Dynamic Factor Models: Structural Breaks, Forecasting, and Structural Analysis

Ze-Yu Zhong

Bachelor of Actuarial Science, Bachelor of Commerce (Hons)

A thesis submitted for the degree of Doctor of Philosophy at Monash University in 2025

Department of Econometrics and Business Statistics

Faculty of Business and Economics

Copyright Notice

Abstract

This thesis contributes to the literature on dynamic factor models, specifically in the areas of structural breaks, forecasting, and structural analysis.

Chapter 1 provides a historical overview of dynamic factor models, tracing their evolution from strict classical foundations to more flexible modern approaches, and highlighting key unresolved issues in the field.

Chapter 2 addresses the challenge of disentangling structural breaks in dynamic factor models. Through a routine normalisation of the factor variance, standard methods for estimating factor models in macroeconomics do not distinguish between breaks of the factor variance and factor loadings, despite their markedly different economic interpretations. To address this, we develop a projection-based decomposition that separates the two and leads to two standard and easy-to-implement Wald tests to disentangle structural breaks. Applying our procedure to U.S. macroeconomic data, we find evidence of both types of breaks associated with the Great Moderation and the Great Recession. Through our projection-based decomposition, we estimate that the Great Moderation is associated with an over 60% reduction in the total factor variance, highlighting the relevance of disentangling breaks in the factor structure.

Chapter 3 explores factor-augmented forecasting subject to structural breaks in the factor structure. Building on the work of Chapter 2, we decompose any break in the factor loading matrix into rotational and shift components. To effectively utilise 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, alongside two competing factor estimators, in the presence of different types of breaks. To leverage the respective advantages of each factor estimator in a data-driven way, we introduce a leave-h-out cross-validation criterion for model averaging. 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 U.S. macroeconomic data and emphasise the importance of incorporating structural breaks into factor-augmented forecasting models.

Chapter 4 outlines a new framework for identifying and estimating dynamic causal effects in structural factor models with external instruments. Unlike traditional structural vector autoregressions, our approach is able to holistically deal with the problems of nonfundamentalness, singularity, and identification validity that plague traditional structural vector autoregression approaches. Using a generalised method-of-moments approach, we allow for the joint use of multiple instruments to sharpen inference, as well as develop overidentification tests for their joint validity and an automatic instrument selection procedure. Simulation results confirm the improvement in estimation accuracy of impulse response functions when more than one valid instrument is used, as well as the size and consistency of the tests and procedures. Applying this methodology to U.S. data, we estimate the effects of a monetary policy shock, showing that popular monetary policy instruments are jointly valid, and their joint use of popular monetary policy instruments leads to more accurate and reasonable estimates.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, or any use of generative artificial intelligence technologies, except where due reference is made in the text of the thesis.

Print Name: Ze-Yu Zhong

Date: March 19, 2025

Acknowledgements

I acknowledge the input of my supervisors Bonsoo Koo and Benjamin Wong for the conception and development of Chapter 2, and their support for the entirety of the thesis.

I acknowledge the input of Xu Han for the conception and development of Chapter 3, and their various comments for the entirety of the thesis.

I 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. I acknowledge the financial support of the Australian Research Council under Grants DP210101440, DP240100970, and DE200100693.

I thank the members of my panel, Catherine Forbes, Xibin Zhang, Bin Peng, and Natalia Bailey for their comments and support.

I thank members of my fellow PhD cohort, including Patrick Li, Fin Yang, Sherry Zhang, and Cynthia Huang, for their friendship and support throughout my journey.

I acknowledge the use of ChatGPT for proofreading and copy-editing. The output from ChatGPT was used to catch minor typographical and grammatical errors.

I acknowledge the assistance provided by Janelle Cruickshank in proofreading the Thesis.

Statement of Contribution

I declare that I was the sole contributor to Chapters 1, 4, and 5.

I declare that I was the main contributor to Chapter 2. Bonsoo Koo assisted in contributing the main projection decomposition idea, editing theoretical proofs, and subsequent editing throughout. Benjamin Wong contributed to subsequent re-writing, particularly in relation to the empirical study.

I declare that I was the main contributor to Chapter 3. Xu Han assisted in contributing with the asymptotic theory, and subsequent re-writing and editing throughout.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Factor Models

With the advancement of informational technology, researchers and practitioners now have an ever-increasing amount of data across more series and longer time spans. Thus, the analysis of large dimensional methods garnered significant attention from theoretical and empirical researchers alike. A wide array of tools is now available for studying large datasets: a non-exhaustive list includes factor models, shrinkage and selection methods and high-dimensional Bayesian vector-autoregressions. Among these, factor models have remained popular due to their relative computational simplicity and strong empirical performance.

We begin by formulating and motivating the basics of factor models. A factor model assumes that a large panel of time series can be effectively summarised by a small number of unobserved factors. Specifically, let N and T denote the cross-sectional and time dimensions, respectively, both of which are allowed to be large. A factor model for x_{it} is given as

$$x_{it} = \lambda_i^\top F_t + e_{it}, \tag{1.1.1}$$

$$= C_{it} + e_{it}, \tag{1.1.2}$$

where F_t are r dimensional unobserved factors, λ_i are the corresponding individual loadings (i.e. weights on each factor), and e_{it} are the idiosyncratic errors. If the errors e_{it} exhibit mild serial and cross-sectional correlation, then what arises is termed the *approximate* factor model. The term C_{it} is commonly referred to as the *common* component, which can be interpreted as the signal component in each series. Such factor models are commonly used in macroeconomics and finance. For example, in economics, the term x_{it} could represent a series such as GDP, where F_t are interpreted as the common shocks to the economy, and e_{it} are the leftover idiosyncratic shocks. Alternatively, in finance, x_{it} could denote the (risk-free) return for asset i in period t, in which case F_t is a vector of systematic risks, λ_i represents the risk exposures, and e_{it} are the leftover idiosyncratic returns.

The only observed quantities in a factor model are the data x_{it}. The factors, loadings, and the number of factors are all unobserved and must be estimated.

The literature on large dimensional factor models is extensive. These models have their origin in the statistics literature as so-called "classical" factor models. Assuming that F_t and e_{it} are uncorrelated and have zero mean, the covariance structure of the static model is

$$\Sigma = \Lambda\Lambda^\top + \Omega, \tag{1.1.3}$$

where Σ and Ω are the $N \times N$ covariance matrices of X_t and e_t, respectively, with the normalisation $E\left(F_t F_t^\top\right) = I_r$ imposed. If Ω is diagonal, then Equation (1.1.1) is the *strict* factor model of Chamberlain (1983). More generally, the properties of Equation (1.1.1) with the assumptions that i) e_t is i.i.d. over t, ii) N is fixed as $T \to \infty$ (or vice versa), and iii) both F_t and e_t are normally distributed are well understood (see Lawley and Maxwell, 1962; Anderson and Rubin, 1956; Anderson, 2003). Despite these "classical" factor models requiring restrictive assumptions, extensions to incorporate the serial correlation prevalent in economic data were pioneered by Geweke (1977) and Sargent and Sims (1977). The "dynamic factor model," now prevalent in the literature, differs from these classical approaches by relaxing these three assumptions.

Perhaps confusingly, the "dynamic factor model" as formulated Equation (1.1.1) specifies a *static* relationship between x_{it} and F_t, and the system is "dynamic" in the sense that the factors F_t are allowed to evolve according to a dynamic vector process

$$A(L)F_t = u_t, \tag{1.1.4}$$

where $A(L)$ is a lag polynomial (of possibly infinite order). More accurately, Equation (1.1.1) is the *static formulation* of the truly *dynamic* factor model, written as

$$x_{it} = \lambda_i(L)^\top f_t + e_{it}, \tag{1.1.5}$$

where $\lambda_i(L) = (1 - \lambda_{i1}L - \ldots \lambda_{is}L^s)$ is a vector of dynamic factor loadings of order s corresponding to *dynamic* factors

$$f_t = C(L)\epsilon_t, \tag{1.1.6}$$

where both f_t are *dynamic factors* with innovations ϵ_t, both q dimensional. The term "dynamic factor model" ambiguously refers to both the static and dynamic formulations, due to the fact that the dynamic formulation can always be rewritten in static form as

$$x_{it} = \lambda_i(L)^\top C(L)\epsilon_t + e_{it}. \tag{1.1.7}$$

There exist two main theoretical frameworks that have been developed for dynamic factor models: 1) the principal components estimator of Stock and Watson (1998) motivated from the time domain and assumes a finite lag order s, and 2) the *generalised* principal components estimator of Forni et al. (2000) (also referred to as "dynamic principal components") motivated from the frequency domain and allows s to be infinite. The focus of this Thesis is on the former, as the properties of the estimated static factors are generally better understood from a theoretical standpoint.

Collectively, these factor models are also known as "large approximate factor models": they are "large" because they explicitly allow for both N and T to approach infinity; and they are "approximate" because they allow for the idiosyncratic error term to be *weakly* correlated across cross-sectionally and temporally. Following these seminal works, broadly, three distinct branches have been developed within the approximate factor model framework, to which the following three main chapters of this Thesis contribute.

The first branch in dynamic factor models is the introduction of structural breaks. Among macroeconomic and financial data, parameter instability is a pervasive problem; a non-exhaustive list of causes includes technological change, policy regimes, changes in data collection methodology, or simply more (less) fortuitous shocks hitting the economy. Such instability, whatever its source, can manifest itself as structural breaks in econometric models. Within dynamic factor models, the literature has typically identified periods such as the Great Moderation (Stock and Watson, 2009; Breitung and Eickmeier, 2011; Baltagi et al., 2021), Great Recession (Stock and Watson, 2012a; Barigozzi and Trapani, 2020; Baltagi et al., 2021; Ma and Su, 2018), and the COVID-19 Pandemic (Bai et al., 2024) as structural breaks in the factor structure. However, the literature in this area has largely overlooked some of the unique challenges presented by dynamic factor models when analysing structural breaks. Specifically, the estimation of a dynamic factor model requires normalising the variance in either the factors or loadings; this normalisation then necessarily subsumes the breaks in one or the other. The literature at large has opted to normalise the factor variance and subsequently interpret the aforementioned periods as breaks in the factor loadings. We find this interpretation difficult to reconcile with the intuition from the broader macroeconomic literature, which has typically emphasised changes in the *variances* of all series as a stylised fact, particularly during the Great Moderation. It is, therefore, unclear whether the breaks being reported by the literature were legitimate breaks, or simply factor heteroskedasticity, a sentiment echoed by Stock and Watson (2016).

Chapter 2, *Disentangling Structural Breaks in Factor Models for Macroeconomic Data* (co-written with Bonsoo Koo and Benjamin Wong), focuses on disentangling the precise source of structural breaks in dynamic factor models. Our contribution is a projection-based decomposition to disentangle structural breaks in the factor variance and the loadings in dynamic factor models. At a high level, the projection-based decomposition reparameterises structural change in the factor structure into a rotational and orthogonal component where each has a natural interpretation as a change in the factor covariance matrix and a change in the factor loadings, respectively. This interpretation of the orthogonal component arises from recognising that breaks in the loadings are orthogonal to the original factor space, and can therefore result in more factors being estimated over the whole sample if ignored. At the same time, rotations can be thought of as some suitable twisting or stretching of the factor space that do not result in more factors appearing if ignored, and so are associated with the factor variance. Once one recognises this insight, the reparameterisation naturally leads to two easy-to-implement structural break tests: (i) a test for a break in the factor covariance matrix, and (ii) a test for a break in the factor loadings. We show that these test statistics have standard chi-squared distributions and reasonable finite sample performance. In an empirical application with U.S. macroeconomic data, our tests detect the Great Moderation as a break in the factor covariance matrix where, through our projection-based decomposition, we estimate an over 60% reduction in the total variance of the factors. Our results complement and reconcile with broader factor model work that associates large breaks in factor loadings with events such as the Great Moderation (e.g. Baltagi et al., 2021), suggesting a more nuanced interpretation of breaks in factor models.

The second branch in the dynamic factor model literature is the use of factors in augmented forecasting models. The possibility of using factors in forecasting was first introduced by Stock and Watson (1998, 2002a) as "diffusion index" forecasts. Since then, factor-augmented forecasts remained popular, being able to produce competitive forecasts at little computational cost and complexity; they are thus often regarded as the prevailing benchmark in macroeconomic forecasting.

In a similar vein to Chapter 2, the majority of the theory underpinning factor-augmented forecasting generally assumes parameter stability, when in reality macroeconomic data is often subject to instability; well-documented episodes include the Great Moderation of the 1980s and the Global Financial Crisis. The consequent breaks in the factor structure which these periods induce can therefore undermine the predictive power and reliability of factor-augmented forecasting.

Of course, the problem of forecasting with structural breaks present is well studied. Pesaran et al. (2006, 2013) are representative papers in this literature, and usually work with the case of observed factors. In this case, introducing suitable breaks in the forecasting equation itself (i.e. variations of ignoring the pre-break data) is sufficient. However, in the case of factor-augmented forecasting, both the regression coefficients and the factor structure itself are subject to possible breaks, and the precise impacts of breaks in the factor structure on forecasting has received significantly less attention. Stock and Watson (1998); Bates et al. (2013) show that the principal components estimator of factors is robust to *small* breaks, but this is evidently violated by periods of larger instability. When a *large* break occurs, the dimensionality of the factor structure tends to increase, which leads to breaks in both the moments of the factors, and possibly breaks in the forecasting regression, (Baltagi et al., 2021). The analysis of the effects of structural breaks in the factor structure is further convoluted by the fact that different *types* of breaks can occur, due to the high dimensionality of the system. As documented in Chapter 2, breaks in the factor structure can be classified and interpreted according to whether they occur in the factors or loadings; there is yet to be any literature that formally investigates their respective effects on forecasting, and in relation to their magnitudes.

Chapter 3, *Factor-augmented Forecasting Subject to Structural Breaks in the Factor Structure* (co-written with Xu Han), aims to fill this gap. We extend the projection decomposition of Chapter 2 and model the post-break loading matrix as 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 L2 distance between the pre- and post-break loading matrices. This ensures that the pre- and post-break factor estimates are subject to the same rotational basis asymptotically. As a result, the pre- and post-break factor estimators become compatible and can be directly combined to utilise the pre-break data, and can thus lead to improved forecasting performance by avoiding the potentially significant bias-variance trade-off of a traditional split-sample approach.

The third branch in dynamic factor models is the extension of factor models to structural analysis in order to improve identification and inference of dynamic causal effects. Traditionally, the identification and inference of dynamic causal effects has largely been done using a structural vector-autoregression (SVAR) framework, following the seminal paper of Sims (1980). However, SVARs suffer from three distinct problems: 1) nonfundamentalness due to their limited ability to include a large set of variables, 2) covariance singularity as a result of the number of variables being driven by a smaller number of shocks, and 3) the validity of the identification strategy being employed. Any of these issues could preclude the practitioner from recovering the true impulse responses, and ideally would be addressed simultaneously. However, the existing literature has typically addressed these issues in separate and distinct ways. The issue of nonfundamentalness is a direct consequence of the lack of information in SVAR models. The possibility of adding estimated latent factors to alleviate this informal insufficiency in a factor-augmented vector autoregression (FAVAR) was first introduced by Bernanke et al. (2005). However, the addition of extra variables in an SVAR, whether they be observed or estimated latent factors (as in a FAVAR), necessarily increases the implicit number of shocks, and therefore the chance that the specified system is singular.

To address this, structural factor models (SFMs) were introduced by Stock and Watson (2005, 2012a), which, in comparison to a typical structural vector autoregression, are explicitly formulated with the assumption of a dynamic factor structure. These models naturally deal with both the nonfundamentalness and singularity issues that plague SVARs. However, the additional use of factors requires many adaptations to the estimation procedure and theoretical results, due to the introduction of many nuisance parameters as a result of the unobserved nature of the factors. This is especially the case given that estimation and theoretical results typically need to be established on a case by case basis according to the choice of identification. These include short/long-run restrictions, heteroskedasticity, sign restrictions and, more recently, the use of external instruments, for which we provide an overview below in the context of factor models.

Some of the earliest forms of identification are the use of short- and long-run exclusion restrictions, which still remain popular. These restrictions achieve identification by restricting the impulse responses of some variables to be zero; it is for this reason that they are alternatively known as "zero" restrictions. As one of the oldest approaches to achieving identification, there have been many developments in this area that use factor models. Indeed, the paper of Bernanke et al. (2005), which introduced the FAVAR, relies on fast-slow identification wherein "slow" variables such as industrial production are assumed to not contemporaneously response to "fast" shocks such as monetary policy variables in order to achieve identification. This approach was later formalised by Yamamoto (2019), who studies structural IRFs in a FAVAR context using Cholesky decomposition with either short- or long-run restrictions. In the structural factor models, these zero restrictions can be naturally parameterised as zero restrictions on the loading matrix. Due to the loading matrix being a high dimensional matrix of factor loadings, these zero restrictions can be set up to result in a system that is just identified (Forni and Gambetti, 2010a; Bai and Wang, 2012), over identified with a fixed number of restrictions (Han, 2018), or possibly highly overidentified with a diverging number of restrictions (Stock and Watson, 2005; Han, 2015). However, the zero assumptions that underpin these approaches are controversial, and have fallen out of favour. As noted by Ramey (2016), they are often difficult to justify, and are at odds with the impulse response estimated by New Keynesian Dynamic Stochastic General Equilibrium (DSGE) models. For example, the model of Smets and Wouters (2007) implies that output, hours and inflation should have contemporaneous responses to monetary policy shocks, in comparison to the DSGE model of Christiano et al. (2005), which assumes that agents cannot react to contemporaneous monetary policy shocks and implies no immediate responses. More recently,

Identification via heteroskedasticity (Rigobon, 2003) is an alternative approach that also imposes a minimal set of identifying restrictions. However, this requires the practitioner to correctly specify and model heteroskedasticity in the shock variances, an issue that is particularly difficult to address in factor models due to the unobserved nature of the factors and loadings (Koo et al., 2023). Indeed, Yamamoto and Hara (2022) formally extend this identification approach to FAVAR models by assuming that the shock variances are the only change in the factor (co)variance and, therefore, factor structure; they nevertheless concede that this assertion is questionable. Indeed, the literature on structural breaks in factor models has typically identified evidence of breaks in the factor loadings and factor covariance matrix (i.e. off-diagonals of the factor covariance matrix) around periods associated with factor heteroskedasticity, (e.g. Stock and Watson, 2012a; Koo et al., 2023).

Sign restrictions were initially proposed by Uhlig (2005) as a way to directly impose non-puzzling responses as an identification condition. Generally, a finite number of sign restrictions is only able to achieve set identification. Most commonly, a Bayesian approach is pursued to deal with this partial identification problem in order to impose prior distributions on the parameters of the model, (Uhlig, 2005). This method, however, has notable drawbacks. Granziera et al. (2018) show that Bayesian credible intervals exhibit incorrect coverage rates under partial identification. Additionally, in the Bayesian context, the resulting estimates are potentially sensitive to the choice of prior; Giacomini and Kitagawa (2021) address this problem, but the resulting robust Bayesian credible intervals tend to be uninformatively wide in practice.

In the context of factor models, sign restrictions have also seen some use, albeit without theoretical justification, (see Eickmeier, 2009; Forni and Gambetti, 2010b,c; Luciani, 2015). Gafarov (2014) focuses on the possibility of achieving identification using sign restrictions and provides the necessary and sufficient conditions for point identification in SVAR and SFM models. More generally, there is no formal asymptotic theory for this identification strategy in the context of factor models.

Among the various identification schemes, only the use of zero/contemporaneous timing restrictions has been extensively studied (see e.g. Forni et al., 2009; Han and Inoue, 2015; Han, 2018). However, these have proven to be controversial; in particular, they are at odds with the responses as later estimated by New Keynesian DSGE models, and have thus fallen out of favour (Ramey, 2016). Fortunately, the macroeconomics literature has developed many excellent measures (or proxies) of structural shocks, (e.g. Romer and Romer, 2004; Gertler and Karadi, 2015; Barakchian and Crowe, 2013; Miranda-Agrippino and Ricco, 2021). Compared to existing approaches, which require structural restrictions which may be difficult to justify, the use of external instruments only requires the typical instrument validity conditions, and is thus in some sense considered to be a minimal set of identifying restrictions. The use of external instruments (or external "proxy" measures of underlying structural shocks) in conjunction with a factor model has not yet been formalised.

Chapter 3, *Identification and Estimation of Structural Factor Models with External Instruments* (sole authored), focuses on the estimation of dynamic causal effects in structural factor models with the use of external instruments. Our framework is designed to holistically address the challenging issues of nonfundamentalness, covariance singularity, and testing of identifying restrictions, which often plague traditional SVAR approaches. Our use of a factor structure addresses the nonfundamentalness problem similar to FAVAR models; unlike FAVARs, we explicitly distinguish between the static and dynamic factors and are able to avoid the issues associated with a singular covariance matrix. Identification is achieved using multiple instruments in a generalised method-of-moments approach, which naturally allows for the joint use of multiple instruments to sharpen inference, and overidentification tests to test the validity of the restrictions, and a moment selection procedure to ensure that the correct instruments are chosen. Simulation results confirm the finite sample performance of the proposed estimators and procedures. We apply the proposed methodology to study the effects of a monetary policy shock using a large quarterly U.S. macroeconomic dataset with the use of popular monetary policy instruments proposed by the literature. We find that all monetary policy instruments are valid, and that their joint use can result in more accurate and reasonable estimates of the impulse responses.

Together, these three Chapters make separate, but related contributions to the distinct areas of structural breaks, forecasting, and structural analysis within the dynamic factor model literature.

Chapter 2

Disentangling Structural Breaks in Factor Models for Macroeconomic Data

2.1 Introduction

Dynamic factor models are increasingly used in empirical macroeconomics and finance (e.g. Aastveit et al., 2015; Alessi and Kerssenfischer, 2019; Barigozzi and Luciani, 2023) as a form of dimension reduction, summarising the dynamics of a large set of time series through a small number of factors. There is growing evidence of structural instability in U.S. macroeconomic time series (see, e.g. Breitung and Eickmeier, 2011; Chen et al., 2014; Stock and Watson, 2016). The analysis of structural changes in factor models presents a unique challenge because breaks in the loadings and breaks in the factor variance cannot be easily disentangled. For simplicity, consider the following representative factor model common in applied macroeconomic work for $x_{it}, t = 1, \ldots, T, i = 1, \ldots, N$:

$$x_{it} = \lambda_i^\top f_t + e_{it}, \tag{2.1.1}$$

where λ_i is an $r \times 1$ vector of individual loadings, f_t is an $r \times 1$ vector of factors, and e_{it} is noise. Because both the factors and loadings are unobserved and enter multiplicatively, a normalisation is needed to separately identify them; this is often done on the variance of f_t.[1] While such normalisations are often innocuous, they matter for studying structural changes in factor models; if changes in the factor variance are ruled out through the normalisation, these changes must manifest as changes in the factor loadings even if the loadings are stable. We note similar concerns have previously been raised (see Stock and Watson, 2016).[2]

Our contribution is a projection-based decomposition to disentangle structural breaks[3] in the factor variance and the loadings in dynamic factor models. At a high level, the projection-based decomposition reparameterises structural change in the factor structure into a rotational and orthogonal component where each has a natural interpretation as a change in the factor covariance matrix and a change in the factor loadings, respectively. This interpretation of the orthogonal component arises from recognising that breaks in the loadings are orthogonal to the original factor

[1] Other normalisations exist, (see Bai and Wang, 2016), but still serve the purpose of pinning down the scale of one quantity (e.g. the factor variance or the loadings) in order to identify the other.

[2] We argue such concerns are empirically relevant because the literature has typically identified periods such as the Great Moderation, the Great Recession and, more recently, the COVID-19 Pandemic as evidence of structural breaks, all periods well known for the data displaying heteroskedasticity (e.g. Breitung and Eickmeier, 2011; Baltagi et al., 2021; Bai et al., 2024).

[3] We use the term "structural break" to refer to a discrete change along the time dimension.

space, and can therefore result in more factors being estimated over the whole sample if ignored. At the same time, rotations can be thought of as some suitable twisting or stretching of the factor space that do not result in more factors appearing if ignored, and so are associated with the factor variance.[4] Once one recognises this insight, the reparameterisation naturally leads to two easy-to-implement structural break tests: (i) a test for a break in the factor covariance matrix, and (ii) a test for a break in the factor loadings. We show that these test statistics have standard chi-squared distributions and reasonable finite sample performance.

Disentangling breaks in the factor variance and loadings is not a mere technical curiosity. From the perspective of the dynamic factor model, these breaks imply vastly different interpretations. While breaks in the loadings relate to changes in how variables relate to the factors, breaks in the factor variance imply breaks in the factor dynamics. These breaks in the factor dynamics can be in the form of a break in the dynamic process generating the factors and/or breaks in the variance of the underlying shocks to the factors. Confining breaks to just the loadings would therefore *a priori* preclude breaks in the factor dynamics as a possible interpretation. While the degree of how misleading such an interpretation is depends on context, existing applied work suggests that breaks in the factor dynamics present a more natural interpretation of at least one historical episode: the Great Moderation. Indeed, in an empirical application with U.S. macroeconomic data, our tests detect the Great Moderation as a break in the factor covariance matrix, where, through our projection-based decomposition, we estimate an over 60% reduction in the total variance of the factors.[5] While this finding should be unsurprising to applied macroeconomists, it nevertheless is an effective proof-of-concept underpinning our basic argument: only by disentangling breaks in factor variances and loadings can one attribute a break in the factor variance as part of the most well-known change in volatility common across multiple macroeconomic time series. Although we still find evidence of breaks in loadings even when controlling for breaks in the factor variance, our results complement and reconcile with broader factor model work that associates large breaks in factor loadings with events such as the Great Moderation (e.g. Baltagi et al., 2021), suggesting a

[4] We also briefly touch on how the case of disappearing/emerging factors can be accommodated by a singular rotation.

[5] Henceforth, we refer to a break in the factor variance as a break in the factor covariance matrix; if we refer to a change in the "total factor variance" (i.e. the trace of the factor covariance matrix), we will make this explicit.

more nuanced interpretation of breaks in factor models.

While there are different methods to estimate dynamic factor models and forms of structural instability, our work is most closely related to work that tests for structural breaks in factor models using the principal components estimator (see Stock and Watson, 2009; Breitung and Eickmeier, 2011; Chen et al., 2014; Han and Inoue, 2015; Baltagi et al., 2017). Given breaks in either the factor variance and/or loadings will manifest as breaks in the loadings in these tests, one could first test for breaks in the factor structure using one of these aforementioned procedures to date the break date, and subsequently use our procedure to disentangle the break. Notable exceptions include the tests of Pelger and Xiong (2022) and Massacci (2021), which are tests for breaks in the factor loadings robust to factor heteroscedasticity. Our test for a break in the factor loadings differs from these in several aspects: compared to Pelger and Xiong (2022), we do not require an estimate of the covariance of the idiosyncratic error term that is only possible by assuming sparsity; and compared to Massacci (2021) who works with a threshold model setup our test statistic is under a structural break setup, it is additionally estimated under the alternative as opposed to the null hypothesis. Furthermore, neither consider separately testing for a break in the factor variance.

We proceed as follows. In Section 2, we first demonstrate how breaks in the factor variance and loadings manifest through an underlying dynamic factor model before introducing our projection-based decomposition. Section 3 presents the theory underpinning our tests followed by Monte Carlo simulations in Section 4. Section 5 presents an empirical application with 124 quarterly U.S. macroeconomic time series. Section 6 concludes.

2.2 Interpreting and Reparameterising Structural Breaks in Dynamic Factor Models

To start, consider the representative dynamic factor model (see Stock and Watson, 2016)

$$X_t = \Lambda f_t + e_t \tag{2.2.1}$$

$$f_t = \sum_{j=1}^{p} \Phi_j f_{t-j} + \eta_t, \quad \eta_t \sim (0, \Sigma_\eta), \tag{2.2.2}$$

where Equation (2.2.1) stacks Equation (2.1.1) across the cross-section, such that X_t and e_t are $N \times 1$, $\Lambda = [\lambda_1, \ldots, \lambda_N]^\top$ is $N \times r$. Equation (2.2.2) describes the dynamics of the factors, where Φ_j are autoregressive coefficients for f_t, and η_t are $q \times 1$ (reduced form) innovations with covariance Σ_η.[6] Equations (2.2.1) and (2.2.2) describe the static form of the dynamic factor model, and clarify how changes in the factor variance and loadings imply vastly different interpretations of the dynamic factor model. In particular, the (unconditional) covariance matrix of f_t is both a function of the Φ_j's and Σ_η. Therefore, breaks in the factor variance require a break in Equation (2.2.2), either in the Φ_j's, Σ_η, or both. Breaks in the λ_i's, on the other hand, are isolated to breaks in the relationship between the different variables and the factors in Equation (2.2.1). As an illustrative example, consider the Great Moderation, an event marked by a reduction of volatility across many macroeconomic variables. Such an interpretation, from the perspective of the dynamic factor model, is more naturally accommodated as a reduction in the total variance (i.e. trace of the covariance matrix) of f_t rather than multiple (proportional) breaks in the λ_i's.

Because the factors are identified up to rotation, one needs to impose some form of normalisation. A common approach is to use the principal components estimator of the factors $\tilde{F} = [\tilde{f}_1, \ldots, \tilde{f}_T]^\top$ and loadings $\tilde{\Lambda} = [\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N]^\top$, which imposes

$$\frac{1}{N}\tilde{\Lambda}^\top\tilde{\Lambda} = V_{NT}, \quad \frac{1}{T}\tilde{F}^\top\tilde{F} = I_r, \tag{2.2.3}$$

where V_{NT} is a diagonal matrix whose entries are the first r eigenvalues of the covariance matrix of X. More generally, there exist many different estimation methods and thus normalisations, but our point still holds as long as one needs to impose a normalisation for estimation. Of these, it is known that the method of principal components is able to consistently estimate the space spanned by the dynamic factors under very general conditions. Therefore, the principal components estimator can be used to estimate both λ_i and f_t, and the fitting and specification of Equation (2.2.2) can occur as a separate step (see Stock and Watson, 2016, for more details). The normalisation one applies for estimation convolutes the interpretation of breaks - structural break tests in the factor loadings often first estimate the factors on the full sample, then test for breaks in $\tilde{\lambda}_i$ (e.g. Stock and

[6] Relative to Stock and Watson (2016), we focus on the case where the number of static and dynamic factors are equal, but this does not result in a loss of generality for the discussion.

Watson, 2009; Breitung and Eickmeier, 2011; Chen et al., 2014; Bai and Han, 2016). The necessary application of normalisations like Equation (2.2.3) when estimating the factors thus makes it unclear whether finding a break in $\tilde{\lambda}_i$ is a break in λ_i, or a break in the variance of factors, since the latter is typically assumed to be unchanging as an identification condition as part of constructing the test statistic, (see Stock and Watson, 2016; Chen et al., 2014).

We note that if the goal is to establish whether there are breaks in the factor *structure* (i.e. the loadings or the factor variance), current methods of normalising the variance, and then finding breaks in the loadings are probably appropriate, since any breaks in the factor variance are subsumed into the factor loadings. However, if one wanted to appropriately interpret breaks, especially from the perspective of the dynamic factor model implied by Equations (2.2.1) and (2.2.2), it becomes important to distinguish between breaks in the loadings and breaks in the factor variance. In what follows, we present a reparameterisaton that aids in disentangling these breaks.

2.2.1 A Projection-based Decomposition to Disentangle Structural Changes

We now introduce structural changes[7] to the dynamic factor model in Equations (2.2.1) and (2.2.2). We only work with Equation (2.2.1) in what follows since the fitting of Equation (2.2.2) can occur separately, and we can consistently estimate the space spanned by the static factors. Let k_t denote an indexing variable, which partitions x_{it} into two regimes at some break point k

$$X_t = \begin{cases} \Lambda_1 f_t + e_t, & \text{for} \quad k_t \leq k, \\ \Lambda_2 f_t + e_t, & \text{for} \quad k_t > k, \end{cases} \tag{2.2.4}$$

where f_t is an $r \times 1$ vector of factors, $\Lambda_1 = (\lambda_{1,1}, \ldots, \lambda_{1,N})^\top$ and $\Lambda_2 = (\lambda_{2,1}, \ldots, \lambda_{2,N})^\top$ are $N \times r$ pre- and post-break loadings, and e_t is noise. In practice, both the number of factors r and the break k can be consistently estimated (or chosen *a priori* if desired), and thus are treated as known throughout the remainder of the paper.

By the formulation in Equation (2.1.1), any changes in the factor variance must be common

[7]We use the term "structural changes" to refer to any kind of instability in the factor structure, including but not limited to threshold models, Markov-based regime switching models where the regimes are identified with probability one, or structural break models.

across all series, whereas changes in the loadings are by definition idiosyncratic. This motivates us to disentangle breaks in the factor loadings from breaks in the factors, by decomposing the change in Λ_2 via a projection

$$\Lambda_2 = \Lambda_1 Z + W \tag{2.2.5}$$

where Z is an $r \times r$ rotational change, and $W = (w_1, \dots, w_N)^\top$ is an $N \times r$ orthogonal shift satisfying $\Lambda_1^\top W = O_p(1)$.[8] We briefly note that the rotational change Z can be a singular matrix if desired, which allows a new set of factors that "replace" the old factors, or a change in the number of factors where some factors disappear (if the rank of W is unchanged or decreases, respectively).[9] We elaborate on this case in the Appendix, but otherwise treat the number of factors to be unchanging, as per the extant literature (e.g. Su and Wang, 2017; Massacci, 2021). Thus, the rotation Z and orthogonal shift W are naturally associated with breaks in the factor covariance matrix and loadings, respectively. Heuristically, this is because a break in the factor variance can always be thought of some suitable twisting or stretching of the factors themselves, i.e. a mathematical rotation. Note that because breaks in Z are breaks in the covariance matrix of the factors, they encompass breaks in both their variances (i.e. the diagonal elements) and their correlations (off-diagonals). In contrast, a change in the loadings is idiosyncratic across series, and thus, geometrically, must lie outside and be orthogonal to the space spanned by the factors, (see Wang and Liu, 2021; Pelger and Xiong, 2022; Massacci, 2021, for similar interpretations).

We emphasise that the projection-based decomposition can hold for *any* generic structural change in the factor structure. In what follows, we demonstrate how the reparameterisaton can be used to disentangle a one-time structural *break*, and naturally lead to test statistics that can disentangle changes in the factor variance and factor loadings.

[8]This implies a rate of $\Lambda_1^\top W/N = O_p\left(\frac{1}{N}\right)$, a definition looser than *exact* orthogonality $\Lambda_1^\top W = 0$ considered by Wang and Liu (2021); Pelger and Xiong (2022); Massacci (2021). It can be further loosened to *uncorrelatedness* i.e. $E[\lambda_{1i}^\top w_i] = 0$ with more complicated conditions on the rates between T and N.

[9]The case of disappearing factors requires more careful specification of the model, which we detail in Section A.1.5.

2.2.2 Structural Break Setup

We apply the projection decomposition in a structural break setup. Conditional on the break fraction π satisfying $k = \lfloor \pi T \rfloor$, which splits the data into two partitions of length $T_1 = \lfloor \pi T \rfloor$ and $T_2 = T - \lfloor \pi T \rfloor$, Equation (2.2.4) can stacked in matrix form:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} F_1 \Lambda_1^\top \\ F_2 \Lambda_2^\top \end{bmatrix} + \begin{bmatrix} e_{(1)} \\ e_{(2)} \end{bmatrix} \tag{2.2.6}$$

where $F_1 = (f_1, \dots, f_{T_1})^\top$ are $T_1 \times r$ pre-break factors, $F_2 = (f_{T_1+1}, \dots, f_T)^\top$ are $T_2 \times r$ post-break factors, Λ_1, Λ_2 are $N \times r$ their respective loadings, $e_{(1)} = (e_1, \dots, e_{T_1})$ and $e_{(2)} = (e_{T_1+1}, \dots, e_T)$, and X_1, X_2 denote the respective partitions of X. By substituting Equation (2.2.5) into Equation (2.2.6), we can formulate an equivalent representation as follows:

$$X = \begin{bmatrix} F_1 \Lambda_1^\top \\ F_2 \left[\Lambda_1 Z + W\right]^\top \end{bmatrix} + \begin{bmatrix} e_{(1)} \\ e_{(2)} \end{bmatrix} \tag{2.2.7}$$

$$= \begin{bmatrix} F_1 & 0 \\ F_2 Z^\top & F_2 \end{bmatrix} \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + \begin{bmatrix} e_{(1)} \\ e_{(2)} \end{bmatrix}$$

$$X = G \Xi^\top + e. \tag{2.2.8}$$

Equation (2.2.8) shows that any rotational changes induced by a non-identity Z are absorbed into the factors, and any orthogonal shifts W will augment the factor space. Equation (2.2.8) re-expresses a factor model with structural breaks in its loadings into an observationally equivalent model with time invariant loadings. Equation (2.2.8) highlights that if one were to ignore the break and use the principal components estimator over the whole sample, the estimator will be consistent for an observationally equivalent model with *pseudo*-factors G and time invariant loadings Ξ. Our formulation aims to complement similar formulations in the literature used to identify and estimate the break point (see, e.g. Han and Inoue, 2015; Baltagi et al., 2017).

Existing methods use an estimate of the pseudo-factors G in order to either test for existence of any breaks (e.g. Han and Inoue, 2015; Chen et al., 2014), and/or estimate the break fraction (e.g. Baltagi et al., 2017, 2021; Duan et al., 2022). However, the structure of G implied by Equation (2.2.8) implies that breaks in either Z *or* W will induce a structural break in the pseudo-factors G: when Z is non-identity the first r columns of G will correspond to the multivariate series $\left[F_1^\top, ZF_2^\top\right]^\top$; when $W \neq \mathbf{0}$, the orthogonality of W will induce the last r columns of G to be $\left[0, F_2^\top\right]^\top$, corresponding to a structural break where extra factors appear. Therefore, in either case G will exhibit a break, and methods utilising the estimated pseudo-factors will necessarily have power against breaks in the factors variance, even if the loadings are time invariant.

The case of a rotational break corresponds to $Z \neq I_r$, and can be naturally interpreted as a change in the factor variance. Indeed, by assuming $\Sigma_F = E(f_t f_t^\top)$, it follows from Equation (2.2.8) that the covariance matrix of the factors pre- and post-break are Σ_F and $Z\Sigma_F Z^\top$, respectively, which are in general different for non-identity Z.[10] Although the presence of a rotational break is indistinguishable from a break in the factor variance, and therefore poses no problem for the PC estimator, in the sense that it is still able to consistently recover the factor space, it is still nevertheless important to formally study this phenomenon, as the case of a break in factor variance is crucial for understanding certain empirical events, such as the Great Moderation in teh context of macroeconomic data.

Given that Z captures changes in the factor variance, it follows that the remaining orthogonal shift where $W \neq \mathbf{0}$ must correspond to breaks in the factor loadings.[11] Mechanically, these breaks lie outside the original factor space, which necessitates the estimation of more factors than necessary if one wants to capture all the information while ignoring the break. It is this orthogonality of breaks in the loadings that cause the "factor augmentation" effect in the pseudo-factors G raised by Breitung and Eickmeier (2011).

Disentangling these breaks naturally entails testing for changes in these parameters: a test for a break in the factor variance corresponds to $\mathcal{H}_0 : \Sigma_F = Z\Sigma_F Z^\top$, and a test for a break in the factor loadings corresponds to $\mathcal{H}_0 : W = \mathbf{0}$. In what follows, we develop these ideas more fully, but for now, note that our tests have standard chi-square distributions. We also note that our

[10]Our rotational changes correspond to the type 2 breaks of Han and Inoue (2015) and Baltagi et al. (2017), and the type B breaks of Duan et al. (2022).

[11]Such breaks correspond to the type 1 break as defined by Han and Inoue (2015), Baltagi et al. (2017), and type A break by Duan et al. (2022), as per respective nomenclatures.

test statistics aim to determine which type of break has occurred, for a candidate break provided either *a priori*, or estimated from the data. In the case of multiple breaks, practitioners can simply partition the data suitably and either separately or sequentially focus on different breaks, as we have in our empirical study in Section 2.5.

Finally, we note that, while we mainly use the reparameterisation to develop tests to disentangle changes in the factor variance and factor loadings, the reparameterisation should have broader applications and utility in modelling structural change in factor models.

2.2.3 Implementing the Test to Disentangle Structural Breaks

We outline how to implement our tests to disentangle breaks in the factor variance and loadings, relegating theoretical justifications to Section 2.3. We first condition on a known break fraction, which can be chosen *a priori* (e.g. Breitung and Eickmeier, 2011; Stock and Watson, 2012a), or obtained from a test such as Baltagi et al. (2021). The estimation of various quantities required for constructing the test statistics then follows as:

1. Estimate the pre- and post-break factors $\tilde{F}_1$ and $\tilde{F}_2$ via principal components, subject to the normalisations $\frac{1}{T_1}\tilde{F}_1^\top \tilde{F}_1 = \frac{1}{T_2}\tilde{F}_2^\top \tilde{F}_2 = I_r$, i.e. $\tilde{F}_1$ is $\sqrt{T_1}$ times the first r eigenvectors of $X_1 X_1^\top$, and $\tilde{F}_2$ is $\sqrt{T_2}$ times the first r eigenvectors of $X_2 X_2^\top$.

2. Conditional on the factors, estimate the pre- and post-break loadings via least squares as $\tilde{\Lambda}_1 = X_1^\top \tilde{F}_1 (\tilde{F}_1^\top \tilde{F}_1)^{-1} = \frac{1}{T_1} X_1^\top \tilde{F}_1$ and $\tilde{\Lambda}_2 = X_2^\top \tilde{F}_2 (\tilde{F}_2^\top \tilde{F}_2)^{-1} = \frac{1}{T_2} X_2^\top \tilde{F}_2$.[12]

3. Estimate the rotational change and orthogonal shift as

$$\tilde{Z} = (\tilde{\Lambda}_1^\top \tilde{\Lambda}_1)^{-1} \tilde{\Lambda}_1^\top \tilde{\Lambda}_2, \tag{2.2.9}$$

$$\tilde{W} = \tilde{\Lambda}_2 - \tilde{\Lambda}_1 \tilde{Z}. \tag{2.2.10}$$

The estimates $\tilde{Z}$ and $\tilde{W}$ absorb the effects of the normalisation bases present in $\tilde{\Lambda}_1$ and $\tilde{\Lambda}_2$, and hence cannot be tested directly. Instead, $\tilde{Z}$ and $\tilde{W}$ take on additional interpretations: post multiplying $\tilde{\Lambda}_1$

[12] This follows because $\tilde{F}_1$ and $\tilde{F}_2$ have unit variance.

by $\tilde{Z}$ rotates its normalisation basis to that of $\tilde{\Lambda}_2$ along with any rotational change Z; the remainder $\tilde{W}$ is the remaining idiosyncratic change.[13] We exploit this property in $\tilde{Z}$ to rotate the post-break factors onto the same basis as the pre-break factors, defined as $\widehat{F} = [\tilde{F}_1^\top, \tilde{Z}\tilde{F}_2^\top]^\top = [\widehat{f}_1, \ldots, \widehat{f}_T]^\top$. The variance of the combined series $\widehat{f}_t$ reflects the effect of any rotational change Z, and thus can be used as a basis for a test.

Test Statistics

We now construct the test statistics, which we label as the Z-test for breaks in the factor variance, and the W-test for breaks in the loadings following Equations (2.2.5) and (2.2.7).

We present the Z-test statistic for changes in the factor variance $\mathcal{H}_0 : \Sigma_F = Z\Sigma_F Z^\top$ as

$$\mathscr{W}_Z(\pi, \widehat{F}) = A_Z(\pi, \widehat{F})^\top \widehat{S}_Z(\pi, \widehat{F})^{-1} A_Z(\pi, \widehat{F}), \tag{2.2.11}$$

where

$$A_Z(\pi, \widehat{F}) = vech\left(\sqrt{T}\left(\frac{1}{\lfloor \pi T \rfloor}\sum_{t=1}^{\lfloor \pi T \rfloor} \widehat{f}_t \widehat{f}_t^\top - \frac{1}{T - \lfloor \pi T \rfloor}\sum_{t=\lfloor \pi T + 1 \rfloor}^{T} \widehat{f}_t \widehat{f}_t^\top\right)\right) \tag{2.2.12}$$

denotes the difference in the subsample means of the second moments (outer product) process of $\widehat{f}_t$ at a given break fraction π, and $vech()$ denotes the column-wise vectorisation of a square matrix with the upper triangle excluded. Its long run variance is estimated as a weighted average of the variance from pre- and post-break data

$$\widehat{S}_Z(\pi, \widehat{F}) = \frac{1}{\pi}\widehat{\Omega}_{Z,(1)}(\pi, \widehat{F}) + \frac{1}{1-\pi}\widehat{\Omega}_{Z,(2)}(\pi, \widehat{F}), \tag{2.2.13}$$

where $\widehat{\Omega}_{Z,(1)}$, $\widehat{\Omega}_{Z,(2)}$ are HAC estimators constructed using the respective subsamples of $vech\left(\widehat{f}_t \widehat{f}_t^\top - I_r\right)$. Alternatively, a bootstrap-based procedure to estimate the variance can be entertained. The statistic $\mathscr{W}_Z(\pi, \widehat{F})$ is a Wald test for whether the subsample means of the second moments process of $\widehat{f}_t$ are the same at the pre-specified break point,[14] and thus has a conventional χ^2 distribution with

[13] The case of a disappearing factor can be accommodated by estimating $\tilde{Z}$ as an $r_1 \times r_2$ "rectangular" matrix and $\tilde{W}$ as an $N \times r_2$ matrix, where r_1 and r_2, are the pre- and post-break number of factors such that $r_2 < r_1$.

[14] As noted by an anonymous referee, this statistic is similar to that constructed by Han and Inoue (2015); Baltagi

$r(r+1)/2$ degrees of freedom.

Next, we present the W-test statistics for changes in the loadings $\mathcal{H}_0 : W = \mathbf{0}$. Defining $\tilde{w}_i$ as the ith row of $\tilde{W}$, the individual statistic for the ith series and its variance are

$$\mathscr{W}_{W,i} = T\tilde{w}_i^\top \tilde{\Omega}_{W,i}^{-1}\tilde{w}_i, \tag{2.2.14}$$

$$\tilde{\Omega}_{W,i} = \frac{1}{\pi}\tilde{\Theta}_{1,i} + \frac{1}{1-\pi}\tilde{\Theta}_{2,i} \tag{2.2.15}$$

which uses pre and post break HAC estimates of the asymptotic variance $\tilde{\Theta}_{1,i}$ and $\tilde{\Theta}_{2,i}$ respectively.[15] Next, define the joint statistic for all variables as:

$$\mathscr{W}_W = (TN)\left(\frac{\sum_{i=1}^N \tilde{w}_i}{N}\right)^\top \tilde{\Omega}_W^{-1}\left(\frac{\sum_{i=1}^N \tilde{w}_i}{N}\right), \tag{2.2.16}$$

where the matrix $\tilde{\Omega}_W = N^{-1}\sum_{i=1}^N \tilde{\Omega}_{W,i}$ is an estimate of the joint variance. Both are Wald tests based on $\tilde{w}_i$, and thus have standard χ^2 distributions with r degrees of freedom.

We emphasise that our test statistics do not maintain any assumptions on their counterparts; that is, the test for $\Sigma_F = Z\Sigma_F Z^\top$ holds irrespective of W, and conversely the test for $W = \mathbf{0}$ holds irrespective of Z. As both breaks in the factor variance and the loadings could occur simultaneously, this necessitates the practitioner to run both the Z-test and W-test to accurately pin down the source(s) of the break. A straightforward Bonferroni-Holm correction suffices to correct for the family wise error rate (see Section 2.4).

2.3 Asymptotic Theory

We discuss the asymptotic theory underpinning the test statistics discussed in Section 2.2.3. All limits are taken as both N and T tend to infinity simultaneously, and δ_{NT} is defined as $\min(\sqrt{T}, \sqrt{N})$. For notation, $\|\cdot\|$ denotes the Frobenius norm of a vector or matrix, $\xrightarrow{p}$ denotes convergence in probability, $\Rightarrow$ denotes weak convergence of stochastic processes, $\xrightarrow{d}$ denotes convergence in distri-

et al. (2021), except we construct the statistic using estimates constructed under the alternative hypothesis.

[15]These are constructed using the residuals $\tilde{e}_{(1),it} = x_{it} - \tilde{\lambda}_{1,i}^\top \tilde{f}_{1,t}$ and $\tilde{e}_{(2),it} = x_{it} - \tilde{\lambda}_{2,i}^\top \tilde{f}_{2,t}$ in the series $\tilde{Z}^\top \tilde{f}_{1,t}\tilde{e}_{(1),it}$ and $\tilde{f}_{2,t}\tilde{e}_{(2),it}$ respectively, (see A.1.4 in the Supplementary Material), or via a bootstrap.

bution, $\lfloor\cdot\rfloor$ denotes the floor operator, M denotes generic finite constants, and $A^{-\top}$ denotes the inverse transpose of any invertible matrix A.

2.3.1 Estimation

We first establish the properties of the estimated rotational and orthogonal shift components by making the following assumptions. Let $\iota_{1t} \equiv \mathbf{1}\{t \leq \lfloor \pi T \rfloor\}$ and $\iota_{2t} \equiv \mathbf{1}\{t \geq \lfloor \pi T \rfloor + 1\}$.

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 $\Sigma_F > 0$.*

Assumption 2. *For $m = 1, 2$, $E\|\lambda_{m,i}\|^4 \leq M$, $\left\|\Lambda_m^\top \Lambda_m / N\right\| - \Sigma_{\Lambda_m} \xrightarrow{p} 0$ for some $\Sigma_{\Lambda_m} > 0$, and $\left\|\Lambda_m^\top \Lambda_m / N - \Sigma_{\Lambda_m}\right\| = O_p\left(N^{-1/2}\right)$. The shift break is orthogonal such that $\Lambda_1^\top W = O_p(1)$.*

Assumption 3. *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 $m = 1, 2$, $\{\lambda_{m,i}\}$, $\{f_t\}$ and $\{e_{it}\}$ are mutually independent groups.*

Assumption 5. *For all T and N:*

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

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

Assumption 6. *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) *For each* t $E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{m,i}e_{it}\right\|^4 \leq M$.

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

Assumptions 1 to 7 are either straight from, or slight modifications of, those in Bai (2003). Assumption 1 is 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 a common identification condition (e.g. Han and Inoue, 2015; Baltagi et al., 2017, and others) which limits the factors to exhibit no heteroskedasticity, but this is not restrictive in our case as changes in Σ_F are characterised by Z. Assumption 2 is Assumption B in Bai (2003), except that it specifies the convergence speed of $\Lambda_m^{\top}\Lambda_m/N$ to be no slower than $1/\sqrt{N}$ for $m = 1, 2$. Assumption 2 allows for the loadings to be random, and relaxes the strict $\Lambda_1^{\top}W = 0$ condition found in Wang and Liu (2021); Pelger and Xiong (2022); Massacci (2021).[16] Assumption 3 allows for weak serial and cross-sectional correlation and corresponds 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 strengthens Assumption 3, and corresponds to Assumption E in Bai (2003). Assumption 6 are Assumptions F1-F2 of 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 estimation. Assumption 6 (c) is slightly stronger than 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 there is infinite data pre- and post-break, and is a weaker version of Assumption 8 in Han and Inoue (2015), who assume that the terms are bounded uniformly in a range of potential π.

[16] Although this is not required for the purposes of estimation and the Z rotation test, it is required for the W orthogonal shift tests, and we therefore combine this assumption for simplicity.

Recall that $\tilde{F}_1$ and $\tilde{F}_2$ satisfy $\tilde{F}_1^\top \tilde{F}_1/T_1 = \tilde{F}_2^\top \tilde{F}_2/T_2 = I_r$, and are thus estimates of $F_1 H_1$ and $F_2 H_2$, where H_1 and H_2 are the respective pre- and post-break normalisation bases.[17]

Theorem 2.1. *Under Assumptions 1 to 8, and as $N, T \to \infty$*

a) $\left\| \tilde{Z} - H_1^\top Z H_2^{-\top} \right\| = O_p\left(\frac{1}{\delta_{NT}^2}\right)$,

b) $\frac{1}{N}\left\| \tilde{W} - W H_2^{-\top} \right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

Theorem 2.1 (a) shows that $\tilde{Z}$ is consistent for Z, but is affected by the normalisation matrices. Since $\tilde{F}_2$ estimates $F_2 H_2$, it follows that the combined series $\widehat{F} = [\tilde{F}_1^\top, \tilde{Z}\tilde{F}_2^\top]^\top$ is on the same normalisation basis both before and after the break, and can thus form the basis of a test for evidence of rotational breaks. Importantly, $\widehat{F}$ is free from the effects of any possible orthogonal shifts induced by W, and thus isolates the rotational change in the factor variance. Theorem 2.1 (b) shows that $\tilde{W}$ estimates the true W up to a normalisation basis, and thus can used to construct test statistics.

2.3.2 Z-test for Rotational Changes

We analyse the Z-test for the null of no break in the factor variance $\mathcal{H}_0 : \Sigma_F = Z\Sigma_F Z^\top$ that holds regardless of W, and is therefore robust to breaks in the factor loadings. We define $\mathscr{W}_Z(\pi, FH_{0,1}) = A_Z(\pi, FH_{0,1})^\top \widehat{S}_Z(\pi, FH_{0,1})^{-1} A_Z(\pi, FH_{0,1})$ as the infeasible analog of $\mathscr{W}_Z(\pi, \widehat{F})$, where $H_{0,1} = \text{plim}(H_1)$,[18] and make the following assumptions.

Assumption 9.

a) The Bartlett kernel of Newey and West (1987) is used, and there exists a constant $K > 0$ such that $b_T, b_{\lfloor \pi T \rfloor}$ and $b_{T-\lfloor \pi T \rfloor}$ are less than $KT^{1/3}$; and

[17] Specifically, their respective bases are $H_1 = \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)\left(\frac{F_1^\top \tilde{F}_1}{T_1}\right) V_{NT,1}^{-1}$, $H_2 = \left(\frac{Z^\top \Lambda_1^\top \Lambda_1 Z}{N} + \frac{W^\top W}{N}\right)\left(\frac{F_2^\top \tilde{F}_2}{T_2}\right) V_{NT,2}^{-1}$, where $V_{NT,1}$ and $V_{NT,2}$ are the diagonal matrices of eigenvalues of the first r eigenvalues of $(NT_1)^{-1} X_1 X_1^\top$ and $(NT_2)^{-1} X_2 X_2^\top$. There exists an alternative parameterization $H_2^\dagger = \left(\Lambda_1^\top \Lambda_1 + Z^{-\top} W^\top W Z^{-1}\right)\left(Z F_2^\top \tilde{F}_2\right)/(NT_2) V_{NT,2}^{-1}$, where the rotation Z is parameterised as part of the factors. It is straightforward to verify that either parameterisation leads to same result, (see Remark A.3 in the Appendix).

[18] The definition of $H_{0,1}$ follows from Bai (2003). Lemma A3 of Bai (2003) shows that $V_{NT,1}$ converges to V_1, a diagonal matrix with the eigenvalues of $\Sigma_{\Lambda_1}^{1/2} \Sigma_F \Sigma_{\Lambda_1}^{1/2}$. Let Υ_1 denote its eigenvectors such that $\Upsilon_1^\top \Upsilon_1 = I_r$. Proposition 1 of Bai (2003) shows that $F_1^\top \tilde{F}_1/T_1$ converges to $\Sigma_{\Lambda_1}^{-1/2} \Upsilon_1 V_1^{1/2} = H_{0,1}^{-\top}$. It follows that $H_1 \xrightarrow{p} \Sigma_{\Lambda_1}^{1/2} \Upsilon_1 V_1^{-1/2} = H_{0,1}$, its probability limit. One can also define $H_{0,2} = \text{plim}(H_2)$ in a similar way.

b) $\frac{T^{2/3}}{N} \to 0$ *as* $N, T \to \infty$.

Assumption 10.

a) $\Omega_Z = \lim_{T\to\infty} Var\left(vech\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_{0,1}^\top f_t f_t^\top H_{0,1} - I_r\right)\right)$ *is positive definite, and* $\|\Omega_Z\| < \infty$. *Its estimators* $\widehat{\Omega}_{Z,(m)}(\pi, FH_{0,1})$ *for* $m = 1, 2$ *are consistent such that* $\left\|\widehat{\Omega}_{Z,(m)}(\pi, FH_{0,1}) - \Omega_Z\right\| = o_p(1)$.

b) $\mathscr{W}_Z(\pi, FH_{0,1}) \Rightarrow Q_p(\pi)$, *where* $Q_p(\pi) = [B_p(\pi) - \pi B_p(1)]^\top [B_p(\pi) - \pi B_p(1)]/(\pi(1-\pi))$, *and* $B_p(\cdot)$ *is a* $p = r(r+1)/2$ *vector of independent Brownian motions on* $[0, 1]$.

Assumption 9 specifies conditions for the Bartlett kernel. Assumption 10 (a) is a standard HAC assumption, and states that the infeasible estimators $\widehat{\Omega}_{Z,(1)}(\pi, FH_{0,1})$ and $\widehat{\Omega}_{Z,(2)}(\pi, FH_{0,1})$ converge to their population counterpart Ω_Z. Assumption 10 (b) is the main result of Theorem 3 of Andrews (1993), and is a necessary hyper-assumption to establish the asymptotic distributions of the test statistics. As stated by Andrews (1993), for any fixed π, $Q_p(\pi)$ is distributed as a $\chi^2_{p=r(r+1)/2}$ random variable. Assumption 10 (b) has been used in Han and Inoue (2015), and one can refer to Chen et al. (2014) for more primitive assumptions to see that it is satisfied for a large class of ARMA processes.

Theorem 2.2. *Under Assumptions 1 to 10, and if* $\frac{\sqrt{T}}{N} \to 0$, *then* $\mathscr{W}_Z(\pi, \widehat{F}) \xrightarrow{d} \chi^2_{r(r+1)/2}$.

Theorem 2.2 shows that the Z-test statistic converges to a chi-squared random variable, conditional on a break fraction.[19] To ensure the Z-test's power under the alternative, we make the following assumptions.

Assumption 11. $\|Z\| < \infty$, *and* $Z\Sigma_F Z^\top \neq \Sigma_F$.

Assumption 12. $plim_{T\to\infty} \inf\left(vech(C)^\top \left[\max(b_{\lfloor \pi T\rfloor}, b_{T-\lfloor \pi T\rfloor})\widehat{S}(\pi, F^* H_{0,1})^{-1}\right] vech(C)\right) > 0$, *where* $F^* = [F_1^\top, ZF_2^\top]^\top$ *and* $C \equiv H_{0,1}^\top(\Sigma_F - Z\Sigma_F Z^\top)H_{0,1}$.

[19] It is also possible to construct an LM-like statistic with a restricted estimate of the variance using all of the data (see Section A.1.5 in the Supplementary Material). However, as noted by Chen et al. (2014) and Han and Inoue (2015), such LM-like statistics have much smaller power than their Wald-type counterparts. Therefore, we focus on the Wald test.

Assumption 11 formalises the definition of a break in factor variance. It rules out the unlikely scenario where $Z = -1$, i.e. all of the loadings switch their signs after the break, and is commonly assumed (see Han and Inoue, 2015; Baltagi et al., 2017, 2021, and others). Assumption 12 regulates the asymptotics of the variance matrices of the statistics under the alternative. Together, these ensure that the subsample means of $\widehat{f}_t\widehat{f}_t^\top$ converge to different limits, and the divergence of $\mathscr{W}_Z(\pi, \widehat{F})$, as summarised in the following theorem.

Theorem 2.3. *Under Assumptions 1 to 9 and 12, and if Z satisfies Assumption 11, then*

a) there exists some non-random matrix $C \neq 0$ such that
$\frac{1}{\pi T}\sum_{t=1}^{\lfloor \pi T\rfloor} \widehat{f}_t\widehat{f}_t^\top - \frac{1}{T-\lfloor \pi T\rfloor}\sum_{t=\lfloor \pi T+1\rfloor}^{T} \widehat{f}_t\widehat{f}_t^\top \xrightarrow{p} C,$

b) $\mathscr{W}_Z(\pi, \widehat{F}) \to \infty$ under the alternative hypothesis that $\Sigma_F \neq Z\Sigma_F Z^\top$.

2.3.3 W-test for Orthogonal Shifts

Next, we analyse the W-test for breaks in the loadings $\mathcal{H}_0 : W = \mathbf{0}$ that holds regardless of Z, and is therefore robust to changes in the factor variance. First note that, because W is an $N \times r$ matrix where $N \to \infty$, traditional tests are infeasible. We note that popular approaches such as Bonferroni test statistics and pooling individual test statistics can suffer from significant size distortions,[20] whereas directly testing for a change in the number of factors[21] requires much stricter assumptions on the error term. This motivates us to formulate an individual test statistic $\mathscr{W}_{W,i}$ for each i, and a joint test statistic $\mathscr{W}_W$ pooled across N to overcome the infinite dimensionality problem. To analyse the test statistics, we make the following additional assumptions.

[20]The Bonferroni test statistic B_w is the maximum of $\mathscr{W}_{W,i}$, and the critical value is $\mathscr{X}^{-1}(0.01 - 0.05/N)$ where $\mathscr{X}$ is the chi-square CDF with r degrees of freedom. The pooled statistic pools the individual test *statistics* into $P_w = \left(\sum_{i=1}^{N} \mathscr{W}_{W,i} - rN\right)/\sqrt{2Nr}$, where the corresponding critical values are ± 1.96. Both of these approaches rely on the sequential limit argument that $\mathscr{W}_{W,i} \to \chi_r^2$, and then $N \to \infty$. However, convergence of $\mathscr{W}_{W,i}$ relies on the joint asymptotics of N and T to ∞, so $\mathscr{W}_{W,i} \to \chi_r^2$, and then $N \to \infty$ cannot be separated into two separate steps. Indeed, the sequential and joint limits are not always equivalent (see Phillips and Moon, 1999), and both Bonferroni and pooled test statistics are known for potentially introducing significant size distortions (see Stock and Watson, 2005; Han, 2015).

[21]This follows by restating the null and alternative hypotheses as $\mathcal{H}_0 : r_w = 0, \mathcal{H}_1 : r_w \neq 0$, where r_w is the number of extra factors augmented by the presence of orthogonal shifts that is implied when using the *pseudo* factor representation. Existing tests such as Onatski (2009) cannot be used without imposing further restrictive assumptions on the errors of the approximate factor model.

Assumption 13. *For all* N, T,

a) For each t, $E(N^{-1/2}\sum_{i=1}^{N} e_{it})^2 \leq M$.

Assumption 14. *For all* N, T, *and* $m = 1, 2$*:*

a) For each i, $E\left\|\frac{1}{\sqrt{NT_m}}\sum_{t=1}^{T}\sum_{k=1}^{N}\left(\lambda_{m,k}[e_{kt}e_{it} - E[e_{kt}e_{it}]]\right)\iota_{mt}\right\|^2 \leq M$,

b) For each t, $E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{m,i}e_{it}\right\|^2 \leq M$,

c) For each i, $E\left\|\frac{1}{\sqrt{T_m}}\sum_{t=1}^{T} f_t e_{it}\iota_{mt}\right\|^4 \leq M$,

d) $E\left\|\frac{1}{\sqrt{NT_m}}\sum_{t=1}^{T}\sum_{i=1}^{N} f_t e_{it}\iota_{mt}\right\|^2 \leq M$.

Assumption 15. *For* $m = 1, 2$*:*

a) $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{m,i} = O_p(1)$,

b) $E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_{m,i}e_{it}^2\right\|^2 \leq M$ *for each* t,

c) $E\left\|\frac{1}{N\sqrt{T_m}}\sum_{t=1}^{T}\sum_{k\neq i}\sum_{i=1}^{N}\lambda_{m,k}e_{kt}e_{it}\cdot\iota_{mt}\right\|^2 \leq M$.

Assumption 16.

a) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t e_{it} \xrightarrow{d} N(0, \Phi_i)$, $(T)^{-1}\sum_{t=1}^{T} f_t f_t^\top e_{it}^2 \xrightarrow{p} \Phi_i$, *each* $\Phi_i > 0$,

b) $\frac{1}{\sqrt{TN}}\sum_{t=1}^{T}\sum_{i=1}^{N} f_t e_{it} \xrightarrow{d} N(0, \Phi_W)$, $(TN)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} f_t f_t^\top e_{it}^2 \xrightarrow{p} \Phi_W$, $\Phi_W > 0$.

Assumption 13 is the pooled version of Assumption 3. Assumption 14 (a) is Assumption 6 (a) but for the loadings, Assumption 14 (b) is already implied by Assumption 6 (c), and Assumption 14 (c) is a strengthened version of Assumption 6 (a). These correspond to Assumptions 6 b), 6 d) and 6 e) in Han (2015), and are not restrictive, because they involve zero mean random variables. Assumption 15 is required to bound the sum of the loadings by $O_p\left(\sqrt{N}\right)$, and is a slightly modified version of the Assumption 7 in Han (2015). This will hold if the loadings are centred around zero, and their sum diverges at the rate of $\sqrt{N}$ by the central limit theorem (CLT). Although this is somewhat stricter than a conventional factor model setup, it seems to hold for empirically used datasets. Assumptions 16 (a) and 16 (b) are CLTs, where the latter is the cross-sectionally averaged version of the CLT in Bai (2003), which somewhat restricts the cross-sectional correlation in e_{it}.

Theorem 2.4. *If $\frac{\sqrt{T}}{N} \to 0$, then:*

a) Under Assumptions 1 to 9, 13, 14, and 16, $\mathscr{W}_{W,i} \xrightarrow{d} \chi_r^2$ for each i, and

b) Under Assumptions 1 to 9 and 13 to 16, $\mathscr{W}_W \xrightarrow{d} \chi_r^2$.

Theorem 2.4 shows that the W-test statistics[22] converge to conventional chi-squared random variables. To analyse the joint W-test under the alternative, we make some further assumptions.

Assumption 17. *There exist constants $0 < \alpha \leq 0.5$ and $C > 0$ such that as $N, T \to \infty$, $Pr\left(\left\|\frac{T^{\alpha/2}}{\sqrt{N}} \sum_{i=1}^N w_i\right\| > C\right) \to 1$.*

Assumption 17 requires $\left\|\frac{T^{\alpha/2}}{\sqrt{N}} \sum_{i=1}^N w_i\right\|$ to be bounded away from zero asymptotically. Note that if $N^{-1} \sum_{i=1}^N w_i \xrightarrow{p} 0$ under the alternative, then $N^{-1/2} \sum_{i=1}^N w_i$ converges in distribution to some Gaussian random variable by the CLT, and hence $\left\|N^{-1/2+\epsilon} \sum_{i=1}^N w_i\right\|$ is diverging as $N \to \infty$ for any positive ϵ. In order for $\left\|\frac{T^{\alpha/2}}{\sqrt{N}} \sum_{i=1}^N w_i\right\|$ to be bounded away from zero, any $\alpha \in (0, 0.5]$ such that $T^{\alpha/2} \geq N^{\epsilon}$ is required, which is not difficult. Assumption 17, therefore, ensures that the joint test statistic diverges under the alternative hypothesis, even if $N^{-1} \sum_{i=1}^N w_i \xrightarrow{p} 0$, as summarised in the following theorem.

Theorem 2.5. *If $\frac{\sqrt{T}}{N} \to 0$, and the alternative $\mathcal{H}_1 : W \neq \mathbf{0}$ holds, then:*

a) Under Assumptions 1 to 9, 13, 14, and 16, and if $w_i \neq 0$, then $\mathscr{W}_{W,i} \to \infty$ as $N, T \to \infty$,

b) Under Assumptions 1 to 9 and 13 to 17, $\mathscr{W}_W \to \infty$ if $\frac{\sqrt{N}}{T^{1-\alpha/2}} \to 0$ as $N, T \to \infty$.

2.3.4 Estimation of Number of Factors and Break Fraction

Our test statistics assume that the number of factors and break fraction are known. In practice, consistent estimators can be used instead, as addressed in the following remarks.

Remark 2.1. *The number of factors r in either subsample can be consistently estimated conditional on consistent estimate of π (e.g. Bai and Ng, 2002; Onatski, 2010; Ahn and Horenstein, 2013; Baltagi et al., 2017). If the pre- and post-break estimates of the number of factors $\tilde{r}_1$ and $\tilde{r}_2$ differ, this can be accommodated by allowing $\tilde{Z}$ to be a $\tilde{r}_1 \times \tilde{r}_2$.*

[22] It is also possible to construct an LM-like test statistic by imposing the null hypothesis of no break, but this results in a statistic with lower power, so we focus on the Wald test again.

Remark 2.2. *The break fraction π can be consistently estimated (e.g. Baltagi et al., 2017; Duan et al., 2022, and others). Theorem 3 of Baltagi et al. (2017) shows that consistent estimators of π are sufficient to obtain the usual $O_p\left(\frac{1}{\delta_{NT}^2}\right)$ consistency rate of the estimated factors and loadings, and therefore our test statistics remain valid.*

2.4 Monte Carlo Study

2.4.1 Simulation Specification

We first simulate two sets of loadings, Λ_1, Λ_2, as a multivariate $N(\mathbf{0}_3, I_3)$, focusing on the case of $r = 3$ factors. Then, we set W to be the residuals of the projection $\Lambda_2 - (\Lambda_1^\top \Lambda_1)^{-1}\Lambda_1^\top \Lambda_2$. The rotation Z is set to I_3 in the case of no break, or a lower triangular matrix with $[2.5, 1.5, 0.5]$ on the main diagonal and its lower triangular entries drawn from $N(0, 1)$, as in Duan et al. (2022). The overarching model we simulate from is

$$x_{it} = \begin{cases} \lambda_{1,i}^\top f_t + \sqrt{\theta} e_{it}, & t = 1, \dots, \lfloor \pi T \rfloor \\ (Z\lambda_{1,i} + \omega w_i)^\top f_t + \sqrt{\theta} e_{it}, & t = \lfloor \pi T \rfloor + 1, \dots, T, \end{cases} \tag{2.4.1}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. The parameter $\theta = 3$ sets the signal to noise ratio to be 50%, and ω controls the "size" of the orthogonal shifts. The factors and errors are generated as:

$$f_{k,t} = \rho f_{k,t-1} + \mu_{it}, \quad \mu_{it} \sim i.i.d.\ N(0, 1 - \rho^2), \tag{2.4.2}$$

$$e_{it} = \alpha e_{i,t-1} + v_{it}, \tag{2.4.3}$$

where $\rho \in \{0, 0.7\}$ controls autocorrelation in the factors, and μ_{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 $t = 1, \dots, T$. The scalar $\alpha \in \{0, 0.3\}$ allows mild serial correlation, and as in Bates et al. (2013) and Baltagi et al. (2017), $\Omega_{ij} = \beta^{|i-j|}$ with $\beta \in \{0, 0.3\}$ allowing mild cross-sectional correlation. The break fraction is set to 0.5 and treated as known.[23] Disentanglement requires running both the Z and W tests, which could lead to a higher family wise

[23] Alternative break fractions of 0.3 and 0.7 do not qualitatively change the results, and are relegated to the Supplementary Material.

Table 2.1: Size of Rotation and Orthogonal Shift Tests, $N = 200, r = 3$

				Z-Test		W-Test		W Individual
T	ρ	α	β	Unadjusted	Adjusted	Unadjusted	Adjusted	
200	0.0	0.0	0.0	0.140	0.088	0.135	0.107	0.027
		0.3	0.3	0.147	0.101	0.092	0.055	0.016
500		0.0	0.0	0.100	0.053	0.003	0.001	0.007
		0.3	0.3	0.108	0.058	0.064	0.032	0.007
200	0.7	0.0	0.0	0.220	0.156	0.125	0.085	0.027
		0.3	0.3	0.215	0.154	0.087	0.062	0.029
500		0.0	0.0	0.136	0.086	0.003	0.001	0.008
		0.3	0.3	0.134	0.085	0.062	0.042	0.011

Note:
Entries report the rejection frequencies for the Z-test for break in factor variance, and W-test for break in loadings. Nominal size is 5%. The parameters α and β denote the degree of serial and cross-sectional correlation in the error respectively, ρ denotes the degree of autocorrelation in the factors. The number of pseudo-factors estimated over the whole sample using $IC_p(2)$ of Bai and Ng (2002) is $\tilde{r}$.

error rate, and to this end we report the unadjusted p values, in addition to the adjusted p values using a Bonferroni-Holm correction.

2.4.2 Simulation Results

We present the size analysis in Table 2.1. In the case of no serial correlation in the factors and $T > N$, the Z-test has a nominal size close to 5% regardless of the serial or cross-sectional correlation in the errors. The Z-test seems to be oversized when there is serial correlation in the factors, but this is alleviated and approaches a rejection rate of 0.09 as T increases.[24] The W-test does not seem to be affected by serial correlation in the factors, and also seems to be overly conservative when there is no serial correlation in the error, but otherwise seems to have good size. Implementation of the Bonferroni-Holm procedure to adjust the p values also seems to correct the oversizing issue, so we advocate for its use.

Table 2.2 presents the power of the Z and W tests across all types of breaks; both have good power and are rejecting correctly only on their respective break types. This contrasts with the tests of Han and Inoue (2015) and Baltagi et al. (2021), which reject across all break types, and thus cannot discern which type of break has occurred.

[24]Increasing T further does seem to make the size approach 5% (see Table A.1 in Supplementary Material).

Table 2.2: Power of Z- and W-Tests, $r = 3$, $N = 200$, $\alpha = \beta = 0.3$

				Z-Test		W-Test					
Break Type	T	ω	ρ	Unadj.	Adj.	Unadj.	Adj.	Individual	HI (2015)	BKW (2021)	$\tilde{r}$
$W \neq \mathbf{0}$	200	1	0.0	0.136	0.129	0.860	0.821	0.849	1.000	1.000	5.928
			0.7	0.244	0.233	0.916	0.896	0.908	1.000	1.000	6.000
	500		0.0	0.079	0.076	0.950	0.939	0.947	1.000	1.000	6.000
			0.7	0.146	0.144	0.968	0.965	0.968	1.000	1.000	6.000
$Z \neq I$	200	0	0.0	1.000	1.000	0.100	0.100	0.026	1.000	1.000	3.000
			0.7	1.000	1.000	0.106	0.106	0.035	1.000	1.000	3.000
	500		0.0	1.000	1.000	0.094	0.094	0.009	1.000	1.000	3.000
			0.7	1.000	1.000	0.096	0.096	0.012	1.000	1.000	3.000
$W \neq \mathbf{0}$ and $Z \neq I$	200	1	0.0	1.000	1.000	0.804	0.803	0.765	1.000	1.000	4.206
			0.7	1.000	1.000	0.867	0.867	0.846	1.000	1.000	5.047
	500		0.0	1.000	1.000	0.919	0.919	0.901	1.000	1.000	4.772
			0.7	1.000	1.000	0.946	0.946	0.938	1.000	1.000	5.511

Note:
Entries denote the rejection rates across different simulated break types; a break type of $W \neq 0$ denotes a break in the loadings, $Z \neq I$ a break in the factor variance, and $W \neq 0$ and $Z \neq I$ denoting a break in both. HI denotes Han and Inoue's (2015) test, and BKW denotes Baltagi et al's (2021) test conducted with a pre-known break fraction. The scalar ω denotes the "size" of the break in the loadings. See Table 2.1 for explanation of α, β, ρ and $\tilde{r}$.

2.5 Empirical Application

We apply our methodology to FRED-QD, a standard U.S. quarterly macroeconomic dataset (see McCracken and Ng, 2020), developed to mimic the Stock and Watson (2012a) dataset. The Stock and Watson (2012a) dataset is widely used in the factor modelling literature, having been used to date and test breaks in factor models (e.g. Chen et al., 2014; Baltagi et al., 2021), marking FRED-QD as well-suited for our application.

We consider the sample 1959Q3-2019Q4 from the FRED-QD dataset, and adopt the suggested data cleaning and transformations, removing top level aggregates to yield a panel of 124 series.[25] We align our start date to extant work, and end before the COVID-19 pandemic (as noted by Ng, 2021; Stock and Watson, 2021, this period is unclear to deal with). We consider two breaks in 1984Q1 and 2008Q3, as estimated by the procedure of Baltagi et al. (2021) using 2-6 factors and a trimming parameter of 0.1, which we henceforth refer to as the Great Moderation and Great Recession break respectively. We note that these breaks align with events associated *a priori*, and existing evidence: the Great Moderation break is consistent with works that use a sample of 1960 to the mid 2000s and find a break in the early 1980s (see Stock and Watson, 2009; Breitung and Eickmeier, 2011;

[25]Additional details of the data can be found in Table A.4 in the Appendix.

Chen et al., 2014; Baltagi et al., 2021), and the Great Recession break has some evidence of breaks in loadings as noted by Stock and Watson (2012a) who work with a very short post-2009 sample.[26] The number of factors in each subsample differs across estimators and is often contradictory;[27] given this known instability, we consider two to six factors for our empirical analysis. In line with the extant literature, we set the number of pre- and post-break factors to be the same; results allowing for a change in the number of factors are qualitatively similar, and are listed in Table A.5 of the Appendix.

2.5.1 Joint test results

Table 2.3 reports the p values for the Z- and joint W-tests with the Great Moderation and the Great Recession as candidate break dates. Across 2 to 6 factors, there is strong evidence of rejection of the null hypothesis of no breaks for both tests across both the Great Moderation and Great Recession, with all p values being less than 0.05, indicating that both types of structural breaks are empirically relevant for factor models of U.S. macroeconomic data.

2.5.2 Were breaks in the factor variance important?

A key tenet of the argument of the importance of distinguishing breaks in the factor variance and loadings is because the routine normalisation applied in factor models rules out the possible interpretation that the factor variance changed; breaks in the factor variance are necessarily subsumed into breaks in the factor loadings. Given the clear rejection of the Z-test, our results suggest that rotational breaks, which may stem from changes in the factor variance, are important when modelling U.S. macroeconomic data using factors.

From our framework, rejection of the Z-test implies that the covariance matrix of the factors has changed, i.e. $\Sigma_F \neq Z\Sigma_F Z^\top$. However, it does not assess the economic significance of these breaks. Our projection-based decomposition also allows us to estimate how much the total factor

[26] Barigozzi et al. (2018) and Cheng et al. (2016) also find a structural break in the factor structure around the Great Recession period.

[27] The eigenvalue edge distribution estimator of Onatski (2010) and eigenvalue ratio estimator of Ahn and Horenstein (2013) often finds 2 to 4 factors in the sub-regimes, while information criteria such as Bai and Ng (2002) often suggest a large number of factors such as 5 to 9. Stock and Watson, on the other hand, always consider 6 factors when studying factor instability (e.g. Stock and Watson, 2009, 2012a).

Table 2.3: Joint Test Results

	Z-Test p values		W-Test p values	
$\tilde{r}$	Unadjusted	Adjusted	Unadjusted	Adjusted
Great Moderation (1984 Q1), 1959 Q3 - 2008 Q3 Sample				
2	0.001	0.001	0.000	0.000
3	0.000	0.000	0.000	0.000
4	0.008	0.007	0.000	0.000
5	0.000	0.000	0.001	0.001
6	0.000	0.000	0.000	0.000
Great Recession (2008 Q3), 1984 Q2 - 2019 Q4 Sample				
2	0.000	0.000	0.004	0.004
3	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000

Note:
Rejection of the Z-test corresponds to a break in the factor covariance matrix, and rejection of the W-test corresponds to a break in the loadings across the entire cross-section.

variance has changed pre- and post-break. Denoting $tr(A)$ as the trace of a square matrix A, $tr(\Sigma_F)$ and $tr(Z\Sigma_F Z^\top)$ are the total variance of the factors pre- and post-break. Therefore, their ratios provide an estimate of how much the total factor variance has (de)inflated post-break. Table 2.4 presents our estimate of the ratio, or $tr(Z\Sigma_F Z^\top)/tr(\Sigma_F)$ along with the 95% confidence interval.[28] We estimate that the Great Moderation was associated with an over 60% reduction in the total variance of the factors, across the specification of 2 to 6 factors. For the Great Recession break, this ratio is close to 1, with 1 always being within the confidence interval. This suggests that despite rejection of the Z-test, breaks in the factor variance were less important for understanding the factor model relative to the Great Moderation break.[29]

Being able to associate the Great Moderation break with a change in the factor variance reconciles how one can understand the underlying dynamic factor model in Equations (2.2.1) and (2.2.2). First, work by, for example, Primiceri (2005), Cogley and Sargent (2005), and Sims and Zha (2006) attach a "Good luck" interpretation of the Great Moderation. "Good luck" interprets the Great

[28]More precisely, we estimate $tr(H_{0,1}^\top Z\Sigma_F Z^\top H_{0,1})/tr(H_{0,1}^\top \Sigma_F H_{0,1})$, see Section A.3.4.

[29]Note that rejection of the Z-test which is due to a break in the factor covariance matrix, may occur even when the total factor variance (trace of the variance matrix) remains stable. This can occur if the individual factors' variances (diagonal elements) break but their sum remains the same, if the correlation between factors (off-diagonals) change, or some combination thereof. Since we can only estimate the space spanned by the factors, the Z-test cannot tell the two cases apart. All we can claim is given the trace appears similar pre- and post-break with the Great Recession, breaks in factor variances are probably less important, relative to the Great Moderation.

Table 2.4: Estimated ratio of the factor variances

r	$tr(\tilde{Z}\tilde{Z}^\top)/tr(I_r)$	95% Bootstrap Confidence Interval
Great Moderation (1984 Q1), 1959 Q3 - 2008 Q3 Sample		
2	0.255	[0.185, 0.269]
3	0.294	[0.189, 0.306]
4	0.347	[0.239, 0.353]
5	0.306	[0.23, 0.324]
6	0.289	[0.223, 0.301]
Great Recession (2008 Q3), 1984 Q2 - 2019 Q4 Sample		
2	0.893	[0.888, 1.089]
3	1.303	[0.762, 1.49]
4	1.208	[0.957, 1.375]
5	1.097	[0.913, 1.158]
6	1.030	[0.894, 1.112]

Note:
The table presents estimates of the ratio of the total factor variance pre and post-break, or $tr\left(\Sigma_F\right)/tr\left(Z\Sigma_F Z^\top\right)$. Values less than 1 indicate that the estimated total variance of the factors pre-break is smaller than the total variance of the factors post-break. Confidence intervals are constructed by a block bootstrap ($m = 8$) to preserve both serial and cross-sectional correlation.

Moderation as arising from smaller shocks hitting the economy, whereas "Good policy" views the Great Moderation as arising from explicit policy choice. From the perspective of the dynamic factor model, the "Good luck" interpretation is only possible through a break in the variance of the underlying shocks, or Σ_η in Equation (2.2.2). In contrast, the "Good policy" interpretation arises from parameters not linked to the variances of shocks. That is, a break in the factor variance is a necessary condition for the dynamic factor model to attach the "Good luck" interpretation. Our estimate of an over 60% decrease in the factor variance from pre- to post-Great Moderation thus allows for the "Good luck" interpretation. To be clear, our tests alone cannot distinguish between the "Good luck" and "Good policy" interpretation because breaks in the factor dynamics in Equation (2.2.2) (i.e. the Φ_j's) without a corresponding break in Σ_η, also lead to a break in the factor variance, and one would probably attach a "Good policy" interpretation akin to Lubik and Schorfheide (2004) for such a case. Second, regardless of how one interprets the Great Moderation, the fact that we can reject our Z-test and estimate an over 60% decrease in the total factor variance tells us that a large break associated with the Great Moderation, when viewed via the dynamic factor model, *must* have occurred in the equations governing the factor dynamics or Equation (2.2.2). Regardless of whether breaks in Equation (2.2.1) occurred or were important, the extant literature, by only working with breaks in the loadings, rules out interpreting possible breaks in Equation (2.2.2) by construction.

2.5.3 Individual Test Results

Besides the joint W-test, we can also test for breaks in loadings individually. Table 2.5 presents the number of series where we can reject at least one of their factor loadings breaking, controlling for a possible break in the factor variance. For comparison, we also apply Breitung and Eickmeier (2011)'s test for breaks in the factor loadings; however, we caution that direct comparison is not straightforward due to their use of *pseudo* factors, and hence non-robustness to changes in the factor variance. Furthermore, both individual tests are run as-is, and are thus subject to the problem of multiple hypothesis testing. We nonetheless suggest two tentative conclusions. First, there is some evidence to suggest that by accounting for changes in the factor variance, one may find fewer breaks in the loadings. This suggests that for the Great Moderation, some of the breaks in factor loadings found when using pseudo-factors may be related to breaks in the factor variance instead of the factor loadings. We base this tentative conclusion on the fact that we find fewer series with a break in their factor loadings than the Breitung and Eickmeier (2011) procedure, but note that this seems to only hold when we consider 2-4 factors. Alternatively, our higher number of rejections for 5-6 factors could be due to the loss of power of Breitung and Eickmeier (2011)'s statistic as the number of factors increases, as documented by Yamamoto and Tanaka (2015). Second, our results suggest that with a longer post-Great Recession sample, we find many more breaks in the loadings than Stock and Watson (2012a) and Breitung and Eickmeier (2011). One possibility is our test is more powerful in detecting breaks in the factor loadings, though we urge caution with this interpretation given that their use of the pseudo-factors means that the procedures can be quite different.

2.5.4 Which variables experienced a break in their loadings?

To further understand the break associated with the Great Moderation and Great Recession, we explored which types of variables had breaks in loadings. In order to understand whether these breaks were important for understanding the variation in variables, we calculate an R^2 measure

Table 2.5: Individual Series Loading Break Test Rejection Counts

$\tilde{r}$	Individual w_i	Breitung and Eickmeier (2011)
Great Moderation (1984 Q1), 1959 Q3 - 2008 Q3 Sample		
2	54	69
3	59	66
4	51	60
5	70	61
6	72	71
Great Recession (2008 Q3), 1984 Q2 - 2019 Q4 Sample		
2	55	25
3	48	32
4	56	42
5	67	45
6	66	52

* Numbers in cells represent the count of rejections of the null hypothesis that the loadings of an individual series broke at the given break date, (5% significance level). Total of 124 series in each subsample.

for each series subject to no restrictions, and with the restriction that there were no breaks in the loadings.[30] Thus, these R^2 statistics should have a large (small) difference if the breaks in the loadings were (un)important.

Figure 2.1 presents the unrestricted and $W = 0$ restricted R^2 statistics averaged across all series by category. We present the R^2 for $r = 3$ but note that these conclusions are very similar for $r = 2$ to 6. For the Great Moderation break, breaks in the loadings appear to be important for prices, earnings, exchange rates, and non-household balances. Two of these categories at least plausibly coincide with extant knowledge: the Great Inflation, which preceded the Great Moderation and affected prices; and the collapse of the Bretton Woods system in the mid-1970s, which affected exchange rate variables. For the Great Recession break, while it appears that breaks in the loadings were important for many variable categories, they appear important for financial variables in categories such as exchange rate, money, and the stock market. Additionally, it appears that breaks in loadings were important for prices, also documented by Stock and Watson (2012a).

[30]This uses Equation (2.2.7) to impose the restriction of no breaks in the loadings. For more details, see the Supplementary Material.

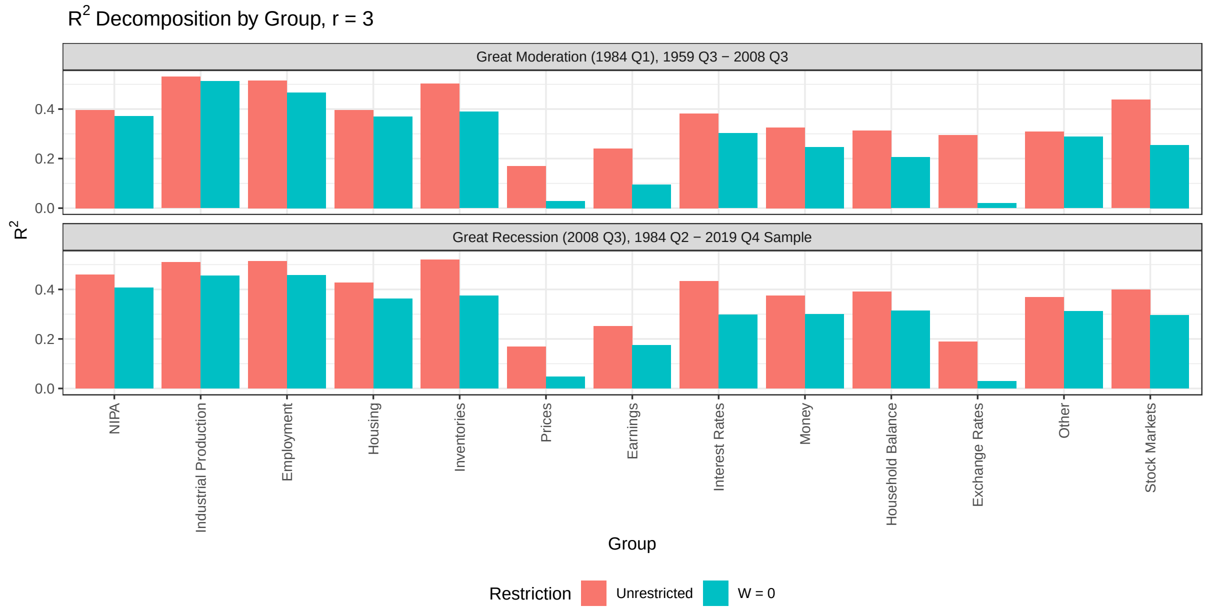

Figure 2.1: R^2 Statistics for unrestricted and restricted common component ($W = \mathbf{0}$) for Great Moderation Subsample, and Great Recession Subsample, for $r = 3$.

2.6 Conclusion

The existing literature on structural breaks in factor models by and large does not distinguish between breaks between the factor variance and loadings, due to the need of a normalisation during estimation. We argue it is important to distinguish between them, as both can lead to different economic interpretations. To address this, we develop a projection-based decomposition of *any* structural break into a rotational and orthogonal shift component, which are naturally interpreted as a change in factor variance and loadings, respectively. The estimators are simple to calculate and lead to two easy-to-implement Wald tests to disentangle structural breaks in the factor variance and loadings. Their finite sample performance is confirmed by a Monte Carlo study. Applying our procedure to U.S. macroeconomic data we find strong evidence of both types of breaks associated with the Great Moderation and Great Recession. Our projection-based decomposition allows us to estimate that the Great Moderation is associated with an over 60% reduction in the factor variance, a result precluded *a priori* if the break is not disentangled, and thus highlights the importance of doing so. Although our framework cannot distinguish between breaks in the factor dynamics and innovations and their respective "Good Policy" and "Good Luck" interpretations within the Great Moderation literature, rejection of our Z-test and the conclusion of a break in the factor variance is a necessary condition for subsequent discussion regarding the factors, a conclusion which cannot be established with existing tools in the literature.

Our framework provides a potential foundation to explore the precise practical and theoretical implications of structural breaks in factor models. For example, a natural question is how different break types can affect subsequent use of factors, such as factor-augmented forecasts (see Stock and Watson, 2002a; Bai and Ng, 2006), and factor-augmented vector auto-regressions (e.g. Bernanke et al., 2005). Indeed, despite suggestions for how factor-augmented forecasting can be done in presence of a structural break (e.g. Stock and Watson, 2009; Baltagi et al., 2021), there is still no formal treatment of this.

Chapter 3

Factor-augmented Forecasting Subject to Structural Breaks in the Factor Structure

3.1 Introduction

Factor-augmented regressions, pioneered by Stock and Watson (2002a, 2012b) 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 assumes structural stability at large, 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 can 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), and focus 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 compared to common empirical factor-based benchmarks, 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).

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.

3.2 Model and Estimation

3.2.1 Model Setup

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

$$y_{t+h} = f_t^\top \beta(L) + z_t^\top \delta + \eta_{t+h}, \tag{3.2.1}$$

[1]Our findings for the effects of structural breaks for factor estimation are standalone, hold regardless of the forecasting setup for y_{T+h}. Their subsequent effects on iterated forecasts are similar, though require more tedious theoretical adjustments and derivations.

where $h \geq 1$ is the forecast horizon, and $\beta(L)$ is a lag polynomial of order q for some $0 \leq q \leq q_{max}$. The term z_t collects all other regressors thought to improve forecasting performance; typically this includes a constant term, y_t itself and its lags. Our theoretical analysis focuses on the case with stationary regressors.

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). We note that breaks in the forecasting equation itself can be accommodated by split-sample estimation - a candidate model which we consider and allow for in our model averaging step; a more comprehensive treatment is left for future research. To this end, the r-dimensional factors f_t are unobserved but related to the panel of time series subject to a one time break in the factor structure

$$x_{it} = \begin{cases} \lambda_{1i}^{\top} f_t + e_{it}, & t = 1, \ldots, \lfloor \pi T \rfloor, \\ \lambda_{2i}^{\top} f_t + e_{it}, & t = \lfloor \pi T \rfloor + 1, \ldots, T, \end{cases} \tag{3.2.2}$$

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

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} F_1 \Lambda_1^{\top} + e_{(1)} \\ F_2 \Lambda_2^{\top} + e_{(2)} \end{bmatrix}, \tag{3.2.3}$$

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

$$\Lambda_2 = \Lambda_1 Z + W \tag{3.2.4}$$

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 $\mathbf{0}$

$$Z = I_r + \frac{R}{N^{1-\nu}}, \tag{3.2.5}$$
$$W = \frac{D}{N^{(1-\alpha)/2}}, \tag{3.2.6}$$

where R is some finite matrix satisfying $\|R\| < M$, $\frac{D^\top \Lambda_1}{N} = O_p\left(\frac{1}{\sqrt{N}}\right)$, 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 breaks, corresponding to the cases of $\nu < 0.5, \nu = 0.5$, and $\nu > 0.5$, as well as the cases of small, moderate, large, and very large shift breaks, corresponding to the cases of $\alpha < 0.5, \alpha = 0.5$, $\alpha \in (0.5, 1)$, and $\alpha = 1$. 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 (3.2.5) implies the following rates

$$\|I_r - Z\| = O_p\left(\frac{N^\nu}{N}\right), \tag{3.2.7}$$
$$\frac{\Lambda_1^\top W}{N} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \tag{3.2.8}$$

where the latter is implied by the Central Limit Theorem for Λ_1 and W which 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 (3.2.8).

Remark 3.1. *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" $r_1 \times r_2$ 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.*

3.2.2 Effects of Structural Breaks on Factor Estimates

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 (3.2.4) implies the following equivalent representation

$$
\begin{aligned}
X &= \begin{bmatrix} F_1 & 0 \\ F_2 Z^\top & F_2 \end{bmatrix} \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + e \\
&= \begin{bmatrix} G_r & G_p \end{bmatrix} \begin{bmatrix} \Lambda_1^\top \\ W^\top \end{bmatrix} + e \\
&= G\Xi^\top + e. \qquad (3.2.9)
\end{aligned}
$$

Equation (3.2.9) 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 potentially noisy estimate of G_r. Thus, the pseudo-factors are subject to contamination by 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 $X_1 X_1^\top/(T_1 N)$ and $\sqrt{T_2}$ times the first r eigenvectors of $X_2 X_2^\top/(T_2 N)$, 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.

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

$$\tilde{Z} = \left(\tilde{\Lambda}_1^\top \tilde{\Lambda}_1\right)^{-1} \tilde{\Lambda}_1^\top \tilde{\Lambda}_2 \tag{3.2.10}$$

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 3.1, and are a preview of the theoretical results in Section 3.3.1.

Table 3.1: 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

3.2.3 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 (3.2.1) as

$$\begin{aligned} y_{t+h} &= c_t^\top \theta + \eta_{t+h} \\ &= \mu_t + \eta_{t+h}, \end{aligned} \tag{3.2.11}$$

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 (3.2.11) 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 3.3.2.

Pseudo-factors and rotated factors are asymptotically equivalent for small shift 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. 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 rotational breaks. Thus, when the rotational break is 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.

3.2.4 Model Averaging and Cross-validation

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 (3.2.1) in scalar and matrix forms, respectively:

$$y_{t+h} = c_t(\mathcal{M})^\top \theta + \eta_{t+h}, \tag{3.2.12}$$

$$Y = C(\mathcal{M})\theta + \eta. \tag{3.2.13}$$

Remark 3.2. *Equation* (3.2.1) *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

$$\tilde{c}_t(m) = \begin{cases} \tilde{c}_{P,t}(m) & m = 1, \ldots, \mathcal{M}, \\ \tilde{c}_{S,t}(m) & m = \mathcal{M}+1, \ldots, 2\mathcal{M}, \\ \tilde{c}_{R,t}(m) & m = 2\mathcal{M}+1, \ldots 3\mathcal{M}, \end{cases} \tag{3.2.14}$$

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

$$\widehat{y}_{t+h|T}(m) = \tilde{c}_t(m)^\top \widehat{\theta}(m). \tag{3.2.15}$$

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

$$\widehat{y}_{t+h|T}(w) = \sum_{m=1}^{3\mathcal{M}} w(m)\widehat{y}_{t+h|T}(m), \tag{3.2.16}$$

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 $\widehat{\eta}_{t+h}(w) = \sum_{m=1}^{3\mathcal{M}} w(m)\widehat{\eta}_{t+h}(m)$.

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 other frequentist approaches such as the Mallows criterion, (Cheng and Hansen, 2015); the presence of a structural break in the regressors 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

$$CV_{h,T}(m) = \frac{1}{\lfloor(1-\pi)T\rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} \tilde{\eta}_{t+h,h}(m)^2, \tag{3.2.17}$$

and the corresponding cross-validation selected model is $\widehat{m} = argmin_{1 \le m \le 3\mathcal{M}} CV_{h,T}(m)$; 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

$$\begin{aligned} CV_{h,T}(w) &= \frac{1}{\lfloor(1-\pi)T\rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} \tilde{\eta}_{t+h,h}(w)^2 \\ &= \frac{1}{\lfloor(1-\pi)T\rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} \left(\sum_{m=1}^{3\mathcal{M}} w(m)\tilde{\eta}_{t+h,h}(m) \right)^2. \end{aligned} \tag{3.2.18}$$

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

$$\widehat{w} = \underset{w \in \mathscr{H}^{3\mathcal{M}}}{\arg\min}\, CV_{h,T}(w), \tag{3.2.19}$$

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.

3.3 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.

3.3.1 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 3.3. *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.*[2]

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

$$H_G = \frac{\Lambda_1^\top \Lambda_1}{N} \frac{G_r^\top \tilde{F}_P}{T} V_{NT,r}^{-1}, \tag{3.3.1}$$

where $V_{NT,r}$ is a diagonal matrix of the first r eigenvalues of $(NT)^{-1} X X^\top$ 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 able to specify a normalisation basis with a valid probability limit[3] as

$$H_{\Xi,r} = \frac{\Xi^\top \Xi}{N} \frac{G^\top \tilde{F}_P}{T} V_{NT,r}^{-1}. \tag{3.3.2}$$

[2] See Sections B.1.5 and B.1.7

[3] See Lemma B.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

$$H_1 = \frac{\Lambda_1^\top \Lambda_1}{N} \frac{F_1^\top \tilde{F}_1}{T_1} V_{NT,1}^{-1}, \quad H_2 = \frac{\Lambda_2^\top \Lambda_2}{N} \frac{F_2^\top \tilde{F}_2}{T_2} V_{NT,2}^{-1}, \tag{3.3.3}$$

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 requires allowing for a break in the forecasting equation. This is algebraically equivalent to using the post-break data to estimate the factors and forecasting equation, 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 3.1. *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 3.1 in provided in Section B.1.4. Because $\tilde{F}_2$ estimates $F_2 H_2$, and $\tilde{F}_1$ estimates $F_1 H_1$, Proposition 3.1 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 3.1. *Under Assumptions 1 to 8, as $N, T \to \infty$,* *a) The pseudo-factors $\tilde{F}_P$ satisfy:*

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

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

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

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

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

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

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

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

Theorem 3.1 (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 3.1 (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 3.1 (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 $F_1 H_1$ and $F_2 H_2$ respectively. Because the normalisation bases H_1 and H_2 generally differ, this necessitates a break in the forecasting equation. Theorem 3.1 (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.

3.3.2 Forecasting Bias-variance Trade-offs

Model and Expansion Results

Next, we provide the precise theoretical analysis of the bias-variance tradeoffs for out-of-sample forecasting using different factor estimators. Without loss of generality, we assume that the lag structure of the forecasting equation is known and only contains one lag of f_t. The general case of $q > 1$ lags of f_t follows at the cost of more complex notation after suitably redefining the regressor matrices, and an extension to an unknown lag structure can be handled by our model averaging framework in Section 3.3.3. 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) $E\left(\eta_{t+h}|\mathcal{F}_t\right) = 0$.

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) $E\|c_t\|^4 \leq M$, $\mathbb{E}\eta_t^4 \leq M$, *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* $\Omega_{CC,\eta} = \sum_{|j|<h} E\left(c_t c_{t-j}^\top \eta_{t+h}\eta_{t+h-j}\right) > 0$.

e) There exists a set $\mathcal{S}$ *of finite cardinality such that* $y \perp\!\!\!\perp \lambda_{1i} e_{iT}$.

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. The variance $\Omega_{CC,\eta}$ incorporates autocovariances up to order less than h because 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 before and after the break. Assumptions 9 (c) and 9 (d) are standard moment conditions and the central limit theorem, the latter of which is satisfied under standard weak dependence conditions. Assumption 9 (e) allows for limited dependence between y_T and the idiosyncratic error, and is looser than independence and zero mean required by Assumption E and Assumption 3c) of Bai and Ng (2006) and Gonçalves and Perron (2014), respectively.

Using the rates derived in Theorem 3.1, we show that the pseudo-, split-sample, and rotated factor methods have the following expressions for their out-of-sample biases and variances.

Proposition 3.2. *Under Assumptions 1 to 9 (d), as $N, T \to \infty$ and under the condition that $N \propto T$, then:*

$$\begin{aligned}
\text{bias}(\tilde{c}_{P,T}^{\top}\widehat{\theta}_P) = &\left[\left((I-Z) - \left(\frac{\Lambda_1^{\top}\Lambda_1}{N}\right)^{-1}\left(\frac{\tilde{F}_P^{\top}G_r}{T}\right)^{-1}\frac{\tilde{F}_P^{\top}G_p}{T}\frac{W^{\top}W}{N}\right)\left(f_T - \frac{G_p^{\top}\tilde{C}_P}{T}\left(\frac{\tilde{C}_P^{\top}\tilde{C}_P}{T}\right)^{-1}\tilde{c}_{P,T}\right)\right. \\
&\left. - \left(\frac{\Lambda_1^{\top}\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^{\top}e_T}{N}\right]^{\top}\beta + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right),
\end{aligned} \tag{3.3.4}$$

$$\text{bias}(\tilde{c}_{S,T}^{\top}\widehat{\theta}_S) = \frac{-e_T^{\top}\Lambda_2}{N}\left(\frac{\Lambda_2^{\top}\Lambda_2}{N}\right)^{-1}\beta + O_p\left(\frac{1}{\delta_{NT}^2}\right), \tag{3.3.5}$$

$$\begin{aligned}
\text{bias}(\tilde{c}_{R,T}^{\top}\widehat{\theta}_R) = &\left((I-Z)\left(f_T - \frac{\tilde{F}_2^{\top}\tilde{C}_{R,2}}{T}\left(\frac{\tilde{C}_R^{\top}\tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\right) - Z\left(\frac{\Lambda_2^{\top}\Lambda_2}{N}\right)^{-1}\frac{\Lambda_2^{\top}e_T}{N}\right)^{\top}\beta \\
&+ O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right),
\end{aligned} \tag{3.3.6}$$

$$\text{var}\left(\tilde{c}_{P,T}^{\top}\widehat{\theta}_P\right) = \text{var}\left(\tilde{c}_{S,T}^{\top}\widehat{\theta}_S\right) = \text{var}\left(\tilde{c}_{R,T}^{\top}\widehat{\theta}_R\right) = O_p\left(T^{-1}\right). \tag{3.3.7}$$

Equations (3.3.4) to (3.3.6) in Proposition 3.2 express the bias in terms of the rotational break $(I - Z)$, shift break $\left(W^{\top}W/N\right)$, and inherent estimation uncertainty. Equation (3.3.7) shows that the variance terms for all three forecasts are of order $O_p\left(T^{-1}\right)$, with their specific forms are relegated to Section B.2 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 different forecasts.

Theorem 3.2. *Under Assumptions 1 to 9 (d), 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 $\nu < 0.5$, and if Assumption 9 (e) additionally holds,*

$$\begin{aligned}
&\left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - c_T^{\top}\theta\right\|^2 \asymp_p \left\|\tilde{c}_{S,T}^{\top}\widehat{\theta}_S - c_T^{\top}\theta\right\|^2, \\
&\left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2 / min\left[\left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2, \left\|\tilde{c}_{S,T}^{\top}\widehat{\theta}_S - c_T^{\top}\theta\right\|^2\right] \xrightarrow{p} 0, && \textit{for} \quad \alpha = 0.5, \\
&\left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2 / \left\|\tilde{c}_{S,T}^{\top}\widehat{\theta}_S - c_T^{\top}\theta\right\|^2 \xrightarrow{p} 0, \\
&\left\|\tilde{c}_{S,T}^{\top}\widehat{\theta}_S - c_T^{\top}\theta\right\|^2 / \left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - c_T^{\top}\theta\right\|^2 \xrightarrow{p} 0, && \textit{for} \quad 0.5 < \alpha < 1, \textit{ and} \\
&\left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2 \asymp_p \left\|\tilde{c}_{S,T}^{\top}\widehat{\theta}_S - c_T^{\top}\theta\right\|^2, \\
&\left\|\tilde{c}_{P,T}^{\top}\widehat{\theta}_P - c_T^{\top}\theta\right\|^2 / max\left[\left\|\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - c_T^{\top}\theta\right\|^2, \left\|\tilde{c}_{S,T}^{\top}\widehat{\theta}_S - c_T^{\top}\theta\right\|^2\right] \xrightarrow{p} \infty, && \textit{for} \quad \alpha = 1,
\end{aligned}$$

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

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

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 3.2 provides the detailed comparisons between different forecasts produced by each set of factor estimates for varying sizes of shift and rotational breaks, which we summarise into four cases. Theorem 3.2 (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 3.2 (b) shows how the rotated factors weakly dominate the pseudo-factors when the rotational break is small. Additionally, it shows that the rotated factors have MSFEs smaller than the split-sample approach for all but very large shift breaks corresponding to $\alpha = 1$. Theorem 3.2 (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 biases in the rotated and pseudo-factors, induced by the break terms, are of the same order of magnitude as the loss in efficiency from using the split-sample factors. Therefore, the specific ranking of each method in this region depends on the data-generating process. Theorem 3.2 (d) shows that 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 3.2 are summarised in Table 3.2.

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

Table 3.2: Summary of Theorem 3.2. 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 3.4. *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.*

3.3.3 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 $\{y_{t+h}, \tilde{c}_t(m) : t = 1-h, \ldots, T-h\}$. Compared to a leave-$h$-out[4] estimator $\tilde{\theta}_{T,h}(m)$ with the last h observations $\{y_{j+h}, \tilde{c}_j(m) : j = T-h+1, \ldots, T+h-1\}$ 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

$$MSFE_T(w) = \mathbb{E}\left(y_{T+h} - \widehat{y}_{T+h|T}(w)\right)^2 = \mathbb{E}\tilde{\eta}^2_{T+h,h}(w)^2. \tag{3.3.8}$$

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

$$\begin{aligned} CV_{h,T}(w) &= \frac{1}{T_2} \sum_{t=T_1+1-h}^{T-h} \tilde{\eta}_{t,h}(w)^\top \tilde{\eta}_{t,h}(w) \\ &= \tilde{L}_{T_2}(w) + \frac{1}{T}\eta_{(2)}^\top \eta_{(2)} + \frac{2}{\sqrt{T}}\tilde{r}_{1T}(w) \end{aligned} \tag{3.3.9}$$

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

$$\begin{aligned} \tilde{L}_{T_2}(w) &= \frac{1}{T_2} \sum_{t=T_1+1-h}^{T-h} \left(\mu_t - \tilde{\mu}_{t,h}(w)\right)^2 \\ &= \frac{1}{T_2}\left(\mu_{(2)} - \tilde{\mu}_{(2),h}(w)\right)^\top \left(\mu_{(2)} - \tilde{\mu}_{(2),h}(w)\right) \end{aligned} \tag{3.3.10}$$

[4]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.

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

$$\begin{aligned}\tilde{r}_{1T}(w) &= \frac{1}{\sqrt{T_2}}\left(\mu_2 - \tilde{\mu}_{2,h}(w)\right)^\top \eta_{(2)} \\ &= \sum_{m=1}^{M} w(m) \frac{1}{\sqrt{T_2}} \sum_{t=T_1+1-h}^{T-h} \left(\mu_t - \tilde{c}_t(m)^\top \tilde{\theta}_{t,h}(m)\right) \eta_{t+h} \\ &= \sum_{m=1}^{M} w(m) \tilde{r}_{1T}(m). \end{aligned} \tag{3.3.11}$$

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 $C(m)$.[5] This allows us to establish the asymptotic negligibility of $\tilde{r}_{1T}(m)$, and therefore legitimacy of the post-break cross-validation criterion.

Proposition 3.3. *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 $Q(m) = \text{plim}_{T\to\infty} \frac{1}{(1-\pi)^2}\frac{1}{T}(\mu_{(2)} - C_{2,H}(m)\theta(m))^\top(\mu_{(2)} - C_{2,H}(m)\theta(m))$, *and* $C_{2,H}(m)$ *are the post break rows of* $C_H(m)$.

Theorem 3.3. *Under Assumptions 1 to 9, we have for any* $h \geq 1$, *fixed* $\mathcal{M}$ *and* w, *and* $N, T \to \infty$,

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

[5] 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.

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

Theorem 3.3 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 3.3 mirrors and extends Theorem 2 of Cheng and Hansen (2015) to allow for the case of a structural break in the factor structure.

Remark 3.5. *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 3.2 (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.*

3.4 Monte Carlo Study

3.4.1 Model Specification

We investigate the finite sample root mean squared forecast error (RMSFE) 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

$$Z = I_r + \frac{R}{N^{1-\nu}}, \quad W = \frac{1.5 \times D}{N^{(1-\alpha)/2}}, \tag{3.4.1}$$

where the elements of R are drawn from a $N\left(0_{r^2}, I_{r^2}\right)$, and the ith row of D is drawn as $d_i \sim N(0, I_r)$. 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

$$x_{it} = \begin{cases} \lambda_{1,i}^{\top} f_t + \sqrt{\theta} e_{it}, & t = 1, \ldots, \lfloor \pi T \rfloor \\ \lambda_{2,i}^{\top} f_t + \sqrt{\theta} e_{it}, & t = \lfloor \pi T \rfloor + 1, \ldots, T, \end{cases} \tag{3.4.2}$$

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% and approximately in line with empirical estimates.

The factors and errors are generated as follows:

$$f_{k,t} = \rho f_{k,t-1} + u_{it}, u_{it} \sim i.i.d.N(0, 1 - \rho^2), \tag{3.4.3}$$

$$e_{it} = \alpha e_{i,t-1} + v_{it}, \tag{3.4.4}$$

where $\rho \in \{0, 0.7\}$ captures the serial correlation in the factors, and ϵ_{it}, v_{it} are mutually independent with $v_t = (v_{1,t}, \ldots, v_{N,t})^{\top}$ being i.i.d. $N(0, \Omega)$ for $t = 1, \ldots, T$. For $t = 1$, $e_{.t} = (e_{1,1}, \ldots, e_{N,1})^{\top}$ 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.[6]

The regression equation for the forecast is

$$y_{t+h} = \beta_1 f_t + \beta_2 f_{t-1} + \beta_3 f_{t-2} + \eta_{t+h} \tag{3.4.5}$$

$$\eta_{t+h} = \sum_{j=1}^{h-1} \kappa^j \varepsilon_{t+h-j} \tag{3.4.6}$$

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

[6] Additional results for $\pi \in \{0.3, 0.7\}$ are similar relegated in Section B.4.

set of possible factors, the set of candidate regressors for the model averaging and model selection is

$$\mathcal{C}_t = (1, \tilde{f}_t^\top, \ldots, \tilde{f}_{t-q_{max}}^\top, y_t, \ldots, y_{t-p_{max}}). \tag{3.4.7}$$

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 RMSFE 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), Bayesian model averaging (BMA),[7] and simple averaging with equal weights. The model selection methods include the proposed leave-h-out cross-validation and Mallows selection.

3.4.2 Results

Across most parameter values and all forecast horizons, the proposed post-break leave-h-out cross-validation averaged forecasts yield the smallest RMSFE and hence the best forecasting performance. The leave-h-out cross-validation tends to perform the best for high degrees of serial dependence in the error term, as expected. For compactness, we report the results of each factor estimator and the model averaging estimators in Section 3.4.2, respectively, with poorly performing models omitted. The RMSFE are normalised by the RMSFE 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,

[7] 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.

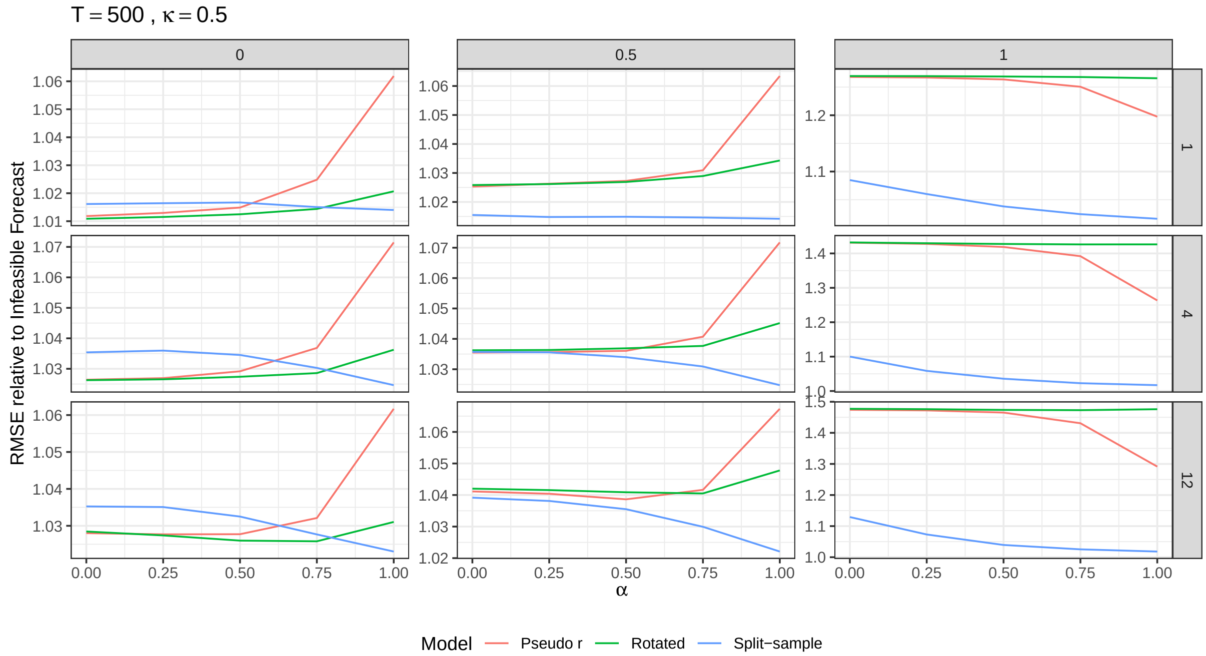

Figure 3.1: Relative MSFE for each factor estimator and proposed post-break cross-validation weighted forecasts, faceted by h (rows) and ν (columns), $\kappa = 0.5$ for moderate serial correlation in errors, $q_{max} = p_{max} = 4$.

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. For multi-step-ahead forecasts, the advantage of the cross-validation method is most prominent when the forecast horizon is long and the serial dependence is strong, similar to Cheng and Hansen (2015). 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.

3.4.3 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 3.4.3

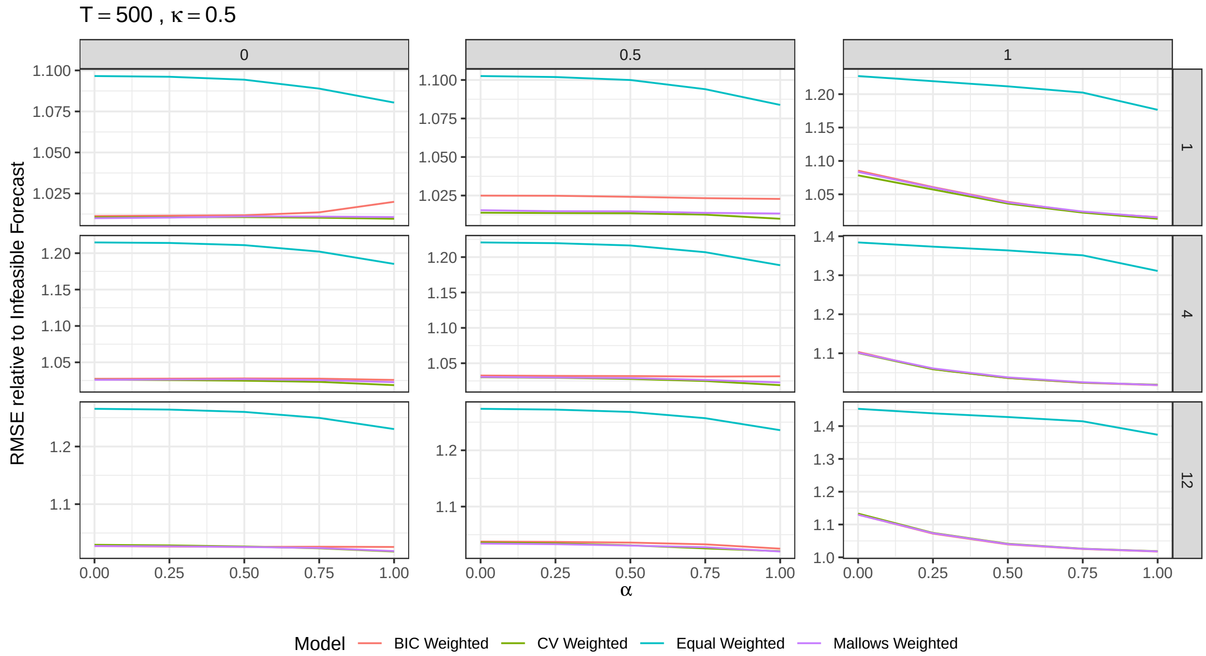

Figure 3.2: Relative MSFE for model averaged forecasts, faceted by h (rows) and ν (columns), $\kappa = 0.5$ for moderate serial correlation in errors, $q_{max} = p_{max} = 4$.

displays the results of this exercise. 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.

3.5 Empirical Study

3.5.1 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.

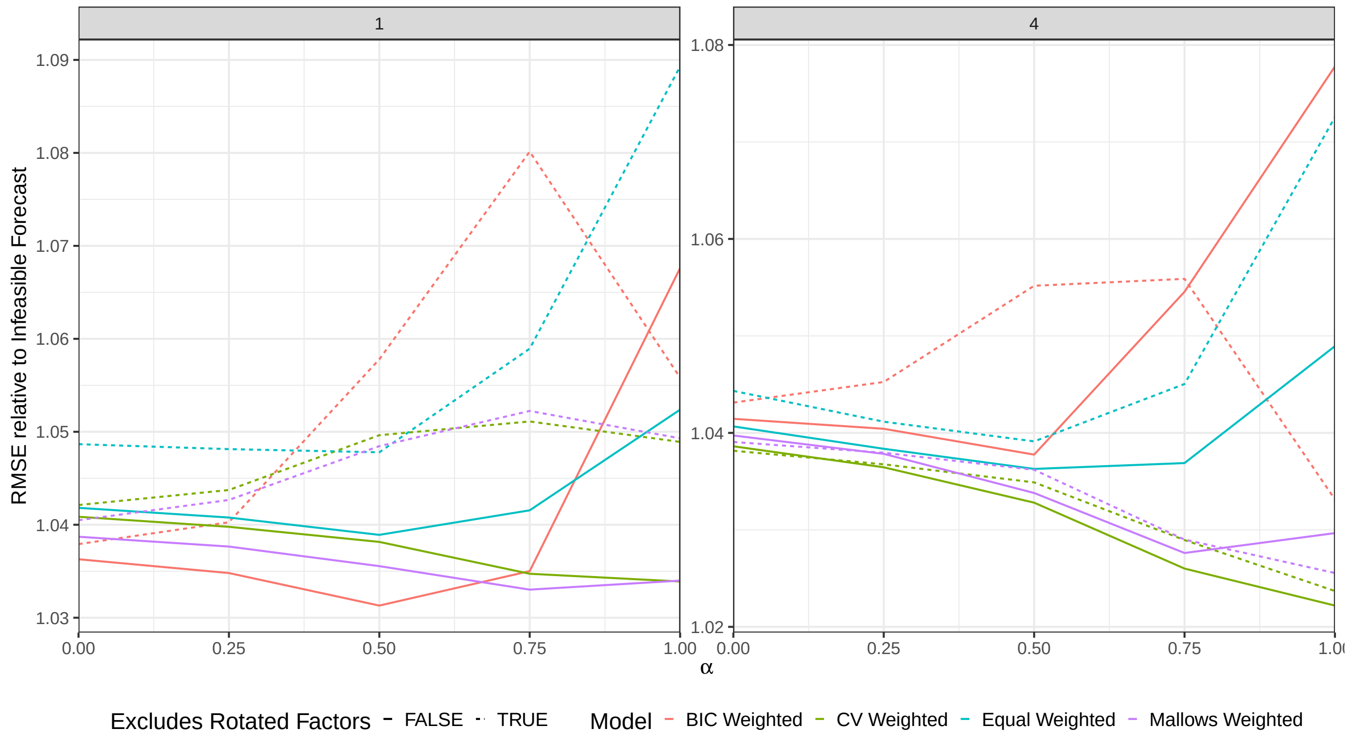

Figure 3.3: Relative MSFE, faceted by h, $\kappa = 0.1$ for mild serial correlation, $q_{max} = p_{max} = 4$. Solid line represents model averaged forecasts where the model set includes the rotated factors $\tilde{F}_R$, dashed line represents model averaged forecasts where the model set excludes $\tilde{F}_R$.

We consider the FRED-QD database of McCracken and Ng (2020), which consists of 235[8] U.S. macroeconomic time series, 124 of which are non-top-level aggregates used to estimate the factors. As our method only deals with one possible structural break, we focus on the subsample of ranging from 1959 Q3 to the 2008 Q4, as existing estimators tend to find evidence of a structural break in 1984 Q3 associated with the Great Moderation, and in 2008 Q4 associated with the Global Financial Crisis (see e.g. Stock and Watson, 2009; Breitung and Eickmeier, 2011; Baltagi et al., 2021).

3.5.2 Methodology

As in Stock and Watson (2012b), 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

[8]This number is obtained by following McCracken and Ng (2020), and suitably removing series which cannot be included after the suggested transformation(s).

$r = 0$ to 10, rather than $r = 0$ to $r = 50$ as in Stock and Watson (2012b).

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 data post-break, the initial length of the expanding window is set to be 24, 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 (2012b) 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 (2012b). Our experience shows that this omission results in forecasts of very similar quality, and is unnecessary.

3.5.3 Results

Table 3.3 reports the percentiles of the distributions of the one-, two-, and four-quarter ahead expanding window out-of-sample RMSFEs over the 235 series for the proposed forecasting methods, where the RMSFEs are relative to the DFM-5 benchmark. These results are remarkable, given that U.S. data over the Great Moderation period experienced reduced volatility and reduced predictability. Indeed, Stock and Watson (2012b) report that many advanced shrinkage methods fail to improve upon the DFM-5 benchmark, while Cheng and Hansen (2015) show that frequentist model averaging yields only modest gains for a minority of the series. Thus, it is particularly noteworthy that our proposed methods, which account for potential breaks in the factor structure provide further - albeit modest - gains.

We find that the model averaged forecasts generally exhibit the least deterioration and show potential for substantial improvement over the benchmark for at least half of the series. Of these, the Bayesian model averaging and cross-validation averaging methods are generally the most competitive. Examining the performance of each factor estimator helps reveal the source of these gains. Generally, of the three factor estimators, the rotated factors performs the best, followed closely by the pseudo-factors. Indeed, using the rotated factors alone yields very competitive forecasting performance. In contrast, the split-sample factors offer the worst performance, representing a significant bias-variance trade-off. This demonstrates that the gains from the model averaged forecasts 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 (2012b) 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.

Table 3.4 breaks down the results of Table 3.3 by category at the median RMSFE relative to the DFM-5 benchmark. Here, we can see that, on a categorical level, the cross-validation weighted forecast typically provides the most stable out of sample performance; the BIC weighted forecast is prone to providing much worse performance in the key categories of NIPA, Industrial Production, and Inventories, though it tends to better at forecasting Prices and Exchange Rates at the shorter one and two quarter ahead horizons. Of note, the equal weighted forecast is also able to outperform the CV weighted forecast in the categories of NIPA, Industrial Production and Employment, though only at the one and two quarter ahead horizons.

Table 3.3: Distributions of relative RMSEs by forecasting method, relative to DFM-5, h = 1, 2, 4, FREDQD Great Moderation Subsample (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
BIC Weighted	0.953	0.984	1.004	0.953	0.979	0.999	0.968	0.991	1.004
CV Select	0.958	0.994	1.017	0.971	0.998	1.025	0.980	1.002	1.025
CV Weighted	0.946	0.979	1.005	0.955	0.988	1.008	0.968	0.997	1.016

Equal Weighted	0.938	0.984	1.015	0.953	0.999	1.031	0.970	1.014	1.049
Mallows Select	0.960	0.991	1.015	0.952	0.980	1.004	0.970	0.992	1.010
Mallows Weighted	0.944	0.980	1.004	0.964	0.987	1.008	0.977	1.000	1.018
Pseudo r	0.987	1.006	1.028	0.982	1.001	1.031	0.986	1.001	1.017
Rotated	0.962	0.990	1.006	0.960	0.986	1.011	0.977	0.997	1.012
Split-sample	1.071	1.132	1.210	1.067	1.136	1.243	1.087	1.200	1.306

Note:

Entries are percentiles of distributions of relative RMSEs over the 235 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.

Table 3.4 breaks down the results of Table 3.3 by the 13 categories present in the FRED-QD dataset.

Table 3.4: Median RMSE by forecasting method and category of series, relative to DFM-5, expanding window forecast estimates, FREDQD Great Moderation Subsample (1959 Q3 - 2008 Q3, 1984 Q1 Break).

Group	BIC Weighted	CV Select	CV Weighted	Equal Weighted	Mallows Select	Mallows Weighted	Pseudo r	Rotated	Split-sample
h = 1									
NIPA	0.987	1.008	0.986	0.963	1.002	0.968	1.011	0.992	1.131
Industrial Production	0.993	0.985	0.974	0.952	1.024	0.959	1.010	0.986	1.167
Employment	0.990	0.999	0.966	0.982	0.999	0.964	1.008	0.974	1.141
Housing	0.991	0.938	0.896	0.909	1.001	0.926	0.917	0.979	1.044
Inventories	0.993	1.012	0.960	0.945	0.986	0.953	1.011	0.991	1.121
Prices	0.965	0.997	0.991	1.002	0.975	0.991	1.002	0.991	1.126
Earnings	0.982	0.984	0.987	0.994	0.999	0.978	1.007	0.986	1.120
Interest Rates	0.978	0.969	0.987	0.990	1.000	0.999	1.033	0.999	1.357
Money	0.966	0.989	0.984	0.988	0.977	0.990	1.000	0.970	1.107
Household Balance	1.006	0.996	1.005	1.009	1.010	1.002	1.014	0.994	1.083
Exchange Rates	0.980	0.985	1.011	1.035	0.981	1.018	0.982	0.982	1.413
Other	0.979	1.010	1.006	1.019	1.018	0.986	1.003	0.969	1.226
Stock Markets	0.995	0.998	0.990	0.959	0.983	0.986	1.024	1.005	1.051
Non Household Balance	1.005	0.973	0.957	0.971	1.003	0.971	0.995	1.012	1.090
h = 2									
NIPA	0.996	1.002	0.975	0.961	1.002	0.983	1.029	0.994	1.130
Industrial Production	0.968	1.008	0.952	0.932	0.965	0.942	1.005	0.992	1.045
Employment	0.984	0.999	0.965	0.955	0.979	0.979	0.999	0.992	1.085
Housing	1.003	0.995	1.004	1.009	1.004	1.011	1.013	1.016	1.154
Inventories	0.984	0.959	0.953	0.933	0.977	0.939	0.986	0.938	1.081
Prices	0.961	0.990	0.993	1.006	0.969	0.987	0.998	0.982	1.131
Earnings	0.983	0.992	0.997	1.044	0.981	1.009	1.007	1.007	1.227
Interest Rates	1.013	1.036	1.017	1.016	1.014	1.022	1.062	1.046	1.238
Money	0.966	0.982	0.976	0.974	0.980	0.981	0.995	0.972	1.190
Household Balance	0.962	1.008	1.004	1.016	0.989	0.992	0.975	0.970	1.115

Exchange Rates	0.983	1.019	0.999	1.012	0.985	0.996	0.993	0.983	1.217
Other	0.993	1.016	1.018	1.018	0.997	1.024	1.015	1.005	1.203
Stock Markets	0.959	1.049	1.054	1.061	0.957	1.012	0.996	0.949	1.243
Non Household Balance	0.980	0.981	0.986	1.005	0.979	0.985	0.981	0.982	1.085
h = 4									
NIPA	0.986	0.998	0.987	1.006	0.983	0.990	1.001	0.987	1.202
Industrial Production	0.982	0.973	0.952	0.980	0.969	0.976	0.999	0.987	1.127
Employment	0.990	1.001	0.980	0.972	0.990	0.995	1.001	0.987	1.170
Housing	0.998	0.997	0.992	1.048	0.995	0.994	1.023	1.004	1.419
Inventories	1.004	1.007	0.971	0.955	1.000	0.977	0.998	1.003	1.101
Prices	0.998	1.002	0.997	1.024	0.998	1.000	1.001	1.001	1.239
Earnings	0.994	1.008	1.004	1.033	0.986	1.009	1.004	1.001	1.185
Interest Rates	0.997	1.028	1.016	1.060	0.994	1.043	1.029	1.012	1.251
Money	0.982	1.003	0.998	0.988	0.979	1.001	1.000	0.996	1.212
Household Balance	0.993	1.024	1.017	0.991	0.995	1.004	0.995	0.998	1.111
Exchange Rates	0.990	1.008	1.001	1.075	0.986	1.006	1.001	0.998	1.528
Other	0.987	1.001	0.995	0.986	0.987	0.990	0.986	1.009	1.095
Stock Markets	0.979	1.116	1.104	1.044	0.980	1.017	0.977	0.978	1.196
Non Household Balance	0.994	1.003	1.005	1.023	0.998	1.014	1.003	1.004	1.213

Note:

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

Table 3.5: Distributions of relative RMSE by forecasting method, relative to DFM-5, h = 1, 2, 4, FREDQD Global Financial Crisis Subsample (1984 Q3 - 2019 Q4, 2008 Q3 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
BIC Weighted	0.968	0.993	1.029	0.950	0.986	1.013	0.923	0.978	1.014
CV Select	0.969	0.998	1.048	0.968	0.998	1.031	0.953	0.993	1.029
CV Weighted	0.968	0.998	1.029	0.961	0.996	1.021	0.948	0.987	1.019
Equal Weighted	0.977	1.013	1.046	0.969	1.026	1.083	0.952	0.999	1.045
Mallows Select	0.967	0.991	1.023	0.952	0.990	1.015	0.928	0.978	1.013
Mallows Weighted	0.981	1.021	1.078	0.970	1.001	1.048	0.949	0.990	1.028
Pseudo r	0.967	0.991	1.023	0.979	0.998	1.029	0.969	0.994	1.052
Rotated	0.963	0.994	1.028	0.965	0.990	1.019	0.942	0.978	1.015
Split-sample	1.095	1.179	1.272	1.152	1.271	1.663	1.042	1.161	1.312

Note:

Entries are percentiles of distributions of relative RMSEs over the 235 variables being forecasts, by series, at the specified forecast horizon. RMSEs are relative to the DFM-5 forecast, using an expanding out of sample exercise. All forecasts are direct.

Table 3.6: Median RMSE by forecasting method and category of series, relative to DFM-5, rolling forecast estimates, FREDQD Global Financial Crisis Subsample (1984 Q3 - 2019 Q4, 2008 Q3 Break)

Group	BIC Weighted	CV Select	CV Weighted	Equal Weighted	Mallows Select	Mallows Weighted	Pseudo r	Rotated	Split-sample
h = 1									
NIPA	0.961	0.960	0.958	0.972	0.960	1.008	0.960	0.959	1.208
Industrial Production	1.093	1.060	0.990	1.005	1.081	0.975	1.081	1.068	1.145
Employment	1.023	1.026	1.025	1.040	1.011	1.042	1.011	1.029	1.180
Housing	0.981	0.983	1.015	1.016	0.979	1.072	0.979	0.985	1.214
Inventories	1.020	1.025	1.025	0.993	1.035	1.029	1.035	1.017	1.232
Prices	0.992	0.997	0.998	1.013	0.992	1.033	0.992	0.993	1.144
Earnings	0.999	1.032	1.013	0.993	0.999	0.994	0.999	1.004	1.109
Interest Rates	0.948	0.949	0.966	1.019	0.946	0.988	0.946	0.948	1.317
Money	1.000	0.982	0.993	1.034	0.973	1.046	0.973	0.965	1.291
Household Balance	0.987	0.987	0.992	0.994	0.987	0.986	0.987	0.987	1.086
Exchange Rates	1.016	1.020	1.012	1.048	1.015	1.016	1.015	1.017	1.171
Other	0.941	0.939	0.942	0.984	0.938	0.960	0.938	0.944	1.243
Stock Markets	0.980	0.979	0.989	1.001	0.974	1.046	0.974	0.973	1.197
Non Household Balance	1.005	1.010	1.011	1.013	1.003	1.015	1.003	1.012	1.099
h = 2									
NIPA	0.911	0.936	0.928	0.919	0.935	0.945	0.993	0.950	1.233
Industrial Production	0.974	0.971	0.975	0.950	0.980	0.966	1.038	1.010	1.071
Employment	1.015	1.046	1.010	1.035	1.024	1.047	1.043	1.033	1.388
Housing	0.999	0.991	1.002	1.035	0.991	1.012	0.991	0.997	1.239
Inventories	0.972	0.996	0.954	0.925	0.961	0.914	1.010	0.982	1.099
Prices	0.987	0.992	1.002	1.022	0.989	0.995	0.992	0.988	1.216
Earnings	0.985	1.008	1.008	1.099	0.980	1.000	0.988	0.990	1.750
Interest Rates	0.991	0.981	0.974	1.082	0.989	1.001	0.981	0.990	1.752
Money	0.964	1.004	0.988	1.253	0.943	1.033	0.977	0.942	2.246
Household Balance	0.951	1.010	1.007	1.020	0.942	1.005	0.987	0.990	1.238

Exchange Rates	0.992	1.013	1.007	0.982	0.999	0.975	0.996	0.990	1.162
Other	0.953	1.012	1.044	1.152	0.933	1.093	0.996	0.946	1.803
Stock Markets	0.982	1.046	1.032	1.073	0.974	1.036	0.979	0.983	1.656
Non Household Balance	0.998	1.001	0.998	1.036	0.995	1.013	0.987	0.997	1.230
h = 4									
NIPA	0.925	0.943	0.952	0.973	0.928	0.967	0.987	0.941	1.251
Industrial Production	0.942	0.978	0.966	0.951	0.958	0.958	0.980	0.964	1.004
Employment	1.013	1.024	1.007	1.017	1.010	1.029	1.082	1.022	1.144
Housing	1.005	0.999	0.995	1.010	1.009	1.013	1.004	0.999	1.210
Inventories	0.980	0.966	0.979	1.003	0.971	1.023	0.986	0.976	1.195
Prices	0.975	0.985	0.984	0.997	0.976	0.982	0.987	0.974	1.111
Earnings	0.991	0.997	1.029	1.021	0.993	0.991	0.989	0.991	1.230
Interest Rates	1.010	1.009	1.005	1.058	1.009	1.013	0.997	1.010	1.296
Money	0.963	1.005	0.990	0.982	0.976	1.004	1.002	0.967	1.292
Household Balance	0.968	1.016	1.003	0.997	0.979	0.981	0.987	0.967	1.190
Exchange Rates	0.989	0.983	0.989	0.997	0.986	0.967	1.000	0.988	1.066
Other	0.986	0.992	0.977	0.970	0.987	0.942	0.987	0.992	1.069
Stock Markets	0.925	0.956	0.913	1.003	0.849	0.895	0.923	0.882	1.566
Non Household Balance	0.917	0.955	0.947	0.998	0.961	1.005	0.981	0.973	1.309

Note:

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

3.5.4 Robustness Check using Stock and Watson (2012b) Data

As a robustness check, we also consider the quarterly dataset used by Stock and Watson (2012b), 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. Evidence of a structural break for the Great Moderation is somewhat weaker for this specific dataset. Our results are available in Section B.5.1, and broadly similar to the results using FRED-QD.

3.6 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 (2012b) mistakenly say there are 109 series.

Chapter 4

Identification and Estimation of Structural Factor Models with External Instruments

4.1 Introduction

Since the seminal paper of Sims (1980), structural vector autoregressive (SVAR) models have remained a popular and indispensable tool for identifying and estimating the effects of macroeconomic shocks. A wide literature on SVAR modelling now exists (see the survery by Kilian and Lütkepohl, 2017), which has documented three challenging limitations that can preclude practitioners from estimating the true impulse responses to structural shocks. First, SVAR models often suffer from nonfundamentalness due to their inherent inability to include large information sets in a small equation system. Indeed, policymakers have access to information sets which span over literally hundreds of variables, and it is thus extremely difficult for the included variables in an SVAR system to span the space spanned by the true structural shocks, (Sims, 1992; Bernanke et al., 2005). Second, economic theory often implies the SVAR system specified is singular; that is, the number of variables included are driven by a smaller number of structural shocks. This point is often overlooked in the literature, as the number of structural shocks is usually implicitly set to be the same number of variables included as part of the specification. Work that does not rule this out show that specific steps must be used to address the problem of covariance singularity (e.g. Han, 2018; Chan et al., 2020). Third, the validity of the identification strategy required for an SVAR may be difficult to formally test for, in addition to each strategy typically requiring the development of a specific estimation—and therefore asymptotic theory—for inferential results. This latter point is particularly important, as this means that strategies such as the inclusion of factors (as in the case of a factor-augmented vector autoregression model) require the careful adaptation of existing theory in order to deal with nuisance parameters.

Ideally, practitioners would employ an estimation framework that addresses all three limitations in a holistic way, as any one of these precludes the ability to recover the true impulse responses. Instead however, these limitations have been typically dealt with in distinct, complex, and oftentimes opposing ways. For example, nonfundamentalness is typically addressed by including more variables that need to be justified to represent theoretical constructs (such as using GDP to represent “economic activity”), or factors estimated from a large dataset in a factor-augmented vector autoregression (FAVAR), (e.g. Bernanke et al., 2005). However, as noted earlier, macroeconomic theory often implies that a small number of structural shocks drive a large number of variables, (e.g. Sargent and Sims, 1977; Geweke, 1977) but including more variables therefore increases the chance of the specified system being singular. This problem is not even resolved by the FAVAR model. As we will show, most FAVARs do not effectively distinguish between the so-called dynamic factors and the static factors (i.e. the stacked lags of dynamic factors) whose dimension is typically larger. This directly implies that the static factors, and therefore the FAVAR system, are necessarily singular, which requires special treatment, often in the form of an additional dimensional reduction step, (e.g. Chan et al., 2020). Indeed, the direct augmentation of factors to a typical SVAR system has been noted by Stock and Watson (2005) to be inconsistent with the primitives of the dynamic factor model (DFM) used to compute factors in the first place, and it is therefore unclear whether this could produce a compelling basis for structural identification at all. Furthermore, these issues are compounded with the often wide, yet controversial, choices of identification strategies available to practitioners. Examples include short-run/long-run exclusion restrictions, sign restrictions, and identification via heteroskedasticity, (Stock and Watson, 2016). Among these, however, the method of using external instruments (or proxies) has increasingly gained popularity for their parsimonious set of identifying assumptions and ability to incorporate further external information into models. However, formal extensions of external instruments to a data-rich environment generally remain rare, and still suffer from limitations.

Our contribution to the literature is to provide a framework that addresses these three challenging limitations in a holistic approach. To do so, we propose the use of a structural factor model (SFM) which naturally deals with the problems of nonfundamentalness and covariance singularity, in conjunction with the use of multiple external instruments, which naturally allow for testing and selecting valid identifying restrictions. The SFMs we work with were introduced by Stock and Watson (2005), who were inspired by their success in macroeconomic forecasting. Unlike FAVAR models, SFMs are directly formulated from the factor structure and aim to combine the attractive features of large dimensional factor models and existing identification strategies employed in SVARs. Since their introduction, they have received increasing attention in the literature for their ability to estimate more reasonable and efficient impulse responses, which have been argued to be a direct consequence of their ability to parsimoniously summarise large information sets - something that is generally impossible or difficult to do within a standard SVAR framework. Similar to FAVAR models, the factor structure in an SFM implies the appearance of many extra nuisance parameters, necessitating the development of estimator and relevant asymptotic theory that is often specific to the identification scheme. To this end, we contribute to the literature by developing an asymptotic theory that explicitly acknowledges the random rotation problem included in factor models, similar to Bai and Ng (2006) and Yamamoto and Hara (2022). Furthermore, unlike existing attempts within the SFM literature, we establish the validity of identification and estimation of impulse responses with the use of multiple different instruments via the use of a generalised method-of-moments framework. Altogether, these provide the familiar theoretical basis for us to develop analytical formulas for statistical inference, such as confidence intervals, and overidentification and automatic moment selection procedures for testing/choosing valid instruments. A Monte Carlo experiment shows that the resulting estimators, tests, and selection criteria exhibit good finite sample performance.

Our work is related to the broader SVAR literature that achieves identification with external instruments, and structural factor models that attempt to extend existing identification strategies with factor models. For the former, the literature on identification in SVAR models is extensive, for which a comprehensive summary of mainstream identification approaches is provided by Kilian and Lütkepohl (2017). Since then, within the external instruments approach there have been further developments. Stock and Watson (2018) provide a comparison of local projection instrument variable (LP-IV) and SVAR-IV estimators, and show that SVAR-IV estimators do not require the strict lead-lag exogeneity assumption of LP-IV. Montiel Olea et al. (2021) derive the asymptotic theory for SVAR-IV. Cheng et al. (2021) derive the asymptotic theory for a generalised method-of-moments estimator that is robust to non-stationarity, but do not pursue overidentification or moment selection procedures. Schlaak et al. (2023) combine identification via heteroskedasticity and external instruments to sharpen inference. However, their framework is focused on using heteroskedasticity to achieve exact identification; overidentification using external instruments then proceeds in a proxy-SVAR framework. Importantly, their empirical study still focuses on using one instrument at a time. For the latter, the literature for SFMs is less developed and has typically focused on older identification strategies. Stock and Watson (2005) and Forni and Gambetti (2010a) use an SFM and employ a slow-fast identification, though neither provide formal theoretical justification. Forni et al. (2009) show that the presence of a factor structure in the data typically implies nonfundamentalness in fixed dimensional SVARs. Han (2015) and Han (2018) develops inferential theory for the identification of impulses using a diverging and finite number of zero restrictions respectively. Yamamoto and Hara (2022) develop inferential theory for the identification of impulse responses using heteroskedasticity in a FAVAR model. Forni and Gambetti (2010c) utilise a SFM with sign restrictions, the possibility of which is investigated by Gafarov et al. (2018).

To the best of our knowledge, there exist only a handful of papers that combine a structural factor model with external instrument identification. Stock and Watson (2012a) focus on one instrument at a time. Stock and Watson (2016), in their review, propose the use of a normalisation scheme that allows for direct application of an SVAR identification scheme with an SFM. Both of these only provide an estimation algorithm with little formal theoretical treatment of the proposed estimators. Han (2024) proposes a unifying framework for the global identification of structural impulse responses in factor models, but assumes that the number of static factors is equal to the number of primitive shocks, ignoring singularity issues that could occur. Our theory differs from the pre-existing literature in that we provide a formal theoretical treatment of the identification of impulse responses through the use of a factor structure that summarises a data-rich environment, a latent factor process that distinguishes between the static factor and primitive shocks, and a generalised method-of-moments approach which allows for the joint use of multiple instruments and leads to standard overidentification and instrument selection procedures to ensure that the identification conditions are valid. These features of our proposed framework allow us to respectively deal with the problems of nonfundamentalness, covariance singularity, and identification issues that plague typical SVAR models, in a holistic fashion.

In an empirical application on quarterly U.S. macroeconomic data, we apply the proposed method to study the dynamic causal effects of a monetary policy shock, and the validity of many popular monetary policy instruments proposed by the literature. We find evidence that all the monetary policy instrument considered are jointly valid, and that their joint use leads to more efficient and reasonable impulse responses. In particular, we show that using one instrument at a time is more prone to recovering puzzling responses.

The rest of the paper is organised as follows. Section 4.2 lays out the model setup, identification strategy, and estimation. Section 4.3 presents the asymptotic theory. Section 4.4 conducts the Monte Carlo study to confirm finite sample behaviour. Section 4.5 presents the empirical application. Section 4.6 concludes. All proofs are relegated to the Appendices. For notation, $P_Z = Z(Z^\top Z)^{-1}Z^\top$ and $M_Z = I - P_Z$ denote the projection and residual maker matrices for any matrix Z, respectively, $\|Z\| = \left[tr(Z^\top Z)\right]^{1/2}$ denotes the Euclidean norm, and $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and distribution, respectively.

4.2 Identification of Dynamic Responses in Structural Factor Models

4.2.1 Model Setup

Consider the following structural model for $t = 1, \dots, T$,

$$X_t = \Lambda F_t + e_t, \tag{4.2.1}$$

$$F_t = \sum_{j=1}^{p} \Phi_j F_{t-j} + G\eta_t, \tag{4.2.2}$$

$$\eta_t = A\zeta_t, \tag{4.2.3}$$

where $X_t = [x_{1t}, \dots, x_{Nt}]^\top$ is an N-dimensional vector, F_t is an r-dimensional set of unobserved factors, Λ is the corresponding $N \times r$ factor-loading matrix, and $e_t = [e_{1t}, \dots, e_{Nt}]^\top$ is an N-dimensional idiosyncratic error term. The matrix G is an $r \times q$ matrix of rank q which maps the q-dimensional reduced form shocks η_t to the lags of the factors, ζ_t are the structural shocks subject to the identification condition $E[\zeta_t \zeta_t^\top] = I_q$, and A is a $q \times q$ nonsingular matrix. Unlike many existing studies, we set focus on the case of $q \leq r$ to allow for dynamic factors. This is important, because the case of $q < r$ corresponds to a singular covariance structure in the static factors, rendering the FAVAR, and even many existing SFM approaches untenable. The assumption of stationarity in F_t

implies

$$(I_r - \Phi_1 L - \cdots - \Phi_p L^p)^{-1} = \sum_{s=1}^{\infty} \Psi_s L^s, \tag{4.2.4}$$

where Ψ_s is the coefficient matrix of the vector-moving average representation of Equation (4.2.2).

Let $\mathcal{F}_t = \left(F_t^\top, \ldots, F_{t-p}^\top\right)^\top$ collect the lags of F_t and $\Phi = [\Phi_1, \ldots, \Phi_p]$ collect the corresponding coefficient matrices. Plugging Equations (4.2.2) and (4.2.3) into Equation (4.2.1), we have

$$\begin{aligned} X_t &= \Pi \mathcal{F}_t + \Theta \eta_t + e_t & (4.2.5) \\ &= \Pi \mathcal{F}_t + \Gamma \zeta_t + e_t, & (4.2.6) \end{aligned}$$

where $\Pi = \Lambda\Phi$, $\Theta = \Lambda G$ and $\Gamma = \Theta A$. The matrix representations of Equations (4.2.5) and (4.2.6) follow as

$$\begin{aligned} X &= \mathcal{F}\Pi^\top + \eta\Theta^\top + e \\ &= \mathcal{F}\Pi^\top + \zeta\Gamma^\top + e & (4.2.7) \end{aligned}$$

where $X = [X_{p+1}, \ldots, X_T]^\top$, $\mathcal{F} = [\mathcal{F}_{p+1}, \ldots, \mathcal{F}_T]^\top$, $\eta = [\eta_{p+1}, \ldots, \eta_T]^\top$, $\zeta = [\zeta_{p+1}, \ldots, \zeta_T]^\top$, and $e = [e_{p+1}, \ldots, e_T]^\top$.

The OLS estimators for $\lambda_i^\top$ and Λ are, respectively,

$$\begin{aligned} \widehat{\lambda}_i &= \frac{1}{T}\sum_{t=1}^{T} \widehat{F}_t X_{it}, \\ \widehat{\Lambda} &= \frac{1}{T}\sum_{t=1}^{T} X_t \widehat{F}_t^\top. & (4.2.8) \end{aligned}$$

To estimate the reduced form shocks, first note that the dynamic factor model in Equation (4.2.7) implies a factor structure in the reduced form shocks, i.e. $X - \mathcal{F}\Pi^\top = \eta\Theta^\top + e$ itself exhibits a factor structure. Let

$$\widehat{X} = M_{\widehat{\mathcal{F}}} X \tag{4.2.9}$$

be a corresponding estimate of $\eta\Theta^\top + e$. It follows that the reduced form shocks can then be estimated via a second-stage principal components estimator. We set the estimated reduced shocks $\widehat{\eta}$ equal to $\sqrt{T-p}$ times the eigenvectors corresponding to the first q eigenvalues of the $(T-p)\times(T-p)$ covariance matrix $\widehat{X}\widehat{X}^\top$. The OLS estimator for G can be computed as

$$\widehat{G} = \widehat{F}^\top\widehat{\eta}\left(\widehat{\eta}^\top\widehat{\eta}\right)^{-1} = \frac{1}{T-p}\sum_{t=p+1}^{T}\widehat{F}_t\widehat{\eta}_t^\top, \tag{4.2.10}$$

which follows because $\widehat{\mathcal{F}}$ and $\widehat{\eta}$ are orthogonal by design, and $\widehat{\eta}^\top\widehat{\eta}/(T-p) = I_q$ by eigenidentity. The estimator for Θ can be computed as

$$\widehat{\Theta} = \widehat{\Lambda}\widehat{G}, \tag{4.2.11}$$

where we additionally use $\widehat{\theta}_i$ to denote the transposition of the ith row of $\widehat{\Theta}$. The estimator for Φ via OLS is given by

$$\widehat{\Phi} = \widehat{F}^\top\widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1} = \left(\sum_{t=p+1}^{T}\widehat{F}_t\widehat{\mathcal{F}}_t\right)\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}. \tag{4.2.12}$$

Given $\widehat{\Phi}$, the estimates for $\widehat{\Psi}_s$, $s = 1, 2, \ldots$ follow by inverting the lag polynomial in Equation (4.2.4).

Remark 4.1. *We focus on the principal components of $\widehat{X}\widehat{X}^\top$ to estimate the reduced form shocks. An alternative way to estimate η_t is to conduct a spectral decomposition of the variance of the residuals $\widehat{\varepsilon}_t = \widehat{F}_t - \widehat{\Phi}\widehat{\mathcal{F}}$ (see, e.g. Forni et al., 2009; Forni and Gambetti, 2010a). Specifically, let $\widehat{\Sigma}_\varepsilon$ be the sample covariance matrix of $\widehat{\varepsilon}$, $\widehat{\mathcal{D}}$ be a diagonal matrix consisting of the first q eigenvalues of $\widehat{\Sigma}_\varepsilon$ in descending order, and $\widehat{\mathcal{S}}$ be the corresponding associated eigenvectors. Then, $\breve{\eta}_t \equiv \widehat{\mathcal{D}}^{-1/2}\widehat{\mathcal{S}}\widehat{\varepsilon}_t$ is an alternative estimator for η_t because $\widehat{\Sigma}_\varepsilon$ is of rank q, asymptotically. Due to the use of a principal components fit, it is likely that the theory developed for this paper can also be adapted for $\breve{\eta}_t$. The simulation results of Han (2018) show that $\widehat{\eta}_t$ tends to produce more accurate estimates for η_t as measured by trace R^2 statistics, and we therefore leave the use of $\breve{\eta}_t$ to future research.*

4.2.2 Identification and Estimation with External Instruments

Suppose, without loss of generality, that we are interested in the effects of the first structural shock. The preceding model setup dictates that the impulse response function (IRF) to the entire N panel of time series X_t to a one unit increase in the first structural shock is given by

$$\frac{\partial X_t}{\partial \zeta_{1,t-s}} = \Lambda \Psi_s G a_1, \tag{4.2.13}$$

where a_1 denotes the first column of A, where its columns are partitioned as $A = \begin{bmatrix} a_1 & \dots & a_q \end{bmatrix}$. In the special case of $s = 0$, we set $\Psi_s = I_r$, so the contemporaneous response simplifies to

$$\frac{\partial X_t}{\partial \zeta_{1,t}} = \Lambda G a_1 = \Theta a_1. \tag{4.2.14}$$

The estimators for Λ, Ψ_s and G are described earlier, and thus it remains to find an appropriate estimator for a_1 to compute the IRF.

It is well known that principal components estimators are only consistent up to a rotation. Specifically, the principal components-based estimator $\widehat{\eta}_t$ is only able to estimate its unobserved counterpart η_t up to a rotation, which we denote as H_η, i.e. $\widehat{\eta}_t$ estimates $H_\eta^\top \eta_t$. The presence of this rotational basis H_η will generally affect the distribution of the individual components, which enter into the expression for the impulse response functions. We show that, however, the resulting estimators of the impulse response functions are not affected by this rotation, and thus the identification of the impulse response functions themselves is not affected. Identification proceeds by requiring $q-1$ restrictions to identify a_1 (assuming that the first element is fixed to unity). Unlike the typical SVAR-IV case, the use of principal components estimator $\widehat{\eta}_t$, which recovers $H_\eta^\top \eta_t$, implies that we are instead identifying $a_1^* = H_\eta^\top a_1$.

We are interested in identifying a_1^* with external instruments $Z_t \in R^k$, which satisfy i) $E(Z_t \zeta_{1t}) = \alpha \neq 0_k$, and ii) $E(Z_t \zeta_{jt}) = 0_k$ for $j \neq 1$, which are the instrument relevance and exogeneity conditions. We emphasise that these conditions are with respect to only the *contemporaneous* shocks - these SVAR-IV conditions permit the instruments to be correlated with lagged values of the non-target shocks, and is thus far less restrictive compared to the Local Projection (LP-IV) approach as noted by Stock and Watson (2018). Under these conditions, the instruments satisfy

$$\begin{aligned} E(H_\eta^\top \eta_t Z_t^\top) &= E(H_\eta^\top A \zeta_t Z_t^\top) \\ &= H_\eta^\top a_1 \alpha^\top \\ &= a_1^* \alpha^\top \in R^{q\times k}, \end{aligned} \tag{4.2.15}$$

and are thus able the identify a_1^* up to a scale. The case of $k = 1$ instrument corresponds to $q - 1$ restrictions and suffices to just identify a_1^*; the system is overidentified if $k > 1$. In the traditional SVAR setting, the reduced form shocks are estimated without the effect of H_η and thus the moment conditions are $E(\eta_t Z_t^\top) = a_1 \alpha^\top$; estimation then proceeds by regressing each reduced form shock on the first using the instrument(s) Z_t via two stage least squares (2SLS) as in Ramey (2016), or a generalised method-of-moments approach as in Cheng et al. (2021).

Without loss of generality, we normalise the first element of a_1^* to be 1.[1] This allows us to remove the constant 1 and define the parameter

$$\delta = \left[a_{12}^*, \ldots, a_{1q}^*\right]^\top \in R^{q-1}. \tag{4.2.16}$$

With $a_1^* = 1$, Equation (4.2.15) is therefore equivalent to the moment conditions

$$\begin{aligned} &E\left[\left((H_\eta^\top \eta)_{-1t} - \delta \eta_{1t}^*\right) \otimes Z_t\right] \\ =&E\left[\left(\eta_{-1t}^* - \delta \eta_{1t}^*\right) \otimes Z_t\right] = \mathbf{0} \in R^{k(q-1)}, \end{aligned} \tag{4.2.17}$$

where η_{1t}^* is the first element of $H_\eta^\top \eta_t$ and η_{-1t}^* is the rest of $H_\eta^\top \eta_t$ with η_{1t}^* removed. Let $\widehat{\eta}_{1t}$ and $\widehat{\eta}_{-1t}$ denote the principal components-based estimated counterparts of η_{1t}^* and η_{-1t}^*, respectively. We estimate δ by minimising the generalised method-of-moments (GMM) criterion

$$\mathcal{Q}_T(\delta) = \bar{g}_T(\delta)^\top W_T \bar{g}_T(\delta) \tag{4.2.18}$$

[1] This corresponds to an additional scale assumption that $H_{\eta,1}^\top \alpha_1 = 1$, and is analogous to the innocuous identification condition of setting the first element of a_1 to one in the case of η_t being observed or estimated without the effects of H_η, as is the case in a traditional SVAR setting. In practice, any normalisation can be used afterwards, such as the unit-effect normalisation.

using the empirical moments

$$\bar{g}_T(\delta) = \frac{1}{T-p}\sum_{t=p+1}^{T}[(\widehat{\eta}_{-1,t} - \delta\widehat{\eta}_{1,t}) \otimes Z_t] \tag{4.2.19}$$

and a weighting matrix W_T. The first-order condition yields the GMM estimator

$$\widehat{\delta} = (\mathcal{A}_T W_T \mathcal{A}_T^\top)^{-1}\mathcal{A}_T W_T \mathcal{G}_T \tag{4.2.20}$$

where

$$\mathcal{A}_T = I_{q-1} \otimes \left(\frac{1}{T-p}\sum_{t=p+1}^{T}\widehat{\eta}_{1,t}Z_t^\top\right), \quad \text{and} \quad \mathcal{G}_T = \frac{1}{T-p}\sum_{t=p+1}^{T}(\widehat{\eta}_{-1,t} \otimes Z_t). \tag{4.2.21}$$

If $W_T = I_{q-1} \otimes \left(\frac{1}{T-p}\sum_{t=p+1}^{T} Z_t Z_t^\top\right)^{-1}$ then $\widehat{\delta}$ corresponds to the equation by equation 2SLS estimator that is typically considered by the literature. By defining V_δ as the variance-covariance matrix of $\mathcal{G}_T$, an optimal two-step GMM estimator $\widehat{\delta}^o$ can be estimated as follows. In the first step, we use either $I_{(q-1)k}$ or $I_{q-1} \otimes \left(T^{-1}\sum_{t=1}^{T} Z_t Z_t^\top\right)^{-1}$ as the weighting matrix and compute the GMM estimator $\widehat{\delta}$. In the second step, we compute the feasible weight estimate as

$$\widehat{V}_\delta = [\mathbb{S}_{\widehat{\delta}} \otimes I_k]\widehat{\Sigma_i^{(1)}}[\mathbb{S}_{\widehat{\delta}} \otimes I_k]^\top, \tag{4.2.22}$$

where $\mathbb{S}_{\widehat{\delta}}$ is a $(q-1) \times q$ matrix such that $\mathbb{S}_{\widehat{\delta}}\widehat{\eta}_t = \widehat{\eta}_{-1t} - \widehat{\delta}\widehat{\eta}_{1t}$, which by definition is equal to $[\widehat{\delta} \vdots I_{q-1}(1:q-1)]$ where $I_{q-1}(1:q-1)$ collects the $q-1$ matrix of I_{q-1}, and $\widehat{\Sigma_i^{(1)}}$ is a feasible estimate of the variance of the instruments, detailed in Section 4.3.4. Note that, due to the effects of the generated regressor, this is different to the implicit 2SLS weighting matrix $I_{q-1} \otimes \left(\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t^\top\right)^{-1}$, even in the absence of conditional heteroskedasticity.

4.2.3 Comparison With Existing Approaches in Factor Models

The identification scheme in this paper uses the information from an external instrument of the structural shock, which is widely considered to be parsimonious in terms of identifying assumptions. The literature on combining identification with external instruments with a factor structure has seen increasing, though still limited, attention.

This approach was initially proposed by Stock and Watson (2012a). However, their methodology in implementing the identification condition differs somewhat - instead of regressing the reduced form shocks on each other in a typical IV regression as we have, they opt to regress the instrument on all remaining reduced form shocks. Although both approaches are consistent at recovering the same structural shock (see Montiel Olea et al., 2021, for a proof in an SVAR-IV context), our adoption of a generalised method-of-moments framework allows researchers to easily adapt the wide array of tools within that literature. This can be seen in how Stock and Watson (2012a) only report the estimated structural shock as estimated by each instrument one at a time, and investigate joint validity of instruments by reporting their correlations, an approach that precludes the ability to formally *test* joint validity. The theoretical validity of this approach in the context of factor models is also unclear; Stock and Watson (2016) provide only an unjustified bootstrap algorithm to calculate inferential quantities such as confidence intervals.

Stock and Watson (2016) justify this by proposing a "named factor" normalisation which allows for the direct implementation of existing SVAR identification methods. However, this requires that the space of the innovations to the first r common components span the space of the innovations of the remaining variables. Although the theoretical assumptions this additionally imposes are mild, in practice this is sensitive to the choice of named factor variables; one needs to ensure that the set of named variables 1) be sufficiently heterogeneous, 2) are sufficiently representative of the remaining groups of variables and 3) have innovations to their common components that sufficiently span the space of the innovations to the factors. As explained by Han (2024), this normalisation is mostly applicable in cases when the large dataset employed does not adequately capture the true structural shock, such as in the case of oil shocks.

4.2.4 Incorporating Structural Breaks

We briefly touch upon how the prevalent issue of structural instability can be incorporated into the system. A structural break in X_t at breakpoint $\lfloor \pi T \rfloor$ where π denotes the break fraction can be introduced as

$$X_t = \begin{cases} \Lambda_1 F_t + e_t, & t = 1, \dots, \lfloor \pi T \rfloor, \\ \Lambda_2 F_t + e_t, & t = \lfloor \pi T \rfloor + 1, \dots, T, \end{cases} \tag{4.2.23}$$

where Λ_1 and Λ_2 are the pre- and post-break factor loading matrices. Following Koo et al. (2023), the change in these matrices can be decomposed as

$$\Lambda_2 = \Lambda_1 Z + W, \tag{4.2.24}$$

where Z is a common rotational component interpreted as a change in the factor variance, and W is a shift component interpreted a change in the factor loadings. In general, both types of breaks can occur independently of one another, and incorporation of the structural break depends on which type of break occurs.

In the case where a shift break is present, i.e. $W \neq 0$, then the change in the factor loadings will in general contaminate the principal components estimator, (e.g. Bates et al., 2013). Let $\tilde{F}_1$ and $\tilde{F}_2$ denote the pre- and post-break principal components based estimates of the static factors, $\tilde{\Lambda}_1$ and $\tilde{\Lambda}_2$ their respective corresponding loadings, and define a set of "rotated factors" as

$$\widehat{F}_R = \left[\tilde{F}_1^\top, \tilde{Z}\tilde{F}_2^\top\right]^\top, \tag{4.2.25}$$

$$\tilde{Z} = \left(\tilde{\Lambda}_1^\top \tilde{\Lambda}_1\right)^{-1} \tilde{\Lambda}_1{}^\top \tilde{\Lambda}_2. \tag{4.2.26}$$

Chapters 2 and 3 show that $\widehat{F}$ effectively "purges out" the effects of structural breaks, and is a much more robust estimate of F_t. Thus, for the variables in X_t that do not experience a structural break, estimating the VAR coefficients in $\widehat{F}_R$ may offer better results.

In the case where a rotational break is present, then the change in the factor variance implies that the dynamics of F_t have changed, which requires a structural break in the VAR coefficients.

4.3 Asymptotic Theory

4.3.1 Assumptions

To analyse the properties of the proposed estimators, we make the following assumptions.

Assumption 1. *There exists a positive constant $M < \infty$ such that:*

a) $E\|f_t\|^4 < M$, $\frac{1}{T}\sum_{t=1}^{T} f_t f_t^\top \xrightarrow{p} \Sigma_F$, *and* $\frac{1}{T}\sum_{t=p+1}^{T} \mathcal{F}_t\mathcal{F}_t^\top \xrightarrow{p} \Sigma_{\mathcal{F}}$ *for some positive definite matrices* Σ_F *and* $\Sigma_{\mathcal{F}}$.

b) $E(\zeta_t\zeta_t^\top) = I_q$, $E\|\zeta_t\|^4 < M$, $E(\zeta_s\zeta_t^\top) = 0$ *for any* $s \neq t$, *and* $\frac{1}{T-p}\sum_{t=p+1}^{T} \zeta_t\zeta_t^\top \xrightarrow{p} I_q$.

c) $E\left\|\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} \zeta_t\mathcal{F}_t^\top\right\|^2 < M$.

Assumption 2. *There exists a positive constant M such that:*

a) $E\|\lambda_i\|^4 \leq M$, $\left\|\Lambda^\top\Lambda/N\right\| - \Sigma_\Lambda \xrightarrow{p} 0$ *for some* $\Sigma_\Lambda > 0$.

b) $\text{rank}(G) = q$, $\|G\| \leq M$, *and* $\|\Phi\| \leq M$.

c) All of the roots of $|I_q - \Phi_1 L - \cdots - \Phi_p L^p| = 0$ *are outside the unit circle.*

d) The matrices $\Sigma_F\Sigma_\Lambda$ *and* $G^\top\Sigma_\Lambda G$ *have distinct eigenvalues.*

Assumption 3. *There exists some 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. *The variables* $\{\lambda_i\}$, $\{\zeta_t\}$, *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 and T:*

a) For each t, $E\Big\|\frac{1}{NT}\sum_{s=1}^{T}\sum_{k=1}^{N} F_s[e_{ks}e_{kt} - E(e_{ks}e_{kt})]\Big\|^2 \leq M$*, and* $E\Big\|\frac{1}{NT}\sum_{s=p+1}^{T}\sum_{k=1}^{N} \zeta_s[e_{ks}e_{kt} - E(e_{ks}e_{kt})]\Big\|^2 \leq M$.

b) $E\Big\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{k=1}^{N} F_t\lambda_k^\top e_{kt}\Big\|^2 \leq M$.

c) $E\Big\|\frac{1}{\sqrt{TN}}\sum_{t=p+1}^{T}\sum_{i=1}^{N} \lambda_i e_{i,t-j}\mathcal{F}_t^\top\Big\|^2 \leq M$ *and* $E\Big\|\frac{1}{\sqrt{TN}}\sum_{t=p+1}^{T}\sum_{i=1}^{N} \lambda_i e_{i,t-j}\zeta_t^\top\Big\|^2 \leq M$ *for* $j = 0, 1, \ldots, p$.

d) For each t, $E\Big\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_i e_{it}\Big\|^2 \leq M$.

e) For each i, $E\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t e_{it}\Big\|^4 \leq M$ *and* $E\Big\|\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} \zeta_t e_{it}\Big\|^4 \leq M$*. For each i and for $j = 1, \ldots, p$,* $E\Big\|\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} F_{t-j}^\top e_{it}\Big\|^4 \leq M$.

Assumption 7. *For $i = 1, \ldots, N$,*

$$\frac{1}{\sqrt{T}}\begin{bmatrix} \mathrm{vec}\left(Z^\top\eta - E(Z^\top\eta)\right) \\ F^\top e_i \\ \mathrm{vec}\left(\mathcal{F}^\top\eta\right) \end{bmatrix} \xrightarrow{d} N\left(\mathbf{0}_{(qk+r+rpq)\times 1}, \Sigma_i\right).$$

Assumption 8. *The structural shock ζ_t is linked to the reduced form error by the linear transformation $\eta_t = A\zeta_t$, for some nonsingular matrix A, and $E(Z_t\zeta_t^\top) = [\alpha, 0_{k\times(q-1)}]$, where $\alpha \neq 0_k$.*

Assumptions 1 to 6 are either straight from, or slight modifications of, Assumptions A-G of Bai (2003) and Assumptions 1-6 of Han (2018). Assumption 1 (a) regulates the moments of the static factors. The positive definiteness of $\Sigma_{\mathcal{F}}$ is the same as Assumption A10 of Amengual and Watson (2007). Assumption 1 (b) restricts the structural shocks to be serially uncorrelated and have an identity covariance matrix. Assumption 1 (c) is not restrictive because the structural shocks ζ_t

and lags of F_t are commonly assumed to be uncorrelated in the VAR literature. Assumption 2 (a) follows from Assumption B of Bai (2003). Assumption 2 (d) is similar to Assumption G of Bai (2003), and ensures the existence of the probability limits of the rotation matrices H_F and H_η. Assumptions 3 and 5 allows for weak serial and cross-sectional correlation in the errors, corresponding to Assumptions C and E of Bai (2003). Assumption 4 is similar to Assumption D of Bai and Ng (2004). Assumption 6 is not stringent because all of the sums involve zero mean random variables. It is close to Assumption F of Bai (2003) and Assumption 6 of Han (2015). Assumption 7 are central limit theorems which can be obtained under primitive assumptions, (e.g. Theorem 5.15 or Theorem 7.1.2 of White, 1984; Brockwell and Davis, 1991, respectively). Assumption 8 formalises the instrument relevance and exogeneity conditions with respect to contemporaneous shocks. Note that $\zeta_t = A^{-1}\eta_t$ implies $E(\zeta_t|Z_{t-1}, Z_{t-2}, \dots) = 0$, i.e. that the structural shock is uncorrelated with lags of the instruments. This is consistent with the structural VAR literature, where the structural shocks are interpreted as unanticipated, and therefore, unpredictable conditional on historical information. Similar to Stock and Watson (2018), we allow Z_t to be correlated with lags of ζ_t, which is a much looser condition than is required for a local projection (LP-IV) approach.

It is well known that the principal components estimator is only able to estimate the true factors up to a rotational basis. Let $X^0 \equiv [X_1, \dots, X_T]^\top$ be the full data matrix. We define the following normalisation bases for F_t and η_t

$$H_F = \left(\frac{\Lambda^\top\Lambda}{N}\right)\left(\frac{F^\top\widehat{F}}{T}\right)\widehat{V}_F^{-1} \quad \text{and} \tag{4.3.1}$$

$$H_\eta = \left(\frac{\Theta^\top\Theta}{N}\right)\left(\frac{\eta^\top\widehat{\eta}}{T-p}\right)\widehat{V}_\eta^{-1} \tag{4.3.2}$$

where $\widehat{V}_F$ is an $r \times r$ diagonal matrix consisting of the first r largest eigenvalues of $X^0X^{0\top}/(NT)$ in descending order, and $\widehat{V}_\eta$ is a $q \times q$ diagonal matrix consisting of the first q eigenvalues of $\widehat{X}\widehat{X}^\top/(N(T-p))$ in descending order. Their corresponding probability limits are

$$\bar{H}_F = \operatorname{plim} H_F \quad \text{and} \tag{4.3.3}$$

$$\bar{H}_\eta = \operatorname{plim} H_\eta, \tag{4.3.4}$$

which can be shown by Lemma A3 and Proposition 1 of Bai (2003).

Analogously, we can define $H_{\mathcal{F}}$ such that $\widehat{\mathcal{F}}$ is a consistent estimator for $H_{\mathcal{F}}^{\top}\mathcal{F}_t$. Recall that $\widehat{\mathcal{F}}_t = \left[\widehat{F}_{t-1}^{\top}, \ldots, \widehat{F}_{t-p}^{\top}\right]^{\top}$ and $\mathcal{F} = \left[F_{t-1}^{\top}, \ldots, F_{t-p}^{\top}\right]^{\top}$. It follows that the rotational basis $H_{\mathcal{F}}$ can be defined as

$$H_{\mathcal{F}} \equiv I_p \otimes H_F \tag{4.3.5}$$

so that $\widehat{\mathcal{F}}$ is a consistent estimator for $H_{\mathcal{F}}^{\top}\mathcal{F}_t$. The probability limit of $H_{\mathcal{F}}$ is

$$\bar{H}_{\mathcal{F}} = I_p \otimes \bar{H}_F. \tag{4.3.6}$$

Remark 4.2. *We focus on the setup where the static factors F_t are unobserved. If some of the factors are treated as observed, then the model becomes a factor-augmented VAR (FAVAR) model, (e.g. Bernanke et al., 2005; Bai et al., 2016). Specifically, in the FAVAR setup, $\widehat{F}_t$ can be constructed by stacking the observed factors (regressors) and the estimated factors. Generally, the introduction of observed factors results in $\frac{1}{T}\sum_{t=1}^{T}\widehat{F}_t\widehat{F}_t^{\top}$ no longer being an identity matrix; the corresponding loading matrix should then be estimated by least squares as $\widehat{\Lambda} = X^{\top}\widehat{F}\left(\widehat{F}^{\top}\widehat{F}\right)^{-1}$. The rotational basis H_F then needs to be redefined as $\begin{bmatrix} I & 0 \\ 0 & H_F^u \end{bmatrix}$ where the identity matrix is the same dimension as the observed factors, and H_F^u is the normalisation basis for the unobserved factors defined in a similar manner to Equation* (4.3.1)*, i.e. H_F is defined in a suitable way that keeps the observed factors unchanged, but rotates the columns of the unobserved factors. Therefore, the theory developed in this paper can also be applied to FAVAR models with some minor adjustments.*

Remark 4.3. *In addition, the X_t series that are used for factor estimation need not be identical to the series whose impulse responses we are interested in. This can occur, for example, if a subset of X_t corresponding to non-aggregate series is used to estimate the factors as is commonly done (e.g. Stock and Watson, 2002a, 2012a, 2016). The corresponding loading matrix is still estimated by least squares as $\widehat{\Lambda}$, and the theory developed in this paper remains applicable.*

4.3.2 Asymptotic Distribution of Structural Parameters

We begin by deriving the asymptotic distribution of $\widehat{\delta}$, which is necessary to analyse $\widehat{a}$ and, therefore, the IRF. Let $\mathbb{S}_\delta$ be the infeasible counterpart of $\mathbb{S}_{\widehat{\delta}}$, i.e. a $(q-1)\times q$ matrix such that

$$\mathbb{S}_\delta \eta_t^* = \eta_{-1t}^* - \delta \eta_{1t}^*, \tag{4.3.7}$$

which by definition is equal to

$$\mathbb{S}_\delta = \left[\delta \;\; \vdots \;\; I_{q-1}(1:q-1)\right], \tag{4.3.8}$$

where $I_{q-1}(1:q-1)$ collects the last $q-1$ matrix of I_{q-1}.

Theorem 4.1. *Under Assumptions 1 to 8, and the conditions that $W_T \xrightarrow{p} W$, and $\sqrt{T}/N \to 0$ as $N, T \to \infty$,*

a) $\widehat{\delta}$ is a consistent estimator of δ, and

$$\sqrt{T}\left(\widehat{\delta}-\delta\right) \xrightarrow{d} \left(\mathcal{A}W\mathcal{A}^\top\right)^{-1}\mathcal{A}WN\left(0_{kq\times 1}, \left(\mathbb{S}_\delta \bar{H}_\eta^\top \otimes I_k\right)\Sigma_i^{(1)}\left(\mathbb{S}_\delta \bar{H}_\eta^\top \otimes I_k\right)^\top\right),$$

where $\mathcal{A} = I_{q-1} \otimes \mathbb{S}_1 \bar{H}_\eta^\top E(\eta_{1t} Z_t^\top)$, $\mathbb{S}_1 = [1, 0_{1\times(q-1)}]$, and $\Sigma_i^{(1)}$ is the upper left block of Σ_i.

b) The optimal choice of the weighting matrix is V_δ^{-1}, where $V_\delta = \mathcal{C}\Sigma_i^{(1)}\mathcal{C}^\top$ and $\mathcal{C} = \left[\mathbb{S}_\delta \bar{H}_\eta^\top \otimes I_k\right]$.

c) $\widehat{V}_\delta \xrightarrow{p} V_\delta$.

d) $\sqrt{T}\left(\widehat{\delta}^o - \delta\right) \xrightarrow{d} N\left(0, [\mathcal{A}V_\delta^{-1}\mathcal{A}^\top]^{-1}\right)$.

Theorem 4.1 shows that $\widehat{\delta}$ is consistent for δ and has a standard asymptotic normal distribution. Theorems 4.1 (b) and 4.1 (c) show the form of the infeasible weight matrix and the consistency of the feasible weight matrix, respectively. The use of the optimal weight matrix results in the optimal two-step GMM estimator $\widehat{\delta}^o$ in Theorem 4.1 (d), which follows from typical GMM arguments.

4.3.3 Asymptotic Distributions of Impulse Response Functions

In this subsection, we present the asymptotic distributions of the estimators of the IRFs. The IRFs are a function of $\widehat{a}_1, \widehat{\Lambda}, \widehat{\Psi}_s, \widehat{G}$ and $\widehat{\Theta}$. Because $\widehat{a}_1 = \left(1, \widehat{\delta}^\top\right)^\top$, the asymptotic properties of $\widehat{\delta}$ in Section 4.3.2 can be used by defining $\bar{\mathbb{S}}_1$ as the last $q-1$ columns of I_q, so that $\widehat{a} - a^* = \bar{\mathbb{S}}_1 \begin{bmatrix} 0 \\ \widehat{\delta} - \delta \end{bmatrix}$. Thus, we derive the asymptotic representations of the remaining terms and then combine these results to obtain the asymptotic distributions of the IRFs.

Proposition 4.1. *Under Assumptions 1 to 6,* $\widehat{G} - H_F^\top G \Sigma_\eta H_\eta = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

Proposition 4.2. *Under Assumptions 1 to 6, if* $\sqrt{T}/N \to 0$ *as* $N, T \to \infty$,

$$\sqrt{T}\left(\widehat{\theta}_i - H_\eta^{-1}\theta_i\right) = \left(H_\eta^{-1} G^\top H_F\right)\sqrt{T}\left(\widehat{\lambda}_i - H_F^{-1}\lambda_i\right) + o_p(1). \tag{4.3.9}$$

Proposition 4.3. *Under Assumptions 1 to 8, if* $\sqrt{T}/N \to 0$ *as* $N, T \to \infty$,

a)

$$\sqrt{T}\begin{bmatrix} \widehat{a}_1 - a_1^* \\ \widehat{\lambda}_i - H_F^{-1}\lambda_i \\ \operatorname{vec}(\widehat{\Psi}_s^\top - H_F^{-1}\Psi_s^\top H_F) \end{bmatrix} = B_s \frac{1}{\sqrt{T}} \begin{bmatrix} \operatorname{vec}\left(Z^\top \eta - E(Z^\top \eta)\right) \\ F^\top e_i \\ \operatorname{vec}\left(\mathcal{F}^\top \eta\right) \end{bmatrix} + o_p(1)$$
$$\xrightarrow{d} N(0_{(q+r+r^2)\times 1}, B_s \Sigma_i B_s^\top), \tag{4.3.10}$$

where

$$B_s = \begin{bmatrix} \bar{\mathbb{S}}_1\left(\mathcal{A}W\mathcal{A}^\top\right)^{-1}\mathcal{A}W(\mathbb{S}_\delta \bar{H}_\eta^\top \otimes I_k) & 0_{q\times r} & 0_{k\times rpq} \\ 0_{r\times qk} & \bar{H}_F^\top & 0_{r\times rpq} \\ 0_{r^2\times qk} & 0_{r^2\times r} & \bar{R}_s\left[\bar{H}_F G \otimes \left(\Sigma_{\mathcal{F}}\bar{H}_{\mathcal{F}}\right)^{-1}\right] \end{bmatrix},$$

$$\bar{R}_s = \sum_{j=1}^{s}\left(\bar{H}_F^\top \Psi_{j-1}\bar{H}_F^{-\top} \otimes \left[\bar{H}_F^{-\top}\Psi_{s-j}^\top \bar{H}_F, \bar{H}_F^{-\top}\Psi_{s-j-1}^\top \bar{H}_F, \ldots, \bar{H}_F^{-\top}\Psi_{s-j-p+1}^\top \bar{H}_F\right]\right),$$

$$\Psi_0 = I_r,$$

$$\Psi_s = 0_{r\times r} \quad \text{for} \quad s < 0,$$

b)

$$\sqrt{T}\begin{bmatrix}\widehat{a}_1 - a_1^* \\ \widehat{\lambda}_i - H_F^{-1}\lambda_i\end{bmatrix} = B_0 \frac{1}{\sqrt{T}}\begin{bmatrix}\text{vec}\left(Z^\top \eta - E(Z^\top \eta)\right) \\ F^\top e_i\end{bmatrix} + o_p(1)$$

$$\xrightarrow{d} N\left(0_{(q+r)\times 1}, B_0 \Sigma_i^{(1)} B_0^\top\right),$$

where $B_0 = \begin{bmatrix}\bar{\mathbb{S}}_1\left(\mathcal{A}W\mathcal{A}^\top\right)^{-1}\mathcal{A}W\left(\mathbb{S}_\delta \bar{H}_\eta^\top \otimes I_k\right) & 0_{q\times r} \\ 0_{r\times qk} & \bar{H}_F^\top\end{bmatrix}.$

Proposition 4.1 shows that $\sqrt{T}\left(\widehat{G} - H_F^\top G \Sigma_\eta H_\eta\right)$ has a degenerate limiting distribution if $\sqrt{T}/N \to 0$, therefore $\widehat{G}$ can be directly replaced by $H_F^\top G \Sigma_\eta H_\eta$ as if $\widehat{G}$ is observed when N is large relative to T. Proposition 4.2 is used for obtaining the asymptotic representations of the contemporaneous impulse responses. Proposition 4.3 (a) implies the asymptotic distribution of the IRFs. Proposition 4.3 (b) is simply Proposition 4.3 (a) but without the effects of $\widehat{\Psi}_s$, and is used for simplifying the results for the contemporaneous IRFs. Theorem 4.1 and Propositions 4.1 to 4.3 together are applied to obtain the asymptotic distributions of the dynamic IRFs over time, as summarised in the following theorem.

Theorem 4.2. *Under Assumptions 1 to 8, and the conditions that* $W_T \xrightarrow{p} W$, *and* $\sqrt{T}/N \to 0$ *as* $N, T \to \infty$,

a) For the contemporaneous IRFs of X_{it} *to* ζ_{1t},

$$\sqrt{T}\left(\widehat{\theta}_i^\top \widehat{a}_1 - \theta_i^\top a_1\right) = \sqrt{T}\bar{Q}_{1,i}\begin{bmatrix}\widehat{a}_1 - a_1^* \\ \widehat{\lambda}_i - H_F^{-1}\lambda_i\end{bmatrix} + o_p(1)$$

$$\xrightarrow{d} N\left(0, \bar{Q}_{1,i} B_0 \Sigma_i B_0^\top \bar{Q}_{1,i}^\top\right),$$

where $\bar{Q}_{1,i} = \left(\theta_i^\top \bar{H}_\eta^{-\top} C_1 + a_1 G^\top \bar{H}_F C_2\right)$, $C_1 = [I_q \vdots 0_{q\times r}]$, *and* $C_2 = [0_{r\times q} \vdots I_r]$.

b) For the IRFs of X_{it} to $\zeta_{1,t-s}, (s \geq 1)$,

$$\sqrt{T}\left(\widehat{\lambda}_i^\top \widehat{\Psi}_s \widehat{G}\widehat{a}_1 - \lambda_i^\top \Psi_s G a_1\right) = \sqrt{T}\bar{Q}_{2,i} \begin{bmatrix} \widehat{a}_1 - a_1^* \\ \widehat{\lambda}_i - H_F^{-1}\lambda_i \\ \mathrm{vec}\left(\widehat{\Psi}_s^\top - H_F^{-1}\Psi_s H_F\right) \end{bmatrix} + o_p(1)$$

$$\xrightarrow{d} N(0, \bar{Q}_{2,i} B_s \Sigma_i B_s^\top \bar{Q}_{2,i}^\top),$$

where

$$\bar{Q}_{2,i} = \lambda_i^\top \Psi_s G \Sigma_\eta \bar{H}_\eta C_3 + a_1^\top G^\top \Psi_s^\top \bar{H}_F C_4 + \left(\lambda_i^\top \bar{H}_F^{-\top} \otimes a_1^\top G^\top \bar{H}_F\right) C_5,$$

and $C_3 = [I_q \vdots 0_{q\times r} \vdots 0_{q\times r^2}]$, $C_4 = [0_{r\times q} \vdots I_r \vdots 0_{r\times r^2}]$ and $C_5 = [0_{r^2\times q} \vdots 0_{r^2\times r} \vdots I_{r^2}]$.

Theorem 4.2 establishes the consistency and asymptotic normality of the dynamic IRFs. Note that despite our estimator $\widehat{a}_1$ recovering $a_1^* = H_\eta^\top a_1$ and therefore being subject to the effect of the principal components rotation, our resulting estimators for the IRFs can consistently estimate the true impulse responses without the effect of any rotations. Additionally, we do not need H_η or H_F to estimate their asymptotic variances. Hence, Theorem 4.2 is the main result necessary for frequentist inference and construction of valid confidence intervals in empirical analysis. We discuss practical implementation of the covariance matrices in the next subsection.

Remark 4.4. *The structural factor model can also offer a way to test conventional identifying restrictions employed in SVAR models, such as short-run and long-run exclusion restrictions on the impulse responses. It is straightforward to implement any tests for zero impulse responses using Theorem 4.2. However, note that this approach does not establish a consistent estimator for A per se; and hence simply gives some theoretical justification to the often ad-hoc practice of checking for "reasonable" impulse responses often employed in the literature, (e.g. Bernanke et al., 2005).*

4.3.4 Covariance Matrix Estimation

We next detail feasible estimation of the covariance matrices for the IRFs in Theorem 4.2. First note that the idiosyncratic errors can be consistently estimated as

$$\widehat{e}_t = X_t - \widehat{\Lambda}\widehat{F}_t, \tag{4.3.11}$$

and let $\widehat{e}_{it}$ denote the ith element of $\widehat{e}_t$. We define an estimator for Σ_i as

$$\widehat{\Sigma}_i = \frac{1}{T-p}\sum_{t=p+1}^{T} \xi_{it}\xi_{it}^{\top}, \tag{4.3.12}$$

where

$$\xi_{it} = \begin{bmatrix} \operatorname{vec}\left(Z_t^{\top}\widehat{\eta}_t - \frac{1}{T-p}(Z^{\top}\widehat{\eta})\right) \\ \widehat{F}_t^{\top}\widehat{e}_{it} \\ \operatorname{vec}\left(\widehat{\mathcal{F}}_t^{\top}\widehat{\eta}_t\right) \end{bmatrix},$$

and

$$\widehat{R}_s = \sum_{j=1}^{s}\left(\widehat{\Psi}_{j-1} \otimes \left[\widehat{\Psi}_{s-j}^{\top}, \widehat{\Psi}_{s-j-1}^{\top}, \ldots, \widehat{\Psi}_{s-j-p+1}^{\top}\right]\right)$$

with $\widehat{\Phi}_0 = I_r$ and $\widehat{\Psi}_s = 0_{r\times r}$ for $s < 0$. Similarly, an estimator for $\Sigma_i^{(1)}$ can be defined as

$$\widehat{\Sigma}_i^{(1)} = \frac{1}{T-p}\sum_{t=p+1}^{T} C_6\xi_{it}\xi_{it}^{\top}C_6^{\top}, \tag{4.3.13}$$

where $C_6 = \begin{bmatrix} I_q & 0_{q\times r} & 0_{q\times rpq} \\ 0_{r\times qk} & I_r & 0_{r\times rpq} \end{bmatrix}$. When e_t is serially correlated, the HAC estimators for the asymptotic variances can be readily constructed following the arguments of Bai (2003) and Han and Inoue (2015); cross-sectional correlation in e_t can be additionally accommodated via a CS-HAC estimator following Bai and Ng (2006) and Gonçalves and Perron (2020). Estimators for B_0 and B_s for $s = 1, \ldots, h$ follow by appropriate replacement of their unknown quantities with their feasible counterparts

$$\widehat{B}_0 = \begin{bmatrix} \bar{\mathbb{S}}_1 \left(\mathcal{A}_T W_T \mathcal{A}_T\right)^{-1} \mathcal{A}_T W_T \left(\mathbb{S}_{\widehat{\delta}} \otimes I_k\right) & 0_{q\times r} \\ 0_{r\times qk} & I_r \end{bmatrix} \tag{4.3.14}$$

$$\widehat{B}_s = \begin{bmatrix} \bar{\mathbb{S}}_1 \left(\mathcal{A}_T W_T \mathcal{A}_T\right)^{-1} \mathcal{A}_T W_T \left(\mathbb{S}_{\widehat{\delta}} \otimes I_k\right) & 0_{q\times r} & 0_{q\times rpq} \\ 0_{q\times qk} & I_r & 0_{r\times rpq} \\ 0_{r^2\times qk} & 0_{r^2\times r} & \widehat{R}_s \left(\widehat{G} \otimes \left(\frac{\widehat{\mathcal{F}}^\top \widehat{\mathcal{F}}}{T-p}\right)^{-1}\right) \end{bmatrix}. \tag{4.3.15}$$

The constant matrices $\bar{Q}_{j,i}$ for $j = 1, 2$ can be estimated by replacing the unknown parameters with their consistent counterparts. Based on Theorems 4.1 and 4.2 and Propositions 4.1 to 4.3, we know that $\widehat{\theta}_i, \widehat{\lambda}_i, \widehat{\Psi}_s, \widehat{G}$, and $\widehat{a}_1$ consistently estimate $\bar{H}_\eta^{-\top}\theta_i, \bar{H}_F^{-1}\lambda_i, \bar{H}_F^\top \Psi_s \bar{H}_F^{-\top}, \bar{H}_F^\top G \Sigma_\eta \bar{H}_\eta$, and $\bar{H}_\eta^\top a_1$, respectively. Hence, we propose the following estimators for the constant matrices

$$\widehat{Q}_{1,i} = \left[\widehat{\theta}_i^\top C_1 + \widehat{a}_1 \widehat{G}^\top C_2\right],$$
$$\widehat{Q}_{2,i} = \left[\widehat{\lambda}_i^\top \widehat{\Psi}_s \widehat{G} C_3 + \widehat{a}_1^\top \widehat{G}^\top \widehat{\Psi}_s^\top C_4 + \left(\widehat{\lambda}_i^\top \otimes \widehat{a}_1^\top \widehat{G}^\top\right) C_5\right].$$

4.3.5 Overidentification and Automatic Selection of External Instruments

A major advantage of the GMM-based framework that we adopt is the possibility of 1) testing the joint validity of external instruments and 2) automatic selection of valid external instruments.

We first present the J-test for the joint validity of external instruments as

$$J_T \equiv T \mathcal{Q}_T(\widehat{\delta}), \tag{4.3.16}$$

where $\mathcal{Q}_T(\widehat{\delta})$ is the GMM-criterion function $\mathcal{Q}_T$ evaluated at $\widehat{\delta}$ where the weight W_T is chosen optimally. We show that J_T has a standard asymptotic $\chi^2_{(k-1)(q-1)}$ distribution.

Following this definition of the J-test statistic, we then propose a series of instrument selection criteria for automatic selection of valid instruments. Let c denote the instrument selection vector, which takes values 0 or 1. The number of overidentifying restrictions is therefore $(|c| - 1)(q - 1)$.

Define the GMM-estimator using the instruments selected by c as

$$\widehat{\delta}(c) = \underset{\delta}{\operatorname{argmin}} \mathcal{Q}_T(\delta(c)) = \underset{\delta}{\operatorname{argmin}} \bar{g}_{T,c}(\delta)^\top W_T(c) \bar{g}_{T,c}(\delta), \tag{4.3.17}$$

where $\bar{g}_{T,c}(\delta)$ and $W_T(c)$ are the empirical moments and their weight matrix, defined using only the instruments selected by c. Thus, the corresponding $J_T(c)$ test can also be written as

$$J_T(c) = T \bar{g}_{T,c}\left(\widehat{\delta}(c)\right)^\top W_T(c) \bar{g}_{T,c}\left(\widehat{\delta}(c)\right). \tag{4.3.18}$$

We consider estimation of c^0 the "correct" selection vector, using an estimator $\widehat{c}$, which has parameter space $\mathscr{C} \in \mathcal{C}$. The space $\mathscr{C}$ contains $c = \mathbf{0}$, and is defined in terms of the selection of the *instruments* in order to exploit the block structure implied by the moment conditions; that is, if the first instrument is invalid, then this implies that all of the first $q-1$ moment conditions are also invalid.

Assumption 9. *Define $\mathscr{Z} = \{c \in \mathcal{C} : c = c^0(\delta) \text{ for some } \delta\}$, the set of selection vectors in $\mathcal{C}$ which select only moment conditions that are zero asymptotically, and $\mathscr{M}\mathscr{Z} = \{c \in \mathscr{Z} : |c| \geq |c^*| \forall c^* \in \mathscr{Z}\}$, the set of selection vectors in $\mathscr{Z}$ that maximise the selected moments out of the selection vectors in $\mathscr{Z}$. We require the following conditions:*

a) $\mathscr{M}\mathscr{Z}$ contains a single element c^0.

b) $\bar{g}_{T,c}(\delta)$ has a unique solution δ.

Assumptions 9 (a) and 9 (b) correspond to Assumptions ICc^0 and ICθ^0 of Andrews (1999). Assumption 9 (a) requires that the correct selection vector uniquely selects the maximal number of moment conditions that equal zero asymptotically. Assumption 9 (b) specifies that δ is the "true" value of δ. As we are in a standard GMM context, $\mathscr{M}\mathscr{Z} = \{1_{q-1}\}$ and thus both assumptions hold.

Similarly, we follow Andrews (1999) and propose a set of information criteria which can be used to select for the correct set of moments (instruments). The GMM moment selection criterion chooses the vector in $\widehat{c}_{GMM_{BIC}}, \widehat{c}_{GMM_{AIC}}, \widehat{c}_{GMM_{HQIC}}$ in $\mathcal{C}$ which, respectively, minimise:

$$GMM_{BIC} = J(c) - (|c| - 1)(q - 1)logT;$$

$$GMM_{AIC} = J(c) - 2(|c| - 1)(q - 1);$$

$$GMM_{HQIC} = J(c) - Q(|c| - 1)(q - 1)loglogT,$$

for some $Q > 2$ (which we set to 2.01), and where $(|c| - 1)(q - 1)$ is the number of identifying restrictions.[2] These criteria are the counterparts of the Bayesian Information Criteria, Akaike Information Criteria, and Hannan-Quinn Information Criteria.

Downwards Testing

As an alternative, we next propose a downwards testing procedure that sequentially tests all combinations of the external instruments and asymptotically yields the correct selection of instruments.[3] As stated by Andrews (1999), this downwards testing procedure formalises the ad-hoc approaches used by empirical researchers.

We describe the downwards testing procedure, which is based on the test statistic $J_T(c)$. Starting with vectors $c \in \mathscr{C}$ for which $|c|$ is the largest, we carry out tests with progressively smaller $|c|$ until we find a test that does not reject the null hypothesis; let $\widehat{k}_{DT}$ denote the value of $|c|$ for the first such test that does not reject. Given $\widehat{k}_{DT}$, the downwards testing estimator of $\widehat{c}_{DT}$ is defined to be the vector that minimises $J_T(c)$ over $c \in \mathscr{C}$ with $|c| = \widehat{k}_{DT}$. The downwards testing moment selection procedure thus progresses from the most to least restrictive model.

The consistency of the J-test, moment selection criteria, and downwards testing procedure are summarised in the following theorem.

Theorem 4.3. *Under Assumptions 1 to 8 and the condition that* $\frac{\sqrt{T}}{N} \to 0$ *as* $N, T \to \infty$,

a) $J_T \equiv T\mathcal{Q}_T(\widehat{\delta}) \xrightarrow{d} \chi^2_{(k-1)(q-1)}$,

b) Additionally under Assumption 9, for $MSC \in \{GMM_{BIC}, GMM_{AIC}, GMM_{HQIC}\}$, $\widehat{c}_{MSC} = c^0$ *w.p.a.* 1,

[2]Note that $GMM - AIC$ is inconsistent.

[3]Note that the upwards testing procedure requires some extra assumptions, so we omit it for brevity.

c) Additionally under Assumption 9, $\widehat{c}_{DT} = c^0$ *w.p.a.* 1.

Theorem 4.3 (a) follows from a standard application of Hansen's J-test. The degrees of freedom corresponds to the fact that each of the k instruments corresponds to $(q-1)$ moments, and only $k = 1$ instrument is required to just identify a_1. Theorem 4.3 (b) establishes that the model selection criteria $\widehat{c}_{GMM_{BIC}}, \widehat{c}_{GMM_{AIC}}$, and $\widehat{c}_{GMM_{HQIC}}$ are consistent at determining when there are no over-identifying restrictions. Theorem 4.3 (c) corresponds to Theorem 2 of Andrews (1999), and establishes that the downwards testing estimator $\widehat{c}_{DT}$ is able to determine when there are no over-identifying restrictions, similar to $\widehat{c}_{MSC}$. In practice, over-rejection of the J test in finite samples tends to lead to a higher probability of using only correct moments, but not necessarily all valid moments.

Remark 4.5. *An upwards testing procedure can also be considered following Andrews (1999). However, this requires an additional assumption on the parameter space* $\mathscr{C}$ *to ensure that it does not stop at too small a value of* $|c|$*. In addition, although both the upwards and downwards testing procedure are consistent, in finite sample the upwards testing procedure will always select fewer moments than the downwards testing procedure if they do not agree. Thus, we focus on the model selection criteria and downwards testing approaches.*

4.4 Monte Carlo

4.4.1 Data Specification

The factor loadings λ_i are drawn from a multivariate normal distribution with mean $\mathbf{0}_r$ and covariance matrix $\Sigma_\Lambda = I_r$.[4] The structural shocks ζ_t are drawn from $N(\mathbf{0}_q, I_q)$.

Similar to Bai and Wang (2015), we specify a VAR process for the dynamic factors as

$$f_t = \phi f_{t-1} + A\zeta_t, \tag{4.4.1}$$

[4]The set of loadings λ_i is set to be a vector of ones, in order to ensure that the first impulse response function is not too small, which can cause some numerical issues when implementing the unit effect normalisation.

where $\phi = 0.7$ and $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$. The static factors are stacked as $F_t = \left[f_t^\top, f_{1,t-1}^\top, \ldots, f_{r-q,t-1}^\top\right]^\top$ where $r \leq 2q$ in order to include some lags of f_t as static factors. Equation (4.4.1) implies

$$F_t = \Phi_1 F_{t-1} + GA\zeta_t, \tag{4.4.2}$$

where $\Phi = \begin{bmatrix} \phi I_r & 0_{q\times(r-q)} \\ I_{(r-q)\times q} & 0_{(r-q)\times(r-q)} \end{bmatrix}$ where $I_{(r-q)\times q}$ denotes the first $(r-q)$ rows of I_q, and $GA = \begin{bmatrix} A \\ 0_{(r-q)\times q} \end{bmatrix}$. We set $r = 5$ and $q = 3$. The observable series are then generated as

$$X_t = \Lambda F_t + e_t. \tag{4.4.3}$$

To investigate the efficiency gains from overidentification and the size of the proposed J-test, we generate the instrument Z_{jt} that is correlated with the first structural shock at time t by

$$\textbf{DGP 1}: Z_{jt} = \sqrt{1-a^2}w_{jt} + a\zeta_{1t} + \zeta_{q,t-1}, \quad \text{for} \quad j = 1, \ldots, k = 4, \tag{4.4.4}$$

where w_{jt} are i.i.d. standard normal random variables, and a is set equal to $\sqrt{1/2}/2$ so that the correlation between Z_{jt} and ζ_{1t} is equal to 0.25. We set the number of instruments as $k = 1, \ldots, 4$ to investigate the benefits of overidentification.

To investigate the power of the overidentification test and consistency of the moment selection procedures, we generate instruments as

$$\begin{aligned} \textbf{DGP 2}: \quad & Z_{jt} = \sqrt{1-a^2}w_{jt} + a\zeta_{1t} + \zeta_{q,t-1}, \quad \text{for} \quad j = 1, 2, \\ & Z_{3t} = \sqrt{1-a^2}w_{3t} + a\zeta_{2t} + \zeta_{q,t-1}, \\ & Z_{4t} = \sqrt{1-a^2}w_{4t} + a\zeta_{3t} + \zeta_{q,t-1}, \end{aligned} \tag{4.4.5}$$

such that the first two instruments are only correlated with the structural shock of interest and hence valid, but the last two instruments are contaminated with the effects of other structural shocks, and hence invalid. In either specification, the instruments are also correlated with the lags of the qth structural shock. The number of replications is $1,000$.

With the observed data X_t, we estimate the static factors $\widehat{F}$ and loadings $\widehat{\Lambda}$. The MA coefficients $\widehat{\Psi}_s$ are computed using the OLS estimate of $\widehat{\Phi}$. The principal components-based estimates of the reduced form shocks $\widehat{\eta}_t$ and the instruments are then used for the estimation of δ. We implement two-step GMM estimation using the weighting matrix $I_q \otimes \left(\frac{1}{T}\sum_{t=1}^{T} Z_t Z_t^\top\right)^{-1}$ in the first step, then re-estimate δ using the optimal weighting matrix $\widehat{V}^{-1}$ in the second step. The confidence intervals for the structural IRFs are computed based on the asymptotic normal distribution in Theorem 4.2 and the proposed consistent estimators of the covariance matrices.

Additionally, for a point of comparison, we also identify and estimate the impulse response using an SVAR-IV estimator using X_{1t} for normalisation, and three other variables in X_t selected randomly. The implementation of this follows Cheng et al. (2021), the difference from the SFM approach being that the SVAR is estimated in X_{jt} for $j = 1, \ldots, 4$ directly, and the reduced form shocks estimated as the residuals of this system. Note that because $q = 3$, this results in a singular VAR system.

4.4.2 Results

Efficiency Gains

Table 4.1 reports the finite-sample RMSEs of the estimated IRFs as measured by a ratio of each overidentified scheme compared to the just-identified scheme using only one instrument. The RMSE ratios are typically decreasing in k for the IRFs, with the effect being more pronounced at the contemporaneous horizon and when $N = T$. This confirms that a more efficient estimate of a_1 via the use of more than one instrument can lead to efficiency gains for the IRFs at both zero and non-zero horizons, and that our asymptotic theory is correct.

Table 4.2 reports the finite-sample coverage rates of the confidence intervals. In general, these coverage rates are acceptable and close to the nominal level of 95%, particularly at the $h = 3$ horizon. There is some evidence that the coverage ratios are slightly under-estimated for contemporaneous

Table 4.1: RMSE ratios

T	N	$h=0$ $k=2$	$h=0$ $k=3$	$h=0$ $k=4$	$h=0$ SVAR-IV ($k=4$)	$h=3$ $k=2$	$h=3$ $k=3$	$h=3$ $k=4$	$h=3$ SVAR-IV ($k=4$)
250	125	0.923	0.911	0.892	6.150	0.968	0.956	0.953	3.688
	250	0.924	0.919	0.889	6.340	0.963	0.951	0.935	3.706
500	125	0.940	0.914	0.902	8.293	0.973	0.969	0.966	4.874
	250	0.941	0.913	0.901	8.711	0.979	0.970	0.963	4.934

Note:
Entries report the RMSE of the estimated IRFs of the overidentified system, compared to the RMSE of the IRFs of the just-identified system. The SVAR-IV is estimated with $k=4$ variables, the first of which is X_{1t} for normalisation, and the rest of which are randomly selected.

Table 4.2: Coverage Probabilities

T	N	$h=0$ $k=1$	$h=0$ $k=2$	$h=0$ $k=3$	$h=0$ $k=4$	$h=3$ $k=1$	$h=3$ $k=2$	$h=3$ $k=3$	$h=3$ $k=4$
250	125	0.919	0.897	0.889	0.881	0.954	0.949	0.947	0.944
	250	0.920	0.904	0.895	0.888	0.954	0.949	0.945	0.944
500	125	0.903	0.890	0.883	0.880	0.946	0.943	0.941	0.940
	250	0.909	0.898	0.892	0.889	0.948	0.945	0.943	0.942

Note:
Entries report the coverage probabilities of the IRFs using the proposed asymptotic distributions (nominal 95%).

horizons. This is a commonly encountered problem in the factor modelling literature (e.g. Yamamoto and Hara, 2022), and is a finite-sample aberration that can be readily addressed by employing a bootstrap procedure similar to that of Yamamoto (2019). We leave this issue for future research. The effective coverage rates all improve at the sample size increases and particularly as N increases, confirming our asymptotic theory.

Table 4.3 reports the finite-sample performance of the proposed J-test to test the null hypothesis of joint instrument exogeneity. In general, the effect size of the proposed tests is acceptable for the sample sizes considered in simulations.

Overidentification Test and Moment Selection

We investigate the results of the proposed overidentification and moment selection procedures, which correspond to DGP 2. Tables 4.4 and 4.5 present the finite sample performance of the J-test for

Table 4.3: Size of J-test

T	N	$k = 2$	$k = 3$	$k = 4$
250	125	0.062	0.053	0.054
	250	0.059	0.061	0.053
500	125	0.058	0.051	0.053
	250	0.063	0.054	0.063

Note:
Entries report the rejection frequencies of the J-test (nominal size of 5%).

Table 4.4: Power of J-test

T	N	Rejection Frequency
250	125	1.000
	250	1.000
500	125	1.000
	250	1.000

Note:
Entries report the rejection frequency of J-test for overidentification, with $k = 4$ instruments.

joint exogeneity of the instruments and accuracy of the moment selection procedures, respectively, under DGP 2. It can be seen that, across all specifications, the J-test has high power, and the moment selection procedures are able to accurately select the correct instruments.

Table 4.5: Accuracy of Moment Selection Procedures

		Information Criteria			Testing	
T	N	GMM_{BIC}	GMM_{AIC}	GMM_{HQIC}	DT	UT
250	125	1.000	1.000	1.000	1.000	1.000
	250	1.000	1.000	1.000	1.000	1.000
500	125	1.000	1.000	1.000	1.000	1.000
	250	1.000	1.000	1.000	1.000	1.000

Note:
Entries report the frequencies of correct instrument selection. DT and UT denote Downwards and Upwards Testing respectively. Correct instruments are Z_1 and Z_2.

4.5 Empirical Application

4.5.1 Data and Instruments

We consider the dataset used by Stock and Watson (2012a). This dataset consists of quarterly observations from 1959Q1 - 2011Q2 on 200 U.S. macroeconomic time series, grouped into 13 categories and suitably transformed to induce stationarity. Of the 200 series available, we only use the 132 disaggregated series to estimate the factors in order to avoid double counting high level aggregates.[5] We restrict our sample to 1980Q1 - 2007Q2; the start is dictated by the data availability of the instruments, while the end is chosen to avoid the onset of the Global Financial Crisis. We focus on identifying and estimating the dynamic causal effect of a monetary policy shock with the use of various monetary policy instruments. These include the narrative-based instrument of Romer and Romer (2004) computed as the residual of a Fed monetary intentions measure on internal Fed forecasts,[6] a model based instrument in the form of the monetary shocks identified from the SVAR of Bernanke and Mihov (1998), and a collection of monetary surprises identified using high(er) frequency data: the changes in federal funds futures around policy announcements using a daily window (Barakchian and Crowe, 2013), a 30-minute window (Gertler and Karadi, 2015), and a 30-minute window with further cleaning of the surprises via a regression on more control variables, (Miranda-Agrippino and Ricco, 2021). This selection of five instruments corresponds to

[5]See Stock and Watson (2012a) for more details on data cleaning.
[6]We use an updated and extended version of the Romer and Romer (2004) shocks, as constructed by Wieland and Yang (2020).

the instrument set used by Schlaak et al. (2023).[7]

4.5.2 Model Specification

We first estimate the number of factors in the dataset. The $IC_p(2)$ criterion of Bai and Ng (2002) suggests $r = 5$ static factors, though the criteria are quite flat for $4 - 12$ factors. As suggested by Stock and Watson (2016), the sixth to twelfth factors can often help in explaining the majority of variation in many important variables such as labour productivity, hourly compensation, the term spread, and exchange rates. The first nine factors explain about 52.1% of the total variance in the dataset; whereas the contributions of the 10th-12th factors only provide marginal gains totalling 6.79% additional explanatory power. As stated by Han (2015, 2018), it is important to set r high enough such that the space spanned by the static factors can be fully recovered; at the same time, setting r too high (usually more than nine factors) tends to introduce too much extra noise.[8] We therefore proceed with the choice of $r = 9$ static factors for our benchmark model. We fit a VAR(2) process to model the dynamics of F_t, corresponding to $p = 2$ lags in our benchmark analysis, as supported by the BIC. The criterion of Bai and Ng (2007) tends to detect three dynamic factors; we thus set $q = 3$ factors in our benchmark model.

We proceed with the identification of the monetary policy shock by implementing the proposed estimators with all five available instruments.

4.5.3 Results

Dynamic Causal Effects of Monetary Policy Shocks

We present the results of our benchmark model. Figure 4.1 shows the cumulative impulse responses of various macroeconomic variables to a standard deviation monetary contraction in the Federal Funds rate as identified, using the benchmark model with all instruments. Although most impulse responses are not statistically significant from zero, it is remarkable that most of the point estimates

[7] We do not consider combining the instruments by taking their first eigenvector, as our generalised method-of-moments framework already achieves this in a data-driven manner.

[8] The procedures of Onatski (2010) and Ahn and Horenstein (2013) tend to estimate $r = 1$ factors, which as noted Forni and Gambetti (2014) is at odds with most macroeconomic theory and the theoretical premise of the SVAR literature.

are consistent with economic theory and the consensus as documented by Christiano et al. (1998). Economic activity as measured by industrial production declines immediately, with the response bottoming after one year. Falls are also evident in earnings, employment, and money variables. The unemployment rate is estimated to increase after a contractionary monetary policy shock, with a persistent effect.

Note that there is a persistent, though generally statistically insignificant puzzle evident. In

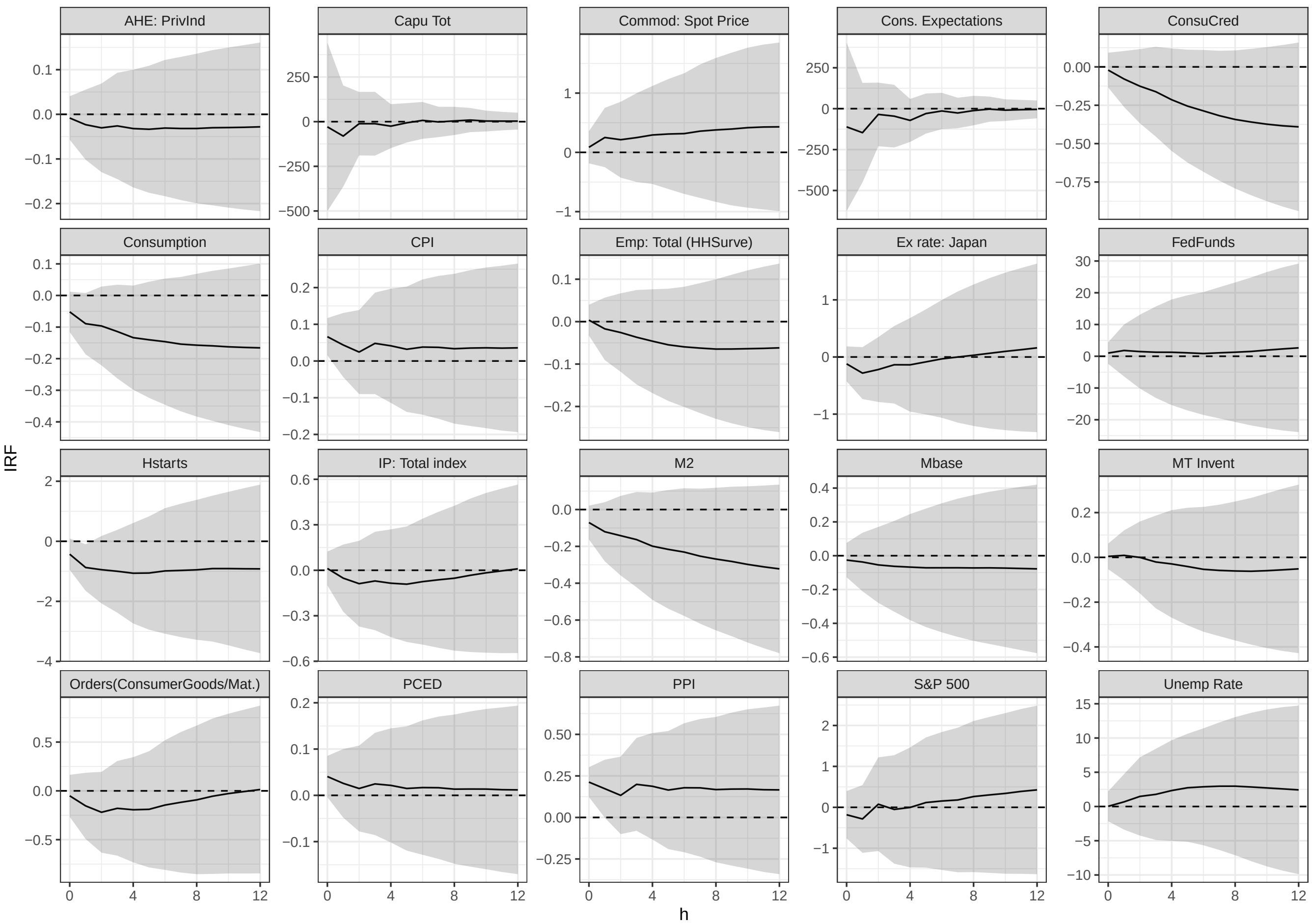

Figure 4.1: Cumulative IRFs after a contractionary monetary policy shock for the over-identified benchmark, normalised to a 100 basis point movement in the Federal Funds rate.

contrast, Figure 4.2 presents the cumulative IRFs after a contractionary monetary policy as (just) identified by using each instrument one at a time, in comparison with the benchmark overidentified model. Note that the impulse responses as identified by the Bernanke and Mihov (1998) model-based instrument are omitted due to scaling issues. It is remarkable that although almost all impulse responses from these just identified schemes lie within the 95% confidence interval of the overidentified case, each of them produces responses that are less efficient and/or more puzzling.

For example, the narrative-based instrument of Romer and Romer (2004) produces responses with the correct signs for series such as earnings, housing starts, industrial production, and employment. However, it is also prone to producing greatly puzzling behaviour in prices, exchange rates, and the S&P500. Next, the model-based instrument of Bernanke and Mihov (1998) identifies a contemporaneous response of the Federal Funds Rate to a monetary policy shock to be near zero. Such a result poses significant numerical problems when imposing the unit-effect normalisation and causes significant scaling issues; this result is additionally highly incompatible with macroeconomic theory. We therefore believe that the impulse responses as identified by this model-based measure to be untenable in any practical sense. Finally, the high frequency-based instruments of Barakchian and Crowe (2013); Gertler and Karadi (2015); Miranda-Agrippino and Ricco (2021) are much more prone to producing puzzling responses. In particular, all high frequency instruments produce puzzling responses in the key variables of earnings, orders, employment, inventories, consumer credit, and crucially industrial production. Although these responses are not statistically significantly different from zero, the point estimates are still broadly incompatible with the macroeconomic consensus as summarised by Christiano et al. (1998). On the other hand, we find that high frequency instruments tend to alleviate, and, in some cases, eliminate price puzzles altogether. Of these, it is noteworthy that the Gertler and Karadi (2015) instrument tends to produce the largest puzzles in real activity and prices, a phenomenon that Miranda-Agrippino and Ricco (2021) attribute to omitted information effects and alleviate via their instrument. These results are not entirely surprising; as noted by Ramey (2016), these popular instrument variables are unstable and can produce puzzling responses to prices and real variables.

Evidently, the responses identified by each instrument have their distinct advantages and disadvantages, which, in practice, make economic reconciliation difficult. Thus, the comparison in Figure 4.2 shows that the proposed overidentification scheme is able to automatically leverage the respective advantages of each instrument, and produce, overall, more reasonable responses.

Which Monetary Policy Instruments are Valid?

Testing the exogeneity and therefore validity of the instruments has so far been largely unresolved

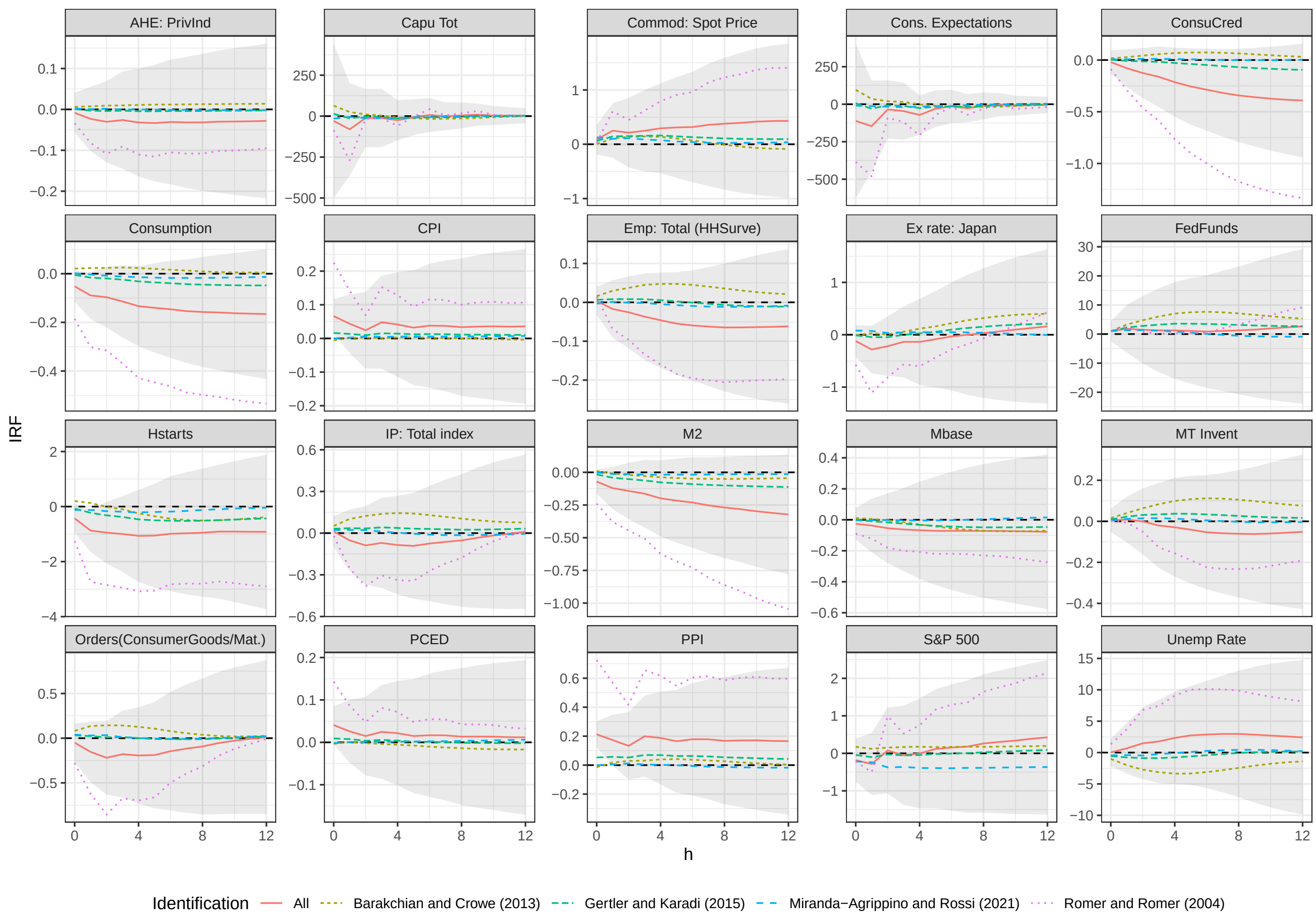

Figure 4.2: Cumulative IRFs after a contractionary monetary policy shock for the over-identified benchmark (solid curve), and just identified setups each using one instrument at a time, normalised to a 100 basis point movement in the Federal Funds rate.

in the literature, but can be addressed with the J-test and automatic moment selection criteria we propose. Table 4.6 presents the results of J-test for joint exogeneity of instruments and automatic moment selection criteria. Across all overidentified specifications, we fail to reject the J-test. Correspondingly, all three model selection criteria are minimised when all five instruments are used; the downwards testing estimator consequently selects all five instruments.

Table 4.6: Results of J-test for Overidentification and GMM_{MSC} Criteria.

GMM_{BIC}	GMM_{HQIC}	J_T	J_{crit}	RR04	GK15	MR21	BM98	BK13
-8.573	-5.502	0.638	5.991	TRUE	TRUE	FALSE	FALSE	FALSE
-8.824	-5.753	0.386	5.991	TRUE	FALSE	TRUE	FALSE	FALSE
-8.836	-5.765	0.375	5.991	FALSE	TRUE	TRUE	FALSE	FALSE
-15.285	-9.143	3.135	9.488	TRUE	TRUE	TRUE	FALSE	FALSE
-9.200	-6.129	0.011	5.991	TRUE	FALSE	FALSE	TRUE	FALSE
-8.760	-5.689	0.451	5.991	FALSE	TRUE	FALSE	TRUE	FALSE
-17.766	-11.623	0.655	9.488	TRUE	TRUE	FALSE	TRUE	FALSE
-8.878	-5.807	0.333	5.991	FALSE	FALSE	TRUE	TRUE	FALSE
-18.035	-11.893	0.386	9.488	TRUE	FALSE	TRUE	TRUE	FALSE
-15.953	-9.810	2.468	9.488	FALSE	TRUE	TRUE	TRUE	FALSE
-23.846	-14.633	3.785	12.592	TRUE	TRUE	TRUE	TRUE	FALSE
-4.998	-1.927	4.212	5.991	TRUE	FALSE	FALSE	FALSE	TRUE
-5.783	-2.712	3.427	5.991	FALSE	TRUE	FALSE	FALSE	TRUE
-14.073	-7.930	4.348	9.488	TRUE	TRUE	FALSE	FALSE	TRUE
-8.450	-5.379	0.761	5.991	FALSE	FALSE	TRUE	FALSE	TRUE
-13.764	-7.622	4.656	9.488	TRUE	FALSE	TRUE	FALSE	TRUE
-14.377	-8.235	4.043	9.488	FALSE	TRUE	TRUE	FALSE	TRUE
-20.511	-11.298	7.120	12.592	TRUE	TRUE	TRUE	FALSE	TRUE
-5.826	-2.754	3.385	5.991	FALSE	FALSE	FALSE	TRUE	TRUE
-14.060	-7.918	4.360	9.488	TRUE	FALSE	FALSE	TRUE	TRUE
-14.155	-8.012	4.266	9.488	FALSE	TRUE	FALSE	TRUE	TRUE
-23.035	-13.822	4.596	12.592	TRUE	TRUE	FALSE	TRUE	TRUE
-14.780	-8.637	3.641	9.488	FALSE	FALSE	TRUE	TRUE	TRUE
-23.044	-13.830	4.587	12.592	TRUE	FALSE	TRUE	TRUE	TRUE
-21.320	-12.107	6.311	12.592	FALSE	TRUE	TRUE	TRUE	TRUE
-28.955	-16.670	7.887	15.507	TRUE	TRUE	TRUE	TRUE	TRUE

Note:
RR04, GK15, MR21, BM98, and BC13 refer to the external instruments of Romer and Romer (2004), Gertler and Karadi (2015), Miranda-Agrippino and Rossi (2021), Bernanke and Mihov (1998), and Barakchian and Crowe (2013)

Is Using More Instruments Better?

Given the evidence that all monetary policy instruments are jointly valid, we next investigate the efficiency gains from using more than one instrument. We do this by estimating and comparing the asymptotic variances of the IRFs. Table 4.7 reports the ratios of the asymptotic standard deviations of the estimated IRFs under the just-identified IRFs obtained by using one instrument compared to the benchmark model overidentified using all instruments; a ratio greater than one means that the overidentified model provides a more efficient estimate. Not all ratios are greater than one, so we additionally compute the means of the ratios to see if overidentification can lead to efficiency gains on average. On average, the benchmark model that uses all instruments tends to produce more efficient estimated for both zero and nonzero horizons compared to all instruments, with the notable exception of Gertler and Karadi (2015). However, although the relative efficiency of the responses of this instrument are on average lower, its behaviour can be quite erratic across different horizons even for the same variable. Therefore, we conclude that the overidentified scheme provides an ideal trade-off between producing the most reasonable impulse responses, and efficiency.

Table 4.7: Ratios of Asymptotic Standard Deviations of Estimated Impulse Response Functions: Just Identified / Over-identified.

		h				
Instrument	Series	0	1	2	3	4
Barakchian and Crowe (2013)	IP: Total index	13.469	15.811	12.019	10.952	8.082
	Commod: Spot Price	2.192	12.585	7.305	2.023	3.937
	CPI	6.216	6.830	2.954	3.994	3.918
	FedFunds	7.750	14.087	13.601	14.745	14.167
	S&P 500	11.052	5.487	2.460	3.784	2.429
	Ex rate: Japan	1.301	9.709	5.268	11.549	13.178
	Consumption	11.975	5.316	2.344	4.717	1.762
	Mean	7.553	8.720	7.585	7.201	8.007
Bernanke and Mihov (1998)	IP: Total index	27.201	26.339	21.741	25.927	21.605
	Commod: Spot Price	32.133	22.162	18.692	21.629	19.736
	CPI	20.598	18.303	21.087	17.794	14.360
	FedFunds	28.214	23.022	22.871	22.970	22.277
	S&P 500	26.318	14.397	38.714	15.720	22.038
	Ex rate: Japan	22.969	18.982	31.212	21.465	22.962
	Consumption	21.169	15.815	15.703	14.187	12.970
	Mean	22.756	20.545	21.151	19.861	19.043
Gertler and Karadi (2015)	IP: Total index	1.048	1.012	0.815	1.355	0.990
	Commod: Spot Price	1.495	1.112	0.708	0.940	0.895
	CPI	0.529	0.653	0.826	0.620	0.530
	FedFunds	0.937	0.982	0.952	1.011	1.141
	S&P 500	1.473	0.606	1.780	0.645	0.844
	Ex rate: Japan	1.027	0.693	1.312	0.759	0.840
	Consumption	0.901	0.578	0.703	0.574	0.555
	Mean	0.925	0.835	0.837	0.827	0.811
Miranda-Agrippino and Rossi (2021)	IP: Total index	5.555	2.558	2.057	12.454	8.651
	Commod: Spot Price	15.735	8.226	3.382	9.605	6.929
	CPI	3.589	3.379	7.183	5.169	2.358
	FedFunds	8.710	2.295	2.349	4.850	6.428
	S&P 500	13.762	1.413	19.213	4.447	8.660
	Ex rate: Japan	10.731	1.790	12.743	2.236	2.181
	Consumption	3.248	2.834	5.699	2.977	1.641
	Mean	6.720	4.188	5.039	5.385	4.288
Romer and Romer (2004)	IP: Total index	3.501	3.458	2.974	3.350	2.889
	Commod: Spot Price	3.748	2.708	2.813	2.733	2.617
	CPI	2.374	2.507	2.605	2.542	2.227
	FedFunds	3.143	3.111	3.161	3.080	3.143
	S&P 500	3.325	2.234	4.387	2.499	2.535
	Ex rate: Japan	3.117	2.443	3.512	2.621	3.075
	Consumption	3.105	2.249	2.553	2.277	2.063
	Mean	2.956	2.765	2.821	2.733	2.638

4.6 Conclusion

This paper develops new estimators for the impulse response functions in structural factor models under overidentifying restrictions by using multiple external instruments. Compared with a typical SVAR-IV approach, our framework is able to simultaneously address the challenging issues of nonfundamentalness, covariance singularity, and validity testing of identification restrictions, any of which can prevent the practitioner from recovering the true impulse responses. We establish the asymptotic distributions of the new estimators, and develop test statistics for the joint validity of instruments, and a downwards testing procedure which automatically selects the correct instruments. Our simulation study confirms that the estimated impulse response functions are more accurate than structural factor models that only use one instrument at a time, and the pre-existing SVAR-IV approach, as well as the finite sample properties of the proposed validity tests and moment selection criteria. We apply the framework to identify and estimate the impacts of a contractionary monetary policy shock on a large quarterly U.S. macroeconomic dataset using five commonly used instruments, including narrative-based measures, model-based measures, and monetary surprises identified with high(er) frequency data. We find that, although all of these instruments are jointly valid, using these instruments one by one as is commonly done in the literature can nevertheless produce puzzling and highly inaccurate responses. Instead, our proposed framework that jointly uses all instruments is able to produce responses which are overall more reasonable, and more accurate.

Chapter 5

Conclusion

This Thesis contributes to the dynamic factor model literature, specifically in the areas of structural breaks, forecasting, and structural analysis. All of these are highly active areas of research.

Although the focus of this Thesis can been on the *static* formulation of the dynamic factor model and the principal components estimator, more generally the insights from this work can apply to other factor models. Chapter 2 establishes a new projection-based decomposition theorem that decomposes any structural break into a rotational and shift component, each respectively associated with a change in the factor variance and the factor loadings. The decomposition is designed to work with any form of structural change, and we show that it leads to two easy-to-implement Wald tests to separately test for evidence of breaks in the factor variance and breaks in the factor loadings. Their finite sample size and consistency are confirmed in a Monte Carlo study. When applied to U.S. macroeconomic data, we find evidence of both breaks occurring, but are able to associate the Great Moderation with a >60% in reduction the total factor variance, after controlling for changes in the factor loadings. Our findings are, overall, more in line with broader macroeconomic intuition, highlighting the need to disentangle structural breaks, and nuance the discussion around these historical episodes.

Chapter 3 studies the effects of structural breaks in the factor structure on factor-augmented forecasting. We propose a new "rotated" factor estimator that is designed to be more robust to shift-type breaks and effectively utilise pre-break data. We explore the asymptotic properties of this estimator and two competing approaches under different types of breaks in a local asymptotic framework. This allows us to provide detailed out of sample MSFE rankings of competing factor-based forecasts. We then propose a post-break cross-validation criterion to automatically select and weight these competing forecasts. The strengths of our approach are confirmed in a simulation study. In an empirical example with U.S. macroeconomic data, we show that our proposed estimators outperform the prevailing benchmark. In particular, we highlight that pre-existing approaches that simply allow for a break in the forecasting equation perform poorly, and one should aim to design more robust estimates of the factors.

Chapter 4 proposes a new estimator for the impulse response functions that leverage both information from a factor structure, and multiple external instruments to achieve (over)identification, allowing us to holistically address the issues of nonfundamentalness, singularity, and identification validity issues that plague SVARs. Specifically, the use of a factor structure provides tools, which naturally avoids the nonfundamentalness and singularity issues that plague traditional SVARs, while our generalised-method-of-moments approach naturally allows for the joint use of multiple instruments, overidentifcation tests for their joint validity, and an automatic moment selection procedure to choose the correct instruments. Simulation confirms the asymptotic theory that the estimators can produce more accurate estimates of the impulse responses, as well as the size and consistency of the procedures. In an empirical study on U.S. macroeconomic data, we find evidence that many proposed monetary policy instruments are jointly valid and can lead to more efficient and reasonable estimates of monetary policy responses.

Several further extensions can follow. Naturally, an extension of projection decomposition in Chapters 2 and 3 to the aforementioned alternative formulations of factor models is possible. The Common Correlated Effects estimator of Pesaran (2006), the two-way matrix factor model of Yu et al. (2022), the latent factor model of Lam et al. (2011), and the factor analysis estimator, which has been recently shown by Fortin et al. (2024) to be well suited for short T applications. The projection decomposition used in Chapter 2 and 3 can also be applied to *any* structural break in the factor structure, and this naturally allows for extension to threshold-type regime switching models following Massacci (2017, 2021), or time varying factor models as in Su and Wang (2017, 2020). The asymptotic framework that was adopted in Chapter 4 allows for many possible extensions. These can include the accommodation of weak instruments such as the use Anderson-Rubin confidence sets as in Montiel Olea et al. (2021). The adoption of a generalised method-of-moments framework specifically allows for combining instrument-based identification to other forms of identification, as long as they can be expressed in terms of moment conditions. For example, identification via heteroskedasticity (Schlaak et al., 2023), narrative restrictions (Antolín-Díaz and Rubio-Ramírez, 2018), and sign restrictions (Granziera et al., 2018) can be re-written in terms of moment conditions. As noted by Han (2024) however, these can typically require the development of new estimation methods in order to ensure that the relevant moment conditions are imposed during estimation. We leave these interesting questions for future research.

Appendix A

Appendices for Chapter 2

A.1 Asymptotic Proofs

A.1.1 Preliminary

First, recall that $V_{NT,1}, V_{NT,2}$ are the $r \times r$ diagonal matrices of the first r largest eigenvalues of the matrices $\frac{1}{T_1 N} X_1 X_1^\top$ and $\frac{1}{T_2 N} X_2 X_2^\top$, respectively. The estimated factor matrices $\tilde{F}_1, \tilde{F}_2$ are $\sqrt{T_1}, \sqrt{T_2}$ times the eigenvectors corresponding to the r largest eigenvalues of $X_1 X_1^\top$ and $X_2 X_2^\top$, respectively, we denote $T_1 = \lfloor \pi T \rfloor$ and $T_2 = T - \lfloor \pi T \rfloor$ for brevity. We therefore have for $m = 1, 2$

$$\frac{1}{NT_m} X_m X_m^\top \tilde{F}_m = \tilde{F}_m V_{NT,m} \tag{A.1.1}$$

Let $\delta_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$. Using $X_m = F_m \Lambda_m^\top + e_{(m)}$ gives:

$$\frac{1}{NT_m}\left(F_m \Lambda_m^\top \Lambda_m F_m^\top + F_m \Lambda_m^\top e_{(m)}^\top + e_{(m)} \Lambda_m F_m^\top + e_{(m)} e_{(m)}^\top\right) \tilde{F}_m V_{NT,m}^{-1} = \tilde{F}_m. \tag{A.1.2}$$

Using the fact that $H_m = (\Lambda_m^\top \Lambda_m/N)(F_m^\top \tilde{F}_m/T_m)V_{NT,m}^{-1}$ yields:

$$\tilde{F}_m - F_m H_m = \frac{1}{NT_m}\left(F_m\Lambda_m^\top e_{(m)}^\top \tilde{F}_m + e_{(m)}\Lambda_m F_m^\top \tilde{F}_m + e_{(m)}e_{(m)}^\top\right)V_{NT,m}^{-1}, \tag{A.1.3}$$

$$\begin{aligned}\tilde{f}_{m,t} - H_m^\top f_t = V_{NT,m}^{-1}\Bigg(&\frac{1}{T_m}\sum_{s=1}^{T}\tilde{f}_{m,s}\gamma_N(s,t)\iota_{mt} + \frac{1}{T_m}\sum_{s=1}^{T}\tilde{f}_{m,s}\zeta_{st}\iota_{mt} \\ &+ \frac{1}{T_m}\sum_{s=1}^{T}\tilde{f}_{m,s}\eta_{m,st}\iota_{mt} + \frac{1}{T_m}\sum_{s=1}^{T}\tilde{f}_s\xi_{m,st}\iota_{mt}\Bigg)\end{aligned} \tag{A.1.4}$$

where $\zeta_{st} = \frac{e_s^\top e_t}{N} - \gamma_N(s,t), \eta_{m,st} = f_s^\top \Lambda_m^\top e_t/N$ and $\xi_{m,st} = f_t^\top \Lambda_m^\top e_s/N$.

We first present some lemmas from Bai (2003), stated for convenience.

Lemma A.1. *For $m = 1, 2$:*

a) Under Assumptions 1 to 4 and 8, $\frac{1}{T_m}\sum_{t=1}^{T}\left\|\left(\tilde{f}_{m,t} - H_m^\top f_t\right)\iota_{mt}\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

b) Under Assumptions 1 to 4 and 8, $\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_{m,i} - H_m^{-\top}\lambda_{m,i}\right\| = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

c) Under Assumptions 1 to 6 and 8, $\frac{1}{T_m}(\tilde{F}_m - F_m H_m)^\top e_{m,i} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

d) Under Assumptions 1 to 6 and 8, $\frac{1}{T_m}(\tilde{F}_m - F_m H_m)^\top F_m = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

e) Under Assumptions 1 to 4 and 8, $\|H_m\| = O_p(1)$.

f) Under Assumptions 1 to 4, 7, and 8, $\operatorname{plim}\left(\frac{\tilde{F}_m^\top F_m}{T_m}\right) = H_{0,m}^{-1}$.

Lemmas A.1 (a) and A.1 (c) to A.1 (f) are the subsample counterparts of Lemmas A.1, B.1, B.2, B.3, A.3 and Proposition 1 of Bai (2003), respectively. Lemma A.1 (b) follows from Lemma A.1 (a) via symmetry.

A.1.2 Consistency Proofs

Proof of Theorem 2.1 (a). By the definition of $\tilde{Z}$, we have:

$$\begin{aligned}\tilde{Z} &= \left(\tilde{\Lambda}_1^\top \tilde{\Lambda}_1\right)^{-1}\tilde{\Lambda}_1^\top \tilde{\Lambda}_2 \\ &= \frac{1}{N}V_{NT,1}^{-1}\frac{1}{T_1}(\tilde{F}_1^\top X_1)^\top \frac{1}{T_2}(\tilde{F}_2^\top X_2),\end{aligned}$$

because for $m = 1, 2$, $\tilde{\Lambda}_m^\top \tilde{\Lambda}_m / N = V_{NT,m}$ by eigen-identity, and $\Lambda_m^\top = (\tilde{F}_m^\top \tilde{F})^\top \tilde{F}^\top X_m = \frac{1}{T_m} \tilde{F}_m^\top X_m$ via a least squares fit. Therefore

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

We shall see that $Z.III$ is the dominating term, and $Z.I, Z.II, Z.IV, Z.V, Z.VI$ are all asymptotically negligible.

Lemma A.2. *Under Assumptions 1 to 6 and 8*

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

b) $Z.II = O_p\left(\frac{1}{N}\right)$,

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

d) $Z.V = O_p\left(\frac{1}{\delta_{NT}^2}\right)$, *and*

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

Proof of Lemma A.2 (a).

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

because of Lemmas A.1 (d) to A.1 (f) and Assumptions 1, 3, and 6 respectively. ■

Proof of Lemma A.2 (b).

$$Z.II = \frac{\tilde{F}_1^\top F_1}{T_1}\frac{\Lambda_1^\top W}{N}\frac{F_2^\top \tilde{F}_2}{T_2} = O_p\left(\frac{1}{N}\right),$$

because $\frac{\Lambda_1^\top W}{N} = O_p\left(\frac{1}{N}\right)$. Optionally, if one is willing to assume strict orthogonality in finite sample, then $\Lambda_1^\top W = 0$, and it follows that $Z.II = 0$. ■

Proof of Lemma A.2 (c).

$$\begin{aligned} Z.IV &= \frac{\tilde{F}_1^\top e_{(1)} e_{(2)}^\top \tilde{F}_2}{NT_1T_2} \\ &= \frac{(\tilde{F}_1 - F_1H_1)^\top}{T_1}\frac{e_{(1)}e_{(2)}^\top}{N}\frac{(\tilde{F}_2 - F_2H_2)}{T_2} + \frac{(F_1H_1)^\top}{T_1}\frac{e_{(1)}e_{(2)}^\top}{N}\frac{(\tilde{F}_2 - F_2H_2)}{T_2} + \\ &\quad \frac{(\tilde{F}_1 - F_1H_1)^\top}{T_1}\frac{e_{(1)}e_{(2)}^\top}{N}\frac{(F_2H_2)}{T_2} + \frac{(F_1H_1)^\top}{T_1}\frac{e_{(1)}e_{(2)}^\top}{N}\frac{(F_2H_2)}{T_2} \\ &= Z.IV.a + Z.IV.b + Z.IV.c + Z.IV.d. \end{aligned}$$

Analysing each of the four terms above, we have:

1.

$$\begin{aligned} \|Z.IV.a\| &\leq \left\|\frac{(\tilde{F}_1 - F_1H_1)}{\sqrt{T_1}}\right\|\left\|\frac{e_{(1)}e_{(2)}^\top}{\sqrt{T_1}\sqrt{T_2}N}\right\|\left\|\frac{(\tilde{F}_2 - F_2H_2)}{\sqrt{T_2}}\right\| \\ &= O_p\left(\frac{1}{\delta_{NT}}\right)O_p\left(\frac{1}{\delta_{NT}}\right)O_p\left(\frac{1}{\delta_{NT}}\right) = O_p\left(\frac{1}{\delta_{NT}^3}\right), \end{aligned}$$

by Lemma A.1 (a) and Assumption 3,

2.

$$\begin{aligned} \|Z.IV.b\| &\leq \left\|\frac{(F_1H_1)}{\sqrt{T_1}}\right\|\left\|\frac{e_{(1)}e_{(2)}^\top}{\sqrt{T_1}\sqrt{T_2}N}\right\|\left\|\frac{(\tilde{F}_2 - F_2H_2)}{\sqrt{T_2}}\right\| \\ &= O_p(1)\,O_p\left(\frac{1}{\delta_{NT}}\right)O_p\left(\frac{1}{\delta_{NT}}\right) = O_p\left(\frac{1}{\delta_{NT}^2}\right), \end{aligned}$$

by Assumptions 1 and 3 and lemmas A.1 (a) and A.1 (e), 3.

$$
\begin{aligned}
\|Z.IV.c\| &\leq \left\|\frac{(\tilde{F}_1 - F_1 H_1)}{\sqrt{T_1}}\right\| \left\|\frac{e_{(1)} e_{(2)}^\top}{\sqrt{T_1}\sqrt{T_2} N}\right\| \left\|\frac{(F_2 H_2)}{\sqrt{T_2}}\right\| \\
&= O_p\left(\frac{1}{\delta_{NT}}\right) O_p\left(\frac{1}{\delta_{NT}}\right) O_p(1) = O_p\left(\frac{1}{\delta_{NT}^2}\right),
\end{aligned}
$$

by Lemmas A.1 (a) and A.1 (e) and Assumptions 1 and 3,

4.

$$
\begin{aligned}
\|Z.IV.d\| &\leq \|H_1\| \left\|\frac{F_1^\top e_{(2)}^\top}{T_1\sqrt{N}}\right\| \left\|\frac{e^\top F_2}{T_2\sqrt{N}}\right\| \|H_2\| \\
&= O_p(1)\, O_p\left(\frac{1}{\sqrt{T}}\right) O_p\left(\frac{1}{\sqrt{T}}\right) O_p(1) = O_p\left(\frac{1}{\delta_{NT}^2}\right),
\end{aligned}
$$

by Lemmas A.1 (e) to A.1 (f) and Assumption 6. Therefore, $Z.IV = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

■

Proof of Lemma A.2 (d).

$$
\begin{aligned}
Z.V &= \frac{\tilde{F}_1^\top}{T_1} \frac{e_{(1)} W}{N} \frac{F_2^\top \tilde{F}_2}{T_2} \\
&\leq \left\|\frac{(\tilde{F}_1 - F_1 H_1)}{\sqrt{T_1}}\right\| \left\|\frac{e_{(1)} W}{N\sqrt{T_1}}\right\| \left\|\frac{F_2^\top \tilde{F}_2}{T_2}\right\| + \|H\| \left\|\frac{F_1^\top e_{(1)} W}{T_1 N}\right\| \left\|\frac{F_2^\top \tilde{F}_2}{T_2}\right\| \\
&= O_p\left(\frac{1}{\delta_{NT}}\right) O_p\left(\frac{1}{\sqrt{N}}\right) O_p(1) + O_p(1)\, O_p\left(\frac{1}{\delta_{NT}^2}\right) O_p(1) \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right),
\end{aligned}
$$

because of Lemmas A.1 (a), A.1 (e), and A.1 (f) and Assumption 3. Note that $\left\|\frac{e_{(1)} W}{N\sqrt{T_1}}\right\| = O_p\left(\frac{1}{\sqrt{N}}\right)$ is implied by Assumption 2, because $\|W\| = \|\Lambda_2 - \Lambda_1 Z\| \leq \|\Lambda_2\|$. ■

Proof of Lemma A.2 (e).

$$\begin{aligned}
Z.VI &= \frac{\tilde{F}_1^\top}{T_1}\frac{e_{(1)}\Lambda_1 Z}{N}\frac{F_2^\top \tilde{F}_2}{T_2} \\
&\leq \left\| \frac{(\tilde{F}_1 - F_1 H_1)}{\sqrt{T_1}} \right\| \left\| \frac{e_{(1)}\Lambda_1}{N\sqrt{T_1}} \right\| \|Z\| \left\| \frac{F_2^\top \tilde{F}_2}{T_2} \right\| + \|H\| \left\| \frac{F_1^\top e_{(1)}\Lambda_1}{T_1 N} \right\| \|Z\| \left\| \frac{F_2^\top \tilde{F}_2}{T_2} \right\| \\
&= O_p\left(\frac{1}{\delta_{NT}}\right) O_p\left(\frac{1}{\sqrt{N}}\right) O_p(1)\, O_p(1) + O_p(1)\, O_p\left(\frac{1}{\delta_{NT}^2}\right) O_p(1)\, O_p(1) \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right),
\end{aligned}$$

because of Lemmas A.1 (a), A.1 (e), and A.1 (f) and Assumption 3, and because $\|Z\| < \infty$ is implied by Assumption 2. ■

Therefore, combining the terms above together, we have:

$$\begin{aligned}
\tilde{Z} &= V_{NT}^{-1}\left(\frac{\tilde{F}_1^\top F_1}{T_1}\right)\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right) Z \left(\frac{F_2^\top \tilde{F}_2}{T_2}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) \\
&= H_1^\top Z H_2^{-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right),
\end{aligned}$$

where the last line follows from the definition of H_1, and the fact that

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

■

The consistency results for $\tilde{Z}$ allows us to extend Lemma A.1 to the case of the rotated factors in the second subsample.

Lemma A.3.

a) *Under Assumptions 1 to 4 and 8,* $\frac{1}{T_2}\left\|\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1)\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

b) *Under Assumptions 1 to 6 and 8,* $\frac{1}{T}(\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1)^\top e_{(2),i} = O_p\left(\frac{1}{\delta_{NT}^2\sqrt{T}}\right)$,

c) *Under Assumptions 1 to 6 and 8,* $\frac{1}{T}(\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1)^\top F_2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

Proof of Lemma A.3. For part a), we have

$$\begin{aligned}
\frac{1}{\sqrt{T_2}}\left\|\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1\right\| &= \frac{1}{\sqrt{T_2}}\left\|\tilde{F}_2\left(\tilde{Z} - H_1^\top Z H_2^{-\top}\right)^\top + \left(\tilde{F}_2 - F_2 H_2\right)\left(H_1^\top Z H_2^{-\top}\right)^\top - F_2 Z^\top H_1\right\| \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right),
\end{aligned}$$

where the last line follows by Theorem 2.1 (a) and Lemma A.1 (a). Squaring both sides proves the result.

For part b), we have

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

For part c), we have

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

■

Before we prove Theorem 2.1 (b), we need the following lemmas.

Lemma A.4. *Under Assumptions 1 to 5 and 8,*

$$\frac{1}{\sqrt{N}}\left\|\tilde{\Lambda}_1\tilde{Z}-\Lambda_1 Z H_2^{-\top}\right\| = O_p\left(\frac{1}{\delta_{NT}}\right)$$

Proof of Lemma A.4.

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\left\|\tilde{\Lambda}_1\tilde{Z}-\Lambda_1 Z H_2^{-\top}\right\| \\
=&\frac{1}{\sqrt{N}}\left\|(\Lambda_1-\Lambda_1 H_1^{-\top})\tilde{Z}+\Lambda_1 H_1^{-\top}\tilde{Z}-\Lambda_1 Z H_2^{-\top}\right\| \\
=&\frac{1}{\sqrt{N}}\left\|(\tilde{\Lambda}_1-\Lambda_1 H_1^{-\top})\tilde{Z}+\Lambda_1 H_1^{-\top}\left(\tilde{Z}-H_1^{\top} Z H_2^{-\top}\right)+\ \Lambda_1 H_1^{-\top}H_1^{\top} Z H_2^{-\top}-\Lambda_1 Z H_2^{-\top}\right\| \\
\leq&\frac{1}{\sqrt{N}}\left\|(\tilde{\Lambda}_1-\Lambda_1 H_1^{-\top})\right\|\left\|\tilde{Z}\right\|+\frac{1}{\sqrt{N}}\left\|\Lambda_1 H_1^{-\top}\right\|\left\|\tilde{Z}-H_1^{\top} Z H_2^{-\top}\right\| \\
=&O_p\left(\frac{1}{\delta_{NT}}\right)O_p(1)+O_p(1)\left[O_p\left(\frac{1}{\delta_{NT}^2}\right)\right] \\
=&O_p\left(\frac{1}{\delta_{NT}}\right),
\end{aligned}$$

where the second last line follows because the first term is $O_p\left(\frac{1}{\delta_{NT}}\right)$ by Lemma A.1 (a) via symmetry, and the second term is $O_p\left(\frac{1}{\delta_{NT}^2}\right)$ by Theorem 2.1 (a). ■

Proof of Theorem 2.1 (b). Next, we prove the consistency of $\tilde{W}$. First, recall that

$$\tilde{W}=\tilde{\Lambda}_2-\tilde{\Lambda}_1\tilde{Z}, \tag{A.1.5}$$

$$W=\Lambda_2-\Lambda_1 Z, \tag{A.1.6}$$

which we can rearrange to form

$$\tilde{W}-W H_2^{-\top}=\tilde{\Lambda}_2-\Lambda_2 H_2^{-\top}-(\tilde{\Lambda}_1\tilde{Z}-\Lambda_1 Z H_2^{-\top}) \tag{A.1.7}$$

Taking the norm of both sides, and dividing by $\sqrt{N}$, we have

$$
\begin{aligned}
\frac{1}{\sqrt{N}}\left\|\tilde{W} - WH_2^{-\top}\right\| &= \frac{1}{\sqrt{N}}\left\|\tilde{\Lambda}_2 - \Lambda_2 H_2^{-\top} - (\tilde{\Lambda}_1\tilde{Z} - \Lambda_1 Z H_2^{-\top})\right\| \\
&\leq \frac{1}{\sqrt{N}}\left(\left\|\tilde{\Lambda}_2 - \Lambda_2 H_2^{-\top}\right\| + \left\|(\tilde{\Lambda}_1\tilde{Z} - \Lambda_1 Z H_2^{-\top})\right\|\right) \\
&\leq \frac{1}{\sqrt{N}}\left\|\tilde{\Lambda}_2 - \Lambda_2 H_2^{-\top}\right\| + \frac{1}{\sqrt{N}}\left\|\tilde{\Lambda}_1\tilde{Z} - \Lambda_1 Z H_2^{-\top}\right\| \\
&= O_p\left(\frac{1}{\delta_{NT}}\right),
\end{aligned}
$$

where the last line follows by Lemma A.4. Squaring both sides results in the usual $O_p\left(\frac{1}{\delta_{NT}^2}\right)$ rate, mirroring the result of Bai (2003). ■

Remark A.3. *We detail how there exists an alternative observationally equivalent parameterisation of the rotation matrix H_2, and how this ultimately does not matter.*

Recall that we originally define $H_2 = \left(\frac{Z^\top \Lambda_1^\top \Lambda_1 Z}{N} + \frac{W^\top W}{N}\right)\left(\frac{F_2^\top \tilde{F}_2}{T_2}\right) V_{NT,2}^{-1}$, and this parameterises all of the change in terms of the loadings, and is what the literature at large does (see Baltagi et al., 2017; Han and Inoue, 2015). It is also possible to parameterise the rotational change explicitly as part of the factors by defining:

$$
\begin{aligned}
H_2^\dagger &= \left(\frac{\Lambda_1^\top \Lambda_1}{N} + \frac{Z^{-\top} W^\top W Z^{-1}}{N}\right)\left(\frac{Z F_2^\top \tilde{F}_2}{T_2}\right) V_{NT,2}^{-1} \\
&= \left(\frac{\Lambda_1^\top \Lambda_1}{N} + \frac{Z^{-\top} W^\top W Z^{-1}}{N}\right)\left(\frac{G_2^\top \tilde{F}_2}{T_2}\right) V_{NT,2}^{-1} \qquad \text{(A.1.8)}
\end{aligned}
$$

where $G_2 = F_2 Z^\top$, the pseudo-factor representation.

With $H_2^\dagger$, the consistency result $\tilde{Z}$ in Theorem 2.1 (a) changes to:

$$
\left\|\tilde{Z} - H_1^\top H_2^{\dagger -\top}\right\| = O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{1}{\sqrt{N}}\right) \qquad \text{(A.1.9)}
$$

where the Z is absorbed into $H_2^\dagger$. Note that this does not affect following results, because we can simply replace $\tilde{F}_2 - F_2 H_2$ with $\tilde{F}_2 - F_2 Z^\top H_2^\dagger$ in all the proofs. Doing so, we get the same results for $\tilde{F}_2\tilde{Z}$. Thus, it does not matter which parameterisation of H_2 we use.

Next, present some lemmas necessary for the proofs of the Z and W-tests.

Lemma A.5. *For $m = 1, 2$:*

a) Under Assumptions 1 to 4 and 8,

$$\frac{1}{T_m}\sum_{t=1}^{T}\left\|(\tilde{f}_{m,t} - H_m^{\top} f_{m,t})\iota_{mt}\right\|^4 = O_p\left(\frac{1}{\delta_{NT}^4}\right),$$

$$\frac{1}{T_2}\sum_{t=\lfloor \pi T+1\rfloor}^{T}\left\|(\tilde{Z}\tilde{f}_{2,t} - H_1^{\top} Z f_t)\right\|^4 = O_p\left(\frac{1}{\delta_{NT}^4}\right),$$

b) Under Assumptions 1 to 4 and 8, $\frac{1}{T_m}\sum_{t=1}^{T}\left\|\tilde{f}_{m,t}\iota_{mt}\right\|^4 = O_p(1)$ and $\frac{1}{T_2}\sum_{t=\lfloor \pi T+1\rfloor}^{T}\left\|\tilde{Z}\tilde{f}_{2,t}\right\|^4 = O_p(1)$.

c) Under Assumptions 1 to 8, $\|H_m - H_{m,0}\| = O_p\left(\frac{1}{\delta_{NT}}\right)$.

d) Under Assumptions 1 to 4 and 8 $\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_{m,i} - H_m^{-1}\lambda_{m,i}\right\|^4 = O_p\left(\frac{1}{\delta_{NT}^4}\right)$.

Proof of Lemma A.5. Lemmas A.5 (a) to A.5 (c) are just the subsample counterparts of Lemmas 5.i), 5.ii), 5.iii) and 6 of Han and Inoue (2015), but with the addition of the use of the rotated set of factors.

The second part of Lemma A.5 (a) follows by

$$\begin{aligned}
&\frac{1}{T_2}\sum_{t=\lfloor \pi T+1\rfloor}^{T}\left\|(\tilde{Z}\tilde{f}_{2,t} - H_1^{\top} Z f_t)\right\|^4 \\
=&\frac{1}{T_2}\sum_{t=\lfloor \pi T+1\rfloor}^{T}\left\|\tilde{Z}\left(\tilde{f}_{2,t} - H_2^{\top} f_t\right) + \left(\tilde{Z} - H_1^{\top} Z H_2^{-\top}\right) H_2^{\top} f_t\right\|^4 \\
\leq&\left\|\tilde{Z}\right\|^4 \frac{1}{T_2}\sum_{t=\lfloor \pi T+1\rfloor}^{T}\left\|\tilde{f}_{2,t} - H_2^{\top} f_t\right\|^4 + \left\|\tilde{Z} - H_1^{\top} Z H_2^{-\top}\right\| \frac{1}{T_2}\sum_{t=\lfloor \pi T+1\rfloor}^{T}\left\|H_2^{\top} f_t\right\|^4 \\
=&O_p\left(\frac{1}{\delta_{NT}^4}\right) + \left[O_p\left(\frac{1}{\delta_{NT}^2}\right)\right]^4 \\
=&O_p\left(\frac{1}{\delta_{NT}^4}\right),
\end{aligned}$$

using Assumption 1, Lemma A.1 (a), and Theorem 2.1 (a).

The second part of Lemma A.5 (b) follows by the fact that $\tilde{Z} = O_p(1)$.

Lemma A.5 (d) follows by symmetry from Lemma A.5 (a).

■

A.1.3 Z-test Proofs

Define the long run variance estimate of the $A_Z(\pi, \widehat{F})$ as $\widehat{S}_Z(\pi, \widehat{F}) = \frac{1}{\pi}\widehat{\Omega}_{Z,(1)}(\pi, \widehat{F}) + \frac{1}{1-\pi}\widehat{\Omega}_{Z,(2)}(\pi, \widehat{F})$, a weighted average of the variance from pre- and post-break data ($m = 1, 2$, respectively)

$$\widehat{\Omega}_{Z,(m)}(\pi, \widehat{F}) = \widehat{\Gamma}_{(m),0}(\pi, \widehat{F}) + \sum_{j=1}^{T_m-1} \mathbf{k}\left(\frac{j}{b_{T_m}}\right)\left(\widehat{\Gamma}_{(m),j}(\pi, \widehat{F}) + \widehat{\Gamma}_{(m),j}(\pi, \widehat{F})^\top\right),$$

$$\widehat{\Gamma}_{(1),j}(\pi, \widehat{F}) = \frac{1}{T_1}\sum_{t=j+1}^{T_1} vech(\widehat{f}_t\widehat{f}_t^\top - I_r)vech(\widehat{f}_t\widehat{f}_t^\top - I_r)^\top, \quad \text{and} \tag{A.1.10}$$

$$\widehat{\Gamma}_{(2),j}(\pi, \widehat{F}) = \frac{1}{T_2}\sum_{t=j+T_1+1}^{T} vech(\widehat{f}_t\widehat{f}_t^\top - I_r)vech(\widehat{f}_t\widehat{f}_t^\top - I_r)^\top, \tag{A.1.11}$$

where $\mathbf{k}(.)$ is a real valued kernel, b is the bandwidth, and its subscripts denotes the size of the (sub)samples used to estimate the long-run variance.

The proof of Theorem 2.2 requires proving the following lemmas:

Lemma A.6.

a) Under Assumptions 1 to 8, if $\frac{\sqrt{T}}{N} \to \infty$, then
$\left\|A_Z(\pi, \widehat{F}) - A_Z(\pi, FH_{0,1})\right\| \xrightarrow{p} 0.$

b) Under Assumptions 1 to 8, and if the conditions in Assumption 9 hold, then
$\left\|\widehat{S}(\pi, \widehat{F}) - \widehat{S}(\pi, FH_{0,1})\right\| \xrightarrow{p} 0.$

Proof of Lemma A.6 (a). Taking the norm of $A_Z(\pi, \widehat{F}) - A_Z(\pi, FH_{0,1})$, we have

$$\begin{aligned}\left\|A_Z(\pi, \widehat{F}) - A_Z(\pi, FH_{0,1})\right\| = \Bigg\| & vech\sqrt{T}\left(\frac{1}{\pi T}\sum_{t=1}^{\lfloor \pi T\rfloor}\widehat{f}_t\widehat{f}_t^\top - \frac{1}{(1-\pi)T}\sum_{t=\lfloor \pi T+1\rfloor}^{T}\widehat{f}_t\widehat{f}_t^\top\right) \\ & - vech\sqrt{T}\left(\frac{1}{\pi T}\sum_{t=1}^{\lfloor \pi T\rfloor}H_{0,1}^\top f_t f_t^\top H_{0,1} - \frac{1}{(1-\pi)T}\sum_{t=\lfloor \pi T+1\rfloor}^{T}H_{0,1}^\top f_t f_t^\top H_{0,1}\right)\Bigg\|.\end{aligned}$$

Because $\widehat{f}_t$ is consistent for $H_1^\top f_t$, and H_1 is consistent for $H_{0,1}$, it suffices to prove that

$$A_Z.I = \sqrt{T}\left\|vech\left(\frac{\sum_{t=1}^{\lfloor \pi T\rfloor}\widehat{f}_t\widehat{f}_t^\top}{\pi T} - \frac{\sum_{t=\lfloor \pi T+1\rfloor}^{T}\widehat{f}_t\widehat{f}_t^\top}{(1-\pi)T}\right) - vech\left(\frac{\sum_{t=1}^{\lfloor \pi T\rfloor}H_1^\top f_t f_t^\top H_1}{\pi T} - \frac{\sum_{t=\lfloor \pi T+1\rfloor}^{T}H_1^\top f_t f_t^\top H_1}{(1-\pi)T}\right)\right\|$$

and

$$A_Z.II = \sqrt{T}\left\| vech\left(\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} H_1^\top f_t f_t^\top H_1}{\pi T} - \frac{\sum_{t=\lfloor \pi T+1\rfloor}^{T} H_1^\top f_t f_t^\top H_1}{(1-\pi)T}\right) - vech\left(\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} H_{0,1}^\top f_t f_t^\top H_{0,1}}{\pi T} - \frac{\sum_{t=\lfloor \pi T+1\rfloor}^{T} H_{0,1}^\top f_t f_t^\top H_{0,1}}{(1-\pi)T}\right)\right\|$$

are both $o_p(1)$. The first term $A_Z.I$ is bounded by

$$A_Z.I \leq \sqrt{T}\left\| vech\left(\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} \widehat{f}_t\widehat{f}_t^\top - H_1^\top f_t f_t^\top H_1}{\pi T}\right)\right\| + \sqrt{T}\left\| vech\left(\frac{\sum_{t=\lfloor \pi T+1\rfloor}^{T} \widehat{f}_t\widehat{f}_t^\top - H_1^\top f_t f_t^\top H_1}{(1-\pi)T}\right)\right\|.$$

We focus on proving that the first term is $o_p(1)$, as the second term can be proved very similarly. The first term is bounded by

$$\begin{aligned}
&\sqrt{T}\left\| vech\left(\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} \widehat{f}_t(\widehat{f}_t^\top - H_1^\top f_t)^\top + \widehat{f}_t(H_1^\top f_t) - H_1^\top f_t f_t^\top H_1}{\pi T}\right)\right\| \\
=&\sqrt{T}\left\| vech\left(\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} \widehat{f}_t(\widehat{f}_t^\top - H_1^\top f_t) + (\widehat{f}_t^\top - f_t H_1) f_t^\top H_1}{\pi T}\right)\right\| \\
=&\sqrt{T}\left\| vech\left(\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} (\widehat{f}_t - H_1^\top f_t)(\widehat{f}_t^\top - f_t^\top H_1) + H_1^\top f_t(\widehat{f}_t^\top - f_t^\top H_1) + (\widehat{f}_t - H_1^\top f_t) f_t^\top H_1}{\pi T}\right)\right\| \\
\leq&\sqrt{T}\left(\left\|\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} (\widehat{f}_t - H_1^\top f_t)(\widehat{f}_t^\top - f_t^\top H_1)}{\pi T}\right\| + 2\left\|\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} H_1^\top f_t(\widehat{f}_t^\top - f_t^\top H_1)}{\pi T}\right\|\right) \\
=&O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) \\
=&o_p(1)
\end{aligned}$$

as $\frac{\sqrt{T}}{N} \to 0$, where each of the terms on the second last line are $O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)$ by Lemma A.1 (a). Next,

$A_Z.II$ can be bounded by

$$\begin{aligned}
&\sqrt{T}\left\|\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} H_1^\top f_t f_t^\top H_1 - H_{0,1}^\top f_t f_t^\top H_{0,1}}{\pi T} - \frac{\sum_{t=\lfloor \pi T+1\rfloor}^{T} H_1^\top f_t f_t^\top H_1 - H_{0,1}^\top f_t f_t^\top H_{0,1}}{(1-\pi)T}\right\| \\
=&\sqrt{T}\left\|\frac{\sum_{t=1}^{\lfloor \pi T\rfloor}(H_1 - H_{0,1})^\top f_t f_t^\top H_1 + H_{0,1}^\top f_t f_t^\top (H_1 - H_{0,1})}{\pi T}\right. \\
&\quad\left. - \frac{\sum_{t=\lfloor \pi T+1\rfloor}^{T}(H_1 - H_{0,1})^\top f_t f_t^\top H_1 + H_{0,1}^\top f_t f_t^\top (H_1 - H_{0,1})}{(1-\pi)T}\right\| \\
\leq&\sqrt{T}\|H_1 - H_{0,1}\|\left\|\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} f_t f_t^\top}{\pi T} - \frac{\sum_{t=\lfloor \pi T+1\rfloor}^{T} f_t f_t^\top}{(1-\pi)T}\right\|(\|H_1\| + \|H_{0,1}\|) \\
=&\sqrt{T} o_p(1) O_p\left(\frac{1}{\sqrt{T}}\right) O_p(1),
\end{aligned}$$

because we use Assumption 8 (b) and $\|H_1 - H_{0,1}\| = o_p(1)$, which is implied by Assumptions 1 to 4 and 7 (see Bai, 2003). ■

To prove Lemma A.6 (b), we need to prove the following lemmas first.

Lemma A.7. *Under Assumptions 1 to 5, for $m = 1, 2$:*

$$\left\|\widehat{\Gamma}_{(m),j}(\pi, \widehat{F}) - \widehat{\Gamma}_{(m),j}(\pi, FH_1)\right\| = O_p\left(\frac{1}{\delta_{NT}}\right).$$

Proof of Lemma A.7. We shall focus on the case of $m = 1$, as the case for $m = 2$ is analogous and thus omitted.

$$
\begin{aligned}
&\left\|\widehat{\Gamma}_{(1),j}(\pi,\widehat{F})-\widehat{\Gamma}_{(1),j}(\pi,FH_1)\right\| \\
\leq&\left\|\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor} vech(\widehat{f}_t\widehat{f}_t^\top-I_r)vech(\widehat{f}_{t-j}\widehat{f}_{t-j}^\top-I_r)^\top-vech(H_1^\top f_tf_t^\top H_1-I_r)vech(H_1^\top f_{t-j}f_{t-j}^\top H_1-I_r)^\top\right\| \\
\leq&\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|vech(\widehat{f}_t\widehat{f}_t^\top-I_r)vech(\widehat{f}_{t-j}\widehat{f}_{t-j}^\top-H_1^\top f_{t-j}f_{t-j}^\top H_1)\right\|+ \\
&\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|vech(\widehat{f}_t\widehat{f}_t^\top-H_1^\top f_tf_t^\top H_1)vech(H_1^\top f_{t-j}f_{t-j}^\top H_1-I_r)^\top\right\| \\
\leq&\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_t\widehat{f}_t^\top\right\|\left\|\widehat{f}_{t-j}\widehat{f}_{t-j}^\top-H_1^\top f_{t-j}f_{t-j}^\top H_1\right\|+ \\
&\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}r\left\|\widehat{f}_{t-j}\widehat{f}_{t-j}^\top-H_1^\top f_{t-j}f_{t-j}^\top H_1\right\|+\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_t\widehat{f}_t^\top-H_1^\top f_tf_t^\top H_1\right\|\left\|H_1^\top f_{t-j}f_{t-j}^\top H_1\right\| \\
&+\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}r\left\|\widehat{f}_j\widehat{f}_j^\top-H_1^\top f_tf_t^\top H_1\right\| \\
=&\Gamma.I+\Gamma.II+\Gamma.III+\Gamma.IV.
\end{aligned}
$$

We proceed by bounding Term $\Gamma.I$:

$$
\begin{aligned}
\Gamma.I&=\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_t\widehat{f}_t^\top\right\|\left\|\widehat{f}_{t-j}\widehat{f}_{t-j}^\top-H_1^\top f_{t-j}f_{t-j}^\top H_1\right\| \\
&\leq\frac{1}{\pi T}\left(\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_t\right\|^4\right)^{1/2}\left(\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_{t-j}(\widehat{f}_{t-j}^\top\widehat{f}_{t-j}^\top H_1)+(\widehat{f}_{t-j}-H_1^\top f_{t-j})f_{t-j}^\top H_1\right\|^2\right)^{1/2} \\
&\leq\left(\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_t\right\|^4\right)^{1/2}\left(\frac{2}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_{t-j}(\widehat{f}_{t-j}^\top-\widehat{f}_{t-j}^\top H_1)\right\|^2+\frac{2}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|(\widehat{f}_{t-j}-\widehat{f}_{t-j}H_1)f_{t-j}^\top H_1\right\|^2\right)^2 \\
&\leq\left(\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_t\right\|^4\right)^{1/2} \\
&\quad\times\left[\left(\frac{2}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_{t-j}\right\|^4\frac{2}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_{t-j}-f_{t-j}H_1\right\|^4\right)^{1/2}\right. \\
&\qquad\left.+\left(\frac{2}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_{t-j}-H_1^\top f_{t-j}\right\|^4\frac{2}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|f_{t-j}H_1\right\|^4\right)^{1/2}\right]^{1/2} \\
&=O_p\left(\frac{1}{\delta_{NT}}\right),
\end{aligned}
$$

under Assumptions 1 to 5, where $\frac{1}{T}\sum_{t=1}^{T}\left\|\widehat{f}_t\right\|^4 = O_p(1)$ by Lemma A.5 (a). Using similar arguments, terms $\Gamma.II$, $\Gamma.III$, and $\Gamma.IV$ can be shown to be $O_p\left(T^{-1/4}\right) + O_p\left(\frac{1}{\sqrt{N}}\right)$. ■

Lemma A.8. *Under Assumptions 1 to 5, for $m = 1, 2$,*

$$\left\|\widehat{\Gamma}_{m,j}(\pi, F_m H_1) - \widehat{\Gamma}_{m,j}(\pi, F_m H_{0,1})\right\| = O_p\left(\frac{1}{\delta_{NT}}\right).$$

Proof of Lemma A.8. We shall only prove the lemma for $m = 1$, because the proof for $m = 2$ is similar and thus omitted.

$$\begin{aligned}
&\left\|\widehat{\Gamma}_{1,j}(\pi, FH_1) - \widehat{\Gamma}_{1,j}(\pi, FH_{0,1})\right\| \\
=&\left\|\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\Big[vech(H_1^\top f_t f_t^\top H_1 - I_r)vech(H_1^\top f_{t-j} f_{t-j}^\top H_1 - I_r)\right. \\
&\quad \left. - vech(H_{0,1}^\top f_t f_t^\top H_{0,1} - I_r)vech(H_{0,1}^\top f_{t-j} f_{t-j}^\top H_{0,1} - I_r)^\top\Big]\right\| \\
\leq&\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|H_1^\top f_t f_t^\top H_1\right\|\left\|H_1^\top f_{t-j} f_{t-j}^\top H_1 - H_{0,1}^\top f_{t-j} f_{t-j}^\top H_{0,1}\right\| + \\
&\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|H_1^\top f_{t-j} f_{t-j}^\top H_1 - H_{0,1}^\top f_{t-j} f_{t-j}^\top H_{0,1}\right\| + \\
&\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|\widehat{f}_t\right\|^2\left\|H_1^\top f_t f_t^\top H_1 - H_{0,1}^\top f_t f_t^\top H_{0,1}\right\| + \\
&\frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|H_1^\top f_t f_t^\top H_1 - H_{0,1}^\top f_t f_t^\top H_{0,1}\right\|\left\|H_{0,1}^\top f_{t-j} f_{t-j}^\top H_{0,1}\right\| \\
=&\Gamma.V + \Gamma.VI + \Gamma.VII + \Gamma.VIII.
\end{aligned}$$

Term $\Gamma.V$ is bounded by:

$$\begin{aligned}
\Gamma.V &= \frac{1}{\pi T}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|H_1^\top f_t f_t^\top H_1\right\|\left\|H_1^\top f_{t-j} f_{t-j}^\top H_1 - H_{0,1}^\top f_{t-j} f_{t-j}^\top H_{0,1}\right\| \\
&\leq \left(\frac{1}{\pi T}\sum_{t=1}^{T}\|f_t H_1\|^4\right)^{1/2}\left(\frac{1}{\pi T}(\|H_1\|^2 + \|H_{0,1}\|^2)\sum_{t=1}^{T}\|f_t\|^4\right)^{1/2}\|H_1 - H_{0,1}\| \\
&= O_p(1)\, O_p\left(\frac{1}{\delta_{NT}}\right)
\end{aligned}$$

by Assumption 1 and Lemma A.5 (c). The proofs of terms $\Gamma.VI, \Gamma.VII, \Gamma.VIII$ are similar and thus omitted. ■

Proof of Lemma A.6 (b). It suffices to show that

$$\left\|\widehat{\Omega}_{Z,m}(\pi,\widehat{F})-\widehat{\Omega}_{Z,m}(\pi,FH_{0,1})\right\|\xrightarrow{p}0,\quad\text{for}\quad m=1,2.$$

For brevity, we will only prove the case for $m=1$, as the case for $m=2$ can be proved similarly. First, see that

$$\begin{aligned}\left\|\widehat{\Omega}_{Z,1}(\pi,\widehat{F})-\widehat{\Omega}_{Z,1}(\pi,FH_{0,1})\right\|\leq&\left\|\widehat{\Omega}_{Z,1}(\pi,\widehat{F})-\widehat{\Omega}_{Z,1}(\pi,FH_1)\right\|+\left\|\widehat{\Omega}_{Z,1}(\pi,\widehat{FH_1})-\widehat{\Omega}_{Z,1}(\pi,FH_{0,1})\right\|\\=&\Omega_Z.I+\Omega_Z.II\end{aligned}$$

For term $\Omega_Z.I$, we have:

$$\begin{aligned}&\left\|\widehat{\Omega}_{Z,1}(\pi,\widehat{F})-\widehat{\Omega}_{Z,1}(\pi,FH_1)\right\|\\\leq&\left\|\widehat{\Gamma}_{(1),0}(\pi,\widehat{F})+\sum_{t=1}^{\lfloor\pi T\rfloor}\mathbf{k}\left(\frac{j}{b_{\lfloor\pi T\rfloor}}\right)\left(\widehat{\Gamma}_{(1),j}(\pi,\widehat{F})+\widehat{\Gamma}_{(1),j}(\pi,\widehat{F})^{\top}\right)\right.\\&\left.-\widehat{\Gamma}_{(1),0}(\pi,FH_1)+\sum_{t=1}^{\lfloor\pi T\rfloor}\mathbf{k}\left(\frac{j}{b_{\lfloor\pi T\rfloor}}\right)\left(\widehat{\Gamma}_{(1),j}(\pi,FH_1)+\widehat{\Gamma}_{(1),j}(\pi,FH_1)^{\top}\right)\right\|.\end{aligned}$$

Recall that $\left|\mathbf{k}\left(\frac{j}{b_{\lfloor\pi T\rfloor}}\right)\right|\leq 1$ and $\mathbf{k}\left(\frac{j}{b_{\lfloor\pi T\rfloor}}\right)=0$ if $j>b_{\lfloor\pi T\rfloor}$ for the Bartlett kernel. Thus,

$$\Omega_Z.I\leq\left\|\widehat{\Gamma}_{1,0}(\pi,\widehat{F})-\widehat{\Gamma}_{1,0}(\pi,FH_1)\right\|+2\sum_{j=1}^{b_{\lfloor\pi T\rfloor}}\left\|\widehat{\Gamma}_{1,j}(\pi,\widehat{F})-\widehat{\Gamma}_{1,j}(\pi,FH_1)\right\|.$$

In the case of the Bartlett kernel, $\Omega_Z.I$ is $O_p\left(\frac{T^{1/3}}{\delta_{NT}}\right)$ by Lemma A.8 and the condition in Assumption 9 (a), which states that $b_{\lfloor\pi T\rfloor}\leq KT^{1/3}$, so $\Omega_Z.I$ is $o_p(1)$ if $\frac{T^{2/3}}{N}\to 0$ as $N,T\to\infty$.

The term $\Omega_Z.II$ can be also be shown to be $o_p(1)$ with similar arguments. ■

To prove the consistency of the Wald test statistic in Theorem 2.2, we present the following lemmas:

Lemma A.9. *Under Assumptions 1 to 9,* *a)* $\left\|\widehat{S}(\pi, FH_{0,1})^{-1}\right\| = O_p(1)$ *and* $\left\|\widehat{S}(\pi, \widehat{F})^{-1}\right\| = O_p(1)$,

b) $\left\|\widehat{S}(\pi, FH_{0,1})^{-1} - \widehat{S}(\pi, \widehat{F})^{-1}\right\| = o_p(1)$.

Proof of Lemma A.9. For Lemma A.9 (a), because $0 < \pi < 1$, this implies that $\left\|\widehat{S}(\pi, FH_{0,1}) - \left(\frac{1}{\pi} + \frac{1}{1-\pi}\Omega\right)\right\| = o_p(1)$. Let ρ_{min}, ρ_{max} denote the minimum and maximum eigenvalues of a symmetric matrix, respectively. Since Ω is positive definite, $\left|\rho_{min}(\widehat{S}(\pi, FH_{0,1})) - \rho_{min}\left(\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\Omega\right)\right| \leq \left\|\widehat{S}(\pi, FH_{0,1})\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\Omega\right\| = o_p(1)$. This means that the eigenvalues of $\widehat{S}(\pi, FH_{0,1})$ are bounded away from zero, so $\left\|\widehat{S}(\pi, FH_{0,1})\right\| = O_p(1)$. For the second part of this lemma, we have $\left\|\widehat{S}(\pi, \widehat{F}) - \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\Omega\right\| \leq \left\|\widehat{S}(\pi, \widehat{F}) - \widehat{S}(\pi, FH_{0,1})\right\| + \left\|\widehat{S}(\pi, FH_{0,1}) - \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\Omega\right\| = o_p(1)$ by Lemma A.6 (b) and Assumption 9 (a). Therefore,

$$\left|\rho_{min}(\widehat{S}(\pi, \widehat{F})) - \rho_{min}\left(\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\Omega\right)\right| \leq \left\|\widehat{S}(\pi, \widehat{F}) - \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\Omega\right\| = o_p(1),$$

which means that the eigenvalues of $\widehat{S}(\pi, \widehat{F})$ are also bounded away from zero, which subsequently implies that $\widehat{S}(\pi, \widehat{F})^1 = O_p(1)$. For Lemma A.9 (b):

$$\begin{aligned}\left\|\widehat{S}(\pi, \widehat{F})^{-1} - \widehat{S}(\pi, FH_{0,1})^{-1}\right\| =& \left\|\widehat{S}(\pi, FH_{0,1})^{-1}\left(\widehat{S}(\pi, FH_{0,1}) - \widehat{S}(\pi, \widehat{F})\right)\widehat{S}(\pi, \widehat{F})^{-1}\right\| \\ \leq& \left\|\widehat{S}(\pi, FH_{0,1})^{-1}\right\|\left\|\widehat{S}(\pi, FH_{0,1}) - \widehat{S}(\pi, \widehat{F})\right\|\left\|\widehat{S}(\pi, \widehat{F})^{-1}\right\| \\ =& O_p(1)\, o_p(1) O_p(1) = o_p(1),\end{aligned}$$

by Lemma A.6 (b) and Assumption 9 (a). ■

Proof of Theorem 2.2.

$$\begin{aligned}\left|\mathscr{W}_Z(\pi, \widehat{F}) - \mathscr{W}_Z(\pi, FH_{0,1})\right| \leq& \left|A_Z(\pi, \widehat{F})^\top\left[\widehat{S}(\pi, \widehat{F})^{-1} - \widehat{S}(\pi, FH_{0,1})^{-1}\right]A_Z(\pi, \widehat{F})\right| \\ &+ \left|\left[A_Z(\pi, \widehat{F}) - A_Z(\pi, FH_{0,1})\right]^\top \widehat{S}(\pi, FH_{0,1})^{-1}A_Z(\pi, \widehat{F})\right| \\ &+ \left|A_Z(\pi, FH_{0,1})^\top \widehat{S}(\pi, FH_{0,1})^{-1}\left[A_Z(\pi, \widehat{F}) - A_Z(\pi, FH_{0,1})\right]\right| \\ =& o_p(1),\end{aligned}$$

using the results of Lemma A.9 and Lemma A.6 (a). ■

Proof of Theorem 2.3. Under the alternative hypothesis, $Z \neq I$, so we have:

$$
\begin{aligned}
&\frac{1}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \widehat{f}_t \widehat{f}_t^\top - \frac{1}{T - \lfloor \pi T \rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} \widehat{f}_t \widehat{f}_t^\top \\
=&\left(\frac{1}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} H_1^\top f_t f_t^\top H_1 - \frac{1}{T - \lfloor \pi T \rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} H_1^\top Z f_t f_t^\top Z^\top H_1 \right) \\
&+ \frac{1}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \left(\widehat{f}_t \widehat{f}_t^\top - H_1^\top f_t f_t^\top H_1 \right) - \frac{1}{T - \lfloor \pi T \rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} \left(\widehat{f}_t \widehat{f}_t^\top - H_1^\top Z f_t f_t^\top Z^\top H_1 \right).
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\frac{1}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \left(\widehat{f}_t \widehat{f}_t^\top - H_1^\top f_t f_t^\top H_1 \right) \\
=&\frac{1}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \left[(\widehat{f}_t - H_1^\top f_t) f_t^\top H_1 + (\widehat{f}_t - H_1^\top f_t)(\widehat{f}_t^\top - f_t^\top H_1) + H_1^\top f_t (\widehat{f}_t^\top - f_t^\top H_1) \right] \\
=&O_p\left(\frac{1}{\delta_{NT}^2} \right),
\end{aligned}
$$

by the arguments in the proof of Lemma A.6 (a). Similarly, $\frac{1}{T-\lfloor \pi T \rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} \left(\widehat{f}_t \widehat{f}_t^\top - H_1^\top Z f_t f_t^\top Z^\top H_1 \right) = O_p\left(\frac{1}{\delta_{NT}^2} \right)$.

Under the alternative hypothesis where there is a rotational break, we have:

$$
H_1^\top \left(\frac{1}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t f_t^\top - \frac{1}{T - \lfloor \pi T \rfloor} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} Z f_t f_t^\top Z^\top \right) H_1 \xrightarrow{p} H_{0,1}^\top (\Sigma_F - Z \Sigma_F Z^\top) H_{0,1} \equiv C,
$$

by Assumption 8, and the definitions of H_1 and $H_{0,1}$. Matrix C contains non-zero entries because $\Sigma_F - Z\Sigma_F Z^\top$ is not zero by Assumption 11, and the fact that $H_{0,1}$ is non-singular. Note that Assumptions 1 to 8 still hold under the alternative hypothesis, and hence Lemma A.6 (b) still holds for the equivalent models under the alternative. Finally, putting the above together we have:

$$\begin{aligned}
\mathscr{W}_Z(\pi,\widehat{F}) &= A_Z(\pi,\widehat{F})^\top \widehat{S}(\pi,\widehat{F})^{-1} A_Z(\pi,\widehat{F}) \\
&= \frac{T}{\max(b_{\lfloor \pi T\rfloor}, b_{T-\lfloor \pi T\rfloor})}\left[\frac{1}{\sqrt{T}} A_Z(\pi,\widehat{F})^\top\right]\left[\max(b_{\lfloor \pi T\rfloor}, b_{T-\lfloor \pi T\rfloor})\widehat{S}(\pi,\widehat{F})^{-1}\right]\left[\frac{1}{\sqrt{T}} A_Z(\pi,\widehat{F})\right] \\
&= \frac{T}{\max(b_{\lfloor \pi T\rfloor}, b_{T-\lfloor \pi T\rfloor})}\left[vech(C)^\top + o_p(1)\right]\left[\max(b_{\lfloor \pi T\rfloor}, b_{T-\lfloor \pi T\rfloor})\widehat{S}(\pi,\widehat{F})^{-1}\right]\left[vech(C) + o_p(1)\right] \\
&\to \infty
\end{aligned}$$

by Assumptions 11 and 12. ■

A.1.4 W-test Proofs

We first recall the following identity for the factor loadings for $m = 1, 2$:

$$\tilde{\lambda}_{m,i} - H_m^{-1}\lambda_{m,i} = \frac{1}{T_m} H_m^\top F_m^\top e_{(m),i} + \frac{1}{T_m}\tilde{F}^\top (F_m - \tilde{F}_m H_m^{-1})\lambda_{m,i} + \frac{1}{T_m}(\tilde{F}_m - F_m H_m)^\top e_{(m),i}, \tag{A.1.12}$$

where Equation (A.1.12) is the subsample version of the asymptotic expansion of the factor loadings considered by Bai (2003) (see the proof of their Theorem 2). The last two terms are both $O_p\left(\frac{1}{\delta_{NT}^2}\right)$ by Lemma A.1 (d) and Lemma A.1 (c), and we therefore have the following lemma.

Lemma A.10. *Under Assumptions 1 to 6, for $m = 1, 2$,*

$$\tilde{\lambda}_{m,i} - H_m^{-1}\lambda_{m,i} = H_m^\top \frac{1}{T_m}\sum_{t=1}^{T} f_{m,t} e_{it} \iota_{mt} + O_p\left(\frac{1}{\delta_{NT}^2}\right) \tag{A.1.13}$$

for each i.

Lemma A.10 is simply the subsample counterpart of Equation B.2 in Bai (2003).

Proof of Theorem 2.4 (a). Recall that $\tilde{W} = \tilde{\Lambda}_2 - \tilde{\Lambda}_1 \tilde{Z}$, which implies that

$$\begin{aligned}
\tilde{\lambda}_{2,i} &= \tilde{Z}^\top \tilde{\lambda}_{1,i} + \tilde{w}_i, \\
\tilde{w}_i &= \tilde{\lambda}_{2,i} - \tilde{Z}^\top \tilde{\lambda}_{1,i}.
\end{aligned} \tag{A.1.14}$$

Substituting in the decompositions in Lemma A.10, we have

$$\tilde{\lambda}_{2,i} - H_2^{-1}\lambda_{2,i} = H_2^\top \frac{1}{(1-\pi)T} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} f_t e_{it} + O_p(\delta_{NT}^{-2})$$

$$(\tilde{Z}^\top \tilde{\lambda}_{1,i} + \tilde{w}_i) - H_2^{-1}(Z^\top \lambda_{1,i} + w_i) = H_2^\top \frac{1}{(1-\pi)T} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} f_t e_{it} + O_p(\delta_{NT}^{-2})$$

$$(\tilde{w}_i - H_2^{-1}w_i) = H_2^\top \frac{1}{(1-\pi)T} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} f_t e_{it} - (\tilde{Z}^\top \tilde{\lambda}_{1,i} - H_2^{-1}Z^\top \lambda_{1,i}) + O_p(\delta_{NT}^{-2}).$$

We now focus on the asymptotic expansion of $\tilde{Z}^\top \tilde{\lambda}_{1,i} - H_2^{-1}Z^\top \lambda_{1,i}$:

$$\begin{aligned}
&\tilde{Z}^\top \tilde{\lambda}_{1,i} - H_2^{-1}Z^\top \lambda_{1,i} \\
=&\tilde{Z}^\top(\tilde{\lambda}_{1,i} - H_1^{-1}\lambda_{1,i}) + \left(\tilde{Z} - H_1^\top Z H_2^{-\top}\right)^\top H_1^{-1}\lambda_{1,i} + \left(H_1^\top Z H_2^{-\top}\right)^\top H_1^{-1}\lambda_{1,i} - H_2^{-1}Z^\top \lambda_{1,i} \\
=&\tilde{Z}^\top(\tilde{\lambda}_{1,i} - H_1^{-1}\lambda_{1,i}) + O_p\left(\frac{1}{\delta_{NT}^2}\right) O_p(1) \\
=&\tilde{Z}^\top(\tilde{\lambda}_{1,i} - H_1^{-1}\lambda_{1,i}) + O_p\left(\frac{1}{\delta_{NT}^2}\right),
\end{aligned}$$

by Theorem 2.1 (a).

Applying Lemma A.10 to expand $(\tilde{\lambda}_{1,i} - H_1^{-1}\lambda_{1,i})$, we have

$$(\tilde{w}_i - H_2^{-1}w_i) = H_2^\top \frac{1}{(1-\pi)T} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} f_t e_{it} - \tilde{Z}^\top \frac{1}{\pi T} \sum_{t=1}^{\lfloor \pi T \rfloor} H_1^\top f_t e_{it} + O_p\left(\frac{1}{\delta_{NT}^2}\right).$$

Multiplying both sides by $\sqrt{T}$ then yields

$$\sqrt{T}(\tilde{w}_i - H_2^{-1}w_i) = H_2^\top \frac{1}{(1-\pi)\sqrt{T}} \sum_{t=\lfloor \pi T+1 \rfloor}^{T} f_t e_{it} - \tilde{Z}^\top \frac{1}{\pi \sqrt{T}} \sum_{t=1}^{\lfloor \pi T \rfloor} H_1^\top f_t e_{it} + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right),$$

where the remainder term is $o_p(1)$ as $\frac{\sqrt{T}}{N} \to 0$. Recognising the CLT random variable terms in Assumption 16, we have

$$\sqrt{T}(\tilde{w}_i - H_2^{-1}w_i) \xrightarrow{d} N(0, \Omega_{W,i})$$

where

$$\Omega_{W,i} = \left(\frac{1}{1-\pi}\right) H_{0,2}^{\top}\Phi_i H_{0,2} + \left(\frac{1}{\pi}\right) H_{0,2}^{-1} Z' \Sigma_F^{-1}\Phi_i\Sigma_F^{-1} Z H_{0,2}^{-\top}$$
$$= \left(\frac{1}{1-\pi}\right)\Theta_{1,i} + \left(\frac{1}{\pi}\right)\Theta_{2,i}.$$

The form of $\Theta_{2,i}$ comes from the fact that $\tilde{Z}$ estimates $H_1 Z H_2^{-\top}$ by Theorem 2.1 (a). By the convergence of H_1 to its limit $H_{0,1}$, we have

$$\tilde{Z}^{\top} H_1^{\top} \xrightarrow{p} H_{0,2}^{-1} Z^{\top} H_{0,1} H_{0,1}^{\top}$$
$$= H_{0,2}^{-1} Z^{\top} \Sigma_F^{-1},$$

where the last line follows from the identity $H_{0,1}H_{0,1}^{\top} = \Sigma_F^{-1}$. To see this, recall that $H_{0,1}^{-1} = V^{1/2}\Upsilon_1^{\top}\Sigma_{\Lambda_1}^{1/2}$. This means that $H_{0,1}^{-1}\Sigma_F^{-1}H_{0,1}^{\top} = V_1^{1/2}\Upsilon_1^{\top}\left(\Sigma_{\Lambda_1}^{1/2}\Sigma_F^{-1}\Sigma_{\Lambda_1}^{1/2}\right)\Upsilon_1 V_1^{1/2} = V_1^{1/2}V_1^{-1}V_1^{1/2} = I$ by eigen-identity, which can then be re-arranged as required.

Their estimators $\tilde{\Theta}_{1,i}, \tilde{\Theta}_{2,i}$ are discussed in Bai (2003), and are given by HAC estimators constructed using the estimated residuals $\tilde{e}_{(1),it} = x_{it} - \tilde{\lambda}_{2,i}^{\top}\tilde{f}_{1,t}$ and $\tilde{e}_{(2),it} = x_{it} - \tilde{\lambda}_{2,i}^{\top}\tilde{f}_{2,t}$ in the series $\tilde{Z}^{\top}\tilde{f}_{1,t}\cdot\tilde{e}_{(1),it}$ and $\tilde{f}_{2,t}\cdot\tilde{e}_{(2),it}$, respectively:

$$\tilde{\Theta}_{1,i} = D_{0,1,i} + \sum_{v=1}^{\lfloor\pi T\rfloor-1}\mathbf{k}\left(\frac{v}{b_{\lfloor\pi T\rfloor}}\right)(D_{1,vi} + D_{1,vi}^{\top}) \tag{A.1.15}$$
$$\tilde{\Theta}_{2,i} = D_{0,2,i} + \sum_{v=1}^{T-\lfloor\pi T\rfloor-1}\mathbf{k}\left(\frac{v}{b_{T-\lfloor\pi T\rfloor}}\right)(D_{2,vi} + D_{2,vi}^{\top}), \tag{A.1.16}$$

where $D_{1,vi} = (T_1)^{-1}\sum_{t=v+1}^{\lfloor\pi T\rfloor}\tilde{f}_{1,t}\tilde{e}_{it}\tilde{e}_{i,t-v}\tilde{f}_{1,t-v}^{\top}$, $D_{2,vi} = (T_2)^{-1}\sum_{t=T_1+v+1}^{T}\tilde{f}_{2,t}\tilde{e}_{it}\tilde{e}_{i,t-v}\tilde{f}_{2,t-v}^{\top}$, and $\mathbf{k}(.)$ is a real valued kernel, such as the Bartlett kernel, satisfying Assumption 9. The consistency of $\tilde{\Theta}_{1,i}$ and $\tilde{\Theta}_{2,i}$ for $\Theta_{1,i} = H_{1,0}^{\top}\Phi_{i,1}H_{1,0}$ and $\Theta_{2,i} = H_{2,0}^{\top}\Phi_{i,2}H_{2,0}$, respectively, can be proved using the argument of Newey and West (1987), as stated by Bai (2003). Recalling that $\tilde{Z}$ estimates $H_1^{\top} Z H_2^{-\top}$ from Theorem 2.1 (a), it follows that we can estimate $\Omega_{W,i}$ using

$$\widehat{\Omega}_{W,i} = \frac{1}{1-\pi}\tilde{\Theta}_{2,i} + \frac{1}{\pi}\tilde{Z}^{\top}\tilde{\Theta}_{1,i}\tilde{Z}.$$

The asymptotic Chi-squared distribution then follows. ■

Before we prove Theorem 2.4 (b), we prove some lemmas that need to be used.

Lemma A.11. *Under Assumptions 1 to 5, 8, and 15, for $m = 1, 2$ we have:*

a) $\frac{\tilde{F}_m^\top (F_m - F_m H_m)}{T_m} \frac{\sum_{i=1}^N \lambda_{m,i}}{\sqrt{N}} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$,

b) $\frac{(\tilde{F}_m - F_m H_m)^\top}{\sqrt{T_m}} \frac{\sum_{i=1}^N e_{m,i}}{\sqrt{T_m N}} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$,

c) Under Assumptions 1 to 8, 13, and 14, and if $\frac{\sqrt{T}}{N} \to 0$*, then for* $m = 1, 2$, $\left\|\tilde{\Theta}_m - \Theta_m\right\| = o_p(1)$, *where* $\Theta_m = \text{plim}(N)^{-1} \sum_{i=1}^N \Theta_{m,i}$.

Proof of Lemma A.11 (a). The first term can be bounded by

$$\begin{aligned}&\frac{\tilde{F}_m^\top (F_m - F_m H_m)}{T_m} \frac{\sum_{i=1}^N \lambda_{m,i}}{\sqrt{N}} \\ &= O_p\left(\frac{1}{\delta_{NT}^2}\right) O_p(1)\end{aligned}$$

by Lemma A.1 (d), and Assumption 15. ■

Proof of Lemma A.11 (b). It suffices to show that $\frac{1}{T_m\sqrt{N}} \sum_{t=1}^T (\tilde{f}_{m,t} - H_m^\top f_{m,t}) \sum_{i=1}^N e_{it} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$. We focus on the case of $m = 1$, as the proof for $m = 2$ is similar and thus omitted. From Equation (A.1.4), we have

$$\begin{aligned}&\frac{1}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} (\tilde{f}_{1,t} - H_1^\top f_t) \frac{1}{\sqrt{N}} \sum_{i=1}^N e_{it} \\ &= V_{NT,1}^{-1} \left(\frac{1}{T_1^2} \sum_{s=1}^{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{f}_{1,s} \gamma_N(s,t) \frac{1}{\sqrt{N}} \sum_{i=1}^N e_{it} + \frac{1}{T_1^2} \sum_{s=1}^{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{f}_{1,s} \left(\frac{e_s^\top e_t}{N} - \gamma_N(s,t) \right) \frac{1}{\sqrt{N}} \sum_{i=1}^N e_{it} \right. \\ &\qquad \left. + \frac{1}{NT_1^2} \sum_{s=1}^{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{f}_{1,t} f_t^\top \Lambda_1 e_t \frac{1}{\sqrt{N}} \sum_{i=1}^N e_{it} + \frac{1}{NT_1^2} \sum_{s=1}^{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{f}_{1,t} e_{(1)}^\top \Lambda_1 f_t \frac{1}{\sqrt{N}} \sum_{i=1}^N e_{it} \right) \\ &= V_{NT}^{-1}(a + b + c + d),\end{aligned}$$

where we shall prove that each of a, b, c, and d are asymptotically negligible.

$$
\begin{aligned}
a &= \frac{1}{T_1^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^{T} H_1^\top f_{1,s} \gamma_N(s,t) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} + \frac{1}{T_1^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^{T} (\tilde{f}_{1,s} - H_1^\top f_{1,s}) \gamma_N(s,t) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \\
&= a.I + a.II
\end{aligned}
$$

We shall prove that each of $a.I$ and $a.II$ are asymptotically negligible. First, the term $a.I$ can be bounded by

$$
\begin{aligned}
a.I &\leq \frac{1}{T_1^2} E\left(\sum_{s=1}^{T} \sum_{t=1}^{\lfloor \pi T \rfloor} f_{1,s} \gamma_N(s,t) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right) \\
&= \frac{1}{T_1^2} \sum_{s=1}^{T} \sum_{t=1}^{\lfloor \pi T \rfloor} |\gamma_N(s,t)| \, E\left(\|f_{1,s}\|^2 E\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right)^2 \right) \\
&\leq \frac{1}{T_1^2} \sum_{s=1}^{T} \sum_{t=1}^{\lfloor \pi T \rfloor} |\gamma_N(s,t)| \, M = O_p\left(\frac{1}{T} \right),
\end{aligned}
$$

by Assumptions 1, 3 (a), and 3 (c), and $E\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right) \leq M$ by Assumption 13. Next, the term $a.II$ can be bounded by

$$
\begin{aligned}
a.II &\leq \frac{1}{\sqrt{T_1}} \left(\frac{1}{T_1} \sum_{s=1}^{\lfloor \pi T \rfloor} \left\| \tilde{f}_{1,s} - H_1 f_s \right\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^{\lfloor \pi T \rfloor} |\gamma_N(s,t)|^2 \frac{1}{T_1} E\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right)^2 \right)^{\frac{1}{2}} \\
&= \frac{1}{\sqrt{T_1}} O_p\left(\frac{1}{\delta_{NT}} \right) O_p(1),
\end{aligned}
$$

where the $O_p(1)$ term follows from Assumption 3 (a), and because $T^{-1} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^{\lfloor \pi T \rfloor} |\gamma_N(s,t)|^2 \leq M$ by Lemma 1(i) of Bai and Ng (2002). Therefore, it follows that $a = O_p\left(\frac{1}{\delta_{NT}\sqrt{T}} \right)$.

Next, term b can be bounded by

$$
\begin{aligned}
b &= \frac{1}{T_1^2} \sum_{s=1}^{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{f}_{1,s} \left(\frac{e_s^\top e_t}{N} - \gamma_N(s,t) \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \\
&= b.I + b.II
\end{aligned}
$$

For $b.I$, we shall define $z_{1,t} = \frac{\sum_{s=1}^{\lfloor \pi T \rfloor} \sum_{i=1}^{N} f_{1,s} [e_{is} e_{it} - E(e_{is} e_{it})]}{\sqrt{TN}}$. By Assumption 6 (a), $E\|z_{1,t}\|^2 < M$, and thus $\frac{1}{T_1}\sum_{t=1}^{\lfloor \pi T \rfloor} \|z_{1,t}\|^2 = O_p(1)$ by Assumption 13. This implies:

$$\frac{1}{\sqrt{NT_1}}\|H_1\| \left(\frac{1}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} \|z_{1,t}\|^2 \frac{1}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right)^2 \right)^{\frac{1}{2}} = \frac{1}{\sqrt{T_1 N}} O_p(1),$$

by Assumptions 3 (a) and 6 (a), and $\|H_1\| = O_p(1)$ due to Lemma A.1 (e). We can bound $b.II$ by

$$\begin{aligned}
&\left(\frac{1}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} \left\| \tilde{f}_{1,t} - H_1^\top f_{1,s} \right\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} \left(\frac{1}{T_1} \sum_{i=1}^{N} \left(\frac{e_s^\top e_t}{N} - \gamma_N(s,t) \right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right)^2 \right)^{\frac{1}{2}} \\
\le & O_p\left(\frac{1}{\delta_{NT}} \right) \left(\frac{1}{T_1^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{i=1}^{N} \left(\frac{e_s^\top e_t}{N} - \gamma_N(s,t) \right)^2 \frac{1}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right)^2 \right)^{\frac{1}{2}} \\
= & O_p\left(\frac{1}{\delta_{NT}} \right) \frac{1}{\sqrt{N}} \left(\frac{1}{T_1^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{is}e_{it} - E(e_{is}e_{it}) \right)^2 \right) \frac{1}{T_1} \left(\sum_{i=1}^{N} \left(\frac{1}{\sqrt{N}} \left(\sum_{i=1}^{N} e_{it} \right)^2 \right) \right)^{\frac{1}{2}} \\
= & O_p\left(\frac{1}{\delta_{NT}} \right) \frac{1}{\sqrt{N}} O_p(1),
\end{aligned}$$

because of Lemma A.1 (a), the $O_p(1)$ term comes from Assumptions 3 (a) and 13. Next, term c can be bounded by

$$\begin{aligned}
c &= \frac{1}{NT_1^2} \sum_{s=1}^{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{f}_{1,s} f_{1,s}^\top \Lambda_1 e_t \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \\
&= O_p(1) \left(\frac{1}{T_1 N} \sum_{i=1}^{N} \left(\frac{1}{\sqrt{N}} \sum_{k=1}^{N} \lambda_{1,k} e_{kt} e_{it} \right) + \frac{1}{T_1 N \sqrt{N}} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{i=1}^{N} \sum_{k \neq i} \lambda_{1,k} e_{kt} e_{it} \right) \\
&= O_p(1) \left(O_p\left(\frac{1}{N} \right) + O_p\left(\frac{1}{\sqrt{TN}} \right) \right) = O_p\left(\frac{1}{\delta_{NT}\sqrt{N}} \right) \qquad \text{(A.1.17)}
\end{aligned}$$

by Assumptions 6 (b) and 15 (b). Term d can be bounded by

$$\begin{aligned}
d &= \frac{1}{NT_1^2} \sum_{s=1}^{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{f}_{1,s} e_s^\top \Lambda_1 f_t \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \\
&= \frac{1}{NT_1^2} \sum_{s=1}^{T} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{f}_{1,s} e_s^\top \Lambda_1 f_t \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} + \frac{1}{NT_1^2} \sum_{s=1}^{T} \sum_{t=1}^{\lfloor \pi T \rfloor} (\tilde{f}_{1,s} - H_1^\top f_{1,s}) e_s^\top \Lambda_1 f_t \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \\
&= d.I + d.II.
\end{aligned}$$

The first term $d.I$ can be bounded by

$$\begin{aligned}
d.I &\leq \frac{1}{\sqrt{T_1 N}} \left\| \frac{1}{\sqrt{T_1 N}} \sum_{s=1}^{T} \sum_{t=1}^{\lfloor \pi T \rfloor} H_1^\top f_{1,s} e_{it} \lambda_{1,i}^\top \right\| \left\| \frac{1}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right\| \\
&\leq \frac{1}{\sqrt{T_1 N}} O_p(1) \left(\frac{1}{T_1} \sum_{t=1}^{T} \|f_t\|^2 \frac{1}{T_1} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right)^2 \right)^{1/2} \\
&= O_p\left(\frac{1}{\sqrt{T_1 N}} \right),
\end{aligned}$$

by Assumptions 1, 6 (a), and 13.

The second term $d.II$ can be bounded by:

$$\begin{aligned}
d.II &\leq \frac{1}{\sqrt{N}} \left\| \frac{1}{T_1} \sum_{s=1}^{\lfloor \pi T \rfloor} (\widehat{f}_{1,s} - H_1^\top f_{1,s}) \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_{1,i}^\top e_{is} \right) \right\| \left\| \frac{1}{T_1} \sum_{s=1}^{\lfloor \pi T \rfloor} f_t \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right\| \\
&\leq \frac{1}{\sqrt{N}} \left(\frac{1}{T_1} \sum_{s=1}^{\lfloor \pi T \rfloor} \left\| \widehat{f}_s - H_1^\top f_s \right\|^2 \frac{1}{T_1} \sum_{s=1}^{\lfloor \pi T \rfloor} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_{1,i}^\top e_{is} \right\|^2 \right)^{1/2} \left(\frac{1}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} \|f_t\|^2 \frac{1}{T_1} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_{it} \right)^2 \right)^{1/2} \\
&= \frac{1}{\sqrt{N}} \left(O_p\left(\frac{1}{\delta_{NT}^2} \right) O_p(1) \right)^{1/2} O_p(1) = \frac{1}{\sqrt{N}} O_p\left(\frac{1}{\delta_{NT}} \right),
\end{aligned}$$

by Assumptions 1 to 6 (c) and Lemma A.1 (a). Therefore, a, b, c, d in the remainder term are all asymptotically negligible. ■

The proof of $\tilde{\Theta}_{m,i} \to \Theta_{m,i}$ for $m = 1, 2$ has been briefly illustrated in Bai (2003), and can be proved by applying a HAC estimator using $\tilde{f}_t \cdot \tilde{e}_{it}$. We therefore focus on the consistency of the pooled version $\tilde{\Theta}_m \to \Theta_m$. Before we present the main proofs for the W covariance matrices, we present some lemmas to be used.

Lemma A.12. *Under Assumptions 1 to 6, 8, 13, and 14, and if $\sqrt{T}/N \to 0$, then for $m = 1, 2$:*

a) $\frac{1}{T_m} \sum_{t=1}^{T} \left| \left(\tilde{e}_{(m),it} - e_{it} \right) \iota_{mt} \right|^4 = o_p(1)$ *for all i,*

b) If additionally, Assumption 15 holds, then $\frac{1}{T_m N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| \left(\tilde{e}_{(m),it} - e_{it} \right) \iota_{mt} \right|^4 = o_p(1)$*,*

c) $\frac{1}{T_m} \sum_{t=1}^{T} (\tilde{e}_{(m),it} \iota_{mt})^4 = O_p(1)$*, and*

d) If additionally, Assumption 15 holds, $\frac{1}{T_m N} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{e}_{(m),it}^4 \iota_{mt} = O_p(1)$*.*

Proof of Lemma A.12 (a). For brevity, we focus on $m = 1$, as the case for $m = 2$ is very similar. We have by definition:

$$\begin{aligned}
\left|\tilde{e}_{(1),it} - e_{it}\right| &= \left|\tilde{\lambda}_{1,i}^\top \tilde{f}_{1,t} - \lambda_{1,i}^\top f_t\right| \\
&= \left|\tilde{\lambda}_{1,i}^\top \tilde{f}_{1,t} - \lambda_{1,i}^\top H_1^{-\top} H_1^\top f_t\right| \\
&= \left|\tilde{\lambda}_{1,i}^\top (\tilde{f}_{1,t} - H_1^\top f_t) + \tilde{\lambda}_{1,i}^\top H_1^\top f_t - (\lambda_{1,i}^\top H_1^{-\top} H_1^\top f_t)\right| \\
&= \left|\tilde{\lambda}_{1,i}^\top (\tilde{f}_{1,t} - H_1^\top f_t) + (\tilde{\lambda}_{1,i}^\top - \lambda_{1,i}^\top H_1^{-\top}) H_1^\top f_t\right| \\
&= |E.I_t + E.II_t|.
\end{aligned}$$

Noting that $\left|\tilde{e}_{(1),it} - e_{it}\right|^4 \leq 64|E.I_t^4 + E.II_t^4|$, it therefore suffices to consider the bounds of $\frac{1}{T_m}\sum_{t=1}^{T} E.I_t^4 \iota_{mt}$ and $\frac{1}{T_m}\sum_{t=1}^{T} E.II_t^4 \iota_{mt}$.

First, $\frac{1}{T_1}\sum_{t=1}^{\lfloor \pi T \rfloor} E.I_t^4$ can be bounded by:

$$\begin{aligned}
\frac{1}{T_1}\sum_{t=1}^{\lfloor \pi T \rfloor} E.I_t^4 &\leq \left\|\tilde{\lambda}_{1,i}\right\|^4 \frac{1}{T_1}\sum_{t=1}^{\lfloor \pi T \rfloor} \left\|\tilde{f}_{1,t} - H_1^\top f_t\right\|^4 \\
&= O_p(1)\, O_p\left(\frac{1}{\delta_{NT}^4}\right),
\end{aligned}$$

where $\left\|\tilde{\lambda}_{1,i}\right\|^4 = O_p(1)$ because each $\tilde{\lambda}_{1,i}$ is bounded by normalisation, and $\frac{1}{T}\left\|\tilde{f}_{1,t} - H_1^\top f_t\right\|^4 = O_p\left(\frac{1}{\delta_{NT}^4}\right)$ by Lemma A.5 (a).

Next, $\frac{1}{T_1}\sum_{t=1}^{T} E.II_t^4$ can be bounded by:

$$\begin{aligned}
\frac{1}{T_1}\sum_{t=1}^{\lfloor \pi T \rfloor} E.II_t^4 &\leq \left\|\tilde{\lambda}_{1,i}^\top - \lambda_{1,i}^\top H_1^{-\top}\right\|^4 \|H_1\|^4 \frac{1}{T_1}\sum_{t=1}^{\lfloor \pi T \rfloor} f_t^4 \\
&= O_p\left(\frac{1}{\delta_{NT}^4}\right) O_p(1)\, O_p(1),
\end{aligned}$$

where $\left\|\tilde{\lambda}_i^\top - \lambda_i^\top H^{-\top}\right\|^4 = O_p\left(\frac{1}{\delta_{NT}^4}\right)$ by Lemma A.5 (d), and $\frac{1}{T_1}\sum_{t=1}^{\lfloor \pi T \rfloor} f_t^4 = O_p(1)$ by Assumption 1.

■

Proof of Lemma A.12 (b). The proof is similar to that of Lemma A.12 (a) - it suffices to show that $\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} E.I_t^4$ and $\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} E.II_t^4$ are both negligible. For brevity, we will focus on the case of $m = 1$, as the case for $m = 2$ is similar. First, $\frac{1}{T_1 N}\sum_{t=1}^{\lfloor \pi T \rfloor}\sum_{i=1}^{N} E.II_t^4$ can be bounded by

$$
\begin{aligned}
\frac{1}{T_1 N}\sum_{t=1}^{\lfloor \pi T \rfloor}\sum_{i=1}^{N} E.I_t^4 =& \frac{1}{T_1 N}\sum_{t=1}^{\lfloor \pi T \rfloor}\sum_{i=1}^{N} \tilde{\lambda}_{1,i}^{\top}(\tilde{f}_{1,t} - H_1^{\top} f_t) \\
=& \frac{1}{T_1 N}\sum_{i=1}^{N}\sum_{t=1}^{\lfloor \pi T \rfloor} \left[\left(\tilde{\lambda}_{1,i} - \lambda_{1,i} H_1^{-\top}\right)\left(\tilde{f}_{1,t} - H_1^{\top} f_t\right) + \lambda_{1,i} H_1^{\top}\left(\tilde{f}_{1,t} - H_1^{\top} f_t\right)\right]^4 \\
\leq& \frac{64}{T_1 N}\sum_{t=1}^{\lfloor \pi T \rfloor}\left(\tilde{f}_{1,t} - H_1^{\top} f_t\right)^4 \frac{1}{N}\left(\tilde{\lambda}_{1,i} - \lambda_i H_1^{-\top}\right)^4 + \frac{64}{T_1 N}\sum_{i=1}^{N}(\lambda_{1,i} H_1^{\top})^4 \sum_{t=1}^{\lfloor \pi T \rfloor}\left(\tilde{f}_{1,t} - H_1^{\top} f_t\right)^4 \\
\leq& \frac{64}{T_1}\sum_{t=1}^{\lfloor \pi T \rfloor}\left\|\tilde{f}_{1,t} - H_1^{\top} f_t\right\|^4 \frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_{1,i} - \lambda_{1,i} H_1^{-\top}\right\|^4 \\
&+ 64\frac{1}{T_1}\sum_{t=1}^{\lfloor \pi T \rfloor}\left\|\tilde{f}_{1,t} - H_1^{\top} f_t\right\|^4 \frac{1}{N}\sum_{i=1}^{N}\|\lambda_{1,i}\|^4\|H_1\|^4 \\
=& o_p(1) o_p(1) + o_p(1) O_p(1) \\
=& o_p(1),
\end{aligned}
$$

where the first $o_p(1)$ term comes from Lemmas A.5 (a) and A.5 (d), and the second term is $o_p(1)$ from Lemmas A.5 (a) and A.1 (e) and Assumption 2. The second term $\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N} E.II_t^4$ can be bounded by

$$
\begin{aligned}
\frac{1}{T_1 N}\sum_{t=1}^{\lfloor \pi T \rfloor}\sum_{i=1}^{N} E.II_t^4 &= \frac{1}{T_1 N}\sum_{t=1}^{\lfloor \pi T \rfloor}\sum_{i=1}^{N}(\tilde{\lambda}_{1,i}^{\top} - \lambda_{1,i}^{\top} H^{-\top}) H_1^{\top} f_t \\
&\leq \frac{1}{T_1 N}\sum_{t=1}^{\lfloor \pi T \rfloor}\sum_{i=1}^{N}\left\|\tilde{\lambda}_{1,i} - \lambda_{1,i} H_1^{-\top}\right\|^4 \|H_1\|^4 \|f_t\|^4 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_{1,i} - \lambda_{1,i} H_1^{-\top}\right\|^4 \frac{1}{T_1}\sum_{t=1}^{\lfloor \pi T \rfloor} f_t^4 \|H_1\|^4 \\
&= o_p(1) O_p(1) O_p(1) \\
&= o_p(1),
\end{aligned}
$$

because of Lemmas A.5 (d) and A.1 (e) and Assumption 1. ■

Proof of Lemmas A.12 (c) and A.12 (d). Lemmas A.12 (c) and A.12 (d) are implications and Lemmas A.12 (a) and A.12 (b) and can be proven in a similar way. Focusing on $m = 1$ as the case for

$m = 2$ is similar, we have

$$\begin{aligned}
\frac{1}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{e}_{it}^4 &\leq \frac{8}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} \left(\tilde{e}_{it} - e_{it}\right)^4 + \frac{8}{T_1} \sum_{t=1}^{\lfloor \pi T \rfloor} e_{it}^4 \\
&= O_p\left(1\right), \quad \text{and} \\
\frac{1}{T_1 N} \sum_{i=1}^{N} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{e}_{it}^4 &\leq \frac{8}{T_1 N} \sum_{i=1}^{N} \sum_{t=1}^{\lfloor \pi T \rfloor} \left(\tilde{e}_{it} - e_{it}\right)^4 + \frac{8}{T_1 N} \sum_{i=1}^{N} \sum_{t=1}^{\lfloor \pi T \rfloor} e_{it}^4 \\
&= O_p\left(1\right).
\end{aligned}$$

■

Lemma A.13. *Under Assumptions 1 to 8 and 13 to 15, for $m = 1, 2$,*

$$\left\| \frac{1}{N} \sum_{i=1}^{N} D_{j,m,i} - D_{j,m}^* \right\| = o_p(1),$$

where $D_{j,m}^* = \operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} D_{j,m,i}^* = \operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T_m} \sum_{t=j+1}^{T} H_m^\top f_t e_{it} e_{i,t-j}^\top f_{t-j}^\top H_m \iota_{mt}$, its infeasible counterpart.

Proof of Lemma A.13. We focus on the case of $m = 1$, as the proof for $m = 2$ is similar and thus omitted. We have

$$\begin{aligned}
&\left\| \frac{1}{N} \sum_{i=1}^{N} D_{j,1,i} - D_{j,1}^* \right\| \\
=&\left\| \frac{1}{T_1 N} \sum_{i=1}^{N} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \tilde{f}_{1,t} \tilde{f}_{1,t-j}^\top \tilde{e}_{(1),it} \tilde{e}_{(1),i,t-j} - H_1^\top \left(\frac{1}{T_1 N} \sum_{i=1}^{N} \sum_{t=j+1}^{\lfloor \pi T \rfloor} f_t f_{t-j}^\top e_{it} e_{i,t-j} \right) H_1 \right\| \\
\leq&\frac{1}{T_1 N} \sum_{i=1}^{N} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left(\tilde{f}_{1,t} \tilde{f}_{1,t-j}^\top - H_1^\top f_t f_{t-j} H_1 \right) \left(\tilde{e}_{it} \tilde{e}_{i,t-j} \right) \\
&\quad + \frac{1}{T_1 N} \sum_{i=1}^{N} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left(H_1^\top f_t f_{t-j} H_1 \right) \left(\tilde{e}_{(1),it} \tilde{e}_{(1),i,t-j} - e_{it} e_{i,t-j} \right) \\
=&D.I + D.II.
\end{aligned}$$

The first term $D.I$ is bounded by

$$
\begin{aligned}
&D.I\\
=&\frac{1}{T_1N}\sum_{i=1}^{N}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left(\tilde{f}_{1,t}\tilde{f}_{1,t-j}^{\top}-H_1^{\top}f_tf_{t-j}^{\top}H_1\right)\left(\tilde{e}_{(1),it}\tilde{e}_{(1),i,t-j}\right)\\
\leq&\left(\frac{1}{T_1}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left\|\tilde{f}_{1,t}\tilde{f}_{1,t-j}^{\top}-H_1^{\top}f_tf_{t-j}^{\top}H_1\right\|^2\frac{1}{T_1}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{(1),it}\tilde{e}_{(1),i,t-j}\right)^2\right)^{\frac{1}{2}}\\
\leq&\left(\frac{1}{T_1}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left\|\tilde{f}_{1,t}(\tilde{f}_{1,t-j}^{\top}-f_{t-j}^{\top}H_1)+(\tilde{f}_{1,t}-H_1^{\top}f_t)(\tilde{f}_{1,t-j}^{\top})\right\|^2\frac{1}{T_1}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{(1),it}\tilde{e}_{(1),i,t-j}\right)^2\right)^{\frac{1}{2}}\\
\leq&\left(\frac{2}{T_1}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left\|\tilde{f}_{1,t}(\tilde{f}_{1,t-j}^{\top}-f_{t-j}^{\top}H_1)\right\|^2+\left\|(\tilde{f}_{1,t}-H_1^{\top}f_t)(\tilde{f}_{1,t-j}^{\top})\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{T_1}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{(1),it}\tilde{e}_{(1),i,t-j}\right)\right)^{\frac{1}{2}}\\
\leq&\left(2\left(\frac{1}{T_1}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left\|\tilde{f}_{1,t}\right\|^4\frac{1}{T_1}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left\|\tilde{f}_{1,t-j}^{\top}-f_{t-j}^{\top}H_1\right\|^4\right)^{\frac{1}{2}}+2\left(\frac{1}{T_1}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left\|\tilde{f}_{1,t-j}\right\|^4+\frac{1}{T_1}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\left\|\tilde{f}_{1,t}-H_1^{\top}f_t\right\|^4\right)^{\frac{1}{2}}\right)^{\frac{1}{2}}\\
&\quad\times\left(\frac{1}{TN}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\sum_{i=1}^{N}\tilde{e}_{(1),it}^2\tilde{e}_{i,t-j}^2\right)^{1/2}\\
\leq&o_p(1)\left(\frac{1}{TN}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\sum_{i=1}^{N}\tilde{e}_{(1),it}^2\tilde{e}_{(1),i,t-j}^2\right)^{1/2}\\
\leq&o_p(1)\left[\left(\frac{1}{TN}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\sum_{i=1}^{N}\tilde{e}_{(1),it}^4\right)^{\frac{1}{2}}\left(\frac{1}{TN}\sum_{t=j+1}^{\lfloor\pi T\rfloor}\sum_{i=1}^{N}\tilde{e}_{(1),i,t-j}^4\right)^{\frac{1}{2}}\right]^{1/2}\\
=&o_p(1)O_p\left(1\right),
\end{aligned}
$$

where the first term is $o_p(1)$ follows from applying Lemmas A.5 (a) and A.5 (c), and the second term follows from Lemma A.12 (d).

The second term $D.II$ is bounded by

$$
\begin{aligned}
D.II =& \frac{1}{T_1 N}\sum_{i=1}^{N}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left(H_1^\top f_t f_{t-j}^\top H_1\right)\left(\tilde{e}_{(1),it}\tilde{e}_{(1),i,t-j} - e_{it}e_{i,t-j}\right)\\
\leq & \left(\frac{1}{T_1}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left\|H_1^\top f_t f_{t-j}^\top H_1\right\|^2 \frac{1}{T_1}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\tilde{e}_{(1),it}\tilde{e}_{(1),i,t-j} - e_{it}e_{i,t-j}\right)\right)^2\right)^{\frac{1}{2}}\\
\leq & O_p(1)\left(\frac{2}{T_1}\sum_{t=j+1}^{\lfloor \pi T\rfloor}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{(1),it}^2(\tilde{e}_{(1),i,t-j} - e_{i,t-j}) + e_{i,t-j}^2(\tilde{e}_{(1),it} - e_{it})^2\right)\right)^{\frac{1}{2}}\\
\leq & O_p(1)\left(\frac{2}{T_1 N}\left(\sum_{t=j+1}^{\lfloor \pi T\rfloor}\sum_{i=1}^{N}\tilde{e}_{(1),it}^4 \sum_{t=1}^{\lfloor \pi T\rfloor}\sum_{i=1}^{N}(\tilde{e}_{(1),i,t-j} - e_{t,i-j})^4\right)^{\frac{1}{2}}\right.\\
&\left. + \frac{2}{T_1 N}\left(\sum_{t=j+1}^{\lfloor \pi T\rfloor}\sum_{i=1}^{N}e_{i,t-j}^4 \sum_{t=j+1}^{\lfloor \pi T\rfloor}\sum_{i=1}^{N}(\tilde{e}_{(1),it} - e_{it})^4\right)^{\frac{1}{2}}\right)\\
=& o_p(1),
\end{aligned}
$$

where the $O_p(1)$ term comes from Assumption 1 and Lemma A.1 (e), and the $o_p(1)$ term comes from Lemma A.12 (b). ■

Proof of Lemma A.11 (c). It suffices to prove that $\left\|\frac{1}{N}\sum_{i=1}^{N}\tilde{\Theta}_{m,i} - \Theta_m\right\| = o_p(1)$ for $m = 1, 2$. For brevity, we will only prove the case for $m = 1$, as the case for $m = 2$ is similar. See that

$$
\begin{aligned}
&\left\|\frac{1}{N}\sum_{i=1}^{N}\tilde{\Theta}_{m,i} - \Theta_m\right\|\\
&\leq \left\|\frac{1}{N}\sum_{i=1}^{N}D_{0,1,i} + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{\lfloor \pi T-1\rfloor}\mathbf{k}\left(\frac{j}{b_{\lfloor \pi T\rfloor}}\right)D_{j,ji} - D_{0,1}^* - \sum_{j=1}^{\lfloor \pi T-1\rfloor}\mathbf{k}\left(\frac{j}{b_{\lfloor \pi T\rfloor}}\right)D_{ji}^*\right\|\\
&\leq \left\|\frac{1}{N}\sum_{i=1}^{N}D_{0,1,i} - D_{0,1}^*\right\| + 2\sum_{j=1}^{b_{\lfloor \pi T\rfloor}}\left\|\frac{1}{N}\sum_{i=1}^{N}D_{1,ji} - D_{ji}^*\right\|\\
&= o_p(1),
\end{aligned}
$$

by Lemma A.13. ■

Proof of Theorem 2.4 (a). We are now considering the asymptotic expansion of $\tilde{\lambda}_{m,i}$ in Lemma A.10, but averaged across the cross-section and inflated by $\sqrt{N}$:

$$\frac{\sum_{i=1}^{N}\tilde{\lambda}_{m,i}-H_1^{-1}\lambda_{m,i}}{\sqrt{N}}=\frac{1}{T_m\sqrt{N}}\sum_{i=1}^{N}\tilde{F}_m^{\top}(F_m-F_mH_m)\lambda_{m,i}+\frac{1}{T_m\sqrt{N}}\sum_{i=1}^{N}\tilde{F}_m-F_mH_m{}^{\top}e_{(m),i}, \tag{A.1.18}$$

where the last two terms are asymptotically negligible because of Lemmas A.11 (a) and A.11 (b).

Similarly, considering the asymptotic expansion of $\tilde{w}_i$ in Equation (A.1.15), taking its cross-sectional mean, and then inflating by $\sqrt{TN}$ on both sides, we have:

$$\begin{aligned}\sqrt{TN}\frac{\sum_{i=1}^{N}(\tilde{w}_i-H_2^{-1}w_i)}{N}=&H_2^{\top}\frac{1}{(1-\pi)\sqrt{TN}}\sum_{i=1}^{N}\sum_{t=\lfloor\pi T+1\rfloor}^{T}f_te_{it}\\&-\tilde{Z}^{\top}\frac{1}{\pi\sqrt{TN}}\sum_{i=1}^{N}\sum_{t=1}^{\lfloor\pi T\rfloor}H_1^{\top}f_te_{it}+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)\\&\xrightarrow{d}N(0,\Omega_W),\end{aligned}$$

where $\Omega_W=\frac{1}{(1-\pi)TN}H_{0,2}^{\top}\sum_{i=1}^{N}\Phi_{i,2}H_{0,2}+\frac{1}{\pi TN}H_{0,2}^{-1}Z'\Sigma_F^{-1}\sum_{i=1}^{N}\Phi_{i,1}\Sigma_F^{-1}ZH_{0,2}^{-\top}$, and the remainder terms are asymptotically negligible by Lemmas A.11 (a) and A.11 (b). The asymptotic distribution then follows by Assumption 16, the convergence of H_1, H_2, and $\tilde{Z}$ to their probability limits, and the consistency of $\tilde{\Omega}_W$ for Ω_W. ■

Proof of Theorem 2.5. We will show that $\mathscr{W}_{W,i}\to\infty$ as $N,T\to\infty$ when $w_i\neq 0$. Recalling the asymptotic expansion of $\tilde{w}_i$, we have:

$$\begin{aligned}\tilde{w}_i=&(H_2^{-1}w_i)+H_2^{-1}\frac{1}{(1-\pi)T}\sum_{t=\lfloor\pi T+1\rfloor}^{T}f_te_{it}-\tilde{Z}^{\top}\frac{1}{\pi T}\sum_{t=1}^{\lfloor\pi T\rfloor}H_1^{\top}f_te_{it}+O_p\left(\frac{1}{\delta_{NT}^2}\right)\\=&H_2^{-1}w_i+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\delta_{NT}^2}\right),\\=&H_2^{-1}w_i+o_p(1).\end{aligned}$$

Because $\tilde{\Omega}_{W,i}\xrightarrow{p}\Omega_{W,i}$ and $\Omega_{W,i}$ is positive definite, it follows that $\tilde{\Omega}_{W,i}^{-1}=O_p(1)$ and we have the desired result:

$$\text{plim}_{N,T\to\infty}\inf\left(W_{W_i}\right) = \text{plim}_{N,T\to\infty}\inf\left(\tilde{w}_i^\top \Omega_{W,i}^{-1}\tilde{w}_i\right) > 0$$

which implies the desired divergence under the alternative hypothesis. We will show that $\mathscr{W}_W \to \infty$ as $N, T \to \infty$ under the alternative. We have:

$$\begin{aligned}
&(T^{\alpha/2})\left(\frac{\sum_{i=1}^N \tilde{w}_i}{\sqrt{N}}\right) \\
=&T^{\alpha/2}\frac{\sum_{i=1}^N (H_2^{-1}w_i)}{\sqrt{N}} + H_2^\top \frac{T^{\alpha/2}}{(1-\pi)T\sqrt{N}}\sum_{i=1}^N\sum_{s=\lfloor\pi T+1\rfloor}^T f_s e_{is} - \tilde{Z}^\top \frac{T^{\alpha/2}}{\pi T\sqrt{N}}\sum_{i=1}^N\sum_{s=1}^{\lfloor\pi T\rfloor} H_1^\top f_s e_{is} + O_p\left(\frac{T^{\alpha/2}}{\delta_{NT}^2}\right) \\
=&H_2^\top T^{\alpha/2}\frac{\sum_{i=1}^N (H_2^{-1}w_i)}{\sqrt{N}} + H_2^\top \frac{1}{(1-\pi)T^{1-\alpha/2}}O_p(1) - \tilde{Z}^\top H_1^\top \frac{1}{\pi T^{1-\alpha/2}}O_p(1) + O_p\left(\frac{T^{\alpha/2}}{\delta_{NT}^2}\right) \\
=&H_2^\top T^{\alpha/2}\frac{\sum_{i=1}^N (H_2^{-1}w_i)}{\sqrt{N}} + O_p\left(\frac{1}{T^{1-\alpha/2}}\right) + O_p\left(\frac{T^{\alpha/2}}{\delta_{NT}^2}\right),
\end{aligned}$$

where the last two terms are $o_p(1)$ because $0 < \alpha \leq 0.5$ and $\sqrt{T}/N \to 0$ as $N, T \to \infty$. Since $\tilde{\Omega}_W \xrightarrow{p} \Omega_W$ and Ω_W is positive definite, it follows that $\tilde{\Omega}_W^{-1} = O_p(1)$ and we have the desired result by Assumption 17:

$$\begin{aligned}
&\text{plim}_{N,T\to\infty}\inf\left(\frac{T^\alpha}{T}\mathscr{W}_W\right) \\
=&\text{plim}_{N,T\to\infty}\inf\left[(T^\alpha)\left(\frac{\sum_{i=1}^N \tilde{w}_i}{\sqrt{N}}\right)^\top (\tilde{\Omega}_W)^{-1}\left(\frac{\sum_{i=1}^N \tilde{w}_i}{\sqrt{N}}\right)\right] > 0,
\end{aligned}$$

which implies the desired divergence under the alternative hypothesis. ■

A.1.5 Singular Z

The case of a singular Z can be further classified into two cases, depending on the column rank of W. In this section, we show that with some suitable adjustments, our test statistics can accommodate these cases.

Replacement of factors: $rank(Z) < rank(W)$

If the column rank of is still r, this represents the case where some of the original factors are "replaced" by an entirely new set of factors. In this case, $\Lambda_2 = \Lambda_1 Z + W$ is still of full rank, and the existing theory can still go through.

Changing number of factors: $rank(Z) = rank(W)$

If Z is singular and has identical (column) rank to W, then this represents the case of a disappearing factor. Note that the case of an emerging factor can always be parameterised in by reversing the pre- and post-break samples, and thus our method can be extended to accommodate a changing number of factors.

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 this 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:

$$\begin{aligned}\Lambda_2 &= (\Lambda_1 + W)\begin{bmatrix} I_{r-r_2} \\ 0 \end{bmatrix} \\ &= \Lambda_1 Z_0 + W_0. \end{aligned} \tag{A.1.19}$$

This allows us to write

$$\begin{aligned} X_2 &= F_2 \Lambda_2^\top + e_{(2)} \\ &= F_2 \begin{bmatrix} I_{r-r_2} \\ 0 \end{bmatrix} \left((\Lambda_1 Z + W) \begin{bmatrix} I_{r-r_2} \\ 0 \end{bmatrix} \right)^\top + e_{(2)} \\ &= F_{2,r-r_2} \left(\Lambda_1 Z_0 + W_0\right)^\top + e_{(2)}, \end{aligned} \tag{A.1.20}$$

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

$$H_{2,r-r_2} = \frac{\Lambda_2^\top \Lambda_2}{N} \frac{F_{2,r-r_2}^\top \tilde{F}_2}{T_2} V_{NT,2,r-r_2}^{-1} \tag{A.1.21}$$

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 A.14. *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) $\frac{1}{T}\left(\tilde{F}_{2,r-r_2} - F_{2,r-r_2}H_{2,r-r_2}\right)^\top F_{2,r-r_2} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$, *and*

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 A.14. These correspond to Theorem of Bai and Ng (2002) and Lemmas B.1 and B.2 of Bai (2003). ■

Lemma A.14 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 A.15. *Under Assumptions 1 to 8, as* $N, T \to \infty$,

$$\tilde{Z} = H_1^\top Z_0 H_{2,r-r_2}^{-\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right).$$

Proof of Lemma A.15. Lemma A.15 is analogous to Theorem 2.1 (a) and can be proved in a similar way. ■

LM-like Z-test

In either cases of a singular Z, the variance of the Z-test can also be estimated with an LM-like estimator that uses the whole data and hence leads to more numerical stability. This is because the variance estimate using the post-break data, i.e. the outer product of $\tilde{F}_2\tilde{Z}^\top$ is singular and therefore a standard HAC estimator applied to this process will fail. Define

$$\widehat{\Omega}_Z(\widehat{F}) = \widehat{\Gamma}(FH_{1,0}) + \sum_{j=1}^{T} \mathbf{k}\left(\frac{j}{S_T}\right)\left(\widehat{\Gamma}_j\left(FH_{1,0}\right) + \widehat{\Gamma}_j\left(FH_{1,0}\right)^{\top}\right),$$

$$\widehat{\Gamma}_j(FH_{1,0}) = \frac{1}{T}\sum_{t=j+1}^{T} vech\left(H_{1,0}^{\top} f_t f_t^{\top} H_{1,0} - I_r\right) vech\left(H_{1,0}^{\top} f_{j-1} f_{j-1}^{\top} H_{1,0} - I_r\right)^{\top},$$

which can be proven to be consistent for their respective infeasible counterparts $\widehat{\Omega}_Z(FH_{1,0})$ and $\widehat{\Gamma}(FH_{1,0})$ in a similar way to Lemmas A.7, A.8, and A.6 (b). These lead to the infeasible estimator using all of the data

$$\tilde{S}_Z(\pi, \widehat{F}) = \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\widehat{\Omega}_Z(\widehat{F}),$$

which can also be proven to be consistent for its infeasible counterpart in a similar way to Lemma A.6 (b). Together with $A_Z(\pi, \widehat{F})$, these can be used to define the LM-like test statistic

$$LM_Z(\pi, \widehat{F}) = A_Z(\pi, \widehat{F})^{\top} \tilde{S}_Z(\pi, \widehat{F})^{-1} A_Z(\pi, \widehat{F}). \tag{A.1.22}$$

Its consistency to its infeasible counterpart and power under the alternative hypothesis can be proven in a similar way to Theorems 2.2 and 2.3.

A.2 Additional Tables

A.2.1 Additional Simulation Results

Table A.1: Size of Rotation and Orthogonal Shift Tests, $r = 3$

						Z Test		W Test		W Individual
T	N	τ	ρ	α	β	Unadjusted	Adjusted	Unadjusted	Adjusted	

200	200	0.3	0.0	0.0	0.0	0.368	0.257	0.001	0.000	0.009
200	200				0.3	0.374	0.280	0.038	0.029	0.010
200	200			0.3	0.0	0.491	0.371	0.004	0.001	0.012
200	200				0.3	0.504	0.397	0.054	0.042	0.014
200	200		0.7	0.0	0.0	0.204	0.147	0.001	0.000	0.007
200	200				0.3	0.206	0.145	0.035	0.020	0.008
200	200			0.3	0.0	0.221	0.156	0.006	0.001	0.021
200	200				0.3	0.225	0.167	0.069	0.047	0.021
200	200	0.5	0.0	0.0	0.0	0.286	0.204	0.004	0.003	0.013
200	200				0.3	0.292	0.215	0.082	0.060	0.014
200	200			0.3	0.0	0.345	0.256	0.004	0.003	0.015
200	200				0.3	0.352	0.281	0.101	0.067	0.017
200	200		0.7	0.0	0.0	0.219	0.144	0.002	0.002	0.012
200	200				0.3	0.221	0.152	0.059	0.039	0.013
200	200			0.3	0.0	0.241	0.167	0.009	0.006	0.026
200	200				0.3	0.238	0.176	0.098	0.072	0.027
200	200	0.7	0.0	0.0	0.0	0.353	0.275	0.000	0.000	0.005
200	200				0.3	0.365	0.282	0.024	0.018	0.006
200	200			0.3	0.0	0.410	0.313	0.000	0.000	0.007
200	200				0.3	0.417	0.335	0.036	0.028	0.007
200	200		0.7	0.0	0.0	0.377	0.289	0.000	0.000	0.005
200	200				0.3	0.385	0.297	0.031	0.020	0.005
200	200			0.3	0.0	0.393	0.318	0.005	0.003	0.014
200	200				0.3	0.398	0.330	0.072	0.051	0.015
500	200	0.3	0.0	0.0	0.0	0.182	0.110	0.001	0.001	0.002
500	200				0.3	0.184	0.112	0.031	0.015	0.003
500	200			0.3	0.0	0.209	0.136	0.001	0.000	0.003
500	200				0.3	0.224	0.135	0.031	0.020	0.003
500	200		0.7	0.0	0.0	0.160	0.101	0.000	0.000	0.002
500	200				0.3	0.158	0.105	0.027	0.018	0.002
500	200			0.3	0.0	0.151	0.093	0.002	0.001	0.005
500	200				0.3	0.148	0.097	0.039	0.029	0.006
500	200	0.5	0.0	0.0	0.0	0.125	0.079	0.007	0.004	0.007
500	200				0.3	0.129	0.085	0.082	0.051	0.007
500	200			0.3	0.0	0.143	0.076	0.007	0.001	0.007
500	200				0.3	0.148	0.098	0.095	0.059	0.008
500	200		0.7	0.0	0.0	0.156	0.080	0.005	0.001	0.006
500	200				0.3	0.155	0.086	0.075	0.052	0.006
500	200			0.3	0.0	0.155	0.088	0.011	0.005	0.010
500	200				0.3	0.159	0.094	0.096	0.068	0.011

500	200	0.7	0.0	0.0	0.0	0.198	0.146	0.000	0.000	0.002
500	200				0.3	0.200	0.150	0.026	0.013	0.002
500	200			0.3	0.0	0.204	0.145	0.000	0.000	0.002
500	200				0.3	0.221	0.156	0.034	0.015	0.002
500	200		0.7	0.0	0.0	0.220	0.148	0.000	0.000	0.002
500	200				0.3	0.220	0.149	0.020	0.012	0.002
500	200			0.3	0.0	0.231	0.151	0.002	0.001	0.004
500	200				0.3	0.232	0.157	0.034	0.020	0.004
1000	200	0.3	0.0	0.0	0.0	0.096	0.061	0.001	0.001	0.001
1000	200				0.3	0.096	0.063	0.027	0.016	0.001
1000	200			0.3	0.0	0.108	0.070	0.000	0.000	0.001
1000	200				0.3	0.118	0.070	0.031	0.013	0.001
1000	200		0.7	0.0	0.0	0.108	0.060	0.000	0.000	0.001
1000	200				0.3	0.104	0.059	0.016	0.005	0.001
1000	200			0.3	0.0	0.091	0.049	0.000	0.000	0.002
1000	200				0.3	0.093	0.054	0.024	0.011	0.002
1000	200	0.5	0.0	0.0	0.0	0.080	0.045	0.005	0.002	0.005
1000	200				0.3	0.081	0.049	0.075	0.047	0.005
1000	200			0.3	0.0	0.086	0.050	0.005	0.002	0.005
1000	200				0.3	0.084	0.055	0.086	0.044	0.005
1000	200		0.7	0.0	0.0	0.091	0.054	0.004	0.002	0.004
1000	200				0.3	0.093	0.055	0.061	0.035	0.004
1000	200			0.3	0.0	0.092	0.053	0.003	0.002	0.006
1000	200				0.3	0.095	0.059	0.080	0.048	0.007
1000	200	0.7	0.0	0.0	0.0	0.122	0.083	0.000	0.000	0.001
1000	200				0.3	0.124	0.087	0.022	0.011	0.001
1000	200			0.3	0.0	0.138	0.091	0.001	0.000	0.001
1000	200				0.3	0.145	0.094	0.031	0.015	0.001
1000	200		0.7	0.0	0.0	0.162	0.121	0.000	0.000	0.001
1000	200				0.3	0.162	0.117	0.015	0.010	0.001
1000	200			0.3	0.0	0.166	0.122	0.000	0.000	0.001
1000	200				0.3	0.165	0.118	0.027	0.009	0.002

Note:

Entries denote the rejection rates for a nominal size of 5%. HI denotes Han and Inoue (2015)'s test, and BKW denotes Baltagi et al (2021)'s test conducted with a pre-known break fraction. The scalar ω denotes the "size" of the break in the loadings. See Table 2.1 for explanation of α, β, and ρ.

Table A.2: Power of Z Rotation Test, $r = 3$

						Z Test		W Test			
T	N	τ	ρ	α	β	Unadjusted	Adjusted	Unadjusted	Adjusted	HI Test	BKW Test
200	100	0.3	0.0	0.0	0.0	1.000	1.000	0.000	0.000	1.000	0.914
200	100				0.3	1.000	1.000	0.038	0.038	1.000	0.913
200	100			0.3	0.0	1.000	0.999	0.001	0.001	1.000	0.915
200	100				0.3	1.000	1.000	0.036	0.036	1.000	0.915
200	100		0.7	0.0	0.0	0.982	0.958	0.000	0.000	0.994	0.739
200	100				0.3	0.982	0.956	0.023	0.023	0.994	0.737
200	100			0.3	0.0	0.978	0.950	0.002	0.002	0.994	0.736
200	100				0.3	0.980	0.954	0.053	0.053	0.995	0.734
200	100	0.5	0.0	0.0	0.0	1.000	1.000	0.003	0.003	1.000	1.000
200	100				0.3	1.000	1.000	0.144	0.144	1.000	1.000
200	100			0.3	0.0	1.000	1.000	0.011	0.011	1.000	1.000
200	100				0.3	1.000	1.000	0.162	0.162	1.000	1.000
200	100		0.7	0.0	0.0	1.000	1.000	0.003	0.003	1.000	0.999
200	100				0.3	1.000	1.000	0.088	0.088	1.000	1.000
200	100			0.3	0.0	1.000	1.000	0.007	0.007	1.000	0.999
200	100				0.3	1.000	1.000	0.146	0.146	1.000	0.999
200	100	0.7	0.0	0.0	0.0	0.998	0.994	0.003	0.003	1.000	0.935
200	100				0.3	0.998	0.993	0.054	0.054	0.999	0.934
200	100			0.3	0.0	0.999	0.994	0.004	0.004	0.998	0.936
200	100				0.3	0.999	0.993	0.092	0.092	0.999	0.937
200	100		0.7	0.0	0.0	0.970	0.927	0.002	0.002	0.979	0.794
200	100				0.3	0.971	0.928	0.036	0.036	0.981	0.795
200	100			0.3	0.0	0.968	0.927	0.004	0.004	0.975	0.791
200	100				0.3	0.972	0.932	0.074	0.073	0.977	0.792
200	200	0.3	0.0	0.0	0.0	1.000	1.000	0.002	0.002	1.000	0.926
200	200				0.3	1.000	1.000	0.026	0.026	1.000	0.922
200	200			0.3	0.0	1.000	1.000	0.000	0.000	1.000	0.926
200	200				0.3	1.000	1.000	0.036	0.036	1.000	0.921
200	200		0.7	0.0	0.0	0.982	0.953	0.001	0.001	0.995	0.760
200	200				0.3	0.981	0.953	0.026	0.026	0.995	0.760
200	200			0.3	0.0	0.979	0.947	0.004	0.004	0.995	0.753
200	200				0.3	0.974	0.952	0.049	0.049	0.996	0.764

200	200	0.5	0.0	0.0	0.0	1.000	1.000	0.005	0.005	1.000	1.000
200	200				0.3	1.000	1.000	0.086	0.086	1.000	1.000
200	200			0.3	0.0	1.000	1.000	0.006	0.006	1.000	1.000
200	200				0.3	1.000	1.000	0.101	0.101	1.000	1.000
200	200		0.7	0.0	0.0	1.000	1.000	0.002	0.002	1.000	1.000
200	200				0.3	1.000	1.000	0.069	0.069	1.000	1.000
200	200			0.3	0.0	1.000	1.000	0.008	0.008	1.000	1.000
200	200				0.3	1.000	1.000	0.103	0.103	1.000	1.000
200	200	0.7	0.0	0.0	0.0	1.000	0.999	0.000	0.000	1.000	0.928
200	200				0.3	1.000	0.999	0.041	0.041	1.000	0.931
200	200			0.3	0.0	1.000	0.999	0.001	0.001	1.000	0.932
200	200				0.3	1.000	0.999	0.048	0.048	1.000	0.930
200	200		0.7	0.0	0.0	0.961	0.933	0.000	0.000	0.970	0.799
200	200				0.3	0.961	0.931	0.035	0.034	0.971	0.801
200	200			0.3	0.0	0.958	0.926	0.003	0.003	0.970	0.798
200	200				0.3	0.962	0.928	0.055	0.055	0.970	0.800
500	100	0.3	0.0	0.0	0.0	1.000	1.000	0.001	0.001	1.000	1.000
500	100				0.3	1.000	1.000	0.030	0.030	1.000	1.000
500	100			0.3	0.0	1.000	1.000	0.001	0.001	1.000	1.000
500	100				0.3	1.000	1.000	0.034	0.034	1.000	1.000
500	100		0.7	0.0	0.0	1.000	1.000	0.000	0.000	1.000	0.998
500	100				0.3	1.000	1.000	0.016	0.016	1.000	0.998
500	100			0.3	0.0	1.000	1.000	0.001	0.001	1.000	0.998
500	100				0.3	1.000	1.000	0.027	0.027	1.000	0.998
500	100	0.5	0.0	0.0	0.0	1.000	1.000	0.008	0.008	1.000	1.000
500	100				0.3	1.000	1.000	0.153	0.153	1.000	1.000
500	100			0.3	0.0	1.000	1.000	0.009	0.009	1.000	1.000
500	100				0.3	1.000	1.000	0.175	0.175	1.000	1.000
500	100		0.7	0.0	0.0	1.000	1.000	0.004	0.004	1.000	1.000
500	100				0.3	1.000	1.000	0.083	0.083	1.000	1.000
500	100			0.3	0.0	1.000	1.000	0.012	0.012	1.000	1.000
500	100				0.3	1.000	1.000	0.114	0.114	1.000	1.000
500	100	0.7	0.0	0.0	0.0	1.000	1.000	0.000	0.000	1.000	1.000
500	100				0.3	1.000	1.000	0.075	0.075	1.000	1.000
500	100			0.3	0.0	1.000	1.000	0.001	0.001	1.000	1.000
500	100				0.3	1.000	1.000	0.093	0.093	1.000	1.000
500	100		0.7	0.0	0.0	0.999	0.998	0.002	0.002	1.000	0.995
500	100				0.3	0.999	0.999	0.042	0.042	1.000	0.995
500	100			0.3	0.0	0.999	0.998	0.001	0.001	1.000	0.993
500	100				0.3	0.999	0.999	0.064	0.064	1.000	0.993

500	200	0.3	0.0	0.0	0.0	1.000	1.000	0.001	0.001	1.000	1.000
500	200				0.3	1.000	1.000	0.019	0.019	1.000	1.000
500	200			0.3	0.0	1.000	1.000	0.000	0.000	1.000	1.000
500	200				0.3	1.000	1.000	0.018	0.018	1.000	1.000
500	200		0.7	0.0	0.0	1.000	1.000	0.000	0.000	1.000	1.000
500	200				0.3	1.000	1.000	0.013	0.013	1.000	1.000
500	200			0.3	0.0	1.000	1.000	0.001	0.001	1.000	1.000
500	200				0.3	1.000	1.000	0.025	0.025	1.000	1.000
500	200	0.5	0.0	0.0	0.0	1.000	1.000	0.005	0.005	1.000	1.000
500	200				0.3	1.000	1.000	0.075	0.075	1.000	1.000
500	200			0.3	0.0	1.000	1.000	0.003	0.003	1.000	1.000
500	200				0.3	1.000	1.000	0.083	0.083	1.000	1.000
500	200		0.7	0.0	0.0	1.000	1.000	0.001	0.001	1.000	1.000
500	200				0.3	1.000	1.000	0.067	0.067	1.000	1.000
500	200			0.3	0.0	1.000	1.000	0.003	0.003	1.000	1.000
500	200				0.3	1.000	1.000	0.090	0.090	1.000	1.000
500	200	0.7	0.0	0.0	0.0	1.000	1.000	0.000	0.000	1.000	1.000
500	200				0.3	1.000	1.000	0.028	0.028	1.000	1.000
500	200			0.3	0.0	1.000	1.000	0.000	0.000	1.000	1.000
500	200				0.3	1.000	1.000	0.032	0.032	1.000	1.000
500	200		0.7	0.0	0.0	1.000	0.999	0.000	0.000	1.000	0.996
500	200				0.3	1.000	0.999	0.030	0.030	1.000	0.996
500	200			0.3	0.0	1.000	0.999	0.000	0.000	1.000	0.995
500	200				0.3	1.000	0.999	0.044	0.044	1.000	0.995

Note:

Entries denote the rejection rates across different simulated break types; a break type of $W \neq 0$ denotes a break in the factor loadings, $Z \neq I$ a break in the factor variance, and $W \neq 0$ and $Z \neq I$ denoting a break in both. HI denotes Han and Inoue (2015)'s test, and BKW denotes Baltagi et al (2021)'s test conducted with a pre-known break fraction. The scalar ω denotes the "size" of the break in the loadings. See Table 2.1 for explanation of α, β and ρ.

Table A.3: Power of Z and W Tests, $r = 3$, $\omega = 1$

						Z Test		W Test					
T	N	τ	ρ	α	β	Unadj.	Adj.	Unadj.	Adj.	Individual	HI	BKW	$\tilde{r}$

200	100			0.0	0.0	1.000	1.000	0.770	0.770	0.796	1.000	1.000	4.069
200	100		0.0		0.3	1.000	1.000	0.814	0.814	0.796	1.000	1.000	4.088
200	100			0.3	0.0	1.000	1.000	0.806	0.806	0.776	1.000	1.000	4.026
200	100				0.3	1.000	1.000	0.793	0.793	0.779	1.000	1.000	4.030
200	100			0.0	0.0	1.000	1.000	0.868	0.868	0.901	1.000	1.000	4.857
200	100		0.7		0.3	1.000	1.000	0.916	0.916	0.905	1.000	1.000	4.872
200	100			0.3	0.0	1.000	1.000	0.830	0.830	0.856	1.000	1.000	4.827
200	100				0.3	1.000	1.000	0.839	0.839	0.859	1.000	1.000	4.846
200	200			0.0	0.0	1.000	1.000	0.817	0.817	0.792	1.000	1.000	4.310
200	200		0.0		0.3	1.000	1.000	0.801	0.801	0.789	1.000	1.000	4.318
200	200	0.3		0.3	0.0	1.000	1.000	0.787	0.787	0.769	1.000	1.000	4.188
200	200				0.3	1.000	1.000	0.815	0.815	0.769	1.000	1.000	4.206
200	200			0.0	0.0	1.000	1.000	0.879	0.879	0.898	1.000	1.000	5.046
200	200		0.7		0.3	1.000	1.000	0.915	0.915	0.899	1.000	1.000	5.053
200	200			0.3	0.0	1.000	1.000	0.870	0.870	0.850	1.000	1.000	5.039
200	200				0.3	1.000	1.000	0.860	0.860	0.851	1.000	1.000	5.047
500	100			0.0	0.0	1.000	1.000	0.896	0.896	0.920	1.000	1.000	4.330
500	100		0.0		0.3	1.000	1.000	0.931	0.931	0.921	1.000	1.000	4.365
500	100			0.3	0.0	1.000	1.000	0.889	0.889	0.910	1.000	1.000	4.168
500	100				0.3	1.000	1.000	0.892	0.892	0.912	1.000	1.000	4.191
500	100			0.0	0.0	1.000	1.000	0.965	0.965	0.968	1.000	1.000	5.049
500	100		0.7		0.3	1.000	1.000	0.985	0.985	0.968	1.000	1.000	5.065
500	100			0.3	0.0	1.000	1.000	0.960	0.960	0.947	1.000	1.000	4.999
500	100				0.3	1.000	1.000	0.929	0.929	0.946	1.000	1.000	5.012
500	200			0.0	0.0	1.000	1.000	0.916	0.916	0.917	1.000	1.000	4.858
500	200		0.0		0.3	1.000	1.000	0.935	0.935	0.916	1.000	1.000	4.863
500	200			0.3	0.0	1.000	1.000	0.911	0.911	0.904	1.000	1.000	4.769
500	200				0.3	1.000	1.000	0.918	0.918	0.907	1.000	1.000	4.772
500	200			0.0	0.0	1.000	1.000	0.959	0.959	0.967	1.000	1.000	5.568
500	200		0.7		0.3	1.000	1.000	0.942	0.942	0.967	1.000	1.000	5.583
500	200			0.3	0.0	1.000	1.000	0.935	0.935	0.944	1.000	1.000	5.491
500	200				0.3	1.000	1.000	0.954	0.954	0.944	1.000	1.000	5.511

200	100			0.0	0.0	1.000	1.000	0.784	0.784	0.792	1.000	1.000	4.822
200	100		0.0		0.3	1.000	1.000	0.803	0.803	0.792	1.000	1.000	4.360
200	100			0.3	0.0	1.000	1.000	0.764	0.764	0.771	1.000	1.000	4.990
200	100				0.3	1.000	1.000	0.785	0.785	0.773	1.000	1.000	4.140
200	100			0.0	0.0	1.000	1.000	0.882	0.882	0.898	1.000	1.000	4.538
200	100		0.7		0.3	1.000	1.000	0.893	0.893	0.898	1.000	1.000	4.205
200	100			0.3	0.0	1.000	1.000	0.834	0.834	0.852	1.000	1.000	4.120
200	100				0.3	1.000	1.000	0.840	0.840	0.853	1.000	1.000	4.085
200	200			0.0	0.0	1.000	1.000	0.801	0.800	0.785	1.000	1.000	4.935
200	200		0.0		0.3	1.000	1.000	0.820	0.820	0.785	1.000	1.000	4.370
200	200	0.5		0.3	0.0	1.000	1.000	0.773	0.772	0.764	1.000	1.000	4.047
200	200				0.3	1.000	1.000	0.804	0.803	0.765	1.000	1.000	4.987
200	200			0.0	0.0	1.000	1.000	0.899	0.899	0.894	1.000	1.000	4.006
200	200		0.7		0.3	1.000	1.000	0.909	0.909	0.893	1.000	1.000	4.046
200	200			0.3	0.0	1.000	1.000	0.860	0.860	0.846	1.000	1.000	4.024
200	200				0.3	1.000	1.000	0.867	0.867	0.846	1.000	1.000	4.468
500	100			0.0	0.0	1.000	1.000	0.903	0.903	0.916	1.000	1.000	4.225
500	100		0.0		0.3	1.000	1.000	0.917	0.917	0.916	1.000	1.000	4.170
500	100			0.3	0.0	1.000	1.000	0.898	0.898	0.905	1.000	1.000	4.318
500	100				0.3	1.000	1.000	0.909	0.909	0.906	1.000	1.000	4.028
500	100			0.0	0.0	1.000	1.000	0.960	0.960	0.964	1.000	1.000	4.541
500	100		0.7		0.3	1.000	1.000	0.960	0.960	0.964	1.000	1.000	4.799
500	100			0.3	0.0	1.000	1.000	0.936	0.936	0.940	1.000	1.000	4.438
500	100				0.3	1.000	1.000	0.938	0.938	0.940	1.000	1.000	4.759
500	200			0.0	0.0	1.000	1.000	0.924	0.924	0.911	1.000	1.000	4.448
500	200		0.0		0.3	1.000	1.000	0.927	0.927	0.912	1.000	1.000	4.393
500	200			0.3	0.0	1.000	1.000	0.913	0.913	0.901	1.000	1.000	4.524
500	200				0.3	1.000	1.000	0.919	0.919	0.901	1.000	1.000	4.331
500	200			0.0	0.0	1.000	1.000	0.964	0.964	0.962	1.000	1.000	4.263
500	200		0.7		0.3	1.000	1.000	0.965	0.965	0.962	1.000	1.000	4.155
500	200			0.3	0.0	1.000	1.000	0.946	0.945	0.938	1.000	1.000	4.346
500	200				0.3	1.000	1.000	0.946	0.946	0.938	1.000	1.000	4.176

200	100		0.0	0.0	0.0	1.000	1.000	0.764	0.764	0.798	1.000	1.000	4.069
200	100				0.3	1.000	1.000	0.825	0.825	0.797	1.000	1.000	4.088
200	100			0.3	0.0	1.000	1.000	0.772	0.772	0.776	1.000	1.000	4.026
200	100				0.3	1.000	1.000	0.759	0.759	0.777	1.000	1.000	4.030
200	100		0.7	0.0	0.0	1.000	1.000	0.886	0.886	0.903	1.000	1.000	4.857
200	100				0.3	1.000	1.000	0.922	0.922	0.903	1.000	1.000	4.872
200	100			0.3	0.0	1.000	1.000	0.828	0.828	0.856	1.000	1.000	4.827
200	100				0.3	1.000	1.000	0.821	0.821	0.858	1.000	1.000	4.846
200	200		0.0	0.0	0.0	1.000	1.000	0.811	0.811	0.788	1.000	1.000	4.310
200	200				0.3	1.000	1.000	0.823	0.823	0.789	1.000	1.000	4.318
200	200	0.7		0.3	0.0	1.000	1.000	0.779	0.779	0.767	1.000	1.000	4.188
200	200				0.3	1.000	1.000	0.812	0.812	0.772	1.000	1.000	4.206
200	200		0.7	0.0	0.0	1.000	1.000	0.869	0.869	0.898	1.000	1.000	5.046
200	200				0.3	1.000	1.000	0.917	0.917	0.898	1.000	1.000	5.053
200	200			0.3	0.0	1.000	1.000	0.823	0.823	0.851	1.000	1.000	5.039
200	200				0.3	1.000	1.000	0.865	0.865	0.850	1.000	1.000	5.047
500	100		0.0	0.0	0.0	1.000	1.000	0.918	0.918	0.922	1.000	1.000	4.330
500	100				0.3	1.000	1.000	0.924	0.924	0.922	1.000	1.000	4.365
500	100			0.3	0.0	1.000	1.000	0.897	0.897	0.911	1.000	1.000	4.168
500	100				0.3	1.000	1.000	0.909	0.909	0.911	1.000	1.000	4.191
500	100		0.7	0.0	0.0	1.000	1.000	0.971	0.971	0.968	1.000	1.000	5.049
500	100				0.3	1.000	1.000	0.938	0.938	0.969	1.000	1.000	5.065
500	100			0.3	0.0	1.000	1.000	0.940	0.940	0.947	1.000	1.000	4.999
500	100				0.3	1.000	1.000	0.945	0.945	0.946	1.000	1.000	5.012
500	200		0.0	0.0	0.0	1.000	1.000	0.935	0.935	0.918	1.000	1.000	4.858
500	200				0.3	1.000	1.000	0.897	0.897	0.917	1.000	1.000	4.863
500	200			0.3	0.0	1.000	1.000	0.927	0.927	0.905	1.000	1.000	4.769
500	200				0.3	1.000	1.000	0.906	0.906	0.907	1.000	1.000	4.772
500	200		0.7	0.0	0.0	1.000	1.000	0.954	0.954	0.968	1.000	1.000	5.568
500	200				0.3	1.000	1.000	0.966	0.966	0.969	1.000	1.000	5.583
500	200			0.3	0.0	1.000	1.000	0.928	0.928	0.942	1.000	1.000	5.491
500	200				0.3	1.000	1.000	0.955	0.955	0.943	1.000	1.000	5.511

Note:

Entries denote the rejection rates across different simulated breaks where both $W \neq 0$ and $Z \neq I$. HI denotes Han and Inoue (2015)'s test, and BKW denotes Baltagi et al (2021)'s test conducted with a pre-known break fraction. The scalar ω denotes the "size" of the break in the loadings. See Table 2.1 for explanation of α, β, and ρ.

A.3 Empirical

A.3.1 Data Description

Table A.4: Data Description

Mnemonic	Description	Group	Trans.
PCDGx	Real personal consumption expenditures: Durable goods (Billions of Chained 2012 Dollars), deflated using PCE	NIPA	5
PCESVx	Real Personal Consumption Expenditures: Services (Billions of 2012 Dollars), deflated using PCE	NIPA	5
PCNDx	Real Personal Consumption Expenditures: Nondurable Goods (Billions of 2012 Dollars), deflated using PCE	NIPA	5
Y033RC1Q027SBEAx	Real Gross Private Domestic Investment: Fixed Investment: Nonresidential: Equipment (Billions of Chained 2012 Dollars), deflated using PCE	NIPA	5
PNFIx	Real private fixed investment: Nonresidential (Billions of Chained 2012 Dollars), deflated using PCE	NIPA	5
PRFIx	Real private fixed investment: Residential (Billions of Chained 2012 Dollars), deflated using PCE	NIPA	5
A014RE1Q156NBEA	Shares of gross domestic product: Gross private domestic investment: Change in private inventories (Percent)	NIPA	1
A823RL1Q225SBEA	Real Government Consumption Expenditures and Gross Investment: Federal (Percent Change from Preceding Period)	NIPA	1
FGRECPTx	Real Federal Government Current Receipts (Billions of Chained 2012 Dollars), deflated using PCE	NIPA	5
SLCEx	Real government state and local consumption expenditures (Billions of Chained 2012 Dollars), deflated using PCE	NIPA	5
EXPGSC1	Real Exports of Goods & Services, 3 Decimal (Billions of Chained 2012 Dollars)	NIPA	5
IMPGSC1	Real Imports of Goods & Services, 3 Decimal (Billions of Chained 2012 Dollars)	NIPA	5
IPDMAT	Industrial Production: Durable Materials (Index 2012=100)	Industrial Production	5
IPNMAT	Industrial Production: Nondurable Materials (Index 2012=100)	Industrial Production	5
IPDCONGD	Industrial Production: Durable Consumer Goods (Index 2012=100)	Industrial Production	5
IPB51110SQ	Industrial Production: Durable Goods: Automotive products (Index 2012=100)	Industrial Production	5
IPNCONGD	Industrial Production: Nondurable Consumer Goods (Index 2012=100)	Industrial Production	5
IPBUSEQ	Industrial Production: Business Equipment (Index 2012=100)	Industrial Production	5
IPB51220SQ	Industrial Production: Consumer energy products (Index 2012=100)	Industrial Production	5

Table A.4: Data Description *(continued)*

Mnemonic	Description	Group	Trans.
TCU	Capacity Utilization: Total Industry (Percent of Capacity)	Industrial Production	1
CUMFNS	Capacity Utilization: Manufacturing (SIC) (Percent of Capacity)	Industrial Production	1
DMANEMP	All Employees: Durable goods (Thousands of Persons)	Employment	5
USCONS	All Employees: Construction (Thousands of Persons)	Employment	5
USEHS	All Employees: Education & Health Services (Thousands of Persons)	Employment	5
USFIRE	All Employees: Financial Activities (Thousands of Persons)	Employment	5
USINFO	All Employees: Information Services (Thousands of Persons)	Employment	5
USPBS	All Employees: Professional & Business Services (Thousands of Persons)	Employment	5
USLAH	All Employees: Leisure & Hospitality (Thousands of Persons)	Employment	5
USSERV	All Employees: Other Services (Thousands of Persons)	Employment	5
USMINE	All Employees: Mining and logging (Thousands of Persons)	Employment	5
USTPU	All Employees: Trade, Transportation & Utilities (Thousands of Persons)	Employment	5
USTRADE	All Employees: Retail Trade (Thousands of Persons)	Employment	5
USWTRADE	All Employees: Wholesale Trade (Thousands of Persons)	Employment	5
CES9091000001	All Employees: Government: Federal (Thousands of Persons)	Employment	5
CES9092000001	All Employees: Government: State Government (Thousands of Persons)	Employment	5
CES9093000001	All Employees: Government: Local Government (Thousands of Persons)	Employment	5
LNS14000012	Unemployment Rate - 16 to 19 years (Percent)	Employment	2
LNS14000025	Unemployment Rate - 20 years and over, Men (Percent)	Employment	2
LNS14000026	Unemployment Rate - 20 years and over, Women (Percent)	Employment	2
UEMPLT5	Number of Civilians Unemployed - Less Than 5 Weeks (Thousands of Persons)	Employment	5
UEMP5TO14	Number of Civilians Unemployed for 5 to 14 Weeks (Thousands of Persons)	Employment	5
UEMP15T26	Number of Civilians Unemployed for 15 to 26 Weeks (Thousands of Persons)	Employment	5
UEMP27OV	Number of Civilians Unemployed for 27 Weeks and Over (Thousands of Persons)	Employment	5
LNS13023621	Unemployment Level - Job Losers (Thousands of Persons)	Employment	5
LNS13023557	Unemployment Level - Reentrants to Labor Force (Thousands of Persons)	Employment	5
LNS13023705	Unemployment Level - Job Leavers (Thousands of Persons)	Employment	5
LNS13023569	Unemployment Level - New Entrants (Thousands of Persons)	Employment	5

Table A.4: Data Description *(continued)*

Mnemonic	Description	Group	Trans.
LNS12032194	Employment Level - Part-Time for Economic Reasons, All Industries (Thousands of Persons)	Employment	5
AWHMAN	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing (Hours)	Employment	1
AWHNONAG	Average Weekly Hours Of Production And Nonsupervisory Employees: Total private (Hours)	Employment	2
AWOTMAN	Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing (Hours)	Employment	2
PERMIT	New Private Housing Units Authorized by Building Permits (Thousands of Units)	Housing	5
HOUSTMW	Housing Starts in Midwest Census Region (Thousands of Units)	Housing	5
HOUSTNE	Housing Starts in Northeast Census Region (Thousands of Units)	Housing	5
HOUSTS	Housing Starts in South Census Region (Thousands of Units)	Housing	5
HOUSTW	Housing Starts in West Census Region (Thousands of Units)	Housing	5
RSAFSx	Real Retail and Food Services Sales (Millions of Chained 2012 Dollars), deflated by Core PCE	Inventories	5
AMDMNOx	Real Manufacturers' New Orders: Durable Goods (Millions of 2012 Dollars), deflated by Core PCE	Inventories	5
ACOGNOx	Real Value of Manufacturers' New Orders for Consumer Goods Industries (Millions of 2012 Dollars), deflated by Core PCE	Inventories	5
AMDMUOx	Real Value of Manufacturers' Unfilled Orders for Durable Goods Industries (Millions of 2012 Dollars), deflated by Core PCE	Inventories	5
ANDENOx	Real Value of Manufacturers' New Orders for Capital Goods: Nondefense Capital Goods Industries (Millions of 2012 Dollars), deflated by Core PCE	Inventories	5
INVCQRMTSPL	Real Manufacturing and Trade Inventories (Millions of 2012 Dollars)	Inventories	5
GPDICTPI	Gross Private Domestic Investment: Chain-type Price Index (Index 2009=100)	Prices	6
IPDBS	Business Sector: Implicit Price Deflator (Index 2009=100)	Prices	6
DMOTRG3Q086SBEA	Personal consumption expenditures: Durable goods: Motor vehicles and parts (chain-type price index)	Prices	6
DFDHRG3Q086SBEA	Personal consumption expenditures: Durable goods: Furnishings and durable household equipment (chain-type price index)	Prices	6
DREQRG3Q086SBEA	Personal consumption expenditures: Durable goods: Recreational goods and vehicles (chain-type price index)	Prices	6
DODGRG3Q086SBEA	Personal consumption expenditures: Durable goods: Other durable goods (chain-type price index)	Prices	6

Table A.4: Data Description *(continued)*

Mnemonic	Description	Group	Trans.
DFXARG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Food and beverages purchased for off-premises consumption (chain-type price index)	Prices	6
DCLORG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Clothing and footwear (chain-type price index)	Prices	6
DGOERG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Gasoline and other energy goods (chain-type price index)	Prices	6
DONGRG3Q086SBEA	Personal consumption expenditures: Nondurable goods: Other nondurable goods (chain-type price index)	Prices	6
DHUTRG3Q086SBEA	Personal consumption expenditures: Services: Housing and utilities (chain-type price index)	Prices	6
DHLCRG3Q086SBEA	Personal consumption expenditures: Services: Health care (chain-type price index)	Prices	6
DTRSRG3Q086SBEA	Personal consumption expenditures: Transportation services (chain-type price index)	Prices	6
DRCARG3Q086SBEA	Personal consumption expenditures: Recreation services (chain-type price index)	Prices	6
DFSARG3Q086SBEA	Personal consumption expenditures: Services: Food services and accommodations (chain-type price index)	Prices	6
DIFSRG3Q086SBEA	Personal consumption expenditures: Financial services and insurance (chain-type price index)	Prices	6
DOTSRG3Q086SBEA	Personal consumption expenditures: Other services (chain-type price index)	Prices	6
WPSFD49502	Producer Price Index by Commodity for Finished Consumer Goods (Index 1982=100)	Prices	6
WPSFD4111	Producer Price Index by Commodity for Finished Consumer Foods (Index 1982=100)	Prices	6
PPIIDC	Producer Price Index by Commodity Industrial Commodities (Index 1982=100)	Prices	6
WPSID61	Producer Price Index by Commodity Intermediate Materials: Supplies & Components (Index 1982=100)	Prices	6
WPU0531	Producer Price Index by Commodity for Fuels and Related Products and Power: Natural Gas (Index 1982=100)	Prices	5
WPU0561	Producer Price Index by Commodity for Fuels and Related Products and Power: Crude Petroleum (Domestic Production) (Index 1982=100)	Prices	5
COMPRMS	Manufacturing Sector: Real Compensation Per Hour (Index 2009=100)	Earnings	5

Table A.4: Data Description *(continued)*

Mnemonic	Description	Group	Trans.
COMPRNFB	Nonfarm Business Sector: Real Compensation Per Hour (Index 2009=100)	Earnings	5
RCPHBS	Business Sector: Real Compensation Per Hour (Index 2009=100)	Earnings	5
OPHMFG	Manufacturing Sector: Real Output Per Hour of All Persons (Index 2009=100)	Earnings	5
OPHNFB	Nonfarm Business Sector: Real Output Per Hour of All Persons (Index 2009=100)	Earnings	5
ULCMFG	Manufacturing Sector: Unit Labor Cost (Index 2009=100)	Earnings	5
ULCNFB	Nonfarm Business Sector: Unit Labor Cost (Index 2009=100)	Earnings	5
UNLPNBS	Nonfarm Business Sector: Unit Nonlabor Payments (Index 2009=100)	Earnings	5
FEDFUNDS	Effective Federal Funds Rate (Percent)	Interest Rates	2
TB3MS	3-Month Treasury Bill: Secondary Market Rate (Percent)	Interest Rates	2
BAA10YM	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity (Percent)	Interest Rates	1
MORTG10YRx	30-Year Conventional Mortgage Rate Relative to 10-Year Treasury Constant Maturity (Percent)	Interest Rates	1
TB6M3Mx	6-Month Treasury Bill Minus 3-Month Treasury Bill, secondary market (Percent)	Interest Rates	1
GS1TB3Mx	1-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)	Interest Rates	1
GS10TB3Mx	10-Year Treasury Constant Maturity Minus 3-Month Treasury Bill, secondary market (Percent)	Interest Rates	1
CPF3MTB3Mx	3-Month Commercial Paper Minus 3-Month Treasury Bill, secondary market (Percent)	Interest Rates	1
BUSLOANSx	Real Commercial and Industrial Loans, All Commercial Banks (Billions of 2009 U.S. Dollars), deflated by Core PCE	Money	5
CONSUMERx	Real Consumer Loans at All Commercial Banks (Billions of 2009 U.S. Dollars), deflated by Core PCE	Money	5
NONREVSLx	Total Real Nonrevolving Credit Owned and Securitized, Outstanding (Billions of Dollars), deflated by Core PCE	Money	5
REALLNx	Real Real Estate Loans, All Commercial Banks (Billions of 2009 U.S. Dollars), deflated by Core PCE	Money	5
REVOLSLx	Total Real Revolving Credit Owned and Securitized, Outstanding (Billions of 2012 Dollars), deflated by Core PCE	Money	5
DRIWCIL	FRB Senior Loans Officer Opions. Net Percentage of Domestic Respondents Reporting Increased Willingness to Make Consumer Installment Loans	Money	1

Table A.4: Data Description *(continued)*

Mnemonic	Description	Group	Trans.
TLBSHNOx	Real Total Liabilities of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE	Household Balance	5
TNWBSHNOx	Real Net Worth of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE	Household Balance	5
TARESAx	Real Assets of Households and Nonprofit Organizations excluding Real Estate Assets (Billions of 2012 Dollars), deflated by Core PCE	Household Balance	5
HNOREMQ027Sx	Real Real Estate Assets of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE	Household Balance	5
TFAABSHNOx	Real Total Financial Assets of Households and Nonprofit Organizations (Billions of 2012 Dollars), deflated by Core PCE	Household Balance	5
VXOCLSx	CBOE S&P 100 Volatility Index: VXO	Stock Markets	1
USSTHPI	All-Transactions House Price Index for the United States (Index 1980 Q1=100)	Housing	5
SPCS10RSA	S&P/Case-Shiller 10-City Composite Home Price Index (Index January 2000 = 100)	Housing	5
SPCS20RSA	S&P/Case-Shiller 20-City Composite Home Price Index (Index January 2000 = 100)	Housing	5
TWEXAFEGSMTHx	Trade Weighted U.S. Dollar Index: Major Currencies (Index March 1973=100)	Exchange Rates	5
EXUSEU	U.S. / Euro Foreign Exchange Rate (U.S. Dollars to One Euro)	Exchange Rates	5
EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	Exchange Rates	5
EXJPUSx	Japan / U.S. Foreign Exchange Rate	Exchange Rates	5
EXUSUKx	U.S. / U.K. Foreign Exchange Rate	Exchange Rates	5
EXCAUSx	Canada / U.S. Foreign Exchange Rate	Exchange Rates	5
UMCSENTx	University of Michigan: Consumer Sentiment (Index 1st Quarter 1966=100)	Other	1
USEPUINDXM	Economic Policy Uncertainty Index for United States	Other	2

Note:

Transformation codes correspond to: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $log(x_t)$; (5) $\Delta log(x_t)$; (6) $\Delta^2 log(x_t)$; (7) $\Delta(x_t/x_{t-1} - 1.0)$

A.3.2 R^2 Variance Decomposition Exercise

We present the details for how to calculate the R^2 figures in Section 2.5.4. Equation (2.2.7) can be used to decompose the common component in the second regime $\widehat{X}_2$ as follows:

$$\widehat{X}_2 = \tilde{F}_2 \left(\tilde{\Lambda}_1 \tilde{Z}\right)^\top + \tilde{F}_2 \tilde{W}^\top, \tag{A.3.1}$$

which allows us to study the effect of the change in the factor loadings where we set $\tilde{W} = \mathbf{0}$ to yield

$$\widehat{X}_{2,W=\mathbf{0}} = \tilde{F}_2 \left(\tilde{\Lambda}_1 \tilde{Z}\right)^\top. \tag{A.3.2}$$

Combined with $\widehat{X}_1 = \tilde{F}_1 \tilde{\Lambda}_1^\top$ in order to produce estimates of the common components with $W = \mathbf{0}$ restriction yields

$$\widehat{X}_{W=\mathbf{0}} = \begin{bmatrix} \widehat{X}_1 \\ \widehat{X}_{2,W=\mathbf{0}} \end{bmatrix}$$

For a given estimate of $\widehat{X}$, restricted or otherwise, the corresponding R^2 is calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^N (\widehat{X}_i - X_i)^2}{\sum_{i=1}^N X_i^2}. \tag{A.3.3}$$

The intuition is as follows. If breaks in the factor loading were important for understanding variation in particular variables, this will induce a large discrepancy between the restricted and unrestricted R^2. Note that in some cases, this metric can be negative if $\sum_{i=1}^N (\widehat{X}_i - X_i)^2 > \sum_{i=1}^N X_i^2$, i.e. the fit is so poor that the residual variation is larger than the variation in the data itself.

A.3.3 Robustness Checks

Table A.5: Great Recession (2008 Q3), 1984 Q2 - 2019 Q4 Sample, allowing changing r

		Z-test p values		W-test p values		
$\tilde{r}_1$	$\tilde{r}_2$	Unadjusted	Adjusted	Unadjusted	Adjusted	w_i test Reject Count

2	4	0.000	0.000	0.080	0.080	56

Note:

Rejection of the Z-test corresponds to a break in the factor covariance matrix, and rejection of the W-test corresponds to a break in the loadings across the entire cross-section. $\tilde{r}_1$ and $\tilde{r}_2$ are estimated using the Bai and Ng (2002) criteria and Onatski's (2010) estimator.

Table A.5 shows the results of the proposed test statistics allowing for a change in the number of factors, for the Great Recession subsample, where $\tilde{r}_1$ and $\tilde{r}_2$ are estimated to be 2 and 4, respectively, according to the criteria of Bai and Ng (2002) and estimator of Onatski (2010). Results for the Great Moderation subsample are omitted, as $\tilde{r}_1$ and $\tilde{r}_2$ are both estimated to be 3 by Bai and Ng (2002) and Onatski (2010).

Figure A.1: R^2 Statistics for unrestricted and restricted common component ($W = 0$) for Great Moderation Subsample, and Global Financial Crisis Subsample, for $r = 2$ to 6 factors.

A.3.4 Estimation of $tr(H_{0,1}^\top \Sigma_F H_{0,1})$ and $tr(H_{0,1}^\top Z \Sigma_F Z^\top H_{0,1})$

We detail how to estimate the quantities $tr(H_{0,1}^\top \Sigma_F H_{0,1})$ and $tr(H_{0,1}^\top Z \Sigma_F Z^\top H_{0,1})$.

Similar to the case of the factors themselves, cannot estimate Σ_F directly and can only estimate $H_{0,1}^\top \Sigma_F H_{0,1}$ which is an r dimensional identity matrix due to the normalisation effect of $H_{0,1}$. This follows from the fact that we estimate $\tilde{f}_{1t}$, which is an estimate of $H_1^\top f_{1t}$, and hence $\frac{1}{T_1} \tilde{F}_1^\top \tilde{F}_1 = \frac{1}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \tilde{f}_{1t} \tilde{f}_{1t}^\top = I_r$. Hence, $tr(\frac{1}{T_1} \tilde{F}_1^\top \tilde{F}_1) = tr(I_r)$ is an estimate of $tr(H_{0,1}^\top \Sigma_F H_{0,1})$.

Similarly, we cannot estimate $Z \Sigma_F Z^\top$ directly, and only up to a normalisation basis. By Lemma A.3 (a), $\tilde{F}_2 \tilde{Z}^\top$ is an estimate of $F_2 Z^\top H_1$, and hence $\frac{1}{T_2} \tilde{Z} \tilde{F}_2^\top \tilde{F}_2 \tilde{Z} = \tilde{Z} \tilde{Z}^\top$ is an estimate of $\frac{1}{T - \lfloor \pi T \rfloor} \sum_{t=\lfloor \pi T + 1 \rfloor}^{T} H_{0,1}^\top Z f_{2t} f_{2t}^\top Z^\top H_{0,1}$, which converges to $Z \Sigma_F Z^\top$. Therefore, $tr(\tilde{Z} \tilde{Z}^\top)$ is an estimate of $tr(H_{0,1}^\top Z \Sigma_F Z^\top H_{0,1})$.

Appendix B

Appendices for Chapter 3

B.1 Factor Model Proofs

B.1.1 Preliminary

We first state some preliminary results used throughout the proofs.

$$\frac{F^\top e}{\sqrt{NT}} = O_p(1) \tag{B.1.1}$$

$$\frac{\Lambda_1^\top e}{\sqrt{NT}} = O_p(1) \tag{B.1.2}$$

$$\frac{W^\top e}{\sqrt{N^\alpha T}} = O_p(1) \tag{B.1.3}$$

$$\frac{ee^\top}{NT} = O_p\left(\frac{1}{\delta_{NT}}\right) \tag{B.1.4}$$

$$\frac{F^\top e\Lambda}{NT} = O_p\left(\frac{1}{\delta_{NT}^2}\right) \tag{B.1.5}$$

$$\frac{F^\top eW}{N^\alpha T} = O_p\left(\frac{1}{\delta_{NT}^2}\right) \tag{B.1.6}$$

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

B.1.2 Pseudo-factors $\tilde{F}_P$

To begin, we make the following expansion

$$\begin{aligned}\tilde{F}_P V_{NT,r} &= \frac{1}{TN} X X^\top \tilde{F}_P \\ \tilde{F}_P &= \frac{1}{TN}\left(G_r\Lambda_1^\top + G_pW^\top + e\right)\left(G_r\Lambda_1^\top + G_pW^\top + e\right)^\top V_{NT,r}^{-1}.\end{aligned} \tag{B.1.7}$$

Proof of Theorem 3.1 (a). Expanding out Equation (B.1.7), we have

$$\begin{aligned}\tilde{F}_P = \frac{1}{TN}\Big(&G_r\Lambda_1^\top\Lambda_1G_r^\top\tilde{F}_P + G_r\Lambda_1^\top e^\top\tilde{F}_P + e\Lambda_1^\top G_r^\top\tilde{F}_P + ee^\top\tilde{F}_P \\ &+eWG_p^\top\tilde{F}_P + G_pW^\top WG_p^\top\tilde{F}_P + G_r\Lambda_1^\top WG_p^\top\tilde{F}_P + G_pW^\top\Lambda_1G_r^\top\tilde{F}_P\Big)V_{NT,r}^{-1}.\end{aligned} \tag{B.1.8}$$

Substituting in H_G and rearranging yields

$$\begin{aligned}\tilde{F}_P - G_rH_G = \frac{1}{TN}\Big(&G_r\Lambda_1^\top e^\top\tilde{F}_P + e\Lambda_1G_r^\top\tilde{F}_P + ee^\top\tilde{F}_P + eWG_p^\top\tilde{F}_P \\ &+G_pW^\top e^\top\tilde{F}_P + G_pW^\top WG_p^\top\tilde{F}_P + G_r\Lambda_1^\top WG_p^\top\tilde{F}_P + G_pW^\top\Lambda_1G_r^\top\tilde{F}_P\Big)V_{NT,r}^{-1}.\end{aligned} \tag{B.1.9}$$

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

$$\begin{aligned}&\frac{1}{\sqrt{T}}\left(\tilde{F}_P - G_rH_G\right) \\ =&\frac{1}{\sqrt{T}}\frac{1}{TN}\Big(G_r\Lambda_1^\top e^\top\tilde{F}_P + e\Lambda_1G_r^\top\tilde{F}_P + ee^\top\tilde{F}_P + eWG_p^\top\tilde{F}_P \\ &+G_pW^\top e^\top\tilde{F}_P + G_pW^\top WG_p^\top\tilde{F}_P + G_r\Lambda_1^\top WG_p^\top\tilde{F}_P + G_pW^\top\Lambda_1G_r^\top\tilde{F}_P\Big)V_{NT,r} \\ =&\left(a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8\right)V_{NT,r}^{-1}.\end{aligned}$$

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

$$\frac{1}{NT\sqrt{T}}\left\|ee^\top \tilde{F}_P\right\| = \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{TN}\sum_{t=1}^{T} e_s^\top e_t \tilde{F}_{P,t}\right\|^2\right)^{1/2},$$

$$\frac{1}{TN}\sum_{t=1}^{T} e_s^\top e_t \tilde{F}_{P,t} = \frac{1}{TN}\sum_{t=1}^{T}\left[e_s^\top e_t - E(e_s^\top e_t)\right]\tilde{F}_{P,t}^\top + \frac{1}{TN}\sum_{t=1}^{T} E(e_s^\top e_t)\tilde{F}_{P,t}^\top,$$

$$\begin{aligned}\frac{1}{TN}\sum_{t=1}^{T}\left[e_s^\top e_t - E(e_s^\top e_t)\tilde{F}_{P,t}\right] &= \frac{1}{TN}\sum_{t=1}^{T}\left[e_s^\top e_t - E(e_s^\top e_t)\right] G_{r,t}^\top H_G \\ &\quad + \frac{1}{TN}\sum_{t=1}^{T}\left[e_s^\top e_t - E(e_s^\top e_t)\right]\left(\tilde{F}_{P,t}^\top - G_{r,t}^\top H_G\right) \\ &= O_p\left(\frac{1}{\sqrt{TN}}\right) + O_p\left(\frac{1}{\sqrt{N}\delta_{NT}} + \frac{N^\alpha}{N\sqrt{N}}\right), \quad \text{and} \end{aligned} \tag{B.1.10}$$

$$\begin{aligned}\frac{1}{TN}\sum_{t=1}^{T} E\left[e_s^\top e_t\right]\tilde{F}_{P,t}^\top &= \frac{1}{T}\sum_{t=1}^{T} E(e_s^\top e_t/N) G_{r,t}^\top H_G + \frac{1}{T}\sum_{t=1}^{T} E(e_s^\top e_t/N)\left(\tilde{F}_{P,t}^\top - G_{r,t} H_G\right) \\ &= O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{T}\delta_{NT}} + \frac{N^\alpha}{\sqrt{T}N}\right). \end{aligned} \tag{B.1.11}$$

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 3.1 (a).

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

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

Theorem 3.1 (a) can then be used to prove the following lemmas for the pseudo-factors $\tilde{F}_P$.

Lemma B.1. *Under Assumptions 1 to 8 and as $N, T \to \infty$ and $\alpha < 1$,* *a)* $\frac{1}{T}\left(\tilde{F}_P - G_r H_G\right)^\top G_r = O_p\left(\frac{1}{\delta^2_{NT}}\right) + O_p\left(\frac{N^\alpha}{N}\right)$, *if* $\alpha < 1$,

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

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

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

Proof of Lemma B.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

$$
\begin{aligned}
a_{13} &= \frac{\tilde{F}_P^\top G_r}{T}\frac{\Lambda_1^\top W}{N}\frac{G_p^\top G_r}{T} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right),\\
a_{14} &= \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top \Lambda}{N}\frac{G_r^\top G_r}{T} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right),\\
a_{15} &= \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N^\alpha}\frac{N^\alpha}{N}\frac{G_p^\top G_r}{T} = O_p\left(\frac{N^\alpha}{N}\right),\\
a_{16} &= \frac{\tilde{F}_P}{\sqrt{T}}\frac{eW}{\sqrt{N^\alpha T}}\frac{G_p^\top G_r}{T}\frac{\sqrt{N^\alpha}}{N} = O_p\left(\frac{\sqrt{N^\alpha}}{N}\right),\\
a_{17} &= \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top e^\top G_r}{N^\alpha T}\frac{N^\alpha}{N} = \frac{N^\alpha}{N}O_p\left(\frac{1}{\delta_{NT}^2}\right),\\
a_{18} &= \frac{\left(\tilde{F}_P - G_r H_G\right)^\top}{\sqrt{T}}\frac{ee^\top}{NT}\frac{G_r}{\sqrt{T}} + \frac{H_G^\top G_r^\top ee^\top G_r}{TNT}\\
&= \left(O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{N^\alpha}{N}\right)\right)O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{1}{T}\right)\\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right) + \frac{N^\alpha}{N}O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{1}{T}\right),\\
a_{19} &= \frac{\tilde{F}_P^\top G_r}{T}\frac{\Lambda_1^\top e^\top G_r}{TN} = O_p\left(\frac{1}{\delta_{NT}^2}\right), \quad \text{and}\\
a_{20} &= \frac{\left(\tilde{F}_P - G_r H_G\right)^\top}{\sqrt{T}}\frac{e\Lambda}{\sqrt{TN}}\frac{G_r^\top G_r}{T}\frac{1}{\sqrt{N}} + \frac{H_G^\top G_r^\top e\Lambda}{TN}\frac{G_r^\top G_r}{T}\\
&= \left(O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{N^\alpha}{N}\right)\right)\frac{1}{\sqrt{N}} + O_p\left(\frac{1}{\delta_{NT}^2}\right)\\
&= O_p\left(\frac{1}{\sqrt{N}\delta_{NT}}\right) + O_p\left(\frac{N^\alpha}{N^{3/2}}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}
$$

Collecting the dominating terms proves the lemmas. ■

Proof of Lemma B.1 (b).

$$\frac{1}{T}\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top G_r = V_{NT,r}^{-1}\frac{1}{T^2N}\left(\tilde{F}_P^\top G_p W^\top e^\top G_r + \tilde{F}_P^\top ee^\top G_r + \tilde{F}_P^\top G_r\Lambda_1^\top e^\top G_r + \tilde{F}_P^\top e\Lambda_1 G_r^\top G_r\right)$$

$$= V_{NT,r}^{-1}\left(a_{21} + a_{22} + a_{23} + a_{24}\right),$$

$$a_{21} = \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top e^\top G_r}{NT} = O_p\left(\frac{1}{\delta_{NT}^2}\right),$$

$$a_{22} = \frac{\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top}{\sqrt{T}}\frac{ee^\top}{NT}\frac{G_r}{\sqrt{T}} + \frac{\left(GH_{\Xi,r}\right)^\top}{\sqrt{T}}\frac{ee^\top}{NT}\frac{G_r}{\sqrt{T}}$$

$$= O_p\left(\frac{1}{\delta_{NT}}\right)O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{1}{T}\right),$$

$$a_{23} = \frac{\tilde{F}_P^\top G_r}{T}\frac{\Lambda_1^\top e^\top G_r}{NT} = O_p\left(\frac{1}{\delta_{NT}^2}\right), \quad \text{and}$$

$$a_{24} = \frac{\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top}{\sqrt{T}}\frac{e\Lambda}{\sqrt{TN}}\frac{G_r^\top G_r}{T} + \frac{H_{\Xi,r}^\top G_r^\top e\Lambda}{NT}\frac{G_r^\top G_r}{T}$$

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

$$\therefore \frac{1}{T}\left(\tilde{F}_P - GH_{\Xi,r}\right)^\top G_r = O_p\left(\frac{1}{\delta_{NT}^2}\right).$$

■

Proof of Lemma B.1 (c).

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

$$+ \tilde{F}_P^\top G_p W^\top W G_p^\top e_i + \tilde{F}_P^\top e W G_p^\top e_i$$

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

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

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

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

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

Proof of Lemma B.1 (d).

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

■

Additionally, Equation (B.1.7) allows us to study the expansion of each $\tilde{g}_{r,t}$. Begin by considering

$\tilde{f}_{P,t} - H_G^\top g_{r,t}$

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

Analysing each term, we have

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

$$
\begin{aligned}
a_{41} &= \frac{\left(\tilde{F}_P^\top - G_r H_G\right)^\top}{T} \frac{eW}{\sqrt{N^\alpha T}} \frac{\sqrt{N^\alpha}}{N} g_{p,t} + \frac{H_G^\top G_r^\top eW}{\sqrt{N^\alpha T}} \frac{\sqrt{N^\alpha}}{N\sqrt{T}} g_{p,t} \\
&= \left(O_p\left(\frac{1}{\delta_{NT}}\right) + O_p\left(\frac{N^\alpha}{N}\right)\right) O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N\sqrt{T}}\right), \\
a_{42} &= \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N^\alpha} \frac{N^\alpha}{N} g_{p,t} \\
&= O_p\left(\frac{N^\alpha}{N}\right), \\
a_{43} &= \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top \Lambda_1}{N} g_{r,t} \\
&= O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \quad \text{and} \\
a_{44} &= \frac{\tilde{F}_P^\top G_r}{T} \frac{\Lambda_1^\top W}{N} g_{p,t} \\
&= O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
\tilde{f}_{P,t} - H_G^\top g_{r,t} &= V_{NT}^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\frac{\Lambda_1^\top e_T}{N} + \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N} g_{p,t}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) \\
&= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{N^\alpha}{N}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right). \qquad \text{(B.1.12)}
\end{aligned}
$$

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

$$
\tilde{f}_{P,T} - H_G^\top f_T = V_{NT}^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\frac{\Lambda_1^\top e_T}{N} + \frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N} g_{p,T}\right) - H_G^\top (I - Z) F_T + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right). \qquad \text{(B.1.13)}
$$

Lemma B.2. *Under Assumptions 1 to 8,*

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

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,* $\Sigma_G = \operatorname{plim} \frac{1}{T} G^\top G$*, and* $+$ *denotes the pseudo inverse.*

Proof of Lemma B.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

$$\frac{\tilde{G}^\top G}{T} \xrightarrow{p} Q_G \equiv V_G^{1/2} \Upsilon_G^\top \Sigma_\Xi^{-1/2}, \tag{B.1.14}$$

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

$$\begin{aligned} \text{plim}\,\frac{\tilde{F}_P^\top G}{T} &= \text{plim}\begin{bmatrix} I_r & 0_r \end{bmatrix} \frac{\tilde{G}^\top G}{T} \\ &= \begin{bmatrix} I_r & 0_r \end{bmatrix} Q_G \\ &= \begin{bmatrix} V_r^{1/2} & 0_r \end{bmatrix} \Upsilon_G \Sigma_\Xi^{-1/2} \equiv Q_{G,r}, \end{aligned} \tag{B.1.15}$$

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

$$\begin{aligned} H_{0,\Xi,r} &= \text{plim}\,\frac{\Xi^\top\Xi}{N}\frac{G^\top \tilde{F}_P}{T}\left(\begin{bmatrix} I_r & 0_r \end{bmatrix} V_{NT,r} \begin{bmatrix} I_r \\ 0_r \end{bmatrix}\right)^{-1} \\ &= \Sigma_\Xi Q_{G,r}^\top V_r^{-1} \\ &= \Sigma_\Xi^{1/2}\Upsilon_G \begin{bmatrix} V_r^{-1/2} \\ 0_r \end{bmatrix} = Q_{G,r}^+, \end{aligned} \tag{B.1.16}$$

where $Q_{G,r}^+$ is the *pseudo* inverse of $Q_{G,r}$, and is a $2r \times r$ matrix.[1] ■

B.1.3 Split-sample Factors $\tilde{F}_S$

Proof of Theorem 3.1 (b). This is simply the subsample version of Theorem 1 of Bai and Ng (2002).

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

Note that Theorem 3.1 (b) can be equivalently stated as

$$\frac{1}{\sqrt{T}}\left\|\left(\tilde{F}_\iota - F_\iota H_\iota\right) - \frac{e_{(\iota)}\Lambda_\iota}{N}\frac{F_\iota^\top \tilde{F}_\iota}{T_\iota}V_{NT,\iota}^{-1}\right\| = O_p\left(\frac{1}{\delta_{NT}^2}\right), \tag{B.1.17}$$

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 3.1 (b) also implies the following lemmas.

Lemma B.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 B.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:

$$\tilde{F}_2 - F_2 H_2 = \frac{1}{T_2 N}\left(F_2\Lambda_2^\top e_{(2)}\tilde{F}_2 + e_{(2)}\Lambda_2 F_2^\top \tilde{F}_2 + e_{(2)}e_{(2)}^\top \tilde{F}_2\right)V_{NT,2}^{-1}, \quad \text{and} \tag{B.1.18}$$

$$\tilde{f}_{2,t} - H_2^\top f_t = V_{NT,2}^{-1}\left(\frac{\tilde{F}_2^\top e_{(2)}^\top \Lambda_2}{T_2 N} f_t + \frac{\tilde{F}_2^\top F_2}{T_2}\frac{\Lambda_2^\top e_t}{N} + \frac{\tilde{F}_2^\top e_{(2)} e_t}{T_2 N}\right), \tag{B.1.19}$$

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.

B.1.4 Rotated Factors $\tilde{F}_R$

Proof of Proposition 3.1. 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\|\left\|\frac{N^\alpha}{N}\right\|\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\|\left\|\frac{N^\alpha}{N}\right\|\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}
$$

$$
\begin{aligned}
Z.VI &= \frac{\tilde{F}_1^\top}{T_1}\frac{e_{(1)}\Lambda_1 Z}{N}\frac{F_2^\top \tilde{F}_2}{T_2} \\
&\leq \left\|\frac{(\tilde{F}_1 - F_1 H_1)}{\sqrt{T_1}}\right\|\left\|\frac{e_{(1)}\Lambda_1}{N\sqrt{T_1}}\right\|\|Z\|\left\|\frac{F_2^\top \tilde{F}_2}{T_2}\right\| + \|H\|\left\|\frac{F_1^\top e_{(1)}\Lambda_1}{T_1 N}\right\|\|Z\|\left\|\frac{F_2^\top \tilde{F}_2}{T_2}\right\| \\
&= O_p\left(\frac{1}{\delta_{NT}}\right)O_p\left(\frac{1}{\sqrt{N}}\right)O_p(1)\,O_p(1) + O_p(1)\,O_p\left(\frac{1}{\delta_{NT}^2}\right)O_p(1)\,O_p(1) \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}
$$

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

$$
\begin{aligned}
\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1 &= \tilde{F}_2(\tilde{Z}^\top - H_2^{-1}Z^\top H_1) + (\tilde{F}_2 H_2^{-1} - F_2)Z^\top H_1, \\
\tilde{F}_2\tilde{Z}^\top - F_2 H_1 &= \tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1 + F_2(Z^\top - I_r)H_1. \qquad \text{(B.1.20)}
\end{aligned}
$$

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

Theorem 3.1 (c) additionally can be used to derive the following lemmas.[2]

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

[2]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.

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 B.4 (a).

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

■

Proof of Lemma B.4 (b).

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

■

Proof of Lemma B.4 (c).

Beginning with $\frac{(\tilde{F}_2\tilde{Z}^\top - F_2 Z^\top H_1)^\top \tilde{F}_2}{T}$, we have

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

Adding and subtracting terms implies

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

■

B.1.5 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

$$\underset{r\times r^*}{H_{G,r^*}} = \frac{\Lambda_1^\top \Lambda_1}{N}\frac{G_r^\top \tilde{F}_{r^*}}{T} V_{NT,r^*}^{-1}, \tag{B.1.21}$$

$$\underset{2r\times r^*}{H_{\Xi,r^*}} = \frac{\Xi^\top \Xi}{N}\frac{G^\top \tilde{F}_{r^*}}{T} V_{NT,r^*}^{-1}, \tag{B.1.22}$$

$$\underset{r\times r^*}{H_{1,r^*}} = \frac{\Lambda_1^\top \Lambda_1}{N}\frac{F_1^\top \tilde{F}_{1,r^*}}{T} V_{NT,1,r^*}^{-1}, \tag{B.1.23}$$

$$\underset{r\times r^*}{H_{2,r^*}} = \frac{\Lambda_2^\top \Lambda_2}{N}\frac{F_2^\top \tilde{F}_{2,r^*}}{T} V_{NT,2,r^*}^{-1}, \tag{B.1.24}$$

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

$$\begin{aligned}
\|H_{G,r^*}\| &\le \left\|\frac{\tilde{F}_{P,r^*}^\top \tilde{F}_{P,r^*}}{T}\right\|^{1/2}\left\|\frac{G_r^\top G_r}{T}\right\|^{1/2}\left\|\frac{\Lambda_1^\top \Lambda_1}{N}\right\|\left\|V_{NT,r^*}^{-1}\right\| = O_p(1),\\
\|H_{\Xi,r^*}\| &\le \left\|\frac{\tilde{F}_{P,r^*}^\top \tilde{F}_{P,r^*}}{T}\right\|^{1/2}\left\|\frac{G^\top G}{T}\right\|^{1/2}\left\|\frac{\Xi^\top \Xi}{N}\right\|\left\|V_{NT,r^*}^{-1}\right\| = O_p(1),\\
\|H_{1,r^*}\| &\le \left\|\frac{\tilde{F}_{1,r^*}^\top \tilde{F}_{1,r^*}}{T}\right\|^{1/2}\left\|\frac{F_1^\top F_1}{T}\right\|^{1/2}\left\|\frac{\Lambda_1^\top \Lambda_1}{N}\right\|\left\|V_{NT,1,r^*}^{-1}\right\| = O_p(1), \quad \text{and}\\
\|H_{2,r^*}\| &\le \left\|\frac{\tilde{F}_{2,r^*}^\top \tilde{F}_{2,r^*}}{T}\right\|^{1/2}\left\|\frac{F_2^\top F_2}{T}\right\|^{1/2}\left\|\frac{\Lambda_2^\top \Lambda_2}{N}\right\|\left\|V_{NT,2,r^*}^{-1}\right\| = O_p(1).
\end{aligned}$$

Therefore, Theorem 3.1 (a) and Lemma B.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

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

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

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

where Q_{G,r^*}^+ is the *pseudo* inverse of Q_{G,r^*}, and is a $2r \times r^*$ matrix. By defining $Q_1 = \operatorname{plim} \frac{\tilde{F}_1^\top F_1}{T}$ and $Q_2 = \operatorname{plim} \frac{\tilde{F}_2^\top F_2}{T}$, 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 3.1 (b) containing the mean square consistency of the split-sample factors $\tilde{F}_S$ are unaffected. Lemma B.3 corresponds to Lemmas B.1 and B.2 of Bai (2003) by applying Theorem 3.1 (b), and therefore also holds.

Similarly, the proof of Proposition 3.1 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 3.1 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 $\left\| \tilde{F}_{2,r^*} - F_2 H_{2,r^*} \right\|^2 = O_p \left(\frac{1}{\delta_{NT}^2} \right)$.

B.1.6 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:

$$\begin{aligned} \Lambda_2 &= (\Lambda_1 + W) \begin{bmatrix} I_{r-r_2} \\ 0 \end{bmatrix} \\ &= \Lambda_1 Z_0 + W_0. \end{aligned} \tag{B.1.25}$$

This allows us to write

$$
\begin{aligned}
X_2 &= F_2\Lambda_2^\top + e_{(2)} \\
&= F_2 \begin{bmatrix} I_{r-r_2} \\ 0 \end{bmatrix} \left((\Lambda_1 Z + W) \begin{bmatrix} I_{r-r_2} \\ 0 \end{bmatrix} \right)^\top + e_{(2)} \\
&= F_{2,r-r_2} \left(\Lambda_1 Z_0 + W_0\right)^\top + e_{(2)},
\end{aligned} \tag{B.1.26}
$$

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

$$
H_{2,r-r_2} = \frac{\Lambda_2^\top \Lambda_2}{N} \frac{F_{2,r-r_2}^\top \tilde{F}_2}{T_2} V_{NT,2,r-r_2}^{-1} \tag{B.1.27}
$$

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

Lemma B.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 B.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 B.6. Lemmas B.6 (a) to B.6 (c) are analogous to Proposition 3.1, Theorem 3.1 (c), and Lemma B.4, and are all proved in a similar way. ■

B.1.7 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

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

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 B.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 B.7 (a) follows from Theorem 1 of Bai and Ng (2002). Lemma B.7 (b) follows by applying the results of Theorem 3.1 (a) to the post-break factors. Lemmas B.7 (c) and B.7 (d) are the counterparts to Lemma B.3. Lemma B.7 also allows us to state the following lemmas.

Next, we focus on the rotated factors. Lemma B.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 B.8. *Under Assumptions 1 to 8, as $N, T \to \infty$*

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

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

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

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

$$\begin{aligned}
\tilde{Z}^* &= \left(\tilde{\Lambda}_1^{*\top} \tilde{\Lambda}_1^*\right)^{-1} \tilde{\Lambda}_1^{*\top} \tilde{\Lambda}_2^* \\
&= \frac{1}{NT_1^* T_2^*} V_{NT,1}^{*-1} \left(\tilde{F}_1^{*\top} X_1^*\right) \left(\tilde{F}_2^{*\top} X_2^*\right)^\top \\
&= \frac{1}{NT_1^* T_2^*} V_{NT,1}^{*-1} \left(\tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top + \tilde{F}_1^{*\top} e_{(1)}\right) \left(\tilde{F}_2^* G^* \Xi^\top + \tilde{F}_2^* e_{(2)}\right)^\top \\
&= \frac{1}{NT_1^* T_2^*} V_{NT,1}^{*-1} \left(\tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top \Xi G^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top e_{(2)}^\top \tilde{F}_2^* + \tilde{F}_1^{*\top} e_{(1)} \Xi^* G^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} e_{(1)} e_{(2)}^\top \tilde{F}_2^*\right) \\
&= \frac{1}{NT_1^* T_2^*} V_{NT,1}^{*-1} \Big(\tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top \Lambda_1 G_r^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top W G_p^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} F_{11}^* \Lambda_1^\top e_{(2)}^\top \tilde{F}_2^* \\
&\qquad + \tilde{F}_1^{*\top} e_{(1)} \Lambda_1 G_r^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} e_{(1)} W G_p^{*\top} \tilde{F}_2^* + \tilde{F}_1^{*\top} e_{(1)} e_{(2)}^\top \tilde{F}_2^*\Big) \\
&= (Z.i + Z.ii + Z.iii + Z.iv + Z.v + Z.vi),
\end{aligned}$$

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 B.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,

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

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 B.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 B.9. Lemma B.9 (a) corresponds to Theorem 3.1 (a) and Theorem 3.1 (b) and can be proven similarly. Lemmas B.9 (b) and B.9 (c) correspond to Lemmas B.1 and B.3 and can be proven similarly. ■

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

Lemma B.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 B.10 is stated separately for the cases of $\alpha < 1$ and $\alpha = 1$.

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

$$
\begin{aligned}
\tilde{Z}^* &= \left(\tilde{\Lambda}_1^{*\top}\tilde{\Lambda}_1^*\right)^{-1}\tilde{\Lambda}_1^{*\top}\tilde{\Lambda}_2^* \\
&= \frac{1}{NT_1^*T_2^*}V_{NT,1}^{*-1}\left(\tilde{F}_1^{*\top}X_1^*\right)\left(\tilde{F}_2^{*\top}X_2^*\right)^\top \\
&= \frac{1}{NT_1^*T_2^*}V_{NT,1}^{*-1}\left(\tilde{F}_1^*G_1^*\Xi^\top + \tilde{F}_1^*e_{(1)}\right)\left(\tilde{F}_2^{*\top}G_2^*\Xi + \tilde{F}_2^{*\top}e_{(2)}\right)^\top \\
&= \frac{1}{NT_1^*T_2^*}V_{NT,1}^{*-1}\left(\tilde{F}_1^*G_1^*\Xi^\top\Xi G_2^{*\top}\tilde{F}_2^* + \tilde{F}_1^*G_1^*\Xi^\top e_{(2)}^\top\tilde{F}_2^* + \tilde{F}_1^*e_{(1)}^\top\Xi G_2^{*\top}\tilde{F}_2^* + \tilde{F}_1^*e_{(1)}^\top e_{(e)}\tilde{F}_2^*\right) \\
&= Z.vi + Z.vii + Z.viii + Z.ix.
\end{aligned}
$$

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

$$
\begin{aligned}
Z.vi = V_{NT,1}^{*-1} &\left(\frac{\tilde{F}_1^{*\top} G_r^*}{T_1^*} \frac{\Lambda_1^\top \Lambda_1 Z}{N} \frac{F_{22}^{*\top} \tilde{F}_2^*}{T_2^*} + \frac{\tilde{F}_1^{*\top} G_p^*}{T_1^*} \frac{W^\top \Lambda_1 Z}{N} \frac{F_{22}^{*\top} \tilde{F}_2^*}{T_2^*} \right. \\
&\qquad \left. + \frac{\tilde{F}_1^{*\top} G_r^*}{T_1^*} \frac{\Lambda_1^\top W}{N} \frac{F_{22}^{*\top} \tilde{F}_2^*}{T_2^*} + \frac{\tilde{F}_1^{*\top} G_p^*}{T_1^*} \frac{W^\top W}{N} \frac{F_{22}^{*\top} \tilde{F}_2^*}{T_2^*} \right) \\
&= H_{1,r}^{*\top} Z H_2^{*-\top} + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) + O_p\left(\frac{N^\alpha}{N}\right),
\end{aligned}
$$

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

B.2 Bias Variance Trade-off Proofs

B.2.1 Out-of-sample Asymptotic Expansions

In this subsection, we provide the precise asymptotic expansions for the out-of-sample forecasts. In what follows, we focus on the case of a DGP that contains only one lag of both the factor and y_t. That is, $Y = [y_1, \ldots, y_T]$ is regressed on $C = [c_{1-h}, \ldots, c_{T-h}]^\top$, where $c_t = \left[f_t^\top, z_t^\top\right]^\top$ are the infeasible regressors, with $z_t = (1, y_t, \ldots, y_{t-p})$, with corresponding forecasting coefficients $\theta = \left(\beta^\top, \delta^\top\right)^\top$. The case of more lags follow by suitably redefining these quantities at the cost of more complex notation. Therefore, our results hold without loss of generality.

Pseudo-factors

In the case where the pseudo-factors $\tilde{F}_P$ are used, the regressor matrix is $\tilde{C}_P = [\tilde{c}_{P,t-1}, \ldots, \tilde{c}_{P,T-h}]^\top$. 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 $C_{G_r} = [c_{G_r,t-h}, \ldots, c_{G_r,T-h}]^\top$, 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

$$c_T^\top \theta - \tilde{c}_{P,T}^\top \widehat{\theta}_P = c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top C\theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P\right)^{-1} \tilde{C}_P^\top \eta. \quad \text{(B.2.1)}$$

The squared norm of this is

$$\left\|c_T^\top\theta - \tilde{c}_{P,T}^\top\widehat{\theta}\right\|^2 = \left\|c_T^\top\theta - \tilde{c}_{P,T}^\top\left(\tilde{C}_P^\top\tilde{C}_P\right)^{-1}\tilde{C}_P^\top C\theta\right\|^2 + \left\|\tilde{c}_{P,T}^\top\left(\tilde{C}_P^\top\tilde{C}_P\right)^{-1}\tilde{C}_P^\top\eta\right\|^2 + 2\eta^\top\tilde{C}_P^\top\left(\tilde{C}_P^\top\tilde{C}_P\right)^{-1}\tilde{c}_{P,T}\tilde{c}_{P,T}^\top\left(\tilde{C}_P^\top\tilde{C}_P\right)^{-1}\tilde{C}_P^\top C\theta = bias^2 + var + cross,. \tag{B.2.2}$$

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 $O_p\left(\frac{1}{T}\right)$, and is mainly of interest in comparing the split-sample factors.

The bias term can be expanded as

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

By the expansion in the proof of Lemma B.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, $\frac{(\tilde{F}_P - FH_G, 0)^\top \tilde{C}_P}{T}$ follows by adding and subtracting

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

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 \\
&\quad + \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. \\
&\quad \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 \\
&= \left(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)\right. \\
&\quad \left. - 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)\right)^\top H_G^{-1} \beta \\
&\quad - \frac{e_T^\top \Lambda_1}{N} \left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \beta + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \\
&= \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. \\
&\quad \left. - \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),
\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

$$
\begin{aligned}
&\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{c}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C \theta \right\|^2 \\
=& \left\| \left[\left((I - Z) - \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T} \right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \right) \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \right. \right. \\
&\qquad \left. \left. - \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \frac{\Lambda_1^\top e_T}{N} \right]^\top \beta \right\|^2 \\
&\quad + O_p\left(N^{\alpha/2+\nu-2} \right) + O_p\left(N^{3\alpha/2} \right) + O_p\left(N^{\alpha/2-3/2} \right).
\end{aligned} \tag{B.2.3}
$$

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

Remark B.1. *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

$$
\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} \frac{\tilde{C}_P^\top \eta}{\sqrt{T}}\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}^\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_P C_{G_r})^\top C_{G_r}H_P}{T}\right)^{-1} \tilde{c}_{P,T}\tilde{c}_{P,T}^\top \left(\frac{(H_P C_{G_r})^\top C_{G_r}H_P}{T}\right)^{-1} \frac{(C_{G_r}H_P)^\top \eta}{\sqrt{T}}\right. \\
&\qquad \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_P C_{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_P C_{G_r})^\top C_{G_r}H_P}{T}\right)^{-1} \frac{(C_{G_r}H_P)^\top \eta}{\sqrt{T}}\right. \\
&\qquad + \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}} \\
&\qquad \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. \\
&\qquad \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}
$$

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 $F_2 H_2$, 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

$$c_T^\top \theta - \tilde{c}_{S,T}^\top \widehat{\theta}_S = c_T^\top \theta - \tilde{c}_{S,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2\right)^{-1} \tilde{C}_2^\top C_2 \theta - \tilde{c}_{S,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2\right)^{-1} \tilde{C}_2^\top \eta_{(2)}. \tag{B.2.4}$$

The out-of-sample forecast fit is then

$$\begin{aligned}&\left\| c_T^\top \theta - \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2\right)^{-1} \tilde{C}_2^\top Y_2 \right\|^2 \\ =&\left\| c_T^\top \theta - \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2\right)^{-1} \tilde{C}_2^\top C_2 \theta \right\|^2 + \left\| \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2\right)^{-1} \tilde{C}_2^\top \eta_{(2)} \right\|^2 \\ &\quad + 2\eta_{(2)}^\top \tilde{C}_2 \left(\tilde{C}_2^\top \tilde{C}_2\right)^{-1} \tilde{c}_{2,T} \left(c_T^\top \theta - \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2\right)^{-1} \tilde{C}_2^\top C_2 \theta \right) \qquad \text{(B.2.5)} \\ =&bias^2 + var + cross.\end{aligned}$$

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 B.3 (b). Therefore, for the bias term, we have

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

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

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

For the variance, we have

$$\begin{aligned}
&\eta_{(2)}^\top \tilde{C}_2 \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{c}_{2,T} \tilde{c}_{2,T}^\top \left(\tilde{C}_2^\top \tilde{C}_2 \right)^{-1} \tilde{C}_2^\top \eta_{(2)} \\
=&\frac{1}{T_2} \left[\frac{\eta_{(2)}^\top \tilde{C}_2}{\sqrt{T_2}} \left(\frac{\tilde{C}_2^\top \tilde{C}_2}{T_2} \right)^{-1} \tilde{c}_{2,T} \tilde{c}_{2,T}^\top \left(\frac{\tilde{C}_2^\top \tilde{C}_2}{T_2} \right)^{-1} \frac{\tilde{C}_2^\top \eta_{(2)}}{\sqrt{T_2}} \right] \\
=&\frac{1}{T_2} \left[\frac{\eta_{(2)}^\top C_2}{\sqrt{T_2}} \left(\frac{C_2^\top C_2}{T_2} \right)^{-1} c_T c_T^\top \left(\frac{C_2^\top C_2}{T_2} \right)^{-1} \frac{C_2^\top \eta_{(2)}}{\sqrt{T_2}} + O_p \left(\frac{\sqrt{T}}{\delta_{NT}^2} \right) \right]. \qquad \text{(B.2.6)}
\end{aligned}$$

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

$$c_T^\top \theta - \tilde{c}_{R,T}^\top \widehat{\theta}_R = c_T^\top \theta - \tilde{c}_{R,T}^\top \left(\tilde{C}_R^\top \tilde{C}_R \right)^{-1} \tilde{C}_R^\top C \theta - \tilde{c}_{R,T}^\top \left(\tilde{C}_R^\top \tilde{C}_R \right)^{-1} \tilde{C}_R^\top \eta. \qquad \text{(B.2.7)}$$

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

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

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

$$
\begin{aligned}
\tilde{Z}\tilde{f}_{2,T} - H_1^\top Z f_T &= H_1^\top Z H_2^{-\top} V_{NT,2}^{-1} \frac{\tilde{F}_2^\top F_2}{T} \frac{\Lambda_2^\top e_T}{N} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) \\
&= H_1^\top Z \left(\frac{\Lambda_2^\top \Lambda_2}{N}\right)^{-1} \frac{\Lambda_2^\top e_T}{N} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right), \qquad \text{(B.2.8)}
\end{aligned}
$$

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

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

$$
\begin{aligned}
&\frac{1}{T}\left(\tilde{C}_R - CH_R\right)^\top \tilde{C}_R \\
&= \begin{bmatrix} \frac{1}{T}\left(\tilde{F}_R - FH_1\right)^\top \tilde{C}_R \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{T}\left(\tilde{F}_1 - F_1H_1\right)^\top \tilde{C}_{R,1} + \frac{1}{T}\left(\tilde{F}_2\tilde{Z}^\top - F_2H_1\right)^\top \tilde{C}_{R,2} \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{T}\left(\tilde{F}_2\tilde{Z}^\top - F_2Z^\top H_1\right)^\top \tilde{C}_{R,2} - \frac{1}{T}H_1^\top(I-Z)F_2^\top \tilde{C}_{R,2} \\ 0 \end{bmatrix} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right) \\
&= \begin{bmatrix} -H_1^\top(I-Z)\frac{F_2^\top \tilde{C}_{R,2}}{T} \\ 0 \end{bmatrix} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^\alpha}}{N}\right).
\end{aligned}
$$

Therefore, the expression for the bias can be expressed as

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

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

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

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.\\
&\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.\\
&+\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)\\
&\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}}\\
&\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}}{\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].
\end{aligned}
$$

B.2.2 Small Shift Break $\alpha < 0.5$

Proof of Theorem 3.2 (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

$$
\begin{aligned}
&\tilde{c}_{R,T}^{\top}\widehat{\theta}_R - \tilde{c}_{P,T}^{\top}\widehat{\theta}_P \\
=&\tilde{c}_{R,T}^{\top}\left(\tilde{C}_R^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_R^{\top}C\theta - \tilde{c}_{P,T}^{\top}\left(\tilde{C}_P^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_P^{\top}C\theta \\
=&\tilde{c}_{R,T}^{\top}\left(C_R^{\top}C_R\right)^{-1}C_R^{\top}C\theta - \tilde{c}_{P,T}^{\top}\left(C_P^{\top}C_P\right)^{-1}C_P^{\top}C\theta + \tilde{c}_{R,T}^{\top}\left[\left(\tilde{C}_R^{\top}\tilde{C}_R\right)^{-1}\tilde{C}_R^{\top}C - \left(C_R^{\top}C_R\right)^{-1}C_R^{\top}C\right]\theta \\
&\quad - \tilde{c}_{P,T}^{\top}\left[\left(\tilde{C}_P^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_P^{\top}C - \left(C_P^{\top}C_P\right)^{-1}C_P^{\top}C\right]\theta \\
=&\tilde{c}_{R,T}^{\top}\left(C_R^{\top}C_R\right)^{-1}C_R^{\top}C\theta - \tilde{c}_{P,T}^{\top}\left(C_P^{\top}C_P\right)^{-1}C_P^{\top}C\theta + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^{\alpha}}{N}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right) \\
=&\left(\tilde{c}_{R,T}^{\top}H_R^{-1} - \tilde{c}_{P,T}H_P^{-1}\right)\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}C\theta + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{N^{\alpha}}{N}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right). \qquad \text{(B.2.9)}
\end{aligned}
$$

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

$$
\begin{aligned}
&\left(\tilde{f}_{R,T}^{\top}H_1^{-1} - f_T^{\top}Z\right) - \left(\tilde{f}_{P,T}^{\top}H_1^{-1} - f_T^{\top}Z\right) \\
=&\tilde{f}_{2,T}^{\top}\left(O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right)\right) + \frac{e_T^{\top}\Lambda_2}{N}\left(\frac{\Lambda_2^{\top}\Lambda_2}{N}\right)^{-1}Z^{\top} - \frac{e_T^{\top}\Lambda_1}{N}\left(\frac{\Lambda_1^{\top}\Lambda_1}{N}\right)^{-1} \\
&\quad - f_T^{\top}\frac{W^{\top}W}{N}\frac{G_p^{\top}\tilde{F}_P}{T}\left(\frac{G_r^{\top}\tilde{F}_P}{T}\right)^{-1}\left(\frac{\Lambda_1^{\top}\Lambda_1}{N}\right)^{-1} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right) \\
=&-f_T^{\top}\frac{W^{\top}W}{N}\frac{G_p^{\top}\tilde{F}_P}{T}\left(\frac{G_r^{\top}\tilde{F}_P}{T}\right)^{-1}\left(\frac{\Lambda_1^{\top}\Lambda_1}{N}\right)^{-1} + O_p\left(\frac{1}{\delta_{NT}^2}\right) + O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right). \qquad \text{(B.2.10)}
\end{aligned}
$$

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 $\alpha < 1/2$:

$$
\tilde{c}_{R,T}^{\top}\left(\tilde{C}_R^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_R^{\top}C\theta - \tilde{c}_{P,T}^{\top}\left(\tilde{C}_P^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_P^{\top}C\theta = o_p\left(N^{-1/2}\right). \qquad \text{(B.2.11)}
$$

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

$$
\begin{aligned}
&\tilde{c}_{R,T}^{\top}\left(\tilde{C}_R^{\top}\tilde{C}_R\right)^{-1}\tilde{C}_R^{\top}\eta-\tilde{c}_{P,T}^{\top}\left(\tilde{C}_P^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_P^{\top}\eta\\
=&\tilde{c}_{R,T}^{\top}\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta-\tilde{c}_{P,T}^{\top}\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta+\tilde{c}_{R,T}^{\top}\left[\left(\tilde{C}_R^{\top}\tilde{C}_R\right)^{-1}\tilde{C}_R^{\top}\eta-\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta\right]\\
&\quad-\tilde{c}_{P,T}^{\top}\left[\left(\tilde{C}_P^{\top}\tilde{C}_P\right)^{-1}\tilde{C}_P^{\top}\eta-\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta\right]\\
=&\tilde{c}_{R,T}^{\top}\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta-\tilde{c}_{P,T}^{\top}\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta+O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{N^{\alpha}}{N}\right)+O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right)\\
=&\left(\tilde{c}_{R,T}^{\top}H_R^{-1}-\tilde{c}_{P,T}^{\top}H_P^{-1}\right)\left(C_{G_r}^{\top}C_{G_r}\right)^{-1}C_{G_r}^{\top}\eta+O_p\left(\frac{1}{\delta_{NT}^2}\right)+O_p\left(\frac{N^{\alpha}}{N}\right)+O_p\left(\frac{\sqrt{N^{\alpha}}}{N}\right)\\
=&o_p\left(N^{-1/2}\right),
\end{aligned}
\tag{B.2.12}
$$

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

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

where the third line uses Lemma B.4 (b), the fourth uses Theorem 3.1 (c), and the fifth line uses

Equation (B.2.8), and

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

where the third line uses Lemma B.1 (c), the fourth uses Theorem 3.1 (a), and the fifth uses Equation (B.1.12).

Combining the bias and variance terms, we have

$$
\tilde{c}_{R,T}^{\top}\widehat{\theta}_R-\tilde{c}_{P,T}^{\top}\widehat{\theta}_P=o_p\left(N^{-1/2}\right). \tag{B.2.13}
$$

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

Proof of Theorem 3.2 (b). We organise the proof in the cases of $\alpha\in[0,0.5)$, $\alpha=0.5$, $\alpha\in(0.5,1)$ and $\alpha=1$.

$\nu \in [0, 0.5)$ **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,

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

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

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

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

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

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

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

$$
\begin{aligned}
&E\left[\frac{1}{T_2}E\left[\frac{\eta_{(2)}^\top C_2}{\sqrt{T_2}}\left(\frac{C_2^\top C_2}{T_2}\right)^{-1}c_T\middle| C\right]^2 - \frac{1}{T}E\left[\frac{\eta^\top C}{\sqrt{T}}\left(\frac{C^\top C}{T}\right)^{-1}c_T\middle| C\right]^2\right]\\
=&E\left[\frac{1}{T}tr\left[\frac{1}{1-\pi}Var\left(\frac{C_2^\top\eta_{(2)}}{\sqrt{T_2}}\middle| C\right)\Sigma_{CC}^{-1} - Var\left(\frac{C^\top\eta}{\sqrt{T}}\middle| C\right)\Sigma_{CC}^{-1}\right]\right]\\
=&\frac{1}{T}tr\left(\Omega_{CC,\eta}\Sigma_C^{-1}\right)\frac{\pi}{1-\pi} + o_p\left(\frac{1}{T}\right) > 0, \qquad \text{(B.2.14)}
\end{aligned}
$$

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 $\nu < 1/2$.

$\nu \in [0, 0.5)$ **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

$$
\begin{aligned}
&\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C \theta \right\|^2 \\
=&\left\| \left[- \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T} \right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) - \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \frac{\Lambda_1^\top e_T}{N} \right]^\top \beta \right\|^2 \\
&\quad + o_p\left(\frac{1}{N} \right) \\
=&\left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T} \right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta \right\|^2 \\
&\quad + \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \frac{\Lambda_1^\top e_T}{N} \right]^\top \beta \right\|^2 + o_p\left(\frac{1}{N} \right) \qquad \text{(B.2.15)}
\end{aligned}
$$

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

$$
\begin{aligned}
&\beta^\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 \\
=&\beta^\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 \\
&- \beta^\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 \\
=&\beta^\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 \\
&- \beta^\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} \\
&\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 \\
&- \beta^\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} \\
&\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 \\
&- \beta^\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} \\
&\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 + 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

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

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 $T^{-1}tr(\Omega_{\eta,CC}\Sigma_F^{-1})\frac{\pi}{1-\pi} \asymp_p N^{-1}$, whereas the pseudo-factor method suffers from an additional squared bias

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

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

$\nu \in [0, 0.5)$ **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 $\left\|\left[\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\left(\frac{\tilde{F}_P^\top G_r}{T}\right)^{-1}\frac{\tilde{F}_P^\top G_p}{T}\frac{W^\top W}{N}\left(f_T - \frac{G_p^\top\tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top\tilde{C}_P}{T}\right)^{-1}\tilde{c}_{P,T}\right)\right]^\top\beta\right\|^2$ for the pseudo-factors becomes the leading term, and therefore

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

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

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

$\nu \in [0, 0.5)$ **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

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

The squared bias of the split-sample method is

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

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.2.4 Moderate Rotational Break $\nu = 0.5$

Proof of Theorem 3.2 (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 $\alpha = 1$.

$\nu = 0.5$ **and** $\alpha \in [0, 0.5)$

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

$$\begin{aligned}
&E\left(\operatorname{plim}_{N,T\to\infty}\left\|c_T^\top\theta - \tilde{c}_{P,T}^\top\left(\tilde{C}_P^\top\tilde{C}_P\right)^{-1}\tilde{C}_P^\top C\theta\right\|^2\right)\\
=&E\left(\operatorname{plim}_{N,T\to\infty}\left\|\left[(I-Z)\left(f_T - \frac{G_p^\top\tilde{C}_P}{T}\left(\frac{\tilde{C}_P^\top\tilde{C}_P}{T}\right)^{-1}\tilde{c}_{P,T}\right) - \left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N}\right]^\top\beta\right\|^2\right.\\
&\left.+o_p\left(\frac{1}{N}\right)\right)\\
=&E\left(\operatorname{plim}_{N,T\to\infty}\left\|\left[(I-Z)\left(f_T - \frac{G_p^\top G_r}{T}\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1}c_T\right)\right]^\top\beta\right\|^2\right.\\
&\left.+\operatorname{plim}_{N,T\to\infty}\left\|\left[\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N}\right]^\top\beta\right\|^2\right)\\
&+o_p\left(\frac{1}{N}\right),
\end{aligned} \tag{B.2.16}$$

which follows because

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

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

$$\begin{aligned}
&E\left(\operatorname{plim}_{N,T\to\infty}\left\|c_T^\top\theta - (\tilde{c}_{R,T})^\top\left(\tilde{C}_R^\top\tilde{C}_R\right)^{-1}\tilde{C}_R^\top C\theta\right\|^2\right)\\
=&E\left(\operatorname{plim}_{N,T\to\infty}\left\|\left[(I-Z)\left(f_T - \frac{\tilde{F}_2^\top\tilde{C}_R}{T}\left(\frac{\tilde{C}_R^\top\tilde{C}_R}{T}\right)^{-1}\tilde{c}_{R,T}\right) - \left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N}\right]^\top\beta\right\|^2 + o_p\left(\frac{1}{N}\right)\right)\\
=&E\left(\operatorname{plim}_{N,T\to\infty}\left\|\left[(I-Z)\left(f_T - \frac{H_2^\top F_2^\top C_{G,r}}{T}\left(\frac{C_{G_r}^\top C_{G_r}}{T}\right)^{-1}c_{G_r,T}\right)\right]^\top\beta\right\|^2\right.\\
&\left.+\left\|\left[\left(\frac{\Lambda_1^\top\Lambda_1}{N}\right)^{-1}\frac{\Lambda_1^\top e_T}{N}\right]^\top\beta\right\|^2\right) + o_p\left(\frac{1}{N}\right),
\end{aligned} \tag{B.2.17}$$

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

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

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 3.2 (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.

$\nu = 0.5$ **and** $\alpha = 0.5$

If $\alpha = 0.5$, then the expansion for the rotated factors remains the same as Equation (B.2.17), 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. \\
&\qquad \left.\left. - \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) \\
=&\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 \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 \right\|^2 \\
&+ \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N}\right)^{-1} \frac{\Lambda_1^\top e_T}{N} \right]^\top \beta \right\|^2 \\
&- 2\beta^\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 \\
&- 2\beta^\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 \\
&- 2\beta^\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 \\
&+ o_p\left(\frac{1}{N}\right),
\end{aligned}
\tag{B.2.18}
$$

where the last two cross terms have an expected probability limit of zero, using arguments similarly employed in Section B.2.4. In this case, whether the rotated or the pseudo-factor method depends on the sign of

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

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.

$\nu = 0.5$ **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.

$\nu = 0.5$ **and** $\alpha = 1$

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

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

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

Proof of Theorem 3.2 (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.

$\nu \in (0.5, 1)$ **and** $\alpha < \nu$

The squared bias of the pseudo-factor method is

$$\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C\theta \right\|^2 = \left\| \left[(I - Z) \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta \right\|^2 + o_p\left(N^{2\nu-2} \right)$$
$$\asymp_p N^{2\nu-2}. \tag{B.2.19}$$

The squared bias of the rotated factor method is

$$\left\| c_T^\top \theta - (\tilde{c}_{R,T})^\top \left(\tilde{C}_R^\top \tilde{C}_R \right)^{-1} \tilde{C}_R^\top C\theta \right\|^2 = \left\| \left[(I - Z) \left(f_T - \frac{\tilde{F}_2^\top \tilde{C}_{R,2}}{T} \left(\frac{\tilde{C}_R^\top \tilde{C}_R}{T} \right)^{-1} \tilde{c}_{R,T} \right) \right]^\top \beta \right\|^2 + o_p\left(N^{2\nu-2} \right)$$
$$\asymp_p N^{2\nu-2}. \tag{B.2.20}$$

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.

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

In this case, the expansion of the rotated method remains the same as Equation (B.2.20), but the expansion for the squared bias of the pseudo method has an additional term due to the shift break. Specifically

$$\begin{aligned}
&\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C\theta \right\|^2 \\
=& \left\| \left[\left((I - Z) - \left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T} \right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \right) \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta \right\|^2 \\
&+ o_p\left(N^{2\nu - 2} \right), \qquad \text{(B.2.21)}
\end{aligned}$$

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.

$\nu \in (0.5, 1)$ and $\alpha \in (\nu, 1)$

The expansion for the rotated factors remains the same as Equation (B.2.20). However, for the pseudo-factors, the bias term induced by the shift break is now the leading term, i.e.

$$\begin{aligned}
&\left\| c_T^\top \theta - \tilde{c}_{P,T}^\top \left(\tilde{C}_P^\top \tilde{C}_P \right)^{-1} \tilde{C}_P^\top C\theta \right\|^2 \\
=& \left\| \left[\left(\frac{\Lambda_1^\top \Lambda_1}{N} \right)^{-1} \left(\frac{\tilde{F}_P^\top G_r}{T} \right)^{-1} \frac{\tilde{F}_P^\top G_p}{T} \frac{W^\top W}{N} \left(f_T - \frac{G_p^\top \tilde{C}_P}{T} \left(\frac{\tilde{C}_P^\top \tilde{C}_P}{T} \right)^{-1} \tilde{c}_{P,T} \right) \right]^\top \beta \right\|^2 + o_p\left(N^{2\alpha - 2} \right) \\
=& \asymp_p N^{2\alpha - 2}. \qquad \text{(B.2.22)}
\end{aligned}$$

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

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

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

$\nu \in (0.5, 1)$ **and** $\alpha = 1$

The expansion of the rotated factor estimator remains the same as Equation (B.2.20). The pseudo-factor method is the least effective estimator, because

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

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

$\nu = 1$ **and** $\alpha < 1$

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

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

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

$\nu = 1$ **and** $\alpha = 1$

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

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

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

B.3 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 B.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 $\tilde{\mu}_{t,h} = (T + 1 - 2h)^{-1} \sum_{|j-t|<h} u_j$,*

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

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

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

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

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

We have

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

Second, since

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

then

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

An application of the Delta method then yields

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

■

Proof of Proposition 3.3 and Theorem 3.3. The term $\tilde{r}_{1T}(m)$ can be decomposed further by directly replacing $\tilde{C}(m)$ with $C_H(m)$:

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

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

$$\tilde{r}_{1T}^0(w)\xrightarrow{d}\xi_1(w)=\sum_{m=1}^{3\mathcal{M}} w(m)S_1(m) \tag{B.3.1}$$

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

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

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 B.1 (a), Lemma B.1 (b), Lemma B.4 (a) and Lemma B.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

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

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

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

The second term is negligible because

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

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 3.3. The result in Theorem 3.3 follows immediately. ■

B.4 Additional Simulation Results

This section collects all the additional simulation results, namely for the break fractions of $\pi=0.3$ and $\pi=0.7$. All results are qualitatively similar to the results presented in Chapter 3.

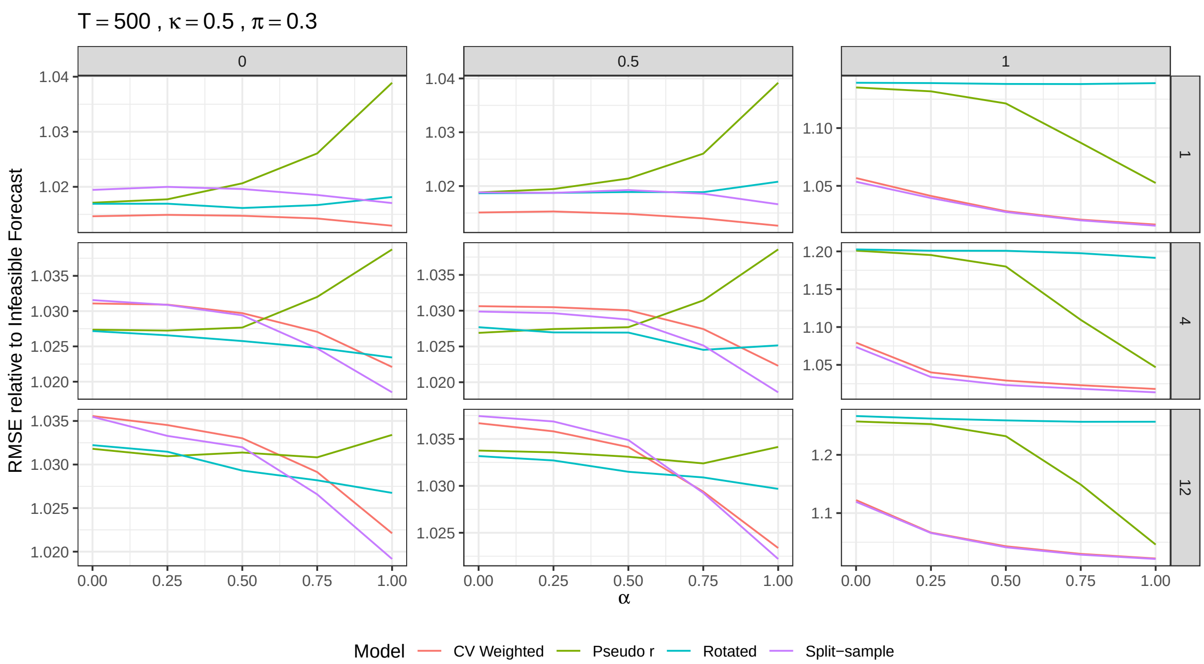

Figure B.1: Relative MSFE for each factor estimator and proposed post-break cross-validation weighted forecasts, faceted by h (rows) and ν (columns), $\kappa = 0.5$ for moderate serial correlation in errors, $q_{max} = p_{max} = 4$, $\pi = 0.3$.

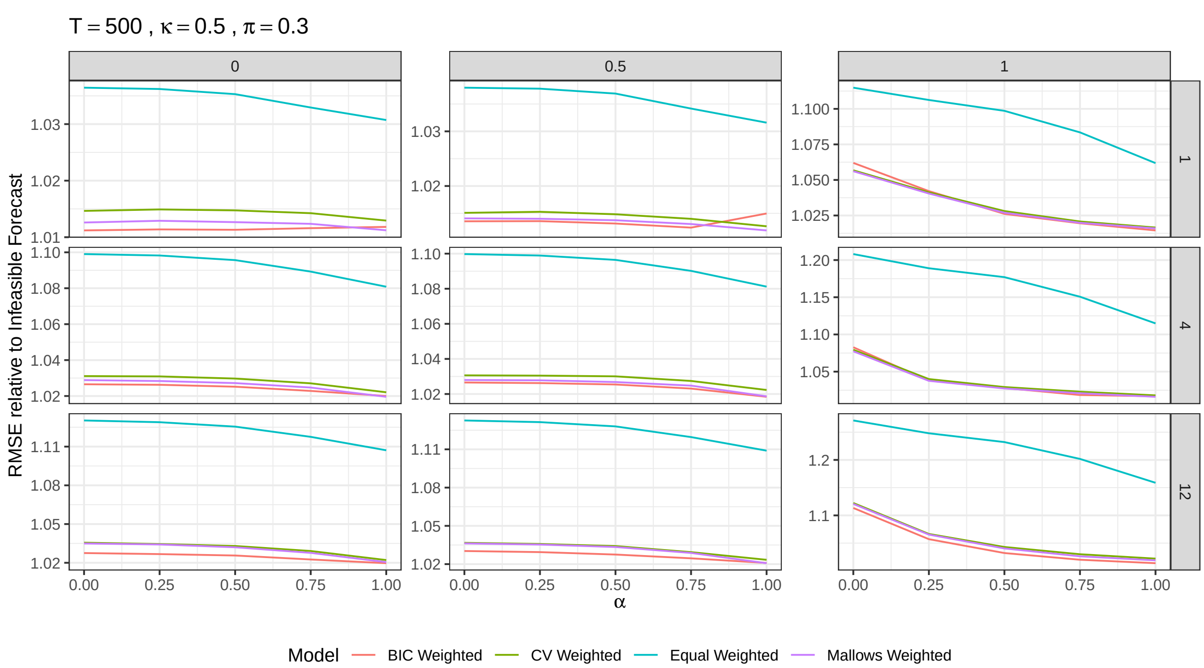

Figure B.2: Relative MSFE for model averaged forecasts, faceted by h (rows) and ν (columns), $\kappa = 0.5$ for moderate serial correlation in errors, $q_{max} = p_{max} = 4$, $\pi = 0.3$.

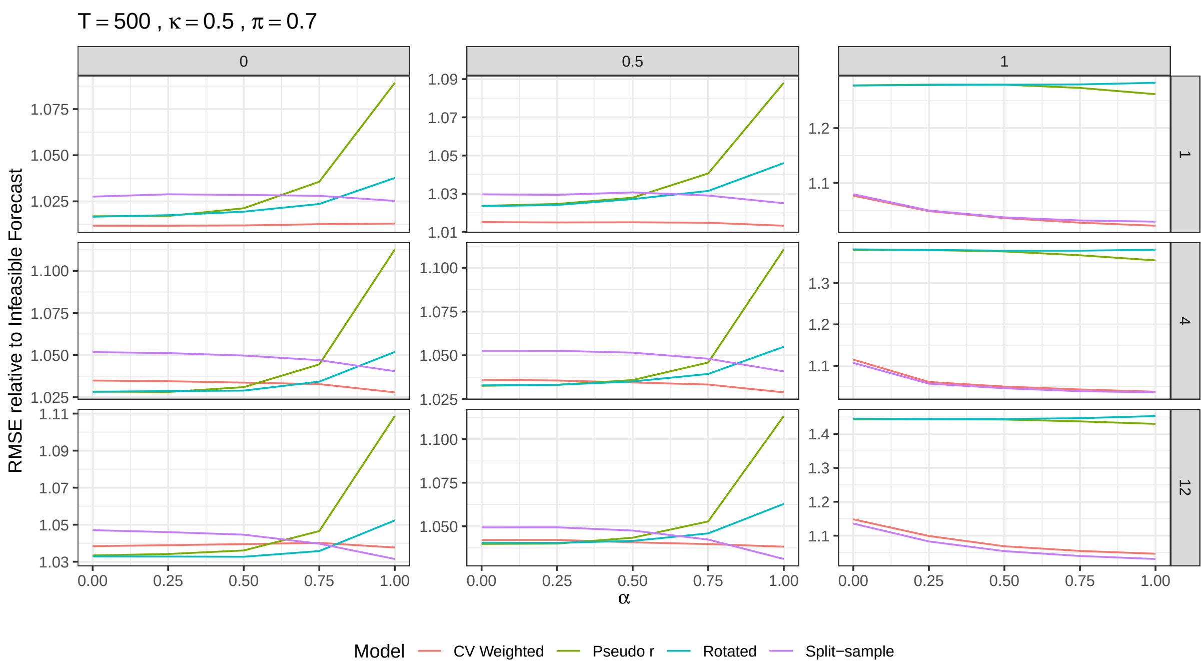

Figure B.3: Relative MSFE for each factor estimator and proposed post-break cross-validation weighted forecasts, faceted by h (rows) and ν (columns), $\kappa = 0.5$ for moderate serial correlation in errors, $q_{max} = p_{max} = 4$, $\pi = 0.7$.

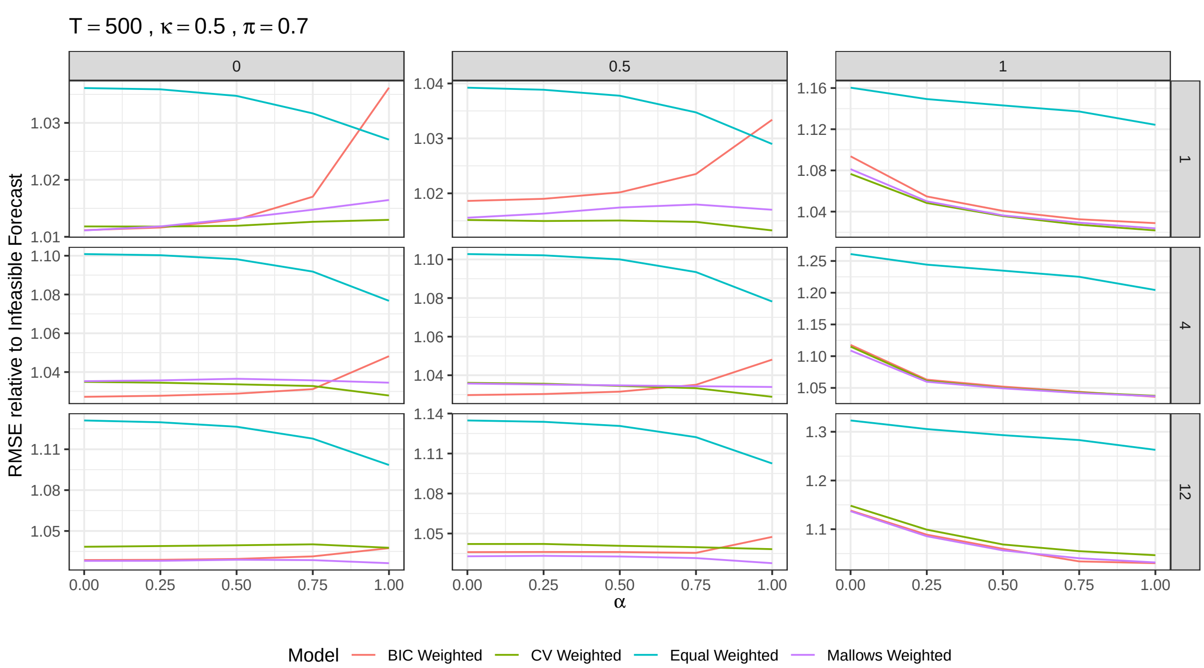

Figure B.4: Relative MSFE for model averaged forecasts, faceted by h (rows) and ν (columns), $\kappa = 0.5$ for moderate serial correlation in errors, $q_{max} = p_{max} = 4$, $\pi = 0.7$.

B.5 Empirical Results

B.5.1 Empirical Robustness Checks

Table B.1 and Table B.2 are the analogous tables for the Stock and Watson (2012b) dataset. The results are qualitatively similar.

Table B.1: 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
BIC Weighted	0.949	0.987	1.014	0.968	0.998	1.018	0.972	0.993	1.008
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 B.2: 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	BIC Weighted	CV Select	CV Weighted	Equal Weighted	Mallows Select	Mallows Weighted	Pseudo r	Rotated	Split-sample
h = 1									
GDP Components	0.990	1.009	1.022	1.035	1.023	1.009	1.025	1.017	1.358
Industrial Production	1.020	1.030	1.023	1.021	1.075	1.000	1.006	1.046	1.238
Employment	0.982	0.972	0.927	0.976	1.064	0.936	0.976	0.956	1.251
Unemployment	1.013	1.005	1.008	0.995	1.042	1.004	1.008	1.020	1.223
Housing	0.987	0.966	0.962	0.954	0.980	0.959	0.961	0.975	1.115
Inventories	0.979	1.032	1.007	1.008	1.040	0.995	1.001	1.035	1.258
Prices	0.947	0.978	0.980	0.976	0.995	0.971	0.995	0.979	1.141
Earnings	0.957	0.998	0.994	0.947	0.995	0.993	0.999	0.984	1.035
Interest Rates	0.986	0.995	1.003	1.080	0.972	1.042	1.080	1.023	1.415
Money	0.969	1.000	1.006	1.008	0.957	1.038	0.995	0.978	1.178
Exchange Rates	0.988	1.005	1.000	1.024	0.987	1.005	0.993	0.991	1.407
Stock Prices	0.967	1.011	1.000	0.974	0.956	0.966	1.004	1.005	1.197
Consumer Expectations	1.018	1.019	1.018	1.113	1.040	1.043	1.007	1.019	1.538
h = 2									
GDP Components	1.008	1.020	1.017	1.039	1.022	1.006	1.010	1.020	1.371
Industrial Production	1.016	1.019	1.019	0.971	1.018	0.980	1.005	1.001	1.110
Employment	0.984	1.015	0.987	0.980	1.064	0.983	0.976	1.007	1.281
Unemployment	1.007	0.999	1.007	1.039	0.999	1.006	1.015	1.016	1.361
Housing	1.026	0.994	0.996	0.979	1.046	0.983	1.000	1.016	1.137
Inventories	0.999	0.983	0.968	0.984	1.035	0.982	1.001	1.016	1.301
Prices	0.986	0.999	1.003	1.032	0.987	1.000	0.999	0.991	1.243
Earnings	0.986	0.993	1.013	1.051	0.990	0.996	0.990	0.986	1.245
Interest Rates	0.969	0.953	0.948	0.991	0.950	0.951	0.990	0.974	1.316
Money	0.998	1.004	0.995	1.003	1.009	0.998	0.998	0.997	1.135
Exchange Rates	0.966	0.996	0.990	1.022	0.960	0.996	0.986	0.981	1.457

Stock Prices	0.983	0.973	0.977	0.964	0.973	0.980	0.993	0.977	1.182
Consumer Expectations	0.962	1.043	1.026	1.051	0.963	1.106	1.035	0.992	1.417
h = 4									
GDP Components	0.989	1.011	1.017	1.048	0.993	1.010	0.999	1.005	1.352
Industrial Production	0.991	1.017	0.979	0.974	0.996	0.969	1.000	0.998	1.150
Employment	0.989	1.031	0.996	0.961	1.027	0.984	0.990	0.989	1.311
Unemployment	1.002	1.016	1.012	0.961	1.025	1.007	1.024	1.022	1.144
Housing	0.983	1.049	1.033	1.026	0.976	1.016	1.047	0.974	1.411
Inventories	0.987	0.976	0.963	0.991	0.982	0.972	1.026	0.985	1.295
Prices	0.993	1.001	1.011	1.045	0.996	1.006	0.999	0.996	1.321
Earnings	1.002	1.007	0.997	1.006	0.996	1.000	1.009	1.014	1.154
Interest Rates	1.003	1.024	1.013	1.136	1.002	1.076	1.069	1.049	1.580
Money	1.001	0.992	0.997	1.027	1.002	1.004	1.002	0.994	1.365
Exchange Rates	0.993	1.047	1.015	1.094	0.988	1.020	1.008	0.994	1.569
Stock Prices	0.984	1.027	0.984	1.007	0.985	1.004	1.016	0.986	1.335
Consumer Expectations	0.990	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.

Appendix C

Appendices for Chapter 4

C.1 Proofs

C.1.1 Preliminary

Lemma C.1. *Under Assumptions 1 to 4,*

a) $\frac{1}{T}\left\|\widehat{F}_t - H_F^\top F_t\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *and* $\frac{1}{T}\left\|\widehat{\mathcal{F}}_f - H_{\mathcal{F}}^\top \mathcal{F}_t\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

b) $\widehat{V}_F \xrightarrow{p} V_F$*, where* V_F *is the diagonal matrix consisting of the eigenvalues of* $\Sigma_F \Sigma_\Lambda$ *in descending order.*

c) H_F *and* $H_{\mathcal{F}}$ *are* $O_p(1)$ *and nonsingular.*

Proof of Lemma C.1.

a) The first equation is the same as Lemma A1 of Bai (2003). The second equation holds by definition of $H_{\mathcal{F}}$ and the first equation.

b) This is Lemma A3 of Bai (2003).

c) $\|H_F\| \leq \left\|\frac{\widehat{F}^\top \widehat{F}}{T}\right\|^{\frac{1}{2}} \left\|\frac{F^\top F}{T}\right\|^{\frac{1}{2}} \left\|\frac{\Lambda^\top \Lambda}{N}\right\| \left\|V_F^{-1}\right\| = O_p(1)$ by Assumptions 1 and 2 and Lemma C.1. The matrices H_F and $H_{\mathcal{F}}$ are nonsingular by Lemma A2 of Han and Inoue (2015).

■

Lemma C.2. *Under Assumptions 1 to 6,*

a) $\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{F}_t - H_F^\top F_t\right)\left[F_t^\top, e_{it}, \eta_t^\top\right] = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *for* $i = 1, \ldots, N$,

b) $\frac{1}{T}\left(\widehat{\mathcal{F}} - \mathcal{F}H_{\mathcal{F}}\right)^\top\left[\mathcal{F}\vdots\eta\right] = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

Proof of Lemma C.2.

a) Lemmas B.1 and B.2 of Bai (2003) imply $\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{F}_t - H_F^\top F_t\right)\left(F_t^\top, e_{it}\right) = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ for $i = 1, \ldots, N$. Lemma 2 a) of Han (2018) shows that $\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{F}_t - H_F^\top F_t\right)\eta_t^\top = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

b) This is the same as Lemma 2 b) of Han (2018).

■

Lemma C.3. *Under Assumptions 1 to 6,* $\frac{1}{T}\sum_{t=p+1}^{T}\left\|\widehat{\eta}_t - H_\eta^\top \eta_t\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

Proof of Lemma C.3. Let $S_1 = \eta\Theta^\top + e$, $S_2 = -P_{\widehat{\mathcal{F}}}(\eta\Theta^\top + e) + M_{\widehat{\mathcal{F}}}\mathcal{F}\Pi^\top$, S_{1t} be the transpose of the tth row of $S1$ and S_{2t} be the transpose of the tth row of S_2. By eigen-identity, we have

$$\widehat{\eta} - \eta H_\eta = \frac{1}{TN}\left(\eta\Theta^\top e^\top + e\Theta\eta^\top + ee^\top + S_1S_2^\top + S_2S_1^\top + S_2S_2\right)\widehat{\eta}\widehat{V}_\eta^{-1} \tag{C.1.1}$$

where $\widehat{V}_\eta$ denotes the diagonal matrix consisting of the first q eigenvalues of $\widehat{X}\widehat{X}^\top/NT$ in descending order. Hence, we obtain

$$\widehat{\eta}_t - H_\eta^\top \eta_t = \frac{1}{TN}\widehat{V}_\eta^{-1}\widehat{\eta}^\top\left(ee_t + \eta\Theta^\top e_t + e\Theta\eta_t + S_2S_{1t} + S_1S_{2t} + S_2S_{2t}\right). \tag{C.1.2}$$

It is sufficient to show that

$$\begin{aligned}
\frac{1}{T}\|c_{lt}\|^2 &= O_p\left(\frac{1}{\delta_{NT}^2}\right) \quad \text{for} \quad l = 1, \ldots, 4,\\
\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{TN}\widehat{\eta}^\top S_2S_{1t}\right\|^2 &= O_p\left(\frac{1}{\delta_{NT}^2}\right),\\
\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{TN}\widehat{\eta}^\top S_1S_{2t}\right\|^2 &= O_p\left(\frac{1}{\delta_{NT}^2}\right), \quad \text{and}\\
\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{TN}\widehat{\eta}^\top S_2S_{2t}\right\|^2 &= O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}$$

The proof of $\frac{1}{T}\sum_{t=1}^{T}\left\|c_{lt}\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ is the same as the proof of Theorem 1 in Bai and Ng (2002).

Note that, since $\widehat{\mathcal{F}}^\top \widehat{X} = 0$, $\widehat{\mathcal{F}}^\top \widehat{\eta} = \frac{1}{TN}\widehat{\mathcal{F}}^\top \widehat{X}\widehat{X}^\top \widehat{\eta}\widehat{V}_\eta^{-1} = 0$, implying that $\widehat{\eta}^\top P_{\widehat{\mathcal{F}}} = 0$ and $\widehat{\eta}^\top M_{\widehat{\mathcal{F}}} = \widehat{\eta}^\top$.

We have the following identities for the terms $\frac{1}{TN}\widehat{\eta}^\top S_2 S_{1t}$, $\frac{1}{TN}\widehat{\eta}^\top S_1 S_{2t}$, and $\frac{1}{TN}\widehat{\eta}^\top S_2 S_{2t}$:

$$\begin{aligned}
\frac{1}{TN}\widehat{\eta}^\top S_2 S_{1t} &= \frac{1}{TN}\widehat{\eta}^\top\left[-P_{\widehat{\mathcal{F}}}\left(\eta\Theta^\top + e\right) + M_{\widehat{\mathcal{F}}}\mathcal{F}\Pi^\top\right](\Theta\eta_t + e_t) \\
&= \frac{1}{TN}\widehat{\eta}^\top \mathcal{F}\Pi^\top\Theta\eta_t + \frac{1}{TN}\widehat{\eta}^\top\mathcal{F}\Pi^\top e_t \\
&= c_{5t} + c_{6t}, \\
\frac{1}{TN}\widehat{\eta}^\top S_1 S_{2t} &= \frac{1}{TN}\widehat{\eta}^\top\left(\eta\Theta^\top + e\right)\left[-\left(\Theta\eta^\top + e^\top\right)\widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t + \Pi\left(\mathcal{F}_t - \mathcal{F}^\top\widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t\right)\right] \\
&= c_{7t} + c_{8t} + c_{9t} + c_{10t} + c_{11t} + c_{12t},
\end{aligned}$$

where

$$\begin{aligned}
c_{7t} &= -\frac{1}{TN}\widehat{\eta}^\top\eta\Theta^\top\Theta\eta^\top\widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t, \\
c_{8t} &= -\frac{1}{TN}\widehat{\eta}^\top\eta\Theta^\top e^\top\widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t, \\
c_{9t} &= \frac{1}{TN}\widehat{\eta}^\top\eta\Theta^\top\Pi\left[\mathcal{F}_t - \widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t\right], \\
c_{10t} &= -\frac{1}{TN}\widehat{\eta}^\top e\Theta\eta^\top\widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t, \\
c_{11t} &= -\frac{1}{TN}\widehat{\eta}^\top ee^\top\widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t, \\
c_{12t} &= \frac{1}{TN}\widehat{\eta}^\top e\Pi\left[\mathcal{F}_t - \widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t\right], \\
\frac{1}{TN}\widehat{\eta}^\top S_2 S_{2t} &= \frac{1}{TN}\widehat{\eta}^\top\mathcal{F}\Pi^\top\left[-\left(\Theta\eta^\top + e^\top\right)\widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t + \Pi\left(\mathcal{F}_t - \widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t\right)\right] \\
&= c_{13t} + c_{14t} + c_{15t} \\
c_{13t} &= -\frac{1}{TN}\widehat{\eta}^\top\mathcal{F}\Pi^\top\Theta\eta^\top\widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t, \\
c_{14t} &= -\frac{1}{TN}\widehat{\eta}^\top\mathcal{F}\Pi^\top e^\top\widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t, \\
c_{15t} &= \frac{1}{TN}\widehat{\eta}^\top\mathcal{F}\Pi^\top\Pi\left[\mathcal{F}_t - \widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top\widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t\right].
\end{aligned}$$

Han (2018) proves that $\frac{1}{T}\sum_{t=1}^{T}\left\|c_{lt}\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ for $l = 5, \ldots, 15$. These quantities are used in subsequent proofs, so are provided here for convenience. ■

Lemma C.4. *Under Assumptions 1 to 6,*

a) $\frac{1}{T}\left(\widehat{\eta}-\eta H_{\eta}\right)^{\top}\eta = O_p\left(\frac{1}{\delta_{NT}^2}\right)$,

b) $\frac{1}{T}\left(\widehat{\eta}-\eta H_{\eta}\right)^{\top}Z = O_p\left(\frac{1}{\delta_{NT}^2}\right)$, *and*

c) $H_{\eta}^{\top} = H_{\eta}^{-1}\Sigma_{\eta}^{-1} + O_p\left(\frac{1}{\delta_{NT}^2}\right)$.

Lemma C.5. *Under Assumptions 1 to 6,* $\frac{1}{T}\sum_{t=1}^{T}\|c_{lt}\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ *for* $l = 1,2,3,4,5,6,8,10,11,12,13,14$ *and* 15 *where each* c_{lt} *are defined in the proof of Lemma C.3.*

Proof of Lemma C.4. Recall that $\widehat{\eta}_t - H_{\eta}^{\top}\eta_t = \widehat{V}_{\eta}^{-1}\sum_{l=1}^{15}c_{lt}$, so it suffices to prove that $\frac{1}{T}\sum_{t=1}^{T}c_{lt}\eta_t = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ for Lemma C.4 (a) and $\frac{1}{T}\sum_{t=1}^{T}c_{lt}Z_t = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ for Lemma C.4 (b), where $l = 1,\dots,15$. Lemma C.4 (a) has been proven by Han (2018), so we focus on Lemma C.4 (b), which can be proven similarly. By the CS-inequality, for $l = 1,2,3,4,5,8,10-15$, we have

$$\left\|\frac{1}{T}\sum_{t=1}^{T}c_{lt}Z_t^{\top}\right\| \leq \left(\sum_{t=1}^{T}\|c_{lt}\|^2\frac{1}{T}\sum_{t=1}^{T}\|Z_t\|^2\right)^{1/2} = O_p\left(\frac{1}{\delta_{NT}^2}\right), \tag{C.1.3}$$

where $\frac{1}{T}\sum_{t=1}^{T}\|c_{lt}\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ for $l = 1,2,3,4,5,6,8,10-15$ by Lemma C.5. Thus, it remains to prove that $\frac{1}{T}\sum_{t=1}^{T}c_{lt}Z_t^{\top} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ for $l = 3,7,9$.

The term $\frac{1}{T}\sum_{t=1}^{T}c_{3t}Z_t^{\top}$ can be bounded by $\frac{1}{TN}\widehat{\eta}^{\top}\eta\Theta^{\top}e_t$

$$\begin{aligned}\left\|\frac{1}{T}\sum_{t=1}^{T}c_{lt}Z_t^{\top}\right\| &\leq \left\|\frac{1}{T}\widehat{\eta}^{\top}\eta\right\|\left\|\frac{1}{TN}\sum_{t=1}^{T}G^{\top}\Lambda e_t Z_t^{\top}\right\| \\ &\leq \left\|\frac{1}{T}\widehat{\eta}^{\top}\eta\right\|\|G\|\left\|\frac{1}{TN}\sum_{t=1}^{T}\sum_{k=1}^{N}\lambda_k e_{kt}Z_t^{\top}\right\| = O_p\left(\frac{1}{TN}\right).\end{aligned}$$

Next, the term $\frac{1}{T}\sum_{t=1}^{T}c_{7t}Z_t^{\top}$ can be rewritten as

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}c_{7t}Z_t^{\top} &= -\frac{\widehat{\eta}^{\top}\eta}{TN}\frac{\Theta^{\top}\Theta}{N}\frac{\eta^{\top}\widehat{\mathcal{F}}}{T}\left(\frac{\widehat{\mathcal{F}}^{\top}\widehat{\mathcal{F}}}{T}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathcal{F}}_t Z_t^{\top} \\ &= O_p\left(\frac{1}{\delta_{NT}}\right)\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathcal{F}}_t Z_t^{\top} \\ &= O_p\left(\frac{1}{\delta_{NT}^2}\right).\end{aligned}$$

Finally, the term $\frac{1}{T}\sum_{t=1}^{T} c_{9t} Z_t^\top$ can be expressed as

$$
\begin{aligned}
&\frac{\widehat{\eta}^\top \eta}{T}\frac{\Theta^\top \Pi}{N}\frac{1}{T}\sum_{t=1}^{T}\left[\mathcal{F}_t - \mathcal{F}^\top \widehat{\mathcal{F}}\left(\widehat{\mathcal{F}}^\top \widehat{\mathcal{F}}\right)^{-1}\widehat{\mathcal{F}}_t\right] Z_t^\top \\
=&O_p(1)\frac{1}{T}\sum_{t=1}^{T} H_{\mathcal{F}}^{-\top}\left(H_{\mathcal{F}}^\top \mathcal{F}_t - \widehat{\mathcal{F}}_t\right) Z_t^\top + O_p(1) H_{\mathcal{F}}^{-\top}\frac{\left(\widehat{\mathcal{F}}^\top - H_{\mathcal{F}}^\top \mathcal{F}^\top\right)\widehat{\mathcal{F}}}{T}\left(\frac{\widehat{\mathcal{F}}^\top \widehat{\mathcal{F}}}{T}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathcal{F}}_t Z_t^\top \\
=&O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}
$$

■

Proof of Lemma C.4 (c). Lemmas C.4 (a) and C.3 imply that

$$
\begin{aligned}
\frac{\widehat{\eta}^\top \widehat{\eta} - H_\eta^\top \eta^\top \eta H_\eta}{T-p} &= \frac{\left(\widehat{\eta} - \eta H_\eta\right)^\top\left(\widehat{\eta} - \eta H_\eta\right) + \left(\widehat{\eta} - \eta H_\eta\right)^\top \eta H_\eta + H_\eta^\top \eta^\top\left(\widehat{\eta} - \eta H_\eta\right)}{T-p} \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\frac{\left(\widehat{\eta}^\top \widehat{\eta} - H_\eta^\top \eta^\top \eta H_\eta\right)}{T-p} &= I_q - \frac{H_\eta^\top \eta^\top \eta H_\eta}{T-p} \\
&= O_p\left(\frac{1}{\delta_{NT}^2}\right).
\end{aligned}
$$

Next, $\eta^\top \eta/(T-p) = \Sigma_\eta$ by Assumption 8, which means

$$
H_\eta^\top = H_\eta^{-1}\Sigma_\eta^{-1} + O_p\left(\frac{1}{\delta_{NT}^2}\right).
$$

■

C.1.2 Main Proofs

Proof of Theorem 4.1. Applying the formula for $\widehat{\delta}$, we have

$$\begin{aligned}\sqrt{T}\left(\widehat{\delta}-\delta\right) &= \left(\mathcal{A}_T W_T \mathcal{A}_T^\top\right)^{-1} \mathcal{A}_T W_T \mathcal{G}_T - \sqrt{T}\delta \\ &= \left(\mathcal{A}_T W_T \mathcal{A}_T^\top\right)^{-1} \mathcal{A}_T W_T \mathcal{B}\end{aligned}$$

where

$$\begin{aligned}\mathcal{A}_T &= I_{q-1} \otimes \left(\frac{1}{T-p}\sum_{t=p+1}^{T} \widehat{\eta}_{1t} Z_t^\top\right), \\ \mathcal{B} &= \frac{1}{\sqrt{T}}\left[\sum_{t=p+1}^{T} \widehat{\eta}_{-1t} \otimes Z_t - \left(I_{q-1} \otimes \sum_{t=1}^{T} Z_t \widehat{\eta}_{1t}\delta\right)\right] \\ &= \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} \left(\widehat{\eta}_{-1t} - \delta\widehat{\eta}_{1t}\right) \otimes Z_t \\ &= \frac{1}{\sqrt{T}}\sum_{t=p+1}^{T} \left(\mathbb{S}_\delta \widehat{\eta}\right) \otimes \left(Z_t - \mu_Z\right).\end{aligned}$$

We need to study the asymptotic distributions of $\mathcal{A}_T$ and $\mathcal{B}$. For $\mathcal{B}$, decompose it into $\mathcal{B} = \mathcal{B}_1 + \mathcal{B}_2$, where

$$\begin{aligned}\mathcal{B}_1 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \left(\mathbb{S}_\delta H_\eta^\top \eta_t\right) \otimes \left(Z_t - \mu_Z\right), \\ \mathcal{B}_2 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \left[\mathbb{S}_\delta \left(\widehat{\eta}_t - H_\eta^\top \eta_t\right)\right] \otimes \left(Z_t - \mu_Z\right).\end{aligned}$$

For $\mathcal{B}_1$, note that $E\left[\mathbb{S}_\delta H_\eta^\top \eta_t \otimes (Z_t - \mu_Z)\right] = 0$ following the moment condition for the estimation of δ and $E(H_\eta^\top \eta_t) = H_\eta^\top E(\eta_t) = 0$. It follows that

$$\begin{aligned}
\mathcal{B}_1 &= \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \left\{ \left(\mathbb{S}_\delta H_\eta^\top \eta_t\right) \otimes (Z_t - \mu_Z) - E\left[\left(\mathbb{S}_\delta H_\eta^\top \eta_t\right) \otimes (Z_t - \mu_Z)\right] \right\} \\
&= \left(\mathbb{S}_\delta H_\eta^\top \otimes I_k\right) \left\{ \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \eta_t \otimes (Z_t - \mu_Z) - E\left[\eta_t \otimes (Z_t - \mu_Z)\right] \right\} \\
&= \left(\mathbb{S}_\delta H_\eta^\top \otimes I_k\right) \frac{1}{\sqrt{T}} \operatorname{vec}\left(Z^\top \eta - E(Z^\top \eta)\right) \\
&\xrightarrow{d} N\left(0_{kq\times 1}, \left(\mathbb{S}_\delta \bar{H}_\eta^\top \otimes I_k\right) \Sigma_i^{(1)} \left(\mathbb{S}_\delta \bar{H}_\eta^\top \otimes I_k\right)^\top\right).
\end{aligned}$$

To study $\mathcal{B}_2$, rewrite it as

$$\begin{aligned}
\mathcal{B}_2 &= \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \left[\mathbb{S}_\delta \left(\widehat{\eta}_t - H_\eta^\top \eta_t\right)\right] \otimes (Z_t - \mu_Z) \operatorname{vec}(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \operatorname{vec}\left((Z_t - \mu_Z) \left(\widehat{\eta}_t - H_\eta^\top \eta_t\right)^\top \mathbb{S}_\delta\right) \\
&= \operatorname{vec}\left(\frac{1}{\sqrt{T}} (Z - \mu_Z)^\top (\widehat{\eta} - \eta H_\eta) \mathbb{S}_\delta\right) \\
&= O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right),
\end{aligned} \tag{C.1.4}$$

where the last equality follows because $\frac{1}{T} Z^\top (\widehat{\eta} - \eta H_\eta) = O_p\left(\frac{1}{\delta_{NT}^2}\right)$. Therefore, $\mathcal{B}_2$ is asymptotically negligible.

To study the limit of $\mathcal{A}_T$, we have

$$\begin{aligned}
\frac{1}{T-p} \sum_{t=p+1}^{T} \widehat{\eta}_{1t} Z_t^\top &= \frac{1}{T-p} \sum_{t=p+1}^{T} \left(H_\eta^\top \eta\right)_{1t} Z_t^\top + \frac{1}{T-p} \sum_{t=p+1}^{T} \left(\widehat{\eta}_{1t} - \left(H_\eta^\top \eta\right)_{1t}\right) Z_t^\top \\
&\xrightarrow{p} E\left(\mathbb{S}_1 H_\eta^\top \eta_t Z_t^\top\right),
\end{aligned}$$

where $\mathbb{S}_1 = \left[1, 0_{1\times(q-1)}\right]$, and $\frac{1}{T-p} \sum_{t=p+1}^{T} \left(\widehat{\eta}_{1t} - (H_\eta^\top \eta)_{1t}\right) Z_t^\top \xrightarrow{p} 0$ using similar arguments used in the proof of $\mathcal{B}_2$. Therefore,

$$\mathcal{A}_T \xrightarrow{p} \mathcal{A} = I_{q-1} \otimes \mathbb{S}_1 \bar{H}_\eta^\top E(\eta_t Z_t^\top). \tag{C.1.5}$$

The weighting matrix W is a full rank matrix and thus invertible. The optimal choice of the weighting matrix follows from standard arguments for GMM estimators.

Collecting the results yields the following distribution as required:

$$\begin{aligned}\sqrt{T}\left(\widehat{\delta}-\delta\right) &\xrightarrow{p} \left(\mathcal{A}\mathcal{W}\mathcal{A}^{\top}\right)^{-1}\mathcal{A}\mathcal{W}\mathcal{B}_1 \\ &\xrightarrow{d} \left(\mathcal{A}\mathcal{W}\mathcal{A}^{\top}\right)^{-1}\mathcal{A}\mathcal{W}N\left(0_{kq\times 1}, \left(\mathbb{S}_{\delta}\bar{H}_{\eta}^{\top}\otimes I_k\right)\Sigma_i^{(1)}\left(\mathbb{S}_{\delta}\bar{H}_{\eta}^{\top}\otimes I_k\right)^{\top}\right),\end{aligned}$$

where $\mathcal{A} = I_{q-1}\otimes \mathbb{S}_1 H_{\eta}^{\top}E(\eta_t Z_t^{\top})$, and $\Sigma_i^{(1)}$ is the upper left block of Σ_i. ■

Proof of Proposition 4.1. By definition,

$$\begin{aligned}\widehat{G} &= \frac{1}{T-p}\sum_{t=p+1}^{T}\widehat{F}_t\widehat{\eta}_t \\ &= \frac{1}{T-p}\sum_{t=p+1}^{T}H_F^{\top}F_t\widehat{\eta}_t + \frac{1}{T-p}\sum_{t=p+1}^{T}\left(\widehat{F}_t - H_F^{\top}F_t\right)\left(\widehat{\eta}_t - H_{\eta}^{\top}\eta_t\right)^{\top} \\ &\quad + \frac{1}{T-p}\sum_{t=p+1}^{T}(\widehat{F}_t - H_F^{\top}F_t)\eta_t^{\top}H_{\eta} \\ &= \frac{1}{T-p}\sum_{t=p+1}^{T}\left(H_F^{\top}\Phi H_{\mathcal{F}}^{-\top}H_{\mathcal{F}}\mathcal{F}_{t-1}\widehat{\eta}_t + H_F^{\top}G\eta_t\widehat{\eta}_t\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right) \\ &= \frac{1}{T-p}\sum_{t=p+1}^{T}\left(H_F^{\top}\Phi H_{\mathcal{F}}^{-\top}H_{\mathcal{F}}\mathcal{F}_{t-1}\widehat{\eta}_t + H_F^{\top}G\eta_t\eta_t^{\top}H_{\eta}\right) + O_p\left(\frac{1}{\delta_{NT}^2}\right).\end{aligned} \tag{C.1.6}$$

Recall that $\eta^{\top}\eta/(T-p)\xrightarrow{p}\Sigma_{\eta}$. By the fact that $\widehat{\mathcal{F}}^{\top}\widehat{\eta}=0$, we have

$$\begin{aligned}\widehat{G} - H_F^{\top}G\Sigma_{\eta}H_{\eta} &= \frac{1}{T-p}H_F^{\top}\Phi H_{\mathcal{F}}^{-\top}\left(\mathcal{F}H_{\mathcal{F}} - \widehat{\mathcal{F}}\right)^{\top} + O_p\left(\frac{1}{\delta_{NT}^2}\right) \\ &= \frac{H_F^{\top}\Phi H_{\mathcal{F}}^{-\top}\left[(\mathcal{F}H_{\mathcal{F}} - \widehat{\mathcal{F}})^{\top}(\widehat{\eta} - \eta H_{\eta}) + (\mathcal{F}H_{\mathcal{F}} - \widehat{\mathcal{F}})^{\top}\eta H_{\eta}\right]}{T-p} \\ &= O_p\left(\frac{1}{\delta_{NT}^2}\right).\end{aligned}$$

■

Proof of Proposition 4.2. Recall that $\widehat{\theta}_i = \widehat{G}^\top \widehat{\lambda}_i$. Therefore,

$$\begin{aligned}
\sqrt{T}\widehat{\theta}_i &= \sqrt{T}\left(H_\eta \Sigma_\eta G^\top H_F^\top \widehat{\lambda}_i\right) + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) \\
&= \left(H_\eta \Sigma_\eta G^\top H_F H_F^{-1} \lambda_i\right) + \left(H_\eta \Sigma_\eta G^\top H_F\right)\sqrt{T}\left(\widehat{\lambda}_i - H_F^{-1}\lambda_i\right) + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right) \\
\sqrt{T}\left(\widehat{\theta}_i - H_\eta^{-1}\theta_i\right) &= \left(H_\eta^{-1} G^\top H_F\right)\sqrt{T}\left(\widehat{\lambda}_i - H_F^{-1}\lambda_i\right) + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right).
\end{aligned}$$

■

Proof of Proposition 4.3 (a).

For the estimator $\widehat{a}_1$, its distribution is based on the distribution of $\widehat{\delta}$. By definition, we have

$$\begin{aligned}
\sqrt{T}(\widehat{a} - a_1^*) &= \sqrt{T}\bar{\mathbb{S}}_1 \begin{bmatrix} 0 \\ \widehat{\delta} - \delta \end{bmatrix} \\
&\xrightarrow{p} \bar{\mathbb{S}}_1 \left(\mathcal{A}\mathcal{W}\mathcal{A}^\top\right)^{-1} \mathcal{A}\mathcal{W}(\mathbb{S}_\delta \bar{H}_\eta^\top \otimes I_k)\frac{1}{\sqrt{T}} \operatorname{vec}\left(Z^\top \eta - E(Z^\top \eta)\right),
\end{aligned}$$

which follows from Theorem 4.1 (a).

For the OLS estimator of the factor loadings $\widehat{\lambda}_i$, Bai (2003) shows that

$$\widehat{\lambda}_i - H_F^{-1}\lambda_i = \frac{1}{T}\bar{H}_F^\top F^\top e_i + O_p\left(\frac{1}{\delta_{NT}^2}\right). \tag{C.1.7}$$

For the estimators $\widehat{\Phi}$ and $\widehat{\Psi}$, Han (2018) shows that

$$\begin{aligned}
\sqrt{T}\operatorname{vec}\left(\widehat{\Phi}^\top - H_{\mathcal{F}}^{-1}\Phi^\top H_F\right) &= \left[H_F^\top G \otimes \left(\frac{H_{\mathcal{F}}^\top \mathcal{F}^\top \mathcal{F} H_{\mathcal{F}}}{T-p}\right)^{-1} H_{\mathcal{F}}^\top\right] \times \frac{\sqrt{T}\sum_{t=p+1}^T vec(\mathcal{F}_t^\top \eta_t)}{T-p} \\
&\quad + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right), \\
\sqrt{T}\operatorname{vec}\left(\widehat{\Psi}_s^\top - H_F^{-1}\Psi_s^\top H_F\right) &= R_s\sqrt{T}\operatorname{vec}\left(\widehat{\Phi}^\top - H_{\mathcal{F}}^{-1}\Phi^\top H_F\right) + o_p(1),
\end{aligned} \tag{C.1.8}$$

where $\bar{R}_s = \sum_{j=1}^s \left(\bar{H}_F^\top \Psi_{j-1}\bar{H}_F^{-\top} \otimes \left[\bar{H}_F^{-\top}\Psi_{s-j}^\top \bar{H}_F, \bar{H}_F^{-\top}\Psi_{s-j-1}^\top \bar{H}_F, \ldots, \bar{H}_F^{-\top}\Psi_{s-j-p+1}^\top \bar{H}_F\right]\right)$ with $\Psi_0 = I_r$ and $\Psi_s = 0_{r\times r}$ for $s < 0$, which follows by (11.7.1) to (11.7.5) of Hamilton (1994).

Combining the above gives

$$\sqrt{T}\begin{bmatrix} \widehat{a}_1 - a_1^* \\ \widehat{\lambda}_i - H_F^{-1}\lambda_i \\ \mathrm{vec}\left(\widehat{\Psi}^\top - H_F^{-1}\Psi^\top H_F\right) \end{bmatrix} = B_s \frac{1}{\sqrt{T}} \begin{bmatrix} \mathrm{vec}\left(Z^\top\eta - E(Z^\top\eta)\right) \\ F^\top e_i \\ \mathrm{vec}\left(\mathcal{F}^\top\eta\right) \end{bmatrix} + o_p(1).$$

■

Proof of Proposition 4.3 (b). This is directly implied by Proposition 4.3 (a). ■

Proof of Theorem 4.2 (a). By adding and subtracting terms, we have

$$\begin{aligned}
&\sqrt{T}\left(\widehat{\theta}_i^\top \widehat{a}_1 - \theta_i a_1\right) \\
&= \sqrt{T}\widehat{\theta}_i^\top\left(\widehat{a} - H_\eta^\top a_1\right) + \sqrt{T}\left(\widehat{\theta}_i^\top - \theta_i^\top H_\eta^{-1}\right) H_\eta^\top a_1 \\
&= \sqrt{T}\widehat{\theta}_i^\top\left(\widehat{a}_1 - a_1^*\right) + \sqrt{T} a_a^\top H_\eta\left(\widehat{\theta}_i - H_\eta^{-1}\theta_i\right) + o_p(1) \\
&= \left[\widehat{\theta}_i^\top [I_q \vdots 0_{q\times r}] + a_1^\top H_\eta H_\eta^{-1} G^\top H_F [0_{r\times qk} \vdots I_r]\right]\sqrt{T}\begin{bmatrix} \widehat{a}_1 - a_1^* \\ \widehat{\lambda}_i - H_F^{-1}\lambda_i \end{bmatrix} + o_p(1).
\end{aligned}$$

Because $\widehat{\theta}_i^\top$ is an estimate of $\theta_i^\top H_\eta^{-\top}$, taking the probability limit and applying Proposition 4.3 (b) yields the result. ■

Proof of Theorem 4.2 (b). By adding and subtracting terms, we have the following asymptotic expansion

$$
\begin{aligned}
&\sqrt{T}\left(\widehat{\lambda}_i^\top \widehat{\Psi}_s \widehat{G}\widehat{a}_1 - \lambda_i^\top H_F^{-\top} H_F^\top \Psi_s H_F^{-\top} H_F^\top G \Sigma_\eta H_\eta H_\eta^{-1} \Sigma_\eta^{-1} a_1\right) \\
&= \sqrt{T}\widehat{\lambda}_i^\top \widehat{\Psi}_s \widehat{G}\left(\widehat{a}_1 - a_1^*\right) + \sqrt{T} a_1^\top G^\top H_F \widehat{\Psi}_s^\top \left(\widehat{\lambda}_i - H_F^{-1}\lambda_i\right) \\
&\quad + \sqrt{T} H_F^{-1} \lambda_i^\top \left(\widehat{\Psi}_s - H_F^\top \Psi_s H_F^{-\top}\right) H_F^\top G a_1^* + o_p(1) \\
&= \sqrt{T}\widehat{\lambda}_i \widehat{\Psi}\widehat{G}\left(\widehat{a}_1 - a_1^*\right) + \sqrt{T} a_1^{*\top} G^\top H_F \widehat{\Psi}_s^\top \left(\widehat{\lambda}_i - H_F^{-1}\lambda_i\right) \\
&\quad + \left(\lambda_i^\top H_F^{-\top} \otimes a_1^{*\top} G^\top H_F\right) \sqrt{T}\, \mathrm{vec}\left(\widehat{\Psi}_s - H_F^\top \Psi_s H_F^{-\top}\right) + o_p(1) \\
&= \Big[\widehat{\lambda}_i^\top \widehat{\Psi}_s \widehat{G}[I_q \vdots 0_{q\times r} \vdots 0_{q \times r^2}] + a_1^{*\top} G^\top H_F \widehat{\Psi}_s^\top [0_{r\times q} \vdots I_r \vdots 0_{r\times r^2}] \qquad \text{(C.1.9)} \\
&\quad + \left(\lambda_i^\top H_F^{-\top}\right) \otimes \left(a_1^{*\top} G^\top H_F\right)[0_{r^2\times q} \vdots 0_{r^2\times r} \vdots I_{r^2}]\Big] \sqrt{T} \begin{bmatrix} \widehat{a}_1 - a_1^* \\ \widehat{\lambda}_i - H_F^{-1}\lambda_i \\ \mathrm{vec}\left(\widehat{\Psi}_s^\top - H_F^{-1}\Psi_s^\top H_F\right) \end{bmatrix} + o_p(1).
\end{aligned}
$$

Hence, we have

$$
\sqrt{T}\left(\widehat{\lambda}_i^\top \widehat{\Psi}_s \widehat{G}\widehat{a}_1 - \lambda_i^\top \Psi_s G a_1\right) \xrightarrow{d} N\left(0, \bar{Q}_{2,i} B_s \Sigma_i B_s^\top \bar{Q}_{2,i}^\top\right),
$$

where

$$
\bar{Q}_{2,i} = \lambda_i^\top \Psi_s G \Sigma_\eta \bar{H}_\eta C_3 + a_1^\top G^\top \Psi_s^\top \bar{H}_F C_4 + \left(\lambda_i^\top \bar{H}_F^{-\top} \otimes a_1^\top G^\top \bar{H}_F\right) C_5,
$$

and $C_3 = [I_q \vdots 0_{q\times r} \vdots 0_{q\times r^2}]$, $C_4 = [0_{r\times q} \vdots I_r \vdots 0_{r\times r^2}]$ and $C_5 = [0_{r^2\times q} \vdots 0_{r^2\times r} \vdots I_{r^2}]$. Applying the distribution in Proposition 4.3 (a) yields the result. ■

Proof of Theorem 4.3 (a). The proof follows standard methods for proving overidentification tests. Note that, the sample moment condition, derivative, and first order condition are respectively

$$
\mathcal{G}_T = \frac{1}{T-p} \sum_{t=p+1}^{T} \left(\widehat{\eta}_{-1t} \otimes Z_t\right), \qquad \text{(C.1.10)}
$$

$$
\mathcal{A}_T = I_{q-1} \otimes \left(\frac{1}{T-p} \sum_{t=p+1}^{T} \widehat{\eta}_{1t} Z_t^\top\right), \qquad \text{(C.1.11)}
$$

$$
\mathcal{A}_T W_T \mathcal{G}_T = 0. \qquad \text{(C.1.12)}
$$

Let $\widehat{\delta}$ denote a GMM estimator obtained with an optimal weight matrix W_T, i.e. $W_T \xrightarrow{p} W$, which is $\mathcal{B}\Sigma_i^{(1)}\mathcal{B}^\top$. Expand the moment condition about $\mathcal{G}_T(\delta)$ to obtain

$$\mathcal{G}_T(\widehat{\delta}) = \mathcal{G}_T(\delta) + \mathcal{A}_T(\delta^*)\left(\widehat{\delta} - \delta\right) + o\left\|\widehat{\delta} - \delta\right\| \tag{C.1.13}$$

where $\|\delta^* - \delta\| \leq \left\|\widehat{\delta} - \delta\right\|$. Substituting this back into the first order condition yields

$$\begin{aligned}\mathcal{A}_T(\widehat{\delta})^\top W_T^{-1}\delta^*\left(\widehat{\delta} - \delta\right) &= -\mathcal{A}_T(\widehat{\delta})^\top W_T^{-1}\mathcal{G}_T(\delta)\\ \left(\widehat{\delta} - \delta\right) &= -\left(\mathcal{A}_T\left(\widehat{\delta}\right)^\top W_T^{-1}\mathcal{A}_T(\delta^*)\right)^{-1}\mathcal{A}_T(\widehat{\delta})^\top W_T^{-1}\mathcal{G}_T(\delta).\end{aligned}$$

Substituting this back into the Taylor expansion gives

$$\begin{aligned}\mathcal{G}_T\left(\widehat{\delta}\right) &= \mathcal{G}_T(\delta) + \mathcal{A}_T(\delta^*)\left(\widehat{\delta} - \delta\right) + o_p(1)\\ &= \mathcal{G}_T(\delta) - \mathcal{A}_T(\delta^*)\left(\mathcal{A}_T\left(\widehat{\delta}\right)^\top W_T^{-1}\mathcal{A}_T(\delta^*)\right)^{-1}\mathcal{A}_T\left(\widehat{\delta}\right)^\top W_T^{-1}\mathcal{G}_T(\delta)\\ &= \left(I - \mathcal{A}_T(\delta^*)\left(\mathcal{A}_T\left(\widehat{\delta}\right)^\top W_T^{-1}\mathcal{A}_T(\delta^*)\right)^{-1}\mathcal{A}_T\left(\widehat{\delta}\right)^\top W_T^{-1}\right)\mathcal{G}_T(\delta).\end{aligned}$$

Because $\mathcal{A}_T\left(\widehat{\delta}\right) \xrightarrow{p} \mathcal{A}_T(\delta)$, $\mathcal{A}_T(\delta^*) \xrightarrow{p} \mathcal{A}_T(\delta)$ by the proof of Theorem 4.1 (a), and $W_T \xrightarrow{p} W$, the Cramer-Wold device and Slutsky's Theorem yield

$$\sqrt{T}W_T^{1/2}\mathcal{G}_T\left(\widehat{\delta}\right) \xrightarrow{p} \left(I - W^{-1/2}\mathcal{A}\left(\mathcal{A}^\top W^{-1}\mathcal{A}\right)^{-1}\mathcal{A}^\top W^{-1/2}\right)\mathbf{Z}_T$$

where $\mathbf{Z}_T = \sqrt{T}W^{-1/2}\mathcal{G}(\delta) \xrightarrow{d} \mathbf{Z} \sim N(0, I)$. It follows that by recognising that $\text{plim}\left|T\mathcal{Q}_T(\delta) - \mathbf{Z}_T^\top\mathbf{Z}_T\right| = 0$,

$$T\mathcal{Q}_T\left(\widehat{\delta}\right) \xrightarrow{d} \mathbf{Z}^\top\left(I - W^{-1/2}\mathcal{A}\left(\mathcal{A}^\top W^{-1}\mathcal{A}\right)^{-1}\mathcal{A}^\top W^{-1/2}\right)\mathbf{Z} = \chi^2_{(k-1)(q-1)}.$$

■

Proof of Theorem 4.3 (b) and Theorem 4.3 (c).

These follow the Proofs of Theorem 1 and 2 of Andrews (1999), respectively. ■

C.2 Additional Simulation Results

Table C.1: Coverage Probabilities

T	N	h	$k=1$	$k=2$	$k=3$	$k=4$
250	125	0	0.919	0.897	0.889	0.881
		1	0.931	0.917	0.909	0.903
		2	0.950	0.942	0.939	0.935
		3	0.954	0.949	0.947	0.944
		6	0.941	0.938	0.938	0.935
		9	0.936	0.931	0.929	0.926
		12	0.929	0.925	0.924	0.922
	250	0	0.920	0.904	0.895	0.888
		1	0.937	0.924	0.916	0.911
		2	0.953	0.946	0.941	0.938
		3	0.954	0.949	0.945	0.944
		6	0.939	0.935	0.932	0.931
		9	0.934	0.927	0.925	0.923
		12	0.931	0.925	0.921	0.919
500	125	0	0.903	0.890	0.883	0.880
		1	0.912	0.901	0.892	0.887
		2	0.943	0.938	0.934	0.932
		3	0.946	0.943	0.941	0.940
		6	0.940	0.937	0.934	0.934
		9	0.931	0.926	0.924	0.923
		12	0.918	0.915	0.913	0.913
	250	0	0.909	0.898	0.892	0.889
		1	0.928	0.922	0.917	0.915
		2	0.947	0.943	0.941	0.940
		3	0.948	0.945	0.943	0.942
		6	0.939	0.935	0.933	0.932
		9	0.926	0.923	0.919	0.918
		12	0.916	0.913	0.911	0.910

Note:
Entries report the coverage rate of the IRFs using the proposed asymptotic distributions (nominal 95%).

Table C.2: RMSE ratios

T	N	h	$k=2$	$k=3$	$k=4$	SVAR-IV ($k=4$)
250	125	0	0.923	0.911	0.892	6.150
		1	0.950	0.942	0.940	5.729
		2	0.967	0.950	0.945	4.554
		3	0.968	0.956	0.953	3.688
		6	0.995	0.994	0.993	2.396
		9	1.008	1.003	1.001	2.151
		12	1.006	1.004	0.998	2.263
	250	0	0.924	0.919	0.889	6.340
		1	0.941	0.935	0.922	5.649
		2	0.960	0.954	0.942	4.528
		3	0.963	0.951	0.935	3.706
		6	0.996	0.978	0.971	2.322
		9	1.009	0.988	0.984	2.118
		12	1.009	0.987	0.990	2.254
500	125	0	0.940	0.914	0.902	8.293
		1	0.965	0.962	0.964	7.935
		2	0.962	0.958	0.959	6.144
		3	0.973	0.969	0.966	4.874
		6	1.002	1.003	0.994	3.148
		9	1.006	1.005	0.995	2.995
		12	1.008	1.012	1.003	3.445
	250	0	0.941	0.913	0.901	8.711
		1	0.963	0.947	0.949	8.017
		2	0.972	0.958	0.957	6.235
		3	0.979	0.970	0.963	4.934
		6	1.005	1.001	0.994	3.085
		9	1.015	1.013	1.003	2.925
		12	1.009	1.009	0.998	3.267

Note:
Entries report the RMSE of the estimated IRFs of the overidentified system, compared to the RMSE of the IRFs of the just-identified system.

Appendix D

Bibliography

Aastveit, K. A., Bjørnland, H. C., and Thorsrud, L. A. (2015). What drives oil prices? Emerging versus developed economies. *Journal of Applied Econometrics*, 30(7):1013–1028.

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue Ratio Test for the Number of Factors. *Econometrica*, 81(3):1203–1227.

Alessi, L. and Kerssenfischer, M. (2019). The response of asset prices to monetary policy shocks: Stronger than thought. *Journal of Applied Econometrics*, 34(5):661–672.

Amengual, D. and Watson, M. W. (2007). Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel. *Journal of Business & Economic Statistics*, 25(1):91–96.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, volume 2. Wiley New York, 3 edition.

Anderson, T. W. and Rubin, H. (1956). Statistical Inference in Factor Analysis. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, page 111. University of California Press.

Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica*, 61(4):821–856.

Andrews, D. W. K. (1999). Consistent Moment Selection Procedures for Generalized Method of Moments Estimation. *Econometrica*, 67(3):543–563.

Antolín-Díaz, J. and Rubio-Ramírez, J. F. (2018). Narrative Sign Restrictions for SVARs. *American Economic Review*, 108(10):2802–2829.

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

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

Bai, J., Duan, J., and Han, X. (2024). The likelihood ratio test for structural changes in factor models. *Journal of Econometrics*, 238(2):105631.

Bai, J. and Han, X. (2016). Structural Changes in High Dimensional Factor Models. *Frontiers of Economics in China*, 11(1):9–39.

Bai, J., Han, X., and Shi, Y. (2020). Estimation and inference of change points in high-dimensional factor models. *Journal of Econometrics*, 219(1):66–100.

Bai, J., Li, K., and Lu, L. (2016). Estimation and Inference of FAVAR Models. *Journal of Business & Economic Statistics*, 34(4):620–641.

Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2004). A PANIC Attack on Unit Roots and Cointegration. *Econometrica*, 72(4):1127–1177.

Bai, J. and Ng, S. (2006). Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2007). Determining the Number of Primitive Shocks in Factor Models. *Journal of Business & Economic Statistics*, 25(1):52–60.

Bai, J. and Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, 24(4):607–629.

Bai, J. and Ng, S. (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*, 235(2):1893–1916.

Bai, J. and Wang, P. (2012). Identification and estimation of dynamic factor models. https://mpra.ub.uni-muenchen.de/38434/.

Bai, J. and Wang, P. (2015). Identification and Bayesian Estimation of Dynamic Factor Models. *Journal of Business & Economic Statistics*, 33(2):221–240.

Bai, J. and Wang, P. (2016). Econometric Analysis of Large Factor Models. *Annual Review of Economics*, 8(1):53–80.

Bailey, N., Kapetanios, G., and Pesaran, M. H. (2021). Measurement of factor strength: Theory and practice. *Journal of Applied Econometrics*, 36(5):587–613.

Baltagi, B. H., Kao, C., and Wang, F. (2017). Identification and estimation of a large factor model with structural instability. *Journal of Econometrics*, 197(1):87–100.

Baltagi, B. H., Kao, C., and Wang, F. (2021). Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics*, 220(2):349–365.

Banerjee, A., Marcellino, M., and Masten, I. (2008). Chapter 4 Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change. In E. Rapach, D. and E. Wohar, M., editors, *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, volume 3 of *Frontiers of Economics and Globalization*, pages 149–194. Emerald Group Publishing Limited.

Barakchian, S. M. and Crowe, C. (2013). Monetary policy matters: Evidence from new shocks data. *Journal of Monetary Economics*, 60(8):950–966.

Barigozzi, M., Cho, H., and Fryzlewicz, P. (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics*, 206(1):187–225.

Barigozzi, M. and Luciani, M. (2023). Measuring the Output Gap using Large Datasets. *The Review of Economics and Statistics*, pages 1–15.

Barigozzi, M. and Trapani, L. (2020). Sequential testing for structural stability in approximate factor models. *Stochastic Processes and their Applications*, 130(8):5149–5187.

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

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

Bernanke, B. S. and Mihov, I. (1998). Measuring Monetary Policy*. *The Quarterly Journal of Economics*, 113(3):869–902.

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

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

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer Science & Business Media.

Chamberlain, G. (1983). Funds, Factors, and Diversification in Arbitrage Pricing Models. *Econometrica*, 51(5):1305.

Chan, J. C. C., Eisenstat, E., and Strachan, R. W. (2020). Reducing the state space dimension in a large TVP-VAR. *Journal of Econometrics*, 218(1):105–118.

Chen, L., Dolado, J. J., and Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, 180(1):30–48.

Cheng, X., Han, X., and Inoue, A. (2021). Instrumental Variable Estimation of Structural Var Models Robust to Possible Nonstationarity. *Econometric Theory*, pages 1–30.

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

Cheng, X., Liao, Z., and Schorfheide, F. (2016). Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities. *The Review of Economic Studies*, 83(4):1511–1543.

Christiano, L., Eichenbaum, M., and Evans, C. (1998). Monetary Policy Shocks: What Have We Learned and to What End? Technical Report w6400, National Bureau of Economic Research, Cambridge, MA.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, 113(1):1–45.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302.

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

Duan, J., Bai, J., and Han, X. (2022). Quasi-maximum likelihood estimation of break point in high-dimensional factor models. *Journal of Econometrics*, page S0304407622000379.

Eickmeier, S. (2009). Comovements and heterogeneity in the euro area analyzed in a non-stationary dynamic factor model. *Journal of Applied Econometrics*, 24(6):933–959.

Forni, M. and Gambetti, L. (2010a). The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics*, 57(2):203–216.

Forni, M. and Gambetti, L. (2010b). Fiscal Foresight and the Effects of Goverment Spending.

Forni, M. and Gambetti, L. (2010c). Macroeconomic Shocks and the Business Cycle: Evidence from a Structural Factor Model.

Forni, M. and Gambetti, L. (2014). Sufficient information in structural VARs. *Journal of Monetary Economics*, 66:124–136.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the Black Box: Structural Factor Models with Large Cross Sections. *Econometric Theory*, 25(5):1319–1347.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. *The Review of Economics and Statistics*, 82(4):540–554.

Fortin, A.-P., Gagliardini, P., and Scaillet, O. (2024). Latent Factor Analysis in Short Panels.

Fu, Z., Su, L., and Wang, X. (2023). Estimation and Inference on Time-Varying FAVAR Models. *Journal of Business & Economic Statistics*, 0(0):1–15.

Gafarov, B. (2014). Identification in Dynamic Models Using Sign Restrictions.

Gafarov, B., Meier, M., and Montiel Olea, J. L. (2018). Delta-method inference for a class of set-identified SVARs. *Journal of Econometrics*, 203(2):316–327.

Gertler, M. and Karadi, P. (2015). Monetary Policy Surprises, Credit Costs, and Economic Activity. *American Economic Journal: Macroeconomics*, 7(1):44–76.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.

Giacomini, R. and Kitagawa, T. (2021). Robust Bayesian Inference for Set-Identified Models. *Econometrica*, 89(4):1519–1556.

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

Gonçalves, S. and Perron, B. (2020). Bootstrapping factor models with cross sectional dependence. *Journal of Econometrics*, 218(2):476–495.

Granziera, E., Moon, H. R., and Schorfheide, F. (2018). Inference for VARs identified with sign restrictions. *Quantitative Economics*, 9(3):1087–1121.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton university press.

Han, X. (2015). Tests for overidentifying restrictions in Factor-Augmented VAR models. *Journal of Econometrics*, 184(2):394–419.

Han, X. (2018). Estimation and inference of dynamic structural factor models with over-identifying restrictions. *Journal of Econometrics*, 202(2):125–147.

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

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

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

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

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Koo, B., Wong, B., and Zhong, Z.-Y. (2023). Disentangling Structural Breaks in High Dimensional Factor Models. https://arxiv.org/abs/2303.00178v1.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

Lawley, D. N. and Maxwell, A. E. (1962). Factor Analysis as a Statistical Method. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 12(3):209–229.

Lubik, T. A. and Schorfheide, F. (2004). Testing for Indeterminacy: An Application to U.S. Monetary Policy. *American Economic Review*, 94(1):190–217.

Luciani, M. (2015). Monetary Policy and the Housing Market: A Structural Factor Analysis. *Journal of Applied Econometrics*, 30(2):199–218.

Ma, S. and Su, L. (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics*, 207(1):1–29.

Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. *Journal of Econometrics*, 197(1):101–129.

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

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

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

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

Miranda-Agrippino, S. and Ricco, G. (2021). The Transmission of Monetary Policy Shocks. *American Economic Journal: Macroeconomics*, 13(3):74–107.

Montiel Olea, J. L., Stock, J. H., and Watson, M. W. (2021). Inference in Structural Vector Autoregressions identified with an external instrument. *Journal of Econometrics*, 225(1):74–87.

Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703–708.

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

Onatski, A. (2009). Testing Hypotheses About the Number of Factors in Large Factor Models. *Econometrica*, 77(5):1447–1479.

Onatski, A. (2010). Determining the Number of Factors from Empirical Distribution of Eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016.

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

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

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). Forecasting Time Series Subject to Multiple Structural Breaks. *The Review of Economic Studies*, 73(4):1057–1084.

Pesaran, M. H., Pick, A., and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177(2):134–152.

Phillips, P. C. B. and Moon, H. R. (1999). Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica*, 67(5):1057–1111.

Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, 72(3):821–852.

Ramey, V. A. (2016). Macroeconomic Shocks and Their Propagation.

Rigobon, R. (2003). Identification Through Heteroskedasticity. *The Review of Economics and Statistics*, 85(4):777–792.

Romer, C. D. and Romer, D. H. (2004). A New Measure of Monetary Shocks: Derivation and Implications. *THE AMERICAN ECONOMIC REVIEW*, 94(4).

Sargent, T. and Sims, C. (1977). Business cycle modeling without pretending to have too much a priori economic theory. Working Paper 55, Federal Reserve Bank of Minneapolis.

Schlaak, T., Rieth, M., and Podstawski, M. (2023). Monetary policy, external instruments, and heteroskedasticity. *Quantitative Economics*, 14(1):161–200.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1):1–48.

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36(5):975–1000.

Sims, C. A. and Zha, T. (2006). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review*, 96(1):54–81.

Smets, F. and Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3):586–606.

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

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

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

Stock, J. H. and Watson, M. W. (2005). Implications of Dynamic Factor Models for VAR Analysis. Working Paper 11467, National Bureau of Economic Research.

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

Stock, J. H. and Watson, M. W. (2012a). Disentangling the channels of the 2007-09 recession. *Brookings Papers on Economic Activity*, (1):81–135.

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

Stock, J. H. and Watson, M. W. (2016). Chapter 8 - Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 415–525. Elsevier.

Stock, J. H. and Watson, M. W. (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *The Economic Journal*, 128(610):917–948.

Stock, J. H. and Watson, M. W. (2021). Comovement of Economic Activity During the Covid Recession. Unpublished manuscript.

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

Su, L. and Wang, X. (2020). Testing for Structural Changes in Factor Models Via a Nonparametric Regression. *Econometric Theory*, 36(6):1127–1158.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

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

White, H. (1984). *Asymptotic Theory for Econometricians*. Academic press.

Wieland, J. F. and Yang, M.-J. (2020). Financial Dampening. *Journal of Money, Credit and Banking*, 52(1):79–113.

Yamamoto, Y. (2019). Bootstrap inference for impulse response functions in factor-augmented vector autoregressions. *Journal of Applied Econometrics*, 34(2):247–267.

Yamamoto, Y. and Hara, N. (2022). Identifying factor-augmented vector autoregression models via changes in shock variances. *Journal of Applied Econometrics*, 37(4):722–745.

Yamamoto, Y. and Tanaka, S. (2015). Testing for factor loading structural change under common breaks. *Journal of Econometrics*, 189(1):187–206.

Yu, L., He, Y., Kong, X., and Zhang, X. (2022). Projected estimation for large-dimensional matrix factor models. *Journal of Econometrics*, 229(1):201–217.