ESTIMATION OF HIGH DIMENSIONAL FACTOR MODELS UNDER GENERAL CONDITIONS

A Dissertation

by

YUTANG SHI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Qi Li
Committee Members,	Dennis W. Jansen
	Li Gan
	Ximing Wu
Head of Department,	Timothy Gronberg

May 2016

Major Subject: Economics

ABSTRACT

High dimensional factor models have drawn attention in both empirical and theoretical studies. Correctly specifying the number of factors (r) is a fundamental issue for the application of factor models. We develop an econometric method to estimate the number of factors in factor models of large dimensions where the number of factors is allowed to increase as the two dimensions, cross-section size (N) and time period (T) increase. Using similar information criteria as proposed by Bai and Ng (2002), we show that the number of factors can be consistently estimated using the criteria. We propose a new procedure that avoids overestimating the number of factors while allowing for one to search for possible number of factors over a wide range of positive integers so that it also avoids underestimation of the number of factors. We conduct Monte-Carlo simulation to investigate the finite sample properties of the proposed approach.

The factor loadings are commonly estimated under the presupposition that they do not depend on time. However, this presumption is easily challenged by structural changes or regime shifts. We investigates high dimensional factor models with structural instability in factor loadings. Our inquiry focuses on how to estimate the unknown common break point and derive its limiting distribution. The least squares method is used to estimate the break point in factor loadings. Several competing methods are compared in the simulation. The results show that our proposed approach outperforms other methods. We further propose a new framework to derive the limiting distribution for the estimated change point. The limiting distribution of the estimated break point is more complex than that of the conventional panel data models, because both factors and factor loadings are unobservable. We show that the estimated factors and estimated factor loadings influence the limiting distribution. Based on the limiting distribution of the estimated break point, one can construct confidence intervals of the underlying true break point. Bootstrap method is also studied. We apply the method to the study of structural changes in financial asset returns and in macroeconomic data.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

1.1 Increasing Number of Factors

Factor models have been widely used in economic analyses such as forecasting economic variables, estimating variance-covariance matrices with high dimension data, and estimating average treatment effects. In practice a few common factors may capture the variations of a large number of economic variables. In the finance literature, the arbitrage pricing theory (APT) of Ross (1976) assumes that a small number of factors can be used to explain a large number of asset returns. Stock and Watson (1998, 1999) consider forecasting inflation with diffusion indices ("factors") constructed from a large number of macroeconomic series. Gregory and Head (1999) and Forni, Hallin, Lippi, and Reichlin (2000) find that cross country variations have common components. Fan, Liao and Mincheva (2011), and Fan, Liao and Mincheva (2013) use factor models to estimate high dimensional variance-covariance matrices. Factor models can also be used to evaluate the impacts of various policies (e.g., Hsiao, Ching and Wan (2012)). By assuming that the cross-sectional correlations for all the units are attributed to the presence of some (unobserved) common factors, Hsiao, Ching and Wan (2012), Ching, Hsiao and Wan (2012) and Bai, Li and Ouyang (2014) use panel data methods to construct the counterfactuals and to measure average treatment effects of some policy interventions based on factor models.

A fundamental issue of factor models is the correct specification of the number of factors, r. When the number of factors is fixed, Bai and Ng (2002), Onatski (2009), Anh and Horenstein (2013), among others, have developed various approaches to consistently estimate the number of factors. But many empirical findings suggest that the number of factors may increase as the dimensions of the data N increases, or T increases. For many empirical analyses, the estimated number of factors ranges from one to more than ten, see Ludvigson and Ng (2009), Giannone, Reichlin and Sala (2005) and Forni and Gambetti (2010). This suggests that the number of factors may depend on sample sizes. One reason that the number of factors may increase with sample size is structural break, new factors may emerge after economic environments change. Using Bai and Ng's (2002) information criteria, Ludvigson and Ng (2007) find that the factor structure of their financial dataset comprising of 172 ($N = 172$) series quarterly financial indicators spanning the first quarter of 1960 through the fourth quarter of 2002 ($T = 172$) can be well described by 8 ($r = 8$) common factors. Jurado, Ludvigson and Ng (2013) update monthly version of the 147 financial time series used in Ludvigson and Ng (2007) and combine them with an updated version of 132 monthly macroeconomic series used in Ludvigson and Ng (2010). They find that 12 ($r = 12$) common factors can capture the variations of this new dataset with 279 series ($N = 279$) spanning the period 1959:01-2011:12 ($T = 636$). Hence, Ludvigson and Ng's (2013) finding supports the argument that the number of factors may increase as sample size increases.

Assuming that the number of factors r is fixed, there are many papers in the literature analyzing the problem of determining the number of factors. Some of them not only fix the number of factors, but also impose restrictions on the dimensions N and T, such as Lewbel (1991), Donald (1997), Cragg and Donald (1997), Connor and Korajczyk (1993), Forni and Reichlin (1998) and Stock and Watson (1998). Imposing no restriction on the relation between N and T except that both N and T are assumed to be large, Bai and Ng (2002) treat the determination of the number of factors as a model selection problem, they propose some criteria and show that the number of factors can be consistently estimated by minimizing the proposed criteria. Onatski (2009) develops a test of the null of k_0 factors against the alternative that the number of factors r satisfies $k_0 < r \leq k_1$ for some finite positive integer k_1. Onatski also describes the asymptotic distribution of the test statistic with critical values tabulated. Onatski (2010) suggests to determine the number of factors from empirical distribution of eigenvalues of sample covariance matrix. Ahn and Horenstein (2013) exploit the fact that the r largest eigenvalues of the variance matrix of N response variables grow unboundedly as N increases, while the other eigenvalues remain bounded to estimate the number of factors. The main difference between our model and the existing work is that we consider the problem of determining the number of factors in a factor model where the number of factors is allowed to increase as N or T increases.

Specifically, Section 2 is designed to provide an approach which enables one to estimate the number of factors consistently when the number of factors is allowed to increase as $N, T \to \infty$. We extend the method of Bai and Ng (2002) to penalize the number of factors with a penalty function which is determined by the sample sizes, N and T, as well as the maximum possible number of factors allowed in the estimation. As the factors are unobserved, the estimation procedure takes two steps. First, assuming the number of factors to be an arbitrary number $1 \leq k \leq k_{max}$, we estimate the factors ($\widehat{F}^k$) using the principal components method, where $k_{max} = k_{max,N,T}$ is the maximum number for possible number of factors, which is assumed to be greater or equal to the true number of factors, whose value is determined by N and T and it increases as N, T increases. Second, we select the number of factors $\hat{k}$ by minimizing a criterion modified from Bai and Ng (2002), which is a function of k and the estimated factors ($\widehat{F}^k$). This criterion depends on the usual trade-off between good fit and parsimony. We show that this method produces a consistent estimator of the number of factors r. However, simulation results show that the selected number of factor $\hat{k}$ can be sensitive to the choice of k_{max} and it tends to choose a $\hat{k}$ that is larger than r when k_{max} is large. We propose using a new ('mode' based) selection procedure to overcome this problem so that the selected $\hat{k}$ is not sensitive to different k_{max} values used in practice.

1.2 Structural Change in Factor Loadings

High dimensional factor models assume that a few number of common factors can represent variation among economic variables. This method of dimension deduction is a powerful statistical tool that has been found useful in forecasting (Stock and Watson, 2002), structural factor-augmented VAR analysis (Bernanke, Boivin, and Eliasz, 2005), reducing the number of instruments (Bai and Ng, 2010), and constructing dynamic stochastic general equilibrium (DSGE) models (Boivin and Giannoni, 2006). Despite the wide use of factor models, factor loadings are commonly estimated under the presupposition that loadings do not depend on time. In reality, however, economists often have to face instable parameters through structural changes or regime shifts, such as technology innovations, policy shift, oil price shock, financial crisis, and so on.

Although factor models are powerful, practitioners must be cautious about the potential structural changes in the high dimensional data sets. In theory, if breaks in the factor loadings are ignored, then the estimated number of factors will be inconsistent and likely overestimated when using any current methods, such as Bai and Ng (2002), Onatski (2009, 2010), and Ahn and Horenstein (2013). For example, in factor models where all factor loadings undergo a break (i.e., pure change case), the estimated number of factors doubles when the break is ignored. Because the pre-break and post-break factor loadings can be equivalently represented by stable factor loadings with extra pseudo factors. The incorrect number of factors causes the estimated factors and estimated factor loadings to be inconsistent. In practice, this concern about a structural break is empirically relevant, because parameter instability is a pervasive phenomenon in time series data (Stock and Watson, 1996). Banerjee and Marcellino (2008) and Yamamoto (2014) provide simulation and empirical evidence that the forecasts based on estimated factors are less accurate if the structural break in the factor loading matrix is ignored. Thus, correctly identifying the break point is an important issue in factor models with structural changes. In Section 3, we assume there is a common break in factor loadings for an approximate factor model. The objective of Section 3 is to estimate the change point and derive its limiting distribution.

When economists investigate the problem of structural changes in factor loadings, the first fundamental and important question is whether there is a break in factor loadings. The theory for structural changes in traditional time series and panel data models has been well developed. However, its application in a factor model is not straightforward because of unobservable latent factors and high dimensionality of the parameter space. Stock and Watson (2008) first considered structural changes in factor loadings. With a given number of factors, they applied the standard Chow statistics to test if the coefficients have a break in the regression of observed data on estimated factors for every cross-section unit i. Breitung and Eickmeier (2011) studied the theoretical properties of this approach and tested all post-break factor loadings equal to 0, based on Wald, LM and LR test statistics. Recently, Tanaka and Yamamoto (2015) formally proved Breitung and Eickmeier's (2011) test is powerful under specified conditions and proposed a modified version of the BE test. Among studies of joint testing the break in factor loadings, Chen, Dolado, and Gonzalo (2014) proposed a regression based test for the no change hypothesis. Han and Inoue (2015) developed a test by comparing the second moments of pre-break and post-break estimated factors. Whereas most papers focused on one-time abrupt structural break in the factor loadings, Su and Wang (2015) considered the case where the factor loadings change smoothly over time.

The rejection of the null hypothesis of no structural breaks naturally leads to the next question of when the break occurred. In comparison to the vast literature on testing structural changes, the corresponding literature for estimating the change point is quite small. Cheng, Liao and Schorfheide (2015) treated the detection of structural changes as a model selection issue. They proposed shrinkage estimation and showed that the number of factors and factor loadings can be consistently estimated in the presence of structural changes. Based on their estimation procedure, the true break fraction can be consistently estimated as a byproduct. Chen (2015) used a least squares estimator of break point and proved the consistency of break fraction. Baltagi, Kao and Wang (2015) proved that estimating the break point in factor loadings can be equivalently represented by estimating break point in the second moment of estimated factors. Massacci (2015) studied least squares estimations of structural changes in factor loadings, according to a threshold principle. The estimator of threshold value, which can be treated as break fraction, is superconsistent. Yet, none of these papers consider limiting distribution of the estimated change point.

In Section 3, we propose using the least squares method to estimate the unknown change point in factor loadings. For any given possible change point k, we calculate the pre-break sum of square residuals ($t = 1, \ldots, k$) and post-break sum of square residuals ($t = k + 1, \ldots, T$). Then we sum these two sums of square residuals. The number which can minimize the sum of two SSR is our estimated change point $\hat{k}$. Massacci (2015) proved that the estimated break fraction $\hat{\tau} = \frac{\hat{k}}{T}$ is superconsistent. His conclusion also indicates the consistency of the estimated change point $\hat{k}$. In practice, however, either because the magnitude of breaks is too small or because of the finiteness of N, we cannot expect $\hat{k}$ to coincide with k_0. A simple simulation shows that the probability of selecting the correct break point is quite low when break size is small. To solve this problem, we propose a new framework to derive the limiting distribution for the estimated change point. The limiting distribution of the estimated break point is more complex than that of the conventional panel data models, because both factors and factor loadings are unobservable. We show that the estimated factors and estimated factor loadings influence the limiting distribution. The random parts depend on *i.i.d.* normal variables and chi square variables. Based on the limiting distribution of the estimated break point, one can construct confidence intervals of the underlying true break point. In the simulation, we compare our least squares method to three competing approaches - including Baltagi, Kao and Wang (2015), Bai (2010), and MLE method. The simulation results show that our proposed method is the most efficient. Under the same data generating process and same combination of N and T, the reported result of the root mean square error by using least squares estimation has the smallest value. Because of the complexity of the limiting distribution, bootstrap method is studied to confirm our theory. Finally, we use an empirical application of the model to study structural changes in financial asset returns and in macroeconomics data.

The rest of this article is organized as follows. Section 2.1 sets up the model and presents the assumptions associated with the model. Section 2.2 presents the estimating procedures and the theoretical properties of the proposed estimators. Section 2.3 reports simulation experiments to examine the finite sample performances of our proposed method when r increases with N or T. All the proofs of Section 2 are given in the Appendix A.

Section 3.1 introduces the factor model with structural instability in factor loadings and describes the least squares method. Section 3.2 considers the limiting distribution of the estimated change point. Section 3.3 reports simulation results. Section 3.4 provides the empirical application. All the proofs of Section 3 are given in the Appendix B.

Section 4 concludes.

2. DETERMINING THE NUMBER OF FACTORS WHEN THE NUMBER OF FACTORS CAN INCREASE WITH SAMPLE SIZE

2.1 Factor Models

We consider the problem of determining the number of factors (r) in a static approximate factor model, allowing $r = r_{N,T} \to \infty$, as $N \to \infty$, or $T \to \infty$, or both $N, T \to \infty$, but with a slower rate than $\min\{N, T\}$, i.e., $\max\{r/N, r/T\} \to 0$, as $N, T \to \infty$.

Let X_{it} denote the response variable for unit i at time t, for $i = 1, \ldots, N$, and $t = 1, \ldots, T$. Our model is of the following form

$$X_{it} = \frac{1}{\sqrt{r}} \lambda_i^{0\prime} F_t^0 + e_{it}, \tag{2.1}$$

where F_t^0 is an $r \times 1$ vector of common factors, λ_i^0 is the $r \times 1$ vector of factor loadings, and e_{it} is the idiosyncratic error of the response variable X_{it}. The factors, factor loadings and idiosyncratic errors are not observed. Without loss of generality, we can assume that $E(X_{it}) = 0$. If this is not the case, we can de-mean the data first.

Note that at the right-hand-side of our model (2.1), we divide $\lambda_i^{0\prime} F_t^0$ by $\sqrt{r}$. This is because we allow for r to diverge when $N, T \to \infty$. If we do not divide $\lambda_i^{0\prime} F_t^0$ by $\sqrt{r}$, then the variance of the systematic part, $\lambda_i^{0\prime} F_t^0$, is proportional to r and the variance of idiosyncratic error e_{it} is finite , the variance of noise part over the variance of information part will go to zero, or equivalently, the goodness-of-fit R^2 will converge to one. By dividing $\lambda_i^{0\prime} F_t^0$ by $\sqrt{r}$, we have $Var(r^{-1/2} \lambda_i^{0\prime} F_t^0) = O(1)$ and we can obtain a reasonable goodness-of-fit that is not too close to one.

Let tr(A) denote the trace of a square matrix A. The norm of a matrix A is defined as $\|A\| = [\text{tr}(A'A)]^{1/2}$. We use M_1 to denote a generic positive constant and use $\mathcal{N}$ to denote the set of natural number. We make the main assumptions as follows:

Assumption A (Factors and loadings):

1. For all t, $r^{-2}E\|F_t^0\|^4 < M_1$;

2. There exists a $r \times r$ positive definite matrix Σ_F such that $\|T^{-1}\sum_{t=1}^T F_t^0 F_t^{0\prime} - \Sigma_F\| \xrightarrow{p} 0$ as $T \to \infty$;

3. $\max_{1\leq i\leq N} r^{-2}E\|\lambda_i^0\|^4 \leq M_1 < \infty$;

4. Let Λ^0 be the $N \times r$ factor loading matrix with its i^{th} row given by λ_i^0. Then there exists a $r \times r$ positive definite matrix D such that $\|N^{-1}\Lambda^{0\prime}\Lambda^0 - D\| \xrightarrow{p} 0$ as $N \to \infty$;

5. Let λ_{il}^0 and F_{tl}^0 be the l^{th} components $(l = 1, ..., r)$ of λ_i^0 and F_t^0, respectively. Then for all (i, t),

   $E\{[r^{-1/2}\sum_{l=1}^r E(\lambda_{il}^0 F_{tl}^0)]^4\} \leq M_1$.

Assumption B (Idiosyncratic Components):

1. For all i and t, $E(e_{it}) = 0, E|e_{it}|^8 \leq M_1$;

2. $E(N^{-1}e_s'e_t) = E(N^{-1}\sum_{i=1}^N e_{is}e_{it}) = \gamma_N(s,t)$, $|\gamma_N(s,s)| \leq M_1$ for all s, and that $T^{-1}\sum_{s=1}^T\sum_{t=1}^T |\gamma_N(s,t)| \leq M_1$;

3. $E(e_{it}e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t; furthermore, $N^{-1}\sum_{i=1}^N\sum_{j=1}^N |\tau_{ij}| \leq M_1$;

4. $E(e_{it}e_{js}) = \tau_{ij,ts}$ and $(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|\tau_{ij,ts}| \leq M_1$;

5. for every (t,s), $E|N^{-1/2}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]|^4 \leq M_1$;

6. We assume that there exist a $T \times T$ matrix L, a $N \times N$ matrix R, and a $T \times N$ matrix ε such that

$$e = L\varepsilon R$$

where L $(T \times T)$ and R $(N \times N)$ are arbitrary non-random positive definite matrices, and $\varepsilon = (\varepsilon_{ti})$ is a $T \times N$ matrix consisting of independent elements with uniformly bounded 7^{th} moment and $E(\varepsilon_{it}) = 0$.

Assumption C

1. Weak Dependence Between Factors and Idiosyncratic Components:

$$E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{Tr}}\sum_{t=1}^{T}F_t^0 e_{it}\right\|^2\right) \leq M_1;$$

2. Weak Dependence Between Factor Loadings and Idiosyncratic Components:

$$E\left(\frac{1}{T}\sum_{t=1}^{N}\left\|\frac{1}{\sqrt{Nr}}\sum_{i=1}^{N}e_{it}\lambda_i^0\right\|^2\right) \leq M_1.$$

Conditions in Assumption A are modified from Assumptions A-B in Bai and Ng (2002) by taking care of the fact that $r \to \infty$ as $N, T \to \infty$. It is easy to see that A1 holds true if $E[(F_{tl}^0)^4] = O(1)$ for all $t = 1, ..., T$ and for all $l = 1, ..., r$. A2 imposes a restriction on the rate of r. For example with $\Sigma_F = E(F_t^0 F_t^{0'})$, it can be easily shown that A2 holds true if $r = o(T^{-1/2})$ and $T^{-1}\sum_{t=1}^{T}\sum_{s\neq t}^{T} Cov(F_{tl}^0 F_{tm}^0, F_{sl}^0 F_{sm}^0) = O(1)$ for all l, $m \in \{1, ..., r\}$. Similarly, A3 holds true if $E[(\lambda_{il}^0)^4] = O(1)$ for all

$i = 1, ..., N$ and for all $l = 1, ..., r$. A4 is similar to A2, it holds true if $r = o(N^{1/2})$ and $N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{T} Cov(\lambda_{il}^0\lambda_{jm}^0, \lambda_{jl}^0\lambda_{jm}^0) = O(1)$ for all l, $m \in \{1, ..., r\}$. A5 requires that $\lambda_{il}^0 F_{tl}^0$ is a weakly dependent process in l because we allow for $r \to \infty$.

Conditions in Assumption B are basically the same as Assumption C in Bai and Ng (2002) because the idiosyncratic error e_{it} is unrelated to r whether r is finite or is allowed to diverge to infinity with the sample size. In particular, B5 is similar to A5 in that it assumes that, for all (t, s), $e_{it}e_{is}$ is a weakly dependent process in i. Assumption B6 puts a structure on the idiosyncratic components. This structure allows heteroscedasticity in both the time and cross-section dimensions, and also limited autocorrelation and cross-sectional correlation in the components.

Finally, assumption C is similar to assumption D in Bai and Ng (2002) except that we modified it by dividing the quantity by $\sqrt{r}$ as r is allowed to diverge as N and T tend to infinity. They allow for limited time-series and cross-section dependence in idiosyncratic component and also weak dependence between factors (factor loadings) and idiosyncratic errors.

2.2 Estimating the Common Factors and the Number of Factors

Following Bai and Ng (2002), we estimate the common factor in a large panel by the principal components method. For $k \in \{1, \ldots, k_{max}\}$, where k_{max} is allowed to increase at a slower speed than $\min\{N, T\}$ such that $k_{max} = o(\min\{N^{1/3}, T\})$. Let λ_i^k and F_t^k denote $k \times 1$ vectors of the loadings and factors with the allowance of k factors in the estimation. The method of principal components minimizes

$$V(k) = \min_{\Lambda^k, F^k} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \frac{1}{\sqrt{k}}\lambda_i^{k\prime} F_t^k)^2 \tag{2.2}$$

over $1 \leq k \leq k_{max}$, subject to the normalization of either $\Lambda^{k\prime}\Lambda^k/N = I_k$ or $F^{k\prime}F^k/T = I_k$, where Λ^k and F^k are the $N \times k$ and $T \times k$ factor loading and factor matrices, respectively.

Let $\text{ev}_{(i)}(A)$ denote the i^{th} largest eigenvalue of matrix A, and $\text{EV}_{(i)}(A)$ is the eigenvector corresponding to the eigenvalue $\text{ev}_{(\text{i})}(A)$ of the matrix A. If one concentrates out Λ^k and uses the normalization that $F^{k\prime}F^k/T = I_k$. The estimated factor matrix is $\tilde{F}^k = \sqrt{T}(\text{EV}_{(1)}(XX)', \ldots, \text{EV}_{(k)}(XX'))$.

Given $\tilde{F}^k$, $\tilde{\Lambda}^{k\prime} = \sqrt{k}(\tilde{F}^{k\prime}\tilde{F}^k)^{-1}\tilde{F}^{k\prime}X = \sqrt{k}\tilde{F}^{k\prime}X/T$ is the corresponding matrix of factor loadings. On the other hand, if one concentrates out F^k and uses the normalization that $\Lambda^{k\prime}\Lambda^k/N = I_k$, the solution to the above problem is given by $(\bar{F}^k, \bar{\Lambda}^k)$, where $\bar{\Lambda}^k = \sqrt{N}(\text{EV}_{(1)}(X'X), \ldots, \text{EV}_{(k)}(X'X))$. The normalization that $\Lambda^{k\prime}\Lambda^k/N = I_k$ implies $\bar{F}^k = \sqrt{k}X\bar{\Lambda}^k/N$.

Define $\widehat{F}^k = \bar{F}^k(\bar{F}^{k\prime}\bar{F}^k/T)^{1/2}$, a rescaled estimator of the factors. This rescaled estimator has the asymptotic properties summarized in the following theorem.

Proposition 2.2.1 *Under the assumptions A - C,*

for any $1 \leq k \leq k_{\max} = o(\min\{N^{1/3}, T\})$ *there exists a* $(r \times k)$ *matrix* H^k *with* $\text{rank} = \min\{k, r\}$ *such that*

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t^k - H^{k\prime}F_t^0\right\|^2 = O_p\left(\max\left\{\frac{k^3 r}{N}, \frac{k^3}{T}\right\}\right). \tag{2.3}$$

Similar to the results of Bai and Ng (2002), Proposition 2.2.1 suggests that the time average of the squared deviations between the estimated factors $\widehat{F}^k$ and those that lie in the true factor space, $H^{k\prime}F_t^0$, will vanish as $N, T \to \infty$. However, the convergence rate depends on not only the panel structure N and T, but also the factor structure r and k.

Given the results of Proposition 2.2.1, we can now analyze the problem of determining the number of factors. Let $V(k, F^k) = \min_\Lambda \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T (X_{it} - \frac{1}{\sqrt{k}}\lambda_i^{k\prime} F_t^k)^2$ be the sum of squared residuals (divided by NT), where the residuals are from regression models of regressing X_i on the k factors for all $i = 1, \ldots, N$, and $X_i = (X_{i1}, X_{i2}, \ldots, X_{iT})'$ is a $T \times 1$ vector of time-series observations for the ith cross-section unit. The selecting criterion modified from those suggested by Bai and Ng (2002) has the form

$$PC(k) = V(k, \widehat{F}^k) + kg(N,T), \tag{2.4}$$

where $g(N,T)$ is the penalty factor satisfying two conditions: (i) $k_{max} \cdot g(N,T) \to 0$ as $N,T \to \infty$, (ii) $C_{N,T,k_{max}}^{-1} g(N,T) \to \infty$ as $N,T \to \infty$, where $C_{N,T,k_{max}} = O_p\left(\max\left\{\frac{k_{max}^3}{\sqrt{N}}, \frac{k_{max}^{5/2}}{\sqrt{T}}\right\}\right)$. As $V(k, \widehat{F}^k)$ is decreasing in k, the criterion above penalizes k with a penalty factor $kg(N,T)$ to select the estimator $\hat{k}$ such that asymptotically under and overparameterized models will not be chosen. Theorem 2.2.1 formally establishes this result.

Theorem 2.2.1 *Let* $1 \le r \le k_{max} = o(\min\{N^{1/17}, T^{1/16}\})$ *and* $\hat{k} = \operatorname{argmin}_{1 \le k \le k_{max}} PC(k)$. *Suppose that Assumptions A-C hold, and that (i)* $k_{max} \cdot g(N,T) \to 0$, *(ii)* $C_{N,T,k_{max}}^{-1} \cdot g(N,T) \to \infty$ *as* $N,T \to \infty$. *Then*

$$\lim_{N,T \to \infty} \text{Prob}[\hat{k} = r] = 1. \tag{2.5}$$

A formal proof of Theorem 2.2.1 is provided in the Appendix A. Conditions (i) and (ii) together define the type of penalty factor that should vanish at an appropriate rate. They are sufficient conditions for estimation consistency so that they may not always be required for consistent estimating the number of factors.

Remark 2.2.1 *Since we often need to divide some quantities by r, we rule out the case that $r = 0$. Allowing for $r = 0$ in our framework will complicate the regularity conditions, notations and proofs. Therefore, we did not consider the case that $r = 0$ in our paper. The $r = 0$ case is covered in Bai and Ng's (2002). Their procedure can be used to select the number of factors even when the true number of factors is* 0. *We also conducted some simulations which show that both Bai and Ng's (2002) original method and the modified method proposed in our paper work well when $r = 0$.*

Note that the condition imposed in k_{max} is asymmetric in (N, T). This result is induced by Proposition 2.2.1. The details can be found in the proof of Proposition 2.2.1 given in the appendix. As a referee correctly points out, if in the proof of Theorem 2.2.1, instead of using the result of Proposition 2.2.1 that $T^{-1}\sum_{t=1}^{T}\left\|\hat{F}_t^k - H^{k'}F_t^0\right\|^2 = O_p\left(\max\left\{\frac{k^3 r}{N}, \frac{k^3}{T}\right\}\right)$, one may use $N^{-1}\sum_{i=1}^{N}\left\|\hat{\lambda}_i^k - \tilde{H}^{k'}\lambda_i^0\right\|^2 = O_p\left(\max\left\{\frac{k^3}{N}, \frac{k^3 r}{T}\right\}\right)$,[1] where $\tilde{H}^k$ is a $r \times k$ matrix with $rank(\tilde{H}^k) = \min\{r, k\}$. Then the condition that $1 \le r \le k_{max} = o(\min\{N^{1/17}, T^{1/16}\})$ in Theorem 2.2.1 will be replaced by $1 \le r \le k_{max} = o(\min\{N^{1/16}, T^{1/17}\})$. The result is still asymmetric in N and T, but the roles of N and T are exchanged.

In fact, it is possible to obtain a symmetric result (of k_{max} in N and T) under some stronger regularity conditions, i.e., one can obtain a symmetric condition for k_{max} as $1 \le r \le k_{max} = o(\min\{N^{1/16}, T^{1/16}\})$ in Theorem 2.2.1 under some stronger assumptions. We state this result in the following proposition.

[1]This result can be proved similar to the proof of Proposition 2.2.1, its proof is available from the authors upon request.

Proposition 2.2.2 *Under the same conditions as in Proposition 2.2.1 except that we strength some conditions as follows: (i) λ_{il} is non-random with $\lambda_{il} \leq \bar{\lambda} < \infty$ for all $i = 1, ..., N$ and $l = 1, ..., r$; (ii) $E(e_{it}e_{jt}) = 0$ for all $t \in \{1, ..., T\}$ and for all $j \neq i$, $i, j \in \{1, ..., N\}$, $E(F^0_{tl}F^0_{tm}) = 0$ for all $t \in \{1, ..., T\}$ and for all $m \neq l$, $l, m \in \{1, ..., r\}$; (iii) e_{it} and F^0_s are independent with each other for all i, t and s. Then*

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t^k - H^{k'}F_t^0\right\|^2 = O_p\left(k^3\left(\frac{1}{N}+\frac{1}{T}\right)\right). \tag{2.6}$$

The proof of Proposition 2.2.2 is given in the appendix. Under Proposition 2.2.2, the condition $1 \leq r \leq k_{max} = o(\min\{N^{1/17}, T^{1/16}\})$ in Theorem 2.2.1 can be replaced by $1 \leq r \leq k_{max} = o(\min\{N^{1/16}, T^{1/16}\})$. That is, we obtain a condition on k_{max} that is symmetric in N and T.

Remark 2.2.2 *The zero correlation assumption on F_{tl} used in Proposition 2.2.2 is quite strong. However, it can be replaced by some weakly dependence assumptions such as ρ-mixing or β-mixing processes with mixing coefficients decay to zero at certain rates. But this will make the presentation (regarding regularity conditions) as well as the proofs of Proposition 2.2.2 much longer. Therefore, we will not pursue a proof of Proposition 2.2.2 under weak regularity conditions in this paper.*

Similar to Bai and Ng (2002) we have the following Corollary.

Corollary 2.2.1 *Under the Assumptions of Theorem 2.2.1, if one replaces $PC(k)$ in Theorem 2.2.1 by the class of criterion defined by*

$$IC(k) = \ln\left(V(k, \widehat{F}^k)\right) + kg(N, T),$$

then the conclusion of Theorem 2.2.1 holds true.

Corollary 2.2.1 states that the class of criterion $IC(k)$ can also be used to consistently estimate the number of factors in factor models where the number of factors possibly increases with the sample size.

Let $\widehat{\sigma}^2$ be a consistent estimate of $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}E(e_{it})^2$. Bai and Ng (2002) generalize the C_p criterion of Mallows (1973) and suggest three PC_p criteria as follows:

$$\begin{aligned} PC_{p1}(k) &= V(k,\widehat{F}^k) + k\cdot\widehat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right), \\ PC_{p2}(k) &= V(k,\widehat{F}^k) + k\cdot\widehat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln(\min\{N,T\}), \\ PC_{p3}(k) &= V(k,\widehat{F}^k) + k\cdot\widehat{\sigma}^2\left(\frac{\ln(\min\{N,T\})}{\min\{N,T\}}\right). \end{aligned} \tag{2.7}$$

It is easy to check that these criteria satisfy the two conditions for the penalty factor in Theorem 2.2.1 if $k_{max} = o\left(\left[\ln\left(\frac{NT}{N+T}\right)\right]^{1/6}\right)$. The three criteria have different finite-sample properties while they are asymptotically equivalent. In applications, Bai and Ng (2002) suggest to replace $\hat{\sigma}^2$ with $V(k_{max},\widehat{F}^{k_{max}}) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2$, where $\hat{e}_{it} = X_{it} - \frac{1}{\sqrt{k}}\widehat{\lambda}_i^{k'_{max}}\widehat{F}_t^{k_{max}}$ for $i = 1,\ldots,N$ and $t = 1,\ldots,T$, the residuals for the linear regression of X on $\widehat{F}^{k_{max}}$. Thus, the number of factors estimated using these three criteria may be sensitive to the selection of k_{max}. We will propose a method that avoids the sensitivity of selected $\hat{k}$ depending on k_{max}.

Corollary 2.2.1 suggests the following three IC_p criteria can also be used to select the number of factors:

$$\begin{aligned} IC_{p1}(k) &= \ln\left(V(k,\widehat{F}^k)\right) + k \cdot \left(\frac{N+T}{NT}\right) \ln\left(\frac{NT}{N+T}\right), \\ IC_{p2}(k) &= \ln\left(V(k,\widehat{F}^k)\right) + k \cdot \left(\frac{N+T}{NT}\right) \ln(\min\{N,T\}), \\ IC_{p3}(k) &= \ln\left(V(k,\widehat{F}^k)\right) + k \cdot \left(\frac{\ln(\min\{N,T\})}{\min\{N,T\}}\right). \end{aligned} \tag{2.8}$$

The main advantage of the three criteria given in (2.8) is that the scaling factor $\hat{\sigma}^2$ is automatically removed by the logarithmic transformation. We do not need to estimate σ^2 before selecting the number of factors. Therefore, the number of factors estimated using IC_p criteria is insensitive to the selection of k_{max}.

As the estimated $\hat{k}$ using PC_p criteria may be sensitive to k_{max}, the selection of k_{max} is an important issue in practice. Bai and Ng (2002) suggest to select k_{max} by setting $k_{max} = 8[(\min\{N,T\}/100)^{1/4}]$ where $[A]$ denotes the integer part of a real number A. But their theoretical result does not cover this case as this k_{max} increases (without bound) with N and T. Using some ad-hoc rules to select k_{max} may lead to $k_{max} < r$, which will lead to an underestimation of the number of factors because if $k_{max} < r$, then we will have $\hat{k} \leq k_{max} < r$. On the other hand, if k_{max} is too large ($k_{max} >> r$), simulations show that the selected $\hat{k}$ tends to overestimate r ($\hat{k} > r$). We propose a new procedure to resolve this problem. We propose to let k_{max} take a wide range of values. For each value of k_{max}, we select a $\hat{k}_{k_{max}}$ that minimizes the PC_p criteria. We then select the value of $\hat{k}$ that appears most times among the different $\hat{k}_{k_{max}}$ values, i.e., we select the mode of $\hat{k}_{k_{max}}$ (over a wide range of k_{max}). We use a specific example to illustrate this selection procedure. We generate a simulated data of $N = 200, T = 60$ with the true number of factors $r = 7$. We let k_{max} take values from $\{1, 2, ..., 40\}$. For each different $1 \leq k_{max} \leq 40$, we select a $\hat{k}_{k_{max}}$ by minimizing

PC_{p1} criterion. The result is presented in Figure A.1. From Figure A.1 we observe that when $k_{max} < r = 7$, we select $\hat{k} = k_{max} < 7$ as expected; when $7 \leq k_{max} \leq 16$, we select $\hat{k} = 7$; when $k_{max} > 16$, the selected $\hat{k} > 7$. Moreover, $\hat{k}$ increases with k_{max}. We also notice that $\hat{k} = 7$ is selected ten times (when $k_{max} = 7, 8, ..., 16$), while all the other values are chosen no more than three times. For example, when $17 \leq k_{max} \leq 19$, the selected $\hat{k}_{k_{max}} = 8$, i.e., $\hat{k}_{k_{max}} = 8$ is selected three times. According to our selection rule, $\hat{k} = 7$ is selected because $\hat{k} = 7$ appears most times (10 times).

Figure A.2 plot $\hat{k}$-k_{max} curves for different N, T and r values. We see that although $\hat{k}$ increases with k_{max} for most cases, our proposed procedure can select the correct number of factors because $\hat{k}_{k_{max}}$ takes value r more often than taking any other values for all cases reported in Figure A.2. Hence, our proposed procedure of selecting $\hat{k}$ is not sensitive to k_{max} provided that one let k_{max} take a wide range of values. Therefore, we suggest letting k_{max} to take values in $\{1, 2, ..., [6 \log(max(N, T))]\}$ where $[A]$ denotes the integer part of a real number A. $[6 \log(max(N, T))]$ is around 41, 45 and 55 when $max(N, T) = 1000$, 2000 and 10000. This setting is also consistent with our simulation since we let $r = [1.5 log(max(N, T))]$ in our simulations in section 1.4.

2.3 Simulations

In this section we conduct Monte Carlo simulations to investigate how our modified criteria of Bai and Ng (2002) perform when the number of factors is allowed to increase with N or T. For simplicity of the comparison with the simulation results in Bai and Ng (2002), we first fix T and allow N and r to increase. When T is fixed as 60, we let $N = 100, 200, 500, 1000, 2000$ and $r = [1.5 \log(N)]$, where $[A]$ denotes the integer part of a real number A; for $T = 100$, we let $N = 40, 60, 100, 200, 500, 1000, 2000$

and $r = [1.5 \log(N)]$. The simulation results for this case are reported in the upper part of each table for each data generating process (DGP). Next, we check the performance of the criteria when N is fixed and T keeps increasing. When $N = 100$, we let $T = 40$, 60,100, 200, 500, 1000, 2000 and $r = [1.5 \log(T)]$; when $N = 60$, we let $T = 100, 200, 500, 1000, 2000$ and $r = [1.5 \log(T)]$. The simulation results for this case are reported in the lower part of each table for each DGP. We replicate the suggested estimating procedure 1000 times and the reported results are the averages of $\hat{k}$ over 1000 replications.

The data generating processes (DGP) have the following form:

$$X_{it} = \frac{1}{\sqrt{r}} \sum_{j=1}^{r} \lambda_{ij} F_{tj} + e_{it},$$

where $\lambda_{ij} \sim i.i.d.\ N(0,1)$, $F_{tj} \sim i.i.d.\ N(0,2)$.

We consider three DGPs here. In the base case, we set the DGP as $e_{it} \sim i.i.d. N(0,1)$. This base DGP is denoted as DGP1. The simulation results for this case are reported in Table 2.1. We see that all information criterion give precise estimates of the number of factors.

For the heterogeneity case of DGP2, we set the idiosyncratic shocks to be heterogeneous. We let $e_{it} = u_{it} + \delta_t \epsilon_{it}$ where $u_{it} \sim i.i.d. N(0,1)$, $\epsilon_{it} \sim i.i.d. N(0,1)$, and $\delta_t = 0$ for even t, $\delta_t = 1$ for odd t. Thus the variance of the idiosyncratic shocks is 1 when t is odd and 2 when t is even. We denote this DGP as DGP2. The estimated values of $\hat{k}$ are reported in Table 2.2 where the boldfaced numbers indicate incorrect selection of the number of factors. Similar to the homogeneous cases, PC_{p1}, PC_{p2}, and PC_{p3} perform well under all kinds of combinations of N and T. The other three criteria IC_{P1} and IC_{p2}, and IC_{p3} also perform well in general, although occasionally they may select $\hat{k}$ that is slightly smaller than the true number of factors r when sample size is small.

For the last case, denoted as DGP3, we allow the idiosyncratic to be autocorrelated. We set $e_{it} = \rho e_{it-1} + v_{it}$, where $\rho = 0.5$ and $v_{it} \sim i.i.d.N(0,1)$. The estimation results are reported in Table 2.3. The results for this case are almost the same as those of the base case except for $(N,T) = (60,200)$ with $r = 7$, IC_{p_1} and IC_{p_2} select $r = 6$. All other four information criteria perform quite well in accurately estimating the number of factors for all (N,T) combinations for DGP3.

Summarizing the results for all the DGPs we observe that PC_{p1}, PC_{p2}, and PC_{p3} have the best overall performance. IC_p1, IC_p2, and IC_{p3} perform well when the sample size is large (min$\{N,T\}$>100).

Table 2.1: Estimated Number of Factors: DGP1

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6	6	6	6	6	6
200	60	7	7	7	7	7	7	7
500	60	9	9	9	9	9	9	9
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	11	11	11
40	100	5	5	5	5	5	5	5
60	100	6	6	6	6	6	6	6
100	100	6	6	6	6	6	6	6
200	100	7	7	7	7	7	7	7
500	100	9	9	9	9	9	9	9
1000	100	10	10	10	10	10	10	10
2000	100	11	11	11	11	11	11	11
100	40	5	5	5	5	5	5	5
100	60	6	6	6	6	6	6	6
100	100	6	6	6	6	6	6	6
100	200	7	7	7	7	7	7	7
100	500	9	9	9	9	9	9	9
100	1000	10	10	10	10	10	10	10
100	2000	11	11	11	11	11	11	11
60	100	6	6	6	6	6	6	6
60	200	7	7	7	7	7	7	7
60	500	9	9	9	9	9	9	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	11	11	11

DGP1: $X_{it} = \frac{1}{\sqrt{r}} \sum_{j=1}^{r} \lambda_{ij} F_{tj} + e_{it}; r = [c * \ln(N)]$ *for the upper part of the table, and* $r = [c * \ln(T)]$ *for the lower part, where c=1.5, and* $[A]$ *denotes the integer part of a real number* A.

Table 2.2: Estimated Number of Factors: Heterogeneity

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6	6	6	**5**	**5**	6
200	60	7	7	7	7	7	7	7
500	60	9	9	9	9	**8**	**8**	**8**
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	11	11	11
40	100	5	5	5	5	5	5	5
60	100	6	6	6	6	6	6	6
100	100	6	6	6	6	6	6	6
200	100	7	7	7	7	7	7	7
500	100	9	9	9	9	9	9	9
1000	100	10	10	10	10	10	10	10
2000	100	11	11	11	11	11	11	11
100	40	5	5	5	5	5	5	5
100	60	6	6	6	6	**5**	**5**	6
100	100	6	6	6	6	6	6	6
100	200	7	7	7	7	7	7	7
100	500	9	9	9	9	9	9	9
100	1000	10	10	10	10	10	10	10
100	2000	11	11	11	11	11	11	11
60	100	6	6	6	6	6	6	6
60	200	7	7	7	7	7	7	7
60	500	9	9	9	9	9	9	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	11	11	11

DGP2: $X_{it} = \frac{1}{\sqrt{r}} \sum_{j=1}^{r} \lambda_{ij} F_{tj} + e_{it}; e_{it} = u_{it} + \delta_t \epsilon_{it}$, *where* $\delta_t = 0$ *for* t *even, and* $\delta_t = 1$ *for* t *odd;* $r = [c \ln(N)]$ *for the upper part of the table, and* $r = [c \ln(T)]$ *for the lower part, where* $[A]$ *denotes taking the integer part of a real number.*

Table 2.3: Estimated Number of Factors: Autocorrelation

N	T	r	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
100	60	6	6	6	6	6	6	6
200	60	7	7	7	7	7	7	7
500	60	9	9	9	9	9	9	9
1000	60	10	10	10	10	10	10	10
2000	60	11	11	11	11	11	11	11
40	100	5	5	5	5	5	5	5
60	100	6	6	6	6	6	6	6
100	100	6	6	6	6	6	6	6
200	100	7	7	7	7	7	7	7
500	100	9	9	9	9	9	9	9
1000	100	10	10	10	10	10	10	10
2000	100	11	11	11	11	11	11	11
100	40	5	5	5	5	5	5	5
100	60	6	6	6	6	6	6	6
100	100	6	6	6	6	6	6	6
100	200	7	7	7	7	7	7	7
100	500	9	9	9	9	9	9	9
100	1000	10	10	10	10	10	10	10
100	2000	11	11	11	11	11	11	11
60	100	6	6	6	6	6	6	6
60	200	7	7	7	7	**6**	**6**	7
60	500	9	9	9	9	9	9	9
60	1000	10	10	10	10	10	10	10
60	2000	11	11	11	11	11	11	11

DGP3: $X_{it} = \frac{1}{\sqrt{r}} \sum_{j=1}^{r} \lambda_{ij} F_{tj} + e_{it}; e_{it} = \rho e_{it-1} + v_{it}$; $\rho = 0.5$; $r = [c \ln (N)]$ *for the upper part of the table, and* $r = [c \ln (T)]$ *for the lower part, where* $[A]$ *denotes taking the integer part of a real number.*

3. STRUCTURAL CHANGE IN HIGH DIMENSIONAL FACTOR MODELS

3.1 The Model and Estimation

We consider the model

$$x_{it} = \begin{cases} \lambda_{i1}' f_t + e_{it} & \text{for t} = 1, 2, \ldots, k_0 \\ \lambda_{i2}' f_t + e_{it} & \text{for t} = k_0 + 1, k_0 + 2, \ldots, T \end{cases} \tag{3.1}$$

$$i = 1, 2, \ldots, N.$$

where x_{it} is the observed data, f_t is an $r \times 1$ vector of unobserved common factors, and e_{it} is the idiosyncratic error for variable i at time t. In this model, each series of factor loadings is subject to structural changes at the true break point k_0, where k_0 is unknown. λ_{i1} is the pre-break factor loading, and λ_{i2} is the post-break factor loading. Both of them are $r \times 1$ vectors. $\lambda_{i1}' = [\lambda_{i1,1}, \ldots, \lambda_{i1,r}]$, $\lambda_{i2}' = [\lambda_{i2,1}, \ldots, \lambda_{i2,r}]$. In matrix form, $\Lambda_1 = [\lambda_{i1}, \ldots, \lambda_{N1}]', \Lambda_2 = [\lambda_{i2}, \ldots, \lambda_{N2}]$, and $F = [f_1, \ldots, f_T]'$. The difference between λ_{i1} and λ_{i2} represents the magnitude of break (break size). We allow for changes in the number of factors, which can be disappearing or emerging factors. For example, some factor loadings are allowed to equal to zero. They can change from zero to nonzero after the break point. These new nonzero factor loadings form the emerging factors.

The study of structural changes in factor loadings is motivated by both theoretical and empirical research. In a standard factor model, $x_{it} = \lambda_i f_t + e_{it}$, the factors are commonly estimated by principal component methods (Bai and Ng, 2002). If the number of factors r is given, then the estimated factors $\hat{F}$ equal to $\sqrt{T}$ times the eigenvectors associated with the r largest eigenvalues of matrix XX', where X is

$T \times N$ data matrix. Given $\hat{F}$, the factor loadings can be estimated by OLS: $\hat{\Lambda} = \frac{X'\hat{F}}{T}$. The pre-break and post-break factors and factor loadings in equation (3.1) are also estimated by principal component methods. Unlike conventional time series models or panel data models with observed regressors, the factors and factor loadings are estimated rather than observed. When the factor loadings are subject to structural changes, the estimation of factors is also affected, depending upon the specifications of the change.

Ludvigson and Lettau (2001) employed an empirical test of CAPM in which the discount factor is approximated as a linear function of the model's fundamental factors. Instead of assuming constant parameters over time, they used a linear factor model with time-varying coefficients. The equation can be written as $\lambda_{it} = a_i + b_i z_t$. We have $\lambda_{it} f_t = a_i f_t + b_i z_t f_t$, and $z_t f_t$ forms the new factors. The number of factors changes because of the unstable factor loadings. Economic events can also destabilize factor loadings. Cheng, Liao and Schorfheide (2015) used Stock and Watson's (2012) data set and showed strong evidence that the factor loadings in the normalized factor model changed because of the 2007-2009 recession, generally implying a stronger co-movement of the series after 2007. More examples may also be given: an oil price shock can influence the coefficients of different countries corresponding to their output, a policy shift of China's exchange rate can affect the investors' strategy, a financial crisis can impact companies' asset returns, and so on. The fundamental issue is to correctly find the break point.

We use the following method to estimate the change point:

$$SSR(k) = \sum_{t=1}^{k} \sum_{i=1}^{N} \left(x_{it} - \hat{\lambda}_{i1}^{(k)\prime} \hat{f}_t^{(k)} \right)^2 + \sum_{t=k+1}^{T} \sum_{i=1}^{N} \left(x_{it} - \hat{\lambda}_{i2}^{(k)\prime} \hat{f}_t^{(T-k)} \right)^2. \qquad (3.2)$$

The least squares estimator for k_0 in the model is defined as

$$\hat{k} = \arg \min_{r \leq k \leq T-r} SSR(k), \tag{3.3}$$

where $\hat{\lambda}_{i1}^{(k)}$ and $\hat{f}_t^{(k)}$ are estimated factor loadings and estimated factors based on the pre-break sample $X^{(k)}$, $X^{(k)} = [X_1, X_2, \ldots, X_k]'$ is of dimension $k \times N$. k is the possible change point. We restrict k to be in $[r, T-r]$ to avoid the singular matrix in subsequent estimation of pre-break and post-break factors and factor loadings. This does not significantly influence the distribution of the estimated break point. $\hat{\lambda}_{i2}^{(k)}$ and $\hat{f}_t^{(T-k)}$ are estimated factor loadings and estimated factors based on the post-break sample $X^{(T-k)}$, $X^{(T-k)} = [X_{k+1}, X_{k+2}, \ldots, X_T]'$ is of dimension $(T-k) \times N$.

When the number of factors is unknown, Chen (2015) proved that the break point can be consistently estimated by using $\tilde{r} - 1$ number of factors, where $\tilde{r}$ is the estimated number of factors ignoring structural breaks. Baltagi et al. (2015) and Massacci (2015) also modified Bai and Ng (2002)'s model selecting method to estimate the number of factors. Our objective is to estimate the break point and derive the limiting distribution of it, so we simply use Chen (2015)'s approach to specify the number of factors.

3.2 Limiting Distribution

Chen (2015) proved the break fraction consistency that $\hat{\tau} - \tau_0 = O_p(max\{\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{T}}\})$, where $\hat{\tau} = \frac{k}{T}, \tau_0 = \frac{k_0}{T}$. Baltagi et al. (2015) showed that the change point of factor loadings is the same as the change point of the second moments of estimated factors. Although the estimated change point is inconsistent ($O_p(1)$), their result remains stronger than Chen (2015). Massacci (2015) proved the consistency of the estimated change point that $\hat{k} - k_0 = O_p(\frac{1}{N^{2\alpha^0 - 1}})$, where $0.5 < \alpha^0 \leq 1$. He required that at least a fraction of $O(N^{\alpha^0})$ series undergo a break. Massacci's (2015) $O_p(\frac{1}{N^{2\alpha^0-1}})$ result is a very strong convergence speed. We followed this conclusion that $\hat{k} - k_0 = o_p(1)$ in our paper. In section 3.3, of the simulation part, we show that the estimated change point equals the true change point in a large sample size. In practice, however, either because of too small magnitudes of breaks or because of the finiteness of N, we cannot expect $\hat{k}$ to coincide with k_0. Although Massacci (2015) gave the restriction on the number of series that were subject to structural changes, he still followed a potential presupposition that

$$\lim_{N \to \infty} \sum_{i=1}^{N} \| \lambda_{i1} - \lambda_{i2} \|^2 = \infty.$$

When the break size is small, the estimated break point may not be reliable in finite samples. A simple simulation can confirm our conjecture. Let factors and errors be *i.i.d.* standard normal variables. The true break point k_0 is $T/2$. The true number of factors is assumed as 1 and 2. The break size is $\lambda_{i1} - \lambda_{i2} = 0.3 * N(0, 1)$. The number of replications is 1,000.

The reported results in Table 3.1 are root mean square errors. The number in parentheses is the percentage of obtaining the correct change point. As shown, the RMSE is large and the probability of finding true break point is very small when the break size is small. When $N = 200$ and $r = 1$, the estimated change point has a 60% probability to be identical to the true break point. It is therefore of interest to study the distribution of $\hat{k}$. Limiting distribution can be used to construct confidence intervals for the true break point. One can select the proper change point based on the confidence intervals. Figures 3.1-3.6 show the simulated distributions of estimated change points. The data generating process is the same as in Table 3.1. All graphs are calculated by 10,000 times. As shown, many estimated change points are away from the true point when the break size is small. However, the estimated change

Table 3.1: Small Break in Factor Loadings: $k_0 = T/2$

N/T	r=1	r=2
10,50	12.1618 (8.00%)	9.4965 (15.50%)
20,50	9.4793 (16.50%)	7.0361 (28.20%)
30,50	7.2205 (24.30%)	4.4679 (36.90%)
40,50	6.2258 (26.50%)	3.1198 (43.60%)
50,50	5.2850 (32.50%)	2.1029 (50.20%)
100,50	2.5931 (45.70%)	1.1014 (68.20%)
200,50	1.2402 (61.50%)	0.6107 (80.80%)

point follows a certain of distribution. As N increases, the estimated change point is more close to the true break point.

We make the main assumptions as follows:

Assumption A (Factors): $E\|f_t\|^4 < M$. Also, there exists an $r \times r$ positive definite matrix Σ_F such that $F'F/T \xrightarrow{p} \Sigma_F$ as $T \to \infty$, $\frac{1}{k_0}\sum_{t=1}^{k_0} f_t f_t' \xrightarrow{p} \Sigma_F$, $\frac{1}{T-k_0}\sum_{t=k_0+1}^{T} f_t f_t' \xrightarrow{p} \Sigma_F$.

Assumption B (Factors Loadings): For $l = 1, 2$, $\max_{1\leq i\leq N} E\|\lambda_{il}\|^4 \leq C < \infty$, and there exists an $r \times r$ positive definite matrix Σ_Λ such that $\|\Lambda_l'\Lambda_l/N - \Sigma_{\Lambda_l}\| \xrightarrow{p} 0$ as $N \to \infty$.

Assumption C (Idiosyncratic Components): As $N, T \to \infty$,

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

2. $E(\frac{e_s' e_t}{N}) = \gamma_N(s,t)$, $|\gamma_N(s,s)| \leq M$ for all $s = 1, \ldots, T$, and

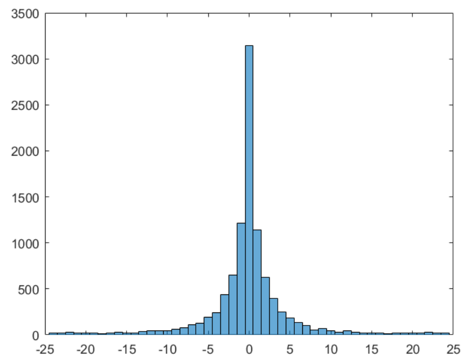


Figure 3.1: Distribution: $N = 50$, $T = 50$, $r = 1$

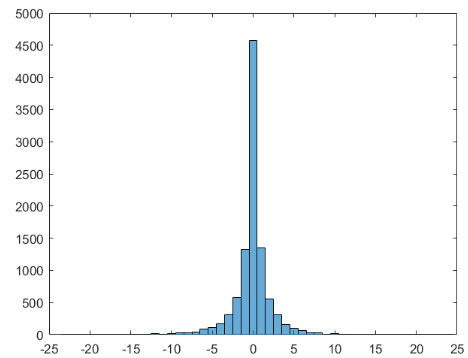


Figure 3.2: Distribution: $N = 100$, $T = 50$, $r = 1$

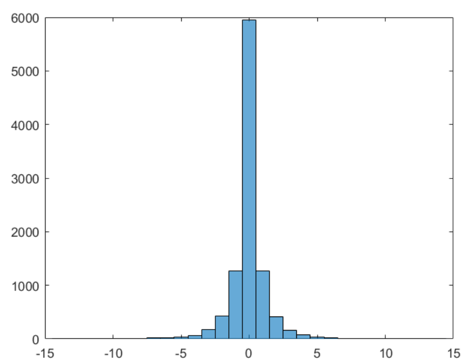


Figure 3.3: Distribution: $N = 200$, $T = 50$, $r = 1$

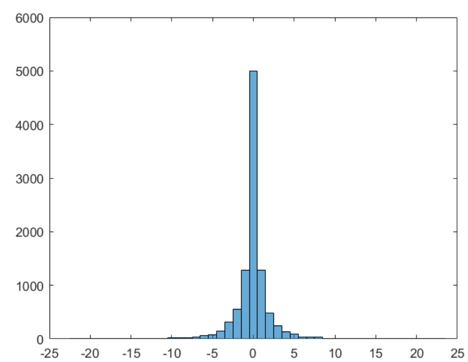


Figure 3.4: Distribution: $N = 50$, $T = 50$, $r = 2$

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

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

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

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

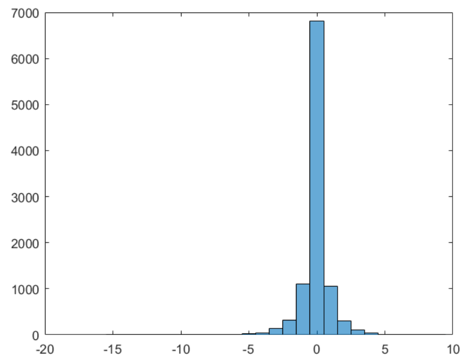


Figure 3.5: Distribution: $N = 100$, $T = 50$, $r = 2$

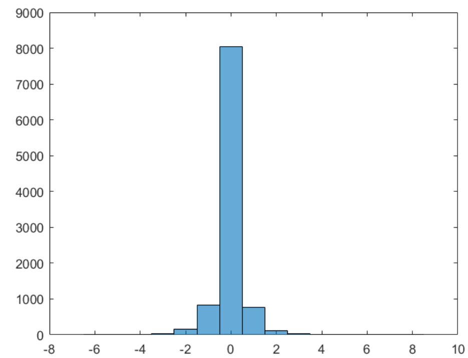


Figure 3.6: Distribution: $N = 200$, $T = 50$, $r = 2$

Assumption D Weak Dependence Between Factors and Idiosyncratic Components:

$$E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{k_0}}\sum_{t=1}^{k_0} f_t e_{it}\right\|^2\right) \leq M, \; and$$
$$E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T-k_0}}\sum_{t=k_0+1}^{T} f_t e_{it}\right\|^2\right) \leq M.$$

Assumption E Small magnitude of break:

$$\lambda_{i1} - \lambda_{i2} = N^{-1/2}\Delta_i, \qquad \text{and}$$
$$\lim_{N\to\infty}\sum_{i=1}^{N}\|\lambda_{i1} - \lambda_{i2}\|^2 = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\|\Delta_i\|^2 = \Sigma_\Delta$$

Assumptions A-D are either from or natural extensions of Assumptions A-D in Bai and Ng (2002). Assumptions A restricts the factors so that the second moments exist. Remarkably, the multiplicative structure of $F\Lambda'$ causes an identification issue. To see the identification issue in the presence of structural changes, suppose that the dynamics of f_t change at k_0, such that the second moment of f_t is doubled afterward. We can establish an observationally equivalent model, where the second moments of factors are constant over time, but the factor loading matrix is scaled by $\sqrt{T}$ after k_0. One way to resolve this indeterminacy. Chen et al. (2014), Cheng et al (2015), Han and Inoue (2014), Baltagi et al. (2015), and Massacci (2015) all assume the uniform convergence conditions that $\frac{1}{k_0}\sum_{t=1}^{k_0} f_t f_t' \xrightarrow{p} \Sigma_F$ and $\frac{1}{T-k_0}\sum_{t=k_0+1}^{T} f_t f_t' \xrightarrow{p} \Sigma_F$. Assumption B ensures that each factor has a nontrivial contribution to the variance of observed data. Both pre- and post-break factor loadings are non stochastic. Assumptions C-D are also similar to Assumptions C-D in Bai and Ng (2002), which allow for limited time-series and cross-section dependence in idiosyncratic components, and also weak dependence between factors (factor loadings) and idiosyncratic errors. Assumption E is that the sum of magnitude of break is small, instead of infinity, as $N \to \infty$. Under Assumptions A-E, we can show that $\hat{k} - k_0 = O_p(1)$. Baltagi et. al (2015) also showed that the estimated change point is statistically bounded ($O_p(1)$) based on their method. Thus, because of the small magnitude of breaks, $\hat{k}$ does not collapse to k_0, leading to a non-degenerate distribution. Nevertheless, this condition specifies that the break fraction can be estimated, because $\hat{\tau} = \hat{k}/T$ remains T consistent for τ_0.

Following these assumptions and equations (3.2) and (3.3), we can state the limiting distribution of the estimated change point in the following form.

Theorem 3.2.1 *Assume e_{it} are uncorrelated over t and i. Under Assumptions A-E,* *as* $N, T \to \infty$,

$$\hat{k} - k_0 \xrightarrow{d} \arg\min_{l} \left[|l| C_1 + C_2 W(l) + C_3 U(l) \right] \tag{3.4}$$

where $W(0) = 0$ *and*

$$W(l) = \sum_{s=-l+1}^{0} Z_s, \qquad l = -1, -2, \ldots$$

$$W(l) = \sum_{s=1}^{l} Z_s, \qquad l = 1, 2, \ldots$$

$$U(l) = \sum_{s=-l+1}^{0} \chi_1^2(s), \qquad l = -1, -2, \ldots$$

$$U(l) = \sum_{s=1}^{l} \chi_1^2(s), \qquad l = 1, 2, \ldots$$

and Z_s, $s = \ldots, -2, -1, 0, 1, 2, \ldots$ are *i.i.d.* standard normal random variables, $\chi_1^2(s)$ is chi-squared random variables with degree of freedom 1.

Corollary 3.2.1: *Assume* e_{it} *are uncorrelated over* t*, under the same assumptions in Theorem 3.2.1,*

$$\hat{k} - k_0 \xrightarrow{d} \arg\min_{l} \left[|l| C_1^* + C_2^* W(l) + C_3^* U(l) \right] \tag{3.5}$$

A formal proof of Theorem 3.2.1 is provided in the Appendix B. The key distinction between our limiting distribution and the limiting distribution in the conventional panel data models with structural changes is that the random parts in Theorem 3.2.1 depend on *i.i.d.* standard normal random variables and chi square variables. The limiting distribution of the estimated break point is more complex than that of the conventional panel data models, because both factors and factor loadings are unobservable. It is affected by the estimated factors and estimated factor loadings. The normality of Z_s results from the central limit theorem applied to product of factors and errors. The chi-squared part results from the central limit theorem applied to the square of factor loadings multiplied errors. Unlike the limiting distribution of estimated change point in panel data models, we cannot obtain the distribution through a simple variable transformation. This limiting distribution is data dependent. For given data, we can compute C_1, C_2, and C_3, and we can simulate the distribution on the right side of Theorem 3.2.1. From the simulated distribution, confidence intervals on k_0 can also be constructed. Based on the confidence intervals, we can select the proper change point in empirical applications. Also note that, C_1, C_2, and C_3 can be estimated consistently.

3.3 Simulations

3.3.1 Comparison

In this section, we conduct Monte Carlo simulation to evaluate the performance of the least squares method in finite sample. Baltagi et al. (2015) proposed that estimating the change point of factor loadings can be converted to estimating change point of the second moments of estimated factors. However, this methodology does not take full advantage of the panel data. To see the advantage of using least squares method, we compare our approach with quasi-maximum likelihood method (QML) from Bai (2010), second moments method from Baltagi et al. (2015), and MLE method. The same data generating processes (DGPs) are applied for each method.

We fix T at 50 and 100 and let N = 20, 30, 40, 50, 100, 500 and 1000. For the true number of factors, we fix r = 1, 2 and $[1.25\log(\max(N, T))]$ (let r increase with sample size, denoted as "increasing r" in the table), where $[A]$ denotes the integer part of a real number A. The break point k_0 is assumed as $[T/2]$, $[T/3]$ and $[T/4]$.

The data generating processes (DGP) have the form as follows:

$$x_{it} = \begin{cases} \lambda_{i1}^{'} f_t + e_{it} & \text{for t} = 1, 2, \ldots, k_0 \\ \lambda_{i2}^{'} f_t + e_{it} & \text{for t} = k_0 + 1, k_0 + 2, \ldots, T \end{cases}$$

$$f_t \sim i.i.d.N(0,1),$$

$$e_{it} \sim i.i.d.N(0,1).$$

Before the break point, we let $\lambda_i \sim i.i.d.N(0,1)$. After the break point, we consider four cases. In Case 1, we redraw λ_i from $N(0,1)$ after the break point k_0. The distribution of post-break factor loadings does not change. This case is denoted as **DGP1**. In Case 2, the distribution of post-break factor loadings changes from $N(0,1)$ to $N(0,0.5)$. We denote this case as **DGP2**. For the third case, the distribution of post-break factor loadings changes from $N(0,1)$ to $N(0.5,1)$. We denote this case as **DGP3**. For the last case, denoted as **DGP4**, we let post-break λ_i keep the same with pre-break λ_i for $i = 1, 2, \ldots, N/2$. For $i = N/2+1, N/2+2, \ldots, N$, we redraw the factor loadings from $N(0,1)$ after break point k_0. All simulations are based on 1000 replications. The reported results are the root mean square error of estimated change point, which

$$RMSE = \sqrt{\frac{1}{1000} \sum_{s=1}^{1000} \left(\hat{k}_0 - k_0\right)^2}.$$

The number in parentheses is the percentage of obtaining correct change point.

3.3.1.1 Least Squares Method

We denote least squares method as **Method 1**. Tables 3.2-3.5 show the results of $RMSE$ and percentages of finding correct change point for each case. All four cases have similar properties. As N increases, the $RMSE$ decreases. When T jumps from 50 to 100, $RMSE$ slightly decreases. When the true number of factors increases, the estimated results significantly improve. We also try r=[1.5log(max(N,T))] (results not showed here), the estimated change points start to diverge when N is very large and T is relatively small. And the percentage of finding correct break point decreases.

3.3.1.2 Unknown Number of Factors

Chen (2015) proposed a consistent estimator for the break fraction by using $\tilde{r}-1$ number of factors to estimate the change point. Because the number of factors tend to be doubled when factor loadings all change, we estimate the pre-break and post-break factor loadings and factors with $2r-1$ number of factors when the number of factors are unknown (denoted as r^* in the table). Then we substitute these estimated factor loadings (N by $r*$) and factors (T by $r*$) into our objective function to estimate the change point. This method is denoted as **Method 2**.

Tables 3.6-3.9 show the results for Method 2. The situation is the same when $r = 1 = \tilde{r}$, thus we consider only $r = 2$ and *increasing* r cases. When $r = 2$, the results have the same trend as in Method 1. However, Method 1's results are more accurate than Method 2 in many cases, because we use the true number of factors. In the "increasing r" case, the estimated change point diverges when N is very large (500 or 1000) and T is relatively small (50). The break point moves further from the middle point, and the ratio of corresponding N and T in which the results diverge decreases. For the following three approaches, we all assume the number of factors is unknown. We use $r* = 2r-1$ instead of r to estimate factors and factor loadings.

Table 3.2: RMSE using Method 1 under DGP1

	k_0=T/4			k_0=T/3			k_0=T/2		
N/T	r=1	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	0.9628 (70.1%)[a]	0.3924 (88.5%)	0.2049 (96.4%)	0.9028 (67.8%)	0.3332 (91.8%)	0.1549 (97.6%)	0.9638 (68.9%)	0.4324 (88.1%)	0.2074 (96.6%)
30,50	0.7280 (74.6%)	0.3376 (92.3%)	0.1517 (97.7%)	0.7596 (72.9%)	0.3162 (92.0%)	0.1095 (98.8%)	0.7752 (74.5%)	0.2933 (92.9%)	0.0894 (99.2%)
40,50	0.6387 (79.1%)	0.2470 (95.0%)	0.0707 (99.5%)	0.6181 (79.7%)	0.2470 (95.1%)	0.0775 (99.4%)	0.6066 (79.0%)	0.2665 (94.1%)	0.0837 (99.3%)
50,50	0.5822 (78.9%)	0.2530 (94.8%)	0.0632 (99.6%)	0.6000 (79.3%)	0.2098 (95.9%)	0.0949 (99.1%)	0.6107 (82.2%)	0.2387 (94.6%)	0.0548 (99.7%)
100,50	0.4427 (87.0%)	0.1342 (98.5%)	0 (100%)	0.4615 (85.7%)	0.1549 (97.6%)	0.0316 (99.9%)	0.4370 (86.6%)	0.1517 (97.7%)	0 (100%)
500,50	0.2757 (93.6%)	0.0316 (99.9%)	0 (100%)	0.2490 (93.8%)	0.0775 (99.4%)	0 (100%)	0.2387 (94.9%)	0.0775 (99.4%)	0 (100%)
1000,50	0.2236 (95.9%)	0.0447 (99.8%)	0 (100%)	0.2366 (95.8%)	0.0447 (99.8%)	0 (100%)	0.2236 (95.6%)	0.0316 (99.9%)	0 (100%)
20,100	0.8849 (69.5%)	0.4025 (89.3%)	0.1643 (97.6%)	0.9508 (66.8%)	0.3728 (88.5%)	0.1225 (98.5%)	0.8000 (69.9%)	0.3886 (89.7%)	0.1140 (98.7%)
30,100	0.7899 (74.8%)	0.3376 (92.4%)	0.0837 (99.6%)	0.7324 (76.6%)	0.3347 (92.2%)	0.0632 (99.6%)	0.6841 (77.3%)	0.3146 (92.2%)	0.0447 (99.8%)
40,100	0.6738 (78.0%)	0.2258 (95.2%)	0.0632 (99.6%)	0.6245 (78.4%)	0.2470 (94.8%)	0.0447 (99.8%)	0.6648 (77.5%)	0.2683 (94.2%)	0.0316 (99.9%)
50,100	0.6293 (79.5%)	0.2236 (95.3%)	0 (100%)	0.5805 (78.9%)	0.2366 (95.2%)	0 (100%)	0.5727 (81.4%)	0.2429 (94.4%)	0 (100%)
100,100	0.4615 (86.0%)	0.1342 (98.2%)	0 (100%)	0.4528 (84.9%)	0.1483 (97.8%)	0 (100%)	0.4231 (86.6%)	0.1378 (98.1%)	0 (100%)
500,100	0.2490 (95.0%)	0.0707 (99.5%)	0 (100%)	0.2569 (94.6%)	0.0707 (99.5%)	0 (100%)	0.2739 (94.0%)	0.0837 (99.3%)	0 (100%)
1000,100	0.2258 (95.5%)	0.0316 (99.9%)	0 (100%)	0.2302 (95.3%)	0.0447 (99.8%)	0 (100%)	0.2214 (96.0%)	0.0447 (99.8%)	0 (100%)

[a] The number in parentheses is the percentage of obtaining correct change point.

Table 3.3: RMSE using Method 1 under DGP2

	k_0=T/4			k_0=T/3			k_0=T/2		
N/T	r=1	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	1.2128 (63.3%)[a]	0.5941 (82.0%)	0.3286 (92.4%)	1.2853 (59.5%)	0.6496 (82.1%)	0.2683 (94.6%)	1.3210 (62.5%)	0.5604 (81.2%)	0.2530 (94.2%)
30,50	0.9110 (69.8%)	0.3873 (90.8%)	0.2025 (97.3%)	0.9252 (68.8%)	0.4135 (87.3%)	0.1817 (96.7%)	0.9659 (70.8%)	0.4219 (87.6%)	0.1789 (97.1%)
40,50	0.8289 (73.6%)	0.3362 (91.4%)	0.1140 (98.7%)	0.7817 (73.9%)	0.3606 (90.3%)	0.1049 (98.9%)	0.7517 (75.1%)	0.3479 (90.8%)	0.1000 (99.0%)
50,50	0.6716 (75.4%)	0.2510 (94.3%)	0.1049 (98.9%)	0.7014 (77.0%)	0.2811 (93.8%)	0.0837 (99.3%)	0.7106 (77.6%)	0.2864 (94.5%)	0.0837 (99.3%)
100,50	0.5273 (85.0%)	0.2121 (96.1%)	0.0316 (99.9%)	0.5376 (82.5%)	0.1612 (97.4%)	0 (100%)	0.5975 (81.5%)	0.1897 (97.0%)	0 (100%)
500,50	0.2966 (92.7%)	0.0548 (99.7%)	0 (100%)	0.2898 (92.8%)	0.0775 (99.4%)	0 (100%)	0.2983 (93.2%)	0.0775 (99.4%)	0 (100%)
1000,50	0.2490 (94.4%)	0.0548 (99.7%)	0 (100%)	0.2627 (93.7%)	0.0837 (99.3%)	0 (100%)	0.2490 (95.0%)	0.0548 (99.7%)	0 (100%)
20,100	1.0968 (63.2%)	0.5639 (82.6%)	0.1871 (96.8%)	1.1623 (65.2%)	0.5604 (83.1%)	0.1975 (96.7%)	1.1480 (64.6%)	0.5301 (83.9%)	0.2191 (95.5%)
30,100	0.8860 (70.4%)	0.4025 (89.4%)	0.1095 (98.8%)	0.8786 (70.5%)	0.3937 (89.4%)	0.1140 (99.0%)	0.8837 (72.7%)	0.3975 (88.6%)	0.0949 (99.1%)
40,100	0.7543 (73.6%)	0.3146 (92.4%)	0.0548 (99.7%)	0.8031 (73.9%)	0.2881 (92.9%)	0.0707 (99.5%)	0.7609 (74.6%)	0.3479 (91.3%)	0.0707 (99.5%)
50,100	0.6557 (79.0%)	0.2828 (94.4%)	0.0632 (99.6%)	0.6450 (77.4%)	0.2793 (93.1%)	0.0707 (99.5%)	0.6535 (78.1%)	0.2864 (93.5%)	0.0316 (99.9%)
100,100	0.4806 (85.6%)	0.1612 (97.4%)	0 (100%)	0.4754 (84.9%)	0.1975 (96.7%)	0 (100%)	0.4754 (83.9%)	0.1975 (96.1%)	0 (100%)
500,100	0.2811 (93.5%)	0.0316 (99.9%)	0 (100%)	0.3225 (92.4%)	0.1000 (99.0%)	0 (100%)	0.3017 (93.0%)	0.0775 (99.4%)	0 (100%)
1000,100	0.2387 (94.9%)	0.0447 (99.8%)	0 (100%)	0.2530 (94.5%)	0.0316 (99.9%)	0 (100%)	0.2302 (95.0%)	0 (100%)	0 (100%)

[a]The number in parentheses is the percentage of obtaining correct change point.

Table 3.4: RMSE using Method 1 under DGP3

	k_0=T/4			k_0=T/3			k_0=T/2		
N/T	r=1	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	0.8826 (71.8%)[a]	0.3987 (89.1%)	0.2121 (95.8%)	0.8888 (71.5%)	0.3886 (87.9%)	0.1703 (97.1%)	0.8056 (72.1%)	0.3606 (90.5%)	0.1732 (97.3%)
30,50	0.7043 (75.0%)	0.2793 (93.1%)	0.1000 (99.0%)	0.7675 (76.9%)	0.2915 (91.8%)	0.1225 (98.5%)	0.6512 (76.4%)	0.3240 (92.3%)	0.0949 (99.1%)
40,50	0.6042 (78.8%)	0.2627 (94.8%)	0.0548 (99.7%)	0.5771 (81.3%)	0.2324 (95.2%)	0.0548 (99.7%)	0.6488 (79.0%)	0.2236 (95.3%)	0.0632 (99.6%)
50,50	0.5376 (81.4%)	0.1817 (97.0%)	0.0548 (99.7%)	0.6427 (80.0%)	0.2191 (96.1%)	0 (100%)	0.5657 (82.2%)	0.2121 (95.8%)	0.0548 (99.7%)
100,50	0.4183 (87.7%)	0.1414 (98.0%)	0 (100%)	0.4159 (87.3%)	0.1265 (98.4%)	0.0316 (99.9%)	0.4243 (86.9%)	0.1304 (98.3%)	0 (100%)
500,50	0.2702 (93.9%)	0.0707 (99.5%)	0 (100%)	0.2720 (93.8%)	0.0837 (99.3%)	0 (100%)	0.2550 (94.1%)	0.0707 (99.5%)	0 (100%)
1000,50	0.1924 (96.3%)	0.0316 (99.9%)	0 (100%)	0.2121 (95.8%)	0.0548 (99.7%)	0 (100%)	0.2191 (95.80%)	0.0447 (99.80%)	0 (100%)
20,100	0.8526 (69.9%)	0.4111 (89.2%)	0.1304 (98.6%)	0.7649 (69.3%)	0.4393 (89.3%)	0.1304 (98.3%)	0.8343 (71.8%)	0.3728 (90.0%)	0.1095 (98.8%)
30,100	0.6819 (76.3%)	0.3114 (92.7%)	0.0548 (99.7%)	0.7085 (77.2%)	0.3146 (92.4%)	0 (100%)	0.7287 (74.4%)	0.2881 (93.2%)	0.0548 (99.7%)
40,100	0.5925 (79.6%)	0.2530 (94.5%)	0.0316 (99.9%)	0.6364 (78.9%)	0.2793 (93.7%)	0.0316 (99.9%)	0.5992 (79.4%)	0.2429 (95.0%)	0.0548 (99.7%)
50,100	0.5797 (82.1%)	0.2168 (96.2%)	0.0447 (99.8%)	0.5710 (81.0%)	0.2191 (95.8%)	0.0316 (99.9%)	0.5040 (83.0%)	0.2025 (96.2%)	0.0447 (99.8%)
100,100	0.4000 (88.4%)	0.1378 (98.1%)	0 (100%)	0.4637 (88.0%)	0.1183 (98.6%)	0 (100%)	0.4615 (87.0%)	0.1342 (98.2%)	0 (100%)
500,100	0.2449 (94.6%)	0.0775 (99.4%)	0 (100%)	0.2449 (94.0%)	0.0632 (99.6%)	0 (100%)	0.2510 (95.1%)	0.0316 (99.9%)	0 (100%)
1000,100	0.2074 (96.3%)	0.0447 (99.8%)	0 (100%)	0.2145 (95.7%)	0.0447 (99.8%)	0 (100%)	0.2000 (96.9%)	0.0447 (99.8%)	0 (100%)

[a] The number in parentheses is the percentage of obtaining correct change point.

Table 3.5: RMSE using Method 1 under DGP4

	k_0=T/4			k_0=T/3			k_0=T/2		
N/T	r=1	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	1.9375 (61.3%)[a]	0.5657 (83.4%)	0.3130 (92.0%)	1.2911 (61.6%)	0.5235 (81.7%)	0.2470 (95.6%)	1.1925 (60.8%)	0.5925 (83.4%)	0.2345 (96.0%)
30,50	0.8888 (72.9%)	0.3795 (89.3%)	0.2000 (97.1%)	1.0756 (71.1%)	0.4243 (88.6%)	0.1549 (97.6%)	0.9263 (70.1%)	0.4427 (89.2%)	0.1732 (97.6%)
40,50	0.8343 (73.0%)	0.3240 (92.7%)	0.1265 (98.4%)	0.7430 (73.4%)	0.2966 (92.9%)	0.0775 (99.4%)	0.7409 (75.9%)	0.3302 (91.5%)	0.1225 (98.5%)
50,50	0.6442 (78.0%)	0.2665 (93.8%)	0.0775 (99.4%)	0.6993 (75.6%)	0.2739 (93.4%)	0.0847 (99.3%)	0.7134 (75.4%)	0.2490 (94.1%)	0.0447 (99.8%)
100,50	0.4837 (84.4%)	0.1449 (97.9%)	0 (100%)	0.5079 (85.0%)	0.1924 (96.6%)	0.0447 (99.8%)	0.5167 (83.7%)	0.1703 (97.4%)	0.0447 (99.8%)
500,50	0.2408 (95.1%)	0.0775 (99.4%)	0 (100%)	0.3178 (91.9%)	0.0775 (99.4%)	0 (100%)	0.2983 (92.3%)	0.0707 (99.5%)	0 (100%)
1000,50	0.2258 (95.2%)	0 (100%)	0 (100%)	0.2025 (96.2%)	0.0447 (99.8%)	0 (100%)	0.2214 (95.7%)	0.0548 (99.7%)	0 (100%)
20,100	1.1815 (64.2%)	0.5865 (82.6%)	0.1844 (97.2%)	1.3442 (65.1%)	0.6099 (81.8%)	0.1732 (97.6%)	1.2333 (62.1%)	0.5486 (83.8%)	0.1844 (97.5%)
30,100	0.9044 (72.0%)	0.3701 (89.7%)	0.1049 (98.9%)	0.8307 (70.9%)	0.3674 (88.9%)	0.0949 (99.1%)	0.8307 (71.9%)	0.3464 (90.6%)	0.1483 (98.4%)
40,100	0.8349 (73.6%)	0.3225 (91.6%)	0.0775 (99.4%)	0.7197 (76.9%)	0.3050 (92.7%)	0.0632 (99.6%)	0.7183 (77.1%)	0.2915 (92.7%)	0.0775 (99.4%)
50,100	0.7078 (74.9%)	0.2915 (93.5%)	0.0447 (99.8%)	0.6701 (78.1%)	0.2702 (93.3%)	0.0316 (99.9%)	0.7204 (78.8%)	0.2864 (93.0%)	0.0632 (99.6%)
100,100	0.4572 (84.5%)	0.1449 (97.9%)	0 (100%)	0.4950 (84.8%)	0.1817 (97.3%)	0 (100%)	0.4690 (84.9%)	0.1581 (97.5%)	0 (100%)
500,100	0.2739 (93.4%)	0.0707 (99.5%)	0 (100%)	0.2702 (93.3%)	0.0837 (99.3%)	0 (100%)	0.2828 (93.2%)	0.0775 (99.4%)	0 (100%)
1000,100	0.2302 (94.7%)	0.0447 (99.8%)	0 (100%)	0.2345 (95.1%)	0.0548 (99.7%)	0 (100%)	0.2236 (95.9%)	0.0447 (99.8%)	0 (100%)

[a] The number in parentheses is the percentage of obtaining correct change point.

Table 3.6: RMSE using Method 2 under DGP1

	k_0=T/4		k_0=T/3		k_0=T/2	
N,T	r=2[a]	increasing r	r=2	increasing r	r=2	increasing r
20,50	0.8390 (77.5%)[b]	1.7627 (56.8%)	0.6656 (77.5%)	1.0188 (71.1%)	0.5831 (83.4%)	0.6033 (81.4%)
30,50	0.6450 (83.3%)	1.1437 (66.7%)	0.5273 (86.6%)	0.5916 (83.8%)	0.4572 (87.1%)	0.3435 (91.4%)
40,50	0.5254 (87.8%)	0.9317 (68.1%)	0.3937 (89.5%)	0.4301 (87.8%)	0.3317 (90.8%)	0.3000 (93.6%)
50,50	0.3847 (90.4%)	0.8894 (74.5%)	0.3507 (93.0%)	0.4099 (92.0%)	0.3302 (92.7%)	0.1844 (97.5%)
100,50	0.2864 (93.5%)	1.1528 (51.3%)	0.2214 (96.2%)	0.4266 (88.3%)	0.1897 (96.7%)	0.1095 (98.8%)
500,50	0.1378 (98.4%)	3.0251 (0%)	0.1049 (98.9%)	1.1747 (44.9%)	0.1000 (99.0%)	0.0316 (99.9%)
1000,50	0.1304 (98.3%)	4.3602 (0%)	0.0775 (99.4%)	2.0425 (9.7%)	0.1000 (99.0%)	0.0632 (99.6%)
20,100	0.5967 (83.5%)	0.9311 (76.2%)	0.5532 (82.6%)	0.6797 (82.2%)	0.5718 (82.6%)	0.4817 (88.1%)
30,100	0.4517 (87.0%)	0.6099 (87.5%)	0.3937 (88.6%)	0.2828 (94.3%)	0.4099 (89.7%)	0.2324 (95.5%)
40,100	0.4324 (89.6%)	0.3435 (92.4%)	0.3271 (92.0%)	0.2366 (95.9%)	0.3209 (92.0%)	0.1673 (97.5%)
50,100	0.2966 (93.3%)	0.2098 (96.2%)	0.2983 (93.1%)	0.1183 (98.6%)	0.2720 (93.5%)	0.1304 (98.3%)
100,100	0.2191 (95.5%)	0.1643 (97.6%)	0.1871 (96.8%)	0.0632 (99.6%)	0.1897 (96.7%)	0.0316 (99.9%)
500,100	0.0837 (99.3%)	0.1871 (96.5%)	0.0707 (99.5%)	0.0316 (99.9%)	0.0949 (99.1%)	0 (100%)
1000,100	0.0447 (99.8%)	0.2846 (93.4%)	0.0632 (99.6%)	0.0316 (99.9%)	0.0837 (99.3%)	0 (100%)

[a] The number of factors we used here is $\tilde{r} = 2 \times r - 1$.
[b] The number in parentheses is the percentage of obtaining correct change point.

Table 3.7: RMSE using Method 2 under DGP2

	k_0=T/4		k_0=T/3		k_0=T/2	
N,T	r=2[a].	increasing r	r=2	increasing r	r=2	increasing r
20,50	1.2633 (67.0%)[b]	2.6734 (43.8%)	1.1122 (70.6%)	1.4426 (62.3%)	0.8050 (75.2%)	0.8591 (74.2%)
30,50	0.7765 (77.2%)	1.5707 (58.5%)	0.6189 (81.4%)	0.9055 (74.9%)	0.6008 (81.4%)	0.4889 (85.2%)
40,50	0.6450 (81.9%)	1.2256 (64.2%)	0.5235 (86.2%)	0.6042 (82.3%)	0.4701 (87.0%)	0.3633 (91.6%)
50,50	0.4940 (84.6%)	1.0469 (68.7%)	0.4648 (87.1%)	0.5357 (85.0%)	0.4062 (89.4%)	0.2408 (94.8%)
100,50	0.3728 (91.3%)	1.2190 (50.1%)	0.3178 (92.3%)	0.4909 (85.6%)	0.2646 (93.3%)	0.1732 (97.6%)
500,50	0.1761 (97.2%)	3.0596 (0%)	0.1789 (97.7%)	1.1713 (44.4%)	0.1517 (97.7%)	0.0316 (99.9%)
1000,50	0.1517 (97.7%)	4.3960 (0%)	0.1378 (98.4%)	2.0540 (9.8%)	0.1049 (98.9%)	0.0632 (99.6%)
20,100	0.9633 (73.7%)	1.6300 (65.5%)	0.7785 (77.1%)	1.0169 (74.2%)	0.7994 (77.0%)	0.5736 (82.0%)
30,100	0.6782 (81.1%)	0.7183 (81.0%)	0.5639 (83.6%)	0.5196 (87.5%)	0.6269 (85.8%)	0.3114 (93.0%)
40,100	0.4615 (87.1%)	0.4743 (88.5%)	0.4899 (87.9%)	0.3332 (93.6%)	0.3987 (88.4%)	0.1897 (97.0%)
50,100	0.4159 (90.6%)	0.3592 (92.9%)	0.3450 (91.5%)	0.2236 (96.5%)	0.3256 (90.9%)	0.1449 (97.7%)
100,100	0.2950 (93.6%)	0.2000 (96.3%)	0.2236 (95.6%)	0.1049 (98.9%)	0.2775 (93.2%)	0.0632 (99.6%)
500,100	0.1449 (97.9%)	0.2258 (95.8%)	0.0949 (99.1%)	0.0548 (99.7%)	0.1049 (98.9%)	0 (100%)
1000,100	0.0894 (99.2%)	0.2846 (92.8%)	0.0775 (99.4%)	0.0447 (99.8%)	0.1000 (99.0%)	0 (100%)

[a]The numbre of factors we used here is $\tilde{r} = 2 \times r - 1$

[b]The number in parentheses is the percentage of obtaining correct change point.

Table 3.8: RMSE using Method 2 under DGP3

	k_0=T/4		k_0=T/3		k_0=T/2	
N,T	r=2[a].	increasing r	r=2	increasing r	r=2	increasing r
20,50	0.6812 (77.0%)[b]	1.6334 (56.7%)	0.6957 (82.3%)	0.9192 (74.8%)	0.6066 (82.6%)	0.6595 (81.9%)
30,50	0.5148 (86.1%)	0.9874 (69.2%)	0.4593 (88.9%)	0.5958 (83.0%)	0.4775 (86.1%)	0.3606 (91.5%)
40,50	0.4219 (89.4%)	0.8562 (72.1%)	0.3592 (91.0%)	0.4817 (88.6%)	0.3302 (90.6%)	0.2449 (94.9%)
50,50	0.3619 (90.5%)	0.7362 (77.8%)	0.2665 (92.9%)	0.3742 (91.1%)	0.2702 (94.4%)	0.2280 (95.7%)
100,50	0.2588 (95.0%)	1.1032 (56.2%)	0.2569 (95.2%)	0.3987 (89.0%)	0.2121 (96.1%)	0.1049 (99.2%)
500,50	0.1140 (98.7%)	2.9751 (0%)	0.1095 (98.8%)	1.1036 (46.7%)	0.0949 (99.1%)	0.0632 (99.6%)
1000,50	0.0775 (99.4%)	4.2872 (0%)	0.1049 (98.9%)	1.8887 (12.4%)	0.0775 (99.4%)	0.0447 (99.8%)
20,100	0.5225 (84.1%)	0.9664 (76.9%)	0.5060 (84.2%)	0.6395 (85.0%)	0.4733 (84.9%)	0.4517 (88.5%)
30,100	0.4000 (89.6%)	0.4324 (88.5%)	0.4000 (89.8%)	0.3114 (93.9%)	0.3987 (89.5%)	0.2214 (95.7%)
40,100	0.3317 (92.5%)	0.3674 (92.3%)	0.2720 (92.9%)	0.2025 (97.1%)	0.3000 (92.8%)	0.1140 (99.0%)
50,100	0.2983 (93.2%)	0.2757 (94.2%)	0.2280 (95.7%)	0.1517 (98.0%)	0.2775 (93.8%)	0.1140 (98.7%)
100,100	0.1673 (97.2%)	0.1378 (98.1%)	0.2258 (95.5%)	0.0316 (99.9%)	0.1871 (96.5%)	0.0447 (99.8%)
500,100	0.1049 (98.9%)	0.1871 (96.8%)	0.0548 (99.7%)	0.0316 (99.9%)	0.1000 (99.0%)	0 (100%)
1000,100	0.0548 (99.7%)	0.2898 (92.8%)	0.0775 (99.4%)	0 (100%)	0.0837 (99.3%)	0 (100%)

[a]The numbre of factors we used here is $\tilde{r} = 2 \times r - 1$

[b]The number in parentheses is the percentage of obtaining correct change point.

Table 3.9: RMSE using Method 2 under DGP4

	k_0=T/4		k_0=T/3		k_0=T/2	
N,T	r=2[a].	increasing r	r=2	increasing r	r=2	increasing r
20,50	1.5162 (70.3%)[b]	2.5118 (48.3%)	1.0663 (72.1%)	1.3210 (66.3%)	0.9017 (72.4%)	0.8994 (75.5%)
30,50	0.7106 (80.9%)	1.3885 (62.3%)	0.6017 (80.8%)	0.8044 (76.8%)	0.5486 (83.0%)	0.4848 (86.9%)
40,50	0.6173 (83.7%)	1.2566 (68.5%)	0.4615 (87.2%)	0.5857 (84.1%)	0.5040 (86.0%)	0.2881 (93.4%)
50,50	0.4722 (87.4%)	0.9706 (71.9%)	0.4074 (87.7%)	0.4266 (88.4%)	0.4123 (88.4%)	0.2510 (94.8%)
100,50	0.2966 (94.0%)	1.2029 (53.7%)	0.2775 (93.8%)	0.4722 (87.5%)	0.2302 (95.3%)	0.1049 (98.9%)
500,50	0.1449 (97.9%)	3.0840 (0%)	0.1265 (98.4%)	1.1367 (45.1%)	0.1140 (98.7%)	0.0316 (99.9%)
1000,50	0.1183 (98.6%)	4.3629 (0%)	0.0837 (99.3%)	1.9728 (10.9%)	0.1000 (99.0%)	0.0707 (99.5%)
20,100	0.8408 (76.0%)	1.3420 (66.5%)	0.7259 (78.1%)	0.9290 (77.7%)	0.7335 (76.3%)	0.6332 (82.9%)
30,100	0.5357 (85.0%)	0.5874 (85.3%)	0.4604 (88.1%)	0.6075 (89.2%)	0.4461 (85.5%)	0.2775 (94.3%)
40,100	0.4012 (87.6%)	0.4290 (90.9%)	0.3860 (88.9%)	0.2950 (94.6%)	0.3937 (89.3%)	0.1673 (97.2%)
50,100	0.4147 (89.5%)	0.3209 (94.0%)	0.3536 (91.0%)	0.1871 (97.4%)	0.3592 (90.7%)	0.1378 (98.4%)
100,100	0.2258 (95.2%)	0.1643 (97.9%)	0.2345 (95.4%)	0.0775 (99.4%)	0.2098 (95.6%)	0 (100%)
500,100	0.1049 (98.9%)	0.1817 (97.0%)	0.1095 (98.8%)	0 (100%)	0.1225 (98.5%)	0 (100%)
1000,100	0.0949 (99.1%)	0.3225 (91.7%)	0.0632 (99.6%)	0 (100%)	0.0775 (99.4%)	0 (100%)

[a]The numbre of factors we used here is $\tilde{r} = 2 \times r - 1$

[b]The number in parentheses is the percentage of obtaining correct change point.

3.3.1.3 QML Method

QML Method is denoted as **Method 3**. Define

$$\bar{\hat{F}}_{j1} = \frac{1}{k}\sum_{t=1}^{k}\hat{F}_t, \qquad \bar{\hat{F}}_{j2} = \frac{1}{T-k}\sum_{t=k+1}^{T}\hat{F}_t.$$

Let

$$\hat{\sigma}_{j1}^2(k) = \frac{1}{k}\sum_{t=1}^{k}\left(\hat{F}_{jt} - \bar{\hat{F}}_{j1}\right)^2, \qquad \hat{\sigma}_{j2}^2(k) = \frac{1}{T-k}\sum_{t=k+1}^{T}\left(\hat{F}_{jt} - \bar{\hat{F}}_{j2}\right)^2,$$

the objective function becomes

$$QML(k) = k\sum_{j=1}^{r}\log\hat{\sigma}_{j1}^2(k) + (T-k)\sum_{j=1}^{r}\log\hat{\sigma}_{j2}^2(k),$$

the estimated break point is defined as $\hat{k} = \arg\min_{1\leq k\leq T}\ QML(k)$.

The QML method is used to estimate change point when there are mean or variance changes in panel data models. Tables 3.10-3.13 show the $RMSE$ and percentage of finding correct break points for four different cases using Method 3. Method 1 outperform Method 3 in all cases. The results are smaller in method 1, with same combination of N and T in the same DPGs. For the results with the same N and different T, we find that $RMSE$ increases when T is 100. Although the percentage of selecting correct break point increases when $T = 100$ and N are the same, it only works when N is small. As N increases, for many cases, the percentage decreases compared to $T = 50$. In DGP4, the estimated results do not seem to converge to the true change point.

Table 3.10: RMSE using Method 3 under DGP1

	k_0=T/4			k_0=T/3			k_0=T/2		
N,T	r=1[a]	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	10.9459 (30.4%)[b]	6.6693 (55.6%)	3.8194 (67.2%)	10.5079 (33.0%)	8.4835 (48.9%)	7.9031 (53.4%)	10.2628 (28.6%)	9.4381 (39.6%)	10.9959 (38.1%)
30,50	8.6173 (41.5%)	5.1951 (64.8%)	1.9336 (78.1%)	9.6766 (37.9%)	6.8758 (57.5%)	4.2466 (65.1%)	9.8063 (36.0%)	9.2190 (42.8%)	9.4166 (46.3%)
40,50	8.4042 (45.1%)	4.6135 (70.3%)	0.7556 (84.7%)	8.1067 (42.9%)	6.5241 (62.8%)	3.3516 (74.0%)	9.4489 (36.9%)	8.7640 (49.6%)	7.8823 (55.2%)
50,50	6.1908 (49.8%)	2.5126 (76.5%)	0.4930 (89.1%)	7.8488 (47.5%)	4.7610 (68.7%)	3.5260 (77.9%)	9.2425 (40.5%)	8.1566 (53.5%)	8.0692 (57.6%)
100,50	5.0168 (59.8%)	1.6034 (84.2%)	0.2049 (96.4%)	6.8402 (55.2%)	4.2656 (76.0%)	0.6083 (89.5%)	7.3848 (50.0%)	6.9246 (62.2%)	5.8330 (69.2%)
500,50	1.7967 (80.4%)	0.5030 (96.6%)	0.0316 (99.9%)	3.7062 (75.4%)	2.0885 (90.4%)	0.1183 (98.9%)	5.4293 (60.6%)	3.4062 (83.3%)	0.2236 (96.4%)
100,50	0.9061 (85.0%)	0.1924 (97.7%)	0 (100%)	2.2832 (82.3%)	0.9198 (95.5%)	0 (100%)	4.8780 (71.7%)	1.8590 (88.7%)	0.0837 (99.3%)
20,100	17.3257 (35.8%)	7.8699 (64.0%)	0.9198 (77.5%)	18.8261 (31.9%)	10.9149 (54.2%)	4.3098 (63.7%)	19.3194 (31.5%)	15.0783 (44.0%)	14.2864 (46.1%)
30,100	14.5148 (45.4%)	3.0176 (71.9%)	0.5621 (86.3%)	15.7772 (37.9%)	9.2596 (63.4%)	1.4255 (75.5%)	18.3141 (32.1%)	13.9754 (45.1%)	11.0530 (52.1%)
40,100	10.6355 (49.6%)	3.0373 (76.9%)	0.4050 (91.2%)	13.7450 (44.8%)	8.6943 (66.2%)	1.3119 (78.8%)	17.5717 (35.1%)	13.3254 (54.8%)	9.9666 (59.7%)
50,100	10.5639 (52.1%)	3.4784 (81.0%)	0.3000 (93.1%)	13.6131 (48.1%)	6.4424 (73.9%)	0.6870 (83.0%)	17.0071 (38.1%)	12.5731 (51.5%)	8.3694 (57.6%)
100,100	3.5364 (65.9%)	0.4648 (88.8%)	0.1483 (98.1%)	9.8759 (58.9%)	5.4958 (80.7%)	0.3435 (91.7%))	15.7710 (38.1%)	10.1534 (61.3%)	5.3707 (71.1%)
500,100	0.6542 (81.9%)	0.2000 (96.8%)	0.0316 (99.9%)	4.1566 (75.7%)	1.4748 (93.6%)	0.0949 (99.1%)	10.2286 (60.0%)	6.7742 (79.0%)	0.3821 (92.9%)
1000,100	0.6124 (87.4%)	0.1612 (97.7%)	0 (100%)	3.8364 (82.4%)	0.2793 (96.5%)	0 (100%)	9.9668 (67.8%)	4.4129 (85.5%)	1.4900 (98.8%)

[a] The number of factors used here is $\tilde{r}=2 \times r - 1$.
[b] The number in parentheses is the percentage of obtaining correct change point.

Table 3.11: RMSE using Method 3 under DGP2

	k_0=T/4			k_0=T/3			k_0=T/2		
N,T	r=1[a]	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	14.2486 (22.0%)[b]	9.0370 (40.4%)	7.9920 (46.4%)	12.5142 (24.9%)	10.3433 (38.2%)	9.6235 (42.7%)	8.4234 (37.1%)	7.5800 (49.2%)	7.9876 (51.3%)
30,50	12.3719 (27.2%)	8.5705 (44.9%)	4.5309 (60.0%)	11.3619 (32.7%)	8.2667 (44.2%)	8.3288 (51.3%)	7.2894 (44.0%)	6.6043 (55.4%)	5.1347 (60.2%)
40,50	12.7169 (31.3%)	8.0969 (52.1%)	2.9375 (67.3%)	10.9142 (35.4%)	7.5721 (51.0%)	6.0412 (56.7%)	5.7792 (49.5%)	5.2931 (64.4%)	4.8266 (70.2%)
50,50	11.3082 (33.5%)	7.2893 (53.7%)	2.8217 (71.5%)	9.4785 (38.9%)	7.6120 (53.0%)	4.9286 (63.4%)	6.4806 (52.1%)	4.7701 (69.6%)	3.8130 (75.1%)
100,50	10.1678 (43.6%)	5.9319 (61.6%)	1.4805 (83.9%)	8.9208 (45.5%)	6.4198 (62.8%)	3.4544 (77.7%)	4.7538 (61.3%)	3.2717 (77.8%)	1.6205 (87.7%)
500,50	6.4380 (62.8%)	3.1716 (83.7%)	0.1140 (98.7%)	6.2742 (62.7%)	3.6340 (83.5%)	0.2470 (97.4%)	3.3362 (80.3%)	1.6377 (92.5%)	0.1095 (98.8%)
100,50	5.8931 (71.5%)	2.2895 (89.4%)	0.0316 (99.9%)	5.8350 (68.0%)	1.5553 (87.9%)	0.0837 (99.3%)	1.4728 (83.6%)	0.7810 (95.3%)	0.0447 (99.8%)
20,100	24.6525 (24.9%)	11.5285 (45.8%)	2.4170 (63.0%)	22.5809 (26.6%)	13.8126 (42.4%)	6.4515 (51.2%)	14.1234 (38.2%)	8.4881 (53.1%)	5.4439 (62.5%)
30,100	23.2325 (30.4%)	9.3630 (51.1%)	1.2083 (69.3%)	20.3752 (32.4%)	10.9027 (49.8%)	3.9028 (58.5%)	11.2422 (43.5%)	7.8386 (62.4%)	2.9918 (71.4%)
40,100	21.1033 (30.7%)	9.2124 (54.8%)	1.2446 (69.8%)	20.8719 (32.1%)	11.9624 (49.9%)	3.5619 (63.4%)	10.7869 (49.7%)	6.3598 (67.8%)	1.9290 (78.9%)
50,100	22.0235 (34.7%)	8.8353 (54.6%)	1.1032 (73.7%)	18.5093 (37.2%)	9.8559 (53.7%)	3.7362 (63.8%)	10.0303 (56.2%)	4.8815 (73.4%)	1.6432 (83.3%)
100,100	17.2568 (42.2%)	7.3403 (63.3%)	0.6745 (83.2%)	18.0897 (42.2%)	10.8116 (60.9%)	1.2751 (70.7%)	6.2267 (64.6%)	3.5188 (80.0%)	0.4195 (91.3%)
500,100	10.5723 (62.2%)	3.5300 (84.6%)	0.2025 (98.0%)	11.3096 (61.8%)	5.9641 (78.9%)	0.3808 (95.2%)	2.9256 (80.3%)	1.5395 (93.5%)	0.0837 (99.3%)
1000,100	9.3535 (68.2%)	1.2562 (87.6%)	0.0894 (99.5%)	9.2396 (66.8%)	3.5436 (86.7%)	0.1265 (98.7%)	2.5213 (85.9%)	0.4648 (96.3%)	0.0316 (99.9%)

[a]The number of factors used here is $\tilde{r}=2 \times r - 1$.
[b]The number in parentheses is the percentage of obtaining correct change point.

Table 3.12: RMSE using Method 3 under DGP3

N,T	k_0=T/4			k_0=T/3			k_0=T/2		
	r=1[a]	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	9.6284 (36.3%)[b]	4.0715 (62.5%)	2.4454 (77.2%)	10.0824 (35.9%)	6.8906 (53.4%)	4.7486 (64.5%)	9.6749 (35.7%)	9.6209 (43.5%)	9.9780 (44.7%)
30,50	7.2813 (44.9%)	3.5382 (72.0%)	0.7266 (84.7%)	8.0922 (40.4%)	5.0391 (64.9%)	2.7199 (74.7%)	9.8117 (35.6%)	7.9611 (48.3%)	8.4342 (51.7%)
40,50	6.9207 (48.8%)	3.1536 (77.6%)	0.4868 (89.0%)	7.2272 (48.1%)	4.2866 (69.4%)	1.5751 (82.6%)	8.7924 (42.0%)	7.6581 (55.4%)	6.9554 (57.8%)
50,50	5.2598 (52.4%)	2.2170 (79.2%)	0.4416 (91.9%)	6.6067 (53.1%)	4.1196 (73.6%)	1.3719 (83.4%)	8.5243 (42.3%)	6.9738 (55.8%)	6.5217 (61.4%)
100,50	3.5377 (67.6%)	1.1269 (89.9%)	0.1703 (98.0%)	4.6789 (60.3%)	3.5575 (82.4%)	0.3688 (93.4%)	7.4223 (48.2%)	5.5766 (67.6%)	3.5130 (79.7%)
500,50	0.9381 (83.6%)	0.1844 (96.9%)	0 (100%)	2.2152 (79.7%)	1.3153 (93.9%)	0.0316 (99.9%)	5.2268 (68.1%)	2.5148 (86.5%)	0.2000 (98.4%)
100,50	0.6964 (86.1%)	0.1817 (99.0%)	0 (100%)	1.8960 (86.7%)	0.1761 (97.2%)	0 (100%)	4.0924 (72.9%)	1.4734 (90.8%)	0.0707 (99.5%)
20,100	13.4052 (42.1%)	4.1747 (68.0%)	0.7596 (83.2%)	15.0674 (41.6%)	7.7683 (61.9%)	1.2506 (71.1%)	18.3794 (34.2%)	13.0522 (46.3%)	11.8084 (53.9%)
30,100	9.8044 (49.9%)	2.5745 (76.9%)	0.4382 (89.3%)	12.5875 (44.5%)	5.4330 (69.6%)	2.1700 (81.5%)	17.8341 (36.2%)	12.1847 (50.6%)	9.6973 (57.0%)
40,100	6.7001 (55.4%)	2.4228 (82.7%)	0.3082 (93.9%)	10.8954 (53.5%)	5.3261 (78.2%)	0.5745 (87.2%)	16.5496 (39.8%)	11.6410 (56.2%)	7.2908 (67.6%)
50,100	9.2614 (58.5%)	0.6309 (85.3%)	0.2214 (96.3%)	8.0862 (57.8%)	6.0548 (79.0%)	0.4359 (90.0%)	16.9919 (39.0%)	10.8285 (60.2%)	5.0019 (70.1%)
100,100	2.7927 (67.2%)	0.3578 (90.6%)	0.1483 (98.9%)	5.6963 (65.0%)	0.7430 (87.3%)	0.2345 (95.6%)	14.9345 (47.2%)	8.1592 (69.9%)	3.8356 (79.4%)
500,100	0.6173 (84.1%)	0.1581 (97.5%)	0.0316 (99.9%)	0.8678 (81.5%)	0.2168 (96.2%)	0.0707 (99.5%)	9.3640 (65.9%)	3.6082 (85.3%)	0.1871 (97.1%)
1000,100	0.4461 (86.9%)	0.1378 (98.7%)	0 (100%)	0.5874 (86.6%)	0.1949 (97.3%)	0.0316 (99.9%)	9.4411 (69.1%)	3.6696 (89.5%)	0.1095 (98.8%)

[a]The number of factors used here is $\tilde{r}$=2 $\times r - 1$.
[b]The number in parentheses is the percentage of obtaining correct change point.

Table 3.13: RMSE using Method 3 under DGP4

	k_0=T/4			k_0=T/3			k_0=T/2		
N,T	r=1[a]	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	18.2826 (11.6%)[b]	9.5616 (34.7%)	7.8022 (49.2%)	16.7801 (10.7%)	12.3671 (28.2%)	11.1028 (34.2%)	15.9283 (6.5%)	14.4722 (17.8%)	15.7963 (17.3%)
30,50	19.4264 (9.2%)	8.7734 (43.1%)	4.0851 (61.9%)	16.0495 (10.8%)	12.0868 (28.9%)	8.4562 (44.0%)	15.4589 (8.2%)	15.2816 (15.1%)	15.2703 (16.5%)
40,50	17.5806 (11.5%)	6.9478 (46.6%)	2.8541 (69.7%)	16.5964 (10.5%)	11.9985 (33.8%)	7.8511 (48.6%)	15.7423 (8.9%)	15.6020 (14.5%)	16.2135 (16.2%)
50,50	17.9982 (9.6%)	6.7668 (46.7%)	1.3308 (74.9%)	17.0357 (7.6%)	12.0050 (29.3%)	7.6692 (50.8%)	16.3844 (6.1%)	16.0074 (11.4%)	16.4037 (15.9%)
100,50	18.2957 (7.9%)	5.4318 (53.2%)	0.6340 (84.8%)	17.0258 (8.6%)	11.5377 (32.8%)	4.6739 (63.7%)	16.7678 (7.3%)	16.8366 (8.4%)	17.3248 (14.4%)
500,50	18.8536 (7.9%)	5.2149 (56.2%)	0.2864 (94.5%)	17.5627 (5.7%)	11.3058 (31.4%)	1.9277 (80.2%)	17.2141 (3.7%)	17.6929 (8.3%)	18.7388 (10.9%)
100,50	18.6644 (7.8%)	6.3185 (52.7%)	0.1817 (97.0%)	17.6529 (7.3%)	11.9099 (32.2%)	1.5840 (84.1%)	17.1359 (4.3%)	17.5394 (7.2%)	18.6656 (11.6%)
20,100	32.3348 (13.5%)	8.0884 (44.8%)	2.3390 (63.6%)	32.3691 (8.4%)	16.5090 (34.5%)	7.9498 (50.0%)	31.3866 (7.9%)	27.8885 (18.2%)	25.1413 (25.1%)
30,100	33.8909 (13.1%)	5.0282 (49.6%)	0.9920 (77.6%)	33.2995 (10.5%)	15.5343 (37.5%)	3.6150 (59.5%)	32.8134 (6.4%)	29.3564 (16.4%)	25.1450 (24.5%)
40,100	33.9013 (10.8%)	4.1609 (55.6%)	0.6156 (82.7%)	32.0494 (10.3%)	15.6424 (38.6%)	3.6985 (63.8%)	32.9190 (6.0%)	30.2861 (14.9%)	25.8062 (22.8%)
50,100	32.9360 (12.2%)	6.1856 (53.5%)	0.5727 (86.3%)	32.8115 (7.6%)	14.7054 (40.3%)	1.6134 (66.5%)	33.4753 (4.4%)	31.1493 (11.9%)	27.1053 (20.9%)
100,100	33.8450 (8.7%)	3.2746 (60.5%)	0.3808 (91.4%)	33.1677 (7.2%)	13.0300 (42.8%)	1.0159 (75.7%)	34.8696 (3.4%)	34.8066 (9.1%)	31.3936 (16.1%)
500,100	33.3804 (7.7%)	3.7232 (61.2%)	0.1949 (97.3%)	33.5857 (5.2%)	14.2954 (40.5%)	0.5206 (86.6%)	35.0513 (3.6%)	36.4184 (6.2%)	35.6213 (11.2%)
1000,100	32.7289 (8.0%)	1.8426 (63.4%)	0.1265 (98.4%)	33.9725 (4.9%)	13.5818 (42.9%)	0.5070 (88.3%)	35.9275 (1.5%)	36.4472 (5.7%)	38.0117 (11.6%)

[a] The number of factors used here is $\tilde{r}=2 \times r - 1$.
[b] The number in parentheses is the percentage of obtaining correct change point.

3.3.1.4 Baltagi, Kao and Wang (2015) Method

Baltagi, Kao and Wang (2015)'s approach is denoted as **Method 4**. Let

$$\hat{\Sigma}_1 = \frac{1}{k}\sum_{t=1}^{k}\hat{F}_t\hat{F}_t', \qquad \hat{\Sigma}_2 = \frac{1}{T-k}\sum_{t=k+1}^{T}\hat{F}_t\hat{F}_t'.$$

Then define the sum of squared residuals as

$$S(k) = \sum_{t=1}^{k}[vec(\hat{F}_t\hat{F}_t' - \hat{\Sigma}_1)]'[vec(\hat{F}_t\hat{F}_t' - \hat{\Sigma}_1)] + \sum_{t=k+1}^{T}[vec(\hat{F}_t\hat{F}_t' - \hat{\Sigma}_2)]'[vec(\hat{F}_t\hat{F}_t' - \hat{\Sigma}_2)],$$

and the least squares of the change point as

$$\hat{k} = \arg\min_{1\leq k\leq T} S(k)$$

Baltagi, Kao and Wang (2015) proved that estimating the break point in factor loadings can be equivalently represented by estimating break point in the second moment of estimated factors. However, this method doesn't take the full advantage of panel data. It doesn't use the information of cross-section units.

Tables 3.14-3.17 show the results for Method 4. In the "increasing r" case, Method 4 performs good, but Method 1 remains more accurate. For $r = 1$ and $r = 2$ cases, it doesn't perform very well. In some cases, the accuracy starts to decrease when N is greater than T.

Table 3.14: RMSE using Method 4 under DGP1

	k_0=T/4			k_0=T/3			k_0=T/2		
N,T	r=1[a]	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	16.6538 (17.9%)[b]	9.5244 (40.2%)	8.9914 (56.1%)	13.6019 (18.9%)	9.1955 (43.0%)	6.7082 (63.0%)	11.4626 (19.4%)	7.6324 (42.1%)	5.6557 (60.5%)
30,50	16.0644 (20.2%)	10.4491 (45.4%)	7.9885 (62.0%)	13.8062 (22.5%)	8.8673 (43.9%)	5.6927 (69.9%)	11.2479 (20.0%)	7.1175 (46.9%)	4.6471 (71.2%)
40,50	16.4514 (22.8%)	9.6009 (47.5%)	6.0247 (67.8%)	13.6644 (22.3%)	7.6766 (50.1%)	4.5936 (74.3%)	11.1297 (18.5%)	6.7324 (51.0%)	4.8930 (72.3%)
50,50	15.7013 (21.5%)	9.2283 (49.3%)	6.9202 (70.6%)	12.5701 (21.8%)	8.2557 (50.8%)	5.5457 (75.9%)	10.7912 (22.2%)	6.3955 (54.8%)	4.3480 (78.0%)
100,50	14.8845 (24.7%)	3.9477 (36.7%)	1.7830 (72.2%)	13.3894 (23.4%)	4.9809 (42.2%)	1.8855 (77.9%)	10.0868 (22.5%)	6.6225 (61.5%)	0.9006 (98.7%)
500,50	14.9480 (26.0%)	3.7107 (42.1%)	1.3031 (78.2%)	12.4631 (27.8%)	4.7080 (45.0%)	0.8826 (89.1%)	9.5254 (27.4%)	4.9900 (68.5%)	0 (100%)
1000,50	14.2557 (23.0%)	3.7049 (39.4%)	1.0144 (82.2%)	12.2068 (27.4%)	5.0758 (44.9%)	0.9203 (90.8%)	9.2395 (27.9%)	5.4446 (66.4%)	0 (100%)
20,100	25.6154 (20.9%)	12.4150 (51.4%)	6.6784 (76.1%)	22.3058 (19.5%)	9.2825 (52.8%)	5.3756 (76.6%)	20.2934 (19.0%)	7.6299 (52.6%)	3.0232 (78.9%)
30,100	26.0419 (23.3%)	11.5932 (56.1%)	5.5006 (84.0%)	21.1578 (24.1%)	10.6804 (54.8%)	3.0525 (84.6%)	20.3851 (17.7%)	8.2056 (55.4%)	2.8525 (84.8%)
40,100	24.0056 (23.6%)	10.0105 (60.3%)	4.7558 (87.1%)	21.8544 (25.0%)	8.6259 (58.7%)	3.7417 (89.3%)	18.3236 (23.1%)	8.3922 (56.2%)	2.2658 (88.6%)
50,100	23.5501 (24.2%)	10.2781 (61.9%)	2.7092 (89.7%)	20.5221 (22.8%)	6.0217 (63.5%)	2.6079 (91.6%)	18.6811 (21.9%)	6.5464 (62.5%)	2.2825 (91.9%)
100,100	22.4109 (26.4%)	9.9549 (65.8%)	1.0464 (94.0%)	19.2773 (27.9%)	7.5248 (67.9%)	0.4950 (93.4%)	16.8994 (23.4%)	6.3520 (64.5%)	1.4663 (94.5%)
500,100	20.3201 (29.2%)	6.3490 (41.2%)	0.7127 (86.4%)	17.6610 (29.0%)	5.2020 (48.9%)	0.4701 (89.4%)	16.1129 (24.8%)	4.7230 (71.4%)	0.0316 (99.9%)
1000,100	21.8523 (27.8%)	5.5923 (42.6%)	0.5666 (87.4%)	17.2761 (28.2%)	5.8756 (48.3%)	0.5753 (91.2%)	15.2762 (25.9%)	5.6996 (72.8%)	0 (100%)

[a]The number of factors used here is $\tilde{r}$=$2 \times r - 1$.
[b]The number in parentheses is the percentage of obtaining correct change point.

Table 3.15: RMSE using Method 4 under DGP2

	k_0=T/4			k_0=T/3			k_0=T/2		
N,T	r=1[a]	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	15.8755 (16.1%)[b]	7.1444 (38.1%)	4.3772 (45.0%)	13.0907 (18.4%)	6.3962 (33.1%)	4.9410 (46.9%)	10.4930 (21.0%)	8.2252 (40.8%)	6.5004 (48.0%)
30,50	15.7149 (17.3%)	5.6165 (43.6%)	3.4861 (52.3%)	12.2920 (20.0%)	6.6273 (38.3%)	4.1473 (48.9%)	10.2248 (24.8%)	7.3903 (40.7%)	6.1815 (53.4%)
40,50	15.2115 (18.4%)	5.8341 (43.1%)	3.1464 (52.7%)	12.3930 (19.9%)	5.9000 (41.2%)	4.1815 (48.4%)	10.0161 (24.3%)	6.9871 (45.5%)	5.7846 (56.1%)
50,50	15.7075 (18.7%)	5.9194 (40.9%)	3.3064 (53.5%)	12.1412 (21.3%)	5.5127 (40.7%)	3.6021 (54.2%)	10.1770 (26.6%)	7.2213 (44.6%)	5.7628 (60.1%)
100,50	15.2227 (21.1%)	4.0153 (39.5%)	1.9728 (69.1%)	11.7997 (24.0%)	4.8350 (43.3%)	1.8679 (79.1%)	9.5097 (27.6%)	5.4709 (64.6%)	0.9094 (98.0%)
500,50	14.2969 (25.0%)	3.9708 (36.7%)	1.2260 (78.8%)	11.0172 (25.3%)	5.0669 (44.6%)	0.9975 (88.0%)	9.3318 (27.7%)	5.1761 (65.5%)	0.0632 (99.6%)
1000,50	13.5037 (25.8%)	3.9628 (39.4%)	1.0445 (80.3%)	11.0583 (24.2%)	4.3295 (48.7%)	0.7308 (92.4%)	9.1777 (27.9%)	5.0417 (69.4%)	0 (100%)
20,100	25.4759 (16.7%)	8.7988 (40.2%)	4.5711 (53.4%)	22.4887 (15.9%)	9.4975 (39.8%)	6.1682 (55.4%)	17.5816 (23.3%)	9.5545 (44.2%)	5.8770 (59.2%)
30,100	26.7876 (18.3%)	8.6697 (38.7%)	3.9575 (60.1%)	21.7310 (19.3%)	7.6630 (43.7%)	5.5464 (60.2%)	16.4013 (25.3%)	8.7602 (46.4%)	5.0791 (64.4%)
40,100	26.4439 (20.1%)	7.0726 (46.5%)	3.7971 (60.8%)	21.3330 (18.8%)	6.8322 (46.3%)	4.0077 (64.2%)	14.5819 (22.9%)	7.0893 (49.0%)	5.3258 (66.1%)
50,100	27.2739 (18.0%)	7.5304 (43.3%)	3.3759 (62.3%)	21.4643 (19.9%)	6.4423 (48.3%)	3.9081 (65.5%)	15.9192 (25.6%)	7.7296 (51.0%)	3.8445 (67.9%)
100,100	24.5661 (23.2%)	6.0255 (44.9%)	3.0699 (65.2%)	20.4334 (22.5%)	7.3890 (45.4%)	3.2965 (67.9%)	12.7932 (27.9%)	7.4533 (53.7%)	2.9519 (73.2%)
500,100	22.4228 (27.1%)	5.9706 (40.8%)	0.9214 (82.5%)	15.6208 (25.7%)	6.6433 (46.5%)	0.6863 (86.9%)	12.8879 (31.0%)	4.4567 (71.2%)	0 (100%)
1000,100	21.1945 (26.0%)	5.9212 (41.2%)	0.8994 (85.6%)	17.6494 (27.1%)	5.7527 (51.5%)	0.3886 (92.7%)	12.3173 (32.1%)	4.1657 (73.8%)	0 (100%)

[a] The number of factors used here is $\tilde{r}$=2 $\times r - 1$.
[b] The number in parentheses is the percentage of obtaining correct change point.

Table 3.16: RMSE using Method 4 under DGP3

	k_0=T/4			k_0=T/3			k_0=T/2		
N,T	r=1[a]	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	16.1276 (19.5%)[b]	11.8590 (41.4%)	10.6655 (51.4%)	13.9640 (18.2%)	10.2228 (44.3%)	7.7475 (57.4%)	10.8707 (18.9%)	7.8535 (43.5%)	5.2604 (59.5%)
30,50	15.3358 (23.2%)	11.7238 (46.4%)	9.4865 (59.4%)	13.8610 (20.1%)	9.5450 (44.3%)	7.2110 (61.3%)	10.5985 (20.4%)	6.3276 (49.0%)	5.1806 (63.0%)
40,50	15.5301 (24.3%)	10.7328 (46.3%)	8.4241 (66.7%)	13.6016 (23.2%)	9.4023 (48.6%)	7.3202 (68.0%)	11.0548 (20.1%)	7.2247 (48.6%)	4.7835 (67.0%)
50,50	14.4117 (21.9%)	11.3663 (46.2%)	8.9140 (65.1%)	13.4694 (24.7%)	8.1288 (50.1%)	6.7415 (69.9%)	10.3095 (22.0%)	7.0730 (51.9%)	4.5383 (70.1%)
100,50	15.7882 (21.6%)	3.8865 (39.1%)	1.8174 (70.3%)	13.4989 (23.3%)	4.7587 (45.9%)	1.6483 (78.1%)	9.8088 (21.9%)	6.1173 (62.4%)	0.6535 (98.1%)
500,50	14.6752 (24.6%)	4.1270 (41.0%)	1.0950 (82.7%)	12.4298 (25.7%)	4.6555 (46.2%)	1.0555 (89.6%)	9.4597 (26.7%)	5.1346 (65.6%)	0.0447 (99.8%)
1000,50	14.6498 (25.0%)	3.6748 (42.3%)	1.0198 (81.1%)	12.1620 (28.5%)	4.7050 (45.1%)	0.6221 (91.2%)	9.2061 (29.2%)	5.4447 (64.7%)	0 (100%)
20,100	25.0800 (22.4%)	13.2590 (49.3%)	10.3406 (69.5%)	23.6131 (20.8%)	11.3191 (53.2%)	6.3690 (68.9%)	19.7643 (18.8%)	9.0838 (50.7%)	5.7552 (68.2%)
30,100	23.4781 (24.7%)	11.6972 (54.8%)	8.6980 (74.3%)	21.9618 (22.5%)	11.8986 (54.6%)	5.4806 (76.7%)	18.7886 (20.1%)	8.3653 (53.1%)	3.6927 (77.7%)
40,100	21.7668 (25.6%)	13.3928 (56.6%)	8.4190 (79.9%)	19.8087 (26.0%)	9.7633 (59.1%)	4.9845 (80.3%)	18.6379 (22.0%)	7.8645 (57.8%)	3.7858 (78.1%)
50,100	22.8922 (27.5%)	10.6944 (60.3%)	7.7473 (79.6%)	19.6209 (23.0%)	8.8468 (58.9%)	5.8354 (80.9%)	17.8228 (21.8%)	7.8388 (57.2%)	3.9939 (78.4%)
100,100	19.1721 (30.6%)	11.4006 (61.5%)	4.9668 (88.5%)	18.1592 (28.3%)	6.6280 (63.6%)	4.7077 (85.9%)	17.5517 (26.8%)	7.1358 (58.5%)	3.5855 (84.1%)
500,100	22.2203 (27.3%)	6.3618 (39.1%)	0.6197 (86.0%)	16.9727 (32.8%)	5.3779 (53.5%)	0.4494 (89.7%)	15.3953 (26.1%)	3.6050 (73.3%)	0.0316 (99.9%)
1000,100	21.5961 (27.6%)	6.1446 (38.0%)	0.6017 (84.4%)	18.2728 (30.3%)	5.8868 (50.0%)	0.3821 (91.0%)	14.4668 (27.8%)	3.8586 (73.3%)	0.0316 (99.9%)

[a]The number of factors used here is $\tilde{r}=2 \times r - 1$.
[b]The number in parentheses is the percentage of obtaining correct change point.

Table 3.17: RMSE using Method 4 under DGP4

	k_0=T/4			k_0=T/3			k_0=T/2		
N,T	r=1[a]	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	18.2890 (10.6%)[b]	13.7889 (26.4%)	11.3045 (43.1%)	16.7752 (10.2%)	11.3008 (31.8%)	8.6306 (46.9%)	13.9177 (8.3%)	9.3345 (31.9%)	7.7487 (47.7%)
30,50	19.2583 (8.8%)	12.4895 (31.1%)	10.1406 (51.4%)	16.7968 (8.4%)	10.3227 (33.6%)	8.1551 (56.4%)	14.1140 (7.9%)	9.1786 (34.7%)	6.0382 (57.4%)
40,50	19.1824 (9.3%)	12.3246 (35.6%)	8.4449 (53.7%)	16.4001 (8.1%)	10.0678 (38.2%)	6.9494 (58.6%)	14.2738 (7.9%)	8.7878 (38.7%)	5.9646 (63.4%)
50,50	19.5346 (9.9%)	12.2678 (35.2%)	9.0816 (58.4%)	16.2826 (8.4%)	10.0862 (37.0%)	6.4424 (65.2%)	14.0441 (7.4%)	8.5213 (39.1%)	5.5366 (67.6%)
100,50	18.4691 (7.8%)	4.0255 (34.4%)	1.9748 (68.5%)	16.7132 (8.9%)	4.8152 (40.9%)	1.9178 (77.8%)	14.3141 (5.8%)	6.7197 (57.3%)	0.1183 (98.6%)
500,50	19.1694 (7.9%)	3.9578 (38.7%)	1.2872 (78.1%)	17.4623 (7.2%)	5.0661 (39.6%)	0.9879 (88.6%)	14.1903 (6.4%)	6.6675 (58.7%)	0 (100%)
1000,50	19.3051 (7.3%)	3.9074 (39.4%)	1.1162 (80.9%)	17.0978 (7.5%)	5.3165 (39.6%)	0.7720 (91.3%)	14.4117 (5.5%)	5.9952 (60.9%)	0 (100%)
20,100	35.7306 (8.7%)	17.8649 (39.8%)	11.9367 (65.6%)	31.5184 (9.4%)	13.6511 (40.1%)	6.6654 (66.9%)	27.6290 (7.9%)	12.6067 (39.2%)	5.7774 (68.8%)
30,100	35.7205 (10.2%)	15.1949 (41.7%)	10.0643 (72.8%)	32.6917 (6.9%)	12.7040 (44.4%)	6.0155 (73.9%)	28.2861 (5.8%)	11.0642 (42.8%)	4.5223 (77.0%)
40,100	35.3307 (8.9%)	17.3808 (44.2%)	6.2227 (77.5%)	31.3847 (9.4%)	11.8578 (43.8%)	4.4170 (80.1%)	28.3528 (5.2%)	13.0323 (44.9%)	3.3929 (82.5%)
50,100	34.3381 (10.5%)	15.1359 (46.2%)	7.6988 (80.7%)	32.3768 (7.3%)	13.1943 (45.9%)	4.7046 (81.9%)	29.2873 (3.7%)	11.4764 (46.2%)	3.0700 (85.4%)
100,100	35.3249 (10.2%)	15.0068 (50.1%)	4.6823 (87.8%)	31.9740 (6.7%)	11.3666 (49.0%)	3.6019 (88.5%)	30.1443 (4.0%)	11.2368 (52.6%)	2.1000 (91.4%)
500,100	34.4176 (9.5%)	6.5851 (38.1%)	1.0909 (80.3%)	32.7654 (5.3%)	6.5210 (43.2%)	0.5595 (88.7%)	30.1733 (3.7%)	4.4395 (70.4%)	0 (100%)
1000,100	35.7563 (8.6%)	6.6246 (35.9%)	0.7570 (86.7%)	32.5690 (6.3%)	7.0556 (40.2%)	0.3674 (92.5%)	30.4479 (2.9%)	5.4832 (69.5%)	0 (100%)

[a] The number of factors used here is $\tilde{r}$=$2 \times r - 1$.
[b] The number in parentheses is the percentage of obtaining correct change point.

3.3.1.5 MLE Method

We denote MLE Method as **Method 5**. Define

$$\hat{\Omega} = E\left((\hat{F} - E(\hat{F}))'(\hat{F} - E(\hat{F}))\right),$$

so $\hat{\Omega}$ is the covariance matrix of estimated factors $\hat{F}_t$. Let $\hat{\Omega}^*$ be the pre-break covariance matrix of $\hat{F}_t$ for $1 \leq t \leq k$. $\hat{\Omega}^{**}$ be the post-break covariance matrix of $\hat{F}_t$ for $k+1 \leq t \leq T$. The objective function becomes

$$MLE(k) = k\sum_{t=1}^{k} \log\left|\hat{\Omega}^*\right| + (T-k)\sum_{t=k+1}^{T} \log\left|\hat{\Omega}_t^{**}\right|$$

where $|A|$ denote the determinant of matrix A. The estimated break point is defined as

$$\hat{k} = \arg\min_{1\leq k\leq T}\ MLE(k).$$

Table 3.18 shows the results for Method 5 under DGP1. For $r = 1$ case, the results are good but still weaker than Method 1. When r increases, the accuracy of estimation declines quickly. The results are not good in MLE method, so we only compute the DGP1 case.

Table 3.18: RMSE using Method 5 under DGP1

	k_0=T/4			k_0=T/3			k_0=T/2		
N,T	r=1[a]	r=2	increasing r	r=1	r=2	increasing r	r=1	r=2	increasing r
20,50	11.5970 (31.0%)[b]	30.0204 (1.4%)	32.7220 (0%)	10.4161 (32.4%)	26.3686 (1.5%)	28.5922 (0%)	10.2591 (33.9%)	23.4654 (2.3%)	24.0000 (0%)
30,50	9.2533 (39.0%)	30.2398 (1.6%)	33.2132 (0%)	9.6325 (37.4%)	26.4777 (3.7%)	28.4722 (0%)	10.1545 (34.4%)	23.0922 (5.5%)	24.0000 (0%)
40,50	8.4973 (46.8%)	29.5957 (6.0%)	33.2704 (0%)	8.5222 (43.8%)	26.2282 (7.2%)	28.6978 (0%)	9.0897 (37.9%)	22.9021 (7.3%)	24.0000 (0%)
50,50	7.3651 (50.2%)	29.6037 (8.6%)	33.8837 (0%)	7.7756 (46.5%)	26.0386 (8.9%)	28.7161 (0%)	8.9520 (41.2%)	22.8660 (8.3%)	23.9972 (0%)
100,50	5.3281 (61.8%)	26.5543 (12.3%)	32.7220 (0%)	7.4223 (56.3%)	24.1728 (11.2%)	27.8299 (0%)	8.2394 (46.2%)	22.4364 (11.6%)	24.0000 (0%)
500,50	1.5624 (81.6%)	27.0304 (12.8%)	34.5945 (0%)	3.8296 (76.7%)	24.0028 (12.8%)	29.3264 (0%)	4.5571 (67.1%)	22.2955 (13.5%)	24.0000 (0%)
1000,50	1.7076 (84.9%)	27.4831 (12.0%)	35.0959 (0%)	1.8138 (81.4%)	24.8079 (13.0%)	29.9820 (0%)	4.6973 (70.3%)	22.5524 (11.6%)	24.0000 (0%)
20,100	16.9526 (37.5%)	58.4697 (8.0%)	66.8326 (0%)	16.6970 (34.3%)	51.2664 (9.6%)	57.9632 (0%)	19.3392 (31.0%)	45.9897 (8.9%)	49.0000 (0%)
30,100	11.9756 (44.3%)	58.4637 (11.0%)	67.0887 (0%)	16.0989 (40.3%)	51.3654 (9.9%)	58.0781 (0%)	17.4761 (31.7%)	45.7309 (10.1%)	49.0000 (0%)
40,100	9.1974 (51.5%)	58.5161 (9.7%)	66.6858 (0%)	14.3300 (42.9%)	50.7987 (12.5%)	58.1354 (0%)	17.9936 (36.7%)	46.1229 (9.8%)	49.0000 (0%)
50,100	10.2594 (54.4%)	57.9495 (13.7%)	68.3550 (0%)	12.1038 (45.2%)	52.7344 (11.0%)	57.9632 (0%)	16.3098 (39.5%)	45.9663 (10.3%)	49.0000 (0%)
100,100	4.9605 (66.4%)	60.2201 (11.3%)	68.2832 (0%)	7.7471 (60.7%)	51.9753 (10.9%)	58.1354 (0%)	15.6878 (44.1%)	45.0962 (14.2%)	49.0000 (0%)
500,100	1.0325 (82.3%)	53.9489 (13.8%)	64.9361 (0%)	4.7111 (78.0%)	48.6173 (14.0%)	56.8898 (0%)	11.2162 (56.2%)	45.9139 (11.6%)	49.0000 (0%)
1000,100	0.8420 (85.1%)	53.6326 (11.4%)	66.3174 (0%)	3.9305 (82.4%)	48.8905 (12.8%)	57.1819 (0%)	9.9331 (68.1%)	45.2289 (14.4%)	49.0000 (0%)

[a] The number of factors used here is $\tilde{r}=2 \times r - 1$.
[b] The number in parentheses is the percentage of obtaining correct change point.

3.3.2 Bootstrap Method

Because of the complexity of our limiting distribution, in this section we use the bootstrap method to verify the performance of our limiting distribution. T is fixed as 50. We let N = 10, 20, 30, 40, and 50. The true number of factors is fixed as 2. And the break should happen at $T/2$. The DGP is similar as in previous simulations. We let $f_t \sim$ *i.i.d.* $N(0,1)$ and $e_{it} \sim$ *i.i.d.* $N(0,1)$. Before the break point, we let $\lambda_{i1} \sim i.i.d.N(0,1)$. For the after break situation, we consider two cases. In Case 1, we let $\lambda_{i2} = \lambda_{i1} + 0.3 * N(0,1)$. This case assumes the break size is small. It is denoted as **DGP B1**. In Case 2, denoted as **DGP B2**, we let post-break factor loadings $\lambda_{i2} = \lambda_{i1} + \frac{2}{\sqrt{N}} * N(0,1)$. When N increases, the break size decreases. This setup is consistent with our small break assumptions.

The following steps explains how we implement the bootstrap method. First, for given simulated data, we estimate the change point $\hat{k}$ by using the least squares method. This estimated change point allows us to obtain estimated errors, estimated factors, estimated pre-break factor loadings, and estimated post-break factor loadings. Second, we treat the estimated factors and estimated factor loadings as real data and construct the bootstrap sample based on the estimated errors. To maintain the serial dependence of the error term, we randomly draw the whole column (T by 1) from estimated errors by N times. The new T by N matrix is our bootstrap sample. Using the estimated factors, estimated factor loadings and the bootstrap sample errors, we obtain a new estimated change point $\hat{k}_{b1}$. Third, by replicating step 2 by 1,000 times, we obtain 1,000 estimated change points. Sorting these 1,000 values from the smallest to the largest, the 25th value and 975th value represent the lower bound and upper bound of the 95% confidence interval. If the true break point k_0 is inside this confidence interval, then it is recorded as 1, otherwise it is 0. Finally, replicating steps 1-3 1,000 times, we obtain 1,000 values equal to 1 or 0. Then we sum these values and divide them by 1,000. This average value is our simulated coverage rate. We compare the actual coverage rates with the nominal rates (90%, 95%, 99%).

Table 3.19 presents the simulated coverage rates. When $N = 10$, the coverage rate is lower than the nominal rate. As N increases, the results significantly improved. The simulated coverage rate is very close to the nominal rate.

Table 3.19: Coverage Rate

	DGP B1			DGP B2		
N/T	90%	95%	99%	90%	95%	99%
10,50	0.7570	0.8690	0.9580	0.8350	0.9020	0.9760
20,50	0.8850	0.9470	0.9890	0.8970	0.9490	0.9900
30,50	0.9030	0.9430	0.9850	0.9190	0.9600	0.9910
40,50	0.9090	0.9570	0.9910	0.9360	0.9680	0.9950
50,50	0.9280	0.9560	0.9920	0.9260	0.9680	0.9910

3.4 Empirical Application

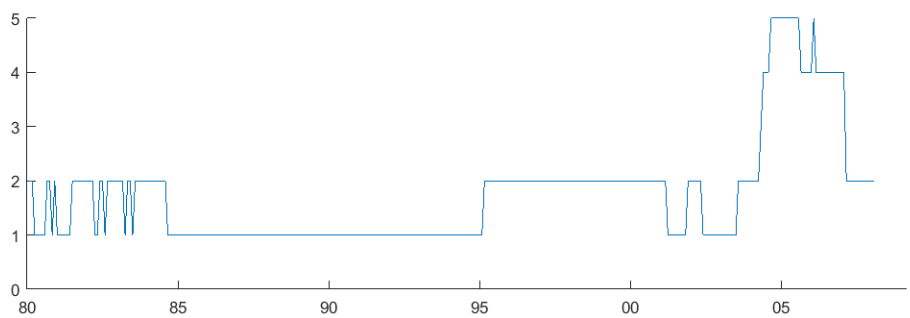
3.4.1 Financial Asset Returns

Data sets of asset returns are typically large, because thousands of companies' stocks are traded daily in the NYSE, AMEX, and NASDAQ. Factor models are statistical tools that can reduce the dimension of and identify representative factors in large data sets. Researchers can then use these representative factors to forecast companies' future returns or estimate the casual effect between dependable variables and factors. Studies of factor models for analyzing asset returns, generally can be categorized into two cases: regression analyses on observable factors, and statistical analyses of unobservable factors. Regression analyses, focus on finding proper factors to explain the dependent variables. For example, in finance literature, people like to use the S&P 500 or other financial indexes to forecast asset returns. This model can be treated as a factor model with observed factors. The financial indexes are observed factors, and the corresponding coefficients often depend on a cross section unit i. A famous study of observed factors is Fama and French's (1993) three factor model, in which the three fundamentals are the overall market return, the performance of small stocks relative to large stocks, and the performance of value stocks relative to growth stocks. Statistical analyses of unobserved factors, focuses on how to estimate the unknown factors and specify the number of factors. As shown in section 3.1, the method of principal component is commonly used to estimate the unknown factors. Bai and Ng (2002) provided six information criterion for selecting the number of factors. They used asset returns of 8,436 stocks between 1994.1 and 1998.12. After deleting all missing data and applying their information criterion, they found two factors.

In this section, we study factor models with structural changes in factor loadings by using financial asset returns data. Factor loadings are unstable and thus tend to overestimate the number of factors. Inconsistency in the estimated number of factors then leads to inconsistency in the estimated factors and estimated factor loadings. We first analyze how the number of factors changes with different sample sizes. We use monthly data for returns traded on the NYSE, AMEX, and NASDAQ between 1980.1 and 2012.12. The data include all live stocks from the first trading day of

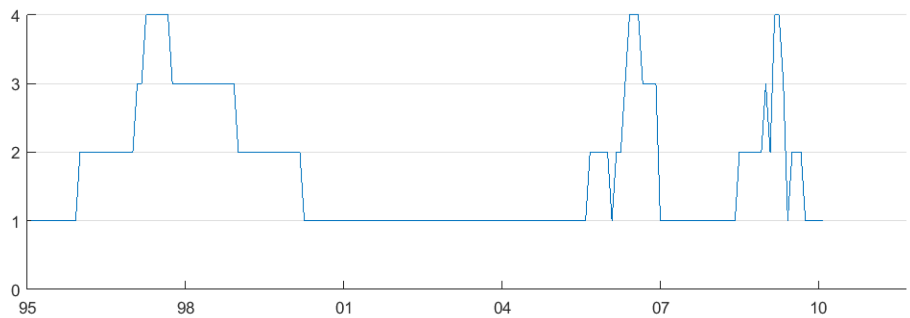
1980 to the last trading day of 2012 and are obtained from the CRSP data base. In this 80-12 monthly data, $T = 396$ and $N = 617$. Bai and Ng (2002) suggested using $PC_{p1}, PC_{p2}, IC_{p1}$, and IC_{p2} information criteria to select the number of factors. They also showed that PC_{p1} and PC_{p2} tend to overestimate the number of factors when N or T is small. Here, we use IC_{p1} information criteria for the entire sample and find **5** number of factors.

Figure 3.7: Moving Window: $T = 60$, $N = 617$, 80-12



Next, we analyze how the estimated number of factors changes using different sample size. We use samples over moving window. T is fixed as 60. We calculate the number of factors for $80.1-85.1, 80.2-85.2, 80.3-85.3, \ldots, 07.12-12.12$. Graph 3.7 report the estimated number of factors by using IC_{p1} information criteria. As shown, the entire sample's estimated number of factors (5) is the largest estimated number of factors among the sub-samples. This result is consistent with the overestimation ignoring structural changes. The shifts of estimated number of factors may due to potential structural changes. The number of factors is less or equal 2 before 2007. This result is same with Bai and Ng's (2002) finding of two factors. When we use sample size including 07-09 economic recession, the estimated number of factors increases significantly. Now we focus on how the estimated number of factors change in 95-12 data.

Figure 3.8: Moving Window: $T = 36$, $N = 1885$, 95-12



We consider only live stocks from 1995.1-2012.12. The 95-12 data set has $T = 216$ and $N = 1885$. By using the IC_{p1} information criteria, we have **4** number of factors. The entire sample's estimated number of factors still equals the largest estimated number of factors among the sub-samples. Figure 3.8 reports the estimated number of factors when using moving window sample where $T = 36$. The volatility of the estimated number of factors is significant after 2004. Finally, we focus on the data between 2005.1 and 2012.12. After deleting all missing data, the sample size is $T = 96$ and $N = 3716$. The estimated number of factors without structural changes is 2. We apply our least squares method and use $\tilde{r} - 1$ number of factors to estimate the unknown break point. Table 3.20 report the result. The estimated change point is 2009:04. This result is reasonable. In finance literature, many studies show that there is a late effect from the 2007 financial crisis. The efficient corresponding to the independent variables does not change immediately. The estimated result is consistent, because $N = 3716$ is very large. Our confidence interval is just the

Table 3.20: Estimated Change Point: Asset Returns, 05-12

$\tilde{r}$	$\hat{k}$	Length of Confidence Interval		
		90%	95%	99%
2	2009.04	[2009.04,2009.04]	[2009.04,2009.04]	[2009.04,2009.04]

estimated point.

3.4.2 Macroeconomic Data

Factor models have been widely used in macroeconomics. Stock and Watson (1998, 1999) considered forecasting inflation with diffusion indices ("factors") constructed from a large number of macroeconomic series. Gregory and Head (1999) and Forni, Hallin, Lippi, and Reichlin (2000) found common components in cross country variations. Bernanke, Boivin, and Eliasz (2005) studied the factor-augmented VAR (FAVAR) model. Ludvigson and Ng (2009) used the factor augmented regression framework to analyze the relation between excess returns and the macro economy. In this part, we study the structural changes in factor loadings using macroeconomic data. We used Ludvigson and Ng's (2013) data. It consists of a panel of 132 U.S. macroeconomic variables from 1960:1 to 2011:12[1]. Unlike the data of financial asset returns, we cannot use Bai and Ng's (2002) information criterion to specify the number of factors, because the cross section correlation is too strong among these macroeconomic variables. Ludvigson and Ng (2013) suggested using eight factors. Here, we follow their conclusion and focus on the structural changes between 2005 and 2011 ($T = 84, N = 132$). Table 3.21 shows the estimated change point and the length of confidence intervals. The estimated change point is 2008:01. The 99%

[1]The detail of this data can be found at *http://www.econ.nyu.edu/user/ludvigsons/*.

Table 3.21: Estimated Change Point: Macroeconomic Data, 05-12

		Length of Confidence Interval		
$\tilde{r}$	$\hat{k}$	90%	95%	99%
8	2008.01	[2008.01,2008.01]	[2007.11,2008.02]	[2007.05,2008,08]

confidence interval of true break point is $[\hat{k}-8, \hat{k}+8]$. Because the number of series is not too large, our estimated change point may not coincide with the true break point. However, our estimated change point is consistent with Cheng, Liao, and Schorfheide's (2015). In their application, they used an updated version of Stock and Watson's (2012) data, which included a set of 200 macroeconomic and financial indicators. Their model selection procedure provides strong evidence that the loadings in the normalized factor model changes, generally implying a stronger co-movement of the series after 2007. Thus, the estimated change point is reasonable by using this data set.

4. SUMMARY

Section 2 considers the problem of determining the number of factors in large factor models where the number of factors is allowed to increase, but with a slower rate, as N or T increases. We extend the analysis of Bai and Ng (2002) to the case that number of factors can increase with the sample size and prove the consistency of a modified Bai and Ng's (2002) procedure in determining the number of factors. We also propose a ('mode' based) new procedure so that our selected number of factors is not sensitive to the choice of k_{max}. Monte Carlo simulation results suggest that the criteria PC_{p1}, PC_{p2} and PC_{p3} all have the overall best performance. Other criteria such as IC_{p1}, IC_{p2} and PC_{p3} can also be used to accurately estimate the number of factors when the data dimensions are relatively large, say $\min\{N, T\} \geq 100$. One possible future research topic is to find alternative criteria that can improve the finite-sample performance of Bai and Ng's (2002) procedure and our modified procedure such that the new criteria can accurately determine the number of factors even in small or medium size samples.

Section 3 considers the structural change in factor loadings in high dimensional factor models. We estimate the unknown break point by using the least squares method. Several competing methods are compared in the simulation. The results show that the least squares method outperforms other approaches. We further propose a new framework to derive the limiting distribution for the estimated change point. The limiting distribution of the estimated break point is more complex than that of the conventional panel data models, because both factors and factor loadings are unobservable. We show that the estimated factors and estimated factor loadings influence the limiting distribution. The random parts depend on *i.i.d.* standard normal variables and chi square variables. Based on the limiting distribution of the estimated break point, one can construct confidence intervals of the underlying true break point. Bootstrap method is also studied. We apply the method to the study of structural changes in financial asset returns and in macroeconomic data.

REFERENCES

Anh, S. and A. Horenstein (2013), Eigenvalue Ratio Test for the Number of Factors, *Econometrica*, 81, 1203-1127.

BAI2003) Bai, J. (2003), Inferential theory for factor models of large dimensions, *Econometrica*, 71, 135-172.

Bai, J. (2010), Common breaks in means and variances for panel data, *Journal of Econometrics*, 164, 92-115.

Bai, C., Q. Li, M. Ouyang (2014), Property Taxes and Home Prices: A Tale of Two Cities, *Journal of Econometrics*, 180, 1-15.

Bai, J. and S. Ng (2002), Determining the Number of Factors in Approximate Factor Models, *Econometrica*, 70, 191-221.

Bai, J. and S. Ng (2010), Instrumental variable estimation in a data rich environment, *Econometric Theory*, 26, 1577-1606.

Banerjee, A. and M. Marcellino (2008), Forecasting Macroeconomic variables using diffusion indexes in short samples with structural change, Emerald Group Publishing Limited, Vol.3, pp. 149-194.

Baltagi, B., C. Kao, and F. Wang (2015), Identification and estimation of a large factor model with structural instability, Manuscript.

Bates, B., M. Plagborg-Moller, J. H. Stock and M. W. Watson (2013), Consistent factor estimation in dynamic factor models with structural instability, *Journal of Econometrics*, 177, 289-304.

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

Boivin, J. and M. Giannoni (2006) DSGE models in a data-rich environment. NBER Working paper, No. 12772.

Breitung, J. and S. Eickmeier (2005), Dynamic factor models, Deutsche Bundesbank Discussion Paper, 38/2005.

Breitung, J. and S. Eickmeier (2011), Testing for structural breaks in dynamic factor models, *Journal of Econometrics*, 163, 71-74.

Chen, L. (2015) Estimation the common break date in large factor models, *Economics Letters*, 131, 70-74.

Chen, L., J. Dolado, and J. Gonzalo (2014), Detecting big structural breaks in large factor models, *Journal of Econometrics*, 180(1), 30-48.

Cheng, X., Z. Liao and F. Schorfheide (2015), Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities, *Manuscript*, University of Pennsylvania and UCLA.

Ching, H. S., C. Hsiao, and S. K. Wan (2012), Impact of CEPA on the labor market of Hong Kong, *China Economic Review*, 23, 975-981.

Connor, G. and R. Korajzcyk (1993), A Test for the Number of Factors in an Approximate Factor Model, *Journal of Finance*, 48, 1263-1291.

Cragg, J. and S. Donald (1997), Inferring the Rank of a Matrix, *Journal of Econometrics*, 76, 223-250.

Donald, S. (1997), Inference Concerning the Number of Factors in a Multivariate Nonparameteric Relationship, *Econometrica*, 65, 103-132.

Fama, E. F. and K. R. French (1993), Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*, Volume 33, Issue 1, 3-56.

Fan, J., Y. Liao, and M. Mincheva (2011), High-Dimensional Covariance Matrix Estimation in Approximate Factor Models, *Annals of Statistics*, 39, 3320-3356.

Fan, J., Y. Liao and M. Mincheva (2013), Large Covariance Estimation by Thresholding Principal Orthogonal Complements, *Journal of the Royal Statistical Society: Series B*, 75, 603-680.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000a), The General Dynamic Factor Model: Identification and Estimation, *Review of Economics and Statistics*, 82, 540-554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), Reference Cycles: The NBER Methodology Revisited, CEPR Discussion Paper 2400.

Forni, M. and L. Gambetti (2010), Macroeconomic Shocks and the Business Cycle: Evidence from a Structural Factor Model fiscal, Working Papers 440, Barcelona Graduate School of Economics.

Forni, M. and L. Reichlin (1998), Lets Get Real: a Factor-Analytic Approach to Disaggregated Business Cycle Dynamics, *Review of Economic Studies*, 65, 453-473.

Giannone, D., L. Reichlin, and L. Sala (2005), Monetary Policy in Real Time, NBER Macroeconomics Annual 2004 e.d. by Mark Gertler and Kenneth Rogoff, 161-200, MIT Press.

Gregory, A. and A. Head (1999), Common and Country-Specific Fluctuations in Productivity, Investment, and the Current Account, *Journal of Monetary Economics*, 44, 423-452.

Han, X. and A. Inoue (2015), Tests for Parameter Instability in Dynamic Factor Models, *Econometric Theory*, Volume 31, Issue 05, 1117-1152.

Hsiao, C., H. S. Ching and S. K. Wan (2012), A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Intergration of Hong Kong with Mainland China, *Journal of Applied Econometrics*, 27, 705-740.

Jurado, K., S. C. Ludvigson, S. Ng (2015), Measuring Uncertainty, *The American Economic Review*, 105(3):1177-1216.

Lettau, M. and S. C. Ludvigson (2001), Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia are Time-Varying, *Journal of Political Economy*, 109(6): 1238-1287.

Lewbel, A. (1991), The Rank of Demand Systems: Theory and Nonparametric Estimation, *Econometrica*, 59, 711-730.

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

Ludvigson, S. and S. Ng (2009), Macro Factors in Bond Risk Premia, *The Review of Financial Studies*, 22, 5027-5067.

Ludvigson, S. and S. Ng (2010), A Factor Analysis of Bond Risk Premia, Handbook of Empirical Economics and Finance, e.d. by Aman Uhla and David E. A. Giles, 313-372, Chapman and Hall, Boca Raton, FL.

Massacci, D. (2015), Least Squares Estimation of Large Dimensional Threshold Factor Models, Manuscript.

Onatski, A. (2009), Testing Hypotheses about the Number of Factors in Large Factor Models, *Econometrica*, 77, 1447-1479.

Onatski, A. (2010), Determining the Number of Factors from Empirical Distribution of Eigenvalues, *Review of Economic and Statistics*, 92, 1004-1016.

Qian, J. and L. Su (2014), Shrinkage Estimation of Common Breaks in Panel Data Models via Adaptive Group Fused Lasso, Manuscript, Singapore Management University.

Ross, S. (1976), The Arbitrage Theory of Capital Asset Pricing, *Journal of Finance*, 13, 341-360.

Stock, J. H. and M. W. Watson (1998), Diffusion Indexes, *NBER Working Papers*, 6702, National Bureau of Economic Research, Inc.

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

Stock, J. H. and M. W. Watson (2002), Macroeconomic Forecasting using Diffusion Indexes, *Journal of Business and Economic Statistics*, 20, 147-162.

Su, L. and X. Wang (2015), On Time-varying Factor Models: Estimation and Inference, Manuscript, Singapore Management University.

Tanaka, S. and Y. Yamamoto (2015), Testing for Factor Loading Structural Change under Common Breaks, Manuscript.

Yamamoto, Y. (2014), Forecasting with Non-spurious factors in U.S. Macroeconomic time series, *Journal of Business and Economic Statistics*, forthcoming.

APPENDIX A

SECTION 2 APPENDIX

A.1 Proofs

A.1.1 Proof of Proposition 2.2.1

We will first prove a lemma (Lemma 1) below which will be used in proving Proposition 2.2.1.

Lemma 1 Under Assumptions A-C, we have for some positive constant $0 < M_2 < \infty$, and for all N and T,

(i) $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_N(s,t)^2 \leq M_2$;
(ii) $E\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}X_{it}X_{is}\right)^2\right) \leq M_2$;

Proof :

(i) See the proof of Lemma 1 (i) in Bai and Ng (2002).

(ii) If suffices to prove that for all (i,t) that $E(X_{it}^4) \leq M$.
Now $E(X_{it}^4) \leq 8r^{-2}E[(\lambda_i^{0\prime}F_t^0)^4] + 8E(e_{it}^4) \leq 16M_1$ by assumption A5 and B1. ■

Proof of Proposition 2.2.1:

Recall that $\hat{F}^k = \frac{\sqrt{k}}{N}X\tilde{\Lambda}^k$ and $\tilde{\Lambda}^k = \frac{\sqrt{k}}{T}X'\tilde{F}^k$. From the normalization $\tilde{F}^{k\prime}\tilde{F}^k/T = I_k$, we also have $(Tk)^{-1}\sum_{t=1}^{T}\|\tilde{F}_t^k\|^2 = 1$. Following Bai and Ng (2002), using $H^{k\prime} = (\tilde{F}^{k\prime}F^0/T)(\Lambda^{0\prime}\Lambda^0/N)$, we have

$$\hat{F}_t^k - H^{k\prime}F_t^0 = \frac{k}{T}\sum_{s=1}^{T}\tilde{F}_s^k\gamma_N(s,t) + \frac{k}{T}\sum_{s=1}^{T}\tilde{F}_s^k\zeta_{st} + \frac{k}{T}\sum_{s=1}^{T}\tilde{F}_s^k\eta_{st} + \frac{k}{T}\sum_{s=1}^{T}\tilde{F}_s^k\xi_{st},$$

where $\zeta_{st} = e_s'e_t/N - \gamma_N(s,t)$, $\eta_{st} = F_s^{0\prime}\Lambda^{0\prime}e_t/(N\sqrt{r})$, and $\xi_{st} = F_t^{0\prime}\Lambda^{0\prime}e_s/(N\sqrt{r}) =$

η_{ts}.

Because $(x+y+z+u)^2 \leq 4(x^2+y^2+z^2+u^2)$, $\|\hat{F}_t^k - H^{k'} F_t^0\|^2 \leq 4(a_t + b_t + c_t + d_t)$, where $a_t = \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k \gamma_N(s,t) \right\|^2$, $b_t = \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k \zeta_{st} \right\|^2$, $c_t = \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k \eta_{st} \right\|^2$ and $d_t = \frac{k^2}{T^2} \left\| \sum_{s=1}^T \tilde{F}_s^k \xi_{st} \right\|^2$. It follows that $(1/T) \sum_{t=1}^T \|\hat{F}_t^k - H^{k'} F_t^0\|^2 \leq (4/T) \sum_{t=1}^T (a_t + b_t + c_t + d_t)$.

By Cauchy's inequality,

we have $\|\sum_{s=1}^T \tilde{F}_s^k \gamma_N(s,t)\|^2 \leq \left(\sum_{s=1}^T \|\tilde{F}_s^k\|^2\right) \cdot \left(\sum_{s=1}^T \gamma_N(s,t)^2\right)$. Thus,

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} a_t &\leq \frac{kk^2}{T}\left(\frac{1}{Tk}\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2\right)\cdot\frac{1}{T}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma_N(s,t)^2\right) \\ &= O_p\left(\frac{k^3}{T}\right) \end{aligned}$$

by Lemma 1(i) and the fact that $(Tk)^{-1}\sum_{t=1}^T \|\tilde{F}_t^k\|^2 = 1$

(this follows from $\tilde{F}^{k'}\tilde{F}^k/T = I_k$).

For b_t, we have that

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} b_t &= \frac{k^2}{T^3}\sum_{t=1}^{T}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\zeta_{st}\right\|^2 \\
&= \frac{k^2}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\tilde{F}_s^{k\prime}\tilde{F}_u^k\zeta_{st}\zeta_{ut} \\
&\leq \frac{k^2}{T}\left(\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}(\tilde{F}_s^{k\prime}\tilde{F}_u^k)^2\right)^{1/2}\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut}\right)^2\right]^{1/2} \\
&\leq \frac{k^3}{T}\left(\frac{1}{Tk}\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2\right)\left[\frac{1}{T^2}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut}\right)^2\right]^{1/2} \\
&= k^3\left[\frac{1}{T^4}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut}\right)^2\right]^{1/2} \\
&= O_p\left(\frac{k^3}{N}\right),
\end{aligned}$$

where the last equality follows from $\left[\frac{1}{T^4}\sum_{s=1}^{T}\sum_{u=1}^{T}\left(\sum_{t=1}^{T}\zeta_{st}\zeta_{ut}\right)^2\right]^{1/2} = O_p(N^{-1})$ as shown in the proof of Theorem 1 of Bai and Ng (2002).

From $E(T^{-1}\sum_{t=1}^{T}\zeta_{st}\zeta_{ut})^2 = E(T^{-2}\sum_{t=1}^{T}\sum_{v=1}^{T}\zeta_{st}\zeta_{ut}\zeta_{sv}\zeta_{uv}) \leq \max_{s,t} E|\zeta_{st}|^4$ and

$$E|\zeta_{st}|^4 = \frac{1}{N^2}E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{it}e_{is} - E(e_{it}e_{is}))\right|^4 \leq \frac{1}{N^2}M_1$$

by Assumption B5, we have

$$\frac{1}{T}\sum_{t=1}^{T} b_t \leq O_p(k^3)\frac{1}{T}\sqrt{\frac{T^2}{N^2}} = O_p\left(\frac{k^3}{N}\right).$$

For c_t, we have

$$\begin{aligned} c_t &= \frac{k^2}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k\eta_{st}\right\|^2 \\ &= \frac{k^2}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k F_s^{0\prime}\Lambda^{0\prime}e_t/N\sqrt{r}\right\|^2 \\ &\leq \frac{k^2}{N^2}\|e_t'\Lambda^0/\sqrt{r}\|^2\left(\frac{k}{Tk}\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2\right)\left(\frac{r}{Tr}\sum_{s=1}^{T}\|F_s^0\|^2\right) \\ &= \frac{k^2}{N^2}\|e_t'\Lambda^0/\sqrt{r}\|^2 O_p(kr) \end{aligned}$$

because $\frac{1}{Tk}\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2 = 1$ and $\frac{1}{Tr}\sum_{s=1}^{T}\|F_s^0\|^2 = O_p(1)$.

It follows that

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}c_t &= O_p(kr)\frac{k^2}{N}\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{e_t'\Lambda^0}{\sqrt{Nr}}\right\|^2 \\ &= O_p\left(\frac{k^3 r}{N}\right) \end{aligned}$$

because $\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{e_t'\Lambda^0}{\sqrt{Nr}}\right\|^2 = O_p(1)$ by assumption C2.

The term $(1/T)\sum_{t=1}^{T}d_t = O_p\left(\frac{k^3 r}{N}\right)$ can be proved similarly. Combining the above results, we have shown that

$$\begin{aligned} (1/T)\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2 &\leq (4/T)\sum_{t=1}^{T}(a_t+b_t+c_t+d_t) \\ &= O_p\left(\frac{k^3 r}{N}\right) + O_p\left(\frac{k^3}{T}\right). \end{aligned}$$

Alternatively, Proposition 2.2.1 can be proved by concentrating out F_t. Following the similar steps, we can show that

$$\begin{aligned}(1/N)\sum_{i=1}^{N}\|\hat{\lambda}_i^k - H^{k'}\lambda_i^0\|^2 &\leq (4/N)\sum_{i=1}^{N}(a_i+b_i+c_i+d_i)\\ &= O_p\left(\frac{k^3}{N}\right)+O_p\left(\frac{k^3 r}{T}\right).\end{aligned}$$

■

A.1.2 Proof of Proposition 2.2.2

Proof From the proof of Proposition 2.2.1 we know that

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k'}F_t^0\|^2 \leq (4/T)\sum_{t=1}^{T}(a_t+b_t+c_t+d_t)$$

and that $T^{-1}\sum_{t=1}^{T} a_t = O_p(k^3/T)$ and $T^{-1}\sum_{t=1}^{T} d_t = O_p(k^3/N)$. Therefore, we only need to show that $T^{-1}\sum_{t=1}^{T} c_t = O_p(k^3/N)$ and $T^{-1}\sum_{t=1}^{T} d_t = O_p(k^3/N)$. Since the proofs are similar. We will only prove for the term related to c_t.

For c_t, we have

$$\begin{aligned}c_t &= \frac{k^2}{T^2}\left\|\sum_{s=1}^{T}\tilde{F}_s^k F_s^{0'}\Lambda^{0'}e_t/N\sqrt{r}\right\|^2\\ &\leq \frac{k^2}{N}\frac{1}{r}\left(\frac{1}{T}\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2\right)\left(\frac{1}{TN}\sum_{s=1}^{T}\|F_s^{0'}\Lambda^{0'}e_t\|^2\right)\\ &= O_p(k^3/N)\frac{1}{r}\left(\frac{1}{TN}\sum_{s=1}^{T}\|F_s^{0'}\Lambda^{0'}e_t\|^2\right)\end{aligned}$$

because $T^{-1}\sum_{s=1}^{T}\|\tilde{F}_s^k\|^2 = O_p(k)$.

Next, we show that $A \stackrel{def}{=} (TN)^{-1}\sum_{s=1}^{T}\|F_s^{0'}\Lambda^{0'}e_t\|^2 = O_p(r)$.

$$
\begin{aligned}
E(|A|) &= \frac{1}{NT}\sum_{s=1}^{T} E\|F_s^{0'}\Lambda^{0'}e_t\|^2 \\
&= \frac{1}{NT}\sum_{s=1}^{T}\sum_{l=1}^{r}\sum_{m=1}^{r}\sum_{i=1}^{N}\sum_{j=1}^{N} E(e_{it}e_{jt})E(F_{sl}^0F_{sm}^0)\Lambda_{il}^0\lambda_{jm}^0 \\
&= \frac{1}{NT}\sum_{s=1}^{T}\sum_{l=1}^{r}\sum_{i=1}^{N} E(e_{it}^2)E((F_{sl}^0)^2)(\Lambda_{il}^0)^2 \\
&= O(r),
\end{aligned}
$$

because of the zero correlation assumptions that $E(e_{it}e_{jt}) = 0$ for $j \neq i$ and $E(F_{sl}^0F_{sm}^0) = 0$ for $m \neq l$. This implies that $A = O_p(r)$. Hence, $c_t = O_p(k^3/N)$. This completes the proof of Proposition 2.2.2. ■

From the above proof we can see that the conclusion of Proposition 2.2.2 still holds true if the zero correlation assumptions are replaced by some weakly dependent assumptions such as $N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N} E(e_{it}e_{jt}c_{1,ijlm}) = O(1)$
and $r^{-1}\sum_{l=1}^{r}\sum_{m\neq l}^{r} E(F_{sl}^0F_{sm}^0c_{2,lmlm}) = O(1)$,
where $c_{1,ijlm}$ and $c_{2,ijlm}$ are some bounded sequences of non-random numbers depending on i, j, l, m.

A.1.3 Proof of Theorem 2.2.1

Lemma 2 Let $D_k = \hat{F}^{k'}\hat{F}^k/T$ and $D_0 = H^{k'}F^{0'}F^0H^k/T$. When $k \leq r$, we have (i) $\|D_k^{-1}\| = O_p(k)$; (ii) $\|D_k^{-1} - D_0^{-1}\| = O_p\left(\max\left\{\frac{k_{max}^4 r^{1.5}}{\sqrt{N}}, \frac{k_{max}^4 r}{\sqrt{T}}\right\}\right)$.

Proof : Following Bai and Ng (2002), we have

$$
\begin{aligned}
D_k - D_0 &= \frac{\hat{F}^{k\prime}\hat{F}^k}{T} - \frac{H^{k\prime}F^{0\prime}F^0H^k}{T} \\
&= \frac{1}{T}\sum_{t=1}^{T}[\hat{F}_t^k\hat{F}_t^{k\prime} - H^{k\prime}F_t^0F_t^{0\prime}H^k] \\
&= \frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t^k - H^{k\prime}F_t^0)(\hat{F}_t^k - H^{k\prime}F_t^0)' + \frac{1}{T}\sum_{t=1}^{T}(\hat{F}_t^k - H^{k\prime}F_t^0)F_t^{0\prime}H^k \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}H^{k\prime}F_t^0(\hat{F}_t^k - H^{k\prime}F_t^0)'.
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
\|D_k - D_0\| &\leq \frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2 + 2\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t^0\|^2\right)^{1/2} \\
&\quad \cdot \left(\frac{1}{T}\sum_{t=1}^{T}\|H^{k\prime}F_t^0\|^2\right)^{1/2} \\
&= O_p\left(\max\left\{\frac{k^3r}{N}, \frac{k^3}{T}\right\}\right) + O_p\left(\max\left\{\frac{\sqrt{k^3r}}{\sqrt{N}}, \frac{\sqrt{k^3}}{\sqrt{T}}\right\}\right) \cdot O_p\left(\sqrt{kr^2}\right) \\
&= O_p\left(\max\left\{\frac{k^2r^{1.5}}{\sqrt{N}}, \frac{k^2r}{\sqrt{T}}\right\}\right)
\end{aligned}
$$

by Proposition 2.2.1 and the fact that $\frac{1}{T}\sum_{t=1}^{T}\|H^{k\prime}F_t^0\|^2 = O_p(kr^2)$, which is shown below.

From weakly dependent process of F_t^0, it is easy to show that

$$
\frac{1}{T}\sum_{t=1}^{T}\|H^{k\prime}F_t^0\|^2 - E\left[\frac{1}{T}\sum_{t=1}^{T}\|H^{k\prime}F_t^0\|^2\right] = O_p\left(\frac{1}{\sqrt{T}}\right).
$$

Also, one can easily show that $||D_k^{-1}|| = O_p(k)$. Then from $D_k^{-1} - D_0^{-1} =$

$D_k^{-1}(D_0 - D_k)D_0^{-1}$, we have

$$\begin{aligned}
\|D_k^{-1} - D_0^{-1}\| &= \|D_k^{-1}(D_o - D_k)D_0^{-1}\| \\
&\leq \|D_k^{-1}\| \cdot \|D_0 - D_k\| \cdot \|D_0^{-1}\| \\
&= k^2 \frac{\|D_k^{-1}\|}{k} \cdot \|D_0 - D_k\| \cdot \frac{\|D_0^{-1}\|}{k} \\
&= k^2 \cdot O_p(1) \cdot O_p\left(\max\left\{\frac{k^2 r^{1.5}}{\sqrt{N}}, \frac{k^2 r}{\sqrt{T}}\right\}\right) \\
&= O_p\left(\max\left\{\frac{k_{max}^4 r^{1.5}}{\sqrt{N}}, \frac{k_{max}^4 r}{\sqrt{T}}\right\}\right).
\end{aligned}$$

■

Lemma 3 For $1 \leq k \leq r$, and the H^k defined in Proposition 2.2.1, we have

$$V(k, \hat{F}^k) - V(k, F^0 H^k) = O_p\left(\max\left\{\frac{k_{max}^5 r^{3.5}}{\sqrt{N}}, \frac{k_{max}^5 r^3}{\sqrt{T}}\right\}\right).$$

Proof : For the true factor matrix with r factors and H^k defined in Proposition 2.2.1, let $M_{FH}^0 = I - P_{FH}^0$ denote the idempotent matrix spanned by null space of $F^0 H^k$, with $P_{FH^0} = F^0 H^k \left(H^{k\prime} F^{0\prime} F^0 H^k\right)^{-1} H^{k\prime} F^{0\prime}$. Correspondingly, let $M_{\hat{F}}^k = I_T - \hat{F}^k (\hat{F}^{k\prime} \hat{F}^k)^{-1} \hat{F}^{k\prime} = I_T - P_{\hat{F}}^k$. Then

$$V(k, \hat{F}^k) = \frac{1}{NT} \sum_{i=1}^{N} \underline{X}_i' M_{\hat{F}}^k \underline{X}_i,$$

$$V(k, F^0 H^k) = \frac{1}{NT} \sum_{i=1}^{N} \underline{X}_i' M_{FH}^0 \underline{X}_i,$$

$$V(k, \hat{F}^k) - V(k, F^0 H^k) = \frac{1}{NT} \sum_{i=1}^{N} \underline{X}_i' (P_{FH}^0 - P_{\hat{F}}^k) \underline{X}_i.$$

Following Bai and Ng (2002), let $D_k = \hat{F}^{k\prime}\hat{F}^k/T$ and $D_0 = H^{k\prime}F^{0\prime}F^0H^k/T$. Then

$$
\begin{aligned}
P^k_{\hat{F}} - P^0_{FH} &= \frac{1}{T}\hat{F}^k\left(\frac{\hat{F}^{k\prime}\hat{F}^k}{T}\right)^{-1}\hat{F}^{k\prime} - \frac{1}{T}F^0H^k\left(\frac{H^{k\prime}F^{0\prime}F^0H^k}{T}\right)^{-1}H^{k\prime}F^{0\prime} \\
&= \frac{1}{T}[\hat{F}^{k\prime}D_k^{-1}\hat{F}^k - F^0H^kD_0^{-1}H^{k\prime}F^{0\prime}] \\
&= \frac{1}{T}\big[(\hat{F}^k - F^0H^k + F^0H^k)D_k^{-1}(\hat{F}^k - F^0H^k + F^0H^k)' \\
&\quad -F^0H^kD_0^{-1}H^{k\prime}F^{0\prime}\big] \\
&= \frac{1}{T}[(\hat{F}^k - F^0H^k)D_k^{-1}(\hat{F}^k - F^0H^k)' + (\hat{F}^k - F^0H^k)D_k^{-1}H^{k\prime}F^{0\prime} \\
&\quad +F^0H^kD_k^{-1}(\hat{F}^k - F^0H^k)' - F^0H^kD_0^{-1}H^{k\prime}F^{0\prime}].
\end{aligned}
$$

Thus, $N^{-1}T^{-1}\sum_{i=1}^N \underline{X}_i'(P^k_{\hat{F}} - P^0_{FH})\underline{X}_i = I + II + III + IV$. We consider each term in turn.

$$
\begin{aligned}
I &= \frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T(\hat{F}^k_t - H^{k\prime}F^0_t)'D_k^{-1}(\hat{F}^k_s - H^{k\prime}F^0_s)X_{it}X_{is} \\
&\leq \left(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T(\hat{F}^k_t - H^{k\prime}F^0_t)'D_k^{-1}(\hat{F}^k_s - H^{k\prime}F^0_s)\right)^{1/2} \\
&\quad \cdot\left(\frac{1}{T^2}\sum_{t=1}\sum_{s=1}\left(\frac{1}{N}\sum_{i=1}^N X_{it}X_{is}\right)^2\right)^{1/2} \\
&\leq \left(\frac{1}{T}\sum_{t=1}^T\|\hat{F}^k_t - H^{k\prime}F^0_t\|^2\right)\cdot\|D_k^{-1}\|\cdot O_P(1) \\
&= O_p\left(\max\left\{\frac{k^3r}{N}, \frac{k^3}{T}\right\}\right)\cdot k\cdot O_p(1) \\
&= O_p\left(\max\left\{\frac{k^4r}{N}, \frac{k^4}{T}\right\}\right)
\end{aligned}
$$

by Proposition 2.2.1, Lemma 1(iii) and Lemma 2(i).

$$
\begin{aligned}
II &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}(\hat{F}_t^k - H^{k'}F_t^0)'D_k^{-1}H^{k'}F_s^0 X_{it}X_{is} \\
&\le \left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\|\hat{F}_t^k - H^{k'}F_t^0\|^2 \cdot \|H^{k'}F_s^0\|^2 \cdot \|D_k^{-1}\|^2\right)^{1/2} \\
&\quad \cdot \left(\frac{1}{T^2}\sum_{t=1}\sum_{s=1}\left(\frac{1}{N}\sum_{i=1}^{N}X_{it}X_{is}\right)^2\right)^{1/2} \\
&\le \left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k'}F_t^0\|^2\right)^{1/2} \cdot \|D_k^{-1}\| \cdot \left(\frac{kr^2}{Tkr^2}\sum_{s=1}^{T}\|H^{k'}F_s^0\|^2\right)^{1/2} \cdot O_p(1) \\
&= O_p\left(\max\left\{\left(\frac{k^3 r}{N}\right)^{1/2}, \left(\frac{k^3}{T}\right)^{1/2}\right\}\right) \cdot k \cdot k^{1/2} r \cdot O_p(1) \\
&= O_p\left(\max\left\{\frac{k^3 r^{1.5}}{\sqrt{N}}, \frac{k^3 r}{\sqrt{T}}\right\}\right).
\end{aligned}
$$

Similarly, one can verify that III is also $O_p\left(\max\left\{\frac{k^3 r^{1.5}}{\sqrt{N}}, \frac{k^3 r}{\sqrt{T}}\right\}\right)$.

$$
\begin{aligned}
IV &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}F_t^{0'}H^k(D_k^{-1} - D_0^{-1})H^{k'}F_s^0 X_{it}X_{is} \\
&\le \|D_k^{-1} - D_0^{-1}\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\|H^{k'}F_t^0\| \cdot |X_{it}|\right)^2 \\
&\le \|D_k^{-1} - D_0^{-1}\|\frac{kr^2}{N}\sum_{i=1}^{N}\left(\frac{1}{T\sqrt{kr}}\sum_{t=1}^{T}\|H^{k'}F_t^0\|\right)^2 \\
&= \|D_k^{-1} - D_0^{-1}\| \cdot kr^2 \cdot O_p(1) \\
&= O_p\left(\max\left\{\frac{k^4 r^{4.5}}{\sqrt{N}}, \frac{k^4 r^4}{\sqrt{T}}\right\}\right),
\end{aligned}
$$

where we used $\|D_k^{-1} - D_0^{-1}\| = O_p\left(\max\left\{\frac{k^3 r^{2.5}}{\sqrt{N}}, \frac{k^3 r^2}{\sqrt{T}}\right\}\right)$ by Lemma 2 (ii).

Thus, we have

$$V(k,\hat{F}^k)-V(k,F^0H^k)=O_p\left(\max\left\{\frac{k_{max}^5 r^{3.5}}{\sqrt{N}},\frac{k_{max}^5 r^3}{\sqrt{T}}\right\}\right).$$

■

Lemma 4 For the matrix H^k defined in Proposition 2.2.1, and for each k with $k<r=r_{N,T}\to\infty$, there exists a positive constant C such that

$$\text{plim}\inf_{N,T\to\infty}\inf_k[V(k,F^0H^k)-V(r,F^0)]\geq C>0.$$

Proof :

$$\begin{aligned}
V(k,F^0H^k)-V(r,F^0) &= \frac{1}{NT}\sum_{i=1}^{N}\underline{X}_i' M_{FH}^0\underline{X}_i-\frac{1}{NT}\sum_{i=1}^{N}\underline{X}_i' M_F^0\underline{X}_i \\
&= \frac{1}{NT}\sum_{i=1}^{N}(\frac{1}{\sqrt{r}}F^0\lambda_i^0+\underline{e}_i)' M_{FH}^0(\frac{1}{\sqrt{r}}F^0\lambda_i^0+\underline{e}_i) \\
&\quad -\frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i' M_F^0\underline{e}_i \\
&= \frac{1}{NTr}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}M_{FH}^0F^0\lambda_i^0+\frac{2}{NT\sqrt{r}}\sum_{i=1}^{N}\underline{e}_i' M_{FH}^0F^0\lambda_i^0 \\
&\quad +\frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i'(P_F^0-P_{FH}^0)\underline{e}_i \\
&= A+B+D.
\end{aligned}$$

Notice that $P_F^0 - P_{FH}^0 \geq 0$, thus $III \geq 0$. For the first term,

$$\begin{aligned} A &= \frac{1}{NTr}\sum_{i=1}^{N}\lambda_i^{0\prime}F^{0\prime}M_{FH}^0F^0\lambda_i^0 \\ &= \frac{1}{NTr}\sum_{i=1}^{N}\left(M_{FH}^0F^0\lambda_i^0\right)' M_{FH}^0F^0\lambda_i^0 \\ &\geq C > 0 \end{aligned}$$

because $k < r$ and $M_{FH}^0F^0\lambda_i^0 \neq 0$.

Next,

$$B = \frac{2}{NT\sqrt{r}}\sum_{i=1}^{N}\underline{e}_i'F^0\lambda_i^0 - \frac{2}{NT\sqrt{r}}\sum_{i=1}^{N}\underline{e}_i'P_{FH}^0F^0\lambda_i^0.$$

Consider the first term

$$\begin{aligned} \left|\frac{1}{NT\sqrt{r}}\sum_{i=1}^{N}\underline{e}_i'F^0\lambda_i^0\right| &= \left|\frac{1}{NT\sqrt{r}}\sum_{i=1}^{N}\sum_{t=1}^{T}\underline{e}_{it}F_t^{0\prime}\lambda_i^0\right| \\ &\leq \left(\frac{1}{Tr}\sum_{t=1}^{T}\|F_t^0\|^2\right)^{1/2}\cdot\sqrt{r} \\ &\quad \cdot\frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{Nr}}\sum_{i=1}^{N}e_{it}\lambda_i^0\right\|^2\right)^{1/2} \\ &= O_p\left(\frac{\sqrt{r}}{\sqrt{N}}\right), \end{aligned}$$

where the last equality follows from assumption C2. The second term is also $o_p(1)$, and hence $B = o_p(1)$. ■

Lemma 5 For any k with $r \leq k \leq k_{max}$,

$$V(k,\hat{F}^k) - V(r,\hat{F}^r) = O_p\left(\max\left\{\frac{k_{max}^2 r}{N}, \frac{k_{max}^2 r^{0.5}}{T}\right\}\right).$$

Proof :

$$\begin{aligned}|V(k,\hat{F}^k)-V(r,\hat{F}^r)| &\le |V(k,\hat{F}^k)-V(r,F^0)|+|V(r,F^0)-V(r,\hat{F}^r)| \\ &\le 2\max_{r\le k}|V(k,\hat{F}^k)-V(r,F^0)|.\end{aligned}$$

Thus, it is sufficient to prove for each k with $r \le k \le k_{max}$,

$$V(k,\hat{F}^k)-V(r,F^0)=O_p\left(\max\left\{\frac{k_{max}r^2}{\sqrt{N}},\frac{k_{max}r^{1.5}}{\sqrt{T}}\right\}\right).$$

Let H^k be as defined in Proposition 2.2.1, with full row rank. Let the $k\times r$ matrix H^{k+} be the generalized inverse of H^k such that $H^kH^{k+}=I_r$. From $\underline{X}_i=\frac{1}{\sqrt{r}}F^0\lambda_i^0+\underline{e}_i$, we have $\underline{X}_i=\frac{1}{\sqrt{r}}F^0H^kH^{k+}\lambda_i^0+\underline{e}_i$. This implies that

$$\begin{aligned}\underline{X}_i &= \frac{1}{\sqrt{r}}\hat{F}^kH^{k+}\lambda_i^0+\underline{e}_i-\frac{1}{\sqrt{r}}(\hat{F}^k-F^0H^k)H^{k+}\lambda_i^0 \\ &= \frac{1}{\sqrt{r}}\hat{F}^kH^{k+}\lambda_i^0+\underline{u}_i,\end{aligned}$$

where $\underline{u}_i=\underline{e}_i-\frac{1}{\sqrt{r}}(\hat{F}^k-F^0H^k)H^{k+}\lambda_i^0$.

Note that

$$
\begin{aligned}
V(k, \hat{F}^k) &= \frac{1}{NT} \sum_{i=1}^{N} \underline{u}_i' M_{\hat{F}}^k \underline{u}_i, \\
V(r, F^0) &= \frac{1}{NT} \sum_{i=1}^{N} \underline{e}_i' M_F^0 \underline{e}_i, \\
V(k, \hat{F}^k) &= \frac{1}{NT} \sum_{i=1}^{N} \left(\underline{e}_i - \frac{1}{\sqrt{r}} (\hat{F}^k - F^0 H^k) H^{k+} \lambda_i^0 \right)' M_{\hat{F}}^k \\
&\quad \left(\underline{e}_i - \frac{1}{\sqrt{r}} (\hat{F}^k - F^0 H^k) H^{k+} \lambda_i^0 \right), \\
&= \frac{1}{NT} \sum_{i=1}^{N} \underline{e}_i' M_{\hat{F}}^k \underline{e}_i - \frac{2}{NT\sqrt{r}} \sum_{i=1}^{N} \lambda_i^{0'} H^{k+'} (\hat{F}^k - F^0 H^k)' M_{\hat{F}}^k \underline{e}_i \\
&\quad + \frac{1}{NTr} \sum_{i=1}^{N} \lambda_i^{0'} H^{k+'} (\hat{F}^k - F^0 H^k)' M_{\hat{F}}^k (\hat{F}^k - F^0 H^k) H^{k+} \lambda_i^0 \\
&= a + b + c.
\end{aligned}
$$

Because $I - M_{\hat{F}}^k$ is positive semi-definite, $x' M_{\hat{F}}^k x \leq x'x$. Thus

$$
\begin{aligned}
c &\leq \frac{1}{NTr} \sum_{i=1}^{N} \lambda_i^{0'} H^{k+'} (\hat{F}^k - F^0 H^k)' (\hat{F}^k - F^0 H^k) H^{k+} \lambda_i^0 \\
&\leq \frac{1}{T} \sum_{t=1}^{T} \| \hat{F}_t^k - H^{k'} F_t^0 \|^2 \cdot \left(\frac{1}{Nr} \sum_{i=1}^{N} \| \lambda_i^0 \|^2 \| H^{k+} \|^2 \right) \\
&= O_p \left(\max \left\{ \frac{k^3 r}{N}, \frac{k^3}{T} \right\} \right) \cdot O_p(kr) \\
&= O_p \left(\max \left\{ \frac{k^4 r^2}{N}, \frac{k^4 r}{T} \right\} \right)
\end{aligned}
$$

by Proposition 2.2.1.

For term b, we use the fact that $|\mathrm{tr}(A)| \leq r\|A\|$ for any $r \times r$ matrix A. Thus

$$\begin{aligned}
b &= \frac{2}{T\sqrt{r}}\mathrm{tr}\left(H^{k+}(\hat{F}^k - F^0H^k)'M^k_{\hat{F}}\left(\frac{1}{N}\sum_{i=1}^{N}\underline{e}_i\lambda_i^0\right)\right) \\
&\leq 2\cdot\|H^{k+}\|\cdot\left\|\frac{\hat{F}^k - F^0H^k}{\sqrt{T}}\right\|\cdot\left\|\frac{1}{\sqrt{TrN}}\sum_{i=1}^{N}\underline{e}_i\lambda_i^0\right\| \\
&\leq 2\cdot\|H^{k+}\|\cdot\left(\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}^k - F^0H^k\|^2\right)^{1/2}\cdot\frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{Nr}}\sum_{i=1}\underline{e}_i\lambda_i^0\right\|^2\right)^{1/2} \\
&= 2\cdot(kr)^{1/2}\cdot O_p\left(\max\left\{\frac{\sqrt{k^3r}}{\sqrt{N}},\frac{\sqrt{k^3}}{\sqrt{T}}\right\}\right)\cdot O_p(1) \\
&= O_p\left(\max\left\{\frac{k^2r}{\sqrt{N}},\frac{k^2r^{0.5}}{\sqrt{T}}\right\}\right)
\end{aligned}$$

by Proposition 2.2.1 and assumption C2. Therefore,

$$V(k,\hat{F}^k) = \frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i'M^k_{\hat{F}}\underline{e}_i + O_p\left(\max\left\{\frac{k^2r}{\sqrt{N}},\frac{k^2r^{0.5}}{\sqrt{T}}\right\}\right).$$

Thus we have

$$\begin{aligned}
V(k,\hat{F}^k) - V(r,F^0) = \frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i'P^0_F\underline{e}_i - \frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i'P^k_{\hat{F}}\underline{e}_i \\
+ O_p\left(\max\left\{\frac{k^2r}{\sqrt{N}},\frac{k^2r^{0.5}}{\sqrt{T}}\right\}\right).
\end{aligned}$$

Note that

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i' P_F^0 \underline{e}_i &\leq \left\| \left(\frac{F^{0\prime}F^0}{T}\right)^{-1} \right\| \cdot \frac{1}{NT^2}\sum_{i=1}^{N}\underline{e}_i' F^0 F^{0\prime}\underline{e}_i \\
&= \left\| \left(\frac{F^{0\prime}F^0}{T}\right)^{-1} \right\| \cdot \frac{1}{NT}\sum_{i=1}^{N}\left\| \frac{1}{\sqrt{Tr}}\sum_{t=1}^{T} F_t^0 \underline{e}_{it} \right\|^2 \cdot r \\
&= r \cdot O_p(1) \cdot \frac{1}{T} \cdot r \cdot O_p(1) \\
&= O_p\left(\frac{r^2}{T}\right) \leq O_p\left(\max\left\{\frac{k^2 r}{\sqrt{N}}, \frac{k^2 r^{0.5}}{\sqrt{T}}\right\}\right).
\end{aligned}$$

$\frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i' P_F^0 \underline{e}_i$ is bounded by the sum of the first k largest eigenvalues of the matrix $A_{NT} = \frac{1}{NT}e'e$, where $e = (e_{ti}), T \times N$. Let $\rho(A)$ denote the largest eigenvalue of a matrix A. Under Assumption B6, as Bai and Ng (2005) shows, $\rho(A_{NT}) = O_p(C_{NT}^{-2})$, where $C_{NT}^2 = \min(N, T)$. Thus,

$$\frac{1}{NT}\sum_{i=1}^{N}\underline{e}_i' P_F^0 \underline{e}_i = O_p\left(\max\left\{\frac{k}{N}, \frac{k}{T}\right\}\right) \leq O_p\left(\max\left\{\frac{k^2 r}{\sqrt{N}}, \frac{k^2 r^{0.5}}{\sqrt{T}}\right\}\right).$$

In summary, we have shown that

$$V(k, \hat{F}^k) - V(r, F^0) = O_p\left(\max\left\{\frac{k_{max}^2 r}{\sqrt{N}}, \frac{k_{max}^2 r^{0.5}}{\sqrt{T}}\right\}\right).$$

■

Proof of Theorem 2.2.1

Proof : We shall prove that $\lim_{N,T\to\infty} P(PC(k) < PC(r)) = 0$ for all $k \neq r$. Since

$$PC(k) - PC(r) = V(k, \hat{F}^k) - V(r, \hat{F}^r) - (r - k)g(N, T),$$

it is sufficient to prove that $P[V(k, \hat{F}^k) - V(r, \hat{F}^r) < (r-k)g(N,T)] \rightarrow 0$ as $N, T, k, r \rightarrow \infty$.

Consider $k < r$. We have the identity:

$$\begin{aligned} V(k, \hat{F}^k) - V(r, \hat{F}^r) &= [V(k, \hat{F}^k) - V(k, F^0H^k)] + [V(k, F^0H^k) - V(r, F^0H^r)] \\ &\quad + [V(r, F^0H^r) - V(r, \hat{F}^r)]. \end{aligned}$$

Lemma 3 implies that the first and the third terms are both $O_p\left(\max\left\{\frac{k_{max}^{8.5}}{\sqrt{N}}, \frac{k_{max}^{8}}{\sqrt{T}}\right\}\right)$. Next, we consider the second item. Because F^0H^r and F^0 span the same column space, $V(r, F^0H^r) = V(r, F^0)$. Thus the second item can be rewritten as $V(k, F^0H^k) - V(r, F^0)$, which has a positive limit by Lemma 4. Hence $P[PC(k) < PC(r)] \rightarrow 0$ if $(r-k)g(N,T) \rightarrow 0$ as $N, T, k, r \rightarrow \infty$.

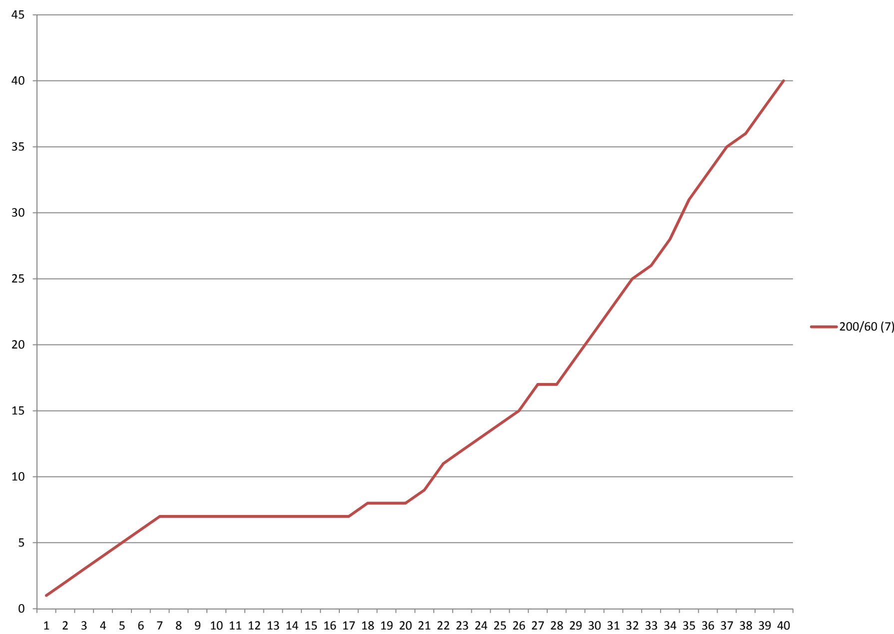
Next, for $k \geq r$,

$$P[PC(k) - PC(r) < 0] = P[V(r, \hat{F}^r) - V(k, \hat{F}^k) > (k-r)g(N,T)].$$

By Lemma 5, $V(r, \hat{F}^r) - V(k, \hat{F}^k) = O_p\left(\max\left\{\frac{k_{max}^{3}}{\sqrt{N}}, \frac{k_{max}^{2.5}}{\sqrt{T}}\right\}\right)$. According to our setting, $(k-r)g(N,T)$ converges to zero at a slower rate than $O_p\left(\max\left\{\frac{k_{max}^{3}}{\sqrt{N}}, \frac{k_{max}^{2.5}}{\sqrt{T}}\right\}\right)$. Thus, for $k > r$, $P[PC(k) < PC(r)] \rightarrow 0$ as $N, T, k, r \rightarrow \infty$. ■

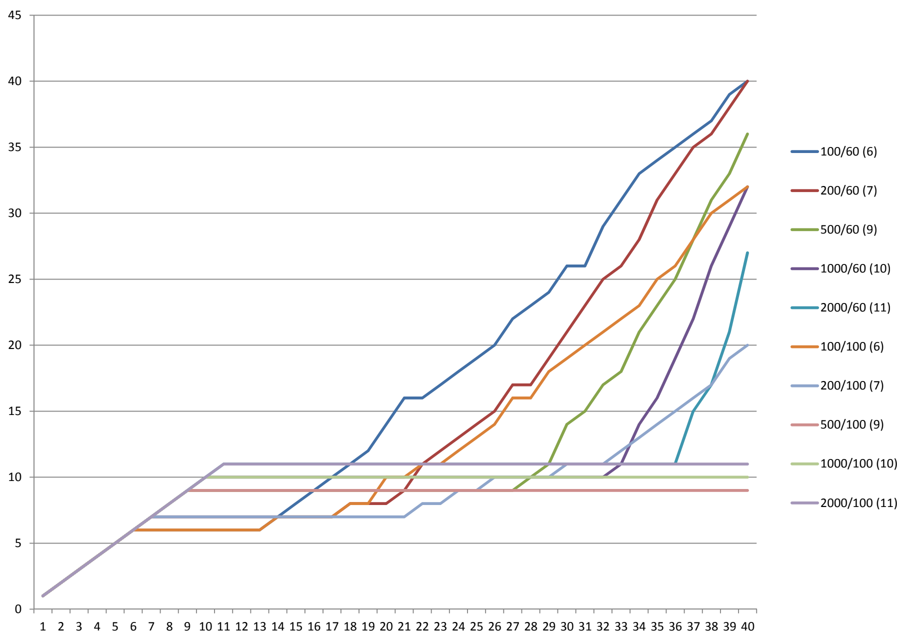
A.2 Figures

Figure A.1: Sensitivity of PC_{p1} Criterion to k_{max}: 200/60 Case



Note:The values of $\hat{k}$ estimated by PC_{p1} for $N = 200, T = 60$ and $r = 7$ with $k_{max} \in [1, 40]$.

Figure A.2: Sensitivity of PC_{p1} Criterion to k_{max}: Multiple Cases



Note: Each line represents $\hat{k}$ estimated by PC_{p1} for each case of different sample size.
The notation in the graph shows the sample size and the true number of factors for each case.
For example, $100/60(6)$ means that $N = 100, T = 60$ and $r = 6$.

APPENDIX B

SECTION 3 APPENDIX

B.1 Proofs

B.1.1 Proof of Theorem 3.2.1

Proof

For $k \leq k_0$,

$$\begin{aligned}
SSR(k) &= \sum_{i=1}^{N}\sum_{t=1}^{k}\left(x_{it} - \hat{\lambda}_{i1}^{(k)\prime}\hat{f}_t^{(k)}\right)^2 + \sum_{i=1}^{N}\sum_{t=k+1}^{T}\left(x_{it} - \hat{\lambda}_{i2}^{(k)\prime}\hat{f}_t^{(T-k)}\right)^2 \\
&= \sum_{i=1}^{N}\sum_{t=1}^{k}\left(x_{it} - \hat{\lambda}_{i1}^{(k)\prime}\hat{f}_t^{(k)}\right)^2 + \sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left(x_{it} - \hat{\lambda}_{i2}^{(k)\prime}\hat{f}_t^{(T-k)}\right)^2 \\
&\quad + \sum_{i=1}^{N}\sum_{t=k_0+1}^{T}\left(x_{it} - \hat{\lambda}_{i2}^{(k)\prime}\hat{f}_t^{(T-k)}\right)^2,
\end{aligned}$$

and

$$\begin{aligned}
SSR(k_0) &= \sum_{i=1}^{N}\sum_{t=1}^{k_0}\left(x_{it} - \hat{\lambda}_{i1}^{(k_0)\prime}\hat{f}_t^{(k_0)}\right)^2 + \sum_{i=1}^{N}\sum_{t=k_0+1}^{T}\left(x_{it} - \hat{\lambda}_{i2}^{(k_0)\prime}\hat{f}_t^{(T-k_0)}\right)^2 \\
&= \sum_{i=1}^{N}\sum_{t=1}^{k}\left(x_{it} - \hat{\lambda}_{i1}^{(k_0)\prime}\hat{f}_t^{(k_0)}\right)^2 + \sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left(x_{it} - \hat{\lambda}_{i1}^{(k_0)\prime}\hat{f}_t^{(k_0)}\right)^2 \\
&\quad + \sum_{i=1}^{N}\sum_{t=k_0+1}^{T}\left(x_{it} - \hat{\lambda}_{i2}^{(k_0)\prime}\hat{f}_t^{(T-k_0)}\right)^2.
\end{aligned}$$

So

$$SSR(k) - SSR(k_0) = \sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left(x_{it} - \hat{\lambda}_{i2}^{(k)\prime}\hat{f}_t^{(T-k)}\right)^2 - \sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left(x_{it} - \hat{\lambda}_{i1}^{(k_0)\prime}\hat{f}_t^{(k_0)}\right)^2$$
$$+ \sum_{i=1}^{N}\sum_{t=1}^{k}\left(x_{it} - \hat{\lambda}_{i1}^{(k)\prime}\hat{f}_t^{(k)}\right)^2 - \sum_{i=1}^{N}\sum_{t=1}^{k}\left(x_{it} - \hat{\lambda}_{i1}^{(k_0)\prime}\hat{f}_t^{(k_0)}\right)^2$$
$$+ \sum_{i=1}^{N}\sum_{t=k_0+1}^{T}\left(x_{it} - \hat{\lambda}_{i2}^{(k)\prime}\hat{f}_t^{(T-k)}\right)^2$$
$$- \sum_{i=1}^{N}\sum_{t=k_0+1}^{T}\left(x_{it} - \hat{\lambda}_{i2}^{(k_0)\prime}\hat{f}_t^{(T-k_0)}\right)^2.$$

We first analyze the property of the following term

$$\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left(x_{it} - \hat{\lambda}_{i2}^{(k)\prime}\hat{f}_t^{(T-k)}\right)^2.$$

We have

$$\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left(x_{it} - \hat{\lambda}_{i2}^{(k)\prime}\hat{f}_t^{(T-k)}\right)^2 = \sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left(\lambda_{1i}'f_t + e_{it} - \hat{\lambda}_{i2}^{(k)\prime}\hat{f}_t^{(T-k)}\right)^2$$
$$= \sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\Bigg[(\lambda_{i1} - \lambda_{i2})' f_t + \left(\lambda_{i2} - \hat{\lambda}_{i2}^{(k)}\right)' f_t$$
$$+ \hat{\lambda}_{i2}^{(k)\prime}\left(f_t - \hat{f}_t^{(T-k)}\right) + e_{it}\Bigg]^2$$
$$= \sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\Bigg\{\left[(\lambda_{i1} - \lambda_{i2})'f_t\right]^2 + \left[(\lambda_{i2} - \hat{\lambda}_{i2}^{(k)})'f_t\right]^2$$
$$+ \left[\hat{\lambda}_{i2}^{(k)\prime}(f_t - \hat{f}_t^{(T-k)})\right]^2 + e_{it}^2 + 2(\lambda_{i1} - \lambda_{i2})'f_t e_{it}$$
$$+ 2(\lambda_{i2} - \hat{\lambda}_{i2}^{k})'f_t e_{it} + 2\hat{\lambda}_{i2}^{(k)\prime}(f_t - \hat{f}_t^{(T-k)})e_{it}$$
$$+ \text{other cross-products are negligible}\Bigg\},$$

e_{it}^2 is canceled out with the corresponding term from $SSR(k_0)$. The drift term is determined by

$$\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left\{\left[(\lambda_{i1}-\lambda_{i2})' f_t\right]^2+\left[(\lambda_{i2}-\hat{\lambda}_{i2}^{(k)})' f_t\right]^2+\left[\hat{\lambda}_{i2}^{(k)\prime}(f_t-\hat{f}_t^{(T-k)})\right]^2\right\}$$

and the random walk part is contributed by

$$\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left[2(\lambda_{i1}-\lambda_{i2})' f_t e_{it}+2(\lambda_{i2}-\hat{\lambda}_{i2}^{(k)})' f_t e_{it}+2\hat{\lambda}_{i2}^{(k)\prime}(f_t-\hat{f}_t^{(T-k)})e_{it}\right]$$

Note that k is close to k_0, using the results of Bai (2003) or Bai and Ng (2013) (ignore high order term), we have

$$\begin{aligned}\lambda_{i2}-\hat{\lambda}_{i2}^{(k)} &= -\left(F_2' F_2\right)^{-1} F_2' e_i\\ &= -\left(F_2' F_2\right)^{-1}\sum_{s=k_0+1}^{T} f_s e_{is},\end{aligned}$$

and

$$\begin{aligned}f_t-\hat{f}_t^{(T-k)} &= -\left(\Lambda_2'\Lambda_2\right)^{-1}\Lambda_2' e_t\\ &= -\left(\Lambda_2'\Lambda_2\right)^{-1}\sum_{j=1}^{N}\lambda_{j2}e_{jt}\end{aligned}$$

Plug in the above two expressions into the random walk parts, then each of the three random walk terms are not negligible (assuming N and T are same order magnitude).

We first check the drift parts, let

$$D_1 = \sum_{i=1}^{N} \sum_{t=k+1}^{k_0} \left[(\lambda_{i1} - \lambda_{i2})' f_t \right]^2,$$

$$D_2 = \sum_{i=1}^{N} \sum_{t=k+1}^{k_0} \left[(\lambda_{i2} - \hat{\lambda}_{i2}^{(k)})' f_t \right]^2,$$

$$D_3 = \sum_{i=1}^{N} \sum_{t=k+1}^{k_0} \left[\hat{\lambda}_{i2}^{(k)'} (f_t - \hat{f}_t^{(T-k)}) \right]^2.$$

Suppose e_{it} is homoscedasticity, i.e. $E(e_{it}^2) = \sigma^2$, and $k_0 - k$ is large, we can show that

$$\begin{aligned} D_1 = \sum_{i=1}^{N} \sum_{t=k+1}^{k_0} \left[(\lambda_{i1} - \lambda_{i2})' f_t \right]^2 &= (k_0 - k) \sum_{i=1}^{N} (\lambda_{i1} - \lambda_{i2})' \\ &\quad \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} f_t f_t' (\lambda_{i1} - \lambda_{i2}) \\ &= (k_0 - k) \frac{1}{N} \sum_{i=1}^{N} \Delta_i' \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} f_t f_t' \Delta_i \\ &\xrightarrow{as\ N \to \infty} (k_0 - k) \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \Delta_i' \Sigma_f \Delta_i \\ &= (k_0 - k) \psi_1, \end{aligned}$$

$$
\begin{aligned}
D_2 &= \sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left[\sum_{s=k_0+1}^{T} e_{is}f_s^{'}\left(F_2^{'}F_2\right)^{-1}f_t\right]^2 \\
&= (k_0-k)\sum_{i=1}^{N}\sum_{s=k_0+1}^{T} e_{is}f_s^{'}\left(F_2^{'}F_2\right)^{-1}\left(\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} f_tf_t^{'}\right)\left(F_2^{'}F_2\right)^{-1}\sum_{s=k_0+1}^{T} f_se_{is} \\
&= (k_0-k)\frac{N}{T-k_0}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T-k_0}}\sum_{s=k_0+1}^{T} e_{is}f_s^{'}\right)\left(\frac{F_2^{'}F_2}{T-k_0}\right)^{-1} \\
&\quad \left(\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} f_tf_t^{'}\right)\left(\frac{F_2^{'}F_2}{T-k_0}\right)^{-1}\left(\frac{1}{\sqrt{T-k_0}}\sum_{s=k_0+1}^{T} f_se_{is}\right) \\
&= (k_0-k)\frac{N}{T-k_0}\frac{1}{N}\sum_{i=1}^{N}\xi_i^{'}\left(\frac{F_2^{'}F_2}{T-k_0}\right)^{-1}\left(\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} f_tf_t^{'}\right)\left(\frac{F_2^{'}F_2}{T-k_0}\right)^{-1}\xi_i \\
&\xrightarrow{by\ LLN} (k_0-k)\frac{N}{T-k_0}\frac{1}{N}\sum_{i=1}^{N}E\left(\xi_i^{'}\Sigma_f^{-1}\Sigma_f\Sigma_f^{-1}\xi_i\right) \\
&= (k_0-k)\frac{N}{T-k_0}\frac{1}{N}\sum_{i=1}^{N}E\left(tr(\xi_i^{'}\Sigma_f^{-1}\xi_i)\right) \\
&= (k_0-k)\frac{N}{T-k_0}\frac{1}{N}\sum_{i=1}^{N}tr\left(E(\xi_i\xi_i^{'}\Sigma_f^{-1})\right) \\
&= (k_0-k)\frac{N}{T-k_0}\sigma^2 \\
&= (k_0-k)\psi_2,
\end{aligned}
$$

$$
\begin{aligned}
D_3 &= \sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\left[\hat{\lambda}_{i2}^{(k)\prime}\left(\Lambda_2^{'}\Lambda_2\right)^{-1}\sum_{j=1}^{N}\lambda_{j2}e_{jt}\right]^2 \\
&= \sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\hat{\lambda}_{i2}^{(k)\prime}\left(\Lambda_2^{'}\Lambda_2\right)^{-1}\sum_{j=1}^{N}\lambda_{j2}e_{jt}\sum_{j=1}^{N}e_{jt}\lambda_{j2}^{'}\left(\Lambda_2^{'}\Lambda_2\right)^{-1}\hat{\lambda}_{i2}^{(k)} \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\hat{\lambda}_{i2}^{(k)\prime}\left(\frac{\Lambda_2^{'}\Lambda_2}{N}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_{j2}e_{jt}\right)\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}e_{jt}\lambda_{j2}^{'}\right) \\
&\quad \left(\frac{\Lambda_2^{'}\Lambda_2}{N}\right)^{-1}\hat{\lambda}_{i2}^{(k)} \\
&= \sigma^2\frac{1}{N}\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\hat{\lambda}_{i2}^{(k)\prime}\left(\frac{\Lambda_2^{'}\Lambda_2}{N}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_{j2}e_{jt}\right)\sigma^{-2}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}e_{jt}\lambda_{j2}^{'}\right) \\
&\quad \left(\frac{\Lambda_2^{'}\Lambda_2}{N}\right)^{-1}\hat{\lambda}_{i2}^{(k)} \\
&\xrightarrow{by\ CLT}\sigma^2\sum_{t=k+1}^{k_0}\chi_1^2(t),
\end{aligned}
$$

so D_3 should be considered with random parts together.

Now let us look at the random walk parts, define

$$
\begin{aligned}
R_1 &= 2\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}(\lambda_{i1}-\lambda_{i2})' f_t e_{it}, \\
R_2 &= 2\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}(\lambda_{i2}-\hat{\lambda}_{i2}^{k})' f_t e_{it}, \\
R_3 &= 2\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\hat{\lambda}_{i2}^{(k)\prime}(f_t-\hat{f}_t^{(T-k)})e_{it}
\end{aligned}
$$

$$
\begin{aligned}
R_1 &= 2\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}(\lambda_{i1}-\lambda_{i2})' f_t e_{it} \\
&= 2\sum_{t=k+1}^{k_0}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta_i^{'} f_t e_{it} \\
&\xrightarrow{by\ CLT} 2\sum_{t=k+1}^{k_0} N\left(0, E(\Delta_i^{'} f_t e_{it}^2 f_t^{'} \Delta_i)\right) \\
&= 2\sum_{t=k+1}^{k_0} N\left(0, \sigma^2 tr\left(E(\Delta_i \Delta_i^{'})E(f_t f_t^{'})\right)\right) \\
&= 2\sum_{t=k+1}^{k_0} N\left(0, \sigma^2 tr\left(\Sigma_\Delta \Sigma_f\right)\right) \\
&= 2\left(\sigma^2 tr\left(\Sigma_\Delta \Sigma_f\right)\right)^{-\frac{1}{2}}\sum_{t=k+1}^{k_0} Z_t \\
&= \phi_1 \sum_{t=k+1}^{k_0} Z_t
\end{aligned}
$$

where Z_t are $i.i.d.$ standard normal random variables.

$$
\begin{aligned}
R_2 &= 2\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}(\lambda_{i2}-\hat{\lambda}_{i2}^{k})' f_t e_{it} = -2\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\sum_{s=k_0+1}^{T} e_{is} f_s^{'}\left(F_2^{'}F_2\right)^{-1} f_t e_{it} \\
&= -2\frac{1}{\sqrt{T-k_0}}\sum_{t=k+1}^{k_0}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T-k_0}}\sum_{s=k_0+1}^{T} e_{is} f_s^{'}\right)\left(\frac{F_2^{'}F_2}{T-k_0}\right)^{-1} f_t e_{it} \\
&= -2\frac{1}{\sqrt{T-k_0}}\sum_{t=k+1}^{k_0}\sum_{i=1}^{N}\xi_i^{'}\left(\frac{F_2^{'}F_2}{T-k_0}\right)^{-1} f_t e_{it} \\
&= -2\frac{\sqrt{N}}{\sqrt{T-k_0}}\sum_{t=k+1}^{k_0}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i^{'}\left(\frac{F_2^{'}F_2}{T-k_0}\right)^{-1} f_t e_{it} \\
&\xrightarrow{by\ CLT} -2\frac{\sqrt{N}}{\sqrt{T-k_0}}\sum_{t=k+1}^{k_0} N\left(0, E(\xi_i^{'}\Sigma_f^{-1} f_t e_{it}^2 f_t^{'}\Sigma_f^{-1}\xi_i)\right) \\
&= -2\frac{\sqrt{N}}{\sqrt{T-k_0}}\sum_{t=k+1}^{k_0} N\left(0, \sigma^2 tr\left(E(\xi_i\xi_i^{'}\Sigma_f^{-1} f_t f_t^{'}\Sigma_f^{-1})\right)\right) \\
&= -2\frac{\sqrt{N}}{\sqrt{T-k_0}}\sum_{t=k+1}^{k_0} N\left(0, \sigma^4\right) \\
&= -2\frac{\sqrt{N}}{\sqrt{T-k_0}}\sigma^{-2}\sum_{t=k+1}^{k_0} Z_t \\
&= \phi_2\sum_{t=k+1}^{k_0} Z_t
\end{aligned}
$$

$$
\begin{aligned}
R_3 &= 2\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\hat{\lambda}_{i2}^{(k)'}(f_t-\hat{f}_t^{(T-k)})e_{it} = -2\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\hat{\lambda}_{i2}^{(k)'}\left(\Lambda_2^{'}\Lambda_2\right)^{-1}\sum_{j=1}^{N}\lambda_{j2}e_{jt}e_{it},\\
&= -2\sum_{t=k+1}^{k_0}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\lambda}_{i2}^{(k)'}e_{it}\right)\left(\frac{\Lambda_2^{'}\Lambda_2}{N}\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_{j2}e_{jt}\right)\\
&= -2\sum_{t=k+1}^{k_0} tr\left(\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\lambda}_{i2}^{(k)'}e_{it}\right)\Sigma_\Lambda^{-1}\Sigma_\Lambda\Sigma_\Lambda^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_{j2}e_{jt}\right)\right)\\
&= -2\sum_{t=k+1}^{k_0} tr\left(\Sigma_\Lambda^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_{j2}e_{jt}\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\lambda}_{i2}^{(k)'}e_{it}\right)\Sigma_\Lambda^{-1}\Sigma_\Lambda\right)\\
&= -2\sum_{t=k+1}^{k_0} tr\left(\Sigma_\Lambda^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_{j2}e_{jt}\right)\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\lambda}_{i2}^{(k)'}e_{it}\right)\Sigma_\Lambda^{-1}(\sum_{jj=1}^{N}\lambda_{jj2}\lambda_{jj2}^{'})\right)\\
&= -2\sigma^2\sum_{t=k+1}^{k_0}\sum_{jj=1}^{N} tr\left(\lambda_{jj2}^{'}\Sigma_\Lambda^{-1}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\lambda_{j2}e_{jt}\right)\sigma^{-2}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\hat{\lambda}_{i2}^{(k)'}e_{it}\right)\Sigma_\Lambda^{-1}\lambda_{jj2}\right)\\
&\xrightarrow{by\ CLT} -2\sigma^2\sum_{t=k+1}^{k_0}\chi_1^2(t),
\end{aligned}
$$

thus D_3 can merge into R_3,

$$
\begin{aligned}
D_3 + R_3 &\xrightarrow{d} -\sigma^2\sum_{t=k+1}^{k_0}\chi_1^2(t)\\
&= C_3\sum_{t=k+1}^{k_0}\chi_1^2(t)
\end{aligned}
$$

In summary, for $k \leq k_0$

$$SSR(k) - SSR(k_0) \xrightarrow{d} (k_0 - k)\psi_1 + (k_0 - k)\psi_2 + \phi_1 \sum_{t=k+1}^{k_0} Z_t + \phi_2 \sum_{t=k+1}^{k_0} Z_t + C_3 \sum_{t=k+1}^{k_0} \chi_1^2(t)$$
$$= (k_0 - k)C_1 + C_2 \sum_{t=k+1}^{k_0} Z_t + C_3 \sum_{t=k+1}^{k_0} \chi_1^2(t)$$

Similarly for $k > k_0$, we can show that

$$SSR(k) - SSR(k_0) \xrightarrow{d} (k - k_0)C_1 + C_2 \sum_{t=k_0+1}^{k} Z_t + C_3 \sum_{t=k_0+1}^{k} \chi_1^2(t)$$

■

B.1.2 Proof of Corollary 3.2.1

Proof

If e_{it} is heteroscedasticity, i.e. $E(e_{it}^2) = \sigma_i^2$, based on our previous results, we can show that

$$
\begin{aligned}
D_{1,hetero} &= \sum_{i=1}^{N} \sum_{t=k+1}^{k_0} \left[(\lambda_{i1} - \lambda_{i2})' f_t \right]^2 \\
&= (k_0 - k) \sum_{i=1}^{N} (\lambda_{i1} - \lambda_{i2})' \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} f_t f_t^{'} (\lambda_{i1} - \lambda_{i2}) \\
&\xrightarrow{N \to \infty} (k_0 - k)\psi_1, \\
D_{2,hetero} &= \sum_{i=1}^{N} \sum_{t=k+1}^{k_0} \left[\sum_{s=k_0+1}^{T} e_{is} f_s^{'} \left(F_2^{'} F_2 \right)^{-1} f_t \right]^2 \\
&\xrightarrow{by\ \ LLN} (k_0 - k) \frac{N}{T - k_0} \frac{1}{N} \sum_{i=1}^{N} tr \left(E(\xi_i \xi_i^{'} \Sigma_f^{-1}) \right) \\
&= (k_0 - k) \frac{N}{T - k_0} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 = (k_0 - k)\psi_2^*, \\
D_{3,hetero} &= \sum_{i=1}^{N} \sum_{t=k+1}^{k_0} \hat{\lambda}_{i2}^{(k)'} \left(\Lambda_2^{'} \Lambda_2 \right)^{-1} \sum_{j=1}^{N} \lambda_{j2} e_{jt} \sum_{j=1}^{N} e_{jt} \lambda_{j2}^{'} \left(\Lambda_2^{'} \Lambda_2 \right)^{-1} \hat{\lambda}_{i2}^{(k)} \\
&= \sum_{t=k+1}^{k_0} \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \lambda_{i2}^{(k)'} \lambda_{i2}^{(k)} \right) \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \lambda_{i2}^{(k)'} \lambda_{i2}^{(k)} \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} \hat{\lambda}_{i2}^{(k)'} \left(\frac{\Lambda_2^{'} \Lambda_2}{N} \right)^{-1} \\
&\left(\frac{1}{\sqrt{N}} \sum_{j=1}^{N} \lambda_{j2} e_{jt} \right) \left(\frac{1}{\sqrt{N}} \sum_{j=1}^{N} e_{jt} \lambda_{j2}^{'} \right) \left(\frac{\Lambda_2^{'} \Lambda_2}{N} \right)^{-1} \hat{\lambda}_{i2}^{(k)} \\
&\xrightarrow{by\ \ CLT} \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \lambda_{i2}^{(k)'} \lambda_{i2}^{(k)} \right) \sum_{t=k+1}^{k_0} \chi_1^2(t)
\end{aligned}
$$

and

$$
\begin{aligned}
R_{1,hetero} &= 2\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}(\lambda_{i1}-\lambda_{i2})' f_t e_{it} \\
&\xrightarrow{by\ CLT} 2\sum_{t=k+1}^{k_0} N\left(0, E(\Delta_i' f_t e_{it}^2 f_t' \Delta_i)\right) \\
&= 2\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 \Delta_i' \Sigma_f \Delta_i \sum_{t=k+1}^{k_0} Z_t \\
&= \phi_1^* \sum_{t=k+1}^{k_0} Z_t \\
R_{2,hetero} &= -2\frac{\sqrt{N}}{\sqrt{T-k_0}}\sum_{t=k+1}^{k_0}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i' \left(\frac{F_2' F_2}{T-k_0}\right)^{-1} f_t e_{it} \\
&\xrightarrow{by\ CLT} -2\frac{\sqrt{N}}{\sqrt{T-k_0}}\sum_{t=k+1}^{k_0} N\left(0, E(\xi_i' \Sigma_f^{-1} f_t e_{it}^2 f_t' \Sigma_f^{-1}\xi_i)\right) \\
&= -2\frac{\sqrt{N}}{\sqrt{T-k_0}}\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^4\right)^{-\frac{1}{2}}\sum_{t=k+1}^{k_0} Z_t \\
&= \phi_2^* \sum_{t=k+1}^{k_0} Z_t \\
R_{3,hetero} &= 2\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}\hat{\lambda}_{i2}^{(k)\prime}(f_t - \hat{f}_t^{(T-k)})e_{it} \\
&\xrightarrow{by\ CLT} -2\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 \lambda_{i2}^{(k)\prime}\lambda_{i2}^{(k)}\right)\sum_{t=k+1}^{k_0}\chi_1^2(t) \\
D_{3,hetero} + R_{3,hetero} &= -\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 \lambda_{i2}^{(k)\prime}\lambda_{i2}^{(k)}\right)\sum_{t=k+1}^{k_0}\chi_1^2(t) \\
&= C_3^* \sum_{t=k+1}^{k_0}\chi_1^2(t)
\end{aligned}
$$

In summary, for $k \le k_0$

$$SSR(k) - SSR(k_0) \xrightarrow{d} (k_0 - k)\psi_1 + (k_0 - k)\psi_2^* + \phi_1^* \sum_{t=k+1}^{k_0} Z_t + \phi_2^* \sum_{t=k+1}^{k_0} Z_t + C_3^* \sum_{t=k+1}^{k_0} \chi_1^2(t)$$

$$= (k_0 - k)C_1^* + C_2^* \sum_{t=k+1}^{k_0} Z_t + C_3^* \sum_{t=k+1}^{k_0} \chi_1^2(t)$$

Similarly for $k > k_0$, we can show that

$$SSR(k) - SSR(k_0) \xrightarrow{d} (k - k_0)C_1^* + C_2^* \sum_{t=k_0+1}^{k} Z_t + C_3^* \sum_{t=k_0+1}^{k} \chi_1^2(t)$$

■