Quasi-maximum likelihood estimation of break point in high-dimensional factor models

Jiangtao Duan\textsuperscript{1}, Jushan Bai\textsuperscript{2}, Xu Han\textsuperscript{3}

\textsuperscript{1}Northeast Normal University, \textsuperscript{2}Columbia University and \textsuperscript{3}City University of Hong Kong

Abstract:

This paper estimates the break point for large-dimensional factor models with a single structural break in factor loadings at a common unknown date. First, we propose a quasi-maximum likelihood (QML) estimator of the change point based on the second moments of factors, which are estimated by principal component analysis. We show that the QML estimator performs consistently when the covariance matrix of the pre- or post-break factor loading, or both, is singular. When the loading matrix undergoes a rotational type of change while the number of factors remains constant over time, the QML estimator incurs a stochastically bounded estimation error. In this case, we establish an asymptotic distribution of the QML estimator. The simulation results validate the feasibility of this estimator when used in finite samples. In addition, we demonstrate empirical applications of the proposed method by applying it to estimate the break points in a U.S. macroeconomic dataset and a stock return dataset.

Key words and phrases: Structural break, High-dimensional factor models, Factor loadings

1. Introduction

Large factor models assume that a few factors can capture the common driving forces of a large number of economic variables. Although factor models are useful, practitioners have to be cautious about the potential structural changes. For example, either the number of factors or the factor loadings may change over time. This concern is empirically relevant because parameter instability is pervasive in large-scale panel data.

So far, many methods have been developed to test structural breaks in factor models (e.g., Stock and Watson (2008), Breitung and Eickmeier (2011), and Chen et al. (2014)). The rejection of the null hypothesis of no structural change leads to the subsequent issues of how to estimate the change point, determine the numbers of pre- and post-break factors, and estimate the factor space. Chen (2015) considers a least-squares estimator of the break point and proves the consistency of the estimated break fraction (i.e., the break date $k$ divided by the full time series $T$). Cheng et al. (2016) propose a shrinkage method to obtain a consistent estimator of the break fraction. Baltagi et al. (2017) develop a least-squares estimator of the change point based on the second moments of the estimated pseudo-factors and show that the estimation
error of the proposed estimator is $O_p(1)$, which indicates the consistency of the estimated break fraction. A few recent studies also explore a consistent estimation of break points, which is technically more challenging. Ma and Su (2018) develop an adaptive fused group Lasso method to consistently estimate all break points under a multibreak setup. Barigozzi et al. (2018) propose a method based on wavelet transformations to consistently estimate the number and locations of break points in the common and idiosyncratic components. Bai et al. (2020) establish the consistency of the least-squares estimator of the break point in large factor models when factor loadings are subjected to a structural break and the size of the break is shrinking as the sample size increases. Although the estimators proposed in these studies are consistent under certain assumptions, the simulation results show that they perform poorly when (1) the number of factors changes after the break or (2) the loading matrix undergoes a rotational type of change.

According to the factor model literature, a factor model with a break in factor loadings is observationally equivalent to that with constant loadings and possibly more pseudo-factors (e.g., Han and Inoue (2015) and Bai and Han (2016)). Thus, the estimation of the change point of factor loadings can be converted into that of the change point of the second moment of the pseudo-factors. We propose a quasi-maximum likelihood (QML) method to estimate the break point based on the second moment of the estimated pseudo-factors; therefore, the number of original factors is not required to be known for computing our estimator. First, we estimate the number of pseudo-factors in an equivalent representation that ignores the break, and then estimate the pre- and post-break second moment matrices of the estimated pseudo-factors for all possible sample splits. The structural break date is estimated by minimizing the QML function among all possible split points.

This paper makes the following contributions to the literature. First, we establish the consistency of the QML break point estimator if the break leads to more pseudo-factors than the original pre- or post-break factors. This occurs when the break augments the factor space or in the presence of disappearing or emerging factors. Under these circumstances, the covariance matrix of loadings on the pre- or post-break pseudo-factors is singular, which is the key condition to establish the consistency of our QML estimator. To the best of our knowledge, this is the first study that links the consistency of the break point estimator to the singularity of covariance matrices of loadings on pre- and post-break pseudo-factors. In addition, we prove that the difference between the estimated and true change points is stochastically bounded when both pre- and post-break loadings on the pseudo-factors have nonsingular covariance matrices. In this case, the loading matrix only undergoes a rotational change, and both the numbers of pre- and post-break original factors are equal to the number of pseudo-factors.

The aforementioned singularity leads to a technical challenge of analyzing the asymptotic property. The singular population covariance matrix of the pre(post)-break loadings has a zero determinant, whose logarithm is not defined appropriately. To resolve this issue, we show that the estimated covariance matrices have nonzero determinants and a well-defined inverse for any given sample size, by obtaining the convergence rate of the lower bound of their smallest eigenvalues. This ensures that the objective function based on the estimated covariance is appropriately defined in any finite sample.

Our second major contribution is that the QML method allows a change in the number of factors. Namely, it allows for disappearing or emerging factors after the break. This is an advantage over the methods developed by Ma and Su (2018) and Bai et al. (2020), who assume that the number of factors remains constant after the break. Our simulation result indicates that the estimator proposed by Bai et al. (2020) is inconsistent when some factors disappear and the remaining factors have time-invariant loadings. Baltagi et al. (2017) allow a change in the number of factors; however, their estimation error was
only stochastically bounded. In contrast, our QML estimator remains consistent under varying number of factors.

Finally, the QML method has a substantial computational advantage over the estimators that iteratively implement high-dimensional principal component analysis (PCA). For example, the estimator proposed by [Bai et al., 2020] runs PCA for pre- and post-split sample covariance matrices for all possible split points. In comparison, our QML runs PCA for the entire sample only once, and thus, is computationally more efficient, especially in large samples.

The rest of this paper is organized as follows. Section 2 introduces the factor model with a single break on the factor loading matrix and describes the QML estimator for the break date. Section 3 presents the assumptions made for this model. Section 4 presents the consistency and asymptotic distribution of the QLM estimator for the break date. Section 5 investigates the finite-sample properties of the QML estimator through simulations. Section 6 implements the proposed method to estimate the break points in a monthly macroeconomic dataset of the United States and a dataset of weekly stock returns of Nasdaq 100 components. Section 7 concludes the study.

2. Model and estimator

Let us consider the following factor model with a common break at \( k_0 \) in the factor loadings for \( i = 1, \cdots, N \):

\[
    x_{it} = \begin{cases} 
    \lambda_{i1} f_t + e_{it} & \text{for } t = 1, 2, \cdots, k_0(T) \\
    \lambda_{i2} f_t + e_{it} & \text{for } t = k_0(T) + 1, \cdots, T,
    \end{cases}
\]

where \( f_t \) is an \( r \)-dimensional vector of unobserved common factors; \( r \) is the number of pseudo-factors; \( k_0(T) \) is the unknown break date; \( \lambda_{i1} \) and \( \lambda_{i2} \) are the pre- and post-break factor loadings, respectively; and \( e_{it} \) is the error term allowed to have serial and cross-sectional dependence as well as heteroskedasticity. \( \tau_0 \in (0, 1) \) is a fixed constant and \([x]\) represents the integer part of \( x \). For notational simplicity, hereinafter, we suppress the dependence of \( k_0 \) on \( T \). Note that we formulate the model using pseudo-factors instead of the original underlying factors. This simplifies the representation of various breaks in a unified framework, which will be clarified in the examples below.

In vector form, model (1) can be expressed as

\[
    x_t = \begin{cases} 
    \Lambda_1 f_t + e_t & \text{for } t = 1, 2, \cdots, k_0 \\
    \Lambda_2 f_t + e_t & \text{for } t = k_0(T) + 1, \cdots, T,
    \end{cases}
\]

where \( x_t = [x_{1t}, \cdots, x_{Nt}]' \), \( e_t = [e_{1t}, \cdots, e_{Nt}]' \), \( \Lambda_1 = [\lambda_{11}, \cdots, \lambda_{N1}]' \), and \( \Lambda_2 = [\lambda_{12}, \cdots, \lambda_{N2}]' \).

For any \( k = 1, \cdots, T - 1 \), we define

\[
    X_k^{(1)} = [x_1, \cdots, x_k]', X_k^{(2)} = [x_{k+1}, \cdots, x_T]', \\
    F_k^{(1)} = [f_1, \cdots, f_k]', F_k^{(2)} = [f_{k+1}, \cdots, f_T], \\
    e_k^{(1)} = [e_1, \cdots, e_k]', e_k^{(2)} = [e_{k+1}, \cdots, e_T],
\]

where the subscript \( k \) denotes the date at which the sample is to be split, and the superscripts (1) and (2) denote the pre-
and post-\( k \) data, respectively. We rewrite (2) using the following matrix representation:

\[
\begin{bmatrix}
X_{k_0}^{(1)} \\
X_{k_0}^{(2)}
\end{bmatrix}
= \begin{bmatrix}
F_{k_0}^{(1)} A_1' \\
F_{k_0}^{(2)} A_2'
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{k_0}^{(1)} \\
\varepsilon_{k_0}^{(2)}
\end{bmatrix}
= \begin{bmatrix}
F_{k_0}^{(1)} (\Lambda B) \\
F_{k_0}^{(2)} (\Lambda C)'
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{k_0}^{(1)} \\
\varepsilon_{k_0}^{(2)}
\end{bmatrix},
\]

\[
= \begin{bmatrix}
F_{k_0}^{(1)} B' \\
F_{k_0}^{(2)} C'
\end{bmatrix} \Lambda' + \begin{bmatrix}
\varepsilon_{k_0}^{(1)} \\
\varepsilon_{k_0}^{(2)}
\end{bmatrix},
\]

\[
= GA + E.
\]

where \( \Lambda \) is an \( N \times r \) matrix with full column rank. The pre- and post-break loadings are modeled as \( \Lambda_1 = \Lambda B \) and \( \Lambda_2 = \Lambda C \), respectively, where \( B \) and \( C \) are some \( r \times r \) matrices. In this model, \( r_1 = rank(B) \leq r \) and \( r_2 = rank(C) \leq r \) denote the numbers of original factors before and after the break, respectively. To distinguish them from the original factors, we refer to \( G \) as the pseudo-factors in (3) and \( rank(G) = r \). Hence, the last line of (3) provides an observationally equivalent representation with constant loadings \( \Lambda \) and \( r \) pseudo-factors \( G \). It is well known that the break can augment the factor space; thus, \( r_1 \leq r \) and \( r_2 \leq r \). \( F_{k_0}^{(1)} \) and \( F_{k_0}^{(2)} \) have dimensions \( k_0 \times r \) and \( (T-k_0) \times r \), respectively, and \( \Lambda_1 \) and \( \Lambda_2 \) have dimension \( N \times r \). Our representation in (3) allows for changes in the factor loadings and the number of factors. Below, several examples are provided to illustrate that the pseudo-factor representation in (3) is general enough to cover three types of breaks.

**Type 1.** Both \( B \) and \( C \) are singular. In this case, the number of original factors is strictly less than that of the pseudo-factors both before and after the break (i.e., \( r_1 < r \) and \( r_2 < r \)). This means that the structural break in the factor loadings augments the dimension of the factor space. Let us consider the following example.

Example (1): Let \( F_{k_0}^{(1)}(k_0 \times r_1) \) and \( F_{k_0}^{(2)}((T-k_0) \times r_2) \) denote the original factors before and after the break, respectively, and \( \Theta_1 \) and \( \Theta_2 \) denote the pre- and post-break loadings on these factors. Thus, this model can be represented and transformed as

\[
\begin{bmatrix}
X_{k_0}^{(1)} \\
X_{k_0}^{(2)}
\end{bmatrix}
= \begin{bmatrix}
F_{k_0}^{(1)} \Theta_1' \\
F_{k_0}^{(2)} \Theta_2'
\end{bmatrix} + \varepsilon = \begin{bmatrix}
\varepsilon_{k_0}^{(1)} \\
\varepsilon_{k_0}^{(2)}
\end{bmatrix} + \begin{bmatrix}
\Theta_1' \\
\Theta_2'
\end{bmatrix} + \varepsilon
\]

\[
= \begin{bmatrix}
[F_{k_0}^{(1)}':]B' \\
[F_{k_0}^{(2)}':]C'
n_{k_0}^{(1)}: \varepsilon_{k_0}^{(2)}
\end{bmatrix} \Lambda' + \begin{bmatrix}
\Theta_1' \\
\Theta_2'
\end{bmatrix} + \varepsilon,
\]

\[
= \begin{bmatrix}
F_{k_0}^{(1)} B' \\
F_{k_0}^{(2)} C'
\end{bmatrix} \Lambda' + \varepsilon,
\]

where \( \Lambda = [\Theta_1, \Theta_2] \), \( B = diag(I_{r_1}, 0_{r_2 \times r_2}) \), \( C = diag(0_{r_1 \times r_1}, I_{r_2}) \), \( F_{k_0}^{(1)} = [F_{k_0}^{(1)}']' \), \( F_{k_0}^{(2)} = [F_{k_0}^{(2)}]' \), and the asterisk denotes some unidentified numbers such that all rows in \( F_{k_0}^{(1)} \) and \( F_{k_0}^{(2)} \) have the same variance (to satisfy Assumption 4 in Section 3). In the special case of \( r_1 = r_2 \), \( \Lambda \) is of full rank \( 2r_1 \) (i.e., the dimension of the pseudo-factor space is twice that of the original factor space) if the shift in the loading matrix \( \Theta_2 - \Theta_1 \) is linearly independent of \( \Theta_1 \). We refer to this special case as the shift type of change, because the augmentation of the factor space is induced by a linearly independent shift in the loading matrix. Hence, Type 1 covers the shift type of change.

**Type 2.** Only \( B \) or \( C \) is singular. In this case, emerging or disappearing factors are present in the model. Let us consider the following example of disappearing factors.
We consider Type 2 separately to emphasize the case of emerging and disappearing factors.

Thus, the last \( r \) factors disappear after the break. Therefore, we can obtain the pseudo-factors by using the following transformation from the original factors \( F \):

\[
\begin{bmatrix}
X^{(1)}_{k_0} \\
X^{(2)}_{k_0}
\end{bmatrix}
= \begin{bmatrix}
F^{(1)}_{k_0} \Theta'_1 \\
F^{(2)}_{k_0} \Theta'_2
\end{bmatrix}
+ \epsilon = \begin{bmatrix}
F^{(1)}_{k_0} \Theta'_1 \\
F^{(2)}_{k_0} C' \Theta'_1
\end{bmatrix}
+ \epsilon
\]

where \( F^{(1)}_{k_0} = F^{(1)}_{k_0} \), \( F^{(2)}_{k_0} = F^{(2)}_{k_0} \cdot \), \( C = diag(I_{r_2}, \Theta_{(r_1-r_2) \times (r_1-r_2)}) \), \( \Lambda = \Theta_1 \), and the asterisk is defined in a similar manner to that in (3). In this example, \( B = I_{r_1} \), \( r = r_1 \), and \( r_2 = \text{rank}(C) < r \). Symmetrically, if \( B \) is singular and \( C = I_{r_2} \), then \( r_2 = r \) and \( r_1 = \text{rank}(B) < r \), which means that certain factors emerge after the break point. Type 2 changes are important in empirical analysis. Please refer to McAlinn et al. (2018) for empirical evidence regarding the varying number of factors in the U.S. macroeconomic dataset. For Types 1 and 2, we obtain a significant result that \( P(k-k_0=0) \rightarrow 1 \) as \( N, T \rightarrow \infty \).

**Type 3.** Both \( B \) and \( C \) are nonsingular. In this case, the loadings on the original factors undergo a rotational change, and the dimension of the original factors is the same as that of the pseudo-factors.

Example (3): Let us assume that \( r_2 = r_1 \) and \( \Theta_2 = \Theta_1 C \) for a nonsingular matrix \( C \). The model with the original factors \( F \) can be transformed into the following pseudo-factor representation:

\[
\begin{bmatrix}
X^{(1)}_{k_0} \\
X^{(2)}_{k_0}
\end{bmatrix}
= \begin{bmatrix}
F^{(1)}_{k_0} \Theta'_1 \\
F^{(2)}_{k_0} C' \Theta'_1
\end{bmatrix}
+ \epsilon = \begin{bmatrix}
F^{(1)}_{k_0} \Theta'_1 \\
F^{(2)}_{k_0} C' \Theta'_1
\end{bmatrix}
+ \epsilon
\]

where \( F^{(1)}_{k_0} = F^{(1)}_{k_0} \), \( F^{(2)}_{k_0} = F^{(2)}_{k_0} \), and \( \Lambda = \Theta_1 \). In this example, \( B = I_{r_1} \) and \( r = r_1 = r_2 \), and the factor dimension remains constant. In the observationally equivalent pseudo-factor representation, the loading is time-invariant and the original post-break factors \( F^{(2)}_{k_0} \) are rotated by \( C \). We refer to this as the rotation type of change.

The above examples show that a factor model with any of these three types of change can be unified and reformulated by the representation in (3) with pseudo-factors. This representation controls the break type by varying the settings for \( B \) and \( C \), and thus, is convenient for our theoretical analysis.

Bai et al. (2020) rule out the rotation type of change because the break date is not identifiable by minimizing the sum of squared residuals. Baltagi et al. (2017) allow changes in the number of factors and rotation type of change; however, the difference between their estimator and the true break point is only stochastically bounded (i.e., their estimator is not consistent). Ma and Su’s (2018) setup requires \( r_1 = r_2 \); thus, Type 2 is ruled out under their assumptions. Our simulation result shows that Ma and Su’s estimator does not perform well under rotational changes (Type 3), whereas our QML method

\footnote{Technically, Types 1 and 2 can be combined into one type that involves singularity, which renders our QML estimator consistent. We consider Type 2 separately to emphasize the case of emerging and disappearing factors.}
can handle changes in all three types discussed above. We obtain a significant result that \( \hat{k} - k_0 = O_p(1) \) if both \( B \) and \( C \) are of full rank (i.e., Type 3) and \( \hat{k} - k_0 = o_p(1) \) if \( B \) or \( C \), or both, is singular (i.e., Type 1 and Type 2).

In this paper, we consider the QML estimator of the break date for model (3):

\[
\hat{k} = \arg \min_{[\tau_1 T] \leq k \leq [\tau_2 T]} U_{NT}(k),
\]

where \([\tau_1 T]\) and \([\tau_2 T]\) denote the prior lower and upper bounds for the real break point \( k_0 \) with \( \tau_1, \tau_2 \in (0, 1) \) and \( \tau_1 \leq \tau_0 \leq \tau_2 \). The QML objective function \( U_{NT}(k) \) is equal to

\[
U_{NT}(k) = k \log(\det(\hat{\Sigma}_2)) + (T - k) \log(\det(\hat{\Sigma}_2)),
\]

where \( \hat{\Sigma}_1 \) and \( \hat{\Sigma}_2 \) can be defined as

\[
\hat{\Sigma}_1 = \frac{1}{k} \sum_{t=1}^{k} \hat{g}_t \hat{g}_t',
\]

\[
\hat{\Sigma}_2 = \frac{1}{T-k} \sum_{t=k+1}^{T} \hat{g}_t \hat{g}_t',
\]

and \( \hat{g}_t \) is the PCA estimator of \( g_t \) (i.e., the transpose of the \( t \)-th row of \( G \)). We define \( \Sigma_{G,1} = E(g_t g_t') \) for \( t \leq k_0 \), \( \Sigma_{G,2} = E(g_t g_t') \) for \( t > k_0 \), and \( \Sigma_G = \tau_0 \Sigma_{G,1} + (1 - \tau_0) \Sigma_{G,2} \), where \( \Sigma_\Lambda \) is the covariance matrix of \( \Lambda \). The PCA estimator \( \hat{g}_t \) is asymptotically close to \( H' g_t \) for a rotation matrix \( H \), and \( H \xrightarrow{p} \Sigma_\Lambda^{1/2} \Phi V^{-1/2} \) as \( (N, T) \to \infty \), where \( V \) and \( \Phi \) are the eigenvalue and eigenvector matrices of \( \Sigma_\Lambda^{1/2} \Sigma_G \Sigma_\Lambda^{1/2} \), respectively. Evidently, the second moment of \( H_0 \hat{g}_t \) shares the same change point as that of \( g_t \). Therefore, we proceed to estimate the pre- and post-break second moments of \( g_t \) by using the estimated factors \( \hat{g}_t \), and then use (7) to obtain the QML break point estimator \( \hat{k}_{QML} \). Similar QML objective functions have been used for multivariate time series with observed data (e.g., [Bas (2000)]).

3. Assumptions

In this section, we state the assumptions made for validating the consistency and asymptotic distribution of the QML estimator.

Assumption 1. (i) \( E\|f_t\|^4 < M < \infty \), \( E(f_t f_t') = \Sigma_F \), where \( \Sigma_F \) is positive definite, and \( \frac{1}{N} \sum_{t=1}^{k_0} f_t f_t' \xrightarrow{p} \Sigma_F \), \( \frac{1}{T-k_0} \sum_{t=k_0+1}^{T} f_t f_t' \xrightarrow{p} \Sigma_F \).

(ii) There exists \( d > 0 \) such that \( \|\Delta\| \geq d > 0 \), where \( \Delta = B \Sigma_F B' - C \Sigma_F C' \) and \( B, C \) are \( r \times r \) matrices.

Assumption 2. \( \|\lambda_{ii}\| \leq \lambda < \infty \) for \( i = 1, 2, \ldots, N \), \( \frac{1}{\delta} \lambda - \Sigma_\Lambda \to 0 \) for some \( r \times r \) positive definite matrix \( \Sigma_\Lambda \).

Assumption 3. There exists a positive constant \( M < \infty \) such that

(i) \( E(e_{it}) = 0 \) and \( E|e_{it}|^\delta \leq M \) for all \( i = 1, \ldots, N \) and \( t = 1, \ldots, T \);

(ii) \( E(\frac{\sum_{i=1}^{N} e_{it} e_{it}}{N}) = E(N^{-1} \sum_{i=1}^{N} e_{it} e_{it}) = \gamma_N(s, t) \) and \( \sum_{s=1}^{T} |\gamma_N(s, t)| \leq M \) for every \( t \leq T \);

(iii) \( E(e_{it} e_{jt}) = \tau_{ij}, t \) with \( |\tau_{ij,t}| < \tau_{ij} \) for some \( \tau_{ij} \) and for all \( t = 1, \ldots, T \) and \( \sum_{j=1}^{N} |\tau_{ij}| \leq M \) for every \( i \leq N \).
(iv) \( E(e_{it}e_{jt}) = \tau_{ij,ts} \),
\[
\frac{1}{NT} \sum_{i,j,t,s=1} |\tau_{ij,ts}| \leq M;
\]

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

**Assumption 4.** There exists a positive constant \( M < \infty \) such that
\[
E \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{k_0}} \sum_{t=1}^{k_0} f_{e_{it}} \right\|^2 \right) \leq M,
\]
\[
E \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T-k_0}} \sum_{t=k_0+1}^{T} f_{e_{it}} \right\|^2 \right) \leq M.
\]

**Assumption 5.** The eigenvalues of \( \Sigma_G \Sigma \) are distinct.

**Assumption 6.** Let us define \( e_t = f_t' - \Sigma_P \). According to the data-generating process (DGP) of factors, the Hájek-Rényi inequality applies to the processes \( \{e_t, t = 1, \cdots, k_0\}, \{e_t, t = k_0, \cdots, 1\} \), \( \{e_t, t = k_0 + 1, \cdots, T\} \), and \( \{e_t, t = T, \cdots, k_0 + 1\} \).

**Remark 1.** Using the Hájek-Rényi inequality on \( e_t \), we can ensure that \( \max_{k_0 < k \leq \lceil T/2 \rceil} \left\| \frac{1}{k_0} \sum_{t=k_0+1}^{T} g(t) \right\| = O_p(1) \), \( \max_{k_0 < k \leq \lceil T/2 \rceil} \left\| \frac{1}{k_0} \sum_{t=k_0+1}^{T} g(t) \right\| = O_p(1) \) in Lemmas 7 and 8.

**Assumption 7.** There exists an \( M < \infty \) such that

(i) For each \( s = 1, \cdots, T \),
\[
E \left( \max_{k < k_0} \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is}e_{it} - E(e_{is}e_{it})] \right\|^2 \right) \leq M,
\]
\[
E \left( \max_{k \leq k_0} \frac{1}{k} \sum_{t=1}^{k} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is}e_{it} - E(e_{is}e_{it})] \right\|^2 \right) \leq M,
\]
\[
E \left( \max_{k > k_0} \frac{1}{k-k_0} \sum_{t=k_0+1}^{k} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is}e_{it} - E(e_{is}e_{it})] \right\|^2 \right) \leq M,
\]
\[
E \left( \max_{k \geq k_0} \frac{1}{T-k} \sum_{t=k+1}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is}e_{it} - E(e_{is}e_{it})] \right\|^2 \right) \leq M.
\]

(ii) 
\[
E \left( \max_{k < k_0} \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} \right\|^2 \right) \leq M,
\]
\[
E \left( \max_{k \leq k_0} \frac{1}{k} \sum_{t=1}^{k} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} \right\|^2 \right) \leq M,
\]
\[
E \left( \max_{k > k_0} \frac{1}{k-k_0} \sum_{t=k_0+1}^{k} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} \right\|^2 \right) \leq M,
\]
\[
E \left( \max_{k \geq k_0} \frac{1}{T-k} \sum_{t=k+1}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_i e_{it} \right\|^2 \right) \leq M.
\]
Assumption 8. There exists an $M < \infty$ such that for all values of $N$ and $T$,

(i) for each $t$,

\[
E \left( \max_{t_1 \leq k < k_0} \frac{1}{T} \sum_{t=k+1}^{k_0} \left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{T} \sum_{i=1}^{N} f_s \left[ e_{is} e_{it} - E(e_{is} e_{it}) \right] \right\|^2 \right) \leq M;
\]

\[
E \left( \max_{k_0 < k \leq \lfloor r_2 T \rfloor} \frac{1}{T-k} \sum_{t=k+1}^{T} \left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^{T} \sum_{i=1}^{N} f_s \left[ e_{is} e_{it} - E(e_{is} e_{it}) \right] \right\|^2 \right) \leq M;
\]

(ii) the $r \times r$ matrix satisfies

\[
E \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} f_i \lambda_t' e_{it} \right\|^2 \leq M;
\]

4. Asymptotic properties of the QML estimator

In this section, we derive the asymptotic properties of the QML estimator for various breaks. In the literature of structural breaks for a fixed-dimensional time series, conventional break point estimators, such as the least-squares (LS) estimator of Bai (1997) or the QML estimator of Qu and Perron (2007), are usually inconsistent. The estimation error of these conventional estimators is $O_p(1)$ when the break size is fixed. To reach consistency, the cross-sectional dimension of the time series must be large (e.g., Bai (2010) and Kim (2011)).

Recall that the observationally equivalent representation in (3) has time-invariant loadings and varying pseudo-factors. Hence, our problem converges to estimating the break point in the $r$-dimensional time series $g_t$, where $r$ is fixed. Theorems 1 and 2 below show that, for rotational breaks (Type 3), the convergence rate and limiting distribution are similar to those available in the literature. However, for Type 1 and 2 breaks, Theorem 3 derives a much more significant result than that available in the literature, according to which our QML estimator is consistent even if our $g_t$ has only a fixed cross-sectional dimension $r$.

Theorem 1. Under Assumptions 1–8 when both $B$ and $C$ are of full rank, $\hat{k} - k_0 = O_p(1)$.

This theorem implies that the difference between the QML estimator and the true change point is stochastically bounded in model (6). Although the estimation errors of BKW and QML methods both are bounded, the QML estimator has much better finite sample properties. To confirm this theoretical result, we conduct a simulation where the factor loadings have rotational change (see DGP 1.B in Section 5). Table 2 presents the MAEs and RMSEs of different estimators. The simulation result shows that the QML estimators have much smaller MAEs and RMSEs than other methods. In addition, $\hat{k}$ does not collapse to $k_0$, leading to a nondegenerate distribution. We will state the limiting distribution in Theorem 2. Nevertheless, this theorem shows that the break point can be appropriately estimated because $\hat{\tau} = \hat{k}/T$ is still consistent for $\tau_0$.

Remark 2. Recall that $\Delta = \check{C}' \Sigma_F C - B' \Sigma_F B$; thus, $\|\Delta\|$ represents the magnitude of the break. Note that the Baltagi et al. (2017) estimator comprises stochastically bounded estimation errors, and is not consistent even if the magnitude of the break is large. In contrast, the QML estimator remains consistent with an increasing $\|\Delta\|$. In fact, the proof indicates that $U(k) - U(k_0) \to \infty$ for $k \neq k_0$ as $\|\Delta\| \to \infty$; thus, the consistency of the QML estimator can be obtained. As it is not common to consider a diverging break size in empirical applications, we do not analyze this case in the present paper.
To make an inference regarding the change point when both $B$ and $C$ are of full rank, we derive the limiting distribution of $\hat{k}$. Let us define

$$
\xi_t = H_0' g_t g_t' H_0 - \Sigma_1 \text{ for } t \leq k_0,
$$
$$
\xi_t = H_0' g_t g_t' H_0 - \Sigma_2 \text{ for } t > k_0,
$$

where $\Sigma_1 = H_0' \Sigma_{G,1} H_0$ and $\Sigma_2 = H_0' \Sigma_{G,2} H_0$ are the pre- and post-breaks of $H_0' E(g_t g_t') H_0$. The limiting distribution of $\hat{k}$ is given by the following theorem:

**Theorem 2.** Under Assumptions 1–8 when both $B$ and $C$ are of full rank,

$$
\hat{k} - k_0 \overset{d}{\to} \arg \min_{\ell} W(\ell),
$$

where

$$
W(\ell) = \sum_{t=k_0+\ell}^{k_0-1} tr((\Sigma_2^{-1} - \Sigma_1^{-1})\xi_t) - (tr(\Sigma_1^{-1} \Sigma_2^{-1}) - r - \log |\Sigma_1^{-1} \Sigma_2^{-1}|) \ell
$$

for $\ell = -1, -2, \ldots$,

$$
W(\ell) = 0 \text{ for } \ell = 0,
$$

$$
W(\ell) = \sum_{t=k_0+1}^{k_0+\ell} tr((\Sigma_2^{-1} - \Sigma_1^{-1})\xi_t) + (tr(\Sigma_1^{-1} \Sigma_2^{-1}) - r - \log |\Sigma_1^{-1} \Sigma_2^{-1}|) \ell
$$

for $\ell = 1, 2, \ldots$.

This result shows that the limiting distribution depends on $\xi_t$. If $\xi_t$ is independent over time, then $W(\ell)$ is a two-sided random walk. If $f_t$ is stationary, then $\xi_t$ is stationary in each regime. Here, the limiting distribution of the estimated break date is dependent on the generation processes of the unobserved factors, and thus, cannot be directly used to construct a confidence interval for a true break point. Bai et al. (2020) propose a bootstrap method to construct a confidence interval for $k_0$ when the change in the factor loading matrix shrinks as $N \to \infty$. However, their bootstrap procedure lacks robustness in the cross-sectional correlation in the error terms. In the current setup, the break magnitude $\|\Sigma_2 - \Sigma_1\|$ is fixed and we leave the case of shrinking break magnitude as a future topic.

Next, we establish a much stronger result than that available in the literature, which states that the QML estimator remains consistent when $B$ or $C$, or both, is singular. We make the following additional assumptions.

**Assumption 9.** With probability approaching one (w.p.a.1), the following inequalities hold:

$$
0 < c \leq \min_{[r_1 T] \leq k \leq k_0} \rho_e \left( \frac{1}{Nk} \sum_{t=1}^{k} \Lambda_0 e_t e'_t \Lambda_0 \right) \leq \overline{c} < +\infty
$$

$$
0 < c \leq \min_{k_0 \leq k \leq [r_2 T]} \rho_e \left( \frac{1}{N(T-k)} \sum_{t=k+1}^{T} \Lambda_0 e_t e'_t \Lambda_0 \right) \leq \overline{c} < +\infty
$$

as $N, T \to \infty$, where $c$ and $\overline{c}$ are some constants.
Note that Assumption 10. matrices. Proposition 1 provides a useful tool to establish the consistency of our QML estimator. Although this technical (i) Assumption 11. result is a byproduct in our analysis, we believe that it is of independent interest and useful in other contexts.

Under Assumptions 1–10, for there exist constants such that \( \rho_j(\hat{\Sigma}_1(k)) \geq \frac{c_L}{N} \) → 1, \( \rho_j(\hat{\Sigma}_2(k)) \geq \frac{c_L}{N} \) → 1, for \( j = r_2 + 1, \ldots, r \).

In proposition 1, the lower bound of the smallest eigenvalues of the estimated sample covariance matrices \( \hat{\Sigma}_1 \) and \( \hat{\Sigma}_2 \) is \( c_L/N \) for a constant \( c_L > 0 \) w.p.a.1. This ensures a lower bound for the determinants of the estimated sample covariance matrices. Proposition 1 provides a useful tool to establish the consistency of our QML estimator. Although this technical result is a byproduct in our analysis, we believe that it is of independent interest and useful in other contexts.

Assumption 11. (i) \( [B, C] \) row full rank.

(ii) \( C^b B f_{k_0} \neq 0 \) when \( r - r_2 = 1 \) and \( B^c C f_{k_0+1} \neq 0 \) when \( r - r_1 = 1 \), where \( \Lambda^\# \) denotes the adjoint matrix for the singular matrix \( \Lambda \).

(iii) \( \|B f_{k_0} - \text{Proj}(B f_{k_0}|C)\| d > 0 \) when \( r - r_2 \geq 2 \) or \( r_2 = 0 \) and \( \|C f_{k_0+1} - \text{Proj}(C f_{k_0+1}|B)\| d > 0 \) when \( r - r_1 \geq 2 \) or \( r_1 = 0 \), where \( \text{Proj}(\Lambda|\mathbb{Z}) \) denotes the projection of \( \Lambda \) onto the \( \mathbb{Z} \) columns and \( d \) is a constant.

\[ \text{rank}(\Sigma_G) = \text{rank} \begin{bmatrix} \sqrt{\sigma B}, \sqrt{1 - \tau_0} C \end{bmatrix} \begin{bmatrix} \Sigma_F, \sqrt{\tau_0} B' \sqrt{1 - \tau_0} C' \end{bmatrix} = \text{rank} \begin{bmatrix} \sqrt{\sigma B}, \sqrt{1 - \tau_0} C \end{bmatrix} \begin{bmatrix} \Sigma_F, \sqrt{1 - \tau_0} C' \end{bmatrix} = \text{rank} \begin{bmatrix} \sqrt{\sigma B}, \sqrt{1 - \tau_0} C \end{bmatrix} \begin{bmatrix} \Sigma_F, \sqrt{1 - \tau_0} C' \end{bmatrix} \]

and \( 1 < \tau_0 < 1 \), \( [B, C] \) row full rank such that \( \Sigma_G \) is a positive definite matrix.
Assumption III(i) implies that $\Sigma_G$ is positive definite, and $B^*C \neq 0$ when $r - r_1 = 1$, and $C^*B \neq 0$ when $r - r_2 = 1$. Assumption III(ii) is to exclude the possibility that $f_{k_0}$ and $f_{k_0+1}$ in the null space of $C^*B$ and $B^*C$, respectively, then the specific low bound of $|\hat{\Sigma}_2(k)|$ with respect to $k_0 - k$ can be obtained when $k < k_0$ and $k_0 - k$ is bounded as $N,T \to \infty$. Similarly, Assumption III(iii) also exclude the possibility that $Bf_{k_0}$ in the column space of $C$ when $r - r_2 \geq 2$ or $r_2 = 0$ and $Cf_{k_0+1}$ in the column space of $B$ when $r - r_1 \geq 2$ or $r_1 = 0$, then the specific low bound of $|\hat{\Sigma}_2(k)|$ with respect to $k_0 - k$ can be obtained when $k < k_0$ and $k_0 - k$ is divergent as $N,T \to \infty$. Assumption III was used to establish Lemma 8, which is useful for validating the consistency result that $\text{Prob}(\hat{k} - k = 0) \to 1$ in the proof of Theorem 3. If the factor $f_{k_0}$ and $f_{k_0+1}$ have continuous probability distribution functions, then Assumption III(ii)-(iii) are to exclude a zero probability event since $C^*B$ and $B^*C$ are not equal to 0. From another perspective, Assumption III(ii)-(iii) allow $f_t$ have various data generating process.

**Theorem 3.** Under Assumptions III and $\frac{N}{T} \to \kappa$, as $N,T \to \infty$ for $0 < \kappa < \infty$, when $B$ or $C$, or both, is singular, $\text{Prob}(\hat{k} - k = 0) \to 1$.

Theorem 3 shows that the estimated change point converges to the true change point w.p.a.1 when $B$ or $C$, or both, is singular (Types 1 and 2 in Section 2). This result is much more significant than that obtained by Baltagi et al. (2017), who showed that the distance between the estimated and true break dates is bounded for Types 1–3. Note that the case in which only $B$ (or $C$) is singular corresponds to Type 2 with emerging (or disappearing) factors. Our QML estimator is consistent under this type of change, whereas Bai et al. (2020) and Ma and Su (2018) rule out this type by assumption. In empirical applications, the conditions of theorem 3 are rather flexible and likely to hold and the consistency of the break date estimator is expected in most economic data for the factor analysis.

**Remark 3.** An important contribution of Theorem 3 is to link the consistency of the QML estimator with the singularity of the covariance matrices of the pre- or post-break factor loadings. The conditions that $B$ or $C$, or both, is singular and $\frac{N}{T} \to \kappa \in (0, \infty)$ are likely to hold in many economic datasets for factor analysis. If both $B$ and $C$ are singular, the break occurs such that the number of pseudo-factors in the entire factor model is larger than that of the factors in the pre- and post-break subsamples. For example, if all factors undergo large breaks in their loadings, the number of factors tends to be doubled (see Breitung and Eickmeier (2011)). If $B$ is of full rank and $C$ is singular, some factors become useless, and thus, the loading coefficients attached to these disappearing factors become zero. For example, in the momentum portfolio, some risks are not part of the firm’s long-run structure as the sorting is only based on recent returns works; the reward is high but disappears within less than a year. If $B$ is singular and $C$ is of full rank, some factors emerge after the break date, which increases the dimension of the post-break factor space. For example, changes in the technology or policy may produce certain new factors.

**Remark 4.** Theorem 3 indicates that $U_{NT}(k)$ can be minimized to consistently estimate $k_0$. The intuition for this is that $U_{NT}(k) - U_{NT}(k_0)$ is always larger than zero, even if $k$ deviates only slightly from the true break point $k_0$, so that $\hat{k}$ must be equal to $k_0$ to minimize $U_{NT}(k) - U_{NT}(k_0)$. For example, in Type 1, when both $B$ and $C$ are singular for $k < k_0$, we can decompose $\hat{\Sigma}_2$ as $\hat{\Sigma}_2 = \frac{1}{T-k} \sum_{t=k+1}^{k_0} \hat{g}_t \hat{g}_t' + \frac{1}{T-k} \sum_{t=k_0+1}^{T} \hat{g}_t \hat{g}_t'$, and the term $\frac{1}{T-k} \sum_{t=k+1}^{k_0} \hat{g}_t \hat{g}_t'$ enlarges the determinant of $\hat{\Sigma}_2$. By symmetry, we obtain a similar result for $k > k_0$. Thus, $U_{NT}(k) - U_{NT}(k_0) > 0$ w.p.a.1 as $N,T \to \infty$ if $k \neq k_0$. 


Remark 5. With the QML estimator, we do not need to know the numbers of original factors \( r_1 \) and \( r_2 \) before and after the break point, but only the number of pseudo-factors in the entire sample. Bai et al. (2020) and Ma and Su (2018) require knowledge of the number of original factors, which is much more difficult to estimate due to the augmented factor space resulting from the break. In practice, the number of pseudo-factors is much easier to estimate by using one of a number of estimators, such as the information criteria developed by Bai and Ng (2002).

5. Simulation

In this section, we consider DGPs corresponding to Types 1–3 to evaluate the finite sample performance of the QML estimator. We compare the QML estimator with three other estimators. As shown below, \( \hat{k}_{BKW} \) is the estimator proposed by Baltagi, Kao, and Wang (2017, BKW hereafter); \( \hat{k}_{BHS} \) is the estimator proposed by Bai, Han, and Shi (2020, BHS hereafter); \( \hat{k}_{MS} \) is the estimator proposed by Ma and Su (2018, MS hereafter); and \( \hat{k}_{QML} \) is the QML estimator by Barigozzi et al. (2018). We develop a change point estimator using wavelet transformation, which exhibits similar performance to that of the estimator proposed by Ma and Su (2018). Hence, the comparison with the estimator proposed by Barigozzi et al. (2018) is not reported here, but the result is available upon request. The DGP roughly follows BKW, which can be used to examine various elements that may affect the finite sample performance of the estimators, and we use this DGP for model \( k_3 \). We calculate the root mean square error (RMSE) and mean absolute error (MAE) of these change point estimators \( \hat{k}_{BKW}, \hat{k}_{BHS}, \) and \( \hat{k}_{QML} \), and each experiment is repeated 1000 times, where

\[
RMSE = \sqrt{\frac{1}{1000} \sum_{s=1}^{1000} (\hat{k}_s - k_0)^2}
\]

\[
MAE = \frac{1}{1000} \sum_{s=1}^{1000} |\hat{k}_s - k_0|.
\]

When \( T \) is small, there is a possibility that Ma and Su’s (2018) method detects no break or multiple breaks; thus, the definition of the estimation error for a single break point in such cases is not straightforward. For a comparison, we compute the RMSE and MAE of the MS estimator by only using the results obtained by the MS estimator when it successfully detects a single break. As the computation of \( \hat{k}_{BHS} \) and \( \hat{k}_{MS} \) requires the number of original factors and that of \( \hat{k}_{BKW} \) and \( \hat{k}_{QML} \) requires the number of pseudo-factors, we set \( \hat{r} = r_0 \) for \( \hat{k}_{BHS} \) and \( \hat{k}_{MS} \) and \( \hat{r} = r \) for \( \hat{k}_{QML} \) and \( \hat{k}_{BKW} \), where \( r_0 \) is the number of original factors and \( r \) is the number of pseudo-factors.

We generate factors and idiosyncratic errors using a DGP similar to that of BKW. Each factor is generated by the following AR(1) process:

\[
f_{t,p} = \rho f_{t-1,p} + u_{t,p}, \quad \text{for} \quad t = 2, \cdots, T; \quad p = 1, \cdots, r_0,
\]

where \( u_t = (u_{t,1}, \cdots, u_{t,r_0})' \) is i.i.d. \( N(0, I_{r_0}) \) for \( t = 2, \cdots, T \) and \( f_1 = (f_{1,1}, \cdots, f_{1,r_0})' \) is i.i.d. \( N(0, \frac{1}{\rho^2}I_{r_0}) \). The scalar \( \rho \) captures the serial correlation of factors, and the idiosyncratic errors are generated by

\[
e_{t,i} = \alpha e_{t-1,i} + v_{t,i}, \quad \text{for} \quad i = 1, \cdots, N \quad t = 2, \cdots, T,
\]

where \( v_t = (v_{t,1}, \cdots, v_{t,N})' \) is i.i.d. \( N(0, \Omega) \) for \( t = 2, \cdots, T \) and \( e_1 = (e_{1,1}, \cdots, e_{N,1})' \) is \( N(0, \frac{1}{1-\alpha^2}\Omega) \). The scalar \( \alpha \) captures the serial correlation of the idiosyncratic errors, and \( \Omega \) is generated as \( \Omega_{ij} = \beta |i-j| \) so that \( \beta \) captures the degree of cross-sectional dependence of the idiosyncratic errors. In addition, \( u_t \) and \( v_t \) are mutually independent for all values of \( t \). We set \( r_0 = 3 \) and \( k_0 = T/2 \). We consider the following DGPs for factor loadings and investigate the performance of the QML estimator for the three types of breaks discussed in Section 2.
DGP 1.A We first consider the case in which $C$ is singular, and set $C = [1, 0, 0; 0, 1, 0; 0, 0, 0]$. This setup aims to model $\mathbf{\Lambda}$. In the pre-break regime, all elements of $\lambda_{i, 1}$ are i.i.d. $N(0, \frac{1}{2} I_{r_0})$ across $i$. In the post-break regime, $\Lambda_2 = (\lambda_{1, 2}, \cdots, \lambda_{N, 2})' = \Lambda_1 C$. This case corresponds to a Type 2 change with a disappearing factor. The number of pseudo-factors is the same as $r_0$, so $r = 3$, and the numbers of pre- and post-break factors are 3 and $\text{rank}(C) = 2$, respectively. Table 4 lists the RMSEs and MAEs of three estimators for different values of $(\rho, \alpha, \beta)$. In all cases, $\hat{k}_{QML}$ has much smaller MAEs and RMSEs than $\hat{k}_{BKW}$ and $\hat{k}_{BHS}$. Moreover, the MAEs and RMSEs of $\hat{k}_{QML}$ tend to decrease as $N$ and $T$ increase. This confirms the consistency of $\hat{k}_{QML}$ established in Theorem 3. In addition, the RMSEs and MAEs of $\hat{k}_{BKW}$ do not converge to zero as $N$ and $T$ increase, which confirms that $\hat{k}_{BKW}$ has a stochastically bounded estimation error. $\hat{k}_{BHS}$ does not appear to be consistent when a factor disappears after the break. Moreover, a larger AR(1) coefficient $\rho$ tends to deteriorate the performance of $\hat{k}_{BKW}$, but does not have much impact on our QML estimator.

DGP 1.B We next consider the case in which $C$ is of full rank. We set $C$ as a lower triangular matrix. The diagonal elements are equal to 0.5, 1.5, and 2.5, and the elements below these diagonal elements are i.i.d. and drawn from a standard normal distribution. Under this DGP, we have $r = r_0$. Table 2 reports the performance of three estimators for different values of $(\rho, \alpha, \beta)$. In all cases, $\hat{k}_{BKW}$ and $\hat{k}_{QML}$ appear to have stochastically bounded estimation errors, which confirms Theorem 1 of BKW and Theorem 1 of this paper. Both $\hat{k}_{QML}$ and $\hat{k}_{BKW}$ are inconsistent under this DGP; however, under all settings, our QML estimator tends to have much smaller RMSEs and MAEs than the estimator of BKW. The MAEs and RMSEs of $\hat{k}_{BHS}$ appear to increase with the sample size; thus, the BHS method cannot handle this case.

DGP 1.C In this case, we set $C = [1, 0, 0; 2, 1, 0; 3, 2, m]$ and $m \in \{1, 0, 8, 0.5, 0.1, 0\}$. As $m$ decreases to zero, the matrix $C$ changes from full rank to singular. We still consider serial correlation in factors and serial correlation and cross-sectional dependence in idiosyncratic errors simultaneously with $N = 100, T = 100$. Table 3 shows that the MAEs and RMSEs of $\hat{k}_{QML}$ decrease with $m$, which confirms our findings in Theorems 4 and 5. In addition, the RMSEs and MAEs of $\hat{k}_{BKW}$ and $\hat{k}_{BHS}$ are much larger than those of $\hat{k}_{QML}$, and do not tend toward zero as $m$ decreases. For each value of $m$, the experiment is repeated 10000 times to more accurately estimate and compare the RMSEs (MAEs) of our QML estimator across different values of $m$.

DGP 1.D This DGP considers a Type 1 break. In the first regime, the last elements of $\lambda_{i, 1}$ are zeros for all $i$, and the first two elements of $\lambda_{i, 1}$ are both i.i.d. $N(0, \frac{1}{2} I_{r_0})$. In the second regime, $\lambda_{i, 2}$ is i.i.d. $N(0, \frac{1}{3} I_{r_0})$ across $i$. As $\lambda_{i, 1}$ and $\lambda_{i, 2}$ are independent, the numbers of factors in the two regimes are $r_1 = 2\text{and} r_2 = 3$, respectively, and the number of pseudo-factors is $r = 5$. Because the numbers of pre- or post-break factors are smaller than that of the pseudo-factors, both $\Sigma_1$ and $\Sigma_2$ are singular matrices. Table 4 reports the MAEs and RMSEs of $\hat{k}_{QML}$, $\hat{k}_{BHS}$, and $\hat{k}_{BKW}$ under this DGP. Table 4 shows the suitable performances of both $\hat{k}_{BHS}$ and our $\hat{k}_{QML}$. Their MAEs (RMSEs) are less than 0.05 (0.25) for all combinations of $N$, $T$, $\rho$, $\alpha$, and $\beta$. Although $\hat{k}_{BHS}$ is consistent under this DGP, our QML estimator still has smaller RMSEs than $\hat{k}_{BHS}$ in most cases reported in Table 4. In addition, $\hat{k}_{BKW}$ performs better under this DGP than DGPs 1.A–1.C. However, its estimation error is much larger than that of our QML estimator. This is not surprising because $\hat{k}_{BKW}$ is not consistent. Finally, a larger AR(1) coefficient $\rho$ tends to yield a larger bias for $\hat{k}_{BKW}$, but does not have much effect on the performances of $\hat{k}_{BHS}$ and $\hat{k}_{QML}$.

In summary, Tables 1 and 2 show that the QML estimator performs much better than $\hat{k}_{BHS}$ under Type 2 and 3 breaks, which are ruled out under the assumptions of [Bai et al. (2020)]. Table 4 shows that the QML estimator tends to slightly
outperform $\hat{k}_{BHS}$, even though the latter is known to be consistent under Type 1 breaks. BHS method is super good for Type 1 changes with smaller breaks. QML method will lose its power when breaks are small like in BHS’s settings in their paper, because the dimension of $G$ (determined by IC criterion in Bai and Ng (2002)) will not be augmented when breaks are small, which means the singularity does not show up in the covariance if breaks are small enough.

Table 1: Simulated mean absolute errors (MAEs) and root mean squared errors (RMSEs) of $\hat{k}_{BKW}$, $\hat{k}_{BHS}$, and $\hat{k}_{QML}$ under DGP 1.A.

| $N, T$      | $\hat{k}_{BKW}$ | $\hat{k}_{BHS}$ | $\hat{k}_{QML}$ |
|-------------|-----------------|-----------------|-----------------|
|             | MAE             | RMSE            | MAE             | RMSE            | MAE             | RMSE            |
| $\rho = 0$  | $\alpha = 0$   | $\beta = 0$    | $\rho = 0$      | $\alpha = 0$   | $\beta = 0$    |
| 100,100     | 6.3130          | 8.9546          | 5.4600          | 7.7325          | 1.6070          | 2.9293          |
| 100,200     | 7.0230          | 11.9053         | 7.9580          | 12.4801         | 1.2990          | 2.3206          |
| 200,200     | 5.6730          | 9.9774          | 6.7150          | 10.8610         | 0.7960          | 1.5218          |
| 200,500     | 4.6940          | 8.5732          | 10.0960         | 17.9778         | 0.7340          | 1.3799          |
| 500,500     | 4.4580          | 8.5789          | 8.6770          | 15.6509         | 0.3890          | 0.8597          |
| $\rho = 0$  | $\alpha = 0$   | $\beta = 0$    | $\rho = 0$      | $\alpha = 0$   | $\beta = 0$    |
| 100,100     | 9.7200          | 12.0612         | 4.5670          | 6.9270          | 1.3570          | 2.7592          |
| 100,200     | 14.3410         | 19.5941         | 7.0110          | 11.1559         | 1.0470          | 2.2070          |
| 200,200     | 13.6260         | 19.1151         | 6.7760          | 10.9099         | 0.5840          | 1.2394          |
| 200,500     | 15.4880         | 27.5716         | 10.5450         | 18.7350         | 0.5190          | 1.1406          |
| 500,500     | 16.9890         | 29.5463         | 8.2030          | 15.1581         | 0.3210          | 0.7944          |
| $\rho = 0$  | $\alpha = 0$   | $\beta = 0$    | $\rho = 0$      | $\alpha = 0$   | $\beta = 0$    |
| 100,100     | 6.5060          | 9.1533          | 6.1520          | 8.6248          | 2.3740          | 4.0635          |
| 100,200     | 7.5490          | 12.4416         | 8.7150          | 13.4473         | 1.6920          | 3.1464          |
| 200,200     | 6.2890          | 10.8337         | 8.4910          | 13.2894         | 1.0230          | 1.9409          |
| 200,500     | 5.1220          | 10.1068         | 11.3960         | 19.4945         | 0.8110          | 1.5156          |
| 500,500     | 4.7580          | 9.5055          | 10.3660         | 18.7453         | 0.4570          | 0.9407          |
| $\rho = 0$  | $\alpha = 0$   | $\beta = 0$    | $\rho = 0$      | $\alpha = 0$   | $\beta = 0$    |
| 100,100     | 6.6620          | 9.2573          | 4.7300          | 6.9593          | 1.7580          | 3.1183          |
| 100,200     | 7.8200          | 12.5561         | 6.1740          | 10.2069         | 1.4930          | 2.6943          |
| 200,200     | 6.4500          | 10.9881         | 5.8020          | 9.7340          | 0.7480          | 1.4276          |
| 200,500     | 4.9340          | 10.3110         | 5.9390          | 10.6041         | 0.7020          | 1.3900          |
| 500,500     | 4.0550          | 7.7718          | 5.8820          | 11.0830         | 0.3660          | 0.8567          |
| $\rho = 0$  | $\alpha = 0$   | $\beta = 0$    | $\rho = 0$      | $\alpha = 0$   | $\beta = 0$    |
| 100,100     | 9.9510          | 12.3063         | 5.4430          | 7.6969          | 1.8080          | 3.4531          |
| 100,200     | 14.2890         | 19.5804         | 7.1810          | 11.7141         | 1.3250          | 2.5367          |
| 200,200     | 14.8820         | 20.3572         | 7.3080          | 11.8072         | 0.7450          | 1.6592          |
| 200,500     | 17.0330         | 29.3210         | 9.2010          | 17.0803         | 0.6680          | 1.3461          |
| 500,500     | 14.7130         | 26.3587         | 10.3800         | 19.2727         | 0.3540          | 0.8331          |
Table 2: Simulated mean absolute errors (MAEs) and root mean squared errors (RMSEs) of $\hat{k}_{BK W}$, $\hat{k}_{BHS}$, and $\hat{k}_{QML}$ under DGP 1.B.

\[
\begin{array}{cccccc}
N, T & \hat{k}_{BK W} & \hat{k}_{BHS} & \hat{k}_{QML} \\
& MAE & RMSE & MAE & RMSE & MAE & RMSE \\
\hline
\rho = 0, & \alpha = 0, & \beta = 0 \\
100,100 & 4.1610 & 6.6934 & 8.7430 & 11.0347 & 1.2180 & 2.3259 \\
100,200 & 4.4450 & 8.4477 & 18.5660 & 22.9913 & 0.9960 & 1.8799 \\
200,200 & 4.9160 & 8.9420 & 19.4440 & 23.6923 & 0.9060 & 1.7082 \\
200,500 & 4.4530 & 8.8368 & 49.3330 & 59.4865 & 0.9130 & 1.7085 \\
500,500 & 3.9420 & 7.2061 & 51.9270 & 61.5507 & 0.8370 & 1.5959 \\
\rho = 0 & \alpha = 0 & \beta = 0 \\
100,100 & 6.4570 & 9.4427 & 10.1710 & 12.3371 & 1.9460 & 3.7691 \\
100,200 & 9.1750 & 14.8115 & 21.3380 & 25.1834 & 1.8480 & 3.6362 \\
200,200 & 9.6310 & 15.0080 & 21.5560 & 25.2723 & 1.7850 & 3.5901 \\
200,500 & 11.4150 & 21.3302 & 51.9850 & 61.9028 & 1.6750 & 3.4218 \\
500,500 & 9.5430 & 18.4598 & 53.6060 & 62.7128 & 1.6490 & 3.5501 \\
\rho = 0 & \alpha = 0.3 & \beta = 0 \\
100,100 & 3.9840 & 6.4778 & 7.9990 & 10.5485 & 1.0910 & 2.1824 \\
100,200 & 4.6820 & 8.6151 & 17.6010 & 22.5002 & 1.0360 & 1.9432 \\
200,200 & 4.6350 & 8.4454 & 21.9190 & 26.0996 & 0.8770 & 1.7306 \\
200,500 & 4.2690 & 8.2870 & 50.1790 & 61.5307 & 0.8600 & 1.6474 \\
500,500 & 4.2040 & 8.3094 & 54.8050 & 64.8615 & 0.8040 & 1.5492 \\
\rho = 0 & \alpha = 0 & \beta = 0.3 \\
100,100 & 4.3220 & 6.9244 & 7.7560 & 10.0601 & 1.0510 & 1.9802 \\
100,200 & 4.7150 & 8.6248 & 14.5640 & 19.4154 & 0.9830 & 1.8571 \\
200,200 & 4.5300 & 8.2421 & 18.7950 & 23.1307 & 0.9090 & 1.7587 \\
200,500 & 3.9080 & 7.3553 & 42.6850 & 54.8098 & 0.8900 & 1.6199 \\
500,500 & 4.3570 & 8.5140 & 49.6030 & 59.6292 & 0.8250 & 1.6843 \\
\rho = 0 & \alpha = 0.3 & \beta = 0.3 \\
100,100 & 6.6990 & 9.6327 & 9.1750 & 11.4037 & 2.0750 & 3.9735 \\
100,200 & 9.4990 & 15.0852 & 18.7450 & 23.3085 & 2.0590 & 4.5305 \\
200,200 & 9.2240 & 14.6721 & 20.4670 & 24.5054 & 1.8140 & 3.8021 \\
200,500 & 12.8110 & 23.1517 & 51.0760 & 61.1890 & 1.7200 & 3.5844 \\
500,500 & 10.0590 & 19.2453 & 52.5400 & 62.2628 & 1.7000 & 3.6521 \\
\end{array}
\]
Table 3: Simulated mean absolute errors (MAEs) and root mean squared errors (RMSEs) of \( \hat{k}_{BKW} \), \( \hat{k}_{BHS} \), and \( \hat{k}_{QML} \) under DGP 1.C with \( N = 100, T = 100 \) among 10000 replications.

| \( m \) | \( \hat{k}_{BKW} \) | \( \hat{k}_{BHS} \) | \( \hat{k}_{QML} \) |
|-------|------------------|------------------|------------------|
|       | MAE   | RMSE  | MAE   | RMSE  | MAE   | RMSE  |
| \( \rho = 0 \) | \( \alpha = 0 \) | \( \beta = 0 \) |
| 1     | 3.9228 | 6.4437 | 7.2579 | 9.7141 | 0.6562 | 1.2903 |
| 0.8   | 3.9425 | 6.4624 | 6.6330 | 9.1145 | 0.6348 | 1.2559 |
| 0.5   | 3.7847 | 6.2139 | 5.4950 | 7.9789 | 0.5420 | 1.0814 |
| 0.1   | 3.8469 | 6.2895 | 4.6050 | 6.9212 | 0.5093 | 1.0568 |
| 0     | 3.8310 | 6.2414 | 4.4915 | 6.8352 | 0.4969 | 1.0315 |
| \( \rho = 0 \) | \( \alpha = 0 \) | \( \beta = 0 \) |
| 1     | 6.0404 | 9.0280 | 9.3131 | 11.5733 | 0.9478 | 2.0680 |
| 0.8   | 6.0063 | 9.0017 | 8.5168 | 10.9192 | 0.8547 | 1.8960 |
| 0.5   | 5.9803 | 8.9390 | 6.5127 | 9.0641 | 0.6925 | 1.5752 |
| 0.1   | 5.9300 | 8.8833 | 6.4894 | 7.9812 | 0.5178 | 1.2335 |
| 0     | 6.0197 | 8.9771 | 6.6440 | 6.9049 | 0.5070 | 1.2057 |
| \( \rho = 0 \) | \( \alpha = 0.3 \) | \( \beta = 0 \) |
| 1     | 3.8349 | 6.2423 | 6.2139 | 8.7224 | 0.6727 | 1.3234 |
| 0.8   | 3.8234 | 6.2331 | 6.6511 | 9.2338 | 0.6535 | 1.2963 |
| 0.5   | 3.8345 | 6.3110 | 5.8040 | 8.3371 | 0.6152 | 1.2362 |
| 0.1   | 3.9127 | 6.4083 | 5.0645 | 7.4846 | 0.5895 | 1.1644 |
| 0     | 3.9188 | 6.4124 | 4.9815 | 7.3974 | 0.5813 | 1.1551 |
| \( \rho = 0 \) | \( \alpha = 0 \) | \( \beta = 0.3 \) |
| 1     | 3.8250 | 6.3150 | 6.2535 | 8.7224 | 0.6622 | 1.3039 |
| 0.8   | 3.8135 | 6.2932 | 5.6808 | 8.1438 | 0.6259 | 1.2379 |
| 0.5   | 3.8253 | 6.3061 | 4.6189 | 6.9328 | 0.5619 | 1.1171 |
| 0.1   | 3.9120 | 6.4147 | 3.9299 | 6.0820 | 0.5424 | 1.0949 |
| 0     | 3.8176 | 6.2881 | 3.8963 | 6.0564 | 0.5199 | 1.0497 |
| \( \rho = 0.7 \) | \( \alpha = 0.3 \) | \( \beta = 0.3 \) |
| 1     | 6.0745 | 9.0347 | 8.0648 | 10.5304 | 1.0515 | 2.2669 |
| 0.8   | 6.0041 | 8.9542 | 7.3126 | 9.8433 | 0.9338 | 2.0173 |
| 0.5   | 6.0519 | 9.0124 | 5.8471 | 8.4490 | 0.7798 | 1.7537 |
| 0.1   | 6.0120 | 8.9694 | 4.6376 | 7.1447 | 0.6100 | 1.4401 |
| 0     | 6.0379 | 8.9861 | 4.5336 | 7.0337 | 0.5850 | 1.3509 |
Table 4: Simulated mean absolute errors (MAEs) and root mean squared errors (RMSEs) of $\hat{k}_{BKW}$, $\hat{k}_{BHS}$, and $\hat{k}_{QML}$ under DGP 1.D.

| $N, T$ | $\hat{k}_{BKW}$ | $\hat{k}_{BHS}$ | $\hat{k}_{QML}$ |
|-------|------------------|------------------|------------------|
|       | MAE RMSE         | MAE RMSE         | MAE RMSE |
|       | $\rho = 0$, $\alpha = 0$, $\beta = 0$ | $\rho = 0$, $\alpha = 0$, $\beta = 0$ | $\rho = 0$, $\alpha = 0$, $\beta = 0$ |
| 100,100 | 0.4330 1.3494 | 0.0370 0.1975 | 0.0260 0.1673 |
| 100,200 | 0.3380 1.0900 | 0.0300 0.1732 | 0.0240 0.1549 |
| 200,200 | 0.2780 0.7668 | 0.0180 0.1342 | 0.0130 0.1140 |
| 200,500 | 0.2850 0.8155 | 0.0070 0.0837 | 0.0100 0.1000 |
|       | $\rho = 0.7$, $\alpha = 0$, $\beta = 0$ |       |       |
| 100,100 | 1.8760 4.8750 | 0.0120 0.1095 | 0.0110 0.1049 |
| 100,200 | 1.1140 4.0007 | 0.0150 0.1225 | 0.0110 0.1140 |
| 200,200 | 0.8700 3.5000 | 0.0050 0.0707 | 0.0020 0.0447 |
| 200,500 | 0.4070 1.3435 | 0.0030 0.0548 | 0.0010 0.0316 |
|       | $\rho = 0$, $\alpha = 0.3$, $\beta = 0$ |       |       |
| 100,100 | 0.4400 1.4519 | 0.0450 0.2302 | 0.0410 0.2258 |
| 100,200 | 0.3590 1.3802 | 0.0440 0.2145 | 0.0340 0.1897 |
| 200,200 | 0.3080 0.8438 | 0.0150 0.1225 | 0.0140 0.1265 |
| 200,500 | 0.2150 0.6656 | 0.0160 0.1265 | 0.0120 0.1095 |
|       | $\rho = 0$, $\alpha = 0$, $\beta = 0.3$ |       |       |
| 100,100 | 0.3710 1.0747 | 0.0380 0.1949 | 0.0360 0.1897 |
| 100,200 | 0.2850 0.7918 | 0.0340 0.1897 | 0.0220 0.1483 |
| 200,200 | 0.3150 0.8972 | 0.0100 0.1000 | 0.0110 0.1049 |
| 200,500 | 0.2380 0.6885 | 0.0120 0.1183 | 0.0050 0.0707 |
|       | $\rho = 0.7$, $\alpha = 0.3$, $\beta = 0.3$ |       |       |
| 100,100 | 1.9420 4.8557 | 0.0260 0.1612 | 0.0180 0.1414 |
| 100,200 | 1.0170 3.6438 | 0.0220 0.1549 | 0.0090 0.0949 |
| 200,200 | 0.9750 3.9242 | 0.0060 0.0775 | 0.0080 0.0894 |
| 200,500 | 0.6390 2.5879 | 0.0080 0.0894 | 0.0050 0.0707 |

Tables 5–8 present the probabilities of the correct estimation of the break date. The results are consistent with those displayed in Tables 1–4: the QML estimator $\hat{k}_{QML}$ can detect the true break date with higher probabilities than others regardless of the values of $(\rho, \alpha, \beta)$. The MS method sometimes detects more than one or no break; hence, we only compute its probability of correctly estimating $k_0$ under the condition that it detects a single break. The probabilities of a correct estimation of the QML method increase with the sample sizes $N$ and $T$ in Tables 5, 6, and 8.

Table 7 shows that the probabilities of correct estimation of the QML estimators increase as $m$ decreases. A smaller $m$ means that $C$ is closer to a singular matrix. Table 7 is consistent with Table 3 and confirms Theorems 1 and 3. To explore in more detail the effect of changes in $m$ on the QML estimator, we vary the value of $m$ using finer grids and find a similar pattern to that shown in Table 7. The results are reported in the supplementary appendix.
Figures 1 and 2 show the frequency of the estimated change points under DGP 1.A for $N = 100, T = 100$ and $N = 500, T = 500$ for 1000 replications. According to these figures, the QML estimators exhibit the highest frequency around the true break under different settings. When we increase the $(N, T)$ value from 100 to 500, the frequency at the true break point increases and the simulated distribution becomes tighter. This indicates that the QML estimators are highly likely to identify the true break point. This is consistent with our theory. However, the other three methods are found to have much larger variation and substantially lower probabilities to correctly estimate the break point. Thus, the QML estimators are advantageous in this case. Moreover, the simulation result indicates that for a sample size exceeding $N = 5000, T = 1000$, the probabilities of correctly estimating the QML estimator exceed 90%.

Recall that BKW and QML only have $O_p(1)$ estimation errors under DGP 1.B. However, Table 6 shows that in all cases, the probabilities of correct estimation by the QML estimator are much higher than those of correct estimation by the BKW estimator. Apparently, the BHS and MS methods cannot accurately estimate the true break point in this case. Figures 3 and 4 show the distributions of the estimated change points under (1.B) for $N = 100, T = 100$ and $N = 500, T = 500$, indicating that BHS and MS cannot handle rotational changes. Although the estimation errors of BKW and QML are bounded under all settings, the QML estimators have a much tighter distribution around the true break point.
Table 5: Probability of correct estimation under DGP 1.A.

| $N,T$ | $k_{BKW}$ | $k_{BHS}$ | $k_{MS}$ | $k_{QML}$ |
|-------|-----------|-----------|----------|-----------|
|       | $\rho = 0$ | $\alpha = 0$ | $\beta = 0$ |           |
| 100,100 | 0.1530   | 0.1440   | 0.1626   | 0.4220   |
| 100,200 | 0.1920   | 0.1510   | 0.1863   | 0.4370   |
| 200,200 | 0.2340   | 0.1780   | 0.1307   | 0.5680   |
| 200,500 | 0.2540   | 0.2030   | 0.2020   | 0.5780   |
| 500,500 | 0.2990   | 0.2100   | 0.2123   | 0.7290   |
|       | $\rho = 0.7$ | $\alpha = 0$ | $\beta = 0$ |           |
| 100,100 | 0.1050   | 0.2050   | 0.2329   | 0.5290   |
| 100,200 | 0.1250   | 0.1850   | 0.1779   | 0.5510   |
| 200,200 | 0.1390   | 0.1920   | 0.1898   | 0.6660   |
| 200,500 | 0.1750   | 0.1890   | 0.2031   | 0.6940   |
| 500,500 | 0.2100   | 0.2420   | 0.2306   | 0.7810   |
|       | $\rho = 0$ | $\alpha = 0.3$ | $\beta = 0$ |           |
| 100,100 | 0.1790   | 0.1300   | 0.1072   | 0.3280   |
| 100,200 | 0.1850   | 0.1380   | 0.1897   | 0.4090   |
| 200,200 | 0.2260   | 0.1650   | 0.1931   | 0.5320   |
| 200,500 | 0.2530   | 0.1730   | 0.1845   | 0.5650   |
| 500,500 | 0.2750   | 0.1920   | 0.1964   | 0.6880   |
|       | $\rho = 0$ | $\alpha = 0$ | $\beta = 0.3$ |           |
| 100,100 | 0.1480   | 0.1700   | 0.1956   | 0.3840   |
| 100,200 | 0.1730   | 0.1810   | 0.1847   | 0.4210   |
| 200,200 | 0.2240   | 0.2110   | 0.2069   | 0.5700   |
| 200,500 | 0.2770   | 0.2250   | 0.2370   | 0.5930   |
| 500,500 | 0.3220   | 0.2790   | 0.2790   | 0.7500   |
|       | $\rho = 0.7$ | $\alpha = 0.3$ | $\beta = 0.3$ |           |
| 100,100 | 0.1070   | 0.1510   | 0.1739   | 0.4670   |
| 100,200 | 0.1210   | 0.1860   | 0.2157   | 0.5030   |
| 200,200 | 0.1370   | 0.1820   | 0.2072   | 0.6360   |
| 200,500 | 0.1670   | 0.2180   | 0.2149   | 0.6520   |
| 500,500 | 0.1900   | 0.2510   | 0.2427   | 0.7640   |
Table 6: Probability of correct estimation under DGP 1.B.

| $N, T$ | $k_{BW}$ | $k_{BHS}$ | $k_{MS}$ | $k_{QML}$ |
|--------|----------|-----------|----------|-----------|
| $\rho = 0 \quad \alpha = 0 \quad \beta = 0$ |          |           |          |           |
| 100,100 | 0.2760   | 0.0690    | 0.0769   | 0.4790    |
| 100,200 | 0.2920   | 0.0540    | 0.0362   | 0.5180    |
| 200,200 | 0.2720   | 0.0320    | 0.0655   | 0.5270    |
| 200,500 | 0.3110   | 0.0140    | 0.0091   | 0.5340    |
| 500,500 | 0.2960   | 0.0100    | 0.0123   | 0.5580    |
| $\rho = 0 \quad \alpha = 0 \quad \beta = 0$ |          |           |          |           |
| 100,100 | 0.2710   | 0.0640    | 0.0909   | 0.4540    |
| 100,200 | 0.2500   | 0.0270    | 0.0398   | 0.4530    |
| 200,200 | 0.2180   | 0.0160    | 0.0200   | 0.4790    |
| 200,500 | 0.2370   | 0.0120    | 0.0144   | 0.4970    |
| 500,500 | 0.2450   | 0.0080    | 0.0080   | 0.5090    |
| $\rho = 0 \quad \alpha = 0.3 \quad \beta = 0$ |          |           |          |           |
| 100,100 | 0.3050   | 0.1060    | 0.1163   | 0.5180    |
| 100,200 | 0.2930   | 0.0740    | 0.0989   | 0.5020    |
| 200,200 | 0.2890   | 0.0390    | 0.0496   | 0.5540    |
| 200,500 | 0.3000   | 0.0230    | 0.0328   | 0.5630    |
| 500,500 | 0.3090   | 0.0090    | 0.0125   | 0.5780    |
| $\rho = 0 \quad \alpha = 0 \quad \beta = 0.3$ |          |           |          |           |
| 100,100 | 0.2740   | 0.0880    | 0.1458   | 0.5000    |
| 100,200 | 0.2970   | 0.0650    | 0.0692   | 0.5220    |
| 200,200 | 0.2870   | 0.0390    | 0.0338   | 0.5390    |
| 200,500 | 0.3100   | 0.0300    | 0.0320   | 0.5290    |
| 500,500 | 0.2940   | 0.0120    | 0.0123   | 0.5810    |
| $\rho = 0.7 \quad \alpha = 0 \quad \beta = 0.3$ |          |           |          |           |
| 100,100 | 0.2210   | 0.1000    | 0.1524   | 0.4330    |
| 100,200 | 0.2400   | 0.0610    | 0.0763   | 0.4640    |
| 200,200 | 0.2370   | 0.0490    | 0.0538   | 0.4810    |
| 200,500 | 0.2230   | 0.0230    | 0.0218   | 0.4770    |
| 500,500 | 0.2420   | 0.0160    | 0.0207   | 0.5100    |
Table 7: Probability of correct estimation under DGP 1.C with \( N = 100, T = 100 \).

| \( m \) | \( k_{BK W} \) | \( k_{BHS} \) | \( k_{MS} \) | \( k_{QML} \) |
|------|-------|-------|-------|-------|
| \( \rho = 0 \) | \( \alpha = 0 \) | \( \beta = 0 \) |       |       |
| 1    | 0.3044 | 0.1075 | 0.1188 | 0.6079 |
| 0.8  | 0.3033 | 0.1252 | 0.1389 | 0.6153 |
| 0.5  | 0.3009 | 0.1736 | 0.1904 | 0.6467 |
| 0.1  | 0.2976 | 0.1998 | 0.2014 | 0.6680 |
| 0    | 0.2977 | 0.2031 | 0.2192 | 0.6705 |
| \( \rho = 0.7 \) | \( \alpha = 0 \) | \( \beta = 0 \) |       |       |
| 1    | 0.2841 | 0.0781 | 0.1040 | 0.6051 |
| 0.8  | 0.2871 | 0.0975 | 0.1135 | 0.6254 |
| 0.5  | 0.2896 | 0.1532 | 0.1620 | 0.6641 |
| 0.1  | 0.2876 | 0.2073 | 0.2369 | 0.7131 |
| 0    | 0.2848 | 0.2194 | 0.2297 | 0.7154 |
| \( \rho = 0 \) | \( \alpha = 0.3 \) | \( \beta = 0 \) |       |       |
| 1    | 0.2973 | 0.1226 | 0.1442 | 0.6063 |
| 0.8  | 0.2981 | 0.1416 | 0.1648 | 0.6134 |
| 0.5  | 0.2988 | 0.1641 | 0.1730 | 0.6219 |
| 0.1  | 0.2993 | 0.1828 | 0.1954 | 0.6316 |
| 0    | 0.2988 | 0.1860 | 0.1995 | 0.6342 |
| \( \rho = 0 \) | \( \alpha = 0 \) | \( \beta = 0.3 \) |       |       |
| 1    | 0.3018 | 0.1344 | 0.1399 | 0.6075 |
| 0.8  | 0.3016 | 0.1549 | 0.1679 | 0.6164 |
| 0.5  | 0.3044 | 0.1927 | 0.2078 | 0.6383 |
| 0.1  | 0.3009 | 0.2141 | 0.2093 | 0.6461 |
| 0    | 0.3036 | 0.2211 | 0.2314 | 0.6566 |
| \( \rho = 0.7 \) | \( \alpha = 0.3 \) | \( \beta = 0.3 \) |       |       |
| 1    | 0.2821 | 0.1519 | 0.1739 | 0.5921 |
| 0.8  | 0.2844 | 0.1710 | 0.1964 | 0.6082 |
| 0.5  | 0.2843 | 0.2327 | 0.2402 | 0.6496 |
| 0.1  | 0.2850 | 0.2889 | 0.2911 | 0.6898 |
| 0    | 0.2868 | 0.2951 | 0.2966 | 0.6951 |
Table 8: Probability of correct estimation under DGP 1.D.

| N,T     | $k_{BKW}$ | $k_{BKS}$ | $k_{MS}$ | $k_{QML}$ |
|---------|-----------|-----------|---------|---------|
| 100,100 | 0.7960    | 0.9640    | 0.9553  | 0.9750  |
| 100,200 | 0.8200    | 0.9700    | 0.9700  | 0.9760  |
| 200,200 | 0.8160    | 0.9820    | 0.9841  | 0.9870  |
| 200,500 | 0.8260    | 0.9930    | 0.9930  | 0.9990  |
| 100,100 | 0.7000    | 0.9880    | 0.9864  | 0.9890  |
| 100,200 | 0.7540    | 0.9850    | 0.9859  | 0.9900  |
| 200,200 | 0.7950    | 0.9950    | 0.9949  | 0.9980  |
| 200,500 | 0.8220    | 0.9970    | 0.9970  | 0.9990  |
| 100,100 | 0.8020    | 0.9580    | 0.9563  | 0.9630  |
| 100,200 | 0.8170    | 0.9570    | 0.9589  | 0.9670  |
| 200,200 | 0.8140    | 0.9850    | 0.9842  | 0.9870  |
| 200,500 | 0.8510    | 0.9840    | 0.9840  | 0.9880  |
| 100,100 | 0.7910    | 0.9620    | 0.9671  | 0.9640  |
| 100,200 | 0.8090    | 0.9670    | 0.9674  | 0.9780  |
| 200,200 | 0.8150    | 0.9900    | 0.9904  | 0.9890  |
| 200,500 | 0.8330    | 0.9890    | 0.9890  | 0.9950  |
| 100,100 | 0.6670    | 0.9740    | 0.9766  | 0.9830  |
| 100,200 | 0.7670    | 0.9790    | 0.9801  | 0.9910  |
| 200,200 | 0.7910    | 0.9940    | 0.9940  | 0.9920  |
| 200,500 | 0.7900    | 0.9920    | 0.9920  | 0.9950  |

6. Empirical Application

6.1 Macroeconomic data

In the first empirical application, we apply our proposed method to a U.S. macroeconomic dataset (Stock and Watson (2012)) to detect the possible structural breaks in the underlying factor model. We use the dataset adopted by Cheng et al. (2016), which comprises monthly observations of 102 U.S. macroeconomic variables. The sample begins after the Great Moderation and ranges from 1985:01 to 2013:01 ($T = 337$). Following Bai et al. (2020), we focus on the subsample period between 2001:12 and 2013:01 ($T = 134$, $N = 102$) because the complete data may have multiple breaks.

Cheng et al. (2016) find that 2007:12 is a single-break date, and that the pre-break and post-break subsamples have one factor and two or three factors, respectively. Following Cheng et al. (2016), Bai et al. (2020) also set the number of factors
Figure 1: Plots of the frequency of the estimated break points among 1000 replications for DGP 1.A and $N = 100, T = 100$. 

(a) $(\rho, \alpha, \beta) = (0.7, 0, 0)$

(b) $(\rho, \alpha, \beta) = (0, 0.3, 0)$

(c) $(\rho, \alpha, \beta) = (0, 0, 0.3)$

(d) $(\rho, \alpha, \beta) = (0.7, 0.3, 0.3)$
Figure 2: Plots of the frequency of the estimated break points among 1000 replications for DGP 1.A and $N = 500, T = 500$. 

- (a) $(\rho, \alpha, \beta) = (0.7, 0, 0)$
- (b) $(\rho, \alpha, \beta) = (0, 0.3, 0)$
- (c) $(\rho, \alpha, \beta) = (0, 0, 0.3)$
- (d) $(\rho, \alpha, \beta) = (0.7, 0.3, 0.3)$
Figure 3: Plots of the frequency of the estimated break points among 1000 replications for DGP 1.B and $N = 100, T = 100$. 

(a) $(\rho, \alpha, \beta) = (0.7, 0, 0)$

(b) $(\rho, \alpha, \beta) = (0, 0.3, 0)$

(c) $(\rho, \alpha, \beta) = (0, 0, 0.3)$

(d) $(\rho, \alpha, \beta) = (0.7, 0.3, 0.3)$
Figure 4: Plots of the frequency of the estimated break points among 1000 replications for DGP 1.B and $N = 500, T = 500$.

(a) $(\rho, \alpha, \beta) = (0.7, 0, 0)$

(b) $(\rho, \alpha, \beta) = (0, 0.3, 0)$

(c) $(\rho, \alpha, \beta) = (0, 0, 0.3)$

(d) $(\rho, \alpha, \beta) = (0.7, 0.3, 0.3)$
equal to one and two for the pre- and post-break subsamples, respectively. Then, they implement the LS estimation and obtain the estimated break point \( \hat{k} = 2008:12 \). To implement our QML method, we first use Bai and Ng’s information criterion IC1 and determine three pseudo-factors in the complete sample. Based on this result, we compute our QML estimator and obtain 2007:07 as the estimated break point, using which we split the sample into pre- and post-break subsamples. IC1 of Bai and Ng (2002) detects two pre-break and three post-break factors. From the numbers of pre- and post-break factors and that of pseudo-factors, one factor appears to emerge over time and the QML estimator is consistent based on Theorem 4.

6.2 Stock data

The second empirical application uses the weekly rate of return for Nasdaq 100 Index from April 18, 2019, to October 1, 2020. As all companies have data starting from April 18, 2019, we choose that as the start date. Traditionally, the index is limited to 100 common-stock issues, with only one issue allowed per user. Now, the index is limited to 100 issuers, some of which may have multiple issues as index components. The current index has 103 components, representing 100 issuers, four of which are from China: Baidu, JD.com, Ctrip, and NetEase. Thus, the sample size is \( T = 76 \) and \( N = 103 \). As IC1 and IC2 of Bai and Ng (2002), the methods proposed by Onatski (2010), Ahn and Horenstein (2013), and Fan et al. (2019) yield different numbers of pseudo-factors for all samples, we use different number of factors \( r = 2, 3, 4, 5, 6, 7 \) to estimate the break date by using the QML method, and find that the estimated break date always falls in the week of February 20, 2020. This result agrees with that obtained using the method developed by Baltagi et al. (2017).

In fact, after receiving a briefing that the COVID-19 epidemic was about to spread, the Senate Intelligence Committee Chairman Richard Burr sold 628,000 to 1.72 million stocks in a one-day transaction on February 13, 2020. A week after this, that is, the week of February 20, 2020, the stock market began to fall sharply, and two weeks later, U.S. stocks halted for the first time. Thus, the factor loading matrix appears to have changed in the early days of the epidemic.

7. Conclusions

We study the QML method for estimating the break point in high-dimensional factor models with a single large structural change. We consider three types of change and develop an asymptotic theory for the QML estimator. We show that the QML estimator is consistent when the covariance matrices of the pre- or post-break factor loadings, or both, are singular. In addition, the estimation error of the QML estimator is \( O_p(1) \) when there is a rotation type of change in the factor loading matrix, and thus, the covariance matrices of the pre- and post-break loadings are both nonsingular. The limiting distribution of the estimated break point can also be derived in this case. The simulation results validate the suitable performance of the QML estimator. We use the proposed method to estimate the break point for U.S. macroeconomic data and stocks data. The estimated break date is July 2007 for the macroeconomic data and February 20, 2020, for the stocks data.

Appendix

In model 4.

\[
X = GA' + \epsilon, \quad (1)
\]

27
\[ G = (g_1, \ldots, g_T)', g_t = Bf_t \text{ for } t \leq k_0, \text{ and } g_t = Cf_t \text{ for } t > k_0. \text{ } \lambda_i \text{ and } f_t \text{ are always } r\text{-dimensional vectors and both } \Lambda_1 \text{ and } \Lambda_2 \text{ have dimension } N \times r. \text{ Let } \hat{G} = (\hat{g}_1, \ldots, \hat{g}_T)' \text{ denote the full-sample PCA estimator for } G:\]

\[
\hat{\Sigma}_1(k) \equiv k^{-1} \sum_{t=1}^{k} \hat{g}_t \hat{g}_t', \\
\hat{\Sigma}_2(k) \equiv (T - k)^{-1} \sum_{t=k+1}^{T} \hat{g}_t \hat{g}_t'.
\]

For notational simplicity, let \( \hat{\Sigma}_1^0 \equiv \hat{\Sigma}_1(k_0) \) and \( \hat{\Sigma}_2^0 \equiv \hat{\Sigma}_2(k_0). \)

The QML objective function can be expressed as

\[
U_{NT}(k) = k \log |\hat{\Sigma}_1| + (T - k) \log |\hat{\Sigma}_2|.
\]

If \( k = k_0 \), the objective function is

\[
U_{NT}(k_0) = k_0 \log |\hat{\Sigma}_1^0| + (T - k_0) \log |\hat{\Sigma}_2^0|,
\]

where \( \hat{\Sigma}_1^0 = k_0^{-1} \sum_{t=1}^{k_0} \hat{g}_t \hat{g}_t' \), \( \hat{\Sigma}_2^0 = (T - k_0)^{-1} \sum_{t=k_0+1}^{T} \hat{g}_t \hat{g}_t' \).

**Representations of \( \hat{g}_t. \)**

The full-sample PCA estimator \( \hat{G} \) satisfies the following identity:

\[
\hat{G} = \frac{1}{NT} XX' \hat{G} V_{NT}^{-1} = GH + \frac{1}{NT\hat{G}'' \hat{G}''^{-1}} + \frac{1}{NT} G' \hat{e} \hat{e}' \hat{G} V_{NT}^{-1} + \frac{1}{NT} \hat{e} \hat{e}' \hat{G} V_{NT}^{-1},
\]

where \( H = \Lambda' \Lambda \hat{G} V_{NT}^{-1} / NT \) and \( V_{NT} \) is a diagonal matrix comprising the eigenvalues of \( XX' / NT \).

Hence, for each period \( t \), we have

\[
\hat{g}_t - H' g_t = V_{NT}^{-1} \left( \frac{\hat{G} G N \hat{e} \hat{e}}{T} + \frac{\hat{G} \hat{e} \Lambda}{NT} g_t + \frac{\hat{G} \hat{e} \hat{e}_2}{NT} \right)
\]

Bai (2003) shows that

\[
\hat{g}_{k+1} - H' g_{k+1} = O_p(\delta^{-1} N T^{-1})
\]

\[
T^{-1} \sum_{t=1}^{T} \|\hat{g}_t - H' g_t\|^2 = O_p(\delta^{-2} N T), \text{ and } T^{-1} (\hat{G}' \hat{G} - H' G'H) = O_p(\delta^{-2} N T)
\]

From (A.1) and Lemma A.2 in Bai (2003), we have the following lemma:

**Lemma 1.** (i) Under Assumptions 7-8

\[
\max_m m^{-1} \sum_{t=k_0 - m}^{k_0} \|\hat{g}_t - H' g_t\|^2 = O_p \left( \frac{1}{N} \right),
\]

\[
\max_m m^{-1} \sum_{t=k_0 + m}^{k_0 + m} \|\hat{g}_t - H' g_t\|^2 = O_p \left( \frac{1}{N} \right).
\]
(ii). Under Assumptions 7–14,

\[
\max_m \frac{m^{-1}}{m} \sum_{t=k_0-m}^{k_0} \|\hat{g}_t - H'g_t\|^2 \leq \frac{c}{N},
\]

where \(c > 0\) is a constant.

**Proof.** See the supplementary appendix. \(\square\)

**Both \(\Sigma_1\) and \(\Sigma_2\) are positive definite matrices.**

We first consider the case in which both \(\Sigma_1\) and \(\Sigma_2\) are positive definite matrices. Following Baltagi et al. (2017), we define

\[
\zeta_t = \hat{g}_t \hat{g}_t' - H_0'g_t g_t'H_0, \text{ for } t = 1, \cdots, T
\]

and

\[
\xi_t = H_0'g_t g_t'H_0 - \Sigma_1 \text{ for } t \leq k_0,
\]

\[
\xi_t = H_0'g_t g_t'H_0 - \Sigma_2 \text{ for } t > k_0,
\]

where \(\Sigma_1 = H_0'\Sigma_{G,1}H_0\) and \(\Sigma_2 = H_0'\Sigma_{G,2}H_0\) are the pre- and post-breaks of \(H_0'E(\hat{g}_t \hat{g}_t')H_0\) and \(H_0\) is the probability limit of \(H\). Thus, we have

\[
\hat{g}_t \hat{g}_t' = \Sigma_1 + \xi_t + \zeta_t \text{ for } t \leq k_0,
\]

\[
\hat{g}_t \hat{g}_t' = \Sigma_2 + \xi_t + \zeta_t \text{ for } t > k_0.
\]

\(H_0\) is nonsingular by Proposition 1 of Bai (2003).

For \(k \leq k_0\),

\[
\hat{\Sigma}_1 = \Sigma_1 + \frac{1}{k} \sum_{t=1}^{\min(k, T)} \xi_t + \frac{1}{k} \sum_{t=1}^{\min(k, T)} \zeta_t,
\]

\[
\hat{\Sigma}_2 = \frac{k_0 - k}{T-k} (\Sigma_1 - \Sigma_2) + \frac{1}{T-k} \sum_{t=k+1}^{T} \xi_t + \frac{1}{T-k} \sum_{t=k+1}^{T} \zeta_t,
\]

Thus,

\[
\hat{\Sigma}_1 - \Sigma_1^0 = \frac{k_0 - k}{k_0} \sum_{t=1}^{k} (\xi_t + \zeta_t) - \frac{1}{k_0} \sum_{t=k+1}^{k_0} (\xi_t + \zeta_t),
\]

\[
\hat{\Sigma}_2 - \Sigma_2^0 = \frac{k - k_0}{T-k} (\Sigma_2 - \Sigma_1) + \frac{1}{T-k} \sum_{t=k+1}^{T} (\xi_t + \zeta_t) + \frac{k - k_0}{(T-k)(T-k_0)} \sum_{t=k+1}^{T} (\xi_t + \zeta_t).
\]

Before analyzing the consistency of the estimated fraction and the boundedness of the estimation error, we need to prove the following lemmas. For any given \(0 < \eta \leq \min(\tau_0, 1 - \tau_0)\) and \(M > 0\), define \(D_\eta = \{k : (\tau_0 - \eta)T \leq k \leq (\tau_0 + \eta)T\}\), \(D_\eta^c\) as the complement of \(D_\eta\), \(\tau_0 = \frac{k_0}{m}\), and \(D_{\eta, M} = \{k : (\tau_0 - \eta)T \leq k \leq (\tau_0 + \eta)T, \ |k_0 - k| > M\}\.
Lemma 2. Under Assumptions\textsuperscript{[4,8]}

\begin{align*}
\text{(i)} & \max_{\left\{ r \right\} \leq k \leq k_0} \left\| \frac{1}{k_0 - k} \sum_{t=1}^{k_0} \xi_t \right\| = O_p\left( \frac{1}{\sqrt{T}} \right), \\
\text{(ii)} & \max_{k \in D, k < k_0} \left\| \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} \xi_t \right\| = O_p(1), \\
\text{(iii)} & \max_{k \in D, k < k_0} \left\| \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} \xi_t \right\| = O_p\left( \frac{1}{\sqrt{T}} \right), \\
\text{(iv)} & \max_{\left\{ r \right\} \leq k \leq k_0} \left\| \frac{1}{T - k} \sum_{t=k+1}^{T} \xi_t \right\| = O_p\left( \frac{1}{\sqrt{T}} \right), \\
\text{(v)} & \max_{\left\{ r \right\} \leq k \leq k_0} \left\| \frac{1}{k_0} \sum_{t=1}^{k_0} \xi_t \right\| = o_p(1), \\
\text{(vi)} & \max_{\left\{ r \right\} \leq k \leq k_0} \left\| \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} \xi_t \right\| = o_p(1), \\
\text{(vii)} & \max_{\left\{ r \right\} \leq k \leq k_0} \left( \frac{1}{T - k} \sum_{t=k+1}^{T} \xi_t \right) = o_p(1),
\end{align*}

where \( \tau \in (0, 1) \) is the prior lower bound for \( \tau_0, \) \( \left\lfloor \tau T \right\rfloor \) denotes the prior lower bound for the real break point \( \left\lfloor \tau_0 T \right\rfloor = k_0, \) and \( \left\lfloor \cdot \right\rfloor \) denotes the integer part of a real number.

\textbf{Proof.} See the supplementary appendix. \( \square \)

Lemma 3. Under Assumptions\textsuperscript{[4,8]}

\[ \max_{k \in D, k < k_0} \left\| \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} \xi_t \right\| = o_p(1) \]

\textbf{Proof.} See the supplementary appendix. \( \square \)

Lemma 4. Under Assumptions\textsuperscript{[4,8]} for \( k \in D, k < k_0, \) if both \( \Sigma_1 \) and \( \Sigma_2 \) are positive definite matrices, then

\[ \frac{k}{k_0 - k} \log |\Sigma_1^{0}\Sigma_1^{-1}| = -\frac{k}{k_0 - k} \sum_{t=k+1}^{k_0} \text{tr}(\xi_t \Sigma_1^{0-1}) + o_p(1), \]

where the \( o_p(1) \) term is uniform in \( k \in D, k < k_0. \)

\textbf{Proof.} See the supplementary appendix. \( \square \)

Lemma 5. Under Assumptions\textsuperscript{[4,8]} for \( |k - k_0| \leq M \) and \( k < k_0, \) if both \( \Sigma_1 \) and \( \Sigma_2 \) are positive definite matrices, then

\[ (T - k) \log |\Sigma_2^{0} \Sigma_2^{-1}| = (k - k_0) \text{tr}(\Sigma_2 - \Sigma_1) \Sigma_2^{-1} + \sum_{t=k+1}^{k_0} \text{tr}(\xi_t \Sigma_2^{-1}) + o_p(1). \]

\textbf{Proof.} See the supplementary appendix. \( \square \)
Proof of $\hat{\tau} - \tau_0 = o_p(1)$

By symmetry, it suffices to study the case of $k < k_0$. Expanding $U_{NT}(k) - U_{NT}(k_0)$ gives

$$U_{NT}(k) - U_{NT}(k_0) = k \log |\hat{\Sigma}_1 \hat{\Sigma}_1^{-1} - (T - k) \log |\hat{\Sigma}_2 \hat{\Sigma}_2^{-1} -(k_0 - k) \log |\hat{\Sigma}_1 \hat{\Sigma}_2^{-1}|.$$  

(12)

To prove $\hat{\tau} - \tau_0 = o_p(1)$, we need to show that for any $\varepsilon > 0$ and $\eta > 0$, $P(|\hat{\tau} - \tau_0| > \eta) < \varepsilon$ as $(N, T) \to \infty$, and that $P(\hat{k} \in D_{\eta}^c) < \varepsilon$. For notational simplicity, we write $U_{NT}(k)$ as $U(k)$.

As $\hat{k} = \arg \min_{k \in D_{\eta}^c} U(k)$, we have $U(\hat{k}) - U(k_0) \leq 0$. If $k \in D_{\eta}^c$, then $\min_{k \in D_{\eta}^c} U(k) - U(k_0) \leq 0$. This implies $P(\hat{k} \in D_{\eta}^c) \leq P(\min_{k \in D_{\eta}^c} U(k) - U(k_0) \leq 0)$; thus, it suffices to show that for any given $\varepsilon > 0$ and $\eta > 0$, $P(\min_{k \in D_{\eta}^c} U(k) - U(k_0) \leq 0) < \varepsilon$ as $(N, T) \to \infty$.

Suppose that $\min_{k \in D_{\eta}^c} U(k) - U(k_0) \leq 0$ and $k^* = \arg \min_{k \in D_{\eta}^c} U(k) - U(k_0)$; then, $U(k^*) - U(k_0) \leq 0$ and $\frac{U(k^*) - U(k_0)}{k^* - k_0} \leq 0$. As $k^* \in D_{\eta}^c$, we have $\min_{k \in D_{\eta}^c} \frac{U(k) - U(k_0)}{k - k_0} \leq \frac{U(k^*) - U(k_0)}{k^* - k_0} \leq 0$. Thus, $\min_{k \in D_{\eta}^c} U(k) - U(k_0) \leq 0$ implies $\min_{k \in D_{\eta}^c} \frac{U(k) - U(k_0)}{k - k_0} \leq 0$. Similary, $\min_{k \in D_{\eta}^c} \frac{U(k) - U(k_0)}{k - k_0} \leq 0$ implies $\min_{k \in D_{\eta}^c} U(k) - U(k_0) \leq 0$. Therefore, the following two events are equivalent:

$$\{w : \min_{k \in D_{\eta}^c} U(k) - U(k_0) \leq 0\} = \{w : \min_{k \in D_{\eta}^c} \frac{U(k) - U(k_0)}{k - k_0} \leq 0\}. \tag{13}$$

Note that

$$P(\min_{x \in X} a(x) + b(x) \leq 0) \leq P(\min_{x \in X} a(x) + b(x) \leq 0) = P(\min_{x \in X} a(x) + o_p(1) \leq 0) \tag{14}$$

if $b(x) = o_p(1)$ uniformly for $x \in X$.

Now, using (12) and (13), we have

$$P(\min_{k \in D_{\eta}^c} U(k) - U(k_0) \leq 0) = P(\min_{k \in D_{\eta}^c} \frac{U(k) - U(k_0)}{k - k_0} \leq 0) = P(\min_{k \in D_{\eta}^c} \frac{U(k) - U(k_0)}{k - k_0} \leq 0) \tag{15}$$

where $\min_{k \in D_{\eta}^c} \frac{1}{k - k_0} \log |\hat{\Sigma}_1 \hat{\Sigma}_1^{-1}| = o_p(1)$ because $||\hat{\Sigma}_1 - \hat{\Sigma}^0||$ is uniformly $o_p(1)$ for $|\tau_1T| \leq k \leq k_0$ by (11) and Lemmas 2 (i), (iii), (v) and (vi). Note that

$$\hat{\Sigma}_2 = \frac{T-k}{T-k} \sum_{T-k+1}^{T} \hat{g}_t \hat{g}_t' = \frac{T-k}{T-k} \sum_{T-k+1}^{T} \hat{g}_t \hat{g}_t' + \frac{T-k}{T-k} \sum_{T-k+1}^{T} \hat{g}_t \hat{g}_t' = \frac{T-k}{T-k} (\hat{\Sigma}_1 + o_p(1)) + \frac{T-k}{T-k}$$

(16)

because $\max_{k \leq k_0 - \eta T} ||\hat{\Sigma}_1 - \hat{\Sigma}^0|| = o_p(1)$ by Lemma 3. Thus, by (14) and (16), we can bound (15) by

$$P(\min_{k \in D_{\eta}^c} \frac{U(k) - U(k_0)}{k - k_0} \leq 0) = P(\min_{k \in D_{\eta}^c} \frac{U(k) - U(k_0)}{k - k_0} \leq 0) \tag{17}$$

Let $g(X) = \frac{T-k}{k_0-k} \sum_{t=k+1}^{T} \log (\frac{T-k}{T-k} I + \frac{k_0-k}{k_0-k} X) - \log |X|$ and $k \in D_{\eta}^c$, $k < k_0$, where $X = \hat{\Sigma}_1 \hat{\Sigma}_2^{-1}$. By the property of a characteristic polynomial, we have

$$g(X) = \frac{T-k}{k_0-k} \sum_{i=1}^{r} \log \left(\frac{T-k}{T-k} I + \frac{k_0-k}{T-k} \rho_i(X) \right) - \sum_{i=1}^{r} \log \rho_i(X). \tag{17}$$
where \( \rho_i(X) \) is the \( i \)-th eigenvalue of \( X \) for \( i = 1, \cdots, r \). On the partial derivative with respect to \( \rho_i(X) \), we have

\[
\frac{\partial g(X)}{\partial \rho_i(X)} = \frac{(T - k_0)(\rho_i(X) - 1)}{[(T - k_0) + (k_0 - k)\rho_i(X)]\rho_i(X)}.
\]

From the derivative with respect to \( \rho_i(X) \), \( \frac{\partial g(X)}{\partial \rho_i(X)} < 0 \) for \( 0 < \rho_i(X) < 1 \) and \( \frac{\partial g(X)}{\partial \rho_i(X)} > 0 \) for \( \rho_i(X) > 1 \). Thus, for \( g(X) \) to achieve its minimum value, all eigenvalues of \( X \) must be one (i.e., all eigenvalues of the symmetric matrix \( \tilde{\Sigma}_{20}^{-1/2}\tilde{\xi}_{20}^{-1/2} \) should be equal to one); thus, \( \tilde{\Sigma}_{20}^{-1/2}\tilde{\xi}_{20}^{-1/2} = I \) and \( \tilde{\xi}_{20} = \tilde{\xi}_2 \). This implies that \( g(I) = 0 \) is a unique minimum of \( g(X) \).

Note that \( \hat{\Sigma} = \Sigma \), \( \Sigma > 0 \) and \( \hat{\Sigma} = \Sigma \). To prove \( \hat{\Sigma} \), \( \Sigma > 0 \) and \( \hat{\Sigma} = \Sigma \). Thus, we have \( \hat{\Sigma} \to \Sigma \). By the consistency of \( \hat{\Sigma} \), \( \Sigma > 0 \) and \( \hat{\Sigma} = \Sigma \). Hence, it suffices to show that for any \( \varepsilon > 0 \) and \( \eta > 0 \), there exists an \( M > 0 \) such that \( P(\hat{\rho} - k_0 > M) < \varepsilon \) as \( (N, T) \to \infty \). For the given \( \eta \) and \( M \), we have \( D_{\eta, M} = \{ k : (\eta - \eta)T \leq k \leq (\eta + \eta)T, |k_0 - k| > M \} \); thus, \( P(\hat{\rho} - k_0 > M) = P(\hat{k} \in D_{\eta, M}) + P(\hat{k} \in D_{\eta, M}) \). Hence, it suffices to show that \( P(\hat{\rho} - k_0 > M) < \varepsilon \) as \( (N, T) \to \infty \).

Again, by symmetry, it suffices to study the case of \( k < k_0 \). Similar to the proof of consistency of \( \hat{\tau} \), we have

\[
P(\min_{k \in D_{\eta, M}, k < k_0} U(k) - U(k_0) \leq 0),
= P(\min_{k \in D_{\eta, M}, k < k_0} \frac{U(k) - U(k_0)}{k_0 - k} \leq 0),
= P(\min_{k \in D_{\eta, M}, k < k_0} \frac{k}{k_0 - k} \log |\tilde{\Sigma}_1\tilde{\xi}_{10}^{-1}| + \frac{T - k_0}{k_0 - k} \log |\tilde{\Sigma}_2\tilde{\xi}_{20}^{-1}| - \log |\tilde{\Sigma}_1\tilde{\xi}_{20}^{-1}| \leq 0)
= P(\min_{k \in D_{\eta, M}, k < k_0} \frac{k}{k_0 - k} \log |\tilde{\Sigma}_1\tilde{\xi}_{10}^{-1}| + \frac{T - k_0}{k_0 - k} \log |\tilde{\Sigma}_2\tilde{\xi}_{20}^{-1}| - \log |\tilde{\Sigma}_1\tilde{\xi}_{20}^{-1}| \leq 0)
= P(\min_{k \in D_{\eta, M}, k < k_0} \frac{k}{k_0 - k} \log |\tilde{\Sigma}_1\tilde{\xi}_{10}^{-1}| + \frac{T - k_0}{k_0 - k} \log |\tilde{\Sigma}_2\tilde{\xi}_{20}^{-1}| - \log |\tilde{\Sigma}_1\tilde{\xi}_{20}^{-1}| \leq 0).
\]

Note that

\[
\frac{k}{k_0 - k} \log |\tilde{\Sigma}_1\tilde{\xi}_{10}^{-1}| = -\frac{k}{k_0 - k} \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} \text{tr}(\tilde{\xi}_t\tilde{\xi}_{10}^{-1}) + o_p(1),
\]

where the \( o_p(1) \) term is uniform in \( k \in D_{\eta, M} \) and \( k < k_0 \) by Lemma 3.
In addition,
\[
\frac{1}{k_0 - k} \sum_{i=1}^{k_0} \hat{g}_i \hat{g}_i' - \Sigma_0 = \frac{1}{k_0 - k} \left( \sum_{i=1}^{k_0} \hat{g}_i \hat{g}_i' - \sum_{i=1}^{k} \hat{g}_i \hat{g}_i' \right) - \Sigma_1
\]
\[
= \frac{k_0}{k_0 - k} \Sigma_0 - \frac{k}{k_0 - k} \hat{\Sigma}_1 - \hat{\Sigma}_1 - \Sigma_1
\]
\[
= \frac{k}{k_0 - k} \left( \Sigma_0 - \hat{\Sigma}_1 - \Sigma_1 \right) \sum_{i=1}^{k_0} \left( \xi_t + \zeta_t \right) - \frac{1}{k_0} \sum_{i=1}^{k_0} \left( \xi_t + \zeta_t \right)
\]
\[
= \frac{k}{k_0 - k} \sum_{i=1}^{k_0} \left( \xi_t + \zeta_t \right) - \frac{1}{k_0} \sum_{i=1}^{k_0} \left( \xi_t + \zeta_t \right)
\]
\[
= \frac{k}{(k_0 - k)k_0} \sum_{i=1}^{k_0} \xi_t + o_p(1),
\]
where the third line uses (11) and the \(o_p(1)\) term in the last line is uniform in \(k \in D_{\eta,M}\) according to Lemmas 2 (i), (v), and (vi).

Let \(v_k\) denote a uniform \(o_p(1)\) term in (22). For any given \(\delta > 0\), (22) implies
\[
P(\max_{k \in D_{\eta,M}, k < k_0} \left\| \frac{1}{k_0 - k} \sum_{i=1}^{k_0} \hat{g}_i \hat{g}_i' - \Sigma_0 \right\| \geq \delta)
\]
\[
\leq P(\max_{k \in D_{\eta,M}, k < k_0} \left\| \frac{1}{k_0 - k} \sum_{i=1}^{k_0} \xi_t \right\| + \|v_k\| \geq \delta)
\]
\[
= P(\max_{k \in D_{\eta,M}, k < k_0} \left\| \frac{1}{k_0 - k} \sum_{i=1}^{k_0} \xi_t \right\| + \|v_k\| \geq \delta, \max_{k \in D_{\eta,M}, k < k_0} \|v_k\| \leq \delta/2)
\]
\[
+ P(\max_{k \in D_{\eta,M}, k < k_0} \left\| \frac{1}{k_0 - k} \sum_{i=1}^{k_0} \xi_t \right\| + \|v_k\| \geq \delta, \max_{k \in D_{\eta,M}, k < k_0} \|v_k\| > \delta/2)
\]
\[
\leq P(\max_{k \in D_{\eta,M}, k < k_0} \left\| \frac{1}{k_0 - k} \sum_{i=1}^{k_0} \xi_t \right\| \geq \delta/2) + o(1)
\]
\[
= P(\max_{(\tau_0 - \eta)T \leq k < k_0 - M} \left\| \frac{1}{k_0 - k} \sum_{i=1}^{k_0} \xi_t \right\| \geq \delta/2) + o(1).
\]

Let \(m = k_0 - k\),
\[
P(\max_{(\tau_0 - \eta)T \leq k < k_0 - M} \left\| \frac{1}{k_0 - k} \sum_{i=1}^{k_0} \xi_t \right\| \geq \delta/2) = P(\max_{M < m \leq \eta T} \left\| \frac{1}{m} \sum_{i=1}^{m} \xi_t \right\| \geq \delta/2)
\]
\[
\leq \frac{4}{\delta^2} \left( 1 + \sum_{i=M+1}^{nT} \frac{1}{i^2} \right)
\]
\[
\leq \frac{4}{\delta^2} \left( \frac{2}{M} - \frac{1}{\eta T} \right)
\]
\[
= \frac{C}{M\delta^2} + o(1) \to 0, \text{ as } M \to \infty,
\]
where \(0 < C < \infty\) is a constant.
Similarly, (21) implies
\[ P(\max_{k \in D_{\eta,M}, k < k_0} \left| \frac{k}{k_0 - k} \log |\hat{\Sigma}_1 \hat{\Sigma}_1^{-1}| \right| \geq \delta) \leq P(\max_{k \in D_{\eta,M}, k < k_0} \sqrt{\frac{1}{k_0 - k}} \sum_{t=0}^{k_0} \xi_t \hat{\Sigma}_t^{0-1} + o_p(1)) \geq \delta) \leq P(\max_{k \in D_{\eta,M}, k < k_0} \sqrt{\frac{1}{k_0 - k}} \sum_{t=0}^{k_0} \xi_t (\hat{\Sigma}_t^{0-1} - \Sigma_1^{-1}) + o_p(1)) \geq \delta \sqrt{\tau}) \leq P(\max_{k \in D_{\eta,M}, k < k_0} \sqrt{\frac{1}{k_0 - k}} \sum_{t=0}^{k_0} \xi_t \Sigma_t^{-1} + o_p(1)) \geq \delta \sqrt{\tau}) \leq \frac{C}{M^{\delta_2}} \to 0, \quad \text{as } M \to \infty, \text{ for some constant } C > 0, \quad (25)\]

where the fourth line follows from \( \max_{k \in D_{\eta,M}, k < k_0} \sum_{t=0}^{k_0} \xi_t (\hat{\Sigma}_t^{0-1} - \Sigma_1^{-1}) = o_p(1) \) by Lemma \( \text{(iii)} \) and the fact that \( ||\hat{\Sigma}_1 - \Sigma_1|| = o_p(1) \). In addition, the last inequality holds through a similar derivation used in \( (23) \) and \( (24) \).

By the continuity of \( g \) defined in \( (17), (19) \) indicates the presence of \( \delta > 0 \) such that \( ||(k_0 - k)^{-1} \sum_{t=t+1}^{t=k_0} \hat{g}_t \hat{g}_t' - \Sigma_0^0|| < \delta \) holds for a sufficiently large \( M \) by \( (24) \) and

\[ \min_{k \in D_{\eta,M}, k < k_0} \frac{T - k}{k_0 - k} \log \left| \frac{k_0 - k}{T - k} \hat{\Sigma}_1 \hat{\Sigma}_1^{-1} + \frac{1}{T - k} \sum_{t=0}^{k_0} \hat{g}_t \hat{g}_t' - \Sigma_0^0 \right| - \log \left| \hat{\Sigma}_2 \hat{\Sigma}_2^{-1} \right| \geq \frac{c_2}{2} > 0, \quad (26)\]

w.p.a.1 as \( N, T \to \infty \). In addition, by \( (25) \), we have

\[ P\left( \left| \min_{k \in D_{\eta,M}, k < k_0} \frac{k}{k_0 - k} \log |\hat{\Sigma}_1 \hat{\Sigma}_1^{-1}| \leq \frac{c_2}{4} \right| \right) \geq P\left( \max_{k \in D_{\eta,M}, k < k_0} \left| \frac{k}{k_0 - k} \log |\hat{\Sigma}_1 \hat{\Sigma}_1^{-1}| \right| \leq \frac{c_2}{4} \right) \geq 1 - \frac{16C}{Mc_2} \to 1 \quad (27)\]

as \( M \to \infty \). Using \( (20) \) and \( (27) \), we can obtain

\[ \min_{k \in D_{\eta,M}, k < k_0} \frac{k}{k_0 - k} \log |\hat{\Sigma}_1 \hat{\Sigma}_1^{-1}| + \frac{T - k}{k_0 - k} \log \left| \frac{k_0 - k}{T - k} \hat{\Sigma}_1 \hat{\Sigma}_1^{-1} + \frac{1}{T - k} \sum_{t=0}^{k_0} \hat{g}_t \hat{g}_t' - \Sigma_0^0 \right| - \log \left| \hat{\Sigma}_2 \hat{\Sigma}_2^{-1} \right| \geq \frac{c_2}{2} \geq \frac{c_2}{4} + \frac{c_2}{4} > 0 \]

w.p.a.1 as \( M \to \infty \). This shows that \( P\left( \min_{k \in D_{\eta,M}, k < k_0} U(k) - U(k_0) \leq 0 < \varepsilon \right) \) for a sufficiently large \( M \). \( \square \)

**Proof of Theorem 2**

Let us recall \( (12) \),

\[ U(k) - U(k_0) = (k_0 - k) \left( \frac{k}{k_0 - k} \log |\hat{\Sigma}_1 \hat{\Sigma}_1^{-1}| + \frac{T - k}{k_0 - k} \log |\hat{\Sigma}_2 \hat{\Sigma}_2^{-1}| - \log |\hat{\Sigma}_1 \hat{\Sigma}_1^{-1}| \right). \]

For the second term in the above equation, we have

\[ (T - k) \log |\hat{\Sigma}_2 \hat{\Sigma}_2^{-1}| = (k - k_0) tr(\Sigma_2 - \Sigma_1) \Sigma_2^{-1} + \sum_{t=k+1}^{k_0} tr(\xi_t \Sigma_2^{-1}) + o_p(1), \]

\[ U(k) - U(k_0) = (k_0 - k) \left( \frac{k}{k_0 - k} \log |\hat{\Sigma}_1 \hat{\Sigma}_1^{-1}| + \frac{T - k}{k_0 - k} \log |\hat{\Sigma}_2 \hat{\Sigma}_2^{-1}| - \log |\hat{\Sigma}_1 \hat{\Sigma}_1^{-1}| \right) + (k - k_0) tr(\Sigma_2 - \Sigma_1) \Sigma_2^{-1} + \sum_{t=k+1}^{k_0} tr(\xi_t \Sigma_2^{-1}) + o_p(1). \]
by Lemma 5. Similarly, by (11) and Lemma 2, we have

\[
\begin{align*}
  k \log |\hat{\Sigma}_1^{10^{-1}}| &= k \log |(\hat{\Sigma}_1 - \hat{\Sigma}_1^{0})\hat{\Sigma}_1^{0-1} + I|, \\
  &= k \log \left[ \frac{k_0 - k}{k k_0} \sum_{t=1}^{k} (\xi_t + \zeta_t) - \frac{1}{k_0} \sum_{t=k+1}^{k_0} (\xi_t + \zeta_t) \right] \hat{\Sigma}_1^{0-1} + I, \\
  &= k \cdot tr\left( \frac{k_0 - k}{k k_0} \sum_{t=1}^{k} (\xi_t + \zeta_t)\hat{\Sigma}_1^{0-1} - k \cdot tr\left( \frac{1}{k_0} \sum_{t=k+1}^{k_0} (\xi_t + \zeta_t)\hat{\Sigma}_1^{0-1} \right) + o_p(1), \\
  &= - \sum_{t=k+1}^{k_0} tr(\xi_t\hat{\Sigma}_1^{-1}) + o_p(1).
\end{align*}
\]

Thus,

\[
U(k) - U(k_0) \to \sum_{t=k+1}^{k_0} tr(\xi_t(\Sigma_2^{-1} - \Sigma_1^{-1})) + (k_0 - k)tr \left( \Sigma_1\Sigma_2^{-1} - r - \log |\Sigma_1\Sigma_2^{-1}| \right)
\]

Similarly, for the case of \( k > k_0 \), the limit can be written as \( \sum_{t=k_0+1}^{k} tr(\xi_t(\Sigma_2^{-1} - \Sigma_1^{-1})) + (k_0 - k)tr \left( \Sigma_1\Sigma_2^{-1} - r - \log |\Sigma_1\Sigma_2^{-1}| \right) \).

\( \Box \)

**\( \Sigma_1 \) or \( \Sigma_2 \), or both, is a singular matrix.**

Before proving the theorem, we need to prove the following lemmas, where \( A^- \) denotes the MP inverse of \( A \), \( \rho_i(A) \) represents the \( i \)-th eigenvalue of matrix \( A \), and \( \sigma_i(A) \) represents the \( i \)-th singular value of matrix \( A \).

**Lemma 6. Under Assumptions**

\[
\max_{k \in [T/2]} \sum_{s=1}^{k} \left\| \sum_{t=1}^{T} \hat{g}_t e'_s e_s / NT \right\|^2 = O_p(\delta N^2)
\]

\[
\max_{k \in [k_0, [2T/2]]} \sum_{s=k+1}^{T} \left\| \sum_{t=1}^{T} \hat{g}_t e'_s e_s / NT \right\|^2 = O_p(\delta N^2).
\]

**Proof:** By symmetry, it is sufficient to focus on the case of \( k \in [k_0, [2T/2]] \).

\[
\frac{1}{N^2T^2(T-k)} \sum_{s=k+1}^{T} \left\| \sum_{t=1}^{T} e'_s e_t \right\|^2 \leq \frac{2}{N^2T^2(T-k)} \sum_{s=k+1}^{T} \left\| \sum_{t=1}^{T} (\hat{g}_t - H'g_t)e'_s e_s \right\|^2 + \frac{2}{N^2T^2(T-k)} \sum_{s=k+1}^{T} \left\| \sum_{t=1}^{T} H'g_t e'_s e_s \right\|^2
\]

Recall that \( E(e'_s e_s) / N \equiv \gamma_N(s, t) \). Consider the equation

\[
\max_{k \in [k_0, [2T/2]]} \frac{2}{N^2T^2(T-k)} \sum_{s=k+1}^{T} \left\| \sum_{t=1}^{T} H'g_t e'_s e_s \right\|^2 \leq \frac{2}{T^3(1 - \tau_2)} \sum_{s=1}^{T} \left\| \sum_{t=1}^{T} H'g_t e'_s e_s - E(e'_s e_s) / N \right\|^2 + \frac{2}{T^3(1 - \tau_2)} \sum_{s=1}^{T} \left\| \sum_{t=1}^{T} H'g_t \gamma_N(s, t) \right\|^2,
\]

35
where the first term can be written as
\[ \frac{2}{T^2(1 - \tau_2)N} \sum_{s=1}^{T} \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} H' g_t (e_i e_s - E(e_i e_s)) \right\|^2 = O_p \left( \frac{1}{NT} \right) \] (28)
under Assumption B(ii) and the second term is \( O_p(T^{-2}) \) because the expectation can be bounded by
\[ \frac{2}{T^3} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{u=1}^{T} E(\|g_t\|\|g_u\|) \gamma_N(s,t) |\gamma_N(s,u)| \]
\[ \leq \frac{2}{T^3} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{u=1}^{T} \max_t E(\|g_t\|^2) \gamma_N(s,t) |\gamma_N(s,u)| \]
\[ \leq \frac{2}{T^3} \sum_{s=1}^{T} \max_t E(\|g_t\|^2) \left( \sum_{t=1}^{T} \gamma_N(s,t) \right)^2 = O(T^{-2}), \] (29)
where we use the facts that \( E(\|g_t\|\|g_u\|) \leq |E(\|g_t\|^2)E(\|g_u\|^2)|^{1/2} \leq \max_t E(\|g_t\|^2) \) under Assumptions A and \( \sum_{t=1}^{T} \gamma_N(s,t) \leq M \) by Assumption B(ii).

Next, consider the term
\[ \max_{k \in [k_0, \tau_2 T]} \frac{1}{N^2 T^2 (T - k)} \sum_{s=k+1}^{T} \left\| \sum_{t=1}^{T} (\hat{g}_t - H' g_t) e_i e_s - E(e_i e_s) \right\|^2 \]
\[ \leq \frac{2}{T^3(1 - \tau_2)} \sum_{s=1}^{T} \sum_{t=1}^{T} \left\| \hat{g}_t - H' g_t \right\|^2 \sum_{i=1}^{T} \sum_{t=1}^{T} \left\| e_i e_s - E(e_i e_s) \right\|^2 \]
\[ = \frac{2}{T^3(1 - \tau_2)} \sum_{s=1}^{T} \sum_{t=1}^{T} \left\| \hat{g}_t - H' g_t \right\|^2 \cdot \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} \left\| e_i e_s - E(e_i e_s) \right\|^2 \]
\[ = O_p \left( \frac{1}{N \delta^2 N_T} \right), \] (30)
by Assumption B(v) and the second term is bounded by
\[ \frac{2}{T^3(1 - \tau_2)} \sum_{s=1}^{T} \left\| \sum_{t=1}^{T} (\hat{g}_t - H' g_t) \gamma_N(s,t) \right\|^2 \leq \frac{2}{T} \sum_{t=1}^{T} \left\| \hat{g}_t - H' g_t \right\|^2 \frac{1}{T^2(1 - \tau_2)} \sum_{s=1}^{T} \sum_{t=1}^{T} |\gamma_N(s,t)| \]
\[ = O_p \left( \frac{1}{T \delta N_T} \right) \] (31)
because \( \sum_{t=1}^{T} |\gamma_N(s,t)| \leq (\sum_{t=1}^{T} |\gamma_N(s,t)|)^2 \leq M^2 \) under Assumption B(ii). Combining the results obtained in (28), (29), we obtain the desired result. \( \square \)

When \( C \) is singular, \( \hat{\Sigma}_2(k) \) converges in probability to a singular matrix for \( k \geq k_0 \). In finite samples, however, the smallest eigenvalue of \( \hat{\Sigma}_2(k) \) is not zero. The following lemma establishes a lower bound for the smallest eigenvalue of \( \hat{\Sigma}_2(k) \), which ensures that it is meaningful to compute the logarithm of \( |\hat{\Sigma}_2(k)| \) in the objective function for any given sample size. Symmetrically, a similar lower bound can be established for the smallest eigenvalue of \( \hat{\Sigma}_1(k) \) for \( k \leq k_0 \) when \( B \) is singular. Because of space restrictions, Proposition A here only states the result for the case of \( \hat{\Sigma}_2(k) \).
**Proposition 1.** Under Assumptions 1–10, for \( k \geq k_0 \) and \( k \leq [T/\tau_2] \), if \( C \) is singular and \( \sqrt{N}/T \to 0 \) as \( N, T \to \infty \), there exists a constant \( c_U \geq c_L > 0 \) such that

\[
P \left( \min_{k \in [k_0, [T/\tau_2]]} \rho_j(\hat{\Sigma}_2(k)) \geq \frac{c_L}{N} \right) \to 1,
\]

\[
P \left( \max_{k \in [k_0, [T/\tau_2]]} \rho_j(\hat{\Sigma}_2(k)) \leq \frac{c_U}{N} \right) \to 1,
\]

for \( j = r_2 + 1, \ldots, r \).

**Proof:**

**Part 1.** For \( k \geq k_0 \), \( \hat{\Sigma}_2(k) = (T - k)^{-1}\hat{G}_2^k \hat{G}_2^k \), where \( \hat{G}_2^k = [\hat{g}_{k+1}, \ldots, \hat{g}_T]' \). Let \( X_2^k = [X_{k+1}, \ldots, X_T]' \), \( e_k^2 = [e_{k+1}, \ldots, e_T]' \), and \( G_2^k = [G_{k+1}, \ldots, G_T]' \).

From \( XX'\hat{G}/NT = \hat{G}V_{NT} \), eq. (2) implies

\[
\hat{G}_2^k = X_2^k X' \hat{G}V_{NT}^{-1}/NT = \frac{1}{NT}(G_2^k \Lambda' + e_2^k)(\Lambda G' + e')\hat{G}V_{NT}^{-1}
\]

\[
\hat{G}_2^k - G_2^k H = \frac{1}{NT} e_2^k e' \hat{G}V_{NT}^{-1} + \frac{1}{NT} G_2^k \Lambda' \hat{G}V_{NT}^{-1} + \frac{1}{NT} G_2^k \Lambda' e' \hat{G}V_{NT}^{-1}.
\]

(32)

In addition, note that

\[
\hat{\Sigma}_2(k) - \frac{1}{T-k} \hat{G}_2^k M_{F_2^k} \hat{G}_2^k = \frac{1}{T-k} \hat{G}_2^k P_{F_2^k} \hat{G}_2^k \geq 0,
\]

where \( P_{F_2^k} = F_2^k (F_2^k F_2^k)'^{-1} F_2^k \), \( M_{F_2^k} = I_{T-k} - P_{F_2^k} \), and \( F_2^k = [f_{k+1}, \ldots, f_T]' \). Thus, Weyl’s inequality for eigenvalues implies

\[
\min_{k \in [k_0, [T/\tau_2]]} \rho_j(\hat{\Sigma}_2(k)) \geq \min_{k \in [k_0, [T/\tau_2]]} \left[ \rho_j \left( \frac{1}{T-k} \hat{G}_2^k M_{F_2^k} \hat{G}_2^k \right) + \rho_j \left( \frac{1}{T-k} \hat{G}_2^k P_{F_2^k} \hat{G}_2^k \right) \right] \geq \min_{k \in [k_0, [T/\tau_2]]} \rho_j \left( \frac{1}{T-k} \hat{G}_2^k M_{F_2^k} \hat{G}_2^k \right), \text{ for } j = r_2 + 1, \ldots, r
\]

(33)

Thus, it suffices to find the lower bound for \( \min_{k \in [k_0, [T/\tau_2]]} \rho_j(\hat{G}_2^k M_{F_2^k} \hat{G}_2^k (T-k)) \). As \( F_2^k C' = G_2^k \) for \( k \geq k_0 \), we have \( M_{F_2^k} G_2^k = 0 \) and

\[
\frac{1}{T-k} \hat{G}_2^k M_{F_2^k} \hat{G}_2^k = \frac{1}{T-k} (\hat{G}_2^k - G_2^k H)' M_{F_2^k} (\hat{G}_2^k - G_2^k H).
\]

(34)

Now, using (32), we can obtain

\[
\frac{1}{\sqrt{T-k}} M_{F_2^k} (\hat{G}_2^k - G_2^k H) = \frac{1}{\sqrt{T-k}} M_{F_2^k} \left( \frac{1}{NT} e_2^k \Lambda G' \hat{G}V_{NT}^{-1} + \frac{1}{NT} G_2^k \Lambda' \hat{G}V_{NT}^{-1} \right)
\]

\[
= a_{1k} + a_{2k} + a_{3k}.
\]

(35)

Let us consider the term \( a_{1k} \) in (35). As \( \sigma_i(A + B) \leq \sigma_i(A) + \sigma_1(B) \), we have

\[
\sigma_1 \left( \frac{e_2^k \Lambda}{\sqrt{NT-k}} (\hat{G}_2^k) V_{NT}^{-1} \right) \leq \sigma_1 \left( M_{F_2^k} \frac{e_2^k \Lambda}{\sqrt{NT-k}} G' \hat{G} V_{NT}^{-1} \right) + \sigma_1 \left( P_{F_2^k} \frac{e_2^k \Lambda}{\sqrt{NT-k}} G' \hat{G} V_{NT}^{-1} \right).
\]
which implies
\[
\begin{align*}
&\min_{k \in [k_0, [r_2T]]} \sigma_j \left( M_{F_2} \frac{\epsilon_k^2}{N \sqrt{T-k}} \frac{G' \hat{G} V_{NT}^{-1}}{T} \right) \\
&\geq \min_{k \in [k_0, [r_2T]]} \sigma_j \left( \frac{\epsilon_k^2}{N \sqrt{T-k}} \frac{G' \hat{G} V_{NT}^{-1}}{T} \right) - \max_{k \in [k_0, [r_2T]]} \sigma_1 \left( P_{F_2} \frac{\epsilon_k^2}{N \sqrt{T-k}} \frac{G' \hat{G} V_{NT}^{-1}}{T} \right) \\
&\geq \frac{1}{\sqrt{N}} \sigma_j \left( \frac{G' \hat{G} V_{NT}^{-1}}{T} \right) \sqrt{\min_{k \in [k_0, [r_2T]]} \rho_j \left( \frac{N \epsilon_k^2}{N \sqrt{T-k}} \right)} \\
&\geq \frac{1}{\sqrt{N}} \sigma_j \left( \frac{G' \hat{G} V_{NT}^{-1}}{T} \right) + O_p \left( \frac{1}{\sqrt{N}} \right) \geq \frac{1}{\sqrt{N}} c, \text{ w.p.a.1 for some } c > 0
\end{align*}
\]  
(36)

where the third and fourth lines use the inequality \( \sigma_j(\mathbb{E}) \geq \sigma_j(\hat{\mathbb{E}}) \sigma_r(\mathbb{E}) \) and the relation \( \rho_r(\mathbb{A}' \mathbb{A})^{1/2} = \sigma_r(\mathbb{A}) \), and the fifth line uses Assumptions 9 and 10 and the fact that \( \frac{G' \hat{G} V_{NT}^{-1}}{T} \) is nonsingular as \( N, T \to \infty \) by Proposition 1 and Lemma A.3 of Bai (2003).

The term \( a_{2k} \) in (35) is zero because \( M_{F_2} G_k^h = 0 \). For term \( a_{3k} \) in (35), we can obtain its upper bound as
\[
\frac{1}{N^2 T^2 (T-k)} \| V_{NT}^{-1} \hat{G}' e_2 e_2' \hat{G} V_{NT}^{-1} \| \leq \frac{1}{N^2 T^2 (T-k)} \sum_{s=k+1}^{T} \sum_{t=1}^{T} \hat{g}_s e_t e_s' \sum_{u=1}^{T} e_u \hat{g}_u V_{NT}^{-1} \|
\leq \frac{1}{T-k} \| V_{NT}^{-1} \| \sum_{s=k+1}^{T} \| \hat{g}_s e_t e_s / NT \| ^2.
\]

Note that
\[
\max_{k \in [k_0, [r_2T]]} (T-k)^{-1} \sum_{s=k+1}^{T} \| \hat{g}_s e_t e_s / NT \| ^2 = O_p(\delta_{NT}^{-4})
\]  
(38)

by Lemma 6.

For the second term in (37), we have
\[
\max_{k \in [k_0, [r_2T]]} \left\| V_{NT}^{-1} \hat{G}' e_2 e_2' \hat{G} V_{NT}^{-1} \right\| = O_p(\delta_{NT}^{-4}).
\]  
(39)

To observe this, note that \( \left\| F_2^h F_2' / (T-k) \right\|^{-1} = O_p(1) \) uniformly over \( k \in [k_0, [r_2T]] \) and
\[
\begin{align*}
\frac{1}{NT(T-k)} \hat{G}' e_2 e_2' \frac{F_2^h}{T} &= \frac{1}{NT(T-k)} \sum_{s=k+1}^{T} \sum_{t=1}^{T} \hat{g}_s e_t e_s' f_s' \\
&= \frac{1}{T(T-k)} \sum_{s=k+1}^{T} \sum_{t=1}^{T} \hat{g}_s e_t e_s / N - \gamma_N(s,t) f_s' + \frac{1}{T(T-k)} \sum_{s=k+1}^{T} \sum_{t=1}^{T} \hat{g}_s e_t e_s / N - \gamma_N(s,t) f_s' = O_p(\delta_{NT}^{-2})
\end{align*}
\]  
38
uniformly over \( k \in [k_0, [r_2 T]] \) following the derivation of terms I and II in Lemma B.2 of Bai (2003).

Thus, combining the results in (37)–(59), we have

\[
\max_{k\in[k_0,[r_2 T]]} \frac{1}{NT \sqrt{T-k}} M_{F_2} e_{F_2} e_G' \tilde{G} V_{NT}^{-1} \leq \max_{k\in[k_0,[r_2 T]]} \frac{1}{NT \sqrt{T-k}} M_{F_2} e_{F_2} e_G' \tilde{G} V_{NT}^{-1} = O_P(\delta_{NT}^{-2}). \tag{40}
\]

Next, rearranging the terms in (55) yields \( a_{1k} = \frac{1}{\sqrt{T-k}} M_{F_2} (\tilde{G}_2^k - G_2^k H) - a_{3k} \), which implies that \( \sigma_j(a_{1k}) \leq \sigma_j \left( \frac{1}{\sqrt{T-k}} M_{F_2} (\tilde{G}_2^k - G_2^k H) \right) + \sigma_1(-a_{3k}) \) and

\[
\frac{1}{\sqrt{T-k}} M_{F_2} (\tilde{G}_2^k - G_2^k H) \geq \min_{k\in[k_0,[r_2 T]]} \sigma_j(a_{1k}) - \max_{k\in[k_0,[r_2 T]]} \sigma_1(a_{3k}) \geq \frac{1}{\sqrt{N}} c, \quad \text{w.p.a.} 1 \text{ for some } c > 0 \tag{41}
\]

because \( \sigma_1(a_{3k}) \) is uniformly \( O_P(\delta_{NT}^{-2}) \) by (40) and dominated by \( \sigma_j(a_{1k}) \) in (53) under the condition that \( \sqrt{N}/T \to 0 \) as \( N, T \to \infty \). Hence, combining (53), (51), and (41) yields

\[
\min_{k\in[k_0,[r_2 T]]} \rho_j(\tilde{\Sigma}_2(k)) \geq \min_{k\in[k_0,[r_2 T]]} \rho_j \left( \frac{1}{\sqrt{T-k}} (\tilde{G}_2^k - G_2^k H)' M_{F_2} (\tilde{G}_2^k - G_2^k H) \right) = \min_{k\in[k_0,[r_2 T]]} \sigma_j \left( \frac{1}{\sqrt{T-k}} M_{F_2} (\tilde{G}_2^k - G_2^k H) \right)^2 \geq \frac{1}{N} c^2 = \frac{1}{N c_U},
\]

w.p.a. 1 if \( \sqrt{N}/T \to 0 \) as \( N, T \to \infty \).

**Part 2.** Note that

\[
\sigma_j(\tilde{G}_2^k/\sqrt{T-k}) \leq \sigma_j(\tilde{G}_2^k H/\sqrt{T-k}) + \sigma_1((\tilde{G}_2^k - G_2^k H)/\sqrt{T-k}).
\]

In addition, \( \sigma_j(\tilde{G}_2^k H/\sqrt{T-k}) = 0 \) for \( r_2 < j \leq r \) and

\[
\max_{k\in[k_0,[r_2 T]]} \sigma_1((\tilde{G}_2^k - G_2^k H)/\sqrt{T-k}) \leq \max_{k\in[k_0,[r_2 T]]} \frac{1}{\sqrt{T-k}} \| \tilde{G}_2^k - G_2^k H \| \leq \frac{1}{\sqrt{(1-\tau_2)T}} \| \tilde{G} - GH \| \leq \frac{c}{\sqrt{N}}
\]

w.p.a. 1 for some \( 0 < c < \infty \) by Lemma 1 (ii) if \( \sqrt{N}/T \to 0 \) as \( N, T \to \infty \). Thus,

\[
\max_{k\in[k_0,[r_2 T]]} \rho_j(\tilde{\Sigma}_2(k)) = \max_{k\in[k_0,[r_2 T]]} \sigma_j((\tilde{G}_2^k/\sqrt{T-k})^2) \leq c_U/N, \quad \text{w.p.a.} 1.
\]

as \( N, T \to \infty \) for \( j = r_2 + 1, \ldots, r \) and some positive constant \( c_U \).

The following lemma yields a bound on the difference between \( |\tilde{\Sigma}_2(k)| \) and \( |\tilde{\Sigma}_2(k)| \) for \( k_0 < k \leq \tau_2 T \) when \( C \) is singular. The same result applies to the difference between \( |\tilde{\Sigma}_1(k)| \) and \( |\tilde{\Sigma}_1(k)| \) for \( \tau_1 T \leq k < k_0 \) when \( B \) is singular.

**Lemma 7.** Under Assumptions 7–10, for \( k > k_0 \) and \( k \leq [r_2 T] \), if \( C \) is singular and \( T/N \to \kappa \) as \( N, T \to \infty \) for \( 0 < \kappa < \infty \), then

\[
\max_{k_0 < k \leq [r_2 T]} \frac{1}{k - k_0} \| \tilde{\Sigma}_2(k) - |\tilde{\Sigma}_2(k)| \| = O_P(T^{-(r-r_2)-1}).
\]

**Proof:**
First, note that
\[
\hat{\Sigma}_2^0 = \frac{1}{T - k_0} \sum_{t=k_0+1}^T \hat{g}_t^2 = \frac{1}{T - k_0} \sum_{t=k+1}^T \hat{g}_t^2 + \frac{1}{T - k_0} \sum_{t=k_0+1}^k \hat{g}_t^2,
\]
where \( A_k \equiv (T - k_0)^{-1} \sum_{t=k+1}^T \hat{g}_t^2 \) and \( \hat{\Sigma} \equiv [\hat{g}_{k+1}, \ldots, \hat{g}_k]^\top \). By (7.10) of Lange (2010), we have
\[
|\hat{\Sigma}_2^0| = |A_k| \cdot |I_{k-k_0} + \frac{1}{T - k_0} \hat{g} A_k^{-1}\hat{g}^\top|, \tag{42}
\]
where \( A_k^{-1} \) is reasonable because the smallest eigenvalue of \( N \cdot A_k \) is bounded away from zero by proposition \( \blacksquare \).

We now analyze the term \((T - k_0)^{-1}\hat{g} A_k^{-1}\hat{g}^\top\), which can be written as
\[
\frac{1}{T - k_0} \hat{g} A_k^{-1}\hat{g}^\top = \frac{1}{|A_k|} \hat{g} A_k^{-1}\hat{g} \left( \frac{1}{T - k_0} \right) = \frac{1}{|A_k|} \hat{g} \left( A_k^{-1} - \frac{1}{T - k_0} \sum_{t=k+1}^T H' g_t g_t H \right) \hat{g}^\top + \frac{1}{|A_k|} \hat{g} \left( \frac{1}{T - k_0} \sum_{t=k+1}^T H' g_t g_t H \right) \hat{g}^\top = S_1 + S_2. \tag{43}
\]
For \( S_1 \),
\[
\max_{k_0 < k \leq T} \| A_k - \frac{1}{T - k_0} \sum_{t=k+1}^T H' g_t g_t H \| = O_p(N^{-1}) \tag{44}
\]
by a uniform version of Lemmas B2 and B3 of Bai (2009). When \( r = r_2 = 1 \), \( \blacksquare \) implies
\[
\max_{k_0 < k \leq T} \left\| \left( T - k_0 \right)^{-1} \sum_{t=k+1}^T H' g_t g_t H \right\| = O_p(N^{-1}).
\]
Then, it follows that
\[
\max_{k_0 < k \leq T} \frac{1}{k - k_0} \| \hat{g} \|^2 \leq \max_{k_0 < k \leq T} \frac{2}{k - k_0} \sum_{t=k_0+1}^k \| \hat{g}_t - H' g_t \|^2 + \frac{2}{k - k_0} \sum_{t=k_0+1}^k \| H' g_t \|^2 = O_p(1)
\]
by Lemma \( \blacksquare \).

When \( r = r_2 \geq 2 \),
\[
\rho_1(A_k) = |A_k|/\rho_r(A_k) = O_p(N^{-(r-2)-1})
\]
onlyx{uniformly over \( k_0 < k \leq \tau T \) by proposition \( \blacksquare \) thus, we have
\[
\| A_k \| \leq \sqrt{r} \rho_1(A_k) = \sqrt{r} \rho_1(A_k) = O_p(N^{-(r-2)-1}) \quad \text{for} \quad r - r_2 \geq 2. \tag{46}
\]
Now, let \( f_t = \Sigma_f^{1/2} \varepsilon_t \) with \( E \varepsilon_t \varepsilon_t^\top = I_r \). Hence, for \( k \geq k_0 \), we have \( g_{k+1} = C \Sigma_f^{1/2} \varepsilon_{k+1} \). By \( \blacksquare \) and
\[
\max_{k_0 < k \leq T} \| \frac{1}{k - k_0} \sum_{t=k+1}^T f_t f_t^\top - \Sigma_f \| = \max_{k_0 < k \leq T} \| \frac{1}{k - k_0} \sum_{t=k+1}^T \varepsilon_t \| = O_p(T^{-1/2}) \quad \text{by Hájek-Rényi inequality, we have}
\max_{k_0 < k \leq T} \| H' g_t g_t H - H' C \Sigma_f C' H \| = O_p(T^{-1/2}) \quad \text{and}
\max_{k_0 < k \leq T} \| A_k = H' C \Sigma_f C' H \| = O_p(\delta^{-1}_N T);
thus,

\[ \max_{k_0 < k \leq \tau_2 T} \| A_k^{1/2} U_k - H' \Sigma_f^{1/2} \| = O_p(\delta_{N_T}^{1/2}), \]  

(47)

where \( U_k = A_k^{1/2} (\Sigma_f^{1/2} C'H)^{-1} \). Therefore, for \( t > k_0 \), we have

\[
\hat{g}_t = H'Cf_t + (\hat{g}_t - H'g_t) = H'C \Sigma_f^{1/2} \epsilon_t + (\hat{g}_t - H'g_t) \\
= A_k^{1/2} U_k \epsilon_t + (H'C \Sigma_f^{1/2} - A_k^{1/2} U_k) \epsilon_t + (\hat{g}_t - H'g_t) \\
= A_k^{1/2} U_k \epsilon_t + O_p(\delta_{N_T}^{1/2}) \epsilon_t + (\hat{g}_t - H'g_t)
\]

by (47) and

\[
\hat{G}' - A_k^{1/2} U_k \mathcal{E}' = O_p(\delta_{N_T}^{1/2}) \cdot \mathcal{E}' + \hat{g}_t - H'g_t
\]

(48)

where \( \mathcal{E}' \equiv [\epsilon_{k_0+1}, \ldots, \epsilon_k] \), \( \mathcal{G}' \equiv [g_{k_0+1}, \ldots, g_k] \), and \( O_p(\delta_{N_T}^{1/2}) \) term is uniform in \( k_0 < k \leq \tau_2 T \). In addition,

\[
\left[ \frac{1}{T - k_0} \sum_{t = k_0 + 1}^T \| H'g_t g'_t H \| \right]^{\#} = 0, \text{ for } r - r_2 \geq 2.
\]

Thus, we have

\[
\frac{1}{T - k_0} \max_{k_0 < k \leq \tau_2 T} \left\| \frac{1}{k - k_0} \hat{G} \left( A_k^{\#} - \left[ \frac{1}{T - k_0} \sum_{t = k_0 + 1}^T \| H'g_t g'_t H \| \right]^{\#} \right) \mathcal{G}' \right\|
\]

\[
= \frac{1}{T - k_0} \max_{k_0 < k \leq \tau_2 T} \left\| \frac{1}{k - k_0} \mathcal{E} U_k' A_k^{1/2} A_k^{\#} A_k^{1/2} U_k \mathcal{E}' + \frac{1}{k - k_0} (\hat{G} - \mathcal{E} U_k' A_k^{1/2} A_k^{\#} A_k^{1/2} U_k \mathcal{E}') \right. \\
+ \frac{1}{k - k_0} \mathcal{E} U_k' A_k^{1/2} A_k^{\#} (\mathcal{G}' - A_k^{1/2} U_k \mathcal{E}') + \frac{1}{k - k_0} (\hat{G} - \mathcal{E} U_k' A_k^{1/2} A_k^{\#} (\mathcal{G}' - A_k^{1/2} U_k \mathcal{E}')) \left. \left\| \right. \right. \\
\leq \frac{1}{T - k_0} \max_{k_0 < k \leq \tau_2 T} \left\| \frac{1}{k - k_0} \| \epsilon_t \|^2 \| U_k \|^2 + \frac{2}{k - k_0} \max_{k_0 < k \leq \tau_2 T} \frac{1}{k - k_0} \| \hat{G} - \mathcal{E} U_k' A_k^{1/2} \| \left\| (A_k^{\#})^{\#} \right\| \left\| U_k \mathcal{E}' \right\| \right\| \right\| \right\| \right\}^{\#} \|
\]

\[
\leq \frac{1}{T - k_0} \max_{k_0 < k \leq \tau_2 T} \frac{1}{k - k_0} \| \hat{G} - \mathcal{E} U_k' A_k^{1/2} \|^2 \| A_k^{\#} \|
\]

(49)

where we use the fact that \( A_k^{\#} = (A_k^{1/2})^{\#} (A_k^{1/2})^{\#} \) and \( (A_k^{1/2})^{\#} A_k^{1/2} = |A_k|^{1/2} I_r \). The definition of \( U_k \) implies that \( U_k = O_p(1) \) uniformly over \( k_0 < k \leq \tau_2 T \). The first term in (49) is \( O_p(N^{-(r - r_2)} T^{-1}) \) because \( |A_k| = O_p(N^{-(r - r_2)}) \) by proposition [1] the second term is \( O_p(N^{-(r - r_2)} \sqrt{N \delta_{N_T}^{-1}} T^{-1}) \) because \( |A_k^{1/2}| = O_p(N^{-(r - r_2)/2}) \), \( (k - k_0)^{-1/2} \| \hat{G} - B_k^{1/2} U \mathcal{E}' \| \) is uniformly \( O_p(\delta_{N_T}^{-1}) \) by \([38]\) and Lemma [1] \( (A_k^{1/2})^{\#} \) is uniformly \( O_p(N^{-(r - r_2 - 1)/2}) \) by \([39]\), and the last term in (49) is \( O_p(\delta_{N_T}^{-2} N^{-(r - r_2 - 1)} T^{-1}) \) by \([10], [38], \) and Lemma [1]. The result in (49) indicates that

\[
\max_{k_0 < k \leq \tau_2 T} \| \frac{1}{k - k_0} S_1 \| = O_p(T^{-1})
\]

(50)

under the condition \( N \propto T \). Recalling (45), we obtain the same rate of \( S_1 \) for both \( r_2 = r - 1 \) and \( r_2 \leq r - 2 \).

Term \( S_2 \) in (43) is zero if \( r_2 \leq r - 2 \) because the adjoint matrix of an \( r \times r \) matrix \( \hat{A} \) is zero when \( \text{rank}(\hat{A}) \leq r - 2 \).
When \( r_2 = r - 1 \), we have

\[
\max_{k_0 < k \leq r_2 T} \frac{1}{|A_k|} \left\| \frac{1}{T - k_0} \sum_{t=k+1}^{T} H' g_t' g_t H \right\| = O_p\left(1\right),
\]

(51)

Thus, combining the results in (43), (45), (51), and (51), we obtain

\[
\frac{1}{T - k_0} \max_{k_0 < k \leq r_2 T} \left\| \frac{1}{k - k_0} \hat{G} A_k^{-1} \hat{G}' \right\| = O_p\left(T^{-1}\right).
\]

(52)

Thus, (12) can be written as

\[
|\tilde{S}_2^0| = |A_k| \prod_{j=1}^{r} \left[ 1 + \rho_j \left( \frac{1}{T - k_0} \hat{G} A_k^{-1} \hat{G}' \right) \right]^{r_j} \leq |A_k| \left[ 1 \right. + \left. 0 \right] \left( \frac{1}{T - k_0} \hat{G} A_k^{-1} \hat{G}' \right)^{r_j} \leq |A_k| \left[ 1 + O_p \left( \frac{k - k_0}{T} \right) \right]^{r_j},
\]

where we use (52) and the fact that \( \rho_1(\hat{G} A_k^{-1} \hat{G}') = \sigma_1(\hat{G} A_k^{-1} \hat{G}') \leq \|\hat{G} A_k^{-1} \hat{G}'\| \). Thus, by proposition (1) we have

\[
0 < \frac{1}{k - k_0} (|\tilde{S}_2^0| - |A_k|) \leq O_p \left( \frac{1}{T} \right) |A_k| = O_p \left( T^{-1} N^{-(r_2 - 1)} \right),
\]

(53)

where the \( O_p\left(T^{-1} N^{-(r_2 - 1)}\right) \) term is uniform over \( k_0 < k \leq r_2 T \).

Next, comparing \( |A_k| \) and \( |\tilde{S}_2(k)| \), we have

\[
\max_{k_0 < k \leq r_2 T} \frac{1}{k - k_0} \left| |A_k| - |\tilde{S}_2(k)| \right|
\]

\[
= \max_{k_0 < k \leq r_2 T} \frac{1}{k - k_0} \left| \frac{1}{T - k_0} \sum_{t=k+1}^{T} \hat{g}_t g_i \right| - \frac{1}{T - k} \sum_{t=k+1}^{T} \hat{g}_t g_i \right| \]

\[
= \max_{k_0 < k \leq r_2 T} \frac{1}{k - k_0} \left( \frac{k - k_0}{T - k_0} \right)^r \left| |\tilde{S}_2(k)| \right| \]

\[
= O_p \left( T^{-1} N^{-(r_2 - 1)} \right),
\]

(54)
where we use the fact that $\max_{k_0 < k < r_2 T} |\tilde{\Sigma}_2(k)| = O_p(N^{-(r-r_2)})$ by proposition 1. As both \[53\] and \[54\] are shown to be $O_p(T^{-1}N^{-(r-r_2)})$, we obtain the desired result for this lemma under the condition $T \propto N$. □

The following lemma yields a lower bound on the difference between $|\tilde{\Sigma}_2(k)|$ and $|\tilde{\Sigma}_2^0|$ for $\tau_1 T \leq k < k_0$ when $C$ is singular and $B$ is either singular or nonsingular. The same result applies to the difference between $|\tilde{\Sigma}_1(k)|$ and $|\tilde{\Sigma}_1^0|$ for $k_0 < k \leq \tau_2 T$ when $B$ is singular.

**Lemma 8.** Under Assumptions 2-11, for $\tau_1 T \leq k < k_0$, if $C$ is singular and $T/N \to \kappa$ as $N, T \to \infty$ for $0 < \kappa < \infty$, then

$$\frac{|\tilde{\Sigma}_2(k) - |\tilde{\Sigma}_2^0|}{|\tilde{\Sigma}_2^0|} \geq c \cdot (k_0 - k) \text{ w.p.a.1}$$

for a constant $c > 0$ as $N, T \to \infty$.

**Proof:**

Let us rewrite $\tilde{\Sigma}_2(k)$ as

$$\tilde{\Sigma}_2(k) = 1 - k \sum_{t=k+1}^T \hat{g}_t \hat{g}_t' = 1 - k \sum_{t=k_0+1}^T \hat{g}_t \hat{g}_t' + 1 - k \sum_{t=k+1}^{k_0} \hat{g}_t \hat{g}_t'$$

$$= D_k + \frac{1}{T-k} \hat{G} \hat{G}'$$

where $D_k \equiv (T-k)^{-1} \sum_{t=k_0+1}^T \hat{g}_t \hat{g}_t'$ and $\hat{G} \equiv [\hat{g}_{k+1}, ..., \hat{g}_{k_0}]'$. By (7.10) of [Lange (2010)], we have

$$|\tilde{\Sigma}_2(k)| = |D_k| \cdot |I_{k_0-k} + \frac{1}{T-k} \hat{G} D_k^{-1} \hat{G}'|$$

$$\geq |D_k| \left[ 1 + \rho_1 \left( \frac{1}{T-k} \hat{G} D_k^{-1} \hat{G}' \right) \right]$$

(55)

We would like to find the lower bound of the largest eigenvalue of matrix $\frac{1}{T-k} \hat{G} D_k^{-1} \hat{G}'$, which can be written as

$$\frac{1}{T-k} \hat{G} D_k^{-1} \hat{G}' = \frac{1}{|D_k|} \left( \frac{T-k}{T-k} \hat{G} D_k^{-1} \hat{G}' \right)$$

$$= \frac{1}{|D_k|(T-k)} \hat{G} \left( D_k^# - \frac{1}{T-k} \sum_{t=k_0+1}^T H' g_t g_t H \right) \hat{G}' + \frac{1}{|D_k|(T-k)} \hat{G} \left( \frac{1}{T-k} \sum_{t=k_0+1}^T H' g_t g_t H \right) \hat{G}'$$

(56)

$$\equiv \frac{1}{|D_k|} (P_1 + P_2).$$

The subsequent proof will be performed in two steps.

**Step 1.** When $r - 1 = r_2 > 0$, we have

$$\max_{\tau_1 T \leq k < k_0} \frac{1}{k_0 - k} \| P_1 \| = O_p(T^{-1}N^{-1})$$

(57)

because $\max_{\tau_1 T \leq k < k_0} \frac{1}{k_0 - k} \| \hat{G} \|^2 = O_p(1)$,
by \[\rho_t(\mathbb{P}_1) \geq -\frac{\max_j |\rho_j(\mathbb{P}_1)|}{k_0-k} = -\frac{\sigma_t(\mathbb{P}_1)}{k_0-k} \geq -\frac{\|\mathbb{P}_1\|}{k_0-k} = O_p(T^{-1}N^{-1}) \] (58)
given the fact that the singular values are absolute values of the eigenvalues of a symmetric matrix.

Next, for the term $\mathbb{P}_2$ in (56), we have

\[
\begin{align*}
\hat{G} \left[ \frac{1}{T-k} \sum_{t=k_0+1}^{T} H' g_t g_t' H \right] & \leq \hat{G} H' C^\dagger \left[ \frac{1}{T-k} \sum_{t=k_0+1}^{T} f_t f_t' \right] C^\dagger H' \hat{G}' \\
& = [(\hat{G} - \mathbb{G}H) + \mathbb{G}H] H' C^\dagger \left[ \frac{1}{T-k} \sum_{t=k_0+1}^{T} f_t f_t' \right] C^\dagger H' \hat{G}' + (\hat{G}' - H' \hat{G}'),
\end{align*}
\] (59)

where $\mathbb{G} \equiv [g_{k+1}, ..., g_{k_0}]'$, and the first line uses the fact that $g_t = C f_t$ for $t \geq k_0$. As $\sigma_1(\mathbb{A} + \mathbb{B}) \leq \sigma_1(\mathbb{A}) + \sigma_1(\mathbb{B})$, we have

\[
\sigma_1(Q_k^{1/2} C^\dagger H' \hat{G}') \leq \sigma_1(Q_k^{1/2} C^\dagger H' \hat{G}') + \sigma_1(-Q_k^{1/2} C^\dagger H' (\hat{G}' - H' \hat{G}'))
\]
(60)

by setting $\mathbb{A} = Q_k^{1/2} C^\dagger H' \hat{G}'$ and $\mathbb{B} = -Q_k^{1/2} C^\dagger H' (\hat{G}' - H' \hat{G}')$. Rearranging the inequality in (60) and using (59), we obtain

\[
\begin{align*}
\frac{1}{\sqrt{k_0-k}} \sigma_1(Q_k^{1/2} C^\dagger H' \hat{G}') & \geq \frac{1}{\sqrt{k_0-k}} \sigma_1(Q_k^{1/2} C^\dagger \hat{G}') \|H\| - \frac{1}{\sqrt{k_0-k}} \sigma_1(Q_k^{1/2} C^\dagger H' (\hat{G}' - H' \hat{G}')) \\
\sqrt{\frac{1}{k_0-k} \rho_1(\hat{G} H' C^\dagger Q_k C^\dagger H' \hat{G}')} & \geq \|H\| \sqrt{\frac{1}{k_0-k} \rho_1(GC' Q_k C^\dagger G') - \frac{1}{\sqrt{k_0-k}} \|Q_k^{1/2} C^\dagger H' (\hat{G}' - H' \hat{G}')\|} \\
& \geq \|H\| \sqrt{\frac{\rho_1(GC' Q_k C^\dagger G')}{k_0-k}} - O_p(N^{-1/2})
\end{align*}
\] (61)

where the first line is based on the fact that $H' H = I_r |H|$, the second line uses the inequality that the maximum singular value is bounded by the Frobenius norm, and the third line follows from the derivation below:

\[
\frac{1}{k_0-k} \|Q_k^{1/2} C^\dagger H' (\hat{G}' - H' \hat{G}')\|^2 = \frac{1}{k_0-k} \text{tr}[\hat{G} - \mathbb{G}H] H' C^\dagger Q_k C^\dagger H' (\hat{G}' - H' \hat{G}')]
\leq \rho_1(Q_k) \frac{1}{k_0-k} \text{tr}[C^\dagger H' (\hat{G}' - H' \hat{G}') (\hat{G} - \mathbb{G}H) H' C^\dagger]
= O_p(N^{-1}) \text{ uniformly over } \tau T \leq k < k_0
\]
given the fact that $\max_{r \leq T \leq k} (k_0-k)^{-1} \|\hat{G}' - H' \hat{G}'\| = O_p(N^{-1})$ by Lemma [h]

Now, it suffices to find the lower bound of $\rho_1(GC' Q_k C^\dagger G')$. Using the definition of $Q_k$ and inequality $\rho_1(\mathbb{A} \mathbb{B}) \geq \rho_1(\mathbb{A}) \rho_1(\mathbb{B})$ for $r \times r$ positive semidefinite matrices $\mathbb{A}$ and $\mathbb{B}$, we have

\[
\rho_1(GC' Q_k C^\dagger G') \geq \rho_r \left( \left( \frac{T - k_0}{T-k} \Sigma_{f,2} \right)^\dagger \right) \rho_1 \left( C^\dagger G' Q_k C^\dagger \right) = \frac{T - k_0}{T-k} \left( \frac{\|\Sigma_f\|}{\rho_1(\Sigma_f)} + o_p(1) \right) \rho_1 \left( C^\dagger G' Q_k C^\dagger \right)
\] (62)

where we use the facts that $\Sigma_{f,2} \equiv \frac{1}{T-k_0} \sum_{t=k_0+1}^{T} f_t f_t' \rightarrow_p \Sigma_f$ and $\rho_r(\Sigma_f^\dagger) = \rho_r(\Sigma_f) = \rho_1(\Sigma_f) = \|\Sigma_f\|/\rho_1(\Sigma_f)$.

For $k_0-k \rightarrow \infty$ as $N, T \rightarrow \infty$,

\[
\rho_1 \left( C^\dagger \frac{\hat{G} G'}{k_0-k} C^\dagger \right) = \rho_1 \left( C^\dagger B \Sigma_f B' C^\dagger + o_p(1) \right) > c_1 \text{ w.p.a.1}
\] (63)
for a constant $c_1 > 0$ as $N, T \to \infty$, because $C^\# B \neq 0$ and $\Sigma_f$ is positive definite according to Assumption 11 (i). As $|\Sigma_f|/\rho_1(\Sigma_f) > 0$ in (62), we have $\lim_{k_0 \to \infty} \rho_1 \left( G C^{\#} Q_k C^{\#} \right) \geq c_2$ w.p.a.1 for a constant $c_2 > 0$ as $N, T \to \infty$.

For $k_0 - k$ being bounded,

$$\rho_1 \left( C^{\#} C^{\#} \right) \geq \rho_1 \left( C^{\#} g_k g_k C^{\#} \right) = \rho_1 \left( C^{\#} B f_k f_k B^{\#} C^{\#} \right) = f_k B^{\#} C^{\#} C^{\#} B f_k > c$$  \hspace{1cm} (64)

for a constant $c > 0$ according to Assumption 11 (ii), where $C^{\#} B f_k \neq 0$. Combining (62), (63), and (64), we have $\rho_1 (G C^{\#} Q_k C^{\#} G^*)/(k_0 - k) > c_2$ w.p.a.1 for a constant $c_2 > 0$. Thus, we can obtain the lower bound for the RHS of (61) as

$$\frac{1}{\sqrt{k_0 - k}} \sigma_1 (Q_k^{1/2} C^\# H^\# \hat{G}^*) \geq |H| c_2^2 - O_p(N^{-1/2})$$

$$\frac{1}{k_0 - k} \rho_1 (\hat{G}^* H^\# C^{\#} Q_k C^{\#} H^\# \hat{G}^*) \geq c_3 \text{ w.p.a.1}$$  \hspace{1cm} (65)

as $N, T \to \infty$ for a constant $c_3 > 0$, because $H$ has a nonsingular limit.

Based on (60) and Weyl’s inequality, we have

$$\frac{1}{k_0 - k} \rho_1 \left( \frac{1}{T - k} \hat{G} D_k^{-1} \hat{G}^* \right) \geq \frac{1}{|D_k|(k_0 - k)} \rho_1 (P_1) + \frac{1}{|D_k|(k_0 - k)} \rho_1 (P_2)$$

$$= \frac{1}{|D_k|(k_0 - k)} \rho_1 (P_1) + \frac{1}{|D_k|(T - k)} \rho_1 (\hat{G} H^\# C^{\#} Q_k C^{\#} H^\# \hat{G}^*)$$

$$\geq O_p(T^{-1}) + \frac{c_4 N}{T - k} \text{ w.p.a.1}$$  \hspace{1cm} (66)

for a constant $c_4 > 0$ as $N, T \to \infty$ by (58) and (65), where the last line follows from proposition 11 and the $O_p(T^{-1})$ term is uniform over $\tau T \leq k < k_0$.

**Step 2.** When $r - r_2 \geq 2$ or $r_2 = 0$, the term $P_2$ in (56) is zero. For the term $P_1$ in (56), we have

$$\hat{G} D_k^{\#} \hat{G}^* = [\hat{G} H + (\hat{G} - \hat{G} H)] D_k^{\#} [H' \hat{G}^* + (\hat{G}' - H' \hat{G}')]$$.

Using similar techniques to those in (60) and (61), we have

$$\frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} \hat{G}^*) \geq \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} H' \hat{G}^*) - \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} (\hat{G}' - H' \hat{G}'))$$

$$\geq \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} H' \hat{G}^*) - \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} (\hat{G}' - H' \hat{G}'))$$

$$\geq \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} H' \hat{G}^*) - O_p(N^{-(r-r_2)/2})$$  \hspace{1cm} (67)

where the last line is based on the fact that

$$\frac{1}{k_0 - k} \|D_k^{1/2} (\hat{G}' - H' \hat{G}')\|^2 = \frac{1}{k_0 - k} \text{tr}[(\hat{G} - G H) D_k^{\#} (\hat{G}' - H' \hat{G}')]$$

$$\leq \rho_1(D_k^{\#}) \text{tr} \left[ (\hat{G}' - H' \hat{G}')(\hat{G} - G H) \right]$$

$$= O_p(N^{-(r-r_2)})$$

uniformly over $\tau T \leq k < k_0$ under the condition $N \propto T$, because of Lemma 11 and the fact that

$$\rho_1(D_k^{\#}) = |D_k|/\rho_c(D_k) = O_p(N^{-(r-r_2)}) O_p(N)$$  \hspace{1cm} (68)
by proposition 1.

Now, it suffices to determine the lower bound of $\rho_1(\mathcal{G}HD_{k}^\# H'\mathcal{G}')$. Similar to (67), we have

$$D_k^{1/2}U_1 - H'C\Sigma_{1/2}^1 = O_p(\delta_{NT}^{-1}),$$

(69)

uniformly over $\tau T \leq k < k_0$, where $U_1 = D_k^{1/2}(\Sigma_j^{1/2}C'H)^-$. For $t \leq k_0$, we have $g_t = Bf_t$; thus,

$$H'g_t = H'Bf_t = D_k^{1/2}U_1\varepsilon_t + (H'Bf_t - D_k^{1/2}U_1\varepsilon_t)$$

and

$$H'\mathcal{G}' = D_k^{1/2}U_1\varepsilon' + (H'\mathcal{G}' - D_k^{1/2}U_1\varepsilon'),$$

(70)

where we set $f_t = \Sigma_j^{1/2}\varepsilon_t$ with $E(\varepsilon_t\varepsilon_t') = I_c$ and $E \equiv [\varepsilon_{k+1}, \ldots, \varepsilon_{k_0}]'$.

First, we consider the case in which $k_0 - k$ is bounded. Note that $D_k^\#$ has $r_2$ eigenvalues of order $O_p(N^{-(r_2-2)})$ and $r - r_2$ eigenvalues of order $O_p(N^{-(r-r_2-1)})$ by (65), and let $v_1(r \times r_2)$ and $v_2(r \times (r - r_2))$ denote the corresponding eigenvectors. By proposition 1 for $t \leq k_0$,

$$\varepsilon_t'U_1D_k^{1/2}D_k^{1/2}U_1\varepsilon_t = |D_k|(\varepsilon_t'U_1\varepsilon_t) = O_p(\delta_{NT}^{-1}),$$

(71)

which implies that $D_k^{1/2}U_1\varepsilon_t$ lies in the space spanned by $v_1$. Thus, for $t \leq k_0$,

$$f_tC'H D_k^\#H'Cf_t = \|D_k^{1/2}(H'C\Sigma_{1/2}^1\varepsilon_t - D_k^{1/2}U_1\varepsilon_t) + D_k^{1/2}U_1\varepsilon_t\|^2$$

$$\leq 2\|D_k^{1/2}(H'C\Sigma_{1/2}^1 - D_k^{1/2}U_1)\varepsilon_l\|^2 + O_p(\delta_{NT}^{-1})$$

$$\leq 2\rho_1(D_k^\#)(|H'C\Sigma_{1/2}^1 - D_k^{1/2}U_1|\varepsilon_l\|^2 + O_p(N^{-(r-r_2)}) = O_p(N^{-(r-r_2)}),$$

(72)

where the second line follows from (71) and the last line follows from (65) and (69) under the condition $N \propto T$.

To bound $\rho_1(\mathcal{G}HD_{k}^\# H'\mathcal{G}')$, we consider

$$D_k^{1/2}H'g_{k_0} = D_k^{1/2}[H'Bf_{k_0} - \text{Proj}(H'Bf_{k_0}|H'C\Sigma_{1/2}^1)] + D_k^{1/2}\text{Proj}(H'Bf_{k_0}|H'C\Sigma_{1/2}^1)$$

$$= D_k^{1/2}[H'Bf_{k_0} - \text{Proj}(H'Bf_{k_0}|D_k^{1/2}U_1) + O_p(\delta_{NT}^{-1})] + D_k^{1/2}\text{Proj}(H'Bf_{k_0}|H'C\Sigma_{1/2}^1)$$

$$= D_k^{1/2}[H'Bf_{k_0} - \text{Proj}(H'Bf_{k_0}|D_k^{1/2}U_1) + O_p(N^{-(r-r_2)/2}),$$

(73)

where $\text{Proj}(A|Z)$ denotes the projection of $A$ onto the columns of $Z$, the second line follows from (65), and the $O_p(N^{-(r-r_2)/2})$ term in the third line follows from (65) and the fact that $D_k^{1/2}\text{Proj}(H'Bf_{k_0}|H'C\Sigma_{1/2}^1) = O_p(N^{-(r-r_2)/2})$ by (72), because $\text{Proj}(H'Bf_{k_0}|H'C\Sigma_{1/2}^1)$ is a linear combination of $H'C\Sigma_{1/2}^1$ columns. Under Assumption (ii)(iii), according to which $|Bf_{k_0} - \text{Proj}(Bf_{k_0}|C)| \geq d > 0$, we have $H'Bf_{k_0} - \text{Proj}(H'Bf_{k_0}|H'C\Sigma_{1/2}^1)$ bounded away from zero. This implies that the term $H'Bf_{k_0} - \text{Proj}(H'Bf_{k_0}|D_k^{1/2}U_1)$ in the last line of (73) is also bounded away from zero and lies in the space spanned by $v_2$, because it is, by design, orthogonal to $D_k^{1/2}U_1$, which lies in the space of $v_1$ by (71). As $v_2$ corresponds to the $O_p(N^{-(r-r_2-1)})$ eigenvalues of $D_k^\#$, we have

$$\rho_1(\mathcal{G}HD_{k}^\# H'\mathcal{G}') \geq g_{k_0}HD_{k}^\# H'g_{k_0} \geq \rho_{r-r_2}(D_k^\#)|H'g_{k_0}|^2 = \frac{|D_k|}{\rho_{r_2+1}(D_k)}|H'g_{k_0}|^2 \geq \frac{N}{c_1}\|H'g_{k_0}\|^2|D_k|$$

w.p.a.1 as $N,T \to \infty$ by proposition 1. Thus, for $k_0 - k$ being bounded,

$$\rho_1(\mathcal{G}HD_{k}^\# H'\mathcal{G}') \geq g_{k_0}HD_{k}^\# H'g_{k_0} \geq c_1N \cdot |D_k|$$

(74)
Thus, in combination with (76), we obtain

\[ GHD_k^\# H'G' = [(GH - EU_1D_k^{1/2}) + EU_1D_k^{1/2}]D_k^{1/2}U_1E' + (H'G' - D_k^{1/2}U_1E'). \]

Based on the same techniques as in (70) and (71), the decomposition in (70) implies

\[ \sigma_1(D_k^{1/2} (H'G' - D_k^{1/2}U_1E')) \leq \sigma_1(D_k^{1/2} H'G') + \sigma_1(-D_k^{1/2} D_k^{1/2}U_1E'), \]

so we have

\[
\frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} H'G') \geq \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} (H'G' - D_k^{1/2}U_1E')) - \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} D_k^{1/2}U_1E') \\
\geq \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} (H'G' - D_k^{1/2}U_1E')) - |D_k^{1/2}| \sigma_1(U_1) \sigma_1 \left( \frac{\mathcal{E}'}{\sqrt{k_0 - k}} \right) \\
= \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} (H'G' - D_k^{1/2}U_1E')) - O_p(N^{-(r-r_2)/2})O_p(1) \tag{75}
\]

where the second inequality is based on the fact that \( \sigma_1(A \mathcal{B}) \leq \sigma_1(A) \sigma_1(\mathcal{B}) \) and the last line follows from the facts that \( |D_k| \) is uniformly \( O_p(N^{-(r-r_2)}) \) by proposition \( \mathcal{B} \) \( \sigma_1(U_1) \leq \|U_1\| = O_p(1) \) by the structure of \( U_1 \), and \( \frac{1}{\sqrt{k_0 - k}} \sigma_1(\mathcal{E}') \leq \sqrt{\|\mathcal{E}'\|^2/(k_0 - k)} = O_p(1) \) uniformly over \( \tau T \leq k < k_0 \).

Next, we need to determine the lower bound of \( \rho_1((GH - EU_1D_k^{1/2})D_k^{1/2}(H'G' - D_k^{1/2}U_1E'))/(k_0 - k) \). From (69), we have

\[
H'G' - D_k^{1/2}U_1E' = H'G' - H'C\Sigma_j^{1/2}E' + (H'C\Sigma_j^{1/2}E' - D_k^{1/2}U_1E') \\
= H'G' - H'C\mathcal{F}' + O_p(\delta_{1/2}^2)E' \\
= H'(B - C)\mathcal{F}' + O_p(\delta_{1/2}^2)E' \tag{76}
\]

where \( \mathcal{F}' \equiv [f_{k+1}, ..., f_{k_0}] = \Sigma_j^{1/2}E' \) in the second line. Again, using the inequality in (69), we have

\[
\sigma_1(D_k^{1/2} (H'G' - H'C\Sigma_j^{1/2}E')) \leq \sigma_1(D_k^{1/2} (H'G' - D_k^{1/2}U_1E')) + \sigma_1(-D_k^{1/2} (H'C\Sigma_j^{1/2}E' - D_k^{1/2}U_1E')).
\]

Thus, in combination with (76), we obtain

\[
\frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} (H'G' - D_k^{1/2}U_1E')) \geq \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} (H'G' - H'C\Sigma_j^{1/2}E')) - \frac{1}{\sqrt{k_0 - k}} \sigma_1(D_k^{1/2} (H'C\Sigma_j^{1/2}E' - D_k^{1/2}U_1E')) \\
\geq \frac{1}{k_0 - k} \rho_1(D_k^{1/2} (H'(B - C)\mathcal{F} + O_p(N^{-(r-r_2)/2}) \tag{77}
\]

w.p.a.1. for a constant \( c_1 > 0 \) as \( N, T \to \infty \) under the condition \( N \ll T \).

Second, we consider the case in which \( k_0 - k \to \infty \) as \( N, T \to \infty \). Using (70), we rewrite \( GHD_k^\# H'G' \) as

\[ GHD_k^\# H'G' = (GH - EU_1D_k^{1/2})D_k^{1/2}U_1E' + (H'G' - D_k^{1/2}U_1E'). \]
w.p.a.1 for a constant $c_2 > 0$ as $N, T \to \infty$, where the $O_p(N^{-(r_2-1)/2})$ term in the third line follows from (65) and the condition $N \propto T$, the last line follows from the fact that $|\rho_1(H^T(B-C)\Sigma_f(B-C)')|^{1/2} \geq c_2 > 0$ because $B-C \neq 0$ and $\Sigma_f$ is positive definite by Assumption 1, and the leading term in the last line is $O_p(N^{-(r_2-1)/2})$ by (65). Hence, combining (65) and (77) yields

$$\frac{1}{\sqrt{k_0-k}}\sigma_1(D_k^{1/2}H'G') \geq \rho_1(D_k^{1/2})^{1/2} c_2 \quad \text{and}$$

$$\frac{1}{k_0-k}\rho_1(\hat{G}HD_k^{1/2}H'G') \geq \frac{|D_k|}{\rho_1(D_k)} c_2^2 \geq \frac{c_2^2}{c_0} N|D_k| = c_3 N \cdot |D_k|$$

(78)

w.p.a.1 for a constant $c_3 > 0$ as $N, T \to \infty$.

According to (67), (74), and (78), there exists a constant $c_4 > 0$ such that

$$\frac{1}{\sqrt{k_0-k}}\sigma_1(D_k^{1/2}G'\hat{G}) \geq \sqrt{c_4 N \cdot |D_k|} - O_p(N^{-(r_2-1)/2}),$$

w.p.a.1 as $N, T \to \infty$; thus, we have

$$\frac{1}{k_0-k}\rho_1(\hat{G}D_k^{1/2}G') \geq c_4 N \cdot |D_k| \quad \text{w.p.a.1}$$

as $N, T \to \infty$. Hence,

$$\frac{1}{k_0-k}\rho_1 \left( \frac{1}{T-k} \hat{G}D_k^{-1}G' \right) = \frac{1}{|D_k|} \left( \frac{1}{T-k} \hat{G}D_k^{-1}G' \right) \geq c_4 \frac{N}{T-k} \quad \text{w.p.a.1}$$

(79)

as $N, T \to \infty$. Summarizing the results in (66) and (79), we obtain the lower bound of $\rho_1 \left( \frac{1}{T-k} \hat{G}D_k^{-1}G' \right)$. Thus, steps 1 and 2 are completed.

Finally, using the lower bound of the largest eigenvalue of matrix $\frac{1}{T-k} \hat{G}D_k^{-1}G'$, we can rewrite (55) as

$$|\hat{\Sigma}_2(k)| = |D_k| \cdot |I_{k_0-k} + \frac{1}{T-k} \hat{G}D_k^{-1}G'| \geq |D_k| \left[ 1 + c_4 \left( \frac{k_0-k}{T-k} \right) \cdot N \right] \geq |D_k| \left[ 1 + \frac{c_4 (k_0-k)}{(1-\tau_1)T} \cdot N \right] \quad \text{w.p.a.1}$$

as $N, T \to \infty$.

Comparing $|D_k|$ and $|\hat{\Sigma}_2^0|$, we have

$$|D_k| = \left| \frac{T-k_0}{T-k} \left( \frac{1}{T-k_0} \sum_{t=k_0+1}^T \hat{g}_t \hat{g}_t' \right) \right| - \left| \frac{1}{T-k_0} \sum_{t=k_0+1}^T \hat{g}_t \hat{g}_t' \right| = \left[ \left( 1 - \frac{k_0-k}{T-k} \right)^r - 1 \right] \left| \hat{\Sigma}_2^0 \right|$$

$$\geq \frac{-r(k_0-k)}{T-k} + \frac{r(r-1)}{2} \left( \frac{k_0-k}{T-k} \right)^2 + \ldots \left| \hat{\Sigma}_2^0 \right|$$

and

$$|D_k| = \frac{c_5 (k_0-k)}{T-k} \left| \hat{\Sigma}_2^0 \right|$$

for some positive constant $c_5 > 0$. In addition,

$$|D_k|/|\hat{\Sigma}_2^0| = \left( \frac{T-k_0}{T-k} \right)^r \geq \left( \frac{1-\tau_1}{1-\tau_1} \right)^r.$$

Thus,

$$\left| \hat{\Sigma}_2^0(k) - |D_k| + |D_k] - |\hat{\Sigma}_2^0| \right| \geq \frac{|D_k|}{|\hat{\Sigma}_2^0|} \left( \frac{c_4 (k_0-k)}{(1-\tau_1)T} \cdot N - \frac{c_5 (k_0-k)}{T-k} \right) \quad \text{w.p.a.1}$$

$$\geq \left( \frac{1-\tau_1}{1-\tau_1} \right)^r \left( \frac{c_4 (k_0-k)}{(1-\tau_1)T} \cdot N - \frac{c_5 (k_0-k)}{T-k} \right) \quad \text{w.p.a.1},$$

leading term

as $N, T \to \infty$, which implies the desired result under the condition $N \propto T$. □
Proof of Theorem 3

We first prove the consistency of \( \hat{\tau} \), then \( \hat{k} - k_0 = O_p(1) \), and finally, \( \hat{k} - k_0 = o_p(1) \). Again, it suffices to study the case of \( k < k_0 \).

To prove \( \hat{\tau} - \tau_0 = o_p(1) \), we need to show that for any \( \varepsilon > 0 \) and \( \eta > 0 \), \( P(|\hat{\tau} - \tau_0| > \eta) < \varepsilon \) as \( N, T \to \infty \). For any given \( 0 < \eta \leq \min(\tau_0, 1 - \tau_0) \), define \( D_\eta = \{ k : (\tau_0 - \eta)T \leq k \leq (\tau_0 + \eta)T \} \) and \( D_\eta^c \) as the complement of \( D_\eta \). Similar to the proof for the consistency of \( \hat{\tau} \) when \( B \) and \( C \) are nonsingular, we need to show that \( P(\hat{k} \in D_\eta^c) < \varepsilon \). Recalling (13) and (14), we have

\[
P(\min_{k \in D_\eta^c, k < k_0} U(k) - U(k_0) \leq 0) = P(\min_{k \in D_\eta^c, k < k_0} \frac{U(k) - U(k_0)}{k_0 - k} \leq 0),
\]

which is based on Lemma 7 and the third line is based on the fact that \( |k_0 - k| > \eta T \).

(1) Consider the first term \( \frac{k}{k_0 - k} \log |\hat{\Sigma}_1 \hat{\Sigma}_2^{-1}| \). When \( \Sigma_1 \) is of full rank, it follows that

\[
\left| \min_{k \in D_\eta^c, k < k_0} \frac{k}{k_0 - k} \log |\hat{\Sigma}_1 \hat{\Sigma}_2^{-1}| \right| = o_p(1)
\]

by the argument used in (15) and Lemma 2. When \( \Sigma_1 \) is singular, we can obtain

\[
\left| \min_{k \in D_\eta^c, k < k_0} \frac{k}{k_0 - k} \log |\hat{\Sigma}_1 \hat{\Sigma}_2^{-1}| \right| = \left| \min_{k \in D_\eta^c, k < k_0} \frac{k}{k_0 - k} \log (\frac{|\hat{\Sigma}_1(k)| - |\hat{\Sigma}_1(k_0)|}{|\hat{\Sigma}_1(k_0)|} + 1) \right|
\]

\[
= \left| \min_{k \in D_\eta^c, k < k_0} \frac{k}{k_0 - k} \log (O_p(T^{-1}(k_0 - k)) + 1) \right|
\]

\[
= O_p(1),
\]

where the second line follows from Lemma (7) and the third line is based on the fact that \( |k_0 - k| > \eta T \).

(2) For the second and third terms, let

\[
f(\hat{\Sigma}_1, \hat{\Sigma}_2^0) = \min_{k \in D_\eta^c, k < k_0} \frac{T - k}{k_0 - k} \log |\hat{\Sigma}_2 \hat{\Sigma}_2^{-1}| - \log |\hat{\Sigma}_1 \hat{\Sigma}_2^{-1}|
\]

\[
= \min_{k \in D_\eta^c, k < k_0} \frac{T - k}{k_0 - k} \sum_{i=1}^r \log \rho_i(\hat{\Sigma}_2 \hat{\Sigma}_2^{-1}) - \sum_{i=1}^r \log \rho_i(\hat{\Sigma}_1 \hat{\Sigma}_2^{-1}).
\]

We show that \( f(\hat{\Sigma}_1, \hat{\Sigma}_2^0) \to +\infty \) at the rate \( \log T \).

When \( \Sigma_1 \) is singular and \( \Sigma_2 \) is a positive definite matrix, (10) implies

\[
\hat{\Sigma}_2 = \frac{k_0 - k}{T - k} \Sigma_1 + \frac{T - k}{T - k} \Sigma_2 + o_p(1),
\]

where the \( o_p(1) \) term is uniform over \( k \in D_\eta^c \) for \( k < k_0 \) by Lemma (iv) and (vii). Together with \( \hat{\Sigma}_2^0 \to_p \Sigma_2^{-1} > 0 \), (84) implies that \( \rho_i(\hat{\Sigma}_2 \hat{\Sigma}_2^{-1}) \) is uniformly \( O_p(1) \) and bounded away from zero; thus,

\[
\left| \frac{T - k}{k_0 - k} \sum_{i=1}^r \log \rho_i(\hat{\Sigma}_2 \hat{\Sigma}_2^{-1}) \right| = O_p(1)
\]

uniformly over \( k \in D_\eta^c \) for \( k < k_0 \). In addition, we have \( \rho_i(\hat{\Sigma}_1 \hat{\Sigma}_2^{-1}) = O_p(T^{-1}) \) uniformly over \( k \) for \( i = r_1 + 1, ..., r \) by proposition (1) when \( N \propto T \); thus, \( \log \rho_i(\hat{\Sigma}_1 \hat{\Sigma}_2^{-1}) \to -\infty \) at the rate of \( \log T \) for \( i = r_1 + 1, ..., r \). Therefore, (83) can be
rewritten as

\[ f(\Sigma_1^0, \Sigma_2^0) = -\sum_{i=1}^{r} \log \rho_i(\Sigma_1^0, \Sigma_2^0) + O_p(1). \]

Thus, it suffices to show that for any \( p \) such that \( \rho(\Sigma_1) > 0 \) diverges at the rate \( T \) by proposition 1 for \( i = 1, ..., r - r \); thus, \( \sum_{i=1}^{r} \log \rho_i(\Sigma_1^0) \to +\infty \) at the rate log \( T \). In addition, when \( \rho(\Sigma_1) > 0 \), we have \( \rho(\Sigma_1) \to_p \rho(\Sigma_1) > 0 \); thus, \( r \log \rho(\Sigma_1) = O_p(1) \). When \( \rho(\Sigma_1) = 0 \) (i.e., \( r_1 = 0 \)), we have \( \rho(\Sigma_1) \cdot T \geq c > 0 \) w.p.a.1 for \( c > 0 \) by proposition 1 thus, \( -r \log \rho(\Sigma_1) \to +\infty \) at the rate log \( T \). For \( \rho(\Sigma_2) \), rearranging the terms in \( \Sigma_2 \) yields

\[ \Sigma_2 = \frac{(k_0 - k)T}{k_0(k_0 - k)} \Sigma_1 + \frac{T - k_0}{k_0 - k} \Sigma_2 + \frac{k T - k_0}{k_0 - k} \rho(1) \]

where \( \tau_0 \Sigma_1 + (1 - \tau_0) \Sigma_2 \) is a positive definite matrix under Assumption 11 (i). Thus, \( \rho(\Sigma_2) \) is \( O_p(1) \) and bounded away from zero w.p.a.1, and

\[ \left| \frac{T - k}{k_0 - k} \right| \log \rho(\Sigma_2) = O_p(1) \]

uniformly over \( k \in D_\eta \) for \( k < k_0 \). Combining the above results, we establish the following result: \( f(\Sigma_1^0, \Sigma_2^0) \to +\infty \) at the rate log \( T \). Together with \( \Sigma_1 \) and \( \Sigma_2 \), we have

\[ \min_{k \in D_\eta, k < k_0} \frac{k}{k_0 - k} \log |\Sigma_1 \Sigma_2^0| + \frac{T - k_0}{k_0 - k} \log |\Sigma_1 \Sigma_2^0| - \log |\Sigma_1 \Sigma_2^0| > 0, \]

w.p.a.1; thus, \( P(\min_{k \in D_\eta, k < k_0} U(k) - U(k_0) \leq 0) \to 0 \) for any \( \eta > 0 \), and hence, \( \tilde{r} \to_p \tau \).

Next, we show that \( k_0 - k = O_p(1) \).

Similar to the proof of Theorem 1 for given \( \eta \) and \( M \), define \( D_{n,M} = \{ \hat{k} : (\tau_0 - \eta)T \leq k \leq (\tau_0 + \eta)T, |k_0 - k| > M \} \), such that \( P(|\hat{k} - k_0| > M) = P(\hat{k} \in D_\eta^0) + P(\hat{k} \in D_{n,M}) \). Hence, it suffices to show that for any \( \varepsilon > 0 \) and \( \eta > 0 \), there exists an \( M > 0 \) such that \( P(\hat{k} \in D_{n,M}) < \varepsilon \) as \( (N, T) \to \infty \). Similar to \( \Sigma_1 \) and \( \Sigma_2 \), it suffices to show that for any given \( \varepsilon > 0 \) and \( \eta > 0 \), there exists an \( M > 0 \) such that

\[ P(\min_{k \in D_{n,M}, k < k_0} \frac{k}{k_0 - k} \log |\Sigma_1 \Sigma_2^0| + \frac{T - k_0}{k_0 - k} \log |\Sigma_1 \Sigma_2^0| - \log |\Sigma_1 \Sigma_2^0| \leq 0) \]

where \( \Sigma_1 \to_p \Sigma_1 \) is of full rank, we have

\[ P(\min_{k \in D_{n,M}, k < k_0} \frac{k}{k_0 - k} \log |\Sigma_1 \Sigma_2^0| \leq c_\Delta) \geq P(\max_{k \in D_{n,M}, k < k_0} \frac{k}{k_0 - k} \log |\Sigma_1 \Sigma_2^0| \leq c_\Delta) \geq 1 - \frac{C}{Mc_\Delta} \to 1 \]
for a constant $C > 0$ under the same arguments as those in \cite{20} and \cite{21}.

When $\Sigma_1$ is singular,
\[
\min_{k \in D_{q,M}, k < k_0} \frac{k}{k_0 - k} \log |\hat{\Sigma}_1^{0,1} - \Sigma_1^{0,1}| = \min_{k \in D_{q,M}, k < k_0} \frac{k}{k_0 - k} \log \left( \frac{|\hat{\Sigma}_1^{0}(k) - \Sigma_1^{0}(k)|}{|\Sigma_1^{0}(k)|} + 1 \right)
\]
\[
= \min_{k \in D_{q,M}, k < k_0} \frac{k}{k_0 - k} \log (O_p(T^{-1}(k_0 - k))) + 1
\]
\[
= O_p(1),
\] (86)

where the second equation holds because of Lemma \cite{22} and the last equality holds because \[\frac{k}{k_0 - k} \log (O_p(T^{-1}(k_0 - k))) = O_p(1)\] whether $k_0 - k$ is bounded or diverging.

For the second and third terms, we consider several cases.

(i). When $\Sigma_1$ is singular and $\Sigma_2$ is positive definite, we have
\[
\min_{k \in D_{q,M}, k < k_0} \frac{T - k}{k_0 - k} \log |\hat{\Sigma}_2^{0,1} - \Sigma_2^{0,1}| = \min_{k \in D_{q,M}, k < k_0} \frac{T - k}{k_0 - k} \log \left( I + \frac{k - k_0}{T - k} + \frac{k_0 - k}{T - k} \hat{\Sigma}_1^{0,1} - \Sigma_1^{0,1} + \frac{k_0 - k}{T - k} \left( 1 + \sum_{t = k+1}^{k_0} \hat{g}_t \hat{g}_t' - \hat{\Sigma}_1^{0} \right) \right) - \log |\hat{\Sigma}_1^{0} - \Sigma_1^{0}| + O_p(1)
\]
\[
= O_p(1) - \log |\hat{\Sigma}_1^{0}| + \log |\Sigma_2^{0}| \to \infty \text{ at the rate } \log T
\] (87)

where the second line is based on the fact that $\hat{\Sigma}_2 = \frac{1}{T - k} \sum_{t = k+1}^{k_0} \hat{g}_t \hat{g}_t' + \frac{T - k}{k_0 - k} \hat{\Sigma}_2^{0}$, the third line follows from the fact that $(k_0 - k)/T \to 0$ through the consistency of $\hat{\tau}$ and the boundedness of $\frac{T}{k_0 - k} \sum_{t = k+1}^{k_0} \hat{g}_t \hat{g}_t' - \hat{\Sigma}_1^{0}$ by \cite{22} and \cite{21}, and the divergence rate in the last line follows from the fact that $- \log |\hat{\Sigma}_1^{0}| \geq \log (c_1 T)$ for some $c_1 > 0$ by proposition \cite{1} under the assumption $N \sim T$ and the fact that $\log |\Sigma_2^{0}| = O_p(1)$ because $\hat{\Sigma}_2^{0} \to_p \Sigma_2$ is positive definite.

(ii). When $\Sigma_2$ is singular and $\Sigma_1$ is either singular or positive definite, we have
\[
\min_{k \in D_{q,M}, k < k_0} \frac{T - k}{k_0 - k} \log |\hat{\Sigma}_2^{0,1} - \Sigma_2^{0,1}| = \min_{k \in D_{q,M}, k < k_0} \frac{T - k}{k_0 - k} \log \left( \frac{|\hat{\Sigma}_2^{0} - \Sigma_2^{0}|}{|\Sigma_2^{0}|} + 1 \right) + \log |\Sigma_2^{0}| - \log |\Sigma_1^{0}|
\]
\[
\geq \min_{k \in D_{q,M}, k < k_0} \frac{T - k}{k_0 - k} \log \left( c(k_0 - k) + 1 \right) + \log |\Sigma_2^{0}| - \log |\Sigma_1^{0}| + O_p(\log T)
\]
\[
\to \infty,
\] (88)

where the inequality in the third line holds because of Lemma \cite{23} the $O_p(\log T)$ term in the third line follows from proposition \cite{1} and the divergence in the last line evidently holds when $k_0 - k \to \infty$ and $(k_0 - k)/T \to 0$, because $\frac{T - k}{k_0 - k} \log (k_0 - k) > \frac{T - k}{k_0 - k} \log \left( \frac{T}{k_0 - k} \right) \to \infty$.

Thus, we have shown that the second and third terms dominate the first term, and hence, \cite{24} holds.

To indicate the consistency of $\hat{k}$, we will show that for any $k < k_0$ and $k_0 - k \leq M$, the objective function $V(k) = U(k) - U(k_0)$ diverges to infinity as $N,T \to \infty$; thus, the minimum $U(k)$ cannot be achieved at a point other than $k_0$. For
the given $M$, define $D_M = \{k : |k_0 - k| \leq M\}$, then
\[
\min_{k \in D_M, k < k_0} \frac{U(k) - U(k_0)}{k_0 - k} = \min_{k \in D_M, k < k_0} \frac{k}{k_0 - k} \log |\hat{\Sigma}_1^{\hat{\Sigma}_1 T_k^T}| + \frac{T - k}{k_0 - k} \log |\hat{\Sigma}_2^{\hat{\Sigma}_2 T_k^T}| - \log |\hat{\Sigma}_1^{\hat{\Sigma}_1 T_k^T}|. \tag{89}
\]

When $\Sigma_1$ is of full rank, the first term in (89) is
\[
\min_{k \in D_M, k < k_0} \frac{k}{k_0 - k} \log |\hat{\Sigma}_1^{\hat{\Sigma}_1 T_k^T}| = \min_{k \in D_M, k < k_0} \frac{k}{k_0 - k} \log |(\hat{\Sigma}_1 - \hat{\Sigma}_1^0)\hat{\Sigma}_1^0 + I|
\]
\[
= \min_{k \in D_M, k < k_0} \frac{k}{k_0 - k} \log \left( \frac{k_0 - k}{k_0} \sum_{t=1}^k (\xi_t + \zeta_t) - \frac{1}{k_0} \sum_{t=k+1}^{k_0} (\xi_t + \zeta_t) \right) |\hat{\Sigma}_1^0 + I|
\]
\[
= \min_{k \in D_M, k < k_0} tr \left( \frac{1}{k_0} \sum_{t=1}^k (\xi_t + \zeta_t) - \frac{k}{k_0(k_0 - k)} \sum_{t=k+1}^{k_0} (\xi_t + \zeta_t) \right) |\hat{\Sigma}_1^0 + O_p(1)
\]
\[
= O_p(1).
\]

Similar to (89), when $\Sigma_1$ is singular, the first term in (89) is
\[
\min_{k \in D_M, k < k_0} \frac{k}{k_0 - k} \log |\hat{\Sigma}_1^{\hat{\Sigma}_1 T_k^T}| = \min_{k \in D_M, k < k_0} \frac{k}{k_0 - k} \log |(\hat{\Sigma}_1^0 - \hat{\Sigma}_1^0)\hat{\Sigma}_1^0 + I| + 1 = O_p(1).
\]

The second and third terms are discussed below.

(i). When $\Sigma_1$ is a singular matrix and $\Sigma_2$ is a positive matrix,
\[
\frac{T - k}{k_0 - k} \log |\hat{\Sigma}_2^{\hat{\Sigma}_2 T_k^T}| - \log |\hat{\Sigma}_1^{\hat{\Sigma}_1 T_k^T}| = O_p(1) + \log(T) \to \infty,
\]
where $\frac{T - k}{k_0 - k} \log |\hat{\Sigma}_2^{\hat{\Sigma}_2 T_k^T}| = O_p(1)$ is similar to (87).

(ii). When $\Sigma_2$ is singular and $\Sigma_1$ is either singular or positive definite, similar to (89), we have
\[
\min_{k \in D_M, k < k_0} \frac{T - k}{k_0 - k} \log |\hat{\Sigma}_2^{\hat{\Sigma}_2 T_k^T}| - \log |\hat{\Sigma}_1^{\hat{\Sigma}_1 T_k^T}|
\]
\[
= \min_{k \in D_M, k < k_0} \frac{T - k}{k_0 - k} \log |\hat{\Sigma}_2^{\hat{\Sigma}_2 T_k^T}| - \log |\hat{\Sigma}_2^{\hat{\Sigma}_2 T_k^T}| + 1 + \log |\hat{\Sigma}_2^{\hat{\Sigma}_2 T_k^T} - \log |\hat{\Sigma}_2^{\hat{\Sigma}_2 T_k^T}|
\]
\[
\geq \min_{k \in D_M, k < k_0} \frac{T - k}{k_0 - k} \log(c(k_0 - k) + 1) + \frac{\log |\hat{\Sigma}_2^{\hat{\Sigma}_2 T_k^T}|}{O_p(\log(T))}
\]
\[
\to \infty,
\]
where the inequality in the third line holds because of Lemma 8 the $O_p(\log(T))$ term in the third line follows from proposition 11 and the divergence in the last line evidently holds when $k_0 - k$ is bounded, because $c(k_0 - k) + 1 > 1$ and by the same argument in (89).

In summary, we can determine $U(k) \to \infty$ when $k < k_0$ and $k_0 - k < M$ as $N, T \to \infty$. Thus, we prove the consistency of $\hat{k}$. □

References

Ahn, S. C. and Horenstein, A. R., (2013). Eigenvalue ratio test for the number of factors. *Econometrica* 81(3), pp. 1203–1227.
Ahn, S. C., Lee, Y. H. and Schmidt, P., (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics 174*, pp. 1–14.

Amemiya, T., (1971). The estimation of the variances in a variance-components model. *International Economic Review 12*, pp. 1–13.

Bai, J., (1997). Estimation Of A Change Point In Multiple Regression Models. *The Review of Economics and Statistics 4*, pp. 551–563.

Bai, J., Perron, P., (1998). Estimating and testing linear models with multiple structural change. *Econometrica 64*, pp. 47–78.

Bai, J., (2000). Vector autoregressive models with structural changes in regression coefficients and variance-covariance matrices. *Annals of Economics and Finance 1*, pp. 301–306.

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

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

Bai, J., (2010). Common breaks in means and variances for panel data, Journal of Econometrics. *Journal of Econometrics 157*, pp. 78–92.

Bai, J. and Han, X., (2016). Structural changes in high dimensional factor models. *Front. Econ. China 11*, pp. 9–39.

Bai, J., Han, X., Shi, Y., (2020). Estimation and inference of change points in high-dimensional factor models. *Journal of Econometrics 219*, pp. 66–100.

Baltagi, B.H., Kao, C., Wang, F., (2017). Identification and estimation of a large factor model with structural instability. *Journal of Econometrics 197*, pp. 87–100.

Baltagi, B.H., Kao, C., Wang, F., (2020). Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.04.005.

Barigozzi, M.,Cho, H.andFryzlewicz, P., (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics 206*, pp. 187–225.

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

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

Chen, L., (2015). Estimating the common break date in large factor models. *Economics Letters 131*, pp. 70–74.
Chen, L., Dolado, J. and Gonzalo, J., (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics* 180, pp. 30–48.

Cheng, X., Liao, Z. and Schorfheide, F., (2016). Shrinkage estimation of high-Dimensional factor models with structural instabilities. *Review of Economic Studies* 83, pp. 1511–1543.

Fa, J., Guo, J. and Zheng, S. (2020). Estimating number of factors by adjusted eigenvaluethresholding. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.2020.1825448.

Han, X. and Inoue, A., (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory* 31, pp. 1117–1152.

Kim, D., (2011). Estimating a common deterministic time trend break in large panels with cross sectional dependence. *Journal of Econometrics* 164, pp. 310–330.

Lange, K. (2010). Numerical Analysis for Statisticians, *New York: Springer Verlag*.

Ma, S. and Su, L., (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics* 207(1), pp. 1–29.

McAlinn, K., Rockova, V. and Saha, E., (2018). Dynamic Sparse Factor Analysis. https://arxiv.org/abs/1812.04187.

Onatski, A., (2010). Determining the number of factors from empirical distribution of eigenvalues. *The review of Economics and Statistics* 92(4), pp. 1004–1016.

Qu, Z. and Perron, P., (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica* 75, pp. 459–502.

Stock, J. H. and Watson, M. W., (2008). Forecasting in dynamic factor models subject to structural instability. UK: Oxford University Press.

Stock, J. H. and Watson, M. W., (2012). Disentangling the Channels of the 2007-09 Recession. Brookings Papers on Economic Activity, pp. 81–156.