University of Windsor

Scholarship at UWindsor

Major Papers | Theses, Dissertations, and Major Papers

October 2018

GROUP-LASSO ESTIMATION IN HIGH-DIMENSIONAL FACTOR MODELS WITH STRUCTURAL BREAKS

Yujie Song
University of Windsor, song12f@uwindsor.ca

Group-Lasso Estimation in High-Dimensional Factor Models with Structural Breaks

by

Yujie Song

A Major Research Paper
Submitted to the Faculty of Graduate Studies
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Master of Science at the
University of Windsor

Windsor, Ontario, Canada

Group-Lasso Estimation in High-Dimensional Factor Models with Structural Breaks

by

Yujie Song

APPROVED BY:

M. Hlynka
Department of Mathematics and Statistics

S. Nkurunziza, Advisor
Department of Mathematics and Statistics

September 19, 2018

Author's Declaration of Originality

I hereby certify that I am the sole author of this major paper and that no part of this major paper has been published or submitted for publication.

I certify that, to the best of my knowledge, my major paper does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my major paper, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my major paper and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my major paper, including any final revisions, as approved by my major paper committee and the Graduate Studies office, and that this major paper has not been submitted for a higher degree to any other University or Institution.

Abstract

In this major paper, we study the influence of structural breaks in the financial market model with high-dimensional data. We present a model which is capable of detecting changes in factor loadings, determining the number of factors and detecting the break date. We consider the case where the break date is both known and unknown and identify the type of instability. For the unknown break date case, we propose a group-LASSO estimator to determine the number of pre- and post-break factors, the break date and the existence of instability of factor loadings when the number of factor is constant. We also present the asymptotic properties of penalized least square estimator with both the cross-sections and the time dimensions tend to infinity.

Further, we develop a cross-validation procedure to obtain the tuning parameters to fine-tune the penalty terms and use the least square approach to estimate the break date after the number of factors is obtained. We also present a Monte Carlo simulation to evaluate the performance of the proposed procedure and analyze real data from 2007-09 Great Recession. The proposed procedure generally detects the break date correctly during the Great Recession while the procedure performs relatively poorly in estimating the number of factors in the pre- and post-break date case.

To my loving parents

Yifan Song and Minglin Li

Acknowledgments

I would like to thank my supervisor Dr. Sévérien Nkurunziza for all his constant support and guidance during my graduate study since I started my M.Sc. program, especially in every stage of this major paper. During my program, he never hesitated to provide me suggestions whenever I had difficulties in study and research. These support and suggestions made me finally succeed in this work.

I would also like to thank Dr. Myron Hlynka for being my department reader. He also taught me the basic knowledge of stochastic processes and regression analysis, which helped me build a solid foundation for further studies in Statistics.

In addition, I would like to thank all faculty and staff members, and the graduate students in the Department of Mathematics and Statistics who helped me in many different ways during my study.

Moreover, my profound gratitude goes to my parents. They have given me constant love, encouragement and support since my birth, and have guided me to face my life positively and go ahead in the right direction. Finally, I would like to thank everyone who directly or indirectly helped me in the completion of this major paper.

Contents

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

This major paper studies the influence of structural breaks in the financial market model with high-dimensional data. In classical economic data sets, statistical models are considered in a low-dimensional data setting since the number of records is larger than the number of covariates. Thus, the classical statistic techniques are only applicable for low-dimensional data. Briefly, high-dimensional statistical analysis refers to the situation where the number of unknown parameters is larger than the number of samples in the data (Peter Bühlmann et al. 2011)[9]. As explained in Sunil Sapra (2015)[22], modern data in economics involves millions of records on individuals. Therefore, the high-dimensional data models have become a necessity in the financial market to analyze data on the massive amount of features for a limited number of individuals.

On top of the high-dimensional data setting, we also consider the scenario where the data have the structural breaks. Indeed, in macroeconomics, the structural breaks demonstrate themselves in time series data for various reasons, for instance, economic crises, policy changes, and regime shifts (S. Chancharat et al. 2007)[14]. We apply the model to panel data in the 2007-2009 Great Recession. The panel data used widely in economics also offers an application of high-dimensional data analysis (Sunil Sapra 2015)[22]. Perron (1989)[20] argues that if the structural breaks are not specified appropriately, we may obtain the spurious results. Indeed, ignoring the break points often leads to unexpected consequences. First, the number of pre- and post-break factors will be overestimated. Second, this action will misinterpret the later analysis on economy associated with the number of factors. Although there exists some work that is related to this topic, our study involves in-depth analysis of structural breaks.

1.2 Existing Studies and Their Limitations

In this subsection, we indicate the limitations of existing studies, which include locating the break date, detecting the change in loadings, determining the number of factors. First, existing methods could determine the number of factors as given in Bai and Ng (2002)[4], Onatski (2010)[19], but the break date is required and the methods are unable to detect the change in loadings. Second, the work of Breitung and Eickmeier (2011)[8] does not provide the estimation of some pre- and post-break factors and is unable to detect a change in the number of factors. Also, we quote the work of Bai (1997)[5] which requires the number of factors to determine the break date by using residual-based procedures. Third, a limitation of the method given in Bai and Ng (2002)[4] is that we have to know the break date. However, if the break date is unknown, the number of factors will be overestimated when applying those methods. Fourth, in the recent work of Bai and Liao (2015)[6], and Caner and Han (2014)[10], they use shrinkage methods to estimate the stable models and detect structural breaks in the model.

The work of Cheng et al. (2016)[13] improves the approach to cope with the unknown break date case compared with the work mentioned above. Assuming that the break date is unknown, the proposed methods simultaneously estimate the number of pre- and post-break factors, and determine changes in factor loading if the number of factors stays constant. Meanwhile, this work does not require the knowledge of the number of pre- and post-break factors for detecting the instability.

Moreover, we show that a structural change is recognizable if the factor loading changes. We determine the break point by the dimensionality of the factor model. As a result, the total number of pre- and post-break factors is minimized when the break date is specified precisely. We also show that as long as the number of pre- and post-break factors have been determined, the location of the break date can be estimated by using sum-of-square residuals criterion.

As mentioned in Cheng et al. (2016)[13], the proposed estimator is developed by minimizing the penalized least square (PLS) criterion function. Then, we apply the Group-Least Absolute Shrinkage and Selection Operator (Group-LASSO) penalties in pre-break factor loadings and changes in factor loadings. The number of non-zero columns in the loading matrices is equal to the number of pre- and post-break factors. When a column of zero in loading matrices becomes non-zero after the break, a new factor appears. We assume that the number of factors is fixed as the sample size increases, and we also assume that the breaks in the loadings do not shrink with the sample size.

1.3 Main Contribution

The main contribution of this major paper is to present an econometric model, which is capable of detecting the type of instability, determining the number of pre- and post-break factors, and detecting the break date simultaneously. We consider the type of factor model instability: changes in the number of strong factors. Indeed, in an economic environment, the break date is usually unknown.

According to Zou (2006)[28], the LASSO is a regularization technique for simultaneous estimation and variable selection. Cheng et al. (2016)[13] extends the results in Zou (2006)[28] in the following way. First, they use a two-step procedure to determine LASSO penalty. Second, they construct the penalty terms for the unknown break date case. The method consists in taking the average of the penalties computed by each potential break date. Third, they develop a cross-validation procedure to fine-tune the LASSO penalties and propose the shrinkage estimation via LASSO. The PLS estimator is a shrinkage estimator because it sets small coefficient estimates equal to zero.

1.4 Organization of the Major Paper

The remainder of this major paper is organized as follows. Chapter 2 describes the statistical models and conditions for the instability. Chapter 3 presents the shrinkage estimation and introduces the model selection. This chapter also addresses the asymptotic theory for the estimator as well as the implementation of the parameter estimation and algorithm. The content presented in Chapter 4 is similar to Chapter 3 but the break date is unknown. Chapter 5 presents the numerical results from the Monte Carlo simulation and the real data set of the Great Recession in the U.S.. The results and interpretation on the finite-sample performance of Group-LASSO estimator and Bootstrap data report are included in this chapter. Finally, Chapter 6 concludes.

Chapter 2

Model Specifications

This chapter introduces the statistical models, notations, and instability with the structural break. It is subdivided into two sections. In Section 2.1, we introduce the statistical model with structural break date changes. In Section 2.2, we identify the existence of instability.

2.1 Statistical Model and Notations

In this section, we introduce the models with respect to the factors and loadings. Consider that we observe the panel data $\{X_{it} \in R : i = 1, \cdots, N, t = 1, \cdots, T_0, \cdots, T\}$. Let $X_t = (X_{1t}, \cdots, X_{Nt})' \in R^{N\times 1}$ be the observations at time t, with $t \in \{1, \cdots, T_0, \cdots, T\}$ where T_0 denotes the break point. Usually, T_0 is unknown. Before T_0, there are r_a unobserved pre-break factors. After T_0, there are no further breaks. We write the pre-break statistical model in matrix notation as

$$X_a = F_a \Lambda^{0\prime} + e_a, \tag{2.1}$$

where $\Lambda^0 \in R^{N \times r_a}$ is the matrix of factor loadings, $X_a = (X_1, \cdots, X_{T_0})' \in R^{T_0 \times N}$, $F_a = (F_1^0, \cdots, F_{T_0}^0)' \in R^{T_0 \times ra}$, and $e_a = (e_1, \cdots, e_{T_0})' \in R^{T_0 \times N}$. Both matrices F_a and Λ^0 are unknown. The post-break statistical model is given by

$$X_b = F_{b,1}(\Lambda^0 + \Gamma_1^0)' + F_{b,2}\Gamma_2^{0\prime} + e_b, \tag{2.2}$$

where $X_b = (X_{T_0+1}, \cdots, X_T)' \in R^{T_1 \times N}$ and $T_1 = T - T_0$, $F_{b,1} = (F_{T_0+1}^0, \cdots, F_T^0)' \in R^{T_1 \times r_a}$, $F_{b,2} = (F_{T_0+1}^*, \cdots, F_T^*)' \in R^{T_1 \times (r_b - r_a)}$, and $e_b = (e_{T_0+1}, \cdots, e_T)' \in R^{T_1 \times N}$. Here, the matrix $F_{b,1}$ spreads the factor in pre-break period to post-break period, and the matrix $F_{b,2}$ collects new factors in post-break period. The matrices Γ_1^0 and Γ_2^0 denote the change in loadings of F_t^0 and loading for new factors F_t^* respectively. Let $\Gamma^0 = (\Gamma_1^0, \Gamma_2^0)$. If the loading for factors doesn't change, then $\Gamma_1^0 = 0$. While if there are no new factors, then $\Gamma_2^0 = 0$. Also, we rewrite the model in (2.2) as

$$X_b = F_b \Psi^{0\prime} + e_b, \tag{2.3}$$

where $F_b = (F_{b,1}, F_{b,2}) \in R^{T_1 \times r_b}$ and $\Psi^0 = (\Lambda^0 + \Gamma_1^0, \Gamma_2^0) \in R^{N \times r_b}$. In (2.1) and (2.2), the product of factors and their loadings are identifiable. However, each term is not identifiable. Thus, we impose normalization restrictions for the factor model. We rewrite the statistical model as

$$\begin{aligned} X_a &= (F_a R_a)(R_a^{-1}\Lambda^{0\prime}) + e_a = F_a^R \Lambda^{R\prime} + e_a, \\ X_b &= (F_b R_b)(R_b^{-1}\Psi^{0\prime}) + e_b = F_b^R \Psi^{R\prime} + e_b, \end{aligned} \tag{2.4}$$

where $F_a^R = F_a R_a$ and $F_b^R = F_b R_b$. The R_a and R_b are transformation matrices.

2.2 Identification of Instability

In this section, we introduce the structural instability and identify it when the break date is known or unknown. We assume that the number of pre-break factors is smaller than the number of post-break factors: $r_a < r_b$. This is called the instability. Under this instability, the new factors appear in the model after the break point T_0. In the mean time, the old factors in the loadings may change. We consider two cases when the break date T_0 is known and unknown, and identify the instability for each case in Chapter 3 and 4.

Breitung and Eickmeier (2011)[8] explain that the subsample of pre- and post-break observations will have one or more additional factors if the break date is misspecified. Specifically, to estimate the break date, we can adjust the potential break date π to minimize the sum of the numbers of pre- and post-break factors.

Chapter 3

Estimation and Modeling in Known Break Date Case

In this chapter, we assume that the break date is known. As mentioned in Tibshirani (1996)[26], the ordinary least squares (OLS) estimators are obtained by minimizing the residual squared error. On the one hand, the OLS estimators often have large variance but low bias. The prediction accuracy can be improved by shrinking the coefficients to zero. On the other hand, with the large number of predictors, we determine a smaller subset which has significant effects. Therefore, we propose the shrinkage estimation via LASSO. It shrinks some coefficients and sets other elements to zero. Then, we determine the number of pre- and post-break factors by the shrinkage estimator. Next, we explain the detailed asymptotic properties of PLS estimator. Finally, we apply a two-step estimation method to improve the finite sample performance with the adjusted penalty weights.

In Section 3.1, we introduce the shrinkage estimator. In Section 3.2, we select the proper model. Section 3.3 describes the post model selection estimation by using the least square method. In Section 3.4, we study the asymptotic theory for the proposed shrinkage estimator and demonstrate related theorems and assumptions. Sections 3.5, 3.6, and 3.7 explain the techniques which are applied on weights and tuning parameters. In real-world applications, the break date is always unknown. We will extend the results to unknown break date case in the next chapter.

3.1 Shrinkage Estimator

The shrinkage estimators dominate the classical estimators in terms of mean squared error (MSE) for a host of statistical models. Ahmed (2014)[1] explains that the shrinkage estimation strategy can be used for both model selection and post estimation. In this section, we introduce the strategy to construct the shrinkage estimator. To give another reference about shrinkage and LASSO estimator, we also quote Nkurunziza et al. (2016)[18] and references there in.

Cheng et al. (2016)[13] rewrite the statistical model in (2.4) as augmented system, because the criterion function needs to be motivated in the shrinkage estimation. Suppose that r_a and r_b are unknown, we choose a upper bound k such that $r_a + r_b \leq k$.

We rewrite statistical model in (2.4) as augmented system

$$
\begin{aligned}
X_a &= \begin{bmatrix} F_a^R & F_{a,1}^{R\perp} & F_{a,2}^{R\perp} \end{bmatrix} \begin{bmatrix} \Lambda^{R'} \\ \mathbf{0}_{(rb-r_a)\times N} \\ \mathbf{0}_{(k-r_b)\times N} \end{bmatrix} + e_a = F_a^{R+}(\Lambda^{R+})' + e_a, \\
X_b &= \begin{bmatrix} F_{b,1}^R & F_{b,2}^R & F_b^{R\perp} \end{bmatrix} \begin{bmatrix} \Lambda^{R'} + \Gamma_1^{R'} \\ \Gamma_2^{R'} \\ \mathbf{0}_{(k-r_b)\times N} \end{bmatrix} + e_b = F_b^{R+}(\Lambda^{R+} + \Gamma^{R+})' + e_b.
\end{aligned} \tag{3.1}
$$

Here, $F_a^{R\perp} = (F_{a,1}^{R\perp}, F_{a,2}^{R\perp})$ and $F_b^R = (F_{b,1}^R, F_{b,2}^R)$. $F_{a,1}^{R\perp} \in R^{T_0\times(r_b-r_a)}$ and $F_{a,2}^{R\perp} \in R^{T_0\times(k-r_b)}$ are sub-matrices of $F_a^{R\perp} \in R^{T_0\times(k-r_a)}$. $F_a^{R\perp}$ denotes the orthogonal complement of $F_a^R \in R^{T_0\times r_a}$. Similarly, $F_b^{R\perp}$ denotes the orthogonal complement of $F_b^R \in R^{T_b\times r_b}$, $\Lambda^{R+} = \begin{bmatrix} \Lambda^R & \mathbf{0}_{N\times(r_b-r_a)} & \mathbf{0}_{N\times(k-r_b)} \end{bmatrix}$ and $\Gamma^{R+} = \begin{bmatrix} \Gamma_1^R & \Gamma_2^R & \mathbf{0}_{N\times(k-r_b)} \end{bmatrix}$. Λ^{R+} and $(\Lambda^{R+} + \Gamma^{R+})$ are the factor loadings in pre- and post-break respectively. F_a^{R+} and F_b^{R+} are augmented matrices. Recall that the number of non-zero columns in the loading matrices is equal to the number of pre- and post-break factors. Thus, r_a and r_b can be estimated. Moreover, the instability can be detected. Note that with the existence of the instability, $r_b > r_a$. We first assume that the break date is known and let $T_a = T_0$. In the following, we introduce the shrinkage estimator.

Yuan and Lin (2006)[27] explain that the shrinkage estimator can be obtained by minimizing the penalized least square (PLS) objective function with group-LASSO penalty, which is defined in terms of the ℓ-th column of the norm of Λ and Γ. A group-LASSO estimator either sets all elements in group equal to zero or estimates those elements as nonzero (Cheng et al. (2016))[13]. We use the group-LASSO for large-scale factor models because the irrelevant factors have zero factor loadings for all series. To estimate the upper bound k, we need to know each principle component estimator in subsample. In particular, assume $j \in \{a, b\}$, let $\widetilde{F}_j \in R^{T_j \times k}$ denote the orthonormalized eigenvectors of $(NT_j)^{-1} X_j X_j'$ with the first k largest eigenvalues. Let $\mathcal{I}_A$ denote the indicator function of the event A. In each subsample, we estimate an over-fitted model with k factors, then we have the unrestricted least square estimators of the factor loading $\widetilde{\Lambda}_{LS} = T_a^{-1} X_a' \widetilde{F}_a$, $\widetilde{\Psi}_{LS} = T_b^{-1} X_b' \widetilde{F}_b$ and $\widetilde{\Gamma}_{LS} = \widetilde{\Psi}_{LS} - \widetilde{\Lambda}_{LS}$. Now, we propose the shrinkage estimator of Λ^{R+} and Γ^{R+} by minimizing the penalized least square (PLS) objective function

$$(\widehat{\Lambda}, \widehat{\Gamma}) = \underset{\Lambda \in R^{N \times k}, \Gamma \in R^{N \times k}}{\operatorname{argmin}} [M(\Lambda, \Gamma) + P_1(\Lambda) + P_2(\Gamma)], \tag{3.2}$$

where

$$\begin{aligned} M(\Lambda, \Gamma) &= (NT)^{-1} \left[\left\| X_a - \widetilde{F}_a(\pi)\Lambda' \right\|^2 + \left\| X_b - \widetilde{F}_b(\Lambda + \Gamma)' \right\|^2 \right], \\ P_1(\Lambda) &= \alpha_{NT} \sum_{\ell=1}^{k} \omega_\ell^\lambda \|\Lambda_\ell\| \text{ and } P_2(\Gamma) = \beta_{NT} \sum_{\ell=1}^{k} \omega_\ell^\gamma \|\Gamma_\ell\|, \end{aligned} \tag{3.3}$$

where $\widetilde{F}_a$ and $\widetilde{F}_b$ are given terms, Λ_ℓ and Γ_ℓ are the ℓ-th column of Λ and Γ respectively. Either α_{NT} or β_{NT} are two coefficients of positive real number which depend on N and T. ω_ℓ^λ and ω_ℓ^γ are weights defined as

$$\begin{aligned} \omega_\ell^\lambda &= \left(N^{-1} \|\widetilde{\Lambda}_\ell\|^2 \mathcal{I}_{\{\widetilde{\Lambda}_\ell \neq 0_{N\times 1}\}} + N^{-1} \|\widetilde{\Lambda}_{\ell,LS}\|^2 \mathcal{I}_{\{\widetilde{\Lambda}_\ell = 0_{N\times 1}\}} \right)^{-2}, \\ \omega_\ell^\gamma &= \left(N^{-1} \|\widetilde{\Gamma}_\ell\|^2 \mathcal{I}_{\{\widetilde{\Gamma}_\ell \neq 0_{N\times 1}\}} + N^{-1} \|\widetilde{\Gamma}_{\ell,LS}\|^2 \mathcal{I}_{\{\widetilde{\Gamma}_\ell = 0_{N\times 1}\}} \right)^{-2}, \end{aligned} \tag{3.4}$$

where $\widetilde{\Lambda} \in R^{N\times K}$ and $\widetilde{\Gamma} \in R^{N\times K}$ are preliminary estimators of factor loading Λ^{+} and Γ^{+}. The weights are used to distinguish the zero and nonzero columns in the loading matrices Λ^{R+} and Γ^{R+}. (Cheng et al. (2016))[13]

3.2 Model Selection Estimator

In this section, we apply the shrinkage estimator defined in Section 3.1 to determine the number of pre- and post-break factors. When the break date T_0 is known, r_a and r_b are known as well. We assume that the inequality $r_b > r_a$ holds, this condition identifies the instability on statistical model. We also detect the existence of the instability by using the shrinkage estimator. Let the break indicator $\mathcal{B}_0 \in \{0,1\}$. If there is no structural break, then $\mathcal{B}_0$=0. With the existence of the instability, $\mathcal{B}_0$=1 and $r_a < r_b$. Here, the estimation of $\mathcal{B}$, r_a and r_b happens in the mean time (Cheng et al. 2016)[13]. Since $\Gamma^0 = (\Gamma_1^0, \Gamma_2^0) = 0$ if and only if $\Gamma^R = (\Gamma_1^R, \Gamma_2^R) = 0$, we rewrite post-break statistical models in (2.4) to determine $\mathcal{B}_0$ as

$$X_b = F_b^R \Psi^{R\prime} + e_b = F_{b,1}^R(\Lambda^R + \Gamma_1^R)' + F_{b,2}^R \Gamma_2^{R\prime} + e_b, \tag{3.5}$$

where $F_b^R = (F_{b,1}^R, F_{b,2}^R)$, $\Psi^R = (\Lambda^R + \Gamma_1^R, \Gamma_2^R)$ and Γ^R=(Γ_1^R,Γ_2^R).

We need to know the column norm of $\widehat{\Lambda}$ and $\widehat{\Gamma}$ to estimate $\mathcal{B}$, r_a and r_b. The estimator of $\mathcal{B}_0$ is defined as

$$\widehat{\mathcal{B}} = \mathcal{I}_{\{\|\widehat{\Gamma}\|>0\}}. \tag{3.6}$$

The estimators of r_a and r_b are given by

$$\begin{aligned} \widehat{r}_a &= \min \{j \geq 1 \text{ and } j : \|\widehat{\Lambda}_\ell\| = 0 \text{ for all } \ell > j\}, \\ \widehat{r}_b &= \max \left(\hat{r}_a,\ \min \{j : \|\widehat{\Gamma}_\ell\| = 0 \text{ for all } \ell > j\}\right). \end{aligned} \tag{3.7}$$

On the one hand, the method can be used to detect a structural break and determine the instability. On the other hand, to detect the structural break in the factor loadings, the method does not require the knowledge of number of pre- and post-break factors.

3.3 Post Model Selection Estimation

In this section, we demonstrate that the shrinkage estimator can provide a estimation of loading matrices Λ and Γ. However, the penalty terms do not estimate the non-zero coefficients. Thus, we propose to re-estimate the loading matrices by using least squares conditional on the estimators $\widehat{\mathcal{B}}$, $\widehat{r}_a$ and $\widehat{r}_b$. The estimator for the post model selection is named PMS estimator.

If $\widehat{\mathcal{B}} = 0$, which means that there is no structural break, we can re-estimate the factor model on the full sample. Specifically, let $\widetilde{F} \in R^{T \times k}$ denote the orthonormalized first k principle components constructed from the full sample. Let $\overline{\Lambda}$ denote the first $\widehat{r}_a$ columns of the full sample least square estimator $\widehat{\Lambda}_{LS}$, where $\widetilde{\Lambda}_{LS} = T^{-1}X'\widetilde{F}$. Therefore, we set $\overline{\Psi} = \overline{\Lambda}$, since the columns of $\widetilde{F}$ are constructed to be orthogonal, $\overline{\Lambda}$ is identical to the OLS estimator obtained by regressing X on the first $\widehat{r}_a{}^*$ columns of

$\widetilde{F}$ (Cheng et al. 2016)[13].

If $\widehat{\mathcal{B}} = 1$, then we need to re-estimate the subsample of factors and loadings. Let $\widetilde{F}_a$ and $\widetilde{F}_b$ denote the factor estimates for the subsample of factors and loadings respectively. In addition, let $\bar{\Lambda}$ denote the first $\widehat{r}_a$ columns of the least square estimator $\widetilde{\Lambda}_{LS} = T^{-1}X_a'\widetilde{F}_a$. Let $\bar{\Psi}$ denote the first $\widehat{r}_b$ columns of the least square estimator $\widetilde{\Psi}_{LS} = T^{-1}X_b'\widetilde{F}_b$. The PMS estimators are defined as

$$\widehat{\Lambda}_{PMS} = (\overline{\Lambda}, \mathbf{0}) \text{ and } \widehat{\Psi}_{PMS} = (\overline{\Psi}, \mathbf{0}), \tag{3.8}$$

where $\mathbf{0}$ is the zero matrix.

3.4 Asymptotic Properties

In this section, we present the asymptotic properties of PLS estimator, the instability and the break date. Bai and Ng (2002)[4] explains that penalty for overfitting must be a function of both the cross-sections (N) and the time dimensions (T) in order to estimate the number of factors. However, the function of N or T, such as AIC or BIC, do not work because both dimensions of the panel data are large. As discussed in Bai and Ng (2002)[4], this major paper assumes that both N and T converge to infinity under empirical application to maintain flexibility. We present some assumptions and theorems on the large sample properties of the preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ and the convergence rates of the sequences α_{NT} and β_{NT} below.

Recall that the notation of $X_n = O_p(a_n)$ means the set of values x_n/a_n is stochastically bounded. More specifically, for any $\epsilon > 0$, there exists a finite $M > 0$ and a finite $N > 0$, such that $\Pr(|X_n/a_n| > M) < \epsilon, \forall n > N$. The following Theorem 3.1 and 3.2 have restricted the preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ on stochastic order, which may change the data-dependent weights ω_ℓ^λ and ω_ℓ^γ in (3.4). We define C_{NT}=min($T^{1/2}$,$N^{1/2}$), where C_{NT} is the convergence rate of the unrestricted least square estimator.

Theorem 3.1. *As N, T→ ∞ with $\sqrt{T}/N \to 0$, the preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ satisfy*

(i) $\Pr(N^{-1}\|\widetilde{\Lambda}_\ell\|^2 \geq C) \to 1$ *for* $\ell = 1, \cdots, r_a$,

$N^{-1}\|\widetilde{\Lambda}_\ell\|^2 = O_p(C_{NT}^{-2})$*for* $\ell = r_a + 1, \cdots, k$;

(ii) If $\Gamma^0 \neq 0$, $\lim\limits_{N,T\to\infty} \Pr(N^{-1}\|\widetilde{\Gamma}_\ell\|^2 \geq C) = 1$ *for* $\ell = 1, \cdots, r_b$,

$N^{-1}\|\widetilde{\Gamma}_\ell\|^2 = O_p(C_{NT}^{-2})$*for* $\ell = r_b + 1, \cdots, k$;

(iii) If $\Gamma^0 = 0$, $N^{-1}\|\widetilde{\Gamma}_\ell\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = 1, \cdots, k$. □

The proof of this theorem follows from the main results in Cheng et al. (2016)[13]. In Theorem 3.1, we separate the column of the preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ into two parts. In the first part, $\lim\limits_{N,T\to\infty} \Pr(N^{-1}\|\widetilde{\Lambda}_\ell\|^2 \geq C) = 1$ and $\lim\limits_{N,T\to\infty} \Pr(N^{-1}\|\widetilde{\Gamma}_\ell\|^2 \geq C) = 1$ such that the data-dependent weights ω_ℓ^λ and ω_ℓ^γ are stochastically bounded. In the second part, $N^{-1}\|\widetilde{\Lambda}_\ell\|^2 = O_p(C_{NT}^{-2})$ and $N^{-1}\|\widetilde{\Gamma}_\ell\|^2 = O_p(C_{NT}^{-2})$ imply that ω_ℓ^λ and ω_ℓ^γ diverge in probability faster than C_{NT}^4.

Theorem 3.2. *As N, T→ ∞ with $\sqrt{T}/N \to 0$, the preliminary estimators $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$ satisfy* *(i)* $\Pr(N^{-1}\|\widetilde{\Lambda}_{LS,\ell}\|^2 \geq C) \to 1$ *for* $\ell = 1, \cdots, r_a$, $N^{-1}\|\widetilde{\Lambda}_{LS,\ell}\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = r_a + 1, \cdots, k$;

(ii) If $\Gamma^0 \neq 0$, $\Pr(N^{-1}\|\widetilde{\Gamma}_{LS,\ell}\|^2 \geq C) \to 1$ *for* $\ell = 1, \cdots, r_b$, $N^{-1}\|\widetilde{\Gamma}_{LS,\ell}\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = r_b + 1, \cdots, k$;

(iii) If $\Gamma^0 = 0$, $N^{-1}\|\widetilde{\Gamma}_{LS,\ell}\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = 1, \cdots, k$. □

The proof of this theorem is given in Cheng et al. (2016)[13]. In Theorem 3.2, if $\widetilde{\Lambda}$ or $\widetilde{\Gamma}$ has zero columns, the data-dependent weights ω_ℓ^λ and ω_ℓ^γ depend on $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Lambda}_{LS}$. Note that $\widetilde{\Lambda}_\ell = 0$ is a special case of $N^{-1}\|\widetilde{\Lambda}_\ell\|^2 = O_p(C_{NT}^{-2})$ in Theorem 3.1, so does $\widetilde{\Gamma}_\ell$. The data-dependent weights ω_ℓ^λ and ω_ℓ^γ determine the relative penalties of different columns in the factor loadings. The tuning parameters α_{NT} and β_{NT} determine the overall penalization. As the tuning parameters vanish asymptotically, we make Assumption 1 for the rates.

Assumption 1. *The tuning parameters* α_{NT} *and* βNT *satisfy*

(i) $\alpha_{NT} = O(N^{-1/2}C_{NT}^{-1})$ *and* $\beta_{NT} = O(N^{-1/2}C_{NT}^{-1})$

(ii) $N^{-1/2}C_{NT}^{-5} = o(\alpha_{NT})$ *and* $N^{-1/2}C_{NT}^{-5} = o(\beta_{NT})$. □

In Assumption 1, the boundaries on the tuning parameters α_{NT} and β_{NT} control the magnitudes of the overall penalization. In Assumption 1(i), we introduce the upper bound to ensure that the penalties on the nonzero columns are small when the weights ω_ℓ^λ and ω_ℓ^γ are stochastically bounded. Also, we propose to shrink the estimators of zero columns to zero. Assumption 1(ii) requires the tuning parameters α_{NT} and β_{NT} converge to zero slowly with the lower bound.

The following Assumptions 2 and 3, respectively, are analogous to Assumptions A and B in Bai and Ng (2002)[4]. For $t > T_0$, let $\bar{F}_t^0 = (F_t^{0\prime}, F_t^{*\prime})' \in R^{r_b}$ denote r_b factors in the post-break period and $C \in \mathbb{R}$ denotes a generic positive constant.

Assumption 2. *(i)*$\mathbb{E}[\|F_t^0\|^4] \leq C$, $\mathbb{E}[\|\overline{F}_t^0\|^4] \leq C$ *and there exists positive definite nonrandom matrices* Σ_F *and* $\Sigma_{\overline{F}}$ *such that* $T_0^{-1}\sum_{t=1}^{T_0} F_t^0 F_t^{0\prime} = \Sigma_F + O_p(T_0^{-1/2})$ *and* $T_1^{-1}\sum_{t=T_0+1}^{T} \overline{F}_t^0 \overline{F}_t^{0\prime} = \Sigma_{\overline{F}} + O_p(T_1^{-1/2})$.
(ii) The positive definite matrices Σ_F *and* $\Sigma_{\overline{F}}$ *are both not related to* N *and* T. □

Assumption 3. *(i)* $\|\lambda_i^0\| \leq C$, $\|\psi_i^0\| \leq C$ *and there exists nonrandom matrices* Σ_Λ, Σ_Ψ *and* $\Sigma_{\Lambda\Psi}$ *such that* $\|\Lambda^{0\prime}\Lambda^0/N - \Sigma_\Lambda\| \to 0$, $\|\Psi^{0\prime}\Psi^0/N - \Sigma_\Psi\| \to 0$ *and* $\|\Lambda^{0\prime}\Psi^0/N - \Sigma_{\Lambda\Psi}\| \to 0$ *as* $N \to \infty$, *where* Σ_Λ *and* Σ_Ψ *are positive definite matrices and are both not related to* N *and* T.
(ii) The matrices $\Sigma_\Lambda\Sigma_F$ *and* $\Sigma_\Psi\Sigma_{\overline{F}}$ *both have distinct eigenvalues.* □

Here, $\Lambda^0 = (\Lambda_1^0, \cdots, \Lambda_N^0)'$, where $\Lambda_i^0 \in R^{r_a \times 1}$ denotes the factor loading for series i before the structural break. Similarly, $\Psi^0 = (\psi_1^0, \cdots, \psi_N^0)'$, where $\psi_i^0 \in R^{r_b \times 1}$ denotes the factor loading for series i after the structural break. We state the following assumptions for additional factors and the changes of factor loadings at T_0.

Suppose $T_0/T \to \tau_0$, $\tau_0 \in (0,1)$ as $T \to \infty$. We extend Assumptions 2 and 3 to Assumptions 4 and 5. Let $e = [e_1, \cdots, e_T] \in R^{N \times T}$ denote the matrix of error terms and e_{it} denote the element of e with series i in period t.

Assumption 4. *(i)* $\mathbb{E}[e_i t] = 0$, $\mathbb{E}[|e_{it}|] \leq C$;
(ii) $\mathbb{E}[N^{-1}\sum_{i=1}^{N} e_{is}e_{it}] = \sigma_N(s,t)$, $|\sigma_N(s,s)| \leq C$ *for all* s,
$T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} |\sigma_N(s,t)| \leq C$;
(iii) $\mathbb{E}[e_{it}e_{jt}] = \tau_{ij,t}$ *with* $|\tau_{ij,t}| \leq |\tau_{ij}|$ *for some* τ_{ij} *and for all* t, *and*

$N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\sigma_N(s,t)| \leq C$;

(iv) $\mathbb{E}[e_{it}e_{js}] = \tau_{ij,ts}$ *and* $(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{N}\sum_{s=1}^{N}|\tau_{ij,ts}| \leq C$;

(v) For every (t,s), $\mathbb{E}[|N^{-1/2}\sum_{i=1}^{N}[e_{is}e_{it} - \mathbb{E}[e_{is}e_{it}]]|^4] \leq C$;

(vi) $\rho_1((NT)^{-1}e_a e_a') = O_p(max[N^{-1}, T^{-1}])$ *and*

$\rho_1((NT)^{-1}e_b e_b') = O_p(max[N^{-1}, T^{-1}])$. □

Assumption 5. $\mathbb{E}[N^{-1}\sum_{i=1}^{N}\|T^{-1/2}(\sum_{t=1}^{T_0}F_t^0 e_{it} + \sum_{t=T_0+1}^{T}\overline{F}_t^0 e_{it})\|^2] \leq C$. □

Assumption 4 models for time-series and cross-sectional weak dependence in the error terms. Assumption 5 models the weak dependence between the factors and error terms. Those assumptions are analogous to Assumptions C and D in Bai and Ng (2002)[4] respectively. With the knowledge of the above theorems and assumptions, we state the asymptotic limits of the PLS estimators $\widehat{\Lambda}$ and $\widehat{\Gamma}$ in the following theorem. Those PLS estimators converge to the coefficients in (2.4). Let the subscript ℓ denote the ℓ-th column of a matrix.

Theorem 3.3. *Suppose that Assumptions 1 to 5 hold. Then,*

(i) Pre-break loadings of relevant factors: $N^{-1}\|\widehat{\Lambda}_\ell - \widehat{\Lambda}_\ell^R\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = 1, \cdots, r_a$;

(ii) Pre-break loadings of irrelevant factors:

$\lim_{N,T\to\infty} \Pr(\|\widehat{\Lambda}_\ell\|^2 = 0 \text{ for } \ell = r_a + 1, \cdots, k) = 1$;

(iii) Post-break changes in loadings of relevant factors: If $\Gamma^0 \neq 0$,

$N^{-1}\|\widehat{\Gamma}_\ell - \Gamma_\ell^R\|^2 = O_p(C_{NT}^{-2})$ *for* $\ell = 1, \cdots, r_b$;

(iv) No-break: If $\Gamma^0 = 0$, $\lim_{N,T\to\infty} \Pr(\|\widehat{\Gamma}_\ell\|^2 = 0 \text{ for } 1, \cdots, r_b) = 1$;

(v) Post-break changes in loadings of irrelevant factors:

$\lim_{N,T\to\infty} \Pr(\|\widehat{\Gamma}_\ell\|^2 = 0 \text{ for } \ell = r_b + 1, \cdots, k) = 1$.

The proof of this theorem follows from the results in Cheng et al. (2016)[13]. In Theorem 3.3 part (i) and (ii), the factor loadings of the irrelevant factors are estimated to zero with probability approaching to one due to the penalization. For $\ell = 1, \cdots, r_a$, the PLS estimators $\widehat{\Lambda}_\ell$ and $\widehat{\Gamma}_\ell$ converge in probability to the factor loadings Λ_ℓ^R and Γ_ℓ^R of the transformed statistical models in (3.1) respectively. Parts (iii) and (v) detect the structural instability. Without the structural instability, as in part (iv), the PLS estimators $\widehat{\Gamma}_\ell$ of change in loadings equal to zero with probability approaching to one. Otherwise, part (v) only applies for $\ell = r_b + 1, \cdots, k$ and ensures the post-break number of factors.

Briefly, the factor loadings of the irrelevant factors are estimated with probability approaching to 1. In addition, without any instability, the changes in loadings of relevant factors are estimated with probability approaching to one as well.

As mentioned in Cheng et al. (2016)[13], to build the model selection for the PLS estimation, it is sufficient to show that the asymptotic limits of $N^{-1}\|\Lambda_\ell^R\|^2$ and $N^{-1}\|\Gamma_\ell^R\|^2$ in Theorem 3.3 part(i) and (iii) are bounded away from zero. We introduce Assumption 6, Lemma 3.1, and the following theorem provides the asymptotic result of $\widehat{\mathcal{B}}$, $\widehat{r}_a$ and $\widehat{r}_b$.

Assumption 6. *One of the following two conditions holds:*

(i) $\text{rank}(\Sigma_{\Lambda\Psi}^+) \geq r_a$;

(ii) $\rho_\ell(\Sigma_F \Sigma_\Lambda) \neq \rho_\ell(\Sigma_{\overline{F}} \Sigma_\Psi)$ *for some* $\ell \leq r_a$. □

Lemma 3.1. *Suppose Assumption 2-5 hold. Then,*

(i) Pre-break factors: $N^{-1}\|\Lambda_\ell^R\|^2 = \rho_\ell(\Sigma_\Lambda \Sigma_F) + o(1)$ *for* $\ell = 1, \cdots, r_a$*;*

(ii) New factors: If $r_b > r_a$, $N^{-1}\|\Gamma_\ell^R\|^2 = \rho_\ell(\Sigma_\Psi \Sigma_{\overline{F}}) + o(1)$ *for* $\ell = r_a + 1, \cdots, r_b$.

The proof of this lemma is given in Appendix A. Note that Lemma 3.1 provides the connection between the Assumption 6 and the statistical model determination.

Theorem 3.4. *Suppose that Assumptions 1-6 hold with the existence of the instability. Then,*

$$\lim_{N,T\to\infty} \Pr(\widehat{r}_a = r_a) = 1; \lim_{N,T\to\infty} \Pr(\widehat{r}_b = r_b) = 1; \lim_{N,T\to\infty} \Pr(\widehat{\mathcal{B}} = \mathcal{B}) = 1. \tag{3.9}$$

The proof of this theorem is given in Appendix A. The model selection procedure holds for any set of preliminary estimators that satisfy Theorem 3.1 and 3.2. Due to Step 1.5 in Algorithm 1 making a transformation of the estimators, we define

$$\mathcal{Z} = \{\ell : N^{-1}\|\Gamma_\ell^R\|^2 = N^{-1}\|\Psi_\ell^R - \Lambda_\ell^R\|^2 \geq C\}. \tag{3.10}$$

We make the following additional assumption.

Assumption 7. *If* $r_a = r_b$*, then* $\inf_{\|W\|=1} N^{-1}\|\Psi^R W - \Lambda_\ell^R\|^2 \geq C$ *for* $\ell \in \mathcal{Z}$. □

Assumption 7 holds as long as Λ_ℓ^R is not in the column space of Ψ^R. Under this assumption, some of the structural factor loadings in unnormalized statistical model (2.2) and (2.3) remain constant, while others change. Moreover, without structural instability, $\mathcal{Z}$ is empty and Assumption 7 is not necessary. The result in Theorem 3.4 can be generalized to the two-step estimation algorithm in later section.

3.5 On Estimation of the Penalty Weights

In this section, we present a practical procedure to choose the tuning parameters α_{NT} and β_{NT}. The penalty functions $P_1(\Lambda)$ and $P_2(\Gamma)$ depend on the weights ω_ℓ^λ and ω_ℓ^γ, they are determined by the tuning parameters α_{NT} and β_{NT}. α_{NT} and β_{NT} which are the penalty weights on the coefficients with respect to X_a and X_b respectively. The tuning parameters are important as they are applied in the two-step shrinkage estimation procedure. They are defined as

$$\alpha_{NT} = \kappa_1 N^{-1/2} C_{NT_a}^{-3} \text{ and } \beta_{NT} = \kappa_2 N^{-1/2} C_{NT_b}^{-3}, \tag{3.11}$$

where $C_{NT_a} = \min\ (N^{1/2}, T_a^{1/2})$, and $C_{NT_b} = \min\ (N^{1/2}, T_b^{1/2})$. Particularly, as mentioned in Cheng et al. (2016)[13], we choose α_{NT} and β_{NT} to fine-tune these two rates and replace the sample size T by the subsample sizes T_a and T_b. κ_1 and κ_2 are based on the PLS estimators with zero solution for some columns in Λ and Γ. Cheng et al. (2016)[13] explains that the criterion function in (3.2) is minimized at 0 if the marginal penalty for deviating from 0 is larger than the marginal gain on the least square criterion function. As mentioned in Bühlmann and van de Geer (2011)[8], $\|\widehat{\Lambda}_\ell\| = 0$ for $\ell > r_a$ if

$$\|e_a'(\widehat{\Lambda})\widetilde{F}_{a,\ell} + e_b'(\widehat{\Lambda} + \widehat{\Gamma})\widetilde{F}_{b,\ell}\| < NT\alpha_{NT}\omega_\ell^\lambda/2, \tag{3.12}$$

where

$$e_a(\Lambda) = X_a - \widetilde{F}_a\Lambda' \text{ and } e_b(\Lambda + \Gamma) = X_b - \widetilde{F}_b(\Lambda + \Gamma)'. \tag{3.13}$$

The reasonable choice of κ_1 is

$$\kappa_1 = \Big\{(NT_a)^{-1/2}\Big\|e_a(\widetilde{\Lambda})\Big\| + (NT_b)^{-1/2}\Big\|e_b(\widetilde{\Lambda}+\widetilde{\Gamma})\Big\|\Big\}. \tag{3.14}$$

To choose κ_2, we have

$$\kappa_2 = (NT_b)^{-1/2}\Big\|e_b(\widetilde{\Lambda}+\widetilde{\Gamma})\Big\| \tag{3.15}$$

where $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ are preliminary estimators and the residual matrices $e_a(\Lambda)$ and $e_b(\Lambda+\Gamma)$ are defined as

$$e_a(\Lambda) = X_a - \widetilde{F}_a\Lambda' \text{ and } e_b(\Lambda+\Gamma) = X_b - \widetilde{F}_b(\Lambda+\Gamma)'. \tag{3.16}$$

We set the constants c_1 and c_2 both equal to 1 as a default. However, we develop a cross-validation procedure to fine-tune these constants over a fixed interval in finite samples.

3.6 Two-Step Estimation Method

In this section, we introduce the two-step estimation procedure, which is designed by Cheng et al. (2016)[13]. Overall, this procedure improves the finite sample performance in two perspective. The tuning parameters α_{NT} and β_{NT} are more precise in the second step. The reason for this is that we obtain $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ in the first-step model selection; thus, the residual matrices $e_a(\Lambda)$ and $e_b(\Lambda+\Gamma)$ are more accurate. The preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ we obtained come from the the rotation of loading matrices Λ^R and Γ^R respectively. Let i=1 and 2 denote the first-step and second-step estimation method respectively. Let $\widetilde{\Lambda}^{(i)}$, $\widetilde{\Gamma}^{(i)}$ and $\widetilde{\Psi}^{(i)}$ denote the preliminary estimators in step i. Let $\widehat{\Lambda}^{(i)}$, $\widehat{\Gamma}^{(i)}$ and $\widehat{\Psi}^{(i)}$ denote the penalty least square (PLS) estimators in step i. Let $\widehat{\Lambda}^{(i)}_{PMS}$, $\widehat{\Gamma}^{(i)}_{PMS}$ and $\widehat{\Psi}^{(i)}_{PMS}$ denote the post model selection (PMS) estimators in step i. The two-step estimation procedures are performed in the following algorithm.

Algorithm 1 (Two-Step Estimation Method)

1. First-Stage Shrinkage Estimation:

   1.1. Compute the unrestricted least square estimators $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$.

   1.2. Set $\widetilde{\Lambda}^{(1)}=\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}^{(1)}=\widetilde{\Gamma}_{LS}$. Calculate ω_ℓ^λ, ω_ℓ^γ, α_{NT} and β_{NT} from (3.4) and (4.8) with $\widetilde{\Lambda}=\widetilde{\Lambda}^{(1)}$ and $\widetilde{\Gamma}=\widetilde{\Gamma}^{(1)}$.

   1.3. Compute the shrinkage estimator $\widetilde{\Lambda}^{(1)}$ and $\widetilde{\Gamma}^{(1)}$ by minimizing the criterion function in (3.2).

   1.4. Estimate r_a and r_b from (3.7) with $\widehat{\Lambda}=\widehat{\Lambda}^{(1)}$ and $\widehat{\Gamma}=\widehat{\Gamma}^{(1)}$. Name the estimator as $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$.

   1.5. Construct $\widehat{\Lambda}^{(1)}_{PMS}$ and $\widehat{\Psi}^{(1)}_{PMS}$ in (3.10). If $\widehat{r}_a^{(1)} = \widehat{r}_b^{(1)}$, then the transformation of the columns of $\overline{\Psi}^{(1)}$ is defined as follow. Let $\overline{\Lambda}^{(1)'}\overline{\Psi}^{(1)}=UDV'$ denote the singular value decomposition of $\overline{\Lambda}^{(1)'}\overline{\Psi}^{(1)}$. The transformed factor loading is defined as

$$\overline{\Psi}^{(1)}_R = \overline{\Psi}^{(1)}Q, \tag{3.17}$$

   where $Q=VU'$. The modified PMS estimator of Ψ is defined as

$$\widehat{\Psi}^{(1)}_{PMS-R} = \left(\overline{\Psi}^{(1)}_R, \mathbf{0}\right) \in R^{N\times K}. \tag{3.18}$$

2. Second-Stage Shrinkage Estimation

   2.1. Let
   $\widetilde{\Lambda}^{(2)} = \widehat{\Lambda}^{(1)}_{PMS}, \widetilde{\Gamma}^{(2)} = \widetilde{\Psi}^{(2)} - \widetilde{\Lambda}^{(2)}, \widetilde{\Psi}^{(2)} = \widehat{\Psi}^{(1)}_{PMS-R}\, \mathcal{I}_{\{\widehat{r}_b^{(1)} = \widehat{r}_a^{(1)}\}} + \widehat{\Psi}^{(1)}_{PMS}\, \mathcal{I}_{\{\widehat{r}_b^{(1)} > \widehat{r}_a^{(1)}\}}$, also calculate ω_ℓ^λ,ω_ℓ^γ, α_{NT}, and β_{NT} from (3.4) and (3.11) with $\widetilde{\Lambda}$=$\widetilde{\Lambda}^{(2)}$ and $\widetilde{\Gamma}$=$\widetilde{\Gamma}^{(2)}$.

   2.2. Compute the shrinkage estimators $\widehat{\Lambda}^{(2)}$ and $\widehat{\Gamma}^{(2)}$ by (3.2).

   2.3. Compute $\mathcal{B}_0^{(2)}$, $\widehat{r}_a^{(2)}$, and $\widehat{r}_b^{(2)}$ from (3.7) with $\widetilde{\Lambda}$=$\widetilde{\Lambda}^{(2)}$ and $\widetilde{\Gamma}$=$\widetilde{\Gamma}^{(2)}$.

   2.4. Construct the PMS estimator $\widehat{\Lambda}^{(2)}_{PMS}$ and $\widehat{\Psi}^{(2)}_{PMS}$ by definition in (3.10) conditional on $\mathcal{B}_0^{(2)}$, $\widehat{r}_a^{(2)}$, and $\widehat{r}_b^{(2)}$.

In this procedure, the preliminary estimator in step one is used to fine-tune the penalty terms in shrinkage estimator of step two. The preliminary estimator in step two is based on the PMS estimator in step one. In Step 1.5, the transformation increases the precision of locating the structural break when $\mathcal{B}_0$=0. The transformation does not have an effect on the asymptotic approach. Specifically, we need to find the orthogonal matrix Q in Step 1.5 such that $\|\overline{\Lambda}^{(1)} - \overline{\Psi}^{(1)}Q\|$ is minimized. The work of Schönemann (1966)[23] proposes similar work and obtained the solution to the orthogonal matrix. We apply this method to Step 1.5 and we find the orthogonal matrix Q=UV'. This leads to the sums of squares of the residual matrix $(\overline{\Lambda}^{(1)} - \overline{\Psi}^{(1)}Q)$ being minimized. It also minimizes the risk of locating an incorrect structural break date.

3.7 Cross Validation

Cross validation is proposed to adjust the constants $c = (c_1, c_2) \in \mathcal{C}$ in the penalty weights in (3.14 and 3.15). Applying the cross validation procedure, we obtain accurate tuning parameters α_{NT} and β_{NT}. In Section 3.5, we mentioned that we set the constants c_1 and c_2 equal to 1. Besides the time series dimension, we consider the sample in the cross-sectional dimension. The procedure is stated as follows: first, we consider the data in the cross-sectional dimension. We create the disjoint subsamples $X_{(-jN)}$ (N-regression) and X_{jN} (N-prediction). Second, we apply the model selection procedure (Section 3.2) to this subsample $X_{(-jN)}$ with a given value of c. We obtain the estimation of the unobserved factors and the model selection estimators. Third, we partition the subsample X_{jN} along the T dimension into regression and prediction samples. If the structural break took place in the model, we need to construct the regression and prediction samples separately for the pre- and post-break periods. We consider the factor estimates from the $X_{(-jN)}$ sample as the observed regressors. Moreover, we estimate the factor loadings based on the regression sample using ordinary least square (OLS).

The cross-validation criterion in this major paper is built on the mean-squared forecast errors (MSFE). The tuning constants κ_1 and κ_2 are chosen to minimize the MSFE for given c. The minimization is performed over a bounded set $\mathcal{C}$.

As mentioned in Cheng et al. (2016)[13], given the estimates of the number of pre- and post-break factors and the loadings, we can generate pseudo-out-of-sample forecasts for the prediction sample. We apply separate rolling pseudo-out-of-sample forecasting schemes for the pre- and post-break samples on the model selection estimators.

Chapter 4

Estimation and Modeling in Unknown Break Date Case

In the previous chapter, we introduced estimation and modeling in the known break case. In this chapter, we generalize the results to account for the unknown break date case. For the unknown break date case, we simply adopt *-superscripts and (π)-arguments to distinguish from the known break date case.

In Section 4.1, we introduce the shrinkage estimator. The model selection is described in Section 4.2. In Section 4.3, we describe the post model selection estimation and estimate the break date by using the least square method. In Section 4.4, we study the asymptotic theory for the proposed shrinkage estimator and present related theorems and assumptions. Section 4.5 presents the method of choosing the tuning parameters and performing the two-step estimation in the unknown break date case.

4.1 Shrinkage Estimator

In this section, we extend the procedure in Section 3.1 to the unknown break date case. When the break date T_0 is unknown, let T be the number of periods in the sample. We introduce a new parameter $\pi_0 = T_0/T$, which denotes the *true* break date. We assume that $\pi_0 \in \Pi$, where Π is some closed subset [0,1]. For any $\pi \in \Pi$, we split the full sample into pre- and post-break subsets $X_a(\pi) = (X_1, \cdots, X_{T_a})' \in R^{T_a \times N}$ and $X_b(\pi) = (X_{T_a+1}, \cdots, X_T)' \in R^{T_b \times N}$, where $T_a = \lfloor T \cdot \pi \rfloor$ denotes the integer part of $T \cdot \pi$ and $T_b = T - T_a$. To obtain the unknown break date π_0, we need to study the number of factors in $X_a(\pi)$ and $X_b(\pi)$. Here we denote $r_a(\pi)$ and $r_b(\pi)$ as number of factors in $X_a(\pi)$ and $X_b(\pi)$ respectively. $r_a(\pi)$ and $r_b(\pi)$ are defined as the number of non-vanishing eigenvalues of $(NT)^{-1}X_a(\pi)'X_a(\pi)$ and $(NT)^{-1}X_b(\pi)'X_b(\pi)$ as $N, T \to \infty$. We propose a range for the break dates such that $\pi \in \Pi = [\underline{\pi}, \overline{\pi}]$, where $\underline{\pi} > 0$ and $\overline{\pi} < 1$.

In practice, the break dates are not supposed to be close to zero or one, because it is not convenient to analyses the factor model in a small time dimension. Cheng et al. (2016)[13] suggest to set $\underline{\pi} \geq 0.15$ and $\overline{\pi} \leq 0.85$ for better model estimation in unknown break date case. Let $\widetilde{F}_a(\pi) \in R^{T_a \times k}$ denote the orthonormalized eigenvectors of $(NT_a)^{-1}X_a(\pi)X_a(\pi)'$ with first k largest eigenvalues. Similarly, let $\widetilde{F}_b(\pi) \in R^{T_b \times k}$ denote the orthonormalized left eigenvectors of $(NT_b)^{-1}X_b(\pi)X_b(\pi)'$ with first k largest eigenvalues. The unrestricted estimators of the factor loadings are $\widetilde{\Lambda}_{LS}(\pi) = T_a^{-1}X_a(\pi)'\widetilde{F}_a(\pi)$ and $\widetilde{\Psi}_{LS}(\pi) = T_b^{-1}X_b(\pi)'\widetilde{F}_b(\pi)$. In addition, $\widetilde{\Gamma}_{LS}(\pi) = \widetilde{\Psi}_{LS}(\pi) - \widetilde{\Lambda}_{LS}(\pi)$.

We replace π_0 by π in Section 3.1. Then, we get a shrinkage estimator with adjusted break date $\pi \in \Pi$ and consistent estimator of $r_a(\pi)$ and $r_b(\pi)$. Since the estimators of $r_a(\pi)$ and $r_b(\pi)$ are sensitive to π, we construct the shrinkage estimator with averaging penalty to maintain a low sensitivity of π in the finite sample. The shrinkage estimator is defined as

$$(\hat{\Lambda}(\pi), \hat{\Gamma}(\pi)) = \underset{\Lambda \in R^{N\times k}, \Gamma \in R^{N\times k}}{\operatorname{argmin}} [M(\Lambda, \Gamma; \pi) + P_1^*(\Lambda) + P_2^*(\Gamma)], \tag{4.1}$$

where

$$M(\Lambda, \Gamma; \pi) = (NT)^{-1} \left[\left\| X_a(\pi) - \widetilde{F}_a(\pi)\Lambda' \right\|^2 + \left\| X_b(\pi) - \widetilde{F}_b(\pi)(\Lambda + \Gamma)' \right\| \right]. \tag{4.2}$$

The averaging penalty functions $P_1^*(\Lambda)$ and$P_2^*(\Lambda)$ are defined as

$$P_1^*(\Lambda) = \sum_{\ell=1}^{k} \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda *}(\xi)]\|\Lambda_\ell\| \text{ and } P_2^*(\Gamma) = \sum_{\ell=1}^{k} \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma *}(\xi)]\|\Gamma_\ell\|, \tag{4.3}$$

where $\mathbb{E}_\xi[.]$ denotes the expectation with respect to ξ. By definition,

$$\begin{aligned} \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda}(\xi)] &= \int_{\underline{\pi}}^{\overline{\pi}} \alpha_{NT}(\xi)\omega_\ell^{\lambda}(\xi)\frac{1}{\overline{\pi}-\underline{\pi}}d\xi \ , \\ \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma}(\xi)] &= \int_{\underline{\pi}}^{\overline{\pi}} \beta_{NT}(\xi)\omega_\ell^{\gamma}(\xi)\frac{1}{\overline{\pi}-\underline{\pi}}d\xi, \end{aligned} \tag{4.4}$$

where $\underline{\pi}$ and $\overline{\pi}$ are lower and upper bounds on Π respectively. They all depend on N and T. $\alpha_{NT}(\pi)$ and $\beta_{NT}(\pi)$ are named tuning parameters and denote the coefficients of constants which depend on N and T for every π. The tuning parameters are not unique since π varies. For $\pi \in \Pi$, let $\widetilde{\Lambda}(\pi)$, $\widetilde{\Psi}(\pi)$ and $\widetilde{\Gamma}(\pi)$ denote some preliminary estimators, then the adaptive weights $\omega_\ell^{\lambda*}(\pi)$ and $\omega_\ell^{\gamma*}(\pi)$ in terms of the above preliminary estimators are defined as:

$$\omega_\ell^{\lambda*}(\pi) = \left(N^{-1}\|\widetilde{\Lambda}_\ell(\pi)\|^2 \mathcal{I}_{\{\widetilde{\Lambda}_\ell(\pi)\neq 0_{N\times 1}\}} + N^{-1}\|\widetilde{\Lambda}_{\ell,LS}(\pi)\|^2 \mathcal{I}_{\{\widetilde{\Lambda}_\ell(\pi)= 0_{N\times 1}\}}\right)^{-2},$$
$$\omega_\ell^{\gamma*}(\pi) = \left(N^{-1}\min\ \left\{\|\widetilde{\Gamma}_\ell(\pi)\|^2, \|\widetilde{\Psi}_\ell(\pi)\|^2\right\}\mathcal{I}_{\{\widetilde{\Gamma}_\ell(\pi)\neq 0_{N\times 1}\}}\right)^{-2} \quad (4.5)$$
$$+ \left(N^{-1}\min\ \left\{\|\widetilde{\Gamma}_{\ell,LS}(\pi)\|^2, \|\widetilde{\Psi}_{\ell,LS}(\pi)\|^2\right\}\mathcal{I}_{\{\widetilde{\Gamma}_\ell(\pi)= 0_{N\times 1}\}}\right)^{-2}.$$

As mentioned in Section 3.1, note that $\omega_\ell^{\lambda*}(\pi_0) = \omega_\ell^{\lambda}$ but $\omega_\ell^{\gamma*}(\pi_0) \neq \omega_\ell^{\gamma}$. When the break date is unknown, it is crucial to use $\omega_\ell^{\gamma*}(\pi)$ for estimation of γ_b. Cheng et al. (2016)[13] explain that for $\pi > \pi_0$ and $\ell > r_b$, we have $N^{-1}\|\widetilde{\Psi}_{\ell,LS}(\pi)\|^2$ converges in probability to zero when $n \to \infty$, but $N^{-1}\|\widetilde{\Gamma}_{\ell,LS}(\pi)\|^2$ may not converge in probability to 0. Thus, the modified adaptive weights can deliver larger penalties, when needed.

4.2 Model Selection Estimator

In this section, the model specification estimators $\widehat{B}^*$, $\widehat{r}_a^*$ and $\widehat{r}_b^*$ are similar to Section 3.2 with *-superscripts and (π)-arguments are adopted. Those estimators can be obtained as follows. First, we let

$$\widehat{\mathcal{B}}^* = \mathcal{I}_{\{sup_{\pi\in\Pi}\|\widehat{\Gamma}(\pi)\|>0\}}. \quad (4.6)$$

Second, the estimators of number of pre- and post-break factors r_a and r_b are defined as

$$\widehat{r}_a^* = \min_{\pi\in\Pi} \widehat{r}_a(\pi) \text{ and } \widehat{r}_b^* = \max_{\pi\in\Pi} \widehat{r}_b(\pi), \tag{4.7}$$

where $\widehat{r}_a(\pi)$ and $\widehat{r}_b(\pi)$ are defined as in (3.7). Here, we replace $\widehat{\Lambda}$ and $\widehat{\Gamma}$ by $\widehat{\Lambda}(\pi)$ and $\widehat{\Gamma}(\pi)$ respectively. The model specification estimators $\widehat{B}^*$, $\widehat{r}_a^*$ and $\widehat{r}_b^*$ can detect instability effectively in a large number of time series for the unknown break date case.

4.3 Post Model Selection Estimation

In Section 3.3, we presented the PMS estimators in the known break date case. For the unknown break date case, the PMS estimators are similar. As mentioned in the beginning of this chapter, we simply adopt *-superscripts and (π)-arguments for the unknown break date case. The PMS estimators are defined as

$$\widehat{\Lambda}_{PMS}(\pi) = (\overline{\Lambda}(\pi), \mathbf{0}) \text{ and } \widehat{\Psi}_{PMS}(\pi) = (\overline{\Psi}(\pi), \mathbf{0}), \tag{4.8}$$

where $\mathbf{0}$ is zero matrix. (Cheng et al. (2016))[13]
Bai (1997)[5] explains that when $\widehat{\mathcal{B}}^* = 1$, one can use the least square objective function to estimate the break date π_0. Let

$$\widehat{\pi} = \operatorname*{argmin}_{\pi\in\Pi} Q_{NT}(\pi; \widehat{r}_a^*, \widehat{r}_b^*), \tag{4.9}$$

where

$$
\begin{aligned}
&Q_{NT}(\pi; \widehat{r}_a^*, \widehat{r}_b^*) \\
&\quad = (NT)^{-1}\left[\left\|X_a(\pi) - \widetilde{F}_a(\pi)\widehat{\Lambda}'_{PMS}(\pi)\right\|^2 + \left\|X_b(\pi) - \widetilde{F}_b(\pi)\widehat{\Psi}'_{PMS}(\pi)\right\|\right].
\end{aligned}
\tag{4.10}
$$

4.4 Asymptotic Properties

In this section, we show that with the support of the averaging penalty in (4.3), the proposed shrinkage estimator with the averaging penalty can be extended to satisfy the unknown-break-date model. The tuning parameters and the two-step estimation method remain the same as in Sections 3.5 and 3.6. We propose the model specification estimators $\widehat{B}^*$, $\widehat{r}_a^*$ and $\widehat{r}_b^*$ directly without establishing the asymptotic behavior of the Group-LASSO estimator $\widehat{\Lambda}(\pi)$ and $\widehat{\Gamma}(\pi)$. The reason is that the shrinkage estimator with averaging penalty does not carry out the estimation of $r_a(\pi)$ and $r_b(\pi)$ for all π. Although the averaging penalty leads to over-penalizing when $\pi \neq \pi_0$, the consistent estimation of r_a and r_b can be obtained eventually since $r_a \leq r_a(\pi)$ and $r_b \leq r_b(\pi)$.

We reinforce Assumption 7 with the averaging penalty in the unknown break date case. For any $\pi \in \Pi$, we rewrite the normalized statistical model as

$$
\begin{aligned}
X_a(\pi) &= F_a^R(\pi)\Lambda^R(\pi)' + e_a(\pi), \\
X_b(\pi) &= F_b^R(\pi)\Psi^R(\pi)' + e_b(\pi),
\end{aligned}
\tag{4.11}
$$

where $F_a^R(\pi) \in R^{T_a \times (r_a + r_b)}$ and $\Lambda^R(\pi) \in R^{N \times (r_a + r_b)}$, and $F_a^R(\pi) \in R^{T_a \times (r_a + r_b)}$ and $\Psi^R(\pi)^{N \times (r_a + r_b)}$.

Assumption 8. *(i) If $r_a = r_b$, then* $\inf_{\pi\in\Pi, \|W\|=1} N^{-1}\|\Psi^R(\pi)W - \Lambda_\ell^R(\pi)\|^2 \geq C$ *for* $\ell \in \mathcal{Z}$;

(ii) If $r_b > r_a$, then $\inf_{\pi>\pi_0} N^{-1}\|\Psi^R(\pi)W - \Lambda_\ell^R(\pi)\|^2 \geq C$ *for* $\ell = r_b$. □

Assumption 8 is the reinforced version of Assumption 7. Part (i) is generalized from Assumption 7 by replacing the break date π to π_0 with any $\pi \in \Pi$. Part (ii) is designed for the unknown break date case because $\Lambda_\ell^R(\pi_0) = 0$ for $\ell = r_b > r_a$. The following theorem indicates that in the unknown break date case, we can obtain the asymptotic result of the estimator of r_a, r_b, and $\mathcal{B}$.

Assumption 9. $\mathbb{E}[\|F_t^0\|^4] \leq C$, $\mathbb{E}[\|\overline{F}_t^0\|^4] \leq C$ *and there exist random positive definite nonrandom matrices* Σ_F *and* $\Sigma_{\overline{F}}$ *such that* $T^{-1}\sum_{t=1}^{\lfloor T\pi \rfloor} F_t^0 F_t^{0\prime} = \pi\Sigma_F + O_p(T^{-1/2})$ *for* $\pi \leq \pi_0$ *and* $T^{-1}\sum_{t=\lfloor T\pi\rfloor+1}^{T} \overline{F}_t^0\overline{F}_t^{0\prime} = (1-\pi)\Sigma_{\overline{F}} + O_p(T^{-1/2})$ *for* $\pi \geq \pi_0$, *where both* $O_p(T^{-1/2})$ *terms are uniform over* $\pi \in \Pi$. □

Assumption 10. *Assumption 4 holds with e_a and e_b replaced by $e_a(\pi)$ and $e_b(\pi)$ and Assumption 4(vi) holds uniformly over $\pi \in \Pi$.* □

Theorem 4.1. *Suppose that Assumptions 3, 5, 6, 8-10 hold with the existence of the instability. Then,*

$$\lim_{N,T\to\infty} \Pr(\widehat{r}_a^* = r_a) = 1; \lim_{N,T\to\infty} \Pr(\widehat{r}_b^* = r_b) = 1; \lim_{N,T\to\infty} \Pr(\widehat{\mathcal{B}}^* = \mathcal{B}) = 1. \tag{4.12}$$

The proof of this theorem is similar to that of Theorem 3.4. When we take into account for the difference between π and π_0, the averaging penalty terms not only tend to over-penalize the loadings, but also set the loadings to zero for $\pi = \pi_0$. This brings out a tendency of underestimating either $r_a(\pi)$ or $r_b(\pi)$ if the conjectured break point is not specified correctly. An estimation of the break date can be identified when we apply the estimates $\widehat{r_a}^*$ and $\widehat{r_b}^*$ to the least squares objective function in (4.9).

In Chapter 6, we conduct Monte Carlo simulation to evaluate the performance of the shrinkage estimators.

4.5 On Estimation of Parameters and Algorithm

In this section, we introduce the tuning parameters, extend the estimation algorithm to the unknown break date case, and adjust the constants in the penalty weights by cross validation procedure for unknown break date case. We choose to use the following tuning parameters

$$\alpha_{NT}(\pi) = \kappa_1(\pi) N^{-1/2} C_{NT_a}^{-3} \text{ and } \beta_{NT}(\pi) = \kappa_2(\pi) N^{-1/2} C_{NT_b}^{-3} \tag{4.13}$$

where $\kappa_1(\pi) \in [\underline{\kappa}_1, \overline{\kappa}_1]$ and $\kappa_2(\pi) \in [\underline{\kappa}_2, \overline{\kappa}_2]$ for some $\underline{\kappa}_1, \overline{\kappa}_2 < \infty$. Note that the parameters in (4.13) are similar to (3.11). Practically, we use the value of $\kappa_1(\pi)$ and $\kappa_2(\pi)$ as defined in (3.14 and 3.15) with $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ replaced by $\widetilde{\Lambda}(\pi)$ and $\widetilde{\Gamma}(\pi)$.

We perform the two-step estimation method as in Section 3.6 to the unknown break date case by plugging the notation (π)-argument and *-subscript into the parameters. First, we set $\widetilde{\Lambda}^{(1)}(\pi) = \widetilde{\Lambda}_{LS}(\pi)$, $\widetilde{\Psi}^{(1)}(\pi) = \widetilde{\Psi}_{LS}(\pi)$ and $\widetilde{\Gamma}^{(1)}(\pi) = \widetilde{\Gamma}_{LS}(\pi)$. Second, we replace ω_ℓ^λ, ω_ℓ^γ, α_{NT} and β_{NT} by $\omega_\ell^{\lambda*}(\pi)$, $\omega_\ell^{\gamma*}(\pi)$, $\alpha_{NT}(\pi)$ and $\beta_{NT}(\pi)$. Third, we replace the PLS criterion function in (3.2) by (4.1). Fourth, we use $\widehat{r}_a^*$ and $\widehat{r}_b^*$ defined in (4.7) to replace those in (3.7). According to the definition in (4.7), the first step number of factors $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$ remains the same no matter how the value of π changes. Thus, we obtain the first-step shrinkage estimators $\widehat{\Lambda}^{(1)}(\pi)$ and $\widehat{\Gamma}^{(1)}(\pi)$ for every $\pi \in \Pi$, and we obtain $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$ in the first-step. Moreover, we obtain the estimators $\widehat{\Lambda}^{(2)}(\pi)$ and $\widehat{\Gamma}^{(2)}(\pi)$ for every $\pi \in \Pi$ in the second step. Finally, we obtain $\widehat{r}_a^*$, $\widehat{r}_b^*$, and $\widehat{\mathcal{B}}^*$ by the two step PLS estimators $\widehat{\Lambda}^{(2)}(\pi)$ and $\widehat{\Gamma}^{(2)}(\pi)$ following the results in Section 3.4.

According to Cheng et al. (2016)[13], the cross validation procedure introduced in Section 3.7 can be applied to the case of unknown break date. We take a common value of c for all possible break dates. For every π, the subsamples $X_{(-jN)}$ are constructed similarly as in Section 4.3; we replace π_0 by π. With the corresponding value of c, we obtain a selected model. Note that by definition, the selected model does not depend on π. For the cross validation subsample $X_{(jN)}$, we avoid the observations located outside of the conjectured break interval Π and then we apply the Step 1.4 of Algorithm 2.

We need to take into account the following perspectives to obtain the proposed Group-LASSO estimator: first, the maximum number of potential factors k. According to Stock and Watson (2012)[24], k is determined by the estimation of the number of factors. However, if the value of k is overestimated, it brings out a large number of potential regressors and drops the efficiency of the shrinkage estimator. If $\widehat{r}_b = k$, then the value of k is set too small. Second, the break date interval Π. The interval of Π is determined by the real world events. For instance, we could set the interval around the year 1984 if we study the breaks of the Great Moderation. In addition, the interval could be set around the year 2007 if we are interested in the Great Recession. Choosing a reasonable length of interval Π would increase the performance of the estimator. Finally, we choose a set of well-behaved $\mathcal{C}$, n_N, and n_T for the Monte Carlo study.

Chapter 5

Numerical Results

In this chapter, we present the Monte Carlo simulation and the results from the experiments. We also analyze the method with the empirical data set of the Great Recession.

5.1 Monte Carlo Simulation

Monte Carlo simulation relies on the repeated random sampling and statistical analysis to compute the results (Raychaudhuri, 2008)[21]. This type of simulation has been widely used for the solution of large, complex systems when analytical approximations are not easy to establish (Cruse, 1997)[15]. In this section, we present the Monte Carlo simulations to evaluate the performance of the proposed estimator $\widehat{r}_a$, $\widehat{r}_b$ and $\mathcal{B}$, the mean squared errors (MSEs) of the proposed shrinkage estimators, and the PMS estimators in finite sample. Section 5.1 presents the statistical model and the estimators used in the experiment. Section 5.2 describes the results and explanation from the simulation.

5.1.1 Design of the Statistical Models

This section describes the design of the factor models and simulations. The statistical models of this major paper refer to the paper of Bates, Plagborg-Møller, Stock and Watson (2013)[3] with improvement on adjusting the structural instability and aiming on the large breaks. The factor models are stated as

$$\begin{aligned} \text{Pre-break:}\quad X_{it} = \lambda_i' F_t + e_{it}, \quad F_{t,\ell} = \rho_a F_{t-1,\ell} + u_{t,\ell}, \\ t = 1, \cdots, \lfloor T\pi_0 \rfloor, \quad \ell = 1, \cdots, r_a, \\ \text{Post-break:}\quad X_{it} = \psi_i' \overline{F}_t + e_{it}, \quad \overline{F}_{t,\ell} = \rho_b \overline{F}_{t-1,\ell} + u_{t,\ell}, \\ t = \lfloor T\pi_0 \rfloor + 1, \cdots, T, \quad \ell = 1, \cdots, r_b, \end{aligned} \tag{5.1}$$

where $i = 1, \cdots, N$, $F_t = (F_{t,1}, \cdots, F_{t,r_a})'$, $\overline{F}_t = (\overline{F}_{t,1}, \cdots, \overline{F}_{t,r_b})'$, and $\{u_{t,\ell} : \ell = 1, \cdots, r_b\}$ with $u_{t,\ell} \sim N(0,1)$. To take into account for the temporal and cross-sectional dependence of the idiosyncratic errors, consider that

$$e_{it} = \alpha e_{it-1} + v_{it},\ v_{it} = (v_{1t}, \cdots, v_{NT})' \sim N(0, \Omega), \tag{5.2}$$

where the (i,j)-th element of Ω is $\beta^{|i-j|}$. Note that the processes are mutually independent and are independent and identically distributed $(i.i.d.)$ across t. Let F_0 and $e_0 = (e_{10}, \cdots, e_{N0})'$ denote the initial values of the factors and the idiosyncratic errors respectively, and they are drawn from their stationary distribution. If $r_b = r_a$, then $\overline{F}_{T_0} = F_{T_0}$. If $r_b > r_a$, then $\overline{F}_{T_0} = (F_{T_0}', F_{T_0}^{*\prime})'$, where each element of $F_{T_0}^*$ is drawn independently from the distribution of $F_{t,\ell}$. The parameters $\{N, T, \pi_0, r_a, r_b, \rho_a, \rho_b, \alpha, \beta\}$ are specified later.

To construct the pre-break factor loadings $\{\lambda_i : i = 1, \cdots, N\}$, let $\lambda_i \sim N(0, \Sigma_i)$, where Σ_i is a diagonal matrix with distinct elements $\sigma_i^2(1), \cdots, \sigma_i^2(r_a)$. The sum of these diagonal elements determines the population regression R^2 of X_{it} on the factors. To select R_i^2 for $i = 1, \cdots, N$, Bai and Ng (2002)[4] explain that R_i^2 is homogeneous and set it equal to 0.5 and it also benchmarks the factor model for the simulations.

Another approach is that R_i^2 is adjusted heterogeneously to match the distribution of R^2 values in the empirical data. To obtain the distribution of R^2, we consider the potential break date of the recent recession of data set before December 2007, and we regress each variable to obtain the empirical distribution of R^2. Then, we draw R_i^2 for $i = 1, \cdots, N$ independently from the empirical distribution to construct the pre-break factor loadings λ_i.

With the existence of the structural instability, we construct the post-break factor loadings ψ_i. ψ_i is similar to the proposed λ_i, except that r_a is replaced by r_b, $\mathbb{E}[(\psi_i' \overline{F}_t)^2]/\mathbb{E}[X_{it}^2]{=}R_i^2$ for $t > T_0$, and R_i^2 is calibrated by post-December 2007 subsample heterogeneously.

The simulated time series are normalized to obtain zero mean and unit variance. Next, we use principal components analysis to extract a maximum of $k = 8$ potential factors from either the subsamples or the full sample. For experiments in the known break date case, the estimator $\widehat{r}_a$, $\widehat{r}_b$, and $\widehat{\mathcal{B}}$ are based on the two-step PLS estimator described in Algorithm 1 and we set $n_N = 5$ and choose $n_T = 10$. Normally, the cross-sectional division is a time consuming process because the model selection procedure has to be performed on each cross-sectional regression sample. Given the selected model, the time-series rolling window forecast is the better choice (Cheng et al. 2016)[13].

For experiments in the unknown break date case, the estimation of the triple estimator depends on the adjusted version of Algorithms 1 described in Section 3.6. We consider Π as a discrete set Π_d and the grid size in Π_d is $\tau = 0.01$, a shift by a quarter for a monthly data set of 300 periods, like the data set in the empirical application. Let $\Pi_d = \{\pi_c - 4\tau, \pi_c - 3\tau, \cdots, \pi_c, \cdots, \pi_c + 3\tau, \pi_c + 4\tau\}$, which spans a two-year interval and is symmetric around the true break point π_0. The post-break subsample for the PMS estimator is obtained by the least square estimator of the break point described in Section 4.3.

We compute the mean-squared errors (MSE) for out-of-sample forecasts (MSFE) generated by the selected model. We set the initial vlaue of $y_1 = X_{iT}$. The series to be forecast is written as

$$\begin{aligned} &\text{Pre-break: } y_{t+1} = \varphi_a' F_t + \epsilon_{t+1}, \quad t = 1, \cdots, T_a, \\ &\text{Post-break: } y_{t+1} = \varphi_b' \overline{F}_t + \epsilon_{t+1}, \quad t = T_a + 1, \cdots, T_a + T_b. \end{aligned} \tag{5.3}$$

Suppose that $\epsilon_1, \epsilon_2, \cdots, \epsilon_{T_a+T_b}$ are iid as N(0,1) and independent with the processes $u_{t,\ell}$ and v_{it}, which are mentioned in Section 5.1.1. The loading vector is generated from the distribution $\varphi_a \sim N(0, I_{r_a})$. If there is no structural break, then we have $\varphi_b = \varphi_a$. Considering the existence of the instability, φ_b is drawn independently based on $\varphi_b \sim N(0, I_{r_b})$. To generate the MSFE, we present the model and the factors based on the full sample X. In the pre-break case, we estimate $\varphi_b = \varphi_a$ on the full sample $t = 1, \cdots, T_a + T_b - 1$ and the evaluation of MSFE is based on the prediction of $y_{T_a+T_b+1}$. In the post-break date case, φ_b is estimated on the subsample $t = T_a + 1, \cdots, T_a + T_b - 1$ and the evaluation of MSFE is based on the prediction of $y_{T_a+T_b+1}$.

$$\begin{aligned} \text{MSFE}_{PMS}\big(\widehat{y}_{T_a+T_b+1}\big) &= \mathbb{E}\big[(y_{T_a+T_b+1} - \widehat{y}_{T_a+T_b+1})^2\big] \\ &= \mathbb{E}\big[(X_{Forecast} - \widehat{\varphi}_b' \overline{F}_{T_a+T_b+1})^2\big]. \end{aligned} \tag{5.4}$$

The full-sample estimator is defined as the first r columns of the full sample least squares estimator $\widetilde{\Lambda}_{LS} = T^{-1}X'\widetilde{F}$, where $r = r_a$ if $\mathcal{B}_0 = 0$, which means no structural break and $r = r_a + r_b$ if $\mathcal{B}_0 \neq 0$, which means there exists a structural break.

$$\text{MSFE}_{Full}\big(\widetilde{\Lambda}_{LS}\big) = \mathbb{E}\big[(X_{Forecast} - \widehat{\varphi}_{Full}' \widetilde{F})^2\big]. \tag{5.5}$$

The relative MSFE depends on the MSFE of the predictor of PMS estimator to the MSFE of the predictor of the full-sample estimation. The calculation of the relative MSFE for full-sample and PMS estimator is summarized as follows:

$$\text{Relative MSFE} = \frac{MSFE_{PMS}\big(\widehat{y}_{T_a+T_b+1}\big)}{MSFE_{Full}\big(\widetilde{\Lambda}_{LS}\big)} \tag{5.6}$$

We expect the values of relative MSFE to be less than 1 because the proposed PMS estimator is more accurate. Moreover, the relative MSFE is less than 1 indicates that the PSM predictor dominates the full-sample predictor.

5.1.2 Results for Shrinkage Estimator

In this section, we illustrate the results from three different types of Monte Carlo experiments in Table 5.1.

Table 5.1: Monte Carlo Experiments

Exp.	π_0	α, β	Break Point
1	0.5	0.2	Known
2	0.8	0.2	Unknown, Known
3	0.5	0.5	Known

In the first experiment, the regression R^2 is homogeneous across all series, we assume that the break date is located at $\pi_0 = 0.5$ and cross-sectional correlation $\alpha = \beta = 0.2$. In the second experiment, we consider the known and unknown break date case. The regression R^2 is heterogeneous across the series and $\pi_0 = 0.8$ indicates that the break occurs at the end of the sample. The third experiment is similar to the first one but the cross-sectional correlation $\alpha = \beta = 0.5$. Overall, we set the temporal correlation to $\rho_a = \rho_b = 0.5$ and all results are based on averages over 1,000 Monte Carlo runs. (Cheng et al. (2016))[13]

First, we display the Monte Carlo results for Experiment 1 in Table 5.2.

Table 5.2: Known Break Point, Homogeneous R^2, $\pi_0 = 0.5$

Model Configuration					Model Selection			Relative
r_a	r_b	w	N	T	$\Pr(\widehat{\mathcal{B}} = \mathcal{B})$	$\Pr(\widehat{r}_a = r_a)$	$\Pr(\widehat{r}_b = r_b)$	MSFE
Panel A. No Break								
3	3	0	115	115	1.00	1.00	1.00	1.00
3	3	0	160	160	1.00	1.00	1.00	1.00
3	3	0	190	190	1.00	1.00	1.00	1.00
Panel B. Type 2-Instability								
1	2	0	115	115	1.00	1.00	1.00	0.96
1	2	0	160	160	1.00	1.00	1.00	1.12
1	2	0	190	190	1.00	1.00	1.00	0.93
3	4	0	115	115	1.00	1.00	1.00	0.90
3	4	0	160	160	1.00	1.00	1.00	0.96
3	4	0	190	190	1.00	1.00	1.00	0.72

Notes: Cross-sectional correlation $\alpha = \beta = 0.2$; temporal correlation $\rho_a = \rho_b = 0.5$.

Table 5.2 contains two panels, corresponding to no break and the instability. Under this instability, we consider the changes of the number of factors from 1 to 2 and 3 to 4, and $w = 0$. Various values of N and T are included in the experiment. We present the probability of correctly estimating $\mathcal{B}$, r_a, r_b. The last column contains the MSFE of the predictor based on the PMS estimator relative to the predictor based on the full-sample least square estimator, where the number of factors is set to r_a for Panel (A), and to $r_a + r_b$ for Panel (B). We expect values less than 1 among relative MSFE because the proposed PMS predictor is more accurate. If the break date is known, the procedure correctly detects the break date, as well as if the break date is located in the middle of the sample ($\pi_0 = 0.5$). We obtain that the probability of correctly estimating $\mathcal{B}$, r_a and r_b equal to 1, which means that the procedure has generally no problem detecting the existence of the instability.

The last column of Table 5.2 shows the relative MSFEs. For the no-break date case, the procedure of estimating r_a, r_b, and $\mathcal{B}$ correctly with probability 1, which indicates that the PMS estimator is identical to the full-sample predictor. Due to the large number of estimated parameters, the predictor is slightly less accurate than the PMS predictor. The proposed PMS predictor is generally accurate since all the relative MSFE values are less than 1, except in the case of $N = T = 160$ in Panel B, where the value of MSFE is slightly greater than 1. Therefore, the PMS predictor weakly dominates the full-sample predictor.

We also give Monte Carlo results for Experiment 2 with unknown break point in Table 5.3.

Table 5.3: Unknown Break Point, Heterogeneous R^2

Model Configuration					Model Selection			Relative
r_a	r_b	w	N	T	$\Pr(\widehat{\mathcal{B}}^* = \mathcal{B})$	$\Pr(\widehat{r}_a^* = r_a)$	$\Pr(\widehat{r}_b^* = r_b)$	MSFE
Panel A. No Break								
3	3	0	100	175	1.00	1.00	1.00	1.00
3	3	0	100	225	1.00	1.00	1.00	1.00
3	3	0	175	275	1.00	1.00	1.00	1.00
Panel B. Instability								
1	2	0	100	175	1.00	1.00	1.00	0.69
1	2	0	100	225	1.00	1.00	1.00	0.23
1	2	0	175	275	1.00	1.00	1.00	1.15
3	4	0	100	175	0.00	0.50	0.00	0.49
3	4	0	100	225	0.50	1.00	0.50	0.75
3	4	0	200	400	1.00	1.00	1.00	0.28

Notes: Cross-sectional correlation $\alpha = \beta = 0.2$; temporal correlation $\rho_a = \rho_b = 0.5$.

Table 5.3 shows that the heterogeneous regression R^2 and the model selection procedure in the unknown break date case is less accurate. When the break date is unknown, the ranking of the PMS estimator and the full-sample predictor is unclear (Cheng et al. 2016[13]). In the no-break case, the procedure correctly determines $\mathcal{B}$,

r_a and r_b for all sample sizes. Indeed, the shrinkage procedure correctly determines the absence of the break and the PMS estimator is the same as the full-sample predictor.

There is no trouble to detect the existence of instability in the procedure when the number of factors changes from 1 to 2 if the break date is unknown. However, when $N = 100$ and $T = 175$ and the number of factors changes from 3 to 4, the probability of estimating $\mathcal{B}$ and r_b are zero and for estimating r_a is 0.5 in Panel B. Once we increase the sample size to $N = 200$ and $T = 400$, the probabilities increase to 1 eventually.

Next, we report the results for the known break date case of Experiment 2.

Table 5.4: Known Break Point, Heterogeneous R^2, $\pi_0 = 0.8$

Model Configuration					Model Selection			Relative
r_a	r_b	w	N	T	$\Pr(\widehat{\mathcal{B}} = \mathcal{B})$	$\Pr(\widehat{r}_a = r_a)$	$\Pr(\widehat{r}_b = r_b)$	MSFE
Panel A. No Break								
3	3	0	100	175	1.00	1.00	1.00	1.00
3	3	0	100	225	1.00	1.00	1.00	1.00
3	3	0	175	275	1.00	1.00	1.00	1.00
Panel B. Instability								
1	2	0	100	175	1.00	1.00	1.00	1.47
1	2	0	100	225	1.00	1.00	1.00	0.17
1	2	0	175	275	1.00	1.00	1.00	1.02
3	4	0	100	175	0.50	1.00	0.50	0.90
3	4	0	100	225	0.50	1.00	0.50	0.57
3	4	0	180	330	1.00	1.00	1.00	0.73

Notes: Cross-sectional correlation $\alpha = \beta = 0.2$; temporal correlation $\rho_a = \rho_b = 0.5$.

Table 5.4 displays the heterogeneous regression R^2 and the model selection procedure in the known break date case is generally accurate. Under the no-break point case, the results are equivalent to Experiment 1 and the PMS estimator is the same as the full-sample predictor.

There is no trouble to detect the existence of instability in the procedure when the number of factors changes from 1 to 2 if the break date is known. However, when the number of factors changes from 3 to 4, $N = 100$, $T = 175$ and $T = 225$, the probability of estimating $\mathcal{B}$ and r_b are 0.5 and for estimating r_a is 1 in Panel B. Once we increase the sample size to $N = 180$ and $T = 330$, the probabilities increase to 1 eventually. Overall, $\mathcal{B}$, r_a and r_b are correctly determined with probability 1. We conclude that the model selection procedure is generally accurate.

The last column of Table 5.4 presents the relative MSFEs. We notice that under the existence of instability, all the relative MSFEs are less than 1, which indicates that the PMS predictor weakly dominates the full-sample predictor. However, for the case of $N = 100$ and $T = 175$, as well as $N = 175$ and $T = 275$, the values of MSFEs are slightly greater than 1. Therefore, the procedure generally has no problem in detecting the existence of the instability if the break date is known and located in the end of the sample.

Finally, we present the results of Experiment 3, which is similar to Experiment 1 but with stronger cross-sectional correlation.

Table 5.5: Known Break Point, Homogeneous R^2, $\pi_0 = 0.5$

Model Configuration					Model Selection			Relative
r_a	r_b	w	N	T	$\Pr(\widehat{\mathcal{B}} = \mathcal{B})$	$\Pr(\widehat{r}_a = r_a)$	$\Pr(\widehat{r}_b = r_b)$	MSFE
Panel A. No Break								
3	3	0	115	115	1.00	1.00	1.00	1.00
3	3	0	160	160	1.00	1.00	1.00	1.00
3	3	0	190	190	1.00	1.00	1.00	1.00
Panel B. Instability								
1	2	0	115	115	1.00	1.00	1.00	1.08
1	2	0	160	160	1.00	1.00	1.00	1.21
1	2	0	190	190	1.00	1.00	1.00	1.12
3	4	0	115	115	0.50	1.00	0.50	0.92
3	4	0	160	160	1.00	1.00	1.00	0.54
3	4	0	190	190	1.00	1.00	1.00	0.67

Notes: Cross-sectional correlation $\alpha = \beta = 0.5$; temporal correlation $\rho_a = \rho_b = 0.5$.

Table 5.5 is similar to Table 5.1 with the same break date $\pi_0 = 0.5$ but different cross-sectional correlation $\alpha = \beta = 0.5$. The results in the no-break date case turn out to be identical to Experiment 1. However, when $N = T = 115$ and the number of factors changes from 3 to 4, the probability of estimating $\mathcal{B}$ and r_b are less than 1. The procedure has generally no problem detecting the existence of the instability if the break date is known and located in the middle of the sample.

The last column of Table 5.5 shows the relative MSFEs. For no-break date case, the probability of model selection equals to 1, which indicates that the PMS estimator is identical to the full-sample predictor. The proposed PMS predictor is generally accurate since half of the relative MSFE values are less than 1, the others are slightly greater than 1.

5.2 Real Data Set: the Great Recession

The Great Recession in the U.S. started in the winter of 2007, after 18 months of recession, growth returned to the U.S. economy in the summer of 2009. As of 2011, the recession was not officially over and kept affecting lives in the form of high employment rate, a host of associated labor-market problems, and ongoing threat of a double-dip recession (Grusky et al. 2011)[17]. It was the longest postwar recession and the associated labor-market dislocations were especially severe. From May 2007 to October 2009, the labor force lost over 7.5 million jobs, and the employment rate climbed from 4.4% to 10.1% (Grusky et al. 2011)[17]. Unlike many other postwar recessions, the disruption of borrowing and lending played an important role in the 2007-2009 recession (Cheng et al. 2016)[13].

We apply Group-LASSO method which developed in previous chapters to investigate the stability of factor loadings and the emergence of new factors. Section 5.2.1 describes the real data set we use for the empirical analysis. Section 5.2.2 presets the empirical results of detecting the break date of the Great Recession.

5.2.1 Some Preliminary Transformations

Stock and Watson (2012)[24] edited a set of 200 macroeconomic and financial indicators. Let X_t denotes the observation of the macroeconomic and financial indicators N, observed over time periods $t = 1, \cdots, T$, where T denotes the number of the months. For instance, those financial indicators are real personal consumption expenditure or unemployment rate and level etc. The list of the description of those financial indicators are presented in the Appendix C.

We use this data set for the empirical analysis. They eliminate 68 replicate indicators from 200 in total to avoid double counting of the data. The new data set is named SW132. We extend the series in the SW132 data set to 2012:M12, using May 2013 data vintages. The first four digits denote the year, the letter M and last two digits denote the certain month. For example, 2012:M12 means the break date is December of 2012. We replace the quarterly series in SW132 by the monthly counterparts, if available. This is possible for the consumption of nondurable, services, and durables; for nonresidential investment; and for 16 price series. We remove the remaining quarterly series for which no monthly observations are available. We add two statistical model components that are available at monthly frequency: change in private inventory and wage and salary disbursements. Following Stock and Watson (2012)[24], we remove local means from all series using a bi-weight kernel with a bandwidth of 100 months, the local means are approximately the same as the ones obtained by a centered moving average of ± 70 months. After making these modifications, the data set consists of $N = 102$ series of monthly macroeconomic and financial indicators. The sample begins after the Great Moderation and ranges from 1985:M1 to 2013:M1 ($T = 337$).

5.2.2 Analyze the Results

The empirical analysis is considered in the unknown break date case. We apply the adjustments of the procedure described in Section 4.5. During the empirical analysis, we fix the number of potential factors to $k = 8$ and use the cross-validation procedure with $n_N = 5$ and $n_T = 10$ (Cheng et al. (2016))[13]. The model selection results are reported in Table 5.6.

Table 5.6: Model Selection, T_c=2012:M12

Interval	Factors		Break Dates	
Size	$\widehat{r}_a$	$\widehat{r}_b$	Least Sq.	Revised
0	1	2	2007:M12	2007:M12
3	1	2	2007:M9	2007:M12
6	1	2	2007:M6	2007:M12
9	1	3	2007:M3	2007:M12

Notes: We center the interval Π at 2007:M12 and use the averaging penalty functions $P_1^*(\Lambda)$ and $P_2^*(\Lambda)$ defined in (4.3) where the average is taken over the interval 2007:M12 ± Size.

Suppose that T_c is the beginning of the Great Recession according to the business cycle dating of the National Bureau of Economic Research (NBER). We select 4 different sets of potential break dates, which are located around the potential break date T_c = 2007:M12. For example, if Size = 0, the set Π corresponds to a single month of the potential break date 2007:M12. In this situation, we obtain 1-month period and consider the break date happens in that time. If Size = 3, the set of potential break dates are located in the range of 2007:M9 and 2008:M3. If Size = 6, the set of potential break dates are located in the range of 2007:M6 to 2008:M6. If Size = 9, the set of potential break dates are located in the range of 2007:M3 to 2008:M9.

For each choice of Π, we obtain $\widehat{r}_a = r_a = 1$ and either $\widehat{r}_b = r_b = 2$ or $\widehat{r}_b = r_b = 3$. Clearly, the procedure provides an evidence of a structural change in the number of factors. In the fourth column of Table 5.6, we present the least squares estimation of the break date by definition in (4.9). In addition, we minimize the least square criterion over the interval Π, which is stated in the first column. It turns out that the minimum is always attained at the boundary of Π (Cheng et al. 2016)[13]. To make the result more precise, we consider the method in Section 2.2 by Breitung and Eickmeier (2011)[8]. They explain that the sum of pre- and post-break factors is minimized at the true break date. Thus, for each break date in a given Π, we compute $\widehat{r}_a + \widehat{r}_b$ and check whether the minimum over the given interval is attained at T_c = 2007:M12. If the minimum is attained at T_c, we set the revised break date equal to the potential break date T_c. If the minimum is not attained at T_c, we consider the revised break date as the date closest to the potential break date T_c and the minimum is attained. Overall, for all choices of Π, the procedure detects the break date correctly so that there is no need to revise the potential break date T_c.

Moreover, we consider another approach to obtain the probability of correctly estimating $\mathcal{B}$, r_a and r_b, as well as the value of relative MSFE. Specifically, the Bootstrap is a proposed method for the case of sampling from a finite population with replacement. It is shown that for a large number of practical situations, the proposed method works as a natural extension of standard bootstrap method. It turns out that bootstrap works for sample mean, sample quantile, t-statistics, empirical processes and some linear combinations of order statistics. (Chao and Lo 1985)[12]

We use the bootstrap method in Matlab R2018a to generate 500 data sets based on the original data set of SW132. Each of the data sets has the same properties as the data set of SW132. To obtain the value of r_a and r_b, we take the average of both $\widehat{r}_a$ and $\widehat{r}_b$ we generated and round to the nearest integer respectively. Table 5.7 presents the bootstrap results.

Table 5.7: Bootstrap Results

Model Configuration		Model Selection			Relative
N	T	$\Pr(\widehat{r}_a^* = r_a)$	$\Pr(\widehat{r}_b^* = r_b)$	$\Pr(\widehat{\mathcal{B}}^* = \mathcal{B})$	MSFE
337	102	0.41	0.52	0.85	0.85

In Table 5.7, the probability of correctly estimating r_a and r_b are less accurate, we obtain the probability only around 41% and 52% respectively. Although the probabilities of correctly estimating the number of factors are lower, the probability of correctly estimating the break date is higher and reaches around 85%. We expect the value of MSFE is less than 1 and obtain the value of MSFE is 0.85, thus, we conclude that the PMS predictor dominates the full-sample predictor. Overall, the procedure generally detects the break date correctly for unknown break date case.

Chapter 6

Concluding Remarks

In this major paper, we develop a high-dimensional econometric model, which is capable of estimating the number of pre- and post-break factors with the existence of the instability. The estimator we developed is robust to the instability when the break date is unknown. In addition, the Group-LASSO estimation procedure can detect the changes in the factor loadings when the number of factors is constant in the sample. We demonstrate that when the number of pre- and post-break factors are determined, the break date can be estimated by the least square approach.

Moreover, by the Monte Carlo simulation, we demonstrate that the Group-LASSO estimation procedure generally has no problem in detecting the existence of the instability. Also, the procedure is designed to determine the number of factors when there is no break in the sample and to detect the break in the factor loadings when the number of factor is known. When the break date is unknown, the procedure can estimate the break date correctly.

In the real data set of the Great Recession in the U.S., we estimate the potential break date precisely by minimizing the sum of the number of the pre- and post-break factors. Under unknown break date case, the break date can be estimated correctly in general. From what has discussed in the empirical analysis, the proposed procedure detects the increase in the number of factors, which provides an evidence of a structural change in the number of factors.

Appendix A

Some Statistical Background

In this appendix, we give some definitions and lemmas used in deriving the main results of this major paper.

Definition A.1 (Casella and Berger (2002)[11]). *A sequence of random variables* $\{X_n\}_{n=1}^{\infty}$ *converges in probability to a random variable* X *if, for every* $\epsilon > 0$, $\lim_{n\to\infty} \Pr(|X_n - X| \geq \epsilon) = 0$. *We denote it as* $X_n \xrightarrow[n\to\infty]{\text{p}} X$.

Definition A.2 (Bickel and Doksum (2001)[7]). *A sequence of random vectors* $Z_n = (Z_{n1}, Z_{n2}, \cdots, Z_{nm})'$ *converges in probability to* $Z = (Z_1, Z_2, \cdots, Z_m)'$ *iff* $Z_{nj} \xrightarrow[n\to\infty]{\text{p}} Z_j$ *for* $1 \leq j \leq m$. *We denote it as* $Z_n \xrightarrow[n\to\infty]{\text{p}} Z$.

Lemma A.1 (Strawderman (1993)[25]). *Let* A_n *be a random sequence of symmetric nonnegative definite* $k \times k$ *matrices where* $k < \infty$. *If a positive definite symmetric* $k \times k$ *matrix* A *with finite elements exists such that* $A_n \xrightarrow[n\to\infty]{\text{p}} A$ *element-wise, then* $\|A_n - A\| \xrightarrow[n\to\infty]{\text{p}} 0$, *where* $\|.\|$ *denotes any proper norm on* $\mathbb{R}^{k\times k}$.

The proof of this lemma is given in Lemma 1 of Strawderman(1999)[24].

Appendix B

Some Proofs

In this appendix, we give the proof of lemma and theorem used in the derivation of the major paper.

Proof of Lemma 3.1. Let $\Upsilon_a \in R^{r_a \times r_a}$ be a matrix of orthonormal eigenvectors, note that $\Upsilon_a' \Upsilon_a = I_{r_a}$ implies that $\Upsilon_a' = \Upsilon_a^{-1}$. Let $\Sigma_a^{1/2}$ be the Cholesky factor of Σ_a, where $\Sigma_a = \Lambda^{0'} \Lambda^0 / N$. By definition,

$$\Upsilon_a' (\Sigma_a^{1/2})' \Sigma_F \Sigma_a^{1/2} \Upsilon_a = V_a. \tag{B.1}$$

Therefore, V_a is a diagonal matrix of eigenvalues, ordered from largest to smallest. Let $\Lambda^R = \Lambda^0 R_a'^{-1}$ and $\Psi^R = \Psi^0 R_b'^{-1}$ and define the transformation matrix $R_a = \Sigma_a^{1/2} \Upsilon_a V_a^{-1/2}$ for pre-break date case. For post-break date case, let $\Sigma_b = \Psi^{0'} \Psi / N \in R^{r_b \times r_b}$, substitute Σ_F in (B.2) by $\Sigma_{\overline{F}}$, and replace a-subscripts by b-subscripts. Thus, the second transformation matrix for R_b is defined as $R_b = \Sigma_b^{1/2} \Upsilon_b V_b^{-1/2}$. Then, we have

$$\frac{\Lambda^{R'}\Lambda^{R}}{N} = V_a^{1/2}\Upsilon_a'\Sigma_a^{-1/2}\frac{\Lambda^{0'}\Lambda^{0}}{N}\Sigma_a^{-1/2}\Upsilon_a V_a^{1/2} = V_a \text{ and } \frac{\Psi^{R'}\Psi^{R}}{N} = V_b \tag{B.2}$$

The ℓ-th diagonal element of V_a is the ℓ-th largest eigenvalue of $\Sigma_a^{1/2}\Sigma_F\Sigma_a^{1/2}$. Let ρ^* denotes the eigenvalue of $\Sigma_a^{1/2}\Sigma_F\Sigma_a^{1/2}$. One can verify that ρ^* is also the eigenvalue of $\Sigma_a\Sigma_F$. Therefore, the ℓ-th diagonal element of V_a is the same as the ℓ-th largest eigenvalue of $\Sigma_a\Sigma_F$. Recall that in Assumption 3, there exists a positive definite matrix $\Sigma_\Lambda \in R^{r_a \times r_a}$ such that $\|\Sigma_a - \Sigma_\Lambda\| \to 0$ as $N \to \infty$. Here, Σ_a is a sequence of symmetric positive definite matrix. By Lemma 1 in Strawderman (1993)[25], we have

$$\Sigma_a \xrightarrow[N\to\infty]{\text{p}} \Sigma_\Lambda. \tag{B.3}$$

Then, we have

$$\Sigma_a\Sigma_F \xrightarrow[N\to\infty]{\text{p}} \Sigma_\Lambda\Sigma_F \tag{B.4}$$

where $\Sigma_F \in R^{r_a \times r_a}$ is positive definite matrix. Carl de Boor (2002)[16] proved that the convergence of matrices is entry-wise such that

$$\left(\Sigma_a\ \Sigma_F\right)_{i,j} \xrightarrow[N\to\infty]{\text{p}} \left(\Sigma_\Lambda\ \Sigma_F\right)_{i,j} \text{ for all } i,j. \tag{B.5}$$

In addition, Alexanderian (2013)[2] mentioned that the characteristic roots, which are eigenvalues, of a polynomial depend continuously on its entries. Then, we have

$$\lambda_\ell(\Sigma_a\Sigma_F) \xrightarrow[N\to\infty]{\text{p}} \lambda_\ell(\Sigma_\Lambda\Sigma_F) \tag{B.6}$$

Therefore, the ℓ-th largest eigenvalue of $\Sigma_a\Sigma_F$ converges to the ℓ-th largest eigenvalue of $\Sigma_\Lambda\Sigma_F$ as $N \to \infty$, denoted by $\rho_\ell(\Sigma_\Lambda\Sigma_F)$. Similarly, the ℓ-th diagonal element of V_b converges to the ℓ-th largest eigenvalue of $\Sigma_\Psi\Sigma_{\overline{F}}$ as $N \to \infty$, denoted by $\rho_\ell(\Sigma_\Psi\Sigma_{\overline{F}})$.

Let a_ℓ be a selection vector that selects the ℓ-th column of a matrix, note that a_ℓ' selects the ℓ-th row of a matrix. Part (i) holds because

$$N^{-1}\|\Lambda_\ell^R\|^2 = N^{-1}(\Lambda_\ell^{R'}\Lambda_\ell^R) = a_\ell'\frac{\Lambda^{R'}\Lambda^R}{N}a_\ell = a_\ell'V_a a_\ell = \rho_\ell(\Sigma_\Lambda\Sigma_F) + o(1). \tag{B.7}$$

To prove part (ii), note that for $r_a < \ell < r_b$, the ℓ-th column of Γ^R is equivalent to the ℓ-th column of Ψ^R. Therefore,

$$N^{-1}\|\Gamma_\ell^R\|^2 = N^{-1}(\Psi_\ell^{R'}\Psi_\ell^R) = a_\ell'\frac{\Psi^{R'}\Psi^R}{N}a_\ell = a_\ell'V_b a_\ell = \rho_\ell(\Sigma_\Psi\Sigma_{\overline{F}}) + o(1). \tag{B.8}$$

This completes the proof. □

Proof of Theorem 3.4. First, we need to prove $\Pr(\widehat{r}_a \geq r_a) \to 1$ as $N, T \to \infty$. Theorem 3.3(i) and Lemma 3.1(i) indicate that

$$N^{-1/2}\|\widehat{\Lambda}_\ell - \Lambda_\ell^R\| = O_p(C_{NT}^{-1}) \text{ for } \ell = r_a \tag{B.9}$$

and

$$N^{-1/2}\|\Lambda_\ell^R\| = [\rho_\ell(\Sigma_\Lambda\Sigma_F)]^{1/2} + o(1) \text{ for } \ell = 1, \cdots, r_a. \tag{B.10}$$

By the triangle inequality, we have

$$O_p(C_{NT}^{-1}) = N^{-1/2}\|\widehat{\Lambda}_\ell - \Lambda_\ell^R\| \geq |N^{-1/2}\|\widehat{\Lambda}_\ell\| - N^{-1/2}\|\Lambda_\ell^R\||. \tag{B.11}$$

Then, we get rid of the absolute value on the right hand side,

$$\begin{aligned} -N^{-1/2}\|\widehat{\Lambda}_\ell - \Lambda_\ell^R\| &\leq N^{-1/2}\|\widehat{\Lambda}_\ell\| - N^{-1/2}\|\Lambda_\ell^R\| \\ N^{-1/2}\|\Lambda_\ell^R\| - N^{-1/2}\|\widehat{\Lambda}_\ell - \Lambda_\ell^R\| &\leq N^{-1/2}\|\widehat{\Lambda}_\ell\| \\ [\rho_\ell(\Sigma_\Lambda\Sigma_F)]^{-1/2} + o(1) &\leq N^{-1/2}\|\widehat{\Lambda}_\ell\| + O_P(C_{NT}^{-1}). \end{aligned} \tag{B.12}$$

Since Σ_Λ and Σ_F are positive definite matrices, we have

$$\Pr(\|\widehat{\Lambda}_\ell\| > 0) \to 1 \text{ as } N, T \to \infty \text{ for } \ell = r_a. \tag{B.13}$$

Here, the r_a-th column of $\widehat{\Lambda}$ has value greater than 0 with probability approaching to 1. By definition of $\widehat{r}_a$ in (3.7), the $\widehat{r}_a$-th column is the largest column where the column of $\widehat{\Lambda}$ has the value not equal to 0. Therefore, $\Pr(\widehat{r}_a \geq r_a) \to 1$ as $N, T \to \infty$.

Second, we need to prove $\Pr(\widehat{r}_a \leq r_a) \to 1$ as $N, T \to \infty$.
Theorem 3.3(ii) indicates that

$$\Pr(\|\widehat{\Lambda}_\ell\| = 0 \text{ for } \ell = r_a + 1, \cdots, k) \to 1 \text{ as } N, T \to \infty. \tag{B.14}$$

Here, the $(r_a + 1)$-th to k-th column of $\widehat{\Lambda}$ have the value of 0 with probability approaching to 1. the definition of $\widehat{r}_a$ in (3.7), Then, we have $\Pr(\widehat{r}_a \leq r_a) \to 1$ as $N, T \to \infty$. Therefore, we have $\lim\limits_{N,T\to\infty} \Pr(\widehat{r}_a = r_a) = 1$.

Third, we consider the case under the existence of the instability and we need to prove $\Pr(\widehat{r}_b \geq r_b) \to 1$ as $N, T \to \infty$. The procedure is similar to the first step. With the existence of the instability where $r_b > r_a$ and $\mathcal{B}_0 = 1$. Theorem 3.3(iii) for $\ell = r_b$ and Lemma 3.1(ii) imply that $\Pr(\|\widehat{\Gamma}_\ell\| > 0) \to 1$ as $N, T \to \infty$ for $\ell = r_b$, together with the definition of $\widehat{r}_b$ in (3.7), hence, $\Pr(\widehat{r}_b \geq r_b) \to 1$ as $N, T \to \infty$.

To prove $\Pr(\widehat{r}_b \leq r_b) \to 1$ as $N, T \to \infty$, the procedure is similar to the second step. Theorem 3.3 (v) and definition of $\widehat{r}_b$ in (3.7) imply that $\Pr(\widehat{r}_b \leq r_b) \to 1$ as $N, T \to \infty$, since with the existence of the instability, $r_b > r_a$ and $\mathcal{B} = 1$ imply $\widehat{r}_b > \widehat{r}_a$ and $\widehat{\mathcal{B}} = 1$. Hence, $\lim_{N,T\to\infty} \Pr(\widehat{r}_b = r_b) = 1$ for a the existence of the instability.

Now, we need to prove $\Pr(\widehat{\mathcal{B}} = 1)$ as $N, T \to \infty$ with the existence of the instability where $r_b > r_a$. We have $\lim_{N,T\to\infty} \Pr(\widehat{r}_b = r_b) = 1$ and $\lim_{N,T\to\infty} \Pr(\widehat{r}_a = r_a) = 1$ proved in second step. By definition of $\widehat{r}_b$ in 3.7 and Theorem 3.3(v), we have

$$\left\{\|\widehat{\Gamma}_{r_b}\| > 0\right\} \subset \left\{\|\widehat{\Gamma}\| > 0\right\}, \tag{B.15}$$

then, we have

$$\Pr(\|\widehat{\Gamma}_{r_b}\| > 0) \leq \Pr(\|\widehat{\Gamma}\| > 0) \leq 1. \tag{B.16}$$

Since the $\widehat{r}_b$-th column is the largest column where the column of $\widehat{\Gamma}$ has value not equal to 0 and we proved that $\Pr(\widehat{r}_b \geq r_b) \to 1$ as $N, T \to \infty$, then, we have

$$\Pr(\|\widehat{\Gamma}_{r_b}\| > 0) \to 1 \text{ as } N, T \to \infty. \tag{B.17}$$

Thus, the inequality in (B.16) can be written as

$$\Pr(\|\widehat{\Gamma}\| > 0) \to 1 \text{ as } N, T \to \infty. \tag{B.18}$$

From (B.15), we also have

$$\begin{aligned} \Big\{\mathcal{I}_{\|\widehat{\Gamma}_{r_b}\|>0} = 1\Big\} &\subset \Big\{\mathcal{I}_{\|\widehat{\Gamma}\|>0} = 1\Big\}, \\ \Pr\Big(\mathcal{I}_{\|\widehat{\Gamma}_{r_b}\|>0} = 1\Big) &\leq \Pr\Big(\mathcal{I}_{\|\widehat{\Gamma}\|>0} = 1\Big) \leq 1. \end{aligned} \tag{B.19}$$

Consider the two events $\Big\{\|\widehat{\Gamma}\| = 0\Big\}$ and $\Big\{\|\widehat{\Gamma}\| > 0\Big\}$. We have $\Big\{\|\widehat{\Gamma}\| = 0\Big\} \cap \Big\{\|\widehat{\Gamma}\| > 0\Big\} = \varnothing$ and $\Pr(\|\widehat{\Gamma}\| = 0) + \Pr(\|\widehat{\Gamma}\| > 0) = 1$. By the law of total probability,

$$\begin{aligned} \Pr\Big(\mathcal{I}_{\|\widehat{\Gamma}_{r_b}\|>0} = 1\Big) = {} & \Pr\Big(\big\{\mathcal{I}_{\|\widehat{\Gamma}_{r_b}\|>0} = 1\big\} \cap \big\{\|\widehat{\Gamma}\| > 0\big\}\Big) + \\ & \Pr\Big(\big\{\mathcal{I}_{\|\widehat{\Gamma}_{r_b}\|>0} = 1\big\} \cap \big\{\|\widehat{\Gamma}\| = 0\big\}\Big), \end{aligned} \tag{B.20}$$

then

$$\begin{aligned} \Pr\Big(\mathcal{I}_{\|\widehat{\Gamma}_{r_b}\|>0} = 1\Big) = {} & \Pr\Big(\big\{\|\widehat{\Gamma}_{r_b}\| > 0\big\} \cap \big\{\|\widehat{\Gamma}\| > 0\big\}\Big) + \\ & \Pr\Big(\big\{\|\widehat{\Gamma}_{r_b}\| > 0\big\} \cap \big\{\|\widehat{\Gamma}\| = 0\big\}\Big), \end{aligned}$$

and then,

$$\begin{aligned} \lim_{N,T\to\infty} \Pr\Big(\mathcal{I}_{\|\widehat{\Gamma}_{r_b}\|>0} = 1\Big) = {} & \lim_{N,T\to\infty} \Pr\Big(\|\widehat{\Gamma}_{r_b}\| > 0\Big) + \\ & \lim_{N,T\to\infty} \Pr\Big(\big\{\|\widehat{\Gamma}_{r_b}\| > 0\big\} \cap \big\{\|\widehat{\Gamma}\| = 0\big\}\Big), \end{aligned}$$

this gives

$$\lim_{N,T\to\infty} \Pr\Big(\mathcal{I}_{\|\widehat{\Gamma}_{r_b}\|>0} = 1\Big) = \lim_{N,T\to\infty} \Pr\Big(\|\widehat{\Gamma}_{r_b}\| > 0\Big) + 0,$$

together with (B.17), therefore,

$$\lim_{N,T\to\infty} \Pr\Big(\mathcal{I}_{\|\widehat{\Gamma}_{r_b}\|>0} = 1\Big) = 1. \tag{B.21}$$

Finally, in (B.19) we have

$$\lim_{N,T\to\infty} \Pr\Big(\mathcal{I}_{\|\widehat{\Gamma}_{r_b}\|>0} = 1\Big) = 1 \leq \lim_{N,T\to\infty} \Pr\Big(\mathcal{I}_{\|\widehat{\Gamma}\|>0} = 1\Big) \leq 1,$$

this gives,

$$\lim_{N,T\to\infty} \Pr\Big(\mathcal{I}_{\|\widehat{\Gamma}\|>0} = 1\Big) = 1, \tag{B.22}$$

which implies that $\lim_{N,T\to\infty} \Pr(\widehat{\mathcal{B}} = 1) = 1$. Therefore, we complete the proof of Theorem 3.3.

Fourth, we need to prove $\Pr(\widehat{r}_b = r_b) \to 1$ as $N, T \to \infty$ and $\Pr(\widehat{\mathcal{B}} = 0) \to 1$ as $N, T \to \infty$ in no break date case, *i.e.*, $r_a = r_b$ and $\mathcal{B} = 0$. Together with the definition of $\widehat{r}_b$ in (3.7) and the fact that $r_a = r_b$, we conclude that $\Pr(\widehat{r}_b = r_b) \to 1$ as $N, T \to \infty$.

By applying the same procedure from (B.19) to (B.22), we have

$$\lim_{N,T\to\infty} \Pr\big(\mathcal{I}_{\|\widehat{\Gamma}\|>0} = 0\big) = 1. \tag{B.23}$$

Thus, by the definition of $\widehat{\mathcal{B}}$ in (3.6), we conclude that

$$\Pr(\widehat{\mathcal{B}} = \mathcal{B} = 0) \to 1 \text{ as } N, T \to \infty \tag{B.24}$$

Therefore, we complete the proof of Theorem 3.3 in no break date case.

This completes the proof. $\square$

Appendix C

Supplemental Table for Empirical Analysis

In this appendix, we present a table of the macroeconomic and financial indicators that data series used from Cheng et al. (2016, see Supplemental Appendix Tables S3-S5)[13].

Table C.1: List of Financial Indicators-Part I

Name	Long Description
Cons: Dur	Real Personal Consumption Expenditures: Durable Goods
Cons: Svc	Real Personal Consumption Expenditures: Services
Cons: NonDur	Real Personal Consumption Expenditures: Non-durable Goods
Real InvtCh	Component for Change in Private Inventories, deflated by JCXFE
Real WageG	Component for Government GDP: Wage and Salary Disbursements by Industry, Government, deflated by JCXFE
IP: DurGds materials	Industrial Production: Durable Materials
IP: NondurGds materials	Industrial Production: Nondurable Materials
IP: DurConsGoods	Industrial Production: Durable Consumer Goods

Table C.2: List of Financial Indicators-Part II

Name	Long Description
IP: Auto	IP: Automotive products
IP: NonDurConsGoods	Industrial Production: Nondurable Consumer Goods
IP: BusEquip	Industrial Production: Business Equipment
IP: EnergyProds	IP: Consumer Energy Products
CapU Tot	Capacity Utilization: Total Industry
CapU Man	Capacity Utilization: Manufacturing (FRED past 1972)
Emp: DurGoods	All Employees: Durable Goods Manufacturing
Emp: Const	All Employees: Construction
Emp: Edu & Health	All Employees: Education & Health Services
Emp: Finance	All Employees: Financial Activities
Emp: Infor	All Employees: Information Services
Emp: Bus Serv	All Employees: Professional & Business Services
Emp: Leisure	All Employees: Leisure & Hospitality
Emp: OtherSvcs	All Employees: Other Services
Emp:Mining/NatRes	All Employees: Natural Resources & Mining
Emp: Trade&Trans	All Employees: Trade, Transportation & Utilities
Emp: Retail	All Employees: Retail Trade
Emp: Wholesal	All Employees: Wholesale Trade
Emp: Gov(Fed)	All Employees: Government: Federal
Emp: Gov (State)	All Employees: Government: State Government
Emp: Gov (Local)	All Employees: Government: Local Government
URate: Age16-19	Unemployment Rate - 16-19 yrs
URate: Age > 20 Men	Unemployment Rate - 20 yrs. & over, Men
URate: Age > 20 Women	Unemployment Rate - 20 yrs. & over, Women
U: Dur < 5wks	Number Unemployed for Less than 5 Weeks
U: Dur 5-14wks	Number Unemployed for 5-14 Weeks
U: Dur > 15-26wks	Civilians Unemployed for 15-26 Weeks
U: Dur > 27wks	Number Unemployed for 27 Weeks & over
U: Job Losers	Unemployment Level - Job Losers
U: LF Reentry	Unemployment Level - Reentrants to Labor Force
U: Job Leavers	Unemployment Level - Job Leavers
U: New Entrants	Unemployment Level - New Entrants
Emp: SlackWk	Employment Level - Part-Time for Economic Reasons, All Industries
AWH Man	Average Weekly Hours: Manufacturing
AWH Privat	Average Weekly Hours: Total Private Industrie
AWH Overtime	Average Weekly Hours: Overtime: Manufacturing
HPermits	New Private Housing Units Authorized by Building Permit

Table C.3: List of Financial Indicators-Part III

Name	Long Description
Hstarts: MW	Housing Starts in Midwest Census Region
Hstarts: NE	Housing Starts in Northeast Census Region
Hstarts: S	Housing Starts in South Census Region
Hstarts: W	Housing Starts in West Census Region
Constr. Contracts	Construction contracts (mil. sq. ft.) (Copyright, McGraw-Hill)
Ret. Sale	Sales of retail stores (mil. Chain 2000 $)
Orders (DurMfg)	Mfrs new orders durable goods industries (bil. chain 2000 $)
Orders (ConsumerGoods/Mat.)	Mfrs new orders, consumer goods and materials (mil. 1982 $)
UnfOrders (DurGds)	Mfrs unfilled orders durable goods indus. (bil. chain 2000 $)
Orders (NonDefCap)	Mfrs new orders, nondefense capital goods (mil. 1982 $)
VendPerf	Index of supplier deliveries vendor performance (pct.)
MT Invent	Manufacturing and trade inventories (bil. Chain 2005 $)
PCED-MotorVec	Motor vehicles and parts
PCED-DurHousehold	Furnishings and durable household equipment
PCED-Recreation	Recreational goods and vehicles
PCED-OthDurGds	Other durable goods
PCED-Food-Bev	Food and beverages purchased for off-premises consumption
PCED-Clothing	Clothing and footwear
PCED-Gas-Enrgy	Gasoline and other energy goods
PCED-OthNDurGds	Other nondurable goods
PCED-Housing-Utilities	Housing and utilities
PCED-HealthCare	Health care
PCED-TransSvg	Transportation services
PCED-RecServices	Recreation services
PCED-FoodServ-Acc.	Food services and accommodations
PCED-FIRE	Financial services and insurance
PCED-OtherServices	Other services
PPI: FinConsGds	Producer Price Index: Finished Consumer Goods
PPI: FinConsGds(Food)	Producer Price Index: Finished Consumer Foods
PPI: IndCom	Producer Price Index: Industrial Commodities
PPI: IntMat	Producer Price Index: Intermediate Materials: Supplies & Components
NAPM ComPrice	NAPM COMMODITY PRICES INDEX (PERCENT)
Real Price: NatGas	PPI: Natural Gas, deflated by PCEPILFE
Real Price: Oil	PPI: Crude Petroleum, deflated by PCEPILFE
FedFunds	Effective Federal Funds Rate

Table C.4: List of Financial Indicators-Part IV

Name	Long Description
TB-3Mth	3-Month Treasury Bill: Secondary Market Rate
BAA-GS10	BAA-GS10 Spread
MRTG-GS10	Mortg-GS10 Spread
TB6m-TB3m	tb6m-tb3m
GS1-TB3m	GS1-Tb3m
GS10-TB3m	GS10-Tb3m
CP-TB Spread	CP-Tbill Spread: CP3FM-TB3MS
Ted-Spread	MED3-TB3MS (Version of TED Spread)
Real C&I Loan	Commercial and Industrial Loans at All Commercial BanksDefl by PCEPILFE
Real ConsLoans	Consumer (Individual) Loans at All Commercial Banks Outlier Code because of change in data in April 2010 see FRB H8 ReleasDefl by PCEPILFE
Real NonRevCredit	Total Nonrevolving Credit Owned and Securitized, OutstandingDefl by PCEPILFE
Real LoansRealEst	Real Estate Loans at All Commercial BanksDefl by PCEPILFE
Real RevolvCredit	Total Revolving Credit OutstandingDefl by PCEPILFE
S&P500	S&PS COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)
DJIA	COMMON STOCK PRICES: DOW JONES INDUSTRIAL AVERAGE
VXO	VXO (Linked by N. Bloom) .. Average daily VIX from 2009
Ex rate: Major	FRB Nominal Major Currencies Dollar Index (Linked to EXRUS in 1973:1)
Ex rate: Switz	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER USD)
Ex rate: Japan	FOREIGN EXCHANGE RATE: JAPAN (YEN PER USD)
Ex rate: UK	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)
EX rate: Canada	FOREIGN EXCHANGE RATE: CANADA (CAD PER USD)
Cons. Expectations	Consumer expectations NSA (Copyright, University of Michigan)

Bibliography

[1] S. Ejaz Ahmed. (2014). Penalty, Shrinkage and Pretest Strategies-Variable Selection and Estimation. Springer Cham Heidelberg New York Dordrecht London.

[2] Alen Alexanderian. (2013). On Continuous Dependence of Roots of Polynomials on Coefficients. The University of Texas at Austin, USA.

[3] Brandon J. Bates, Mikkel Plagborg-Møller, James H. Stock, and Mark W. Watson. (2012). Consistent Factor Estimation in Dynamic Factor Models with Structural Instability. *Journal of Econometrics*, 177(2), 289-304

[4] Jushan Bai and Serena Ng. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, Volume 70, 191-221

[5] Jushan Bai. (1997). Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics*, 79(4), 551-563

[6] Jushan Bai and Yuan Liao. (2015). Efficient Estimation of Approximate Factor Models via Regularized Maximum Likelihood. *Journal of Econometrics*, 191(2016), 1-18

[7] Peter J. Bickel and Kjell A. Doksum. (2001). Mathematical Statistics, Basic Ideas and Selected Topics, Volume 1, Second Edition. New Jersey: Prentice Hall.

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

[9] Peter Bühlmann and Sara van de Geer. (2011). Statistics for High-Dimensional Data. Springer Heidelberg Dordrecht London New York.

[10] Mehmet Caner and Xu Han. (2014). Selecting the Correct Number of Factors in Approximate Factor Models: the Large Panel Case with Group Bridge Estimators. *Journal of Business & Economic Statistics*, 32(3), 359-374

[11] George Casella and Roger L. Berger. (2002). Statistical Inference. Second Edition. Duxbury Press.

[12] Min-Te Chao and Shaw-Hwa Lo. (1985). A Bootstrap Method for Finite Population. *Sankhyā:The Indian Journal of Statistics*, Volume 47, Series A, 399-405

[13] Xu Cheng, Zhipeng Liao and Frank Schorfheide. (2016). Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities. Oxford University Press, Volume 83(4), 1511-1543

[14] S. Chancharat and Abbas Valadkhani. (2007). Structural breaks and testing for the random walk hypothesis in international stock prices. *Journal of the Korean Economy* 8(1),21-38

[15] Thomas A. Cruse. (1997). Reliability-Based Mechanical Design. Marcel Dekker.

[16] Carl de Boor. (2002). Applied Linear Algebra. UW-Madison Department of Mathematics Publications.

[17] David B. Grusky, Bruce Western, and Christopher Wimer. (2011). The Great Recession. Russell Sage Foundation.

[18] Sévérien Nkurunziza, Marwan Al-Momani, and Eric Yu Yin Lin (2016). Shrinkage and LASSO Strategies in high dimensional Heteroscedastic models. *Communications in Statistics Theory and Methods*, volume 45, 15, 4454-4470

[19] Alexei Onatski. (2010). Determining the Number of Factors from Empirical Distribution of Eigenvalues. *Review of Economics and Statistics*, 92(4), 1004-1016

[20] Pierre Perron. (1989). The Great Crash, the Oil Price Shock and the Unit Root Hypothesis. *Econometrica* 57(6), 1361-1401

[21] Samik Raychaudhuri. (2008). Introduction to Monte Carlo Simulation. Proceedings of the 2008 Winter Simulation Conference.

[22] Sunil Sapra. (2015). High-Dimensional Statistical Analysis in Business and Economics. *Business and Economics Journal*, Volume 6, Issue 3, 6:3

[23] Peter H. Schönemann. (1996). A Generalized Solution of the Orthogonal Procrustes Problem. *Psychometrika*, 31(1), 1-10

[24] James H. Stock and Mark W. Watson. (2012). Disentangling the Channels of the 2007-09 Recession. *Brookings Papers on Economic Activity*, 81-156

[25] Robert L. Strawderman. (1993). A note on necessary and sufficient conditions of proving that a random symmetric matrix converges to a given limit. *Statistics & Probability Letters*, 21(1994), 367-370.

[26] Robert Tibshirani. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 58(1), 276-288

[27] Ming Yuan and Yi Lin. (2006). Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 68(1), 49-67

[28] Hui Zou. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476), 1418-1429

Vita Auctoris

Mr. Yujie Song was born in 1990, Xi'an, Shaanxi, China. He graduated from Georges Vanier Secondary School in 2011. From there he went to University of Toronto, where he obtained a Bachelor of Science with a double major in Economics and Statistics. He is currently a candidate for the Master of Science degree in Statistics at the University of Windsor. He hopes to graduate in August 2018.