as a potential break. The results are qualitatively similar.

Table 7: Distributions of relative RMSE by forecasting method, relative to DFM-5, h = 1, 2, 4, for Stock and Watson (2012) Dataset (1959 Q3 - 2008 Q3, 1984 Q1 Break)

Percentile	h = 1			h = 2			h = 4		
Model	0.250	0.500	0.750	0.250	0.500	0.750	0.250	0.500	0.750
CV Select	0.962	1.000	1.030	0.970***	1.001	1.035	0.986	1.010	1.044
CV Weighted	0.956*	0.996***	1.017*	0.969**	0.999	1.023***	0.979**	1.002	1.030
Equal Weighted	0.959***	1.001	1.043	0.970***	1.010	1.050	0.981***	1.017	1.065
Mallows Select	0.973	1.004	1.045	0.973	0.997**	1.035	0.981***	0.997*	1.017*
Mallows Weighted	0.957**	0.992*	1.024***	0.967*	0.995*	1.016*	0.973*	1.001***	1.027
Pseudo r	0.981	0.999	1.020**	0.982	0.998***	1.021**	0.988	1.002	1.023***
Rotated	0.963	0.995**	1.033	0.977	1.000	1.026	0.981***	0.998**	1.021**
Split-sample	1.100	1.225	1.367	1.151	1.262	1.391	1.155	1.312	1.496

Note:

Entries are percentiles of distributions of relative RMSEs over the 143 variables being forecasts, by series, at the specified forecast horizon. RMSEs are relative to the DFM-5 forecast, as an expanding window exercise. All forecasts are direct.

Table 8: Median RMSE by forecasting method and category of series, relative to DFM-5, rolling forecast estimates for Stock and Watson (2012) Dataset (1959 Q3 - 2008 Q3, 1984 Q1 Break).

Group	CV Select	CV Weighted	Equal Weighted	Mallows Select	Mallows Weighted	Pseudo r	Rotated	Split-sample
h = 1								
GDP Components	1.009*	1.022	1.035	1.023	1.009*	1.025	1.017***	1.358
Industrial Production	1.030	1.023	1.021***	1.075	1.000*	1.006**	1.046	1.238
Employment	0.972	0.927*	0.976	1.064	0.936**	0.976	0.956***	1.251
Unemployment	1.005***	1.008	0.995*	1.042	1.004**	1.008	1.020	1.223
Housing	0.966	0.962	0.954*	0.980	0.959**	0.961***	0.975	1.115
Inventories	1.032	1.007***	1.008	1.040	0.995*	1.001**	1.035	1.258
Prices	0.978***	0.980	0.976**	0.995	0.971*	0.995	0.979	1.141
Earnings	0.998	0.994	0.947*	0.995	0.993***	0.999	0.984**	1.035
Interest Rates	0.995**	1.003***	1.080	0.972*	1.042	1.080	1.023	1.415
Money	1.000	1.006	1.008	0.957*	1.038	0.995***	0.978**	1.178
Exchange Rates	1.005	1.000	1.024	0.987*	1.005	0.993***	0.991**	1.407
Stock Prices	1.011	1.000	0.974***	0.956*	0.966**	1.004	1.005	1.197
Consumer Expectations	1.019***	1.018**	1.113	1.040	1.043	1.007*	1.019***	1.538
h = 2								
GDP Components	1.020	1.017***	1.039	1.022	1.006*	1.010**	1.020	1.371
Industrial Production	1.019	1.019	0.971*	1.018	0.980**	1.005	1.001***	1.110
Employment	1.015	0.987	0.980**	1.064	0.983***	0.976*	1.007	1.281
Unemployment	0.999*	1.007	1.039	0.999*	1.006***	1.015	1.016	1.361
Housing	0.994***	0.996	0.979*	1.046	0.983**	1.000	1.016	1.137
Inventories	0.983***	0.968*	0.984	1.035	0.982**	1.001	1.016	1.301
Prices	0.999***	1.003	1.032	0.987*	1.000	0.999***	0.991**	1.243
Earnings	0.993	1.013	1.051	0.990**	0.996	0.990**	0.986*	1.245

Interest Rates	0.953	0.948*	0.991	0.950**	0.951***	0.990	0.974	1.316
Money	1.004	0.995*	1.003	1.009	0.998***	0.998***	0.997**	1.135
Exchange Rates	0.996	0.990	1.022	0.960*	0.996	0.986***	0.981**	1.457
Stock Prices	0.973**	0.977	0.964*	0.973**	0.980	0.993	0.977	1.182
Consumer Expectations	1.043	1.026***	1.051	0.963*	1.106	1.035	0.992**	1.417
h = 4								
GDP Components	1.011	1.017	1.048	0.993*	1.010	0.999**	1.005***	1.352
Industrial Production	1.017	0.979***	0.974**	0.996	0.969*	1.000	0.998	1.150
Employment	1.031	0.996	0.961*	1.027	0.984**	0.990	0.989***	1.311
Unemployment	1.016	1.012***	0.961*	1.025	1.007**	1.024	1.022	1.144
Housing	1.049	1.033	1.026	0.976**	1.016***	1.047	0.974*	1.411
Inventories	0.976***	0.963*	0.991	0.982	0.972**	1.026	0.985	1.295
Prices	1.001	1.011	1.045	0.996*	1.006	0.999***	0.996*	1.321
Earnings	1.007	0.997**	1.006	0.996*	1.000***	1.009	1.014	1.154
Interest Rates	1.024***	1.013**	1.136	1.002*	1.076	1.069	1.049	1.580
Money	0.992*	0.997***	1.027	1.002	1.004	1.002	0.994**	1.365
Exchange Rates	1.047	1.015	1.094	0.988*	1.020	1.008***	0.994**	1.569
Stock Prices	1.027	0.984*	1.007	0.985**	1.004	1.016	0.986***	1.335
Consumer Expectations	1.003	1.001	1.011	0.990**	0.993	0.992***	0.989*	1.214

Note:

Entries are median RMSEs, relative to DFM-5, for the row category of variables.

References

Bai, Jushan (1997), “Estimation of a Change Point in Multiple Regression Models.” *Review of Economics and Statistics*, 79 (4), 551–563.

Bai, Jushan (2003), “Inferential Theory for Factor Models of Large Dimensions.” *Econometrica*, 71 (1), 135–171.

Bai, Jushan, Jiangtao Duan, and Xu Han (2024), “The likelihood ratio test for structural changes in factor models.” *Journal of Econometrics*, 238 (2), 105631.

Bai, Jushan, Xu Han, and Yutang Shi (2020), “Estimation and inference of change points in high-dimensional factor models.” *Journal of Econometrics*, 219 (1), 66–100.

Bai, Jushan and Serena Ng (2002), “Determining the Number of Factors in Approximate Factor Models.” *Econometrica*, 70 (1), 191–221.

Bai, Jushan and Serena Ng (2006), “Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions.” *Econometrica*, 74 (4), 1133–1150.

Bai, Jushan and Serena Ng (2009), “Boosting diffusion indices.” *Journal of Applied Econometrics*, 24 (4), 607–629.

Bai, Jushan and Serena Ng (2023), “Approximate factor models with weaker loadings.” *Journal of Econometrics*, 235 (2), 1893–1916.

Bailey, Natalia, George Kapetanios, and M. Hashem Pesaran (2021), “Measurement of factor strength: Theory and practice.” *Journal of Applied Econometrics*, 36 (5), 587–613.

Baltagi, Badi H., Chihwa Kao, and Fa Wang (2017), “Identification and estimation of a large factor model with structural instability.” *Journal of Econometrics*, 197 (1), 87–100.

Baltagi, Badi H., Chihwa Kao, and Fa Wang (2021), “Estimating and testing high dimensional factor models with multiple structural changes.” *Journal of Econometrics*, 220 (2), 349–365.

Banerjee, Anindya, Massimiliano Marcellino, and Igor Masten (2008), “Chapter 4 Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change.” In *Forecasting in the Presence of Structural Breaks and Model Uncertainty* (David E. Rapach and Mark E. Wohar, eds.), volume 3 of *Frontiers of Economics and Globalization*, 149–194, Emerald Group Publishing Limited.

Bates, Brandon J., Mikkel Plagborg-Møller, James H. Stock, and Mark W. Watson (2013), "Consistent factor estimation in dynamic factor models with structural instability." *Journal of Econometrics*, 177 (2), 289–304.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz (2005), "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach*." *The Quarterly Journal of Economics*, 120 (1), 387–422.

Boivin, Jean and Marc P Giannoni (2006), "Has Monetary Policy Become More Effective?" *The Review of Economics and Statistics*, 88 (3), 445–462.

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

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

Cheng, Xu, Zhipeng Liao, and Frank Schorfheide (2016), "Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities." *The Review of Economic Studies*, 83 (4), 1511–1543.

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

Duan, Jiangtao, Jushan Bai, and Xu Han (2022), "Quasi-maximum likelihood estimation of break point in high-dimensional factor models." *Journal of Econometrics*, S0304407622000379.

Fu, Zhonghao, Liangjun Su, and Xia Wang (2023), "Estimation and Inference on Time-Varying FAVAR Models." *Journal of Business & Economic Statistics*, 0 (0), 1–15.

Han, Xu (2024), "Global Identification, Estimation and Inference of Structural Impulse Response Functions in Factor Models: A Unified Framework."

Han, Xu and Atsushi Inoue (2015), "Tests for Parameter Instability in Dynamic Factor Models." *Econometric Theory*, 31 (5), 1117–1152.

Hansen, Bruce E. (2007), "Least Squares Model Averaging." *Econometrica*, 75 (4), 1175–1189.

Hansen, Bruce E. (2009), "Averaging Estimators for Regressions with a Possible Structural Break." *Econometric Theory*, 25 (6), 1498–1514.

Hansen, Bruce E. (2010), "Multi-step forecast model selection." In *20th Annual Meetings of the Midwest Econometrics Group*.

Koo, Bonsoo, Benjamin Wong, and Ze-Yu Zhong (2023), "Disentangling Structural Breaks in High Dimensional Factor Models." https://arxiv.org/abs/2303.00178v1.

Massacci, Daniele (2019), "Unstable Diffusion Indexes: With an Application to Bond Risk Premia." *Oxford Bulletin of Economics and Statistics*, 81 (6), 1376–1400.

Massacci, Daniele (2021), "Testing for Regime Changes in Portfolios with a Large Number of Assets: A Robust Approach to Factor Heteroskedasticity." *Journal of Financial Econometrics*.

Massacci, Daniele and George Kapetanios (2024), "Forecasting in factor augmented regressions under structural change." *International Journal of Forecasting*, 40 (1), 62–76.

McCracken, Michael W. and Serena Ng (2016), "FRED-MD: A Monthly Database for Macroeconomic Research." *Journal of Business & Economic Statistics*, 34 (4), 574–589.

McCracken, Michael W. and Serena Ng (2020), "FRED-QD: A Quarterly Database for Macroeconomic Research." Technical report.

Ng, Serena (2021), "Modeling Macroeconomic Variations After COVID-19."

Pelger, Markus and Ruoxuan Xiong (2022), "State-Varying Factor Models of Large Dimensions." *Journal of Business & Economic Statistics*, 40 (3), 1315–1333.

Pesaran, M. Hashem (2006), "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure." *Econometrica*, 74 (4), 967–1012.

Pesaran, M. Hashem, Davide Pettenuzzo, and Allan Timmermann (2006), "Forecasting Time Series Subject to Multiple Structural Breaks." *The Review of Economic Studies*, 73 (4), 1057–1084.

Pesaran, M. Hashem, Andreas Pick, and Mikhail Pranovich (2013), "Optimal forecasts in the presence of structural breaks." *Journal of Econometrics*, 177 (2), 134–152.

Stock, James H. and Mark W. Watson (1998), "Diffusion Indexes." Working Paper 6702, National Bureau of Economic Research.

Stock, James H. and Mark W. Watson (2002a), "Forecasting Using Principal Components From a Large Number of Predictors." *Journal of the American Statistical Association*, 97 (460), 1167–1179.

Stock, James H. and Mark W. Watson (2002b), "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business & Economic Statistics*, 20 (2), 147–162.

Stock, James H. and Mark W. Watson (2009), "Forecasting in Dynamic Factor Models Subject to Structural Instability." In *The Methodology and Practice of Econometrics: Festschrift in Honor of D.F. Hendry*, 65.

Stock, James H. and Mark W. Watson (2012), "Generalized Shrinkage Methods for Forecasting Using Many Predictors." *Journal of Business & Economic Statistics*, 30 (4), 481–493.

Su, Liangjun and Xia Wang (2017), "On time-varying factor models: Estimation and testing." *Journal of Econometrics*, 198 (1), 84–101.

Wan, Alan T. K., Xinyu Zhang, and Guohua Zou (2010), "Least squares model averaging by Mallows criterion." *Journal of Econometrics*, 156 (2), 277–283.

Wang, Yiru and Ruiqi Liu (2021), "Identification and Estimation of Parameter Instability in a High Dimensional Approximate Factor Model." Unpublished manuscript.

Zhang, Xiaomeng and Xinyu Zhang (2023), "Optimal model averaging based on forward-validation." *Journal of Econometrics*, 237 (2, Part C), 105295.