Change point estimation in large heterogeneous panels

Badi H. Baltagi* Chihwa Kao† Fa Wang‡

March 24, 2015

Abstract

This paper considers least squares estimation of a common break in heterogeneous large N and large T panel data models. Consistency and limiting distribution of the estimated common change point are established under general conditions. A new Hajek-Renyi inequality is introduced to solve the fundamental issue that for random variables $X_{iT} = O_p(1)$ (or $o_p(1)$) as $T \to \infty$, $\frac{1}{N}\sum_{i=1}^{N} X_{iT}$ is not necessarily $O_p(1)$ (or $o_p(1)$ correspondingly) as N and T go to infinity jointly. Both weak and strong cross-sectional dependence are considered. In the former case the least squares estimator is consistent as the number of subjects tends to infinity while in the latter case a two step estimator is proposed and consistency can be recovered once estimated factors are used to control the cross-sectional dependence. The limiting distribution is derived allowing the error process to be serially dependent and heteroskedastic of unknown form, and inference can be made based on the simulated distribution.

Keywords: Common break, Heterogeneous panel, Large panel, Cross-sectional dependence

JEL Classification: C23, C33

*Badi H. Baltagi: Department of Economics, Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020; tel: 315-443-1630; fax: 315-443-1081; e-mail: bbaltagi@maxwell.syr.edu.

†Chihwa Kao: Department of Economics, Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020; tel: 315-443-3233; fax: 315-443-1081; e-mail: cdkao@maxwell.syr.edu.

‡Fa Wang: Department of Economics, 106 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020; e-mail: fwang09@syr.edu.

1 Introduction

Recently, the econometrics literature has witnessed a wave of development in large panel data models (large T), mainly due to its capability of handling cross-sectional dependence. A peak of this wave is Pesaran (2006) and Bai (2009), in which by imposing a multifactor error structure, not only cross-sectional dependence of the errors but also potential correlation between the regressors and the unobservable effects are controlled for. In parallel, spatial econometrics has been extended to panel data settings, see for example Yu, De Jong and Lee (2008) and Lee and Yu (2010a, 2010b). Large panel also enables testing cross-sectional dependence, see Ng (2006), Pesaran (2004, 2012), Pesaran, Ullah and Yamagata (2008) and Baltagi, Feng and Kao (2011, 2012), to mention a few. However, for such panels with a long time span, there is a substantial risk that the underlying data generating process has experienced structural breaks at some unknown time due to various factors. Examples include important economic events such as the European debt crisis, or political events such as the end of the cold war, or gradual but fundamental changes in economic structure due to technological progress, or policy change such as the end of China's one-child policy, to mention a few. If we ignore the parameter changes, standard estimators will be inconsistent and statistical inference will be misleading. Instead, if we explicitly take them into account, the result will be useful for analyzing and evaluating the effect of a policy change, for uncovering the underlying factors that lead to structural change, and for determining whether the response of economic variables are immediate or gradual. This paper therefore studies the parameter change problem in large panel data model with unknown change point.

Change point estimation in linear regression model with single change is analyzed in Bai (1997). Bai and Perron (1998) extends Bai (1997) to the case with multiple changes and also proposes tests for the presence of structural change and the number of changes. Similar works are done in Qu and Perron (2007) for a system of equations and in Bai, Lumsdaine and Stock (1998) for multivariate time series. For other works on structural changes in finite dimensional setup, see the comprehensive survey Perron (2006). Bai et al. (1998) also finds that the number of series is positively related to the accuracy of the change point estimator. To formally analyze this phenomenon, Bai (2010) studies the asymptotic properties of the change point estimator in a panel mean shift setup allowing the number of series N go to infinity jointly with the sample size T. Based on Bai (2010), Baltagi, Kao and Liu (2014) studies the change point estimation in a homogeneous panel setup. Kim (2011) generalizes Bai (2010) to the case with either mean shift or time trend break or both. Kim (2011) also shows that both cross-sectional and serial dependence of the errors deteriorate the asymptotic behavior of the change point estimator and when the errors have a common factor structure, it reduces to the univariate case. To recover the consistency, Kim (2014) moves one step further from Kim (2011) by estimating the change point jointly with the factors and factor loadings.

This paper considers least squares estimation of a common change point in a large heterogeneous panel data model, allowing the cross-sectional dependence to be either weak or strong. The heterogeneous framework is general enough to include most popular panel data models as special cases, so that the results derived here could be applied to these cases with or without minor adjustment. We first focus on some fundamental difficulties in extending Bai (1997, 2010) to the panel regression setup. The key problem is for random variables $X_{iT} = O_p(1)$ (or $o_p(1)$) as $T \to \infty$, $\frac{1}{N}\sum_{i=1}^{N} X_{iT}$ is not necessarily $O_p(1)$ (or $o_p(1)$ correspondingly) as N and T go to infinity jointly. A simple counterexample is that X_{iT} is identically distributed over T, independent over i, mean zero and variance i^2. This problem is partially solved in Bai (2010) and Kim (2011) by utilizing the specificity of the regressors. In the mean shift setup, $x_{it} = 1$ for all i and t and in the time trend setup, $x_{it} = t$ for all i. However, in the general heterogeneous panel regression setup, it becomes especially troublesome and unavoidable. We solve this problem by introducing new technique, a general Hajek-Renyi inequality proposed recently in Fazekas and Klesov (2001). An example is given to illustrate how to calculate the order of the expectation of sup-type terms, which in fact is intrinsically related to the uniform law of large number. In view of its power, we believe this new tool will also be useful in other places.

We then establish the consistency of the estimated common change point under various sets of conditions on the change magnitude and N-T ratio, allowing N and T go to infinity jointly. As in Kim (2011), we consider both weak and strong cross-sectional dependence of the errors. In the former case, the change point is consistent as the number of series tends to infinity while in the latter case, we propose a two step estimator and show that consistency can be recovered once estimated factors are used to control the cross-sectional dependence. It is also worth noting that because of the powerful tool, our assumptions on the data generating process is fairly general. Rather than assuming specific DGP, e.g., linear process, we only require Doob's maximal inequality to be applicable plus some uniformly bounded moments conditions.

The limiting distribution is derived under the same asymptotic framework as Bai (2010), i.e., shrinking break in the N dimension, but allowing the errors to be cross-sectionally weakly dependent and serially dependent and heteroskedastic of unknown form. The limiting distribution in Bai

(2010) is derived assuming the errors are cross-sectionally and serially independent, thus our result is more general and empirically relevant than Bai (2010). This step is nontrivial, see the Appendix for details. Our proof is rigorous and self-contained. Also, our result does not require the DGP to be stationary even within each regime. This feature is not shared by the change point estimator in finite dimensional setup. Based on our result, further parametric assumption can be imposed on the DGP to consistently estimate the parameters in the limiting distribution, and then the distribution can be simulated and inference can be made based on this simulated distribution. In addition, a problem worth clarification is when the change point estimator $\hat{k}$ is consistent, the convergence rate of $\hat{k}$ will be arbitrary, and it's meaningless to derive the limiting distribution of $\hat{k} - k_0$ by multiplying $\hat{k} - k_0$ by some magnifying speed as we normally do, see Remark 2 for more details.

Finally, this paper is parallel with Baltagi, Feng and Kao (2014) which also studies the parameter change problem in large heterogeneous panels. While Baltagi et al. (2014) focuses on the asymptotic properties of the estimated regression coefficients and only proves consistency of the change point estimator, this paper clarifies and studies some fundamental issues in the joint limit asymptotics of change point estimation and the proof of consistency in Baltagi et al. (2014) is based on solving these issues. Furthermore, this paper derives the limiting distribution of the change point estimator, so that inference of the change point can be made. Also, to control cross-sectional dependence, while Baltagi et al. (2014) uses cross-sectional averages of the dependent variable and the regressors as Pesaran (2006), this paper uses estimated factors as Bai (2009).

The rest of the paper is organized as follows. Section 2 introduces the model setup and notations. Section 3 considers least squares estimation of the change point and related fundamental issues. Section 4 studies the asymptotic properties of the least squares estimator when cross-sectional dependence is weak. Section 5 considers estimation of the change point when cross-sectional dependence is strong. Section 6 reports simulation results, while Section 7 concludes. The proofs are given in the Appendix.

2 Model and notations

Consider the following panel data model with a common structural break at k_0:

$$y_{it} = \begin{cases} x_{it}\beta_i + e_{it}, \ \ for \ i = 1, ..., N \ and \ t = 1, ..., k_0, \\ x_{it}\beta_i + z_{it}\delta_i + e_{it}, \ \ for \ i = 1, ..., N \ and \ t = k_0 + 1, ..., T, \end{cases} \tag{1}$$

where y_{it} is the dependent variable, x_{it} is p dimensional vector of regressors, β_i is p dimensional vector of unknown coefficients, z_{it} is q dimensional vector of regressors whose coefficients experienced a structural change, δ_i is q dimensional vector of unknown break magnitude, $z_{it} = R'x_{it}$ and $R = (0_{q\times(p-q)}, I_{q\times q})'$ so that $p > q$ and $p = q$ correspond to partial change and pure change respectively, e_{it} is the error term allowed to have weak cross-sectional and serial dependence as well as heteroskedasticity, both N and T are large. In case cross-sectional dependence is strong, a common factor structure is imposed and the model becomes:

$$y_{it} = \begin{cases} x_{it}\beta_i + F_t^0\lambda_i + e_{it}, \; for \; i = 1, ..., N \; and \; t = 1, ..., k_0, \\ x_{it}\beta_i + z_{it}\delta_i + F_t^0\lambda_i + e_{it}, \; for \; i = 1, ..., N \; and \; t = k_0 + 1, ..., T, \end{cases} \tag{2}$$

where F_t^0 is s dimensional vector of unobservable common factors, λ_i is s dimensional vector of unobservable factor loadings and the other parts maintain the same. In matrix form, the model can be written as

$$Y_i = X_i\beta_i + Z_{0i}\delta_i + e_i, \; for \; i = 1, ..., N, \tag{3}$$

in case the cross-sectional dependence is weak and

$$Y_i = X_i\beta_i + Z_{0i}\delta_i + F^0\lambda_i + e_i, \; for \; i = 1, ..., N, \tag{4}$$

in case the cross-sectional dependence is strong, where $Z_{0i} = (0_{q\times k_0}, z_{i,k_0+1}, ..., z_{i,T})'$. Also, for any possible change point k, define $Z_{1i} = (z_{i,1}, ..., z_{i,k}, 0_{q\times(T-k)})'$, $Z_{2i} = (0_{q\times k}, z_{i,k+1}, ..., z_{i,T})'$ and $Z_{\Delta i} = (Z_{2i} - Z_{0i})sgn(k_0 - k)$, it follows $Z_{0i} = X_{0i}R$, $Z_{1i} = X_{1i}R$, $Z_{2i} = X_{2i}R$ and $Z_{\Delta i} = X_{\Delta i}R$ once X_{0i}, X_{1i}, X_{2i} and $X_{\Delta i}$ are defined similarly. To study the asymptotic behavior of the change point estimator, the whole set of possible change point, $[1, T]$, is divided into several different regions. Define

$$\begin{aligned} K &= \{k : |k - k_0| \leq T\eta\}, \\ K^c &= \{k : |k - k_0| > T\eta, 1 \leq k \leq T\}, \\ K(k_0) &= \{k : k \neq k_0, |k - k_0| < T\eta)\}, \end{aligned}$$

for some $\eta \in (0, \min\{\tau_0, 1 - \tau_0\})$, where $\tau_0 = k_0/T$ is the change fraction, and for some $C > 0$,

$$K(C) = \{k : |k - k_0| > C\} \cap K.$$

Throughout the paper, $\|A\| = (trAA')^{\frac{1}{2}}$ denotes the Frobenius norm, $\|A\|_{op}$ denotes the operator norm, $\rho_{\min}(A)$ and $\rho_{\max}(A)$ denote the minimum and maximum eigenvalue of A, $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution, c represents a typical constant, $(N, T) \to \infty$ denotes N and T going to infinity jointly.

3 Least squares estimation of the change point

For each possible change point k, the sum of squared residuals is:

$$SSR(k) = \sum_{i=1}^{N} SSR_i(k) = \sum_{i=1}^{N} Y_i' M_{X_i,Z_{2i}} Y_i, \tag{5}$$

where $M_{X_i,Z_{2i}} = I - P_{X_i,Z_{2i}}$ and $P_{X_i,Z_{2i}}$ is the projection matrix of (X_i, Z_{2i}). The change point estimator is obtained by minimizing the sum of squared residuals:

$$\hat{k} = \arg\min SSR(k).$$

From the identity $Y_i' M_{X_i,Z_{2i}} Y_i = Y_i' M_{X_i} Y_i - \hat{\delta}_i'(k)(Z_{2i}' M_{X_i} Z_{2i})\hat{\delta}_i(k)$, where $(\hat{\beta}_i'(k), \hat{\delta}_i'(k))'$ is the least squares estimator of $(\beta_i', \delta_i')'$ by regressing Y_i on X_i and Z_{2i}, we have

$$SSR(k) = \sum_{i=1}^{N} Y_i' M_{X_i} Y_i - \sum_{i=1}^{N} \hat{\delta}_i'(k)(Z_{2i}' M_{X_i} Z_{2i})\hat{\delta}_i(k). \tag{6}$$

For simplicity, M_{X_i} is replaced by M_i henceforth. Define $V_i(k) = \hat{\delta}_i'(k)(Z_{2i}' M_i Z_{2i})\hat{\delta}_i(k)$, then $SSR(k) = \sum_{i=1}^{N} Y_i' M_i Y_i - \sum_{i=1}^{N} V_i(k)$ and $SSR(k) - SSR(k_0) = \sum_{i=1}^{N}[V_i(k_0) - V_i(k)]$, it follows that

$$\hat{k} = \arg\min SSR(k) - SSR(k_0) = \arg\max \sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \tag{7}$$

We consider the asymptotic behavior of $\hat{k}$ under different sets of assumptions. Define $\hat{\tau} = \hat{k}/T$ as the estimated break fraction. To show $\hat{\tau} - \tau_0 = o_p(1)$ as $(N,T) \to \infty$, we need to show $P(\hat{k} \in K^c) < \epsilon$ as $(N,T) \to \infty$, and to show $\hat{k} - k_0 = O_p(1)$, we need to show $P(\hat{k} \in K(C)) < \epsilon$ as $(N,T) \to \infty$ additionally, or $P(\hat{k} \in K(k_0)) < \epsilon$ as $(N,T) \to \infty$ additionally if we want to show $\hat{k}$ is consistent for k_0. Let O represent certain possible region of change point, e.g., K^c. By definition of $\hat{k}$, $\sum_{i=1}^{N}[V_i(\hat{k}) - V_i(k_0)] \geq 0$, hence if $\hat{k} \in O$, then $\sup_{k \in O} \sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0$. This implies $P(\hat{k} \in O) \leq P(\sup_{k \in O} \sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0)$, hence to show the former is asymptotically negligible, it suffices to show the latter. In the appendix, we show that the set $\{\omega : \sup_{k \in O} \sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0\}$ is exactly the same as the set $\{\omega : \sup_{k \in O} \frac{1}{|k-k_0|} \sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0\}$, hence it suffices to show $P(\sup_{k \in O} \frac{1}{|k-k_0|} \sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0) < \epsilon$ as $(N,T) \to \infty$.

Remark 1 *The above argument embodies the essence of least squares estimation and appears explicitly, or implicitly, or intrinsically, or partially in previous change point studies. In fact, the proof of the consistency of β in Bai (2009) is also based on this argument. The difference is here the supremum is taken with respect to k while in Bai (2009) the supremum is taken with respect to*

$F'F/T = I$. This argument also can be further generalized and polished to handle other problems featured by the presence of infinity number of nuisance parameters, by replacing the sum of squared residuals with other criterion function and taking supremum over their corresponding parameter subspaces. Here we formalize this argument so that it can be easily modified to fit other problems.

Next, plug in

$$\begin{aligned} \hat{\delta}_i(k) &= (Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iY_i) = (Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_i(X_i\beta_i + Z_{0i}\delta_i + e_i)) \\ &= (Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})\delta_i + (Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i), \end{aligned} \tag{8}$$

we have

$$\begin{aligned} \hat{\delta}_i'(k)(Z_{2i}'M_iZ_{2i})\hat{\delta}_i(k) &= \delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})\delta_i \\ &\quad +2\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) \\ &\quad +(e_i'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i). \end{aligned} \tag{9}$$

$$\begin{aligned} &\sum\nolimits_{i=1}^N V_i(k) - V_i(k_0) \\ = &\sum\nolimits_{i=1}^N [\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})\delta_i - \delta_i'Z_{0i}'M_iZ_{0i}\delta_i] \\ &+\sum\nolimits_{i=1}^N [2\delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) - 2\delta_i'Z_{0i}'M_ie_i] \\ &+\sum\nolimits_{i=1}^N [(e_i'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) - (e_i'M_iZ_{0i})(Z_{0i}'M_iZ_{0i})^{-1}(Z_{0i}'M_ie_i)]. \end{aligned} \tag{10}$$

Define $G_i(k)$ as the first term divided by $-|k_0 - k|$ for $k \neq k_0$ and $H_i(k)$ as the last two terms within the summation, then

$$\frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N [V_i(k) - V_i(k_0)] = -\sum\nolimits_{i=1}^N G_i(k) + \frac{1}{|k_0-k|}\sum\nolimits_{i=1}^N H_i(k). \tag{11}$$

Thus $\sup_{k\in O} \frac{1}{|k-k_0|}\sum_{i=1}^N [V_i(k) - V_i(k_0)] \geq 0$ implies $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| \geq \inf_{k\in O}\sum_{i=1}^N G_i(k)$, and it suffices to show $P(\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| \geq \inf_{k\in O}\sum_{i=1}^N G_i(k)) < \epsilon$ as $(N,T) \to \infty$.

It is worth noting the technical difficulty here. We need to show that the left hand side will be dominated by the right hand side asymptotically. Write out the left hand side,

$$\begin{aligned} &\frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N H_i(k) \\ = \; &2\frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N \delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) - 2\frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N \delta_i'Z_{0i}'M_ie_i \\ &+\frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N e_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_ie_i - \frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N e_i'M_iZ_{0i}(Z_{0i}'M_iZ_{0i})^{-1}Z_{0i}'M_ie_i \end{aligned} \tag{12}$$

and consider the second term as a representative example. To calculate the stochastic order of $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^{N}\delta_i' Z_{0i}' M_i e_i\right|$, if $N=1$, we are back to Bai (1997) and Hajek-Renyi inequality (Hajek and Renyi (1955)) is applicable. However, if N and T go to infinity jointly, Hajek-Renyi inequality is no longer directly applicable. Noting that $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^{N}\delta_i' Z_{0i}' M_i e_i\right| \leq \sum_{i=1}^{N}\sup_{k\in O}\left|\frac{1}{|k-k_0|}\delta_i' Z_{0i}' M_i e_i\right|$, we may tend to conclude $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^{N}\delta_i' Z_{0i}' M_i e_i\right| = O_p(NB_{NT})$, if $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\delta_i' Z_{0i}' M_i e_i\right| = O_p(B_{NT})$ for each i, where B_{NT} represents certain speed. However, this is not necessarily true. We provide three representative counterexamples.

Counterexample 1: X_{iT} is iid over i, $X_{iT}=O_p(1)$, but $E(X_{iT})\to\infty$ as $T\to\infty$.

Suppose $P(X_{iT}=0)=1-\frac{1}{T}$ and $P(X_{iT}=T^2)=\frac{1}{T}$, then $E(X_{iT})=T$, $Var(X_{iT})=T^3-T^2$, $X_{iT}\xrightarrow{p}0$ as $T\to\infty$ for each i and for each T, $\frac{1}{N}\sum_{i=1}^{N}X_{iT}\xrightarrow{p}\frac{1}{N}\sum_{i=1}^{N}E(X_{iT})=T$ as $N\to\infty$. This implies when both N and T are large, $\frac{1}{N}\sum_{i=1}^{N}X_{iT}$ will be close to a large number with high probability. This contradicts with $\frac{1}{N}\sum_{i=1}^{N}X_{iT}=O_p(1)$.

Counterexample 2: X_{iT} is independent over i, $X_{iT}=O_p(1)$ and $E(X_{iT})$ is bounded as $T\to\infty$, but $E(X_{iT})$ is not uniformly bounded over i.

Suppose X_{iT} follows $\chi^2(i)$ for all T and is independent over i, then $E(\frac{1}{N}\sum_{i=1}^{N}X_{iT})=\frac{N+1}{2}$ and $Var(\frac{1}{N}\sum_{i=1}^{N}X_{iT})=\frac{N+1}{N}$, and it follows $\frac{1}{N}\sum_{i=1}^{N}X_{iT}=O_p(N)$.

Counterexample 3: X_{iT} is independent over i, $X_{iT}=O_p(1)$ and $E(X_{iT})$ is uniformly bounded over i and T, but $Var(X_{iT})$ is not uniformly bounded over i.

Suppose X_{iT} follows $N(0,i^2)$ for all T and is independent over i, then $E(X_{iT})=0$ for all i and T, $E(\frac{1}{N}\sum_{i=1}^{N}X_{iT})=0$ and $Var(\frac{1}{N}\sum_{i=1}^{N}X_{iT})=\frac{(N+1)(2N+1)}{6N}\approx\frac{N}{3}$, and it follows $\frac{1}{N}\sum_{i=1}^{N}X_{iT}=O_p(\sqrt{N})$.

In Bai (2010), Kim (2011) and Kim (2014), this problem is partially solved by utilizing the specificity of the regressors. In the mean shift setup, $x_{it}=1$ for all i and t and in the time trend setup, $x_{it}=t$ for all i. These two cases do not belong to any of the above counterexamples and for such special regressors, the second term of (12) (also the other terms) can be further algebraically simplified so that calculating the stochastic order is feasible. In current setup with general regressors, new method is required. Inspired by the above counterexamples, a feasible solution is to show $E(\sup_{k\in O}\left|\frac{1}{|k-k_0|}\delta_i' Z_{0i}' M_i e_i\right|)\leq MB_{NT}$ for some $M<\infty$ and all i and T. Once this is done, it follows by Markov inequality that for large constant C,

$$P(\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^{N}\delta_i' Z_{0i}' M_i e_i\right|>C)\leq P(\sum_{i=1}^{N}\sup_{k\in O}\left|\frac{1}{|k-k_0|}\delta_i' Z_{0i}' M_i e_i\right|>C)\leq\frac{NMB_{NT}}{C},$$

so that $\sup_{k\in O}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i' Z_{0i}' M_i e_i\right| = O_p(NB_{NT})$. Thus to implement this method, the key step is to control the expectation of sup-type terms uniformly over both i and T. For this, we introduce a more powerful tool:

Lemma 1 *General Hajek-Renyi inequality (Theorem 1.1 of Fazekas and Klesov (2001)): Let* β_1*,* β_2*,...,* β_T *be a sequence of nondecreasing positive numbers. Let* α_1*,* α_2*,...,* α_T *be a sequence of nonnegative numbers. Let* r *be a fixed positive number. Let* $\{x_t, t = 1, ...\}$ *be a sequence of random variables and* $S_l = \sum_{t=1}^l x_t$*. Assume that for each* m *with* $1 \le m \le T$*,* $E(\sup_{1\le l\le m} |S_l|^r) \le \sum_{l=1}^m \alpha_l$*, then* $E(\sup_{1\le l\le T}\left|\frac{S_l}{\beta_l}\right|^r) \le 4\sum_{l=1}^T \frac{\alpha_l}{\beta_l^r}$*.*

This lemma permits calculating the order of the expectation of sup-type terms, rather than just the stochastic order of sup-type terms. Note that no dependence structure of x_t is assumed. This lemma also permits controlling the expectation uniformly over i if we assume the r-th moment is uniformly bounded over i. Consider the following representative example.

Example 1 *Suppose for each* i*,* $\{x_{it}, t = 1, ...\}$ *is a sequence of random variables and* $S_{il} = \sum_{t=1}^l x_{it}$*. If Doob's maximal inequality is applicable, then for each* i *and each* m *with* $1 \le m \le T$*,* $E(\sup_{1\le l\le m} |S_{il}|^r) \le (\frac{r}{r-1})^r E(|S_{im}|^r)$*. Take* $r = 2$ *and assume* $E(S_{im}^2) = O(m)$ *uniformly over* i*, i.e., there exists* $M > 0$ *such that* $E(S_{im}^2) \le mM$ *for all* i*, we can take* $\alpha_{il} = 4M$ *so that* $E(\sup_{1\le l\le m} |S_{il}|^2) \le \sum_{l=1}^m \alpha_{il}$ *for each* i*. If we take* $\beta_l = \sqrt{l}$*, it follows from this lemma that for each* i*,*

$$E(\sup_{1\le l\le T}\left|\frac{1}{\sqrt{l}}S_{il}\right|^2) \le 4\sum_{l=1}^T \frac{\alpha_{il}}{l} = 16M\sum_{l=1}^T \frac{1}{l} \approx 16M\log T,$$

and for each i *and some* $\eta > 0$*,*

$$\begin{aligned}
E(\sup_{T\eta+1\le l\le T}\left|\frac{1}{\sqrt{l}}S_{il}\right|^2) &\le 16M\sum_{l=T\eta+1}^T \frac{1}{l} = 16M(\sum_{l=1}^T \frac{1}{l} - \sum_{l=1}^{T\eta}\frac{1}{l}) \\
&= 16M[(\sum_{l=1}^T \frac{1}{l} - \log T) - (\sum_{l=1}^{T\eta}\frac{1}{l} - \log T\eta) + (\log T - \log T\eta)] \\
&\to 16M[\gamma - \gamma + (\log T - \log T\eta)] = 16M\log\frac{1}{\eta},
\end{aligned}$$

where γ *is the Euler-Mascheroni constant, thus* $E(\sup_{T\eta+1\le l\le T}\left|\frac{1}{\sqrt{l}}S_{il}\right|^2)$ *is uniformly bounded over* i*. If we take* $r = 4$ *and* $\alpha_{il} = (\frac{4}{3})^4(l^2 - (l-1)^2)M$*, we can also show that for each* i*,*

$$E(\sup_{1\le l\le T}\left|\frac{1}{\sqrt{l}}S_{il}\right|^4) \le 4\sum_{l=1}^T \frac{\alpha_{il}}{l^2} = 4(\frac{4}{3})^4 M\sum_{l=1}^T \frac{2l-1}{l^2} \approx 8(\frac{4}{3})^4 M(\log T).$$

This example illustrates how we calculate the order of the expectation of sup-type terms in the Appendix and shows the power of Lemma 1.

4 Asymptotics with weak cross-sectional dependence

This section considers the asymptotic properties of the least squares estimator when cross-sectional dependence is weak. We first present some regularity conditions.

Assumption 1 *$\tau_0 = k_0/T \in (0,1)$.*

The change point is assumed to be bounded away from 1 and T such that the size of each subsample is a positive fraction of the total sample size. This is a conventional assumption in change point literature.

Assumption 2 *(1) $E(x_{it}x_{it}') = \Sigma_i^X$ and for all i, $0 < \rho_1 < \rho_{\min}(\Sigma_i^X) < \rho_{\max}(\Sigma_i^X) < \rho_2 < \infty$.*

(2) There exists $\rho_0 > 0$ such that for some $\eta > 0$ and all T and i, $\inf\limits_{k>T(\tau_0-\eta)} \rho_{\min}(\frac{X_{1i}'X_{1i}}{k}) > \rho_0$ and $\inf\limits_{k<T(\tau_0+\eta)} \rho_{\min}(\frac{X_{2i}'X_{2i}}{T-k}) > \rho_0$.

(3) (Doob's maximal inequality) Define $\{R_i(1,k) = \sum_{t=1}^{k}(x_{it}x_{it}' - \Sigma_i^X)\}$, $\{R_i(k,k_0) = \sum_{t=k+1}^{k_0}(x_{it}x_{it}' - \Sigma_i^X)\}$, $\{R_i(k_0+1,k) = \sum_{t=k_0+1}^{k}(x_{it}x_{it}' - \Sigma_i^X)\}$, $\{R_i(k,T) = \sum_{t=k+1}^{T}(x_{it}x_{it}' - \Sigma_i^X)\}$ and $R_{ijm}(1,k)$, $R_{ijm}(k,k_0)$, $R_{ijm}(k_0+1,k)$, $R_{ijm}(k,T)$ as the j-th row and m-th column of $R_i(1,k)$, $R_i(k,k_0)$, $R_i(k_0+1,k)$, $R_i(k,T)$ respectively, then for $1 \le j \le p$, $1 \le m \le p$ and $1 < r < \infty$,

$E(\sup\limits_{1\le l\le k} |R_{ijm}(1,l)|^r) \le (\frac{r}{r-1})^r E(|R_{ijm}(1,k)|^r)$ *for all $1 \le i \le N$ and $1 \le k \le T$,*

$E(\sup\limits_{k\le l\le k_0-1} |R_{ijm}(l,k_0)|^r) \le (\frac{r}{r-1})^r E(|R_{ijm}(k,k_0)|^r)$ *for all $1 \le i \le N$ and $0 \le k \le k_0 - 1$,*

$E(\sup\limits_{k_0+1\le l\le k} |R_{ijm}(k_0+1,l)|^r) \le (\frac{r}{r-1})^r E(|R_{ijm}(k_0,k)|^r)$ *for all $1 \le i \le N$ and $k_0 + 1 \le k \le T$,*

$E(\sup\limits_{k\le l\le T-1} |R_{ijm}(l,T)|^r) \le (\frac{r}{r-1})^r E(|R_{ijm}(k,T)|^r)$ *for all $1 \le i \le N$ and $0 \le k \le T - 1$.*

(4) There exists $M > 0$ such that for $r = 2, 4$, $1 \le j \le p$ and $1 \le m \le p$,

$E(|R_{ijm}(1,k)|^r) < k^{\frac{r}{2}} M$ *for all $1 \le i \le N$ and $1 \le k \le T$,*

$E(|R_{ijm}(k,k_0)|^r) < (k_0 - k)^{\frac{r}{2}} M$ *for all $1 \le i \le N$ and $0 \le k \le k_0 - 1$,*

$E(|R_{ijm}(k_0+1,k)|^r) < (k - k_0)^{\frac{r}{2}} M$ *for all $1 \le i \le N$ and $k_0 + 1 \le k \le T$,*

$E(|R_{ijm}(k,T)|^r) < (T - k)^{\frac{r}{2}} M$ *for all $1 \le i \le N$ and $0 \le k \le T - 1$.*

(5) Define $\lambda_N = \sum_{i=1}^{N} \delta_i'\delta_i$ and $\xi = \lim\limits_{N\to\infty} \frac{1}{\lambda_N} \sum_{i=1}^{N} \delta_i'\Sigma_i^{ZZ}\delta_i$, for each t, $\frac{1}{\lambda_N} \sum_{i=1}^{N} \delta_i' z_{it} z_{it}' \delta_i \xrightarrow{p} \xi$ as $N \to \infty$.

Part (1) requires Σ_i^X to be positive definite and bounded uniformly over i. When Σ_i^X is the same for all i, this condition is directly satisfied. Part (2) requires $\frac{X_{1i}'X_{1i}}{k}$ and $\frac{X_{2i}'X_{2i}}{T-k}$ to be uniformly positive definite over i and over $k > T(\tau_0 - \eta)$ and $k < T(\tau_0 + \eta)$ respectively, so that $\left\|(\frac{X_{1i}'X_{1i}}{k})^{-1}\right\|$

and $\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}\right\|$ are uniformly bounded over i. If strong law of large number is applicable, and together with part (1), part (2) is true almost surely. Part (3) assumes that Doob's maximal inequality is applicable to the process $R_{ijm}(1,k)$, $R_{ijm}(k,k_0)$, $R_{ijm}(k_0+1,k)$ and $R_{ijm}(k,T)$ for $1 \leq i \leq N$ and $1 \leq j,m \leq p$. Doob's maximal inequality has proved to be applicable to various process, including i.i.d. sequence, martingale and submartingale. For economic data, this condition can be easily satisfied. Part (4) further requires the r-th moment of $R_{ijm}(1,k)$, $R_{ijm}(k,k_0)$, $R_{ijm}(k_0+1,k)$ and $R_{ijm}(k,T)$ to be $O(k^{\frac{r}{2}})$, $O((k_0-k)^{\frac{r}{2}})$, $O((k-k_0)^{\frac{r}{2}})$ and $O((T-k)^{\frac{r}{2}})$ uniformly over i, respectively. This will be satisfied if the regressors are weakly dependent over t. Parts (3) and (4) together enables using Lemma 1 to calculate the order of sup-type terms. Note that here we do not assume specific data generating process, thus our assumptions are quite general. Part (5) assumes weak law of large number is applicable to $\frac{1}{\lambda_N}\sum_{i=1}^N \delta_i' z_{it} z_{it}' \delta_i$ for each t.

Assumption 3 *(1) e_{it} is independent with x_{js} for all i,t,j,s.*

(2) (Doob's maximal inequality) Define $\{S_i(1,k)=\sum_{t=1}^k x_{it}e_{it}\}$, $\{S_i(k,k_0)=\sum_{t=k+1}^{k_0} x_{it}e_{it}\}$, $\{S_i(k_0+1,k)=\sum_{t=k_0+1}^{k} x_{it}e_{it}\}$, $\{S_i(k,T)=\sum_{t=k+1}^{T} x_{it}e_{it}\}$ and $S_{ij}(1,k)$, $S_{ij}(k,k_0)$, $S_{ij}(k_0+1,k)$, $S_{ij}(k,T)$ as the j-th element of $S_i(1,k)$, $S_i(k,k_0)$, $S_i(k_0+1,k)$, $S_i(k,T)$ respectively, then for $1 \leq j \leq p$ and $1 < r < \infty$,

$$E(\sup_{1\leq l\leq k}|S_{ij}(1,l)|^r) \leq (\tfrac{r}{r-1})^r E(|S_{ij}(1,k)|^r) \text{ for all } 1\leq i\leq N \text{ and } 1\leq k\leq T,$$
$$E(\sup_{k\leq l\leq k_0-1}|S_{ij}(l,k_0)|^r) \leq (\tfrac{r}{r-1})^r E(|S_{ij}(k,k_0)|^r) \text{ for all } 1\leq i\leq N \text{ and } 0\leq k\leq k_0-1,$$
$$E(\sup_{k_0+1\leq l\leq k}|S_{ij}(k_0+1,l)|^r) \leq (\tfrac{r}{r-1})^r E(|S_{ij}(k_0,k)|^r) \text{ for all } 1\leq i\leq N \text{ and } k_0+1\leq k\leq T,$$
$$E(\sup_{k\leq l\leq T-1}|S_{ij}(l,T)|^r) \leq (\tfrac{r}{r-1})^r E(|S_{ij}(k,T)|^r) \text{ for all } 1\leq i\leq N \text{ and } 0\leq k\leq T-1.$$

(3) There exists $M>0$ such that for $r=2,4$ and for $1\leq j\leq p$,

$$E(|S_{ij}(1,k)|^r) < k^{\frac{r}{2}}M \text{ for all } 1\leq i\leq N \text{ and } 1\leq k\leq T,$$
$$E(|S_{ij}(k,k_0)|^r) < (k_0-k)^{\frac{r}{2}}M \text{ for all } 1\leq i\leq N \text{ and } 0\leq k\leq k_0-1,$$
$$E(|S_{ij}(k_0+1,k)|^r) < (k-k_0)^{\frac{r}{2}}M \text{ for all } 1\leq i\leq N \text{ and } k_0+1\leq k\leq T,$$
$$E(|S_{ij}(k,T)|^r) < (T-k)^{\frac{r}{2}}M \text{ for all } 1\leq i\leq N \text{ and } 0\leq k\leq T-1.$$

(4) Define $\eta_{Nt}=\frac{1}{\sqrt{\lambda_N}}\sum_{i=1}^N \delta_i' z_{it} e_{it}$, there exists $M>0$ such that

$$E(\sup_{k\leq l\leq k_0-1}\left|\sum_{t=l+1}^{k_0}\eta_{Nt}\right|^2) \leq 4E(\left|\sum_{t=k+1}^{k_0}\eta_{Nt}\right|^2) \leq (k_0-k)M \text{ for all } N \text{ and } 0\leq k\leq k_0-1,$$
$$E(\sup_{k_0+1\leq l\leq k}\left|\sum_{t=k_0+1}^{l}\eta_{Nt}\right|^2) \leq 4E(\left|\sum_{t=k_0+1}^{k}\eta_{Nt}\right|^2) \leq (k-k_0)M \text{ for all } N \text{ and } k_0+1\leq k\leq T.$$

(5) Define $\phi_{st}=\lim_{N\to\infty}E(\frac{1}{\lambda_N}\sum_{i=1}^N\sum_{j=1}^N \delta_i' z_{is} z_{jt}' \delta_j e_{is} e_{jt})$ as the limit of the covariance of η_{Ns} and η_{Nt}. For any fixed $C>0$, $(\eta_{N,k_0-C},...,\eta_{N,k_0+C})' \xrightarrow{d} (Z_{-C},...,Z_C)'$ as $N\to\infty$, where $(Z_{-C},...,Z_C)'$

follows multivariate normal distribution with mean zero and covariance $\phi_{st}, k_0 - C \leq s, t \leq k_0 + C$.

Part (1) assumes the error terms are independent of the regressors. Parts (2) and (3) are analogous to parts (3) and (4) of Assumption 2. Part (2) requires Doob's maximal inequality to be applicable to the process $S_{ij}(1,k)$, $S_{ij}(k,k_0)$, $S_{ij}(k_0+1,k)$ and $S_{ij}(k,T)$ for $1 \leq i \leq N$ and $1 \leq j \leq p$. Part (3) requires weak serial dependence of $x_{it}e_{it}$ for each i. Part (4) is a combination of parts (2) and (3), but imposed on the weighted cross-sectional average. Part (5) assumes central limit theorem is applicable to the fixed dimensional random vector $\{\eta_{Nt}, t = k_0 - C, ..., k_0 + C\}$, thus cross-sectional dependence of e_{it} can be allowed but need to be weak.

Given the above regularity conditions on the DGP, it is easy to see that asymptotic properties of $\hat{k}$ should depend on the change magnitude, λ_N, and the N-T ratio as $(N,T) \to \infty$. We consider three sets of conditions.

Assumption 4 *Assume* $\max\limits_{1 \leq i \leq N} \delta_i' \delta_i = O(\frac{1}{N})$ *and as* $(N,T) \to \infty$,

(a) $\lambda_N \to \lambda < \infty$ *and* $\frac{N}{\sqrt{T}} \to 0$.

(b) $\lambda_N \to \infty$ *and* $\frac{N}{\sqrt{T}} \to 0$.

(c) $\liminf\limits_{N \to \infty} \frac{\lambda_N}{N} > 0$.

Similar sets of conditions are also considered in Bai (2010). $\max\limits_{1 \leq i \leq N} \delta_i' \delta_i = O(\frac{1}{N})$ is imposed to ensure the change magnitude of each series is of similar order so that no series will be dominant.

Theorem 1 *Under Assumptions 1-3 and 4(a) or 4(b) or 4(c),* $\hat{\tau}$ *is consistent as* $(N,T) \to \infty$.

This result is mainly of theoretical importance. Recall that the least squares estimator is searched in the whole set $[1,T]$, given the consistency of $\hat{\tau}$, the search region can be narrowed down to a local region of k_0. Within this local region, the order of sup-type terms can be established more accurately so that we can move one step further to improve the convergence rate.

Theorem 2 *Under Assumptions 1-3 and 4(a),* $\hat{k} - k_0 = O_p(1)$ *as* $(N,T) \to \infty$.

When $\lambda_N \to \lambda$, the change magnitude is of the same order as that of the univariate case, thus not surprisingly the result is also the same, see Bai (1997). Here the extra condition $\frac{N}{\sqrt{T}} \to 0$[1] is imposed to deal with the nuisance parameters $\beta_i, i = 1, ..., N$[2]. With $\hat{\beta}_i$ plugged in the least

[1] The condition $\frac{N}{\sqrt{T}} \to 0$ is slightly stricter than that of Bai (2010), $\frac{N \log T}{T} \to 0$, but the spirit is the same. Bai (2010) considers the mean shift setup, in current setup we do not have the algebraical specificity of the regressors.

[2] In terms of estimating the change point, $\beta_i, i = 1, ..., N$ are nuisance parameters.

squares criterion function, for each i, the difference $\hat{\beta}_i - \beta_i$ would result in an extra source of noise. It can be shown each noise is $O(\frac{1}{\sqrt{T}})$, hence when $\frac{N}{\sqrt{T}} \to 0$, T is large enough to control the total noise resulting from nuisance parameters. If we let $\lambda_N \to \infty$ while still maintain $\frac{N}{\sqrt{T}} \to 0$, then we will have consistency of $\hat{k}$.

Theorem 3 *Under Assumptions 1-3 and 4(b) or 4(c), $\hat{k}$ is consistent as $(N, T) \to \infty$.*

While consistency under Assumption 4(b) still relies on $\frac{N}{\sqrt{T}} \to 0$, consistency under Assumption 4(c) only requires $T \to \infty$. This is because when $\lambda_N = O(N)$, the change magnitude is large enough to overwhelm the nuisance parameters problem. Assumption 4(c) is satisfied when the change magnitude of each series is nonnegligible, thus this result confirms Bai (2010) and Kim (2011) in current regression setup that increasing the number of series helps to identify the change point when cross-sectional dependence of the error terms is weak.

Remark 2 *It's worth point out that once $\hat{k}$ is consistent, the convergence rate of $\hat{k}$ is not well defined since $\hat{k}$ has to be integers. If $\hat{\tau}$ is defined as $\hat{k}/T$, $\hat{\tau}$ has the same problem since $T\hat{\tau}$ has to be integers. In the textbook definition, for a sequence of random variables $\{X_n, n = 1, ...\}$ and a sequence of positive numbers $\{C_n, n = 1, ...\}$, $X_n = O_p(C_n)$ is defined in the sense that X_n/C_n is bounded in probability. In most cases, we then derive the limiting distribution of X_n/C_n. However, when X_n is restricted to be integers, this definition is no longer appropriate. Suppose X_n is consistent for some integer θ, i.e., $P(|X_n - \theta| = 0) \to 1$, then for any C_n, $P(|X_n - \theta| / C_n = 0) = P(|X_n - \theta| = 0) \to 1$. This implies the convergence rate of X_n is arbitrary and the limiting distribution of X_n/C_n is meaningless. Coming back to $\hat{k}$, the convergence rate of $\hat{k}$ will be arbitrary once $\hat{k}$ is consistent, and it's meaningless to derive the limiting distribution of $\hat{k} - k_0$ by multiplying $\hat{k} - k_0$ by some magnifying speed, say, N.*

Except for the above theoretical concern, in practice the change magnitude maybe small and some series may not have structural change. Therefore, we will derive the limiting distribution of $\hat{k}$ under Assumption 4(a).

Theorem 4 *Under Assumptions 1-3 and 4(a),*

$$\hat{k} - k_0 \xrightarrow{d} \arg\max W(m),$$

where $W(m)$ is a partial sum process,

$$W(m)=\begin{cases} -|m|\,\lambda\xi+2\sqrt{\lambda}\sum_{t=m+1}^{0}Z_t, & for\ m\le -1,\\ 0, & for\ m=0,\\ -|m|\,\lambda\xi-2\sqrt{\lambda}\sum_{t=1}^{m}Z_t, & for\ m\ge 1,\end{cases} \tag{13}$$

and $\{Z_t, t=-(k_0-1),...,0,...,T-k_0\}$ is a discrete time Gaussian process with mean zero and autocovariance $\{\phi_{st}, 1\le s,t\le T\}$.

The key feature of this distribution is that it's free of the underlying DGP so that inference of the change point can be made. Different from the univariate case in which normality comes from applying function central limit theorem to the weighted serial average $v_T\sum_{t=k+1}^{k_0}\delta_0' z_t e_t$, where $\delta_T=\delta_0 v_T$ and $v_T\to 0$ as $T\to\infty$, here the normality comes from applying central limit theorem to the weighted cross-sectional average $\frac{1}{\sqrt{\lambda_N}}\sum_{i=1}^{N}\delta_i' z_{it}e_{it}$, also see Yao (1987), Bai (1997), Bai (2010) and Kim (2011). However, the essence of these two frameworks are the same. A second feature is that this distribution is derived allowing $z_{it}e_{it}$ to be dependent over t, while in Bai (2010) $z_{it}e_{it}$ is assumed to be uncorrelated over t. Thus our result is more general and empirically relevant. This step is nontrivial, see Appendix for details, our proof is self-contained. Also note that the DGP is not required to be stationary even within each regime. The autocovariance function ϕ_{st} could be of any form, as long as parts (4) and (5) of Assumption 3 are satisfied.

It remains to estimate the parameters in the limiting distribution. λ and ξ can be estimated by $\hat{\lambda}_N=\sum_{i=1}^{N}\hat{\delta}_i'\hat{\delta}_i$ and $\hat{\xi}=\frac{1}{T}\frac{1}{\hat{\lambda}_N}\sum_{t=1}^{T}\sum_{i=1}^{N}\hat{\delta}_i' z_{it}z_{it}'\hat{\delta}_i$, where $\hat{\delta}_i$ and $\hat{e}_{is}$ can be obtained by least squares estimation of each subsample split at $\hat{k}$, and it will not difficult to show the consistency of $\hat{\lambda}_N$ and $\hat{\xi}$. ϕ_{st} can be estimated by $\hat{\phi}_{st}=\frac{1}{\hat{\lambda}_N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\delta}_i' z_{is}z_{it}'\hat{\delta}_i\hat{e}_{is}\hat{e}_{it}$ and if we assume the DGP is independent over i, $\hat{\phi}_{st}$ can be simplified to $\frac{1}{\hat{\lambda}_N}\sum_{i=1}^{N}\hat{\delta}_i' z_{is}z_{it}'\hat{\delta}_i\hat{e}_{is}\hat{e}_{it}$. For each (s,t), it will not be difficult to show the consistency of $\hat{\phi}_{st}$. However, the limiting distribution relies on the consistency of the whole estimated covariance matrix $\{\hat{\phi}_{st}, 1\le s,t\le T\}$. If we impose further assumption on $z_{it}e_{it}$, e.g., AR(1) or martingale difference, then consistency of $\{\hat{\phi}_{st}, 1\le s,t\le T\}$ also will not be difficult. Once these estimated parameters are available, we can simulate the distribution directly and inference can be made based on this simulated distribution.

5 Estimation with strong cross-sectional dependence

This section considers estimating the change point when cross-sectional dependence is strong due to common factors. When factors are observable, once explicitly incorporated in the model, we are back to the case with weak cross-sectional dependence. When factors are unobservable and we estimate the change point ignoring the factors, the least squares estimator typically will be inconsistent even under Assumption 4(c). This is because when the cross-sectional dependence is strong, increasing the number of series no longer help to identify the change point, Kim (2011) also discussed this phenomenon in the time trend break setup. In this case, a feasible way to recover consistency is using estimated factors to control cross-sectional dependence, similar method also can be found in Bai (2009) and Kim (2014).

We first present some regularity conditions.

Assumption 5 *(1)* $E\left\|F_t^0\right\|^4 < M < \infty$, $E(F_t^0) = 0$, $E(F_t^0 F_t^{0\prime}) = \Sigma_F$ *and* Σ_F *is positive definite.*

(2) (Doob's maximal inequality) Define $\{Q(1,k) = \sum_{t=1}^{k}(F_t^0 F_t^{0\prime} - \Sigma_F)\}$, $\{Q(k,k_0) = \sum_{t=k+1}^{k_0}(F_t^0 F_t^{0\prime} - \Sigma_F)\}$, $\{Q(k_0+1,k) = \sum_{t=k_0+1}^{k}(F_t^0 F_t^{0\prime} - \Sigma_F)\}$, $\{Q(k,T) = \sum_{t=k+1}^{T}(F_t^0 F_t^{0\prime} - \Sigma_F)\}$ *and* $Q_{jm}(1,k)$, $Q_{jm}(k,k_0)$, $Q_{jm}(k_0+1,k)$, $Q_{jm}(k,T)$ *as the j-th row and m-th column of* $Q(1,k)$, $Q(k,k_0)$, $Q(k_0+1,k)$, $Q(k,T)$ *respectively, then for* $1 \le j,m \le s$ *and* $1 < r < \infty$,

$E(\sup_{1\le l\le k} |Q_{jm}(1,l)|^r) \le (\frac{r}{r-1})^r E(|Q_{jm}(1,k)|^r)$ *for* $1 \le k \le T$,

$E(\sup_{k\le l\le k_0-1} |Q_{jm}(l,k_0)|^r) \le (\frac{r}{r-1})^r E(|Q_{jm}(k,k_0)|^r)$ *for* $0 \le k \le k_0 - 1$,

$E(\sup_{k_0+1\le l\le k} |Q_{jm}(k_0+1,l)|^r) \le (\frac{r}{r-1})^r E(|Q_{jm}(k_0,k)|^r)$ *for* $k_0 + 1 \le k \le T$,

$E(\sup_{k\le l\le T-1} |Q_{jm}(l,T)|^r) \le (\frac{r}{r-1})^r E(|Q_{jm}(k,T)|^r)$ *for* $0 \le k \le T-1$.

(3) There exists $M > 0$ *such that for* $r = 2,4$ *and* $1 \le j,m \le s$,

$E(|Q_{jm}(1,k)|^r) < k^{\frac{r}{2}} M$ *for* $1 \le k \le T$,

$E(|Q_{jm}(k,k_0)|^r) < (k_0-k)^{\frac{r}{2}} M$ *for* $0 \le k \le k_0 - 1$,

$E(|Q_{jm}(k_0+1,k)|^r) < (k-k_0)^{\frac{r}{2}} M$ *for* $k_0 + 1 \le k \le T$,

$E(|Q_{jm}(k,T)|^r) < (T-k)^{\frac{r}{2}} M$ *for* $0 \le k \le T-1$.

Part (1) mainly assumes factors have uniformly bounded fourth moment. Parts (2) and (3) are analogous to parts (3) and (4) of Assumption 2. Part (2) requires Doob's maximal inequality to be applicable to the process $Q_{jm}(1,k)$, $Q_{jm}(k,k_0)$, $Q_{jm}(k_0+1,k)$ and $Q_{jm}(k,T)$ for $1 \le j,m \le s$. Part (3) requires the factors to be serially weakly dependent, hence integrated factors are not allowed. Part (3) also implies $\frac{1}{k_0}\sum_{t=1}^{k_0} F_t^0 F_t^{0\prime} \xrightarrow{p} \Sigma_F$ and $\frac{1}{T-k_0}\sum_{t=k_0+1}^{T} F_t^0 F_t^{0\prime} \xrightarrow{p} \Sigma_F$.

Assumption 6 *(1)* x_{it} *is independent with* F_t^0 *for all* i,t.

(2) Define $w_{it} = (x_{it}', F_t^{0\prime})'$, $W_{1i} = (w_{i1},...,w_{ik},0,...,0)'$ *and* $W_{2i} = (0,...,0,w_{i,k+1},...,w_{iT})'$, *there exists* $\rho_0 > 0$ *such that for some* $\eta > 0$ *and all* T *and* i, $\inf_{k>T(\tau_0-\eta)} \rho_{\min}(\frac{W_{1i}'W_{1i}}{k}) > \rho_0$ *and*

$\inf_{k<T(\tau_0+\eta)} \rho_{\min}(\frac{W_{2i}'W_{2i}}{T-k}) > \rho_0.$

Part (1) is assumed to simplify analysis, since our emphasis is the effect of cross-sectional dependence on the asymptotic property of the change point estimator. Part (2) is analogous to part (2) of Assumption 2 and has similar interpretation.

Assumption 7 $\|\lambda_i\| \leq \bar{\lambda} < \infty$, $\left\|\frac{1}{N}\Lambda'\Lambda - \Sigma_\Lambda\right\| \to 0$ *for some positive definite matrix* Σ_Λ.

Assumption 8 *The eigenvalues of* $\Sigma_F \Sigma_\Lambda$ *are distinct.*

Assumption 9 *(1)* e_{it} *is independent with* F_s^0 *for all* i, t, s.

(2) (Doob's maximal inequality) Define $\{P_i(1,k) = \sum_{t=1}^{k} F_t^0 e_{it}\}$, $\{P_i(k,k_0) = \sum_{t=k+1}^{k_0} F_t^0 e_{it}\}$, $\{P_i(k_0+1,k) = \sum_{t=k_0+1}^{k} F_t^0 e_{it}\}$, $\{P_i(k,T) = \sum_{t=k+1}^{T} F_t^0 e_{it}\}$ *and* $P_{ij}(1,k)$, $P_{ij}(k,k_0)$, $P_{ij}(k_0+1,k)$, $P_{ij}(k,T)$ *as the j-th element of* $P_i(1,k)$, $P_i(k,k_0)$, $P_i(k_0+1,k)$, $P_i(k,T)$ *respectively, then for* $1 \leq j \leq s$ *and* $1 < r < \infty$,

$E(\sup_{1\leq l\leq k} |P_{ij}(1,l)|^r) \leq (\frac{r}{r-1})^r E(|P_{ij}(1,k)|^r)$ *for all* $1 \leq i \leq N$ *and* $1 \leq k \leq T$,

$E(\sup_{k\leq l\leq k_0-1} |P_{ij}(l,k_0)|^r) \leq (\frac{r}{r-1})^r E(|P_{ij}(k,k_0)|^r)$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq k_0 - 1$,

$E(\sup_{k_0+1\leq l\leq k} |P_{ij}(k_0+1,l)|^r) \leq (\frac{r}{r-1})^r E(|P_{ij}(k_0,k)|^r)$ *for all* $1 \leq i \leq N$ *and* $k_0 + 1 \leq k \leq T$,

$E(\sup_{k\leq l\leq T-1} |P_{ij}(l,T)|^r) \leq (\frac{r}{r-1})^r E(|P_{ij}(k,T)|^r)$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq T - 1$.

(3) There exists $M > 0$ *such that for* $r = 2, 4$ *and for* $1 \leq j \leq s$,

$E(|P_{ij}(1,k)|^r) < k^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $1 \leq k \leq T$,

$E(|P_{ij}(k,k_0)|^r) < (k_0 - k)^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq k_0 - 1$,

$E(|P_{ij}(k_0+1,k)|^r) < (k - k_0)^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $k_0 + 1 \leq k \leq T$,

$E(|P_{ij}(k,T)|^r) < (T - k)^{\frac{r}{2}} M$ *for all* $1 \leq i \leq N$ *and* $0 \leq k \leq T - 1$.

Assumption 10 *There exists a positive constant* $M < \infty$ *such that:*

1 $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$, *for all* $i = 1, ..., N$, *and* $t = 1, ..., T$,

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

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

4 $E(e_{it}e_{js}) = \tau_{ij,ts}$ *for* $i, j = 1, ..., N$, *and* $t, s = 1, ..., T$, *also*

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |\tau_{ij,ts}| \leq M,$$

5 For every $(t, s = 1, ..., T)$, $E\left|\frac{1}{\sqrt{N}} \sum_{i=1}^{N} [e_{is}e_{it} - E(e_{is}e_{it})]\right|^4 \leq M$, *6 For each* $u=1,...,T$, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|cov(e_{iu}e_{it},e_{ju}e_{js})|\leq M$ *and for each* $k=1,...,N$, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|cov(e_{it}e_{kt},e_{js}e_{ks})|\leq M$.

Assumption 7 and Assumption 8 are standard in the factor literature. Assumption 9 is analogous to parts (1)-(3) of Assumption 3. Assumption 10 requires weak serial and cross-sectional dependence and heteroskedasticity is allowed. Similar conditions are also assumed in Bai (2009), see the discussion therein for detailed interpretation.

Assumption 11 *There exists* $M<\infty$ *such that:*

1 For each $t=1,...,T$, $E(\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{k_0}\sum_{i=1}^{N}F_s^0[e_{is}e_{it}-E(e_{is}e_{it})]\right\|^2)\leq M$,
and $E(\left\|\frac{1}{\sqrt{NT}}\sum_{s=k_0+1}^{T}\sum_{i=1}^{N}F_s^0[e_{is}e_{it}-E(e_{is}e_{it})]\right\|^2)\leq M$;

2 $E(\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{k_0}\sum_{i=1}^{N}F_t^0\lambda_i'e_{it}\right\|^2)\leq M$ *and* $E(\left\|\frac{1}{\sqrt{NT}}\sum_{t=k_0+1}^{T}\sum_{i=1}^{N}F_t^0\lambda_i'e_{it}\right\|^2)\leq M$;

3 For each $t=1,...,T$, $E(\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_ie_{it}\right\|^4)\leq M$.

Assumption 12 *There exists* $M<\infty$ *such that:*

1 For every $s=1,...,T$,

$E(\sup_{k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it}-E(e_{is}e_{it})]\right|^2)\leq M$,

$E(\sup_{k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it}-E(e_{is}e_{it})]\right|^2)\leq M$,

$E(\sup_{k>k_0}\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it}-E(e_{is}e_{it})]\right|^2)\leq M$,

$E(\sup_{k\geq k_0}\frac{1}{T-k}\sum_{t=k+1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it}-E(e_{is}e_{it})]\right|^2)\leq M$,

2 $E(\sup_{k<k_0}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_ie_{it}\right\|^2)\leq M$,

$E(\sup_{k\leq k_0}\frac{1}{k}\sum_{t=1}^{k}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_ie_{it}\right\|^2)\leq M$,

$E(\sup_{k>k_0}\frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_ie_{it}\right\|^2)\leq M$,

$E(\sup_{k\geq k_0}\frac{1}{T-k}\sum_{t=k+1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_ie_{it}\right\|^2)\leq M$.

Assumption 13 *There exists* $M<\infty$ *such that:*

1 $E(\sup_{k<k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k+1}^{k_0}\sum_{i=1}^{N}F_t^0\lambda_i'e_{it}\right\|^2)\leq M$,

$E(\sup_{k\leq k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{k}\sum_{i=1}^{N}F_t^0\lambda_i'e_{it}\right\|^2)\leq M$,

$E(\sup_{k>k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k_0+1}^{k}\sum_{i=1}^{N}F_t^0\lambda_i'e_{it}\right\|^2)\leq M$,

$E(\sup_{k\geq k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k+1}^{T}\sum_{i=1}^{N}F_t^0\lambda_i' e_{it}\right\|^2)\leq M.$

2 For each $j=1,...,T$,

$E(\sup_{k<k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k+1}^{k_0}\sum_{i=1}^{N}\lambda_i' e_{it}e_{jt}\right\|^2)\leq M,$

$E(\sup_{k\leq k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{k}\sum_{i=1}^{N}\lambda_i' e_{it}e_{jt}\right\|^2)\leq M,$

$E(\sup_{k>k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k_0+1}^{k}\sum_{i=1}^{N}\lambda_i' e_{it}e_{jt}\right\|^2)\leq M,$

$E(\sup_{k\geq k_0}\left\|\frac{1}{\sqrt{NT}}\sum_{t=k+1}^{T}\sum_{i=1}^{N}\lambda_i' e_{it}e_{jt}\right\|^2)\leq M.$

Assumptions 11-13 are not restrictive since the summands are zero mean random variables. If Hajek-Renyi inequality were applicable, these conditions are directly satisfied. If further parametric assumptions are made on the factors, factor loadings and errors, it also will not be difficult to verify these conditions. Here we simply lay them out so that these conditions are in their original form.

Assumption 14 *Assumption A(iii), B and C in Song (2013) hold*[3]*.*

In the proof, we will utilizing results in Han and Inoue (2014), Baltagi, Kao and Wang (2015) and Song (2013) as intermediate steps. It can be verified that the assumptions in these three papers are satisfied given all the above assumptions.

To recover consistency, we will use estimated factors as extra regressors to control the cross-sectional dependence. If the true change point k_0 were known, the factors can be estimated globally with the coefficients β_i as in Song (2013). Song (2013) shows that β_i will be $\sqrt{T}$-consistent for each i and the estimated factor space will be consistent. Without knowing k_0, a feasible way is to use $\hat{k}$, the estimated change point ignoring factors.

Theorem 5 *Under Assumptions 1-3, 4(c), 5 and 6,* $\hat{k}-k_0=O_p(1)$ *as* $(N,T)\to\infty$*.*

This result confirms Kim (2011) in current regression setup that when cross-sectional dependence is strong, more series do not increase the accuracy of the change point estimator. Nevertheless, $\hat{k}-k_0=O_p(1)$ is good enough to estimate the factor space. It can be verified that with $O_p(1)$ estimation error, results in Song (2013) remain the same. Once the estimated factors are available and incorporated in the model as extra regressors, consistency of the least squares estimator can be recovered. Define $\tilde{k}$ as the change point estimator in the second step and $\tilde{\tau}=\tilde{k}/T$ as the estimated change fraction, we first show $\tilde{\tau}$ is consistent.

[3] ϵ_{it} in Song (2013) corresponds to e_{it} here.

Theorem 6 *Under Assumptions 1-3, 4(c) and 5-14, $\tilde{\tau} - \tau_0 = o_p(1)$ as $(N,T) \to \infty$ and $\frac{\sqrt{T}}{N} \to 0$.*

Similar to Theorem 1, this result is mainly of theoretical interest and serves as a intermediate step to show the consistency of $\tilde{k}$. The condition $\frac{\sqrt{T}}{N} \to 0$ is required to guarantee the effect of using estimated factors is asymptotically negligible and appears frequently in the factor literature, see for example Bai and Ng (2006).

Theorem 7 *Under Assumption 1-3, 4(c) and 5-14, $\tilde{k}$ is consistent as $(N,T) \to \infty$ and $\frac{\sqrt{T}}{N} \to 0$.*

Again, $\frac{\sqrt{T}}{N} \to 0$ is required to eliminate the effect of using estimated factors. Note that in Theorem 4, $\frac{N}{\sqrt{T}} \to 0$ is required to eliminate the noise resulting from nuisance parameters, $\beta_i, i = 1, ..., N$. These two conditions are in conflict with each other, and consequently it's infeasible to derive the limiting distribution of $\tilde{k}$. Intuitively speaking, for the factors, T is the dimension and N is the sample size while for β_i, N is the dimension and T is the sample size. If we also regard the factors as nuisance parameters, the effect of these two sets of nuisance parameters will not disappear simultaneously. In some sense, this is the cost of using heterogeneous coefficients in panel data.

6 Simulations

In this section we evaluate the limiting distribution derived in Section 4 and examine the effect of serial correlation. To simplify analysis, we assume x_{it} is i.i.d. $N(1,1)$ over both i and t, $e_{it} = \rho e_{i,t-1} + \sigma_i \eta_{it}$ where η_{it} is i.i.d.$N(0,1)$ over both i and t and σ_i^2 is i.i.d.$\chi_2^2/2$ over i, and δ_i is i.i.d.$U(-1,1)$. For this DGP, $\{Z_t, t = -(k_0 - 1), ..., 0, ..., T - k_0\}$ is a Gaussian process with variance $\phi = \frac{1}{\lambda_N} \sum_{i=1}^N \delta_i' E(z_{it} z_{it}') \delta_i E(e_{it}^2)$ and correlation coefficient $\alpha_{st} = \rho^{|s-t|} \frac{\sum_{i=1}^N \delta_i' E(z_{is} z_{it}') \delta_i \sigma_i^2}{\sum_{i=1}^N \delta_i' E(z_{it} z_{it}') \delta_i \sigma_i^2} = \frac{1}{2} \rho^{|s-t|}$. For given values of N, $\lambda\phi$, $\lambda\xi$ and ρ, we can simulate the distribution of $\arg\max W(m)$ and in current case $\lambda\phi = \sum_{i=1}^N \delta_i' E(z_{it} z_{it}') \delta_i E(e_{it}^2) \approx 2NE(\delta_i^2) E(\sigma_i^2) = \frac{2}{3} N$ and $\lambda\xi = \sum_{i=1}^N \delta_i' E(z_{it} z_{it}') \delta_i = 2NE(\delta_i^2) = \frac{2}{3} N$. Figures 1-2 are the simulated distributions obtained from 2000 replications with $T = 100$, $k_0 = 50$, $N = 1, 5$, 10 and 20 and $\rho = 0$, 0.4 and 0.8 respectively. When $\rho = 0$, the distribution is well shaped, but when $\rho > 0$, the distribution is no longer bell-shaped and becomes highly nonstandard. The probability of taking both ends and the true change point are high while the probability of taking the other points are approximately the same. Here $(\lambda\xi)^2 / \lambda\phi = 2NE(\delta_i^2)$, if $E(\delta_i^2)$ is smaller, the nonstandardness will be more severe. Also note that $\alpha_{st} = \frac{1}{2} \rho^{|s-t|}$, even when $\rho = 0.8$, α_{st} is no more than 0.4. If $\frac{\sum_{i=1}^N \delta_i' E(z_{is} z_{it}') \delta_i \sigma_i^2}{\sum_{i=1}^N \delta_i' E(z_{it} z_{it}') \delta_i \sigma_i^2}$ is larger, the nonstandardness will also be more severe. Furthermore, with $E(\delta_i^2)$ fixed down, while large N increases the probability of $\hat{k} = k_0$, it does not make the distribution more bell-shaped.

For such nonstandard distribution, it maybe better to base inference directly on the distribution, rather than on the constructed confidence intervals. Consider the case $N = 20$ and $\rho = 0.8$ for example. Although the probability of $\hat{k} = k_0$ is already around 0.55, the 90% confidence interval is $[2, 99]$! Therefore, we suggest simulating the distribution directly using the estimated parameters and making inference based on this simulated distribution. For example, in current setup the parameters λ, ξ, ϕ and α_{st} can be estimated by $\hat{\lambda}_N = \sum_{i=1}^N \hat{\delta}_i' \hat{\delta}_i$, $\hat{\xi} = \frac{1}{T}\frac{1}{\hat{\lambda}_N}\sum_{t=1}^T \sum_{i=1}^N \hat{\delta}_i' z_{it} z_{it}' \hat{\delta}_i$, $\hat{\phi} = \frac{1}{\hat{\lambda}_N}\sum_{i=1}^N \hat{\delta}_i' (\frac{1}{NT}\sum_{t=1}^T \sum_{i=1}^N z_{it} z_{it}')\hat{\delta}_i (\frac{1}{T}\sum_{t=1}^T \hat{e}_{it}^2)$ and $\hat{\alpha}_{st} = \hat{\rho}^{|s-t|} \frac{\sum_{i=1}^N [\hat{\delta}_i' (\frac{1}{NT}\sum_{t=1}^T \sum_{i=1}^N z_{it})]^2 (\frac{1}{T}\sum_{t=1}^T \hat{e}_{it}^2)}{\sum_{i=1}^N \hat{\delta}_i' (\frac{1}{NT}\sum_{t=1}^T \sum_{i=1}^N z_{it} z_{it}')\hat{\delta}_i (\frac{1}{T}\sum_{t=1}^T \hat{e}_{it}^2)}$, where $\hat{\delta}_i$ and $\hat{e}_{is}$ can be obtained by least squares estimation of each subsample split at $\hat{k}$ and $\hat{\rho} = \frac{1}{N}\sum_{i=1}^N (\sum_{t=2}^T \hat{e}_{it}\hat{e}_{i,t-1} / \sum_{t=2}^T \hat{e}_{i,t-1}^2)$.

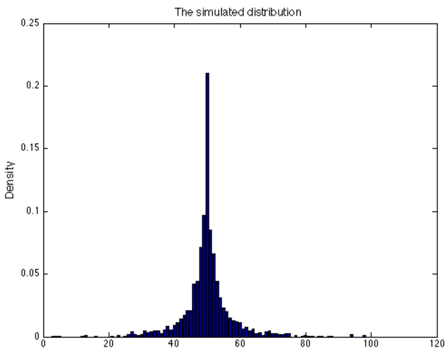
7 Conclusion

This paper studies the joint limit asymptotics of the least squares estimator of a common change point in large heterogeneous panel data models. A general Hajek-Renyi inequality is introduced to solve the fundamental issue that for random variables $X_{iT} = O_p(1)$ (or $o_p(1)$) as $T \to \infty$, $\frac{1}{N}\sum_{i=1}^N X_{iT}$ is not necessarily $O_p(1)$ (or $o_p(1)$ correspondingly) as N and T go to infinity jointly. This new technique is quite powerful and will also be useful in other places. Consistency of the least squares estimator is then established under various sets of conditions on the change magnitude and N-T ratio. Both weak and strong cross-sectional dependence of the errors are considered and in the latter case estimated factors are used to control the cross-sectional dependence. The limiting distribution is derived allowing the errors to be cross-sectionally weakly dependent and serially dependent and heteroskedastic of unknown form, and making inference is therefore feasible based on the simulated distribution using estimated parameters.

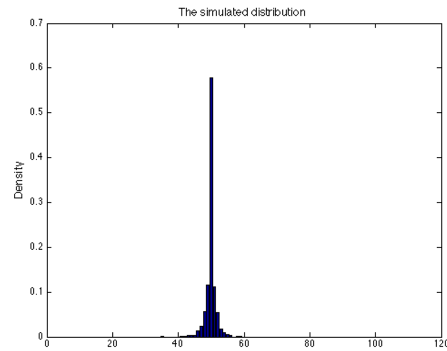
References

[1] Bai, J., 1997. Estimation of a change point in multiple regression models. Review of Economics and Statistics 79, 551–563.

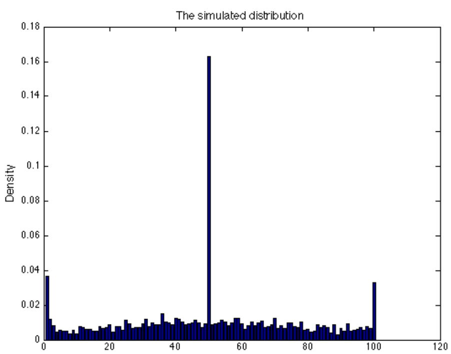
[2] Bai, J., Ng, S., 2006. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. Econometrica 74, 1133-1150.



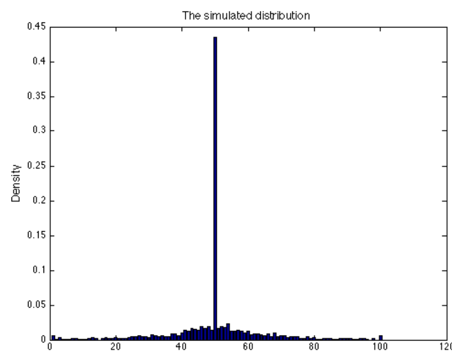
$T = 100,\ k_0 = 50,\ N = 1$ and $\rho = 0$



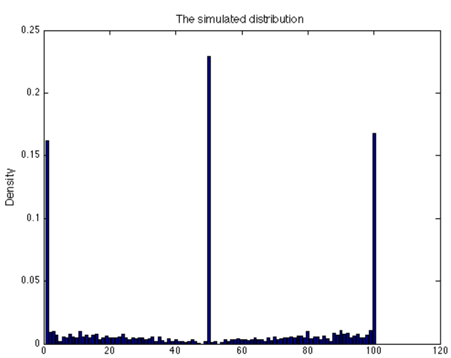
$T = 100,\ k_0 = 50,\ N = 5$ and $\rho = 0$



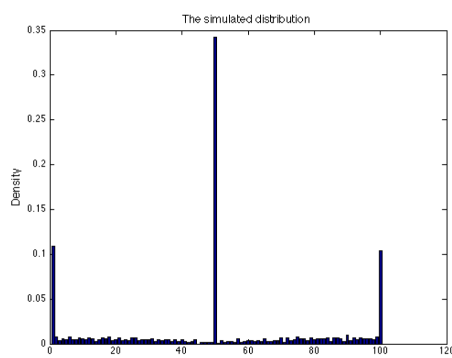
$T = 100,\ k_0 = 50,\ N = 1$ and $\rho = 0.4$



$T = 100,\ k_0 = 50,\ N = 5$ and $\rho = 0.4$

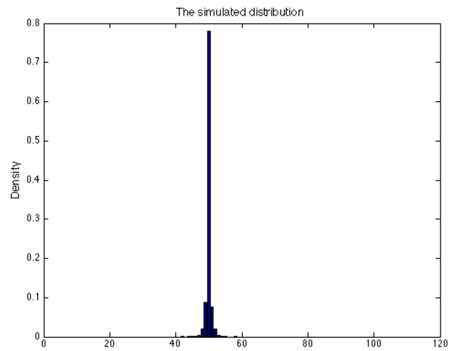


$T = 100,\ k_0 = 50,\ N = 1$ and $\rho = 0.8$

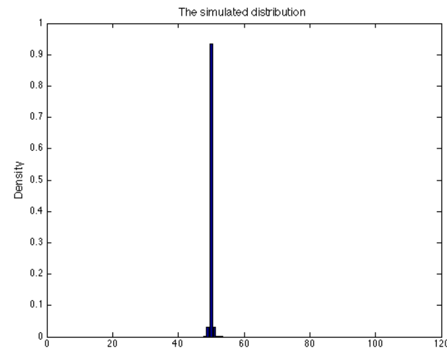


$T = 100,\ k_0 = 50,\ N = 5$ and $\rho = 0.8$

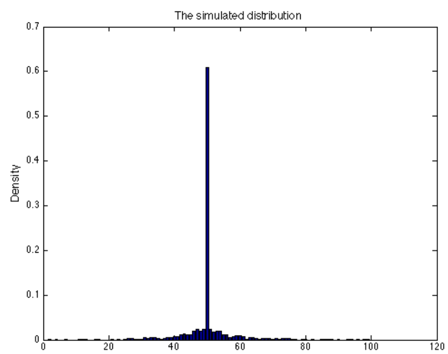
Figure 1: Simulated distribution of $argmaxW(m)$ for $T = 100, k_0 = 50, N = 1$ and 5



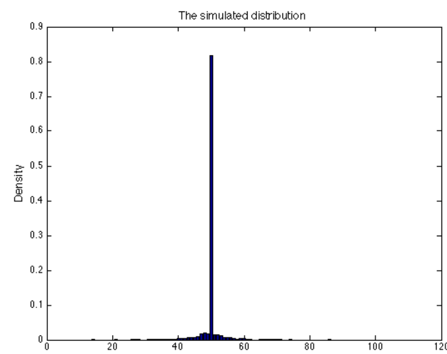
$T = 100,\ k_0 = 50,\ N = 10$ and $\rho = 0$



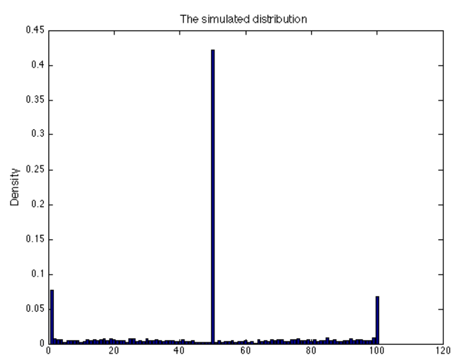
$T = 100,\ k_0 = 50,\ N = 20$ and $\rho = 0$



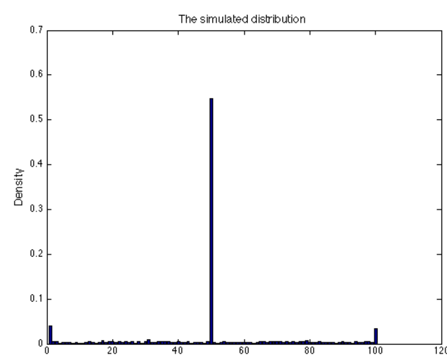
$T = 100,\ k_0 = 50,\ N = 10$ and $\rho = 0.4$



$T = 100,\ k_0 = 50,\ N = 20$ and $\rho = 0.4$



$T = 100,\ k_0 = 50,\ N = 10$ and $\rho = 0.8$



$T = 100,\ k_0 = 50,\ N = 20$ and $\rho = 0.8$

Figure 2: Simulated distribution of $argmaxW(m)$ for $T = 100, k_0 = 50, N = 10$ and 20

[3] Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica, 77, 1229-1279.

[4] Bai, J., 2010. Common breaks in means and variances for panel data. Journal of Econometrics 157, 78–92.

[5] Bai, J., Lumsdaine, R.L., Stock, J.H., 1998. Testing for and dating breaks in multivariate time series. Review of Economic Studies 65, 395–432.

[6] Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

[7] Baltagi, B.H., Feng, Q., Kao, C., 2011. Testing for sphericity in a fixed effects panel data model. Econometrics Journal 14, 25–47.

[8] Baltagi, B.H., Feng, Q., Kao, C., 2012. A Lagrange Multiplier test for cross-sectional dependence in a fixed effects panel data model. Journal of Econometrics 170, 164-177.

[9] Baltagi, B.H., Feng, Q., Kao, C., 2014. Estimation of heterogeneous panels with structural breaks. Working Paper, Syracuse University.

[10] Baltagi, B.H., Kao, C., Liu, L., 2014. Estimation and identification of change points in panel models with nonstationary or stationary regressors and error terms. Working Paper, Syracuse University.

[11] Baltagi, B.H., Kao, C., Wang, F., 2012. Identification and estimation of a large factor model with structural instability. Working paper, Syracuse University.

[12] Fazekas, I., Klesov, O., 2001. A general approach to the strong law of large numbers. Theory of Probability & Its Applications 45, 436-449.

[13] Hájek, J., Rényi, A., 1955. Generalization of an inequality of Kolmogorov. Acta Mathematica Hungarica 6, 281-283.

[14] Han, X., Inoue, A., 2011. Tests for parameter instability in dynamic factor models. Econometric Theory, 1-36.

[15] Kim, D., 2011. Estimating a common deterministic time trend break in large panels with cross sectional dependence. Journal of Econometrics 164, 310-330.

[16] Kim, D., 2014. Common breaks in time trends for large panel data with a factor structure. Econometrics Journal 17, 301-337.

[17] Lee, L.F., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics 154, 165-185.

[18] Lee, L.F., Yu, J., 2010. Some recent developments in spatial panel data models. Regional Science and Urban Economics 40, 255-271.

[19] Ng, S., 2006. Testing cross-section correlation in panel data using spacings. Journal of Business and Economic Statistics 24, 12-23.

[20] Perron, P, 2006. Dealing with structural breaks. Palgrave Handbook of Econometrics 1, 278-352.

[21] Pesaran, M.H., 2004. General diagnostic test for cross section dependence in panels. Working Paper, University of Cambridge & USC.

[22] Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with multifactor error structure. Econometrica 74, 967-1012.

[23] Pesaran, M.H., 2012. Testing weak cross-sectional dependence in large panels. IZA Discussion Paper No. 6432, March 2012.

[24] Pesaran, M.H., Ullah, A., Yamagata, T., 2008. A bias-adjusted LM test of error cross section independence. Econometrics Journal 11, 105-127.

[25] Qu, Z., Perron, P., 2007. Estimating and testing structural changes in multivariate regressions. Econometrica 75, 459–502.

[26] Song, M., 2013. Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. Manuscript, Columbia University.

[27] Yao, Y.C., 1987. Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. Annals of Statistics 15, 1321–1328.

[28] Yu, J., De Jong, R., Lee, L.F. 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. Journal of Econometrics 146, 118-134.

APPENDIX

Lemma 2 *For each i and $k < k_0$,*

$$(Z_{0i}'M_iZ_{0i}) - (Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i}) \geq R'[(X_{\Delta i}'X_{\Delta i})(X_{2i}'X_{2i})^{-1}(X_{0i}'X_{0i})]R.$$

Proof. See Bai (1997) Lemma A.1. ■

Lemma 3 *Under Assumptions 1-3, there exists $M > 0$ such that for all N and T, for each i,*

(1) $E(\left\|\frac{X_i'e_i}{\sqrt{T}}\right\|^2) \leq M,$

(2) $E(\sup_{k<k_0-C}\left\|\frac{(X_{\Delta i}'X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX}\right\|)^4 \leq \frac{M}{(C+1)^2},$

(3) $E(\sup_{k<k_0}\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1}\right\|)^2 \leq \frac{M}{T},$

(4) $E(\left\|\frac{X_{0i}'X_{0i}}{T-k_0} - \Sigma_i^{XX}\right\|)^4 \leq \frac{M}{T^2},$

(5) $E(\sup_{k<k_0}\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|^4) \leq M,$

(6) $E(\left\|(\frac{X_i'X_i}{T})^{-1}\right\|^4) \leq M,$

(7) $E(\sup_{k<k_0}\left\|\frac{e_i'M_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4) \leq M\log T,$

(8) $E(\sup_{k<k_0}\left\|\frac{e_i'M_iZ_{\Delta i}}{|k-k_0|}\right\|^4) \leq M,$

(9) $E(\left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|^4) \leq M,$

(10) $E(\sup_{k<k_0}\left\|\frac{Z_{2i}'M_ie_i}{\sqrt{T-k}}\right\|^4) \leq M,$

(11) $E(\sup_{k<k_0}\left\|\frac{Z_{\Delta i}'M_iZ_{2i}}{|k-k_0|}\right\|^4) \leq M,$

(12) $E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|^2) \leq \frac{M}{T},$

(13) $E(\left\|(\frac{Z_{0i}'M_iZ_{0i}}{T-k_0})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|^2 \leq \frac{M}{T},$

(14) $\sup_{k\in K,k\leq k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}\right\| \leq M,$

(15) $\sup_{k\in K,k\leq k_0}\left\|[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\| \leq M,$

(16) $\sup_{k\in K^c,k<k_0}\left\|(\frac{Z_{1i}'M_iZ_{1i}}{k})^{-1}\right\| \leq M,$

(17) $E(\sup_{k\in K^c,k<k_0}\left\|\frac{e_i'M_iZ_{1i}}{\sqrt{k}}\right\|^2) \leq M\log T,$

(18) $E(\sup_{k\in K^c,k<k_0}\left\|\frac{Z_{0i}'M_iZ_{1i}}{k}\right\|) \leq M.$

Proof. (1)

$$E(\left\|\frac{X_i'e_i}{\sqrt{T}}\right\|^2) = E(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T x_{it}e_{it}\right\|^2) = \sum_{j=1}^p \frac{1}{T}E[S_{ij}(1,T)]^2 \leq pM,$$

where the last inequality follows from part (3) of Assumption 3.

(2) Take $r = 4$ in part (3) of Assumption 2, we have for each $1 \leq j \leq p$, $1 \leq m \leq p$, $1 \leq i \leq N$ and $0 \leq k \leq k_0 - 1$,

$$E(\sup_{k\leq t\leq k_0-1} |R_{ijm}(t,k_0)|^4) \leq (\frac{4}{3})^4 E(|R_{ijm}(k,k_0)|^4) \leq (\frac{4}{3})^4 (k_0-k)^2 M.$$

Next, using Lemma 1 with $r = 4$, $S_l = R_{ijm}(k_0 - C - l, k_0)$, $\beta_{k_0-k} = k_0 - k$ and

$$\alpha_{k_0-k} = \begin{cases} (\frac{4}{3})^4 (k_0-k)^2 M \ for\ k_0 - k = C+1 \\ (\frac{4}{3})^4 [(k_0-k)^2 - (k_0-k-1)^2] M \ for\ C+2 \leq k_0 - k \leq T(\tau_0 - \eta) \end{cases},$$

we have for each $1 \leq j \leq p$, $1 \leq m \leq p$ and $1 \leq i \leq N$,

$$\begin{aligned} E(\sup_{k\in K(C),k<k_0} \left|\frac{1}{k_0-k} R_{ijm}(k,k_0)\right|^4) &\leq 4[\frac{(\frac{4}{3})^4 (C+1)^2 M}{(C+1)^4} + \sum\nolimits_{k_0-k=C+2}^{T\tau_0} \frac{(\frac{4}{3})^4 [2(k_0-k)-1]M}{(k_0-k)^4}] \\ &\leq 4(\frac{4}{3})^4 M[\frac{1}{(C+1)^2} + 2\sum\nolimits_{k_0-k=C+2}^{\infty} \frac{1}{(k_0-k)^3}] \\ &\leq \frac{12(\frac{4}{3})^4 M}{(C+1)^2}, \end{aligned}$$

where the last inequality follows from

$$\begin{aligned} \sum\nolimits_{i=C+2}^{\infty} \frac{1}{i^3} &< \sum\nolimits_{i=C+2}^{\infty} \frac{1}{i}(\frac{1}{i-1} - \frac{1}{i}) < \frac{1}{C+2} \sum\nolimits_{i=C+2}^{\infty} (\frac{1}{i-1} - \frac{1}{i}) \\ &= \frac{1}{(C+2)(C+1)} \leq \frac{1}{(C+1)^2}. \end{aligned}$$

Thus,

$$\begin{aligned} & E(\sup_{k\in K(C),k<k_0} \left\|\frac{1}{k_0-k} \sum\nolimits_{t=k+1}^{k_0} (x_{it}x_{it}' - \Sigma_i^{XX})\right\|^4) \\ \leq\ & p^2 \sum\nolimits_{j=1}^{p} \sum\nolimits_{m=1}^{p} E(\sup_{k\in K(C),k<k_0} \left|\frac{1}{k_0-k} R_{ijm}(k,k_0)\right|^4) \\ \leq\ & \frac{12(\frac{4}{3})^4 p^4 M}{(C+1)^2}. \end{aligned}$$

(3) Noting that $A^{-1} - B^{-1} = B^{-1}(B-A)A^{-1}$,

$$\begin{aligned} & E(\sup_{k<k_0} \left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1}\right\|)^2 \\ \leq\ & E(\left\|(\Sigma_i^{XX})^{-1}\right\|^2 \sup_{k<k_0} \left\|\frac{X_{2i}'X_{2i}}{T-k} - \Sigma_i^{XX}\right\|^2 \sup_{k<k_0} \left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}\right\|^2). \end{aligned}$$

By parts (1) and (2) of Assumption 2, the first and the third term are bounded, hence it suffices to show $E(\sup_{k<k_0} \left\|\frac{X_{2i}'X_{2i}}{T-k} - \Sigma_i^{XX}\right\|)^2 = O(\frac{1}{T})$ uniformly over i. Take $r = 2$ in part (3) of Assumption

2, we have for each $1 \leq j \leq p$, $1 \leq m \leq p$, $1 \leq i \leq N$ and $0 \leq k \leq T-1$,

$$E(\sup_{k \leq t \leq T-1} |R_{ijm}(t,T)|^2) \leq 4E(|R_{ijm}(k,T)|^2) \leq 4(T-k)M,$$

then using Lemma 1 with $r=2$, $S_l = R_{ijm}(k_0-l,T)$, $\beta_{T-k} = T-k$ and

$$\alpha_{T-k} = \begin{cases} 4(T-k_0+1)M \ for \ T-k = T-k_0+1 \\ 4M \ for \ T-k_0+2 \leq T-k \leq T \end{cases},$$

we have for each $1 \leq j \leq p$, $1 \leq m \leq p$ and $1 \leq i \leq N$,

$$\begin{aligned} E(\sup_{k<k_0} \left| \frac{1}{T-k} R_{ijm}(k,T) \right|^2) &\leq 4[\frac{4(T-k_0+1)M}{(T-k_0+1)^2} + \sum\nolimits_{T-k=T-k_0+2}^{T} \frac{4M}{(T-k)^2}] \\ &\leq 16M[\frac{1}{T-k_0+1} + \sum\nolimits_{T-k=T-k_0+2}^{T} \frac{1}{(T-k)^2}] \\ &\leq \frac{32M}{T-k_0+1}. \end{aligned}$$

Thus,

$$E(\sup_{k<k_0} \left\| \frac{1}{T-k} \sum\nolimits_{t=k+1}^{T} (x_{it}x_{it}' - \Sigma_i^{XX}) \right\|^2) \leq \sum\nolimits_{j=1}^{p} \sum\nolimits_{m=1}^{p} E(\sup_{k<k_0} \left| \frac{1}{T-k} R_{ijm}(k,T) \right|^2)$$
$$\leq \frac{32p^2M}{T-k_0+1}.$$

(4)

$$\begin{aligned} E(\left\| \frac{X_{0i}'X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|)^4 &= E\left\| \frac{1}{T-k_0} \sum\nolimits_{t=k_0+1}^{T} (x_{it}x_{it}' - \Sigma_i^{XX}) \right\|^4 \\ &= \frac{1}{(T-k_0)^4} E[\sum\nolimits_{j=1}^{p} \sum\nolimits_{m=1}^{p} R_{ijm}^2(k_0,T)]^2 \\ &\leq \frac{1}{(T-k_0)^4} E[p^2 \sum\nolimits_{j=1}^{p} \sum\nolimits_{m=1}^{p} R_{ijm}^4(k_0,T)] \\ &= \frac{p^2}{(T-k_0)^4} \sum\nolimits_{j=1}^{p} \sum\nolimits_{m=1}^{p} E[R_{ijm}^4(k_0,T)] \\ &\leq \frac{p^4M}{(T-k_0)^2}, \end{aligned}$$

where the last inequality follows from part (4) of Assumption 2.

(5)

$$\left\| \frac{Z_{\Delta i}'X_i}{|k-k_0|} \right\|^2 \leq \left\| \frac{X_{\Delta i}'X_i}{|k-k_0|} \right\|^2 = \left\| \frac{X_{\Delta i}'X_{\Delta i}}{|k-k_0|} \right\|^2 \leq 2 \left\| \frac{X_{\Delta i}'X_{\Delta i}}{|k-k_0|} - \Sigma_i^{XX} \right\|^2 + 2\left\| \Sigma_i^{XX} \right\|^2,$$

hence

$$\sup_{k<k_0} \left\| \frac{Z_{\Delta i}'X_i}{|k-k_0|} \right\|^4 \leq 8 \sup_{k<k_0} \left\| \frac{X_{\Delta i}'X_{\Delta i}}{|k-k_0|} - \Sigma_i^{XX} \right\|^4 + 8\left\| \Sigma_i^{XX} \right\|^4.$$

Take $C=0$ in part (2), the proof is finished.

(6) Under part (2) of Assumption 2, the proof is obvious.

(7)

$$\sup_{k<k_0}\left\|\frac{e_i'M_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4 \leq 8\sup_{k<k_0}\left\|\frac{e_i'Z_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4 + 8\sup_{k<k_0}\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4.$$

For the first term, take $r=4$ in part (2) of Assumption 3, we have for each $1\leq j\leq p$, $1\leq i\leq N$ and $0\leq k\leq k_0-1$,

$$E(\sup_{k\leq t\leq k_0-1}|S_{ij}(t,k_0)|^4) \leq (\frac{4}{3})^4E(|S_{ij}(k,k_0)|^4)\leq(\frac{4}{3})^4(k_0-k)^2M.$$

Using Lemma 1 with $r=4$, $S_l=S_{ij}(k_0-l,k_0)$, $\beta_{k_0-k}=\sqrt{k_0-k}$ and $\alpha_{k_0-k}=(\frac{4}{3})^4[(k_0-k)^2-(k_0-k-1)^2]M$ for $1\leq k_0-k\leq T\tau_0$, we have for each $1\leq j\leq p$ and $1\leq i\leq N$,

$$\begin{aligned} E(\sup_{k<k_0}\left|\frac{1}{\sqrt{k_0-k}}S_{ij}(k,k_0)\right|^4) &\leq 4\sum_{k_0-k=1}^{T\tau_0}\frac{(\frac{4}{3})^4[2(k_0-k)-1]M}{(k_0-k)^2}\\ &\leq 8(\frac{4}{3})^4M\sum_{k_0-k=1}^{T\tau_0}\frac{1}{k_0-k}\leq 8(\frac{4}{3})^4MO(\log T).\end{aligned}$$

Thus,

$$E(\sup_{k<k_0}\left\|\frac{e_i'Z_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4)\leq p\sum_{j=1}^{p}E(\sup_{k<k_0}\left|\frac{1}{\sqrt{k_0-k}}S_{ij}(k,k_0)\right|^4)\leq 8(\frac{4}{3})^4p^2MO(\log T).$$

For the second term,

$$\begin{aligned}\sup_{k<k_0}\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4 &\leq \left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^4\left\|(\frac{X_i'X_i}{T})^{-1}\right\|^4\sup_{k<k_0}\left\|\frac{X_i'Z_{\Delta i}}{|k-k_0|}\right\|^4\\ &\leq M^4\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^4\sup_{k<k_0}\left\|\frac{X_i'Z_{\Delta i}}{|k-k_0|}\right\|^4,\end{aligned}$$

where the last inequality follows from part (2) of Assumption 2. Hence,

$$E(\sup_{k<k_0}\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4)\leq M^4[E(\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^8)E(\sup_{k<k_0}\left\|\frac{X_i'Z_{\Delta i}}{|k-k_0|}\right\|^8)]^{\frac{1}{2}}=O(1),$$

in which $E(\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^8)=O(1)$ and $E(\sup_{k<k_0}\left\|\frac{X_i'Z_{\Delta i}}{|k-k_0|}\right\|^8)=O(1)$ can be proved following the same procedure as part (1) and part (5) respectively.

(8) The proof is similar to part (7). For the first term, the difference is $\beta_{k_0-k}=k_0-k$ and thus

$$\begin{aligned} E(\sup_{k<k_0}\left\|\frac{e_i'Z_{\Delta i}}{|k-k_0|}\right\|^4) &\leq p\sum_{j=1}^{p}E(\sup_{k<k_0}\left|\frac{1}{|k-k_0|}S_{ij}(k,k_0)\right|^4)\\ &\leq p\sum_{j=1}^{p}4\sum_{k_0-k=1}^{T\tau_0}\frac{(\frac{4}{3})^4[2(k_0-k)-1]M}{(k_0-k)^4}\\ &\leq 8(\frac{4}{3})^4p^2M\sum_{k_0-k=1}^{T\tau_0}\frac{1}{(k_0-k)^3}=O(1).\end{aligned}$$

For the second term, the difference is

$$E(\sup_{k<k_0}\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}}{|k-k_0|}\right\|^4)\le\frac{1}{T^2}M^4[E(\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^8)E(\sup_{k<k_0}\left\|\frac{X_i'Z_{\Delta i}}{|k-k_0|}\right\|^8)]^{\frac{1}{2}}=O(\frac{1}{T^2}).$$

(9)

$$E(\left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|^4)\le 8E(\left\|\frac{e_i'Z_{0i}}{\sqrt{T-k_0}}\right\|^4)+8E(\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_iZ_{0i}}{\sqrt{T-k_0}}\right\|^4).$$

Under part (3) of Assumption 3, the first term is $O(1)$. For the second term,

$$\begin{aligned}E(\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_iZ_{0i}}{\sqrt{T-k_0}}\right\|^4) &\le E(\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^4\left\|(\frac{X_i'X_i}{T})^{-1}\right\|^4\left\|\frac{X_i'Z_{0i}}{T-k_0}\right\|^4)\\ &\le M^4[E(\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^8)E(\left\|\frac{X_i'Z_{0i}}{T-k_0}\right\|^8)]^{\frac{1}{2}}=O(1).\end{aligned}$$

(10) The proof is also similar to part (7).

$$\sup_{k<k_0}\left\|\frac{e_i'M_iZ_{2i}}{\sqrt{T-k}}\right\|^4\le 8\sup_{k<k_0}\left\|\frac{e_i'Z_{2i}}{\sqrt{T-k}}\right\|^4+8\sup_{k<k_0}\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_iZ_{2i}}{\sqrt{T-k}}\right\|^4.$$

For the first term, take $r=4$ in part (2) of Assumption 3, we have for each $1\le j\le p$, $1\le i\le N$ and $0\le k\le T-1$,

$$E(\sup_{k\le t\le T-1}|S_{ij}(t,T)|^4)\le(\frac{4}{3})^4E(|S_{ij}(k,T)|^4)\le(\frac{4}{3})^4(T-k)^2M.$$

Using Lemma 1 with $r=4$, $S_l=S_{ij}(k_0-l,T)$, $\beta_{T-k}=\sqrt{T-k}$ and $\alpha_{T-k}=(\frac{4}{3})^4[(T-k)^2-(T-k-1)^2]M$ for $T-k_0+1\le T-k\le T$, we have for each $1\le j\le p$ and $1\le i\le N$,

$$\begin{aligned}E(\sup_{k<k_0}\left|\frac{1}{\sqrt{T-k}}S_{ij}(k,T)\right|^4) &\le 4\sum\nolimits_{T-k=T-k_0+1}^{T}\frac{(\frac{4}{3})^4[2(T-k)-1]M}{(T-k)^2}\\ &\le 8(\frac{4}{3})^4M\sum\nolimits_{T-k=T-k_0+1}^{T}\frac{1}{T-k}\\ &\to 8(\frac{4}{3})^4M\log\frac{1}{1-\tau_0},\end{aligned}$$

since

$$\begin{aligned}\sum\nolimits_{T-k=T-k_0+1}^{T}\frac{1}{T-k} &= \sum\nolimits_{i=1}^{T}\frac{1}{i}-\sum\nolimits_{i=1}^{T-k_0}\frac{1}{i}\\ &= (\sum\nolimits_{i=1}^{T}\frac{1}{i}-\log T)-(\sum\nolimits_{i=1}^{T-k_0}\frac{1}{i}-\log(T-k_0))+(\log T-\log(T-k_0))\\ &\to \gamma-\gamma+\log\frac{1}{1-\tau_0}=\log\frac{1}{1-\tau_0},\end{aligned}$$

where γ is Euler-Mascheroni constant. Thus,

$$E(\sup_{k<k_0}\left\|\frac{e_i'Z_{2i}}{\sqrt{T-k}}\right\|^4)\le p\sum\nolimits_{j=1}^{p}E(\sup_{k<k_0}\left|\frac{1}{\sqrt{T-k}}S_{ij}(k,T)\right|^4)\le 8(\frac{4}{3})^4p^2M\log\frac{1}{1-\tau_0}=O(1).$$

For the second term,

$$E(\sup_{k<k_0}\left\|\frac{e_i'X_i(X_i'X_i)^{-1}X_iZ_{2i}}{\sqrt{T-k}}\right\|^4) \leq E(\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^4\left\|(\frac{X_i'X_i}{T})^{-1}\right\|^4\sup_{k<k_0}\left\|\frac{X_i'Z_{2i}}{T-k}\right\|^4)$$
$$\leq M^4[E(\left\|\frac{e_i'X_i}{\sqrt{T}}\right\|^8)E(\sup_{k<k_0}\left\|\frac{X_i'Z_{2i}}{T-k}\right\|^8)]^{\frac{1}{2}} = O(1),$$

in which $E(\sup_{k<k_0}\left\|\frac{X_i'Z_{2i}}{T-k}\right\|^8) = O(1)$ can be proved following the same procedure as part (5).

(11)

$$E(\sup_{k<k_0}\left\|\frac{Z_{\Delta i}'M_iZ_{2i}}{|k-k_0|}\right\|^4) \leq 8E(\sup_{k<k_0}\left\|\frac{Z_{\Delta i}'Z_{2i}}{|k-k_0|}\right\|^4) + 8E(\sup_{k<k_0}\left\|\frac{Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'Z_{2i}}{|k-k_0|}\right\|^4).$$

The first term is $O(1)$ based on $\left\|\frac{Z_{\Delta i}'Z_{2i}}{|k-k_0|}\right\| = \left\|\frac{Z_{\Delta i}'Z_{\Delta i}}{|k-k_0|}\right\| \leq \left\|\frac{X_{\Delta i}'X_{\Delta i}}{|k-k_0|}\right\|$ and part (2). For the second term,

$$E(\sup_{k<k_0}\left\|\frac{Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'Z_{2i}}{|k-k_0|}\right\|^4) \leq E(\left\|(\frac{X_i'X_i}{T})^{-1}\right\|^4\sup_{k<k_0}\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|^4\sup_{k<k_0}\left\|\frac{X_i'Z_{2i}}{T-k}\right\|^4)$$
$$\leq M^4[E(\sup_{k<k_0}\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|^8)E(\sup_{k<k_0}\left\|\frac{X_i'Z_{2i}}{T-k}\right\|^8)]^{\frac{1}{2}} = O(1).$$

(12) Noting that $A^{-1} - B^{-1} = B^{-1}(B-A)A^{-1}$,

$$E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|)^2$$
$$\leq E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}\right\|^2\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'M_iZ_{2i}}{T-k} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]\right\|^2$$
$$\sup_{k\in K(k_0),k<k_0}\left\|[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|^2).$$

By parts (14) and (15) below, the first and the third term are both bounded. For the middle term, we have

$$E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k} - [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}])\right\|^2)$$
$$= E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'Z_{2i}}{T-k} - \Sigma_i^{ZZ}) - (\frac{Z_{2i}'X_i(X_i'X_i)^{-1}X_i'Z_{2i}}{T-k} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ})\right\|^2)$$
$$\leq 2E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'Z_{2i}}{T-k} - \Sigma_i^{ZZ}\right\|^2)$$
$$+2E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'X_i(X_i'X_i)^{-1}X_i'Z_{2i}}{T-k} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}\right\|^2).$$

The first term is $O(\frac{1}{T})$. For the second term,

$$
\begin{aligned}
& E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'X_i(X_i'X_i)^{-1}X_i'Z_{2i}}{T-k}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}\right\|^2) \\
= \ & E(\sup_{k\in K(k_0),k<k_0}(\frac{T-k}{T})^2\left\|\frac{Z_{2i}'X_i}{T-k}(\frac{X_i'X_i}{T})^{-1}\frac{X_i'Z_{2i}}{T-k}-\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}\right\|^2) \\
\leq \ & E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'X_i}{T-k}(\frac{X_i'X_i}{T})^{-1}\frac{X_i'Z_{2i}}{T-k}-\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}\right\|^2)=O(\frac{1}{T}),
\end{aligned}
$$

since $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{2i}'X_i}{T-k}-\Sigma_i^{ZX}\right\|^2)=O(\frac{1}{T})$ and $E(\left\|(\frac{X_i'X_i}{T})^{-1}-(\Sigma_i^{XX})^{-1}\right\|^2)=O(\frac{1}{T})$.

(13) Following the same procedure as part (12), the proof is straightforward.

(14)

$$
\begin{aligned}
Z_{2i}'M_iZ_{2i} &= R'[X_{2i}'X_{2i}-X_{2i}'X_i(X_i'X_i)^{-1}X_i'X_{2i}]R \\
&= R'[X_{2i}'X_{2i}-X_{2i}'X_{2i}(X_i'X_i)^{-1}X_{2i}'X_{2i}]R \\
&= R'[X_{2i}'X_{2i}-X_{2i}'X_{2i}(X_i'X_i)^{-1}(X_i'X_i-X_{1i}'X_{1i})]R \\
&= R'[X_{2i}'X_{2i}(X_i'X_i)^{-1}X_{1i}'X_{1i}]R \\
&= R'[(X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1}]^{-1}R,
\end{aligned}
$$

where the last equality follows from

$$
\begin{aligned}
[X_{2i}'X_{2i}(X_i'X_i)^{-1}X_{1i}'X_{1i}]^{-1} &= (X_{1i}'X_{1i})^{-1}(X_i'X_i)(X_{2i}'X_{2i})^{-1} \\
&= (X_{1i}'X_{1i})^{-1}(X_{1i}'X_{1i}+X_{2i}'X_{2i})(X_{2i}'X_{2i})^{-1} \\
&= (X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\rho_{\min}(Z_{2i}'M_iZ_{2i}) &= \rho_{\min}(R'[(X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1}]^{-1}R) \\
&\geqslant \rho_{\min}([(X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1}]^{-1}) \\
&= \frac{1}{\rho_{\max}((X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1})},
\end{aligned}
$$

and thus

$$
\begin{aligned}
\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}\right\| &\leq \sqrt{q}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}\right\|_{op}=\frac{\sqrt{q}(T-k)}{\rho_{\min}(Z_{2i}'M_iZ_{2i})} \\
&\leq \sqrt{q}(T-k)\rho_{\max}((X_{1i}'X_{1i})^{-1}+(X_{2i}'X_{2i})^{-1}) \\
&\leq \sqrt{q}(T-k)[\rho_{\max}((X_{1i}'X_{1i})^{-1})+\rho_{\max}((X_{2i}'X_{2i})^{-1})] \\
&= \sqrt{q}(\frac{T-k}{k}\left\|(\frac{X_{1i}'X_{1i}}{k})^{-1}\right\|+\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}\right\|).
\end{aligned}
$$

By part (2) of Assumption 2, both $\sup_{k\in K,k\le k_0}\left\|(\frac{X_{1i}'X_{1i}}{k})^{-1}\right\|$ and $\sup_{k\in K,k\le k_0}\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}\right\|$ are bounded, the proof is thus finished.

(15) First, noting that $\Sigma_i^{ZX}=(R'\Sigma_i^{XX})$, we have $\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}=(R'\Sigma_i^{XX})(\Sigma_i^{XX})^{-1}(\Sigma_i^{XX}R)=R'\Sigma_i^{XX}R=\Sigma_i^{ZZ}$. Thus,

$$\begin{aligned}
&\sup_{k\in K,k\le k_0}\left\|[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|\\
=\;&\sup_{k\in K,k\le k_0}\left\|(\frac{k}{T}\Sigma_i^{ZZ})^{-1}\right\|\le\frac{\sqrt{q}}{\tau_0-\eta}\left\|(\Sigma_i^{ZZ})^{-1}\right\|_{op}\le\frac{\sqrt{q}}{\tau_0-\eta}\frac{1}{\rho_{\min}(\Sigma_i^{ZZ})}\\
\le\;&\frac{\sqrt{q}}{\tau_0-\eta}\frac{1}{\rho_{\min}(\Sigma_i^{XX})}\le\frac{\sqrt{q}}{\tau_0-\eta}\frac{1}{\rho_1},
\end{aligned}$$

where the second inequality follows from $\rho_{\min}(\Sigma_i^{ZZ})\geqslant\rho_{\min}(\Sigma_i^{XX})$.

(16) Noting that $Z_{2i}'M_iZ_{2i}=Z_{1i}'M_iZ_{1i}$, the proof is the same as part (14), except for

$$\left\|(\frac{Z_{1i}'M_iZ_{1i}}{k})^{-1}\right\|\le\frac{\sqrt{q}k}{\rho_{\min}(Z_{1i}'M_iZ_{1i})}=\sqrt{q}(\left\|(\frac{X_{1i}'X_{1i}}{k})^{-1}\right\|+\frac{k}{T-k}\left\|(\frac{X_{2i}'X_{2i}}{T-k})^{-1}\right\|).$$

(17) The proof is similar to part (7).

(18) The proof is similar to part (11). ■

Lemma 4 *Under Assumptions 1-3 and assume* $\max_{1\le i\le N}\frac{\delta_i'\delta_i}{\lambda_N}=O(\frac{1}{N})$*, there exists* $\alpha>0$ *such that for any* $\epsilon>0$*, there exist* $N^*>0$ *and* $T^*>0$ *such that for* $N^*>N$ *and* $T>T^*$*,* $P(\inf_{k<k_0}\sum_{i=1}^N G_i(k)\ge\alpha\lambda_N)>1-\epsilon$*.*

Proof. We will prove by two steps.

Step 1: There exists $\alpha_1>0$ such that for any $\epsilon>0$, there exist $C>0$ and $T^*>0$ such that for $T>T^*$, $P(\inf_{k<k_0-C}\sum_{i=1}^N G_i(k)\ge\alpha_1\lambda_N)>1-\epsilon$.

Step 2: There exists $\alpha_2>0$ such that for any given $C>0$ and $\epsilon>0$, there exist $N^*>0$ and $T^*>0$ such that for $N>N^*$ and $T>T^*$, $P(\inf_{k_0-C\le k<k_0}\sum_{i=1}^N G_i(k)\ge\alpha_2\lambda_N)>1-\epsilon$.

Based on Step 1 and Step 2 and take $\alpha=\min\{\alpha_1,\alpha_2\}$, we have for any $\epsilon>0$, there exist $N^*>0$ and $T^*>0$ such that for $N>N^*$ and $T>T^*$, $P(\inf_{k<k_0}\sum_{i=1}^N G_i(k)<\alpha\lambda_N)\le P(\inf_{k<k_0-C}\sum_{i=1}^N G_i(k)<\alpha\lambda_N)+P(\inf_{k_0-C\le k<k_0}\sum_{i=1}^N G_i(k)<\alpha\lambda_N)\le 2\epsilon$, thus $P(\inf_{k<k_0}\sum_{i=1}^N G_i(k)<\alpha\lambda_N)>1-2\epsilon$.

Proof of Step 1: Define $A_i(k) = \frac{(X'_{\Delta i}X_{\Delta i})(X'_{2i}X_{2i})^{-1}(X'_{0i}X_{0i})}{|k_0-k|}$, then by Lemma 2 we have

$$\begin{aligned}
&\inf_{k<k_0-C}\sum_{i=1}^{N} G_i(k) \\
= &\inf_{k<k_0-C}\sum_{i=1}^{N} \frac{\delta_i'[(Z'_{0i}M_iZ_{0i})-(Z'_{0i}M_iZ_{2i})(Z'_{2i}M_iZ_{2i})^{-1}(Z'_{2i}M_iZ_{0i})]\delta_i}{|k_0-k|} \\
\geq &\inf_{k<k_0-C}\sum_{i=1}^{N} \frac{\delta_i'R'[(X'_{\Delta i}X_{\Delta i})(X'_{2i}X_{2i})^{-1}(X'_{0i}X_{0i})]R\delta_i}{|k_0-k|} \\
= &\inf_{k<k_0-C}\sum_{i=1}^{N} \delta_i'R'A_i(k)R\delta_i \\
\geq &\inf_{k<k_0-C}\sum_{i=1}^{N} \delta_i'R'(\frac{T-k_0}{T-k}\Sigma_i^{XX})R\delta_i - \sup_{k<k_0-C}\left|\sum_{i=1}^{N} \delta_i'R'(A_i(k)-\frac{T-k_0}{T-k}\Sigma_i^{XX})R\delta_i\right| \\
\geq &\sum_{i=1}^{N} \delta_i'R'\Sigma_i^{XX}R\delta_i - \sup_{k<k_0-C}\left|\sum_{i=1}^{N} \delta_i'R'(A_i(k)-\frac{T-k_0}{T-k}\Sigma_i^{XX})R\delta_i\right|.
\end{aligned}$$

By Assumption 2,

$$\sum_{i=1}^{N} \delta_i'R'\Sigma_i^{XX}R\delta_i \geq \sum_{i=1}^{N} \rho_{\min}(\Sigma_i^{XX})\delta_i'\delta_i \geq \rho\lambda_N,$$

thus it suffices to show for any $\epsilon > 0$ and $\eta > 0$, there exists $C > 0$ and $T^* > 0$ such that for $T > T^*$, $P(\sup_{k<k_0-C}\left|\sum_{i=1}^{N} \delta_i'R'(A_i(k)-\frac{T-k_0}{T-k}\Sigma_i^{XX})R\delta_i\right| > \eta\lambda_N) < \epsilon$. With assumption $\lim_{N\to\infty}\max_{1\leq i\leq N} \frac{N\delta_i'\delta_i}{\lambda_N} < \infty$,

$$\begin{aligned}
&\sup_{k<k_0-C}\left|\sum_{i=1}^{N} \delta_i'R'(A_i(k)-\frac{T-k_0}{T-k}\Sigma_i^{XX})R\delta_i\right| \\
\leq &\lambda_N \sup_{k<k_0-C}\frac{1}{N}\sum_{i=1}^{N}\frac{N\delta_i'\delta_i}{\lambda_N}\left\|A_i(k)-\frac{T-k_0}{T-k}\Sigma_i^{XX}\right\| \\
\leq &\lambda_N(\lim_{N\to\infty}\max_{1\leq i\leq N}\frac{N\delta_i'\delta_i}{\lambda_N})\sup_{k<k_0-C}\frac{1}{N}\sum_{i=1}^{N}\left\|A_i(k)-\frac{T-k_0}{T-k}\Sigma_i^{XX}\right\| \\
\leq &\lambda_N(\lim_{N\to\infty}\max_{1\leq i\leq N}\frac{N\delta_i'\delta_i}{\lambda_N})\frac{1}{N}\sum_{i=1}^{N}\sup_{k<k_0-C}\left\|A_i(k)-\frac{T-k_0}{T-k}\Sigma_i^{XX}\right\|,
\end{aligned}$$

thus by Markov inequality it suffices to show for any $\epsilon > 0$, there exist $C < \infty$ and $T^* > 0$ such that for $T > T^*$, $E(\sup_{k<k_0-C}\left\|A_i(k)-\frac{T-k_0}{T-k}\Sigma_i^{XX}\right\|) < \epsilon$ for all i. For each i and any given $C > 0$,

$$\begin{aligned}
&\sup_{k<k_0-C}\left\|A_i(k)-\frac{T-k_0}{T-k}\Sigma_i^{XX}\right\| \\
= &\sup_{k<k_0-C}\left\|\frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|}\frac{T-k_0}{T-k}(\frac{X'_{2i}X_{2i}}{T-k})^{-1}(\frac{X'_{0i}X_{0i}}{T-k_0})-\frac{T-k_0}{T-k}\Sigma_i^{XX}\right\| \\
\leq &\sup_{k<k_0-C}\left\|\frac{(X'_{\Delta i}X_{\Delta i})}{|k_0-k|}(\frac{X'_{2i}X_{2i}}{T-k})^{-1}(\frac{X'_{0i}X_{0i}}{T-k_0})-\Sigma_i^{XX}\right\| \\
= &\sup_{k<k_0-C}\left\|\begin{array}{c}(\frac{X'_{\Delta i}X_{\Delta i}}{|k_0-k|}-\Sigma_i^{XX}+\Sigma_i^{XX})[(\frac{X'_{2i}X_{2i}}{T-k})^{-1}-(\Sigma_i^{XX})^{-1} \\ +(\Sigma_i^{XX})^{-1}](\frac{X'_{0i}X_{0i}}{T-k_0}-\Sigma_i^{XX}+\Sigma_i^{XX})-\Sigma_i^{XX}\end{array}\right\| \\
\leq &I+II+III+IV+V+VI+VII.
\end{aligned}$$

Consider each term one by one. By part (2) and part (3) of Lemma 3, as $C \to \infty$ and $T \to \infty$, for all i

$$\begin{aligned} E(I) &= E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i} X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\| \sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\| \left\| \Sigma_i^{XX} \right\|) \\ &\leq [E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i} X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\|)^2 E(\sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\|)^2]^{\frac{1}{2}} \left\| \Sigma_i^{XX} \right\| \\ &< \epsilon. \end{aligned}$$

By part (2), part (3) and part (4) of Lemma 3, as $C \to \infty$ and $T \to \infty$, for all i,

$$\begin{aligned} & E(II) = E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i} X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\| \sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\| \left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|) \\ \leq\ & [E(\sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\|)^2]^{\frac{1}{2}} [E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i} X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\|)^4 E(\left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|)^4]^{\frac{1}{4}} \\ <\ & \epsilon. \end{aligned}$$

By part (2) of Lemma 3, as $C \to \infty$, for all i,

$$E(III) = E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i} X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\|) \left\| (\Sigma_i^{XX})^{-1} \right\| \left\| \Sigma_i^{XX} \right\| < \epsilon.$$

By part (2) and part (4) of Lemma 3, as $C \to \infty$ and $T \to \infty$, for all i,

$$\begin{aligned} E(IV) &= E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i} X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\| \left\| (\Sigma_i^{XX})^{-1} \right\| \left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|) \\ &\leq [E(\sup_{k<k_0-C} \left\| \frac{(X'_{\Delta i} X_{\Delta i})}{|k_0-k|} - \Sigma_i^{XX} \right\|)^2 E(\left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|)^2]^{\frac{1}{2}} \left\| (\Sigma_i^{XX})^{-1} \right\| \\ &< \epsilon. \end{aligned}$$

By part (3) of Lemma 3, as $T \to \infty$, for all i,

$$E(V) = E(\sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\|) \left\| \Sigma_i^{XX} \right\|^2 < \epsilon.$$

By part (3) and part (4) of Lemma 3, as $T \to \infty$, for all i,

$$\begin{aligned} E(VI) &= E(\sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\| \left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\| \left\| \Sigma_i^{XX} \right\|) \\ &\leq [E(\sup_{k<k_0-C} \left\| (\frac{X'_{2i}X_{2i}}{T-k})^{-1} - (\Sigma_i^{XX})^{-1} \right\|)^2 E(\left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|)^2]^{\frac{1}{2}} \left\| \Sigma_i^{XX} \right\| \\ &< \epsilon. \end{aligned}$$

By part (4) of Lemma 3, as $T \to \infty$, for all i,

$$E(VII) = E(\left\| \frac{X'_{0i}X_{0i}}{T-k_0} - \Sigma_i^{XX} \right\|) < \epsilon.$$

Proof of Step 2: There exists $\alpha_2 > 0$ such that for any given $C > 0$ and $\epsilon > 0$, there exist $N^* > 0$ and $T^* > 0$ such that for $N > N^*$ and $T > T^*$, $P(\inf_{k_0-C\leq k<k_0} \sum_{i=1}^N G_i(k) \geq \alpha_2\lambda_N) > 1-\epsilon$.

$$\begin{aligned}
\sum_{i=1}^N G_i(k) &= \frac{\sum_{i=1}^N \delta_i'[(Z_{0i}'M_iZ_{0i}) - (Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})]\delta_i}{|k_0-k|} \\
&= \frac{\sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{\Delta i})\delta_i - \sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{\Delta i})\delta_i}{|k_0-k|} \\
&= \frac{\sum_{i=1}^N \delta_i'Z_{\Delta i}'Z_{\Delta i}\delta_i}{|k_0-k|} - \frac{\sum_{i=1}^N \delta_i'Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}\delta_i}{|k_0-k|} \\
&\quad - \frac{\sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{\Delta i})\delta_i}{|k_0-k|},
\end{aligned}$$

thus

$$\begin{aligned}
\inf_{k_0-C\leq k<k_0} \sum_{i=1}^N G_i(k) &\geq \inf_{k_0-C\leq k<k_0} \frac{\sum_{i=1}^N \delta_i'Z_{\Delta i}'Z_{\Delta i}\delta_i}{|k_0-k|} - \sup_{k_0-C\leq k<k_0} \frac{\sum_{i=1}^N \delta_i'Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}\delta_i}{|k_0-k|} \\
&\quad - \sup_{k_0-C\leq k<k_0} \frac{\sum_{i=1}^N \delta_i'Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{\Delta i})\delta_i}{|k_0-k|}.
\end{aligned}$$

Consider the first term. By part (5) of Assumption 2, we have for each t, $\frac{1}{\lambda_N}\sum_{i=1}^N \delta_i'z_{it}z_{it}'\delta_i \xrightarrow{p} \xi$ as $N\to\infty$. For given C, $\{\frac{\frac{1}{\lambda_N}\sum_{i=1}^N \delta_i'Z_{\Delta i}'Z_{\Delta i}\delta_i}{|k_0-k|}, k_0-C\leq k<k_0\}$ is finite dimensional random vector, hence $\{\frac{\frac{1}{\lambda_N}\sum_{i=1}^N \delta_i'Z_{\Delta i}'Z_{\Delta i}\delta_i}{|k_0-k|}, k_0-C\leq k<k_0\} \xrightarrow{p} (\xi,...,\xi)'$ as $N\to\infty$. It follows by continuous mapping theorem that $\inf_{k_0-C\leq k<k_0}\frac{\frac{1}{\lambda_N}\sum_{i=1}^N \delta_i'Z_{\Delta i}'Z_{\Delta i}\delta_i}{|k_0-k|} \xrightarrow{p} \xi$ as $N\to\infty$. Next consider the last two terms.

$$\begin{aligned}
&E(\sup_{k_0-C\leq k<k_0} \frac{\sum_{i=1}^N \delta_i'Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}\delta_i}{|k_0-k|}) \\
\leq\ & E(\frac{|k_0-k|}{T}\sup_{k_0-C\leq k<k_0}\sum_{i=1}^N \left\|\frac{Z_{\Delta i}'X_i}{|k_0-k|}\right\|^2 \left\|(\frac{X_i'X_i}{T})^{-1}\right\| \|\delta_i\|^2) \\
\leq\ & \frac{C\lambda_N}{T}(\lim_{N\to\infty}\max_{1\leq i\leq N}\frac{N\delta_i'\delta_i}{\lambda_N})\frac{1}{N}\sum_{i=1}^N E(\sup_{k_0-C\leq k<k_0}\left\|\frac{Z_{\Delta i}'X_i}{|k_0-k|}\right\|^2 \left\|(\frac{X_i'X_i}{T})^{-1}\right\|) \\
=\ & O(\frac{\lambda_N}{T}),
\end{aligned}$$

where the last equality follows from part (2) of Assumption 2 and part (5) of Lemma 3. Similarly,

$$\begin{aligned}
&E(\sup_{k_0-C\leq k<k_0} \frac{\sum_{i=1}^N \delta_i'Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{\Delta i})\delta_i}{|k_0-k|}) \\
\leq\ & \frac{C\lambda_N}{T-k_0}(\lim_{N\to\infty}\max_{1\leq i\leq N}\frac{N\delta_i'\delta_i}{\lambda_N})\frac{1}{N}\sum_{i=1}^N E(\sup_{k_0-C\leq k<k_0}\left\|\frac{Z_{\Delta i}'M_iZ_{2i}}{|k_0-k|}\right\|^2 \left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}\right\|) \\
=\ & O(\frac{\lambda_N}{T}),
\end{aligned}$$

where the last equality follows from parts (11) and (14) of Lemma 3. Taking together, the proof is finished. ■

Lemma 5 *Under Assumptions 1-3,*

(1) $\sup_{k\in K(k_0),k<k_0}|A| = \sup_{k\in K(k_0),k<k_0}\left|2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i' Z_{\Delta i}' e_i\right| = O_p(\sqrt{\lambda_N})$ *as* $(N,T)\to\infty$;

(2) $\sup_{k\in K(k_0),k<k_0}|B| = \sup_{k\in K(k_0),k<k_0}\left|-2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i' Z_{\Delta i}' X_i(X_i'X_i)^{-1}X_i'e_i\right| = O_p(\frac{\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N})$ *as* $(N,T)\to\infty$;

(3) $\sup_{k\in K(k_0),k<k_0}|C| = \sup_{k\in K(k_0),k<k_0}\left|-2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i)\right| = O_p(\frac{\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N})$ *as* $(N,T)\to\infty$;

(4) $\sup_{k\in K(k_0),k<k_0}|D| = \sup_{k\in K(k_0),k<k_0}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{\Delta i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{\Delta i}'M_ie_i\right| = O_p(\frac{N\log T}{T})$ *as* $(N,T)\to\infty$;

(5) $\sup_{k\in K(k_0),k<k_0}|E| = \sup_{k\in K(k_0),k<k_0}\left|2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{0i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{\Delta i}'M_ie_i\right| = O_p(\frac{N}{\sqrt{T}})$ *as* $(N,T)\to\infty$;

(6) $\sup_{k\in K(k_0),k<k_0}|F| = \sup_{k\in K(k_0),k<k_0}\left|\frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{0i}[(Z_{2i}'M_iZ_{2i})^{-1}-(Z_{0i}'M_iZ_{0i})^{-1}]Z_{0i}'M_ie_i\right| = O_p(\frac{N}{\sqrt{T}})$ *as* $(N,T)\to\infty$.

Proof. (1) Under part (4) of Assumption 3, there exists $M>0$ such that $E(\sup_{k\le l<k_0}\left|\sum_{t=l+1}^{k_0}\eta_{Nt}\right|^2)\le 4E(\left|\sum_{t=k+1}^{k_0}\eta_{Nt}\right|^2)\le (k_0-k)M$ for all N and $1\le k<k_0$. Using Lemma 1 and take $r=2$, $\alpha_{k_0-k}=M$, $\beta_{k_0-k}=k_0-k$ for $k_0-k=1,...,T\eta$, we have

$$\begin{aligned}
E(\sup_{k\in K(k_0),k<k_0}|A|)^2 &= 4\lambda_N E(\sup_{k\in K(k_0),k<k_0}\left|\frac{1}{|k-k_0|}\sum_{t=k+1}^{k_0}\eta_{Nt}\right|^2) \\
&\le 16\lambda_N M\sum_{k=T(\tau_0-\eta)}^{k_0-1}\frac{1}{(k_0-k)^2}\le 32\lambda_N M.
\end{aligned}$$

(2)

$$\begin{aligned}
&\sup_{k\in K(k_0),k<k_0}|B| \\
=\ &\sup_{k\in K(k_0),k<k_0}\left|\frac{2}{|k-k_0|}\sum_{i=1}^N \delta_i'Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'e_i\right| \\
\le\ &\frac{2\sqrt{\lambda_N}}{\sqrt{T}}\sup_{k\in K(k_0),k<k_0}\sum_{i=1}^N\left\|\frac{\delta_i}{\sqrt{\lambda_N}}\right\|\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|\left\|(\frac{X_i'X_i}{T})^{-1}\right\|\left\|\frac{X_i'e_i}{\sqrt{T}}\right\| \\
\le\ &\frac{2\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N}(\lim_{N\to\infty}\max_{1\le i\le N}\left\|\frac{\sqrt{N}\delta_i}{\sqrt{\lambda_N}}\right\|)\frac{1}{N}\sum_{i=1}^N(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}'X_i}{|k-k_0|}\right\|)\left\|(\frac{X_i'X_i}{T})^{-1}\right\|\left\|\frac{X_i'e_i}{\sqrt{T}}\right\|.
\end{aligned}$$

Using parts (1), (5) and (6) of Lemma 3,

$$\begin{aligned}
& \frac{1}{N}\sum\nolimits_{i=1}^{N} E((\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{\Delta i}X_i}{|k-k_0|}\right\|)\left\|(\frac{X'_iX_i}{T})^{-1}\right\|\left\|\frac{X'_ie_i}{\sqrt{T}}\right\|) \\
\leq\ & \frac{1}{N}\sum\nolimits_{i=1}^{N}[E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{\Delta i}X_i}{|k-k_0|}\right\|^4)E(\left\|(\frac{X'_iX_i}{T})^{-1}\right\|^4)]^{\frac{1}{4}}[E(\left\|\frac{X'_ie_i}{\sqrt{T}}\right\|^2)]^{\frac{1}{2}} \\
=\ & O(1),
\end{aligned}$$

hence by Markov inequality $\sup\limits_{k\in K(k_0),k<k_0}|B| = O_p(\frac{\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N})$ as $(N,T)\to\infty$.

(3)

$$\begin{aligned}
& \sup_{k\in K(k_0),k<k_0}|C| \\
=\ & \sup_{k\in K(k_0),k<k_0}\left|\frac{2}{\sqrt{T-k}}\sum\nolimits_{i=1}^{N}\delta'_i(\frac{Z'_{\Delta i}M_iZ_{2i}}{|k-k_0|})(\frac{Z'_{2i}M_iZ_{2i}}{T-k})^{-1}(\frac{Z'_{2i}M_ie_i}{\sqrt{T-k}})\right| \\
\leq\ & \frac{2\sqrt{\lambda_N}}{\sqrt{T\tau_0}}\sum\nolimits_{i=1}^{N}\left\|\frac{\delta_i}{\sqrt{\lambda_N}}\right\|\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{\Delta i}M_iZ_{2i}}{|k-k_0|}\right\|\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{2i}M_ie_i}{\sqrt{T-k}}\right\| \\
& \sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z'_{2i}M_iZ_{2i}}{T-k})^{-1}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\| \\
& +\frac{2\sqrt{\lambda_N}}{\sqrt{T\eta}}\sum\nolimits_{i=1}^{N}\left\|\frac{\delta_i}{\sqrt{\lambda_N}}\right\|\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{\Delta i}M_iZ_{2i}}{|k-k_0|}\right\|\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{2i}M_ie_i}{\sqrt{T-k}}\right\| \\
& \left\|[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\| \\
\leq\ & \frac{2\sqrt{N}\sqrt{\lambda_N}}{\sqrt{T\tau_0}}(\lim_{N\to\infty}\max_{1\leq i\leq N}\left\|\frac{\sqrt{N}\delta_i}{\sqrt{\lambda_N}}\right\|)[\frac{1}{N}\sum\nolimits_{i=1}^{N}\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{\Delta i}M_iZ_{2i}}{|k-k_0|}\right\|\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{2i}M_ie_i}{\sqrt{T-k}}\right\| \\
& \sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z'_{2i}M_iZ_{2i}}{T-k})^{-1}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|] \\
& +\frac{2\sqrt{N}\sqrt{\lambda_N}}{\sqrt{T\tau_0}}(\lim_{N\to\infty}\max_{1\leq i\leq N}\left\|\frac{\sqrt{N}\delta_i}{\sqrt{\lambda_N}}\right\|)[\frac{1}{N}\sum\nolimits_{i=1}^{N}\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{\Delta i}M_iZ_{2i}}{|k-k_0|}\right\| \\
& \sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{2i}M_ie_i}{\sqrt{T-k}}\right\|\left\|[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|] \\
=\ & \frac{2\sqrt{N}\sqrt{\lambda_N}}{\sqrt{T\tau_0}}(\lim_{N\to\infty}\max_{1\leq i\leq N}\left\|\frac{\sqrt{N}\delta_i}{\sqrt{\lambda_N}}\right\|)(C_1+C_2)
\end{aligned}$$

Using parts (10), (11) and (12) of Lemma 3,

$$
\begin{aligned}
E(C_1) &\leq \frac{1}{N}\sum_{i=1}^{N}[E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z'_{\Delta i}M_iZ_{2i}}{|k-k_0|}\right\|^4)E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{2i}M_ie_i}{\sqrt{T-k}}\right\|^4)]^{\frac{1}{4}} \\
&\quad [E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z'_{2i}M_iZ_{2i}}{T-k})^{-1}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|^2)]^{\frac{1}{2}} \\
&= O(\frac{1}{\sqrt{T}}),
\end{aligned}
$$

and using parts (10), (11) and (15) of Lemma 3,

$$
\begin{aligned}
E(C_2) &\leq \frac{1}{N}\sum_{i=1}^{N}[E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z'_{\Delta i}M_iZ_{2i}}{|k-k_0|}\right\|^2)E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z'_{2i}M_ie_i}{\sqrt{T-k}}\right\|^2)]^{\frac{1}{2}} \\
&\quad \sup_{k\in K(k_0),k<k_0}\left\|[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\| \\
&= O(1),
\end{aligned}
$$

thus by Markov inequality $\sup_{k\in K(k_0),k<k_0}|C| = O_p(\frac{\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N})$ as $(N,T)\to\infty$.

(4)

$$
\begin{aligned}
&\sup_{k\in K(k_0),k<k_0}|D| \\
= &\sup_{k\in K(k_0),k<k_0}\left|\frac{N}{T-k}\frac{1}{N}\sum_{i=1}^{N}\frac{e'_iM_iZ_{\Delta i}}{\sqrt{|k-k_0|}}(\frac{Z'_{2i}M_iZ_{2i}}{T-k})^{-1}\frac{Z'_{\Delta i}M_ie_i}{\sqrt{|k-k_0|}}\right| \\
\leq &\frac{N}{T\tau_0}(\sup_{k\in K(k_0),k<k_0}\left|\frac{1}{N}\sum_{i=1}^{N}\frac{e'_iM_iZ_{\Delta i}}{\sqrt{|k-k_0|}}[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\frac{Z'_{\Delta i}M_ie_i}{\sqrt{|k-k_0|}}\right| \\
&+\sup_{k\in K(k_0),k<k_0}\left|\frac{1}{N}\sum_{i=1}^{N}\frac{e'_iM_iZ_{\Delta i}}{\sqrt{|k-k_0|}}[(\frac{Z'_{2i}M_iZ_{2i}}{T-k})^{-1}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}]\frac{Z'_{\Delta i}M_ie_i}{\sqrt{|k-k_0|}}\right|) \\
\leq &\frac{N}{T\tau_0}(\frac{1}{N}\sum_{i=1}^{N}\sup_{k\in K(k_0),k<k_0}\left\|\frac{e'_iM_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^2\sup_{K(k_0)}\left\|[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\| \\
&+\frac{1}{N}\sum_{i=1}^{N}\sup_{k\in K(k_0),k<k_0}\left\|\frac{e'_iM_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^2 \\
&\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z'_{2i}M_iZ_{2i}}{T-k})^{-1}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|) \\
= &\frac{N}{T\tau_0}(D_1+D_2),
\end{aligned}
$$

Using parts (7) and (15) of Lemma 3,

$$
\begin{aligned}
E(D_1) &= \frac{1}{N}\sum_{i=1}^{N}E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{e'_iM_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|)^2\sup_{k\in K(k_0),k<k_0}\left\|[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\| \\
&= O(\log T),
\end{aligned}
$$

and using parts (7) and (12) of Lemma 3,

$$\begin{aligned}
E(D_2) &\leq \frac{1}{N}\sum\nolimits_{i=1}^{N}[E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{e_i'M_iZ_{\Delta i}}{\sqrt{|k-k_0|}}\right\|^4) \\
&\quad E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|^2)]^{\frac{1}{2}} \\
&= O(\sqrt{\frac{\log T}{T}}),
\end{aligned}$$

thus by Markov inequality $\sup_{k\in K(k_0),k<k_0}|D| = O_p(\frac{N\log T}{T})$ as $(N,T)\to\infty$.

(5)

$$\begin{aligned}
&\sup_{k\in K(k_0),k<k_0}|E| \\
\leq\ & \sup_{k\in K(k_0),k<k_0}\left|\frac{2N\sqrt{T-k_0}}{T-k}\frac{1}{N}\sum\nolimits_{i=1}^{N}\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\frac{Z_{\Delta i}'M_ie_i}{|k-k_0|}\right| \\
&+\sup_{k\in K(k_0),k<k_0}\left|\begin{array}{c}\frac{2N\sqrt{T-k_0}}{T-k}\frac{1}{N}\sum_{i=1}^{N}\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}[(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1} \\ -[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}]\frac{Z_{\Delta i}'M_ie_i}{|k-k_0|}\end{array}\right| \\
\leq\ & \frac{2}{\sqrt{1-\tau_0}}\frac{N}{\sqrt{T}}(\frac{1}{N}\sum\nolimits_{i=1}^{N}\left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}'M_ie_i}{|k-k_0|}\right\| \\
&\sup_{k\in K(k_0),k<k_0}\left\|[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|) \\
&+\frac{2}{\sqrt{1-\tau_0}}\frac{N}{\sqrt{T}}(\frac{1}{N}\sum\nolimits_{i=1}^{N}\left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}'M_ie_i}{|k-k_0|}\right\| \\
&\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|) \\
=\ & \frac{2}{\sqrt{1-\tau_0}}\frac{N}{\sqrt{T}}(E_1+E_2).
\end{aligned}$$

Using parts (8), (9) and (15) of Lemma 3,

$$\begin{aligned}
E(E_1) &\leq \frac{1}{N}\sum\nolimits_{i=1}^{N}[E(\left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|^2)E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}'M_ie_i}{|k-k_0|}\right\|^2)]^{\frac{1}{2}} \\
&\quad \sup_{k\in K(k_0),k<k_0}\left\|[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\| \\
&= O(1),
\end{aligned}$$

and using parts (8), (9) and (12) of Lemma 3,

$$\begin{aligned}
E(E_2) &\leq \frac{1}{N}\sum\nolimits_{i=1}^{N}[E(\left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|^4)E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{Z_{\Delta i}'M_ie_i}{|k-k_0|}\right\|^4)]^{\frac{1}{4}} \\
&\quad [E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|^2)]^{\frac{1}{2}} \\
&= O(\frac{1}{\sqrt{T}}),
\end{aligned}$$

thus by Markov inequality $\sup_{k\in K(k_0),k<k_0}|E| = O_p(\frac{N}{\sqrt{T}})$ as $(N,T)\to\infty$.

(6)

$$\begin{aligned}
&\sup_{k\in K(k_0),k<k_0}|F| \\
&= \sup_{k\in K(k_0),k<k_0}\left|\frac{1}{|k-k_0|}\sum\nolimits_{i=1}^{N}e_i'M_iZ_{0i}[\frac{1}{T-k}(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}-\frac{1}{T-k_0}(\frac{Z_{0i}'M_iZ_{0i}}{T-k_0})^{-1}]Z_{0i}'M_ie_i\right| \\
&\leq \sup_{k\in K(k_0),k<k_0}\left|\begin{array}{c}\frac{1}{|k-k_0|}\sum_{i=1}^{N}e_i'M_iZ_{0i}[\frac{1}{T-k}(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1} \\ -\frac{1}{T-k}[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}]Z_{0i}'M_ie_i\end{array}\right| \\
&\quad +\sup_{k\in K(k_0),k<k_0}\left|\begin{array}{c}\frac{1}{|k-k_0|}\sum_{i=1}^{N}e_i'M_iZ_{0i}[\frac{1}{T-k}[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \\ -\frac{1}{T-k_0}[\Sigma_i^{ZZ}-\frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}]Z_{0i}'M_ie_i\end{array}\right| \\
&\quad +\left|\begin{array}{c}\frac{1}{|k-k_0|}\sum_{i=1}^{N}e_i'M_iZ_{0i}[(\frac{1}{T-k_0}(\frac{Z_{0i}'M_iZ_{0i}}{T-k_0})^{-1} \\ -\frac{1}{T-k_0}[\Sigma_i^{ZZ}-\frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}]Z_{0i}'M_ie_i\end{array}\right| \\
&= \sup_{k\in K(k_0),k<k_0}|F_1| + \sup_{k\in K(k_0),k<k_0}|F_2| + |F_3|.
\end{aligned}$$

$$\begin{aligned}
&\sup_{k\in K(k_0),k<k_0}|F_1| \\
&= \sup_{k\in K(k_0),k<k_0}\left|\begin{array}{c}\frac{1}{|k-k_0|}\frac{T-k_0}{T-k}\sum_{i=1}^{N}\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}[(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1} \\ -[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}]\frac{Z_{0i}'M_ie_i}{\sqrt{T-k_0}}\end{array}\right| \\
&\leq \sum\nolimits_{i=1}^{N}\left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|^2\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|.
\end{aligned}$$

Using parts (9) and (12) of Lemma 3,

$$\begin{aligned}
&E(\sup_{k\in K(k_0),k<k_0}|F_1|) \\
&\leq \sum\nolimits_{i=1}^{N}[E(\left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|^4)E(\sup_{k\in K(k_0),k<k_0}\left\|(\frac{Z_{2i}'M_iZ_{2i}}{T-k})^{-1}-[\Sigma_i^{ZZ}-\frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|^2)]^{\frac{1}{2}} \\
&= O(\frac{N}{\sqrt{T}}),
\end{aligned}$$

hence by Markov inequality $\sup\limits_{k\in K(k_0),k<k_0} |F_1| = O_p(\frac{N}{\sqrt{T}})$ as $(N,T)\to\infty$.

$$\begin{aligned}\sup_{k\in K(k_0),k<k_0} |F_2| &\leq \sup_{k\in K(k_0),k<k_0} \left| \begin{array}{c} \frac{1}{|k-k_0|}\sum_{i=1}^N e_i' M_i Z_{0i}\{\frac{1}{T-k}[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \\ -\frac{1}{T-k_0}[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\} Z_{0i}' M_i e_i \end{array}\right| \\ &\quad + \sup_{k\in K(k_0),k<k_0} \left| \begin{array}{c} \frac{1}{|k-k_0|}\sum_{i=1}^N e_i' M_i Z_{0i}\{\frac{1}{T-k_0}[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \\ -\frac{1}{T-k_0}[\Sigma_i^{ZZ} - \frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\} Z_{0i}' M_i e_i \end{array}\right| \\ &= \sup_{k\in K(k_0),k<k_0} |F_{21}| + \sup_{k\in K(k_0),k<k_0} |F_{22}|.\end{aligned}$$

$$\begin{aligned}& E(\sup_{k\in K(k_0),k<k_0} |F_{21}|) \\ = \; & E(\sup_{k\in K(k_0),k<k_0} \left|\frac{1}{T-k}\sum_{i=1}^N \frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\frac{Z_{0i}' M_i e_i}{\sqrt{T-k_0}}\right|) \\ \leq \; & \frac{N}{T\tau_0}[\frac{1}{N}\sum_{i=1}^N E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}\right\|^2)\sup_{k\in K(k_0),k<k_0}\left\|[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|].\end{aligned}$$

Using parts (9) and (15) of Lemma 3 and Markov inequality, $\sup\limits_{k\in K(k_0),k<k_0} |F_{21}| = O_p(\frac{N}{T})$ as $(N,T)\to\infty$.

$$\begin{aligned}& E(\sup_{k\in K(k_0),k<k_0} |F_{22}|) \\ \leq \; & \sum_{i=1}^N E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}\right\|^2)\sup_{k\in K(k_0),k<k_0}\frac{1}{|k-k_0|}\left\| \begin{array}{c} [\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \\ -[\Sigma_i^{ZZ} - \frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1} \end{array}\right\|.\end{aligned}$$

Using part (9) of Lemma 3, the first term is $O(1)$. Noting that $A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}$, the second term is not larger than

$$\begin{aligned}&\sup_{k\in K(k_0),k<k_0}\left\|[\Sigma_i^{ZZ} - \frac{T-k}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|\left\|\frac{1}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}\right\| \\ &\left\|[\Sigma_i^{ZZ} - \frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|.\end{aligned}$$

Part (15) of Lemma 3 implies this term is $O(\frac{1}{T})$, thus by Markov inequality $\sup\limits_{k\in K(k_0),k<k_0} |F_{22}| = O_p(\frac{N}{T})$ as $(N,T)\to\infty$.

$$|F_3| \leq \sum_{i=1}^N \left\|\frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}\right\|^2 \left\|(\frac{Z_{0i}' M_i Z_{0i}}{T-k_0})^{-1} - [\Sigma_i^{ZZ} - \frac{T-k_0}{T}\Sigma_i^{ZX}(\Sigma_i^{XX})^{-1}\Sigma_i^{XZ}]^{-1}\right\|)$$

Using parts (9) and (13) of Lemma 3,

$$\begin{aligned}E(F_3) &\leq \sum_{i=1}^N [E(\left\|\frac{e_i' M_i Z_{0i}}{\sqrt{T-k_0}}\right\|^4)E(\left\|(\frac{Z_{0i}' M_i Z_{0i}}{T-k_0})^{-1} - (\Sigma_i^{ZX}\Sigma_i^{XX}\Sigma_i^{XZ})^{-1}\right\|^2)]^{\frac{1}{2}} \\ &= O(\frac{N}{\sqrt{T}}),\end{aligned}$$

thus by Markov inequality $F_3 = O_p(\frac{N}{\sqrt{T}})$ as $(N,T) \to \infty$. Taking together, $\sup_{k \in K(k_0), k<k_0} |F| = O_p(\frac{N}{\sqrt{T}}) + O_p(\frac{N}{T}) + O_p(\frac{N}{\sqrt{T}}) = O_p(\frac{N}{\sqrt{T}})$ as $(N,T) \to \infty$. ■

Lemma 6 *Under Assumptions 5-14, given $|k - k_0| \le C$, there exists $M > 0$ such that*

(1) $E(\left\| \frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t - H'F_t^0)(\tilde{F}_t - H'F_t^0)' \right\|) \le \frac{1}{\delta_{NT}^2}M$,

(2) $E(\left\| \frac{1}{T}\sum_{t=1}^{T}(\tilde{F}_t - H'F_t^0)e_{it} \right\|) \le (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2})M$ *for each* i,

(3) $E(\left\| \frac{1}{T}\sum_{t=1}^{T} H'F_t^0(\tilde{F}_t - H'F_t^0)' \right\|) \le (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2})M$,

(4) $E(\sup_{k \in K(k_0), k<k_0} \left\| \frac{1}{k_0-k}\sum_{t=k+1}^{k_0}(\tilde{F}_t - H'F_t^0)(\tilde{F}_t - H'F_t^0)' \right\|) \le M$,

(5) $E(\sup_{k \in K(k_0), k<k_0} \left\| \frac{1}{k_0-k}\sum_{t=k+1}^{k_0}(\tilde{F}_t - H'F_t^0)e_{it} \right\|) \le M$ *for each* i,

(6) $E(\sup_{k \in K(k_0), k<k_0} \left\| \frac{1}{k_0-k}\sum_{t=k+1}^{k_0} H'F_t^0(\tilde{F}_t - H'F_t^0)' \right\|) \le M$,

(7) $E(\sup_{k \in K(k_0), k<k_0} \left\| \frac{1}{T}\sum_{t=k+1}^{k_0}(\tilde{F}_t - H'F_t^0)(\tilde{F}_t - H'F_t^0)' \right\|) \le \frac{1}{\delta_{NT}^2}M$,

(8) $E(\sup_{k \in K(k_0), k<k_0} \left\| \frac{1}{T}\sum_{t=k+1}^{k_0}(\tilde{F}_t - H'F_t^0)e_{it} \right\|) \le (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2})M$ *for each* i,

(9) $E(\sup_{k \in K(k_0), k<k_0} \left\| \frac{1}{T}\sum_{t=k+1}^{k_0} H'F_t^0(\tilde{F}_t - H'F_t^0)' \right\|) \le (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2})M$,

(10) $E(\sup_{k \in K(k_0), k\le k_0} \left\| \frac{1}{T-k}\sum_{t=k+1}^{T}(\tilde{F}_t - H'F_t^0)(\tilde{F}_t - H'F_t^0)' \right\|) \le \frac{1}{\delta_{NT}^2}M$,

(11) $E(\sup_{k \in K(k_0), k\le k_0} \left\| \frac{1}{T-k}\sum_{t=k+1}^{T}(\tilde{F}_t - H'F_t^0)e_{it} \right\|) \le (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2})M$ *for each* i,

(12) $E(\sup_{k \in K(k_0), k\le k_0} \left\| \frac{1}{T-k}\sum_{t=k+1}^{T} H'F_t^0(\tilde{F}_t - H'F_t^0)' \right\|) \le (\frac{1}{\sqrt{T}} + \frac{1}{\delta_{NT}^2})M$.

Proof. We will show that terms in the parenthesis have the indicated stochastic order. Given our assumptions on the factor process and the error process and using Holder's inequality, $E\|fg\| \le (E\|f\|^2)^{\frac{1}{2}}(E\|g\|^2)^{\frac{1}{2}}$ repeatedly, it is easy to show their expectation have the same order.

First note that $[\frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\hat{\beta}_i(k) - Z_{2i}\hat{\delta}_i(k))(Y_i - X_i\hat{\beta}_i(k) - Z_{2i}\hat{\delta}_i(k))']\tilde{F} = \tilde{F}V_{NT}$, where V_{NT} is a diagonal matrix consists of the r largest eigenvalues of the matrix in the bracket. Define

$$u_i = X_i(\beta_i - \hat{\beta}_i(k)) + Z_{0i}(\delta_i - \hat{\delta}_i(k)) - (Z_{2i} - Z_{0i})\hat{\delta}_i(k),$$

then $Y_i - X_i\hat{\beta}_i(k) - Z_{2i}\hat{\delta}_i(k) = u_i + F^0\lambda_i + e_i$. Expanding terms, we have

$$\begin{aligned}\tilde{F}V_{NT} - \frac{1}{NT}\sum_{i=1}^{N} F^0\lambda_i\lambda_i'F^{0\prime}\tilde{F} &= \frac{1}{NT}\sum_{i=1}^{N} u_iu_i'\tilde{F} + \frac{1}{NT}\sum_{i=1}^{N} u_i\lambda_i'F^{0\prime}\tilde{F} + \frac{1}{NT}\sum_{i=1}^{N} u_ie_i'\tilde{F} \\ &\quad + \frac{1}{NT}\sum_{i=1}^{N} F^0\lambda_iu_i'\tilde{F} + \frac{1}{NT}\sum_{i=1}^{N} e_iu_i'\tilde{F} \\ &\quad + \frac{1}{NT}\sum_{i=1}^{N} F^0\lambda_ie_i'\tilde{F} + \frac{1}{NT}\sum_{i=1}^{N} e_i\lambda_i'F^{0\prime}\tilde{F} + \frac{1}{NT}\sum_{i=1}^{N} e_ie_i'\tilde{F} \\ &= I_1 + ... + I_8. \qquad (14)\end{aligned}$$

Define $H = \frac{1}{NT}\sum_{i=1}^{N}\lambda_i\lambda_i' F^{0\prime}\tilde{F}V_{NT}^{-1}$, then $(\tilde{F} - F^0 H)V_{NT} = I_1 + ... + I_8$.

Parts (1)-(3) correspond to part (ii) of Proposition A.1, part (i) of Lemma A.4 and part (i) of Lemma A.3 respectively in Bai (2009), and can be proved in a similar manner. A key step is to calculate $\left\|\frac{1}{\sqrt{T}}u_i\right\|$. In Bai (2009), $\left\|\frac{1}{\sqrt{T}}u_i\right\| = O_p(\left\|\hat{\beta} - \beta\right\|)$ while in current case $\left\|\frac{1}{\sqrt{T}}u_i\right\| = O_p(\left\|\hat{\beta}_i(k) - \beta_i\right\|) + O_p(\left\|\hat{\delta}_i(k) - \delta_i\right\|) + \frac{1}{\sqrt{T}}\left\|(Z_{2i} - Z_{0i})\hat{\delta}_i(k)\right\|$. If $k = k_0$, Song (2013) shows that $\beta_i - \hat{\beta}_i$ and $\delta_i - \hat{\delta}_i$ are $O_p(\frac{1}{\sqrt{T}})$. It can be verified that this result still holds for $|k - k_0| \leq C$. Moreover, given our assumptions on the regressors and factors, this $O_p(\frac{1}{\sqrt{T}})$ is uniform over i. For the last term,

$$\begin{aligned}\frac{1}{\sqrt{T}}\left\|(Z_{2i} - Z_{0i})\hat{\delta}_i(k)\right\| &= \frac{1}{\sqrt{T}}\left\|(Z_{2i} - Z_{0i})(\hat{\delta}_i(k) - \delta_i)\right\| + \frac{1}{\sqrt{T}}\left\|(Z_{2i} - Z_{0i})\delta_i\right\| \\ &= O_p(\frac{1}{\sqrt{T}}) + \frac{1}{\sqrt{T}}O_p(1) = O_p(\frac{1}{\sqrt{T}}),\end{aligned}$$

where the second equality follows from $E\left\|(Z_{2i} - Z_{0i})\delta_i\right\|^2 = |k - k_0|\, E(\frac{1}{|k-k_0|}\sum_{t=k+1}^{k_0}\left\|z_{it}\right\|^2) = O(1)$ for $|k - k_0| \leq C$. Thus, $\left\|\frac{1}{\sqrt{T}}u_i\right\| = O_p(\frac{1}{\sqrt{T}})$.

Next consider parts (4)-(9). Each term in parts (4)-(9) can be decomposed into eight terms according to equation (14). After decomposition, the proof of the last three terms can be found in existing literature. For part (4), the last three terms together is $O_p(\frac{1}{\delta_{NT}^2})$, see part (5) of Lemma 5 of Baltagi et al. (2015). For part (5), the last three terms together is $O_p(\frac{1}{\delta_{NT}})$, see part (4) of Lemma 5 of Baltagi et al. (2015), replacing F_t^0 by e_{it} does not change the result. For part (6), the last three terms together is $O_p(\frac{1}{\delta_{NT}})$, see part (4) of Lemma 5 of Baltagi et al. (2015). For part (7), the last three terms together is $O_p(\frac{1}{\delta_{NT}^2})$, see part (5) of Lemma 5 of Baltagi et al. (2015), which is a stronger result. For part (8), the last three terms together is $O_p(\frac{1}{\delta_{NT}^2})$, see Lemma 3 of Han and Inoue (2014), replacing $\sum_{t=1}^{\pi T}$ by $\sum_{t=k+1}^{k_0}$ and F_t^0 by e_{it} does not change the result. For part (9), the last three terms together is $O_p(\frac{1}{\delta_{NT}^2})$, see Lemma 3 of Han and Inoue (2014), replacing $\sum_{t=1}^{\pi T}$ by $\sum_{t=k+1}^{k_0}$ does not change the result. Assumptions in Baltagi et al. (2015) and Han and Inoue (2014) can be verified given Assumptions 5–14. For the first five terms, a key intermediate result is $\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}u_{it}^2\right\| = O_p(1)$ for parts (4)-(6) and $\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{T}\sum_{t=k+1}^{k_0}u_{it}^2\right\| = O_p(\frac{1}{T})$ for parts (7)-(9), which follows directly from $\left\|\frac{1}{\sqrt{T}}u_i\right\| = O_p(\frac{1}{\sqrt{T}})$. Based on this, it is easy to see part (4) is $O_p(1) + O_p(\frac{1}{\delta_{NT}^2}) = O_p(1)$, parts (5) and (6) are both $O_p(1) + O_p(\frac{1}{\delta_{NT}}) = O_p(1)$, part (7) is $O_p(\frac{1}{T}) + O_p(\frac{1}{\delta_{NT}^2}) = O_p(\frac{1}{\delta_{NT}^2})$, parts (8) and (9) are both $O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\delta_{NT}^2})$.

(10)-(12) can be proved following the same procedure as (7)-(9). ■

Lemma 7 *Under Assumptions 5-14, given $|k - k_0| \leq C$, there exists $M > 0$ such that*

(1) $E(\left\|\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t e_{it}\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each* i,

(2) $E(\left\|\frac{1}{T}\sum_{t=1}^{T}x_{it}(\tilde{F}_t - H'F_t^0)\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each* i,

(3) $E(\left\|\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t(\tilde{F}_t - H'F_t^0)'\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$,

(4) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\tilde{F}_t e_{it}\right\|) \leq M$ *for each* i,

(5) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}x_{it}(\tilde{F}_t - H'F_t^0)\right\|) \leq M$ *for each* i,

(6) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\tilde{F}_t(\tilde{F}_t - H'F_t^0)'\right\|) \leq M$,

(7) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{T}\sum_{t=k+1}^{k_0}\tilde{F}_t e_{it}\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each* i,

(8) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{T}\sum_{t=k+1}^{k_0}x_{it}(\tilde{F}_t - H'F_t^0)\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each* i,

(9) $E(\sup_{k\in K(k_0),k<k_0}\left\|\frac{1}{T}\sum_{t=k+1}^{k_0}\tilde{F}_t(\tilde{F}_t - H'F_t^0)'\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$,

(10) $E(\sup_{k\in K(k_0),k\leq k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}\tilde{F}_t e_{it}\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each* i,

(11) $E(\sup_{k\in K(k_0),k\leq k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}x_{it}(\tilde{F}_t - H'F_t^0)\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$ *for each* i,

(12) $E(\sup_{k\in K(k_0),k\leq k_0}\left\|\frac{1}{T-k}\sum_{t=k+1}^{T}\tilde{F}_t(\tilde{F}_t - H'F_t^0)'\right\|) \leq (\frac{1}{\sqrt{T}}+\frac{1}{\delta_{NT}^2})M$.

Proof. The proof of parts (2), (5), (8) and (11) are similar to parts (2), (5), (8) and (11) of Lemma 6. Other terms can be easily shown using Lemma 6. ■

A Proof of Theorem 1

Proof. To prove $\hat{\tau} - \tau_0 = o_p(1)$ as $(N,T) \to \infty$, we need to show for any $\epsilon > 0$ and $\eta \in (0, \min\{\tau_0, 1-\tau_0\})$, $P(\left|\hat{k} - k_0\right| > T\eta) < \epsilon$ as $(N,T) \to \infty$, i.e., we need to show $P(\hat{k} \in K^c) < \epsilon$ as $(N,T) \to \infty$. $\hat{k} = \arg\max \sum_{i=1}^{N}[V_i(k) - V_i(k_0)]$, hence $\sum_{i=1}^{N}[V_i(\hat{k}) - V_i(k_0)] \geq 0$. If $\hat{k} \in K^c$, then $\sup_{k\in K^c}\sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0$. This implies $P(\hat{k} \in K^c) \leq P(\sup_{k\in K^c}\sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0)$, hence it suffices to show for any given $\epsilon > 0$ and $\eta \in (0, \min\{\tau_0, 1-\tau_0\})$, $P(\sup_{k\in K^c}\sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0) < \epsilon$ as $(N,T) \to \infty$. If $\omega \in \{\omega : \sup_{k\in K^c}\sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0\}$ and $\arg\max_{k\in K^c}\sum_{i=1}^{N}[V_i(k) - V_i(k_0)] = k^*$, then $\sum_{i=1}^{N}[V_i(k^*) - V_i(k_0)] \geq 0$. This implies $\frac{\sum_{i=1}^{N}[V_i(k^*)-V_i(k_0)]}{|k^*-k_0|} \geq 0$ and it follows $\sup_{k\in K^c}\frac{\sum_{i=1}^{N}[V_i(k)-V_i(k_0)]}{|k-k_0|} \geq \frac{\sum_{i=1}^{N}[V_i(k^*)-V_i(k_0)]}{|k^*-k_0|} \geq 0$. This implies $\omega \in \{\omega : \sup_{k\in K^c}\frac{\sum_{i=1}^{N}[V_i(k)-V_i(k_0)]}{|k-k_0|} \geq 0\}$, hence $\{\omega : \sup_{k\in K^c}\sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0\} \subseteq \{\omega : \sup_{k\in K^c}\frac{\sum_{i=1}^{N}[V_i(k)-V_i(k_0)]}{|k-k_0|} \geq 0\}$. Similarly, $\{\omega : \sup_{k\in K^c}\frac{\sum_{i=1}^{N}[V_i(k)-V_i(k_0)]}{|k-k_0|} \geq 0\} \subseteq \{\omega : \sup_{k\in K^c}\sum_{i=1}^{N}[V_i(k) - V_i(k_0)] \geq 0\}$. Thus, $\{\omega : \sup_{k\in K^c}\sum_{i=1}^{N}[V_i(k) -$

$V_i(k_0)] \geq 0\} = \{\omega : \sup_{k\in K^c} \frac{\sum_{i=1}^N [V_i(k)-V_i(k_0)]}{|k-k_0|} \geq 0\}$ and it suffices to show for any given $\epsilon > 0$ and $\eta \in (0, \min\{\tau_0, 1-\tau_0\})$, $P(\sup_{k\in K^c} \frac{\sum_{i=1}^N [V_i(k)-V_i(k_0)]}{|k-k_0|} \geq 0) < \epsilon$ as $(N,T) \to \infty$. Note that $\frac{\sum_{i=1}^N [V_i(k)-V_i(k_0)]}{|k-k_0|} = -\sum_{i=1}^N G_i(k) + \frac{1}{|k_0-k|}\sum_{i=1}^N H_i(k)$ for $k \neq k_0$, thus $\sup_{k\in K^c} \frac{\sum_{i=1}^N [V_i(k)-V_i(k_0)]}{|k-k_0|} \geq 0$ implies $\sup_{k\in K^c} \left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| \geq \inf_{k\in K^c} \sum_{i=1}^N G_i(k)$, it suffices to show that for any $\epsilon > 0$ and $\eta \in (0, \min\{\tau_0, 1-\tau_0\})$, $P(\sup_{k\in K^c} \left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| \geq \inf_{k\in K^c} \sum_{i=1}^N G_i(k)) < \epsilon$ as $(N,T) \to \infty$. Due to symmetry, it suffices to study the case $k < k_0$.

Consider the left hand side first.

$$\begin{aligned}
& \frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k) \\
= \; & 2\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) - 2\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'Z_{0i}'M_ie_i \\
& + \frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_ie_i - \frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{0i}(Z_{0i}'M_iZ_{0i})^{-1}Z_{0i}'M_ie_i
\end{aligned}$$

For the third term, noting that $M_i(Z_{1i}+Z_{2i}) = M_iZ_i = 0$, we have

$$\begin{aligned}
& \sup_{k\in K^c, k<k_0} \frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{2i}(Z_{2i}'M_iZ_{2i})^{-1}Z_{2i}'M_ie_i \\
= \; & \sup_{k\in K^c, k<k_0} \frac{1}{|k-k_0|}\sum_{i=1}^N e_i'M_iZ_{1i}(Z_{1i}'M_iZ_{1i})^{-1}Z_{1i}'M_ie_i \\
\leq \; & \frac{1}{T\eta}\sum_{i=1}^N \sup_{k\in K^c, k<k_0} \left\|\frac{e_i'M_iZ_{1i}}{\sqrt{k}}\right\|^2 \sup_{k\in K^c, k<k_0} \left\|(\frac{Z_{1i}'M_iZ_{1i}}{k})^{-1}\right\|,
\end{aligned}$$

thus by parts (16) and (17) of Lemma 3 and Markov inequality, this term is $O_p(\frac{N\log T}{T})$. Similarly, the fourth term is not larger than $\frac{1}{T\eta}\sum_{i=1}^N \left\|\frac{e_i'M_iZ_{0i}}{\sqrt{T-k_0}}\right\|^2 \left\|(\frac{Z_{0i}'M_iZ_{0i}}{T-k_0})^{-1}\right\|$, and by parts (9) and (14) of Lemma 3 and Markov inequality, this term is $O_p(\frac{N}{T})$. For the first term, the expectation is not larger than

$$\begin{aligned}
& \frac{2\sqrt{N\lambda_N}}{T\eta}(\lim_{N\to\infty}\max_{1\leq i\leq N} \frac{N\delta_i'\delta_i}{\lambda_N})^{\frac{1}{2}} \frac{1}{N}\sum_{i=1}^N E(\sup_{k\in K^c, k<k_0} \|Z_{0i}'M_iZ_{1i}(Z_{1i}'M_iZ_{1i})^{-1}Z_{1i}'M_ie_i\|) \\
\leq \; & \frac{2\sqrt{N\lambda_N}}{\sqrt{T}\eta}(\lim_{N\to\infty}\max_{1\leq i\leq N} \frac{N\delta_i'\delta_i}{\lambda_N})^{\frac{1}{2}} \frac{1}{N}\sum_{i=1}^N E(\sup_{k\in K^c, k<k_0} \left\|\frac{Z_{0i}'M_iZ_{1i}}{k}\right\| \\
& \sup_{k\in K^c, k<k_0} \left\|(\frac{Z_{1i}'M_iZ_{1i}}{k})^{-1}\right\| \sup_{k\in K^c, k<k_0} \left\|\frac{Z_{1i}'M_ie_i}{\sqrt{k}}\right\|),
\end{aligned}$$

thus by parts (16), (17) and (18) of Lemma 3 and Markov inequality, this term is $O_p(\sqrt{\frac{N\lambda_N\log T}{T}})$. For the second term, using part (9) of Lemma 3, it is easy to see it's $O_p(\sqrt{\frac{N\lambda_N}{T}})$.

Next consider the right hand side. Using Lemma 4, there exists $\alpha > 0$ such that for any $\epsilon > 0$, $P(\inf_{k\neq k_0} \sum_{i=1}^N G_i(k) \geq \alpha\lambda_N) > 1-\epsilon$ as $(N,T)\to\infty$. Noting that $\inf_{k\in K^c, k<k_0} \sum_{i=1}^N G_i(k) \geq \inf_{k\neq k_0} \sum_{i=1}^N G_i(k)$, under Assumption 4(a), or 4(b), or 4(c), $\sup_{k\in K^c, k<k_0} \left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right|$ will be dominated by $\inf_{k\in K^c, k<k_0} \sum_{i=1}^N G_i(k)$ as $(N,T)\to\infty$. ■

B Proof of Theorem 2

Proof. To prove $\hat{k}-k_0 = O_p(1)$ as $(N,T)\to\infty$ and $\frac{N}{\sqrt{T}}\to 0$, we need to show for any $\epsilon > 0$, there exist $C<\infty$, $N^*>0$, $T^*>0$ and $\gamma>0$, such that for $N>N^*$, $T>T^*$ and $\frac{N}{\sqrt{T}}<\gamma$, $P(\left|\hat{k}-k_0\right| > C) < \epsilon$. Since $P(\left|\hat{k}-k_0\right| > C) < P(\sup_{|k-k_0|>C} \sum_{i=1}^N [V_i(k)-V_i(k_0)] \geq 0)$, it suffices to show for any $\epsilon>0$, there exist $C<\infty$, $N^*>0$, $T^*>0$ and $\gamma>0$, such that for $N>N^*$, $T>T^*$ and $\frac{N}{\sqrt{T}}<\gamma$, $P(\sup_{|k-k_0|>C} \sum_{i=1}^N [V_i(k)-V_i(k_0)] \geq 0) < \epsilon$. Since $\hat{\tau}$ is consistent, $P(\hat{k}\in K^c)\to 0$ as $(N,T)\to\infty$. Noting that $K(C) = \{k : |k-k_0| > C\}\cap K$, it suffices to show for any $\epsilon>0$, there exist $C<\infty$, $N^*>0$, $T^*>0$ and $\gamma>0$, such that for $N>N^*$, $T>T^*$ and $\frac{N}{\sqrt{T}}<\gamma$, $P(\sup_{k\in K(C)} \sum_{i=1}^N [V_i(k)-V_i(k_0)] \geq 0) < \epsilon$. Since $\sup_{k\in K(C)} \sum_{i=1}^N [V_i(k)-V_i(k_0)] \geq 0$ implies $\sup_{k\in K(C)} \left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| \geq \inf_{k\in K(C)} \sum_{i=1}^N G_i(k)$, it suffices to show that for any $\epsilon>0$, there exist $C<\infty$, $N^*>0$, $T^*>0$ and $\gamma>0$, such that for $N>N^*$, $T>T^*$ and $\frac{N}{\sqrt{T}}<\gamma$, $P(\sup_{k\in K(C)} \left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| \geq \inf_{k\in K(C)} \sum_{i=1}^N G_i(k)) < \epsilon$. Again by symmetry, it suffices to study the case $k<k_0$, i.e. $P(\sup_{k\in K(C), k<k_0} \left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| \geq \inf_{k\in K(C), k<k_0} \sum_{i=1}^N G_i(k)) < \epsilon$. By Lemma 4, there exists $\alpha>0$ such that for any $\epsilon>0$, there exist $N^*>0$, $T^*>0$ such that for $N>N^*$, $T>T^*$, $P(\inf_{k\neq k_0} \sum_{i=1}^N G_i(k) \geq \alpha\lambda_N) > 1-\epsilon$. Noting that $\inf_{k\in K(C), k<k_0} \sum_{i=1}^N G_i(k) \geq \inf_{k\neq k_0} \sum_{i=1}^N G_i(k)$, it suffices to show for any $\epsilon>0$, there exist $C<\infty$, $N^*>0$, $T^*>0$ and $\gamma>0$, such that for $N>N^*$, $T>T^*$ and $\frac{N}{\sqrt{T}}<\gamma$, $P(\sup_{k\in K(C), k<k_0} \left|\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)\right| \geq \alpha\lambda_N) < \epsilon$. The first two terms of $\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)$ is

$$
\begin{aligned}
& 2\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'(Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) - 2\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'Z_{0i}'M_ie_i \\
= \; & [2\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'Z_{\Delta i}'M_ie_i - 2\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i)]sgn(k_0-k) \\
= \; & 2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'Z_{\Delta i}'e_i - 2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'e_i \\
& -2sgn(k_0-k)\frac{1}{|k-k_0|}\sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_ie_i) \\
= \; & A+B+C.
\end{aligned}
$$

The last two terms of $\frac{1}{|k-k_0|}\sum_{i=1}^N H_i(k)$ is

$$\begin{aligned}
& \frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N e_i' M_i Z_{2i}(Z_{2i}' M_i Z_{2i})^{-1} Z_{2i}' M_i e_i - \frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N e_i' M_i Z_{0i}(Z_{0i}' M_i Z_{0i})^{-1} Z_{0i}' M_i e_i \\
= & \frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N e_i' M_i Z_{\Delta i}(Z_{2i}' M_i Z_{2i})^{-1} Z_{\Delta i}' M_i e_i \\
& +2sgn(k_0-k)\frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N e_i' M_i Z_{0i}(Z_{2i}' M_i Z_{2i})^{-1} Z_{\Delta i}' M_i e_i \\
& +\frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N e_i' M_i Z_{0i}[(Z_{2i}' M_i Z_{2i})^{-1} - (Z_{0i}' M_i Z_{0i})^{-1}] Z_{0i}' M_i e_i \\
= & D+E+F.
\end{aligned}$$

Thus by Lemma 5,

$$\begin{aligned}
\sup_{k\in K(C),k<k_0}\left|\frac{1}{|k-k_0|}\sum\nolimits_{i=1}^N H_i(k)\right| \leq & \sup_{k\in K(C),k<k_0}|A| + \sup_{k\in K(C),k<k_0}|B| + \sup_{k\in K(C),k<k_0}|C| \\
& + \sup_{k\in K(C),k<k_0}|D| + \sup_{k\in K(C),k<k_0}|E| + \sup_{k\in K(C),k<k_0}|F| \\
\leq & \sup_{k\in K(C),k<k_0}|A| + \sup_{k\in K(k_0),k<k_0}|B| + \sup_{k\in K(k_0),k<k_0}|C| \\
& + \sup_{k\in K(k_0),k<k_0}|D| + \sup_{k\in K(k_0),k<k_0}|E| + \sup_{k\in K(k_0),k<k_0}|F| \\
= & \sup_{k\in K(C),k<k_0}|A| + O_p(\frac{\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N}) + O_p(\frac{N\log T}{T}) + O_p(\frac{N}{\sqrt{T}}).
\end{aligned}$$

Under Assumption 4(a), the last three terms are all $o_p(1)$. For the first term, similar to the proof of part (1) of Lemma 5, for all N we have,

$$\begin{aligned}
E(\sup_{k\in K(C),k<k_0}|A|)^2 & \leq 4\lambda_N E(\sup_{k\in K(C),k<k_0}\left|\frac{1}{|k-k_0|}\sum\nolimits_{t=k+1}^{k_0}\eta_{Nt}\right|^2) \\
& \leq 16\lambda_N M\sum\nolimits_{k=T(\tau_0-\eta)}^{k_0-C-1}\frac{1}{(k_0-k)^2} \leq \frac{16\lambda_N M}{C} < \epsilon,
\end{aligned}$$

if C is large enough. The proof is finished. ■

C Proof of Theorem 3

Proof. The proof is similar to the proof of Theorem 2. Based on Theorem 1, $\hat{\tau}$ is consistent under Assumption 4(b) or 4(c), i.e., for any $\epsilon > 0$ and $\eta > 0$, $P(\hat{k} \in K^c) < \epsilon$ as $(N,T) \to \infty$, hence it suffices to show for any $\epsilon > 0$ and $\eta > 0$, $P(\hat{k} \in K(k_0)) < \epsilon$ as $(N,T) \to \infty$ under Assumption 4(b) or 4(c). By Lemma 4, there exists $\alpha > 0$ such that for any $\epsilon > 0$, there exist $N^* > 0$, $T^* > 0$ such that for $N^* > N$, $T > T^*$, $P(\inf_{k\in K(k_0)} \sum_{i=1}^N G_i(k) \geq \alpha\lambda_N) > 1-\epsilon$. By Lemma 5,

$$\begin{aligned}
\sup_{k\in K(k_0)} \left| \frac{1}{|k-k_0|} \sum_{i=1}^N H_i(k) \right| &\leq \sup_{k\in K(k_0)} |A| + \sup_{k\in K(k_0)} |B| + \sup_{k\in K(k_0)} |C| \\
&\quad + \sup_{k\in K(k_0)} |D| + \sup_{k\in K(k_0)} |E| + \sup_{k\in K(k_0)} |F| \\
&= O_p(\sqrt{\lambda_N}) + O_p(\frac{\sqrt{N}}{\sqrt{T}}\sqrt{\lambda_N}) + O_p(\frac{N\log T}{T}) + O_p(\frac{N}{\sqrt{T}}).
\end{aligned}$$

Under Assumption 4(b) or 4(c), all these four terms will be dominated by $\alpha\lambda_N$, the proof is thus finished. ■

D Proof of Theorem 4

Proof. Define $V_{NT}(k) = \sum_{i=1}^N [V_i(k) - V_i(k_0)]$, $U_{NT}(k) = -\sum_{i=1}^N \delta_i' Z_{\Delta i}' Z_{\Delta i}\delta_i + 2sgn(k_0-k)\sum_{i=1}^N \delta_i' Z_{\Delta i}' e_i$, both $V_{NT}(k)$ and $U_{NT}(k)$ are countable dimensional random vector. For any fixed constant $C < \infty$, define $V_{NT}^C(k) = V_{NT}(k)$ for $|k_0 - k| < C$, $U_{NT}^C(k) = U_{NT}(k)$ for $|k_0-k| < C$, $W^C(m) = W(m)$ for $|m| < C$. $V_{NT}^C(k)$, $U_{NT}^C(k)$ and $W^C(m)$ are all finite dimensional random vector.

Step 1: Under Assumption 4(a), $V_{NT}^C(k) \xrightarrow{p} U_{NT}^C(k)$ for any fixed $C < \infty$.

Again due to symmetry, it suffices to show the case $k < k_0$.

For $k \neq k_0$, $V_{NT}(k) = -|k_0-k|\sum_{i=1}^N G_i(k) + \sum_{i=1}^N H_i(k)$, where

$$\begin{aligned}
-|k_0-k|\sum_{i=1}^N G_i(k) &= -\sum_{i=1}^N \delta_i'[(Z_{0i}'M_iZ_{0i}) - (Z_{0i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{0i})]\delta_i \\
&= -\sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{\Delta i})\delta_i + \sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{\Delta i})\delta_i \\
&= -\sum_{i=1}^N \delta_i'Z_{\Delta i}'Z_{\Delta i}\delta_i + \sum_{i=1}^N \delta_i'Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}\delta_i \\
&\quad + \sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{\Delta i})\delta_i,
\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^N H_i(k) &= |k_0-k|(A+B+C+D+E+F) \\
&= 2sgn(k_0-k)\sum_{i=1}^N \delta_i'Z_{\Delta i}'e_i + |k_0-k|(B+C+D+E+F).
\end{aligned}$$

Hence for $k \neq k_0$,

$$\begin{aligned}
V_{NT}(k) - U_{NT}(k) &= \sum_{i=1}^N \delta_i'Z_{\Delta i}'X_i(X_i'X_i)^{-1}X_i'Z_{\Delta i}\delta_i \\
&\quad + \sum_{i=1}^N \delta_i'(Z_{\Delta i}'M_iZ_{2i})(Z_{2i}'M_iZ_{2i})^{-1}(Z_{2i}'M_iZ_{\Delta i})\delta_i \\
&\quad + |k_0-k|(B+C+D+E+F),
\end{aligned}$$

and for $k = k_0$, $V_{NT}(k) - U_{NT}(k) = 0$. As proved in Step 2 of Lemma 4, the first two terms are both $O_p(\frac{1}{T})$ uniformly over $k_0 - C \leq k < k_0$ as $(N,T) \to \infty$. For the last five terms, using Lemma 5,

$$\begin{aligned}
& \sup_{k_0 - C \leq k < k_0} \left| |k_0 - k| \left(B + C + D + E + F \right) \right| \\
\leq \ & C(\sup_{k \in K(k_0), k<k_0} |B| + \sup_{k \in K(k_0), k<k_0} |C| + \sup_{k \in K(k_0), k<k_0} |D| + \sup_{k \in K(k_0), k<k_0} |E| + \sup_{k \in K(k_0), k<k_0} |F|) \\
= \ & o_p(1),
\end{aligned}$$

as $(N,T) \to \infty$ and $\frac{N}{\sqrt{T}} \to 0$. Taking together, we have $\sup_{k_0 - C \leq k < k_0} |V_{NT}(k) - U_{NT}(k)| \xrightarrow{p} 0$ as $(N,T) \to \infty$ and $\frac{N}{\sqrt{T}} \to 0$.

Step 2: For any fixed $C < \infty$, as finite dimensional random vectors, $U_{NT}^C(k) \xrightarrow{d} W^C(k - k_0)$ as $N \to \infty$.

Note that

$$U_{NT}(k) = \begin{cases} -\sum_{t=k+1}^{k_0} \sum_{i=1}^{N} \delta_i' z_{it} z_{it}' \delta_i + 2 \sum_{t=k+1}^{k_0} \sum_{i=1}^{N} \delta_i' z_{it} e_{it}, \ for \ k - k_0 \leq -1, \\ -\sum_{t=k_0+1}^{k} \sum_{i=1}^{N} \delta_i' z_{it} z_{it}' \delta_i - 2 \sum_{t=k_0+1}^{k} \sum_{i=1}^{N} \delta_i' z_{it} e_{it}, \ for \ k - k_0 \geq 1. \end{cases}$$

Under part (5) of Assumption 3, part (5) of Assumption 2 and Assumption 4(a), $\sum_{i=1}^{N} \delta_i' z_{it} z_{it}' \delta_i \xrightarrow{p} \lambda \xi$ for each t and as a random vector, $(\frac{1}{\sqrt{\lambda_N}} \sum_{i=1}^{N} \delta_i' z_{i,k_0} e_{i,k_0}, ..., \frac{1}{\sqrt{\lambda_N}} \sum_{i=1}^{N} \delta_i' z_{i,k_0-C} e_{i,k_0-C})' \xrightarrow{d} (Z_0, ..., Z_{-C})'$. Since

$$\begin{aligned}
& (\sum_{t=k_0}^{k_0} \sum_{i=1}^{N} \delta_i' z_{it} e_{it}, ..., \sum_{t=k_0-C}^{k_0} \sum_{i=1}^{N} \delta_i' z_{it} e_{it})' \\
= \ & Q(\sum_{i=1}^{N} \delta_i' z_{i,k_0} e_{i,k_0}, ..., \sum_{i=1}^{N} \delta_i' z_{i,k_0-C} e_{i,k_0-C})',
\end{aligned}$$

where Q is a $(C+1) \times (C+1)$ lower triangular matrix with all nonzero element equal to one, we have

$$\begin{aligned}
& (\sum_{t=k_0}^{k_0} \sum_{i=1}^{N} \delta_i' z_{it} e_{it}, ..., \sum_{t=k_0-C}^{k_0} \sum_{i=1}^{N} \delta_i' z_{it} e_{it})' \xrightarrow{d} Q(Z_0, ..., Z_{-C})' \\
= \ & (\sum_{t=0}^{0} Z_t, ..., \sum_{t=-C}^{0} Z_t)'.
\end{aligned}$$

Similarly,

$$(\sum_{t=k_0}^{k_0} \sum_{i=1}^{N} \delta_i' z_{it} z_{it}' \delta_i, ..., \sum_{t=k_0-C}^{k_0} \sum_{i=1}^{N} \delta_i' z_{it} z_{it}' \delta_i)' \xrightarrow{p} (\lambda \xi, ..., (C+1) \lambda \xi)'.$$

For the second half of $U_{NT}(k)$, we have similar result. Taking together, we have $U_{NT}^C(k) \xrightarrow{d} W^C(k - k_0)$ as $N \to \infty$.

Step 3: $V_{NT}^{C}(k) \xrightarrow{d} W^{C}(k-k_0)$ as $(N,T) \to \infty$ and $\frac{N}{\sqrt{T}} \to 0$ for any fixed $C < \infty$.

Based on Step 1 and Step 2 and using Slutsky's Lemma for random vectors, $V_{NT}^{C}(k) \xrightarrow{d} W^{C}(k-k_0)$.

Step 4: $\arg\max V_{NT}^{C}(k) - k_0 \xrightarrow{d} \arg\max W^{C}(m)$ uniformly as $(N,T) \to \infty$ and $\frac{N}{\sqrt{T}} \to 0$ for any fixed $C < \infty$.

Step 4.1: If $W(m)$ does not have unique maximizer, then these exist $m \neq m'$ such that $W(m) = W(m')$. Consider the case $m' > m \geq 1$, $P(W(m) = W(m')) = P((m'-m)\xi + 2\sqrt{\lambda}\sum_{t=m}^{m'} Z_t = 0) = 0$. Other cases can be proved similarly. Since the number of integer pairs (m, m') is countable and sum of countable zero is still zero, the probability that $W(m)$ does not have unique maximizer is zero. Therefore, with probability one $\arg\max W(m)$ is unique.

Step 4.2: Based on Step 3 and using continuous mapping theorem, $\arg\max V_{NT}^{C}(k) \xrightarrow{d} \arg\max W^{C}(m)$. Note that for a finite dimensional vector X, $Y = \arg\max X$ is a continuous function. By definition of convergence of distribution, for any $\epsilon > 0$ and any $1 \leq j \leq C$, there exist $N_j^* > 0$, $T_j^* > 0$ and $\gamma_j > 0$ such that if $N > N_j^*$, $T > T_j^*$ and $\frac{N}{\sqrt{T}} < \gamma_j$, then $\left|P(\arg\max V_{NT}^{C}(k) - k_0 = j) - P(\arg\max W^{C}(m) = j)\right| < \epsilon$. Take $N^* = \max\{N_j^*, 1 \leq j \leq C\}$, $T^* = \max\{T_j^*, 1 \leq j \leq C\}$ and $\gamma = \min\{\gamma_j, 1 \leq j \leq C\}$. Since $C < \infty$, we have $N^* < \infty$, $T^* < \infty$ and $\gamma > 0$. For $N > N^*$, $T > T^*$ and $\frac{N}{\sqrt{T}} < \gamma$, $\left|P(\arg\max V_{NT}^{C}(k) - k_0 = j) - P(\arg\max W^{C}(m) = j)\right| < \epsilon$ for all $1 \leq j \leq C$.

Step 5: $\hat{k} - k_0 \xrightarrow{d} \arg\max W(m)$.

Step 5.1:

$\hat{k} - k_0 = O_p(1)$ as $(N,T) \to \infty$ and $\frac{N}{\sqrt{T}} \to 0$, hence for any $\frac{\epsilon}{3} > 0$, there exist $C_1 < \infty$, $N_1 > 0$, $T_1 > 0$ and $\gamma_1 > 0$, such that for $N > N_1$, $T > T_1$ and $\frac{N}{\sqrt{T}} < \gamma_1$, $P(\left|\hat{k} - k_0\right| > C_1) < \frac{\epsilon}{3}$.

Step 5.2: $\hat{m} = \arg\max W(m) = O_p(1)$.

By strong law of large number, $W(m) \xrightarrow{a.s.} -\infty$ as $|m| \to \infty$. Thus $P(\lim_{C\to\infty} \sup_{|m|>C} W(m) = -\infty) = 1$ and this implies $\lim_{C\to\infty} P(\sup_{|m|>C} W(m) \geq 0) = P(\lim_{C\to\infty} \sup_{|m|>C} W(m) \geq 0) = 0$. Therefore, for any $\frac{\epsilon}{3} > 0$, there exists $C_2 < \infty$ such that $P(\sup_{|m|>C_2} W(m) \geq 0) < \frac{\epsilon}{3}$. Since $W(0) = 0$, $\sup W(m) \geq 0$, and $P(|\hat{m}| > C_2) \leq P(\sup_{|m|>C_2} W(m) \geq 0) < \frac{\epsilon}{3}$.

Step 5.3:

Take $C = \max\{C_1, C_2\}$ in Step 4, then for any $\frac{\epsilon}{3} > 0$, there exist $N_2 > 0$, $T_2 > 0$ and $\gamma_2 > 0$, such that for $N > N_2$, $T > T_2$ and $\frac{N}{\sqrt{T}} < \gamma_2$, $\left|P(\arg\max V_{NT}^{C}(k) - k_0 = j) - P(\arg\max W^{C}(m) = j)\right| < \frac{\epsilon}{3}$ for all $1 \leq j \leq C$.

Step 5.4:

Take $N^* = \max\{N_1, N_2\}$, $T^* = \max\{T_1, T_2\}$ and $\gamma = \min\{\gamma_1, \gamma_2\}$. For any $N > N^*$, $T > T^*$ and $\frac{N}{\sqrt{T}} < \gamma$, if $|j| > C$,

$$\begin{aligned}
\left|P(\hat{k} - k_0 = j) - P(\hat{m} = j)\right| &< P(\hat{k} - k_0 = j) + P(\hat{m} = j) \\
&< P(\left|\hat{k} - k_0\right| > C) + P(|\hat{m}| > C) \\
&< P(\left|\hat{k} - k_0\right| > C_1) + P(|\hat{m}| > C_2) \\
&< \frac{\epsilon}{3} + \frac{\epsilon}{3} < \epsilon;
\end{aligned}$$

if $|j| \le C$, $\hat{k} - k_0 = j$ implies $\arg\max V_{NT}^C(k) - k_0 = j$, hence

$$P(\hat{k} - k_0 = j) \le P(\arg\max V_{NT}^C(k) - k_0 = j),$$

and $\arg\max V_{NT}^C(k) - k_0 = j$ implies $\hat{k} - k_0 = j$ or $\left|\hat{k} - k_0\right| > C$, hence

$$P(\arg\max V_{NT}^C(k) - k_0 = j) < P(\hat{k} - k_0 = j) + P(\left|\hat{k} - k_0\right| > C).$$

Therefore,

$$\left|P(\hat{k} - k_0 = j) - P(\arg\max V_{NT}^C(k) - k_0 = j)\right| < P(\left|\hat{k} - k_0\right| > C) < \frac{\epsilon}{3},$$

and similarly

$$\left|P(\hat{m} = j) - P(\arg\max W^C(m) = j)\right| < P(|\hat{m}| > C) < \frac{\epsilon}{3}.$$

It follows that

$$\begin{aligned}
\left|P(\hat{k} - k_0 = j) - P(\hat{m} = j)\right| &< \left|P(\hat{k} - k_0 = j) - P(\arg\max V_{NT}^C(k) - k_0 = j)\right| \\
&\quad + \left|P(\arg\max V_{NT}^C(k) - k_0 = j) - P(\arg\max W^C(m) = j)\right| \\
&\quad + \left|P(\hat{m} = j) - P(\arg\max W^C(m) = j)\right| \\
&< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} < \epsilon.
\end{aligned}$$

We've proved that for any $\epsilon > 0$, there exist $N^* > 0$, $T^* > 0$ and $\gamma > 0$, such that for $N > N^*$, $T > T^*$ and $\frac{N}{\sqrt{T}} < \gamma$, $\left|P(\hat{k} - k_0 = j) - P(\hat{m} = j)\right| < \epsilon$ for all j. By definition, $\hat{k} - k_0 \xrightarrow{d} \arg\max W(m)$. ■

E Proof of Theorem 5

Proof. The proof of Theorem 1 does not rely on weak cross-sectional dependence, hence $\hat{\tau} - \tau_0$ is still consistent when cross-sectional dependence is strong. The rest of the proof follows the same procedure as Theorem 2. The difference is when cross-sectional dependence is strong, $\sup_{k\in K(k_0),k<k_0} |A|$ is $O_p(\sqrt{N\lambda_N})$, which is of the same order as $\inf_{k\in K(k_0)} \sum_{i=1}^N G_i(k)$ given $\lambda_N = O(N)$. And similar to the proof of Theorem 2,

$$
\begin{aligned}
E(\sup_{k\in K(C),k<k_0} |A|) &\leq \sum\nolimits_{i=1}^N E(\sup_{k\in K(C),k<k_0} \left|\frac{1}{|k-k_0|}\sum\nolimits_{t=k+1}^{k_0} \delta_i' z_{it} e_{it}\right|) \\
&\leq 16\sqrt{N\lambda_N}M\sum\nolimits_{k=T(\tau_0-\eta)}^{k_0-C-1} \frac{1}{(k_0-k)^2} \leq \frac{16\sqrt{N\lambda_N}M}{C} < \alpha\lambda_N,
\end{aligned}
$$

if C is large enough. ■

F Proof of Theorem 6 and Theorem 7

Proof. To prove Theorem 6, what we need to show is for any $\epsilon > 0$ and $\eta > 0$, there exist $N^* > 0$ and $T^* > 0$ such that for $N > N^*$ and $T > T^*$, $P(\left|\tilde{k} - k_0\right| > T\eta) < \epsilon$. First note that

$$P(\left|\tilde{k}-k_0\right| > T\eta) = P(\left|\tilde{k}-k_0\right| > T\eta, \left|\hat{k}-k_0\right| > C) + \sum\nolimits_{j=-C}^{C} P(\left|\tilde{k}-k_0\right| > T\eta, \left|\hat{k}-k_0\right| = j).$$

Since $\left|\hat{k}-k_0\right| = O_p(1)$, there exists $C > 0$ such that $P(\left|\hat{k}-k_0\right| > C) < \frac{\epsilon}{2}$ for large N and large T, it follows that the first term is less than $\frac{\epsilon}{2}$ for large N and large T. Since $P(\left|\tilde{k}-k_0\right| > T\eta, \left|\hat{k}-k_0\right| = j)$ is no larger than $P(\left|\tilde{k}-k_0\right| > T\eta, given\ |k-k_0| = j)$, it suffices to show for each $j = -C, ..., C$, $P(\left|\tilde{k}-k_0\right| > T\eta, given\ |k-k_0| = j) < \frac{\epsilon}{2(2C+1)}$ for large N and large T. By Symmetry, it suffices to show for each $j = -C, ..., C$, $P(\tilde{k} \in K^c\ and\ \tilde{k} < k_0, given\ |k-k_0| = j) < \frac{\epsilon}{2(2C+1)}$ for large N and large T.

Similarly, to prove Theorem 7, it suffices to show for each $j = -C, ..., C$, $P(\tilde{k} \neq k_0, given\ |k-k_0| = j) < \frac{\epsilon}{2(2C+1)}$ for large N and large T. Theorem 6 shows that $\left|\tilde{k}-k_0\right| = o_p(T)$, hence it suffices to show for each $j = -C, ..., C$, $P(\tilde{k} \in K(k_0), given\ |k-k_0| = j) < \frac{\epsilon}{2(2C+1)}$ for large N and large T. By symmetry, it suffices to show for each $j = -C, ..., C$, $P(\tilde{k} \in K(k_0)\ and\ \tilde{k} < k_0, given\ |k-k_0| = j) < \frac{\epsilon}{2(2C+1)}$ for large N and large T.

The rest of the proof follow the same procedure as Theorem 1 and Theorem 3 respectively, but in current case we have extra regressors $\tilde{F}$ and extra error $(F^0H - \tilde{F})H^{-1}\Lambda$. The proof of Theorem 1 relies on Lemma 3 and Lemma 4, which further relies on Lemma 3. The proof of Theorem 3 relies on Lemma 4 and Lemma 5, which further rely on Lemma 3. Thus, to prove Theorem 6 and Theorem 7, it suffices to reestablish Lemma 3 with the presence of extra regressors $\tilde{F}$ and extra error $(\tilde{F} - F^0H)H^{-1}\Lambda$. Based on Lemma 6, Lemma 7 and our assumptions on the factor process and error process, this can be easily done following the same procedure as proving Lemma 3. Also note that with $\frac{\sqrt{T}}{N} \to 0$, the effect of using estimated factors disappears asymptotically. ■