Common Local Breaks in Time Trends for Large Panel Data

Dukpa Kim[*]

University of Virginia

September 26, 2010

Abstract

This paper analyzes the issues related to the estimation of common time trend breaks in large panel data. The break parameters are specified to be local to zero. In that case, the common components creating strong cross equation dependence can be consistently estimated without knowing the break date. Subtracting these common components estimates from the original observations removes the cross equation correlation asymptotically. The common break date estimate obtained by minimizing the sum of squared residuals over all permissible break dates after the subtraction of the common components estimates achieves a faster rate of convergence than the one obtained without the subtraction. The limiting distribution of the common break date estimate is provided so that confidence intervals can be formed. Some Monte Carlo simulation results are reported to support the asymptotic results.

JEL Classification Number: **C33**
Keywords: structural break, deterministic time trend, large panel data, common factor, local asymptotics

[*]Department of Economics, University of Virginia, Monroe Hall, McCormick Road, Charlottesville, VA 22903 (dk4p@virginia.edu)

1 Introduction

Exploiting panel structure for time series analysis has become very popular and successful in econometrics, especially in the area of unit root and cointegration tests. Such an effort is relatively scarce but growing in the area of structural breaks. The purpose of this paper is to propose a new estimation method for a common structural break in time series panel data.

A distinct feature of time series panel data is that there is not necessarily a cross sectional relationship among individuals but they still possess in common a certain aspect of their time series properties, for example, the sum of the autoregressive coefficients or the timing of a structural break. Panel data naturally offers a more efficient method of inference for this common aspect. The efficiency gain is often the greatest when the cross sectional units are independent. However, numerous empirical studies have reported there exist common factors in many time series panels such as multicountry or multistate data. Common factors typically create strong cross equation dependence and make inference more challenging.

Kim (2010) proposed an estimation procedure for a common deterministic trend break in large panels with strong cross equation dependence. The estimation method proposed is simply to minimize the sum of squared residuals for all permissible break dates. Kim (2010) reports the rate of convergence and the limiting distribution of this simple common break date estimate under various sets of assumptions on the error process. One of the main results in Kim (2010) is that, in the presence of strong cross equation dependence, the rate of convergence of the common break date estimate is only as fast as the one obtained with one time series.

This paper continues from where Kim (2010) has stopped. Hence, the model will be essentially the same, that is, the dependent variable in each equation consists of a deterministic trend and an error component. The deterministic trend is assumed to have a change in the slope, intercept or both. The error process has the common factor structure so that the cross equation correlation is of very strong form.

However, we modify the model in two important ways in this paper. First, we assume that the break date in each equation is a random draw from a common distribution and thus it can vary across equations. The identical break date assumption in Kim (2010) is only a special case of the current setup. Since the individual break date varies across equations, its mean is referred to as the common break date, and the main focus of paper is on the estimation of this mean.

The second modification specifies the break parameters (i.e. changes in the intercept and slope parameters) to be local to zero. Under this local to zero parameterization, the trend breaks are small enough not to interfere with the estimation of the common components but large enough for the estimation of the common break date. In particular, we take the principal component estimate analyzed by Bai and Ng (2002, 2004) and Bai (2003). We show that these estimates for the common components are consistent even if the breaks in the trend functions are ignored in the estimation procedure, and thus can effectively eliminate the strong cross sectional dependence when subtracted from the original observations. The common break date estimate obtained after the subtraction achieves a faster rate of convergence than the one obtained without the subtraction.

Another implication of the local break assumption is that our asymptotic results may not be suitable when the breaks are large. However, this does not diminish the usefulness of the new procedure. If the breaks are large, the common components estimates will not be consistent anymore and cannot remove the cross sectional dependence properly. However, subtraction of the common components estimates preserves the break date, and the common break date estimate from the data after subtraction still performs well, since the breaks are anyway large. Also, our Monte Carlo simulation results suggest that the new procedure offers meaningful improvements over a wide range of break parameters, although the advantage is especially pronounced for small to medium size breaks with which Kim's (2010) simple estimator may show less than mediocre performance.

The limiting distribution of the new break date estimate resembles that of Kim (2010) in that it is normal when only a slope change is allowed and somewhat non-standard when both intercept and slope changes are allowed. However, the biggest difference is that the limiting distribution of the common break date estimate depends on the distribution from which the individual break dates are drawn. Hence, this limiting distribution can be used with an assumption on the distribution of the individual break date, including the one that all break dates are identical.

We suggest a convenient way to form asymptotically valid confidence intervals based on the limiting distributions derived in the paper. The validity of the limiting distributions depends on the adequacy of the local to zero break assumption. When the breaks are indeed large, alternative limiting distributions can be derived. However, the confidence interval formed following our suggestion is asymptotically valid regardless of the magnitude of the breaks. Hence, empirical researchers do not need to know if the local to zero assumption is proper or not for a given panel of data.

The remainder of the paper is organized as follows. Section 2 presents the models and assumptions. Section 3 contains the details of the proposed estimation procedure and the main theoretical results including the asymptotic distribution of the common break date estimate. Section 4 shows some Monte Carlo experiment results. Section 5 offers a brief empirical illustration. Section 6 concludes. All proofs and technical derivations are collected in the appendix.

2 Models and Assumptions

We consider the models that are analyzed in Kim (2010). The dependent variable in each equation consists of a deterministic time trend and an error component:

$$y_{ti} = d_{ti} + u_{ti},\ (i = 1, \ldots, N \text{ and } t = 1, \ldots, T)$$

The deterministic trend is assumed to have a break and we consider three cases:

$$d_{ti} = \begin{cases} \mu_i + \beta_i t + \gamma_i B_{ti} & \text{Model I (Joint Broken Trend)} \\ \mu_i + \beta_i t + \theta_i C_{ti} + \gamma_i B_{ti} & \text{Model II (Local Disjoint Trend)} \\ \mu_i + \beta_i t + \theta_i C_{ti} & \text{Model III (Mean Shift)} \end{cases}$$

where

$$C_{ti} = \begin{cases} 0 & \text{if } t \le T_i \\ 1 & \text{if } t > T_i \end{cases} \text{ and } B_{ti} = \begin{cases} 0 & \text{if } t \le T_i \\ t - T_i & \text{if } t > T_i \end{cases}.$$

Note that the break dates T_i $i = 1, \ldots, N$, can be different across equations, but we assume that they are drawn from the same distribution.

Assumption 1 *For each i, the true break date T_i is such that $T_i = T_0 + \Delta T_i$ where ΔT_i is an integer valued random variable identically and independently distributed with zero mean and finite variance σ_b^2, and the break fraction $\lambda_0 = T_0/T \in [\pi, 1-\pi]$, $\pi \in (0, 1/2)$ is fixed for all T.*

The mean of the individual break date, T_0 will be referred to as the common break date, and is the main object of interest in this paper. Kim (2010) assumed that the break dates are exactly identical in all equations. This is equivalent to the case where $\sigma_b^2 = 0$ in our model, and thus the current assumption is more general. The motivation for this generalization is that even if all breaks are caused by one common event, the time lag from the onset of the cause to the actual occurrence of a trend change can vary across equations. Another implication of the above assumption is that all the breaks occur at the same time if they are viewed as a fraction of the entire time span, that is, $\sup_i |T_i/T - \lambda_0| \xrightarrow{p} 0$, as $(N, T) \to \infty$ and $N/T^2 \to 0$.

The trimming of the location of λ_0 by π is a simple device to ensure that the regressor matrix be of full column rank and π can be arbitrarily small in practice. Kim (2010) points out an identification issue in Model II that if $\theta_i = -\gamma_i$ or 0, there are two break dates that can generate exactly the same time trend. We assume in the following that θ_i is neither $-\gamma_i$ nor 0.

The error component u_{ti} is such that

$$u_{ti} = h_i' F_t + e_{ti} \tag{1}$$

where F_t is a $r \times 1$ vector of latent common factors, h_i is a factor loading and e_{ti} is an individual specific error. We make the following assumptions on the error component:

Assumption 2 *(i) The vector of common factors F_t is such that $F_t = C(L)w_t$ where $w_t \sim iid(0, I_r)$, $E||w_t||^4 < \infty$ and $C(L) = \sum_{j=0}^{\infty} C_j L^j$ with $\sum_{j=0}^{\infty} j\,||C_j|| < M$ and $\det(C(z)) \neq 0$ for all $|z| \leq 1$.*
(ii) For each equation i, the individual specific error e_{ti} is such that $e_{ti} = d_i(L)\varepsilon_{ti}$ where $d_i(L) = \sum_{j=0}^{\infty} d_{ij} L^j$ with $d_{i0} = 1$, $\sum_{j=0}^{\infty} j\,|d_{ij}| < M$ and $d_i(z) \neq 0$ for all $|z| \leq 1$. Furthermore, e_{ti} is independent across i and for each i, $\varepsilon_{ti} \sim iid(0, \sigma_i^2)$ where $\sigma_i^8 < M$.
(iii) F_t and e_{ti} are independent.

This set of assumptions is identical to the one in Kim (2010) except that it requires higher moments of w_t and ε_{ti} to exist for the estimation of common components. We also make the following assumptions for the break parameters and the factor loadings. Define $H = [h_1, \ldots, h_N]$, $\theta = (\theta_1, \ldots, \theta_N)$ and $\gamma = (\gamma_1, \ldots, \gamma_N)$.

Assumption 3 *(i) $\gamma = N^{-1/2}\dot{\gamma}$ and $\theta = N^{-1/2}\dot{\theta}$.*

(ii) $\theta\theta' \to \dot{A}_{\theta\theta} \neq 0$, $\gamma\gamma' \to \dot{A}_{\gamma\gamma} \neq 0$, $\theta\gamma' \to \dot{A}_{\theta\gamma}$, $N^{-1}HD\Sigma_\varepsilon D\dot{\gamma}' \to S_{H\dot{\gamma}}$, $N^{-1}HD\Sigma_\varepsilon DH' \to S_{HH} \neq 0$ and $\gamma D\Sigma_\varepsilon D\gamma' \to \dot{S}_{\gamma\gamma} \neq 0$, where $\Sigma_\varepsilon = diag\{\sigma_1^2, \ldots, \sigma_N^2\}$ and $D = diag\{d_1(1), \ldots, d_N(1)\}$.

(iii) $\max\{\dot{\theta}_1^2, \ldots, \dot{\theta}_N^2\} = O(1)$, $\max\{\dot{\gamma}_1^2, \ldots, \dot{\gamma}_N^2\} = O(1)$.

(iv) Let $\bar{\Xi}_1 = \lim_N N^{-1}\sum_{i=1}^N (\bar{\theta}_i^2 + \sigma_b^2 \dot{\gamma}_i^2)\Gamma_i$ and $\bar{\Xi}_2 = \lim_N N^{-1}\sum_{i=1}^N \bar{\gamma}_i^2 \Gamma_i$ where $\bar{\theta}_i$ is the i^{th} element of $(I - H'(HH')^{-1}H)\dot{\theta}'$, $\bar{\gamma}_i$ is the i^{th} element of $(I - H'(HH')^{-1}H)\dot{\gamma}'$ and Γ_i is any finite size matrix whose (p, q) element is the autocovariance of e_{ti} at $p - q$ lag.

(v) h_i is such that $N^{-1}H\dot{\gamma}' \to A_{H\dot{\gamma}} \neq 0$, $N^{-1}HH' \to A_{HH} \neq 0$, and $N^{-1}H\dot{\theta}' \to A_{H\dot{\theta}} \neq 0$, where $A_{H\gamma}$, A_{HH}, and $A_{H\dot{\theta}}$ are some fixed matrices.

(vi) $N^{-1}\sum_{i=1}^N \dot{\gamma}_i^4 \to \dot{A}_{\gamma\gamma\gamma\gamma}$, $N^{-1}\sum_{i=1}^N \dot{\gamma}_i^3 h_i \to A_{H\dot{\gamma}\dot{\gamma}\dot{\gamma}}$, and $N^{-1}\sum_{i=1}^N \dot{\gamma}_i^2 h_i h_i' \to A_{H\dot{\gamma}\dot{\gamma}H}$

The break parameters γ and θ are local to zero at rate $\sqrt{N}$. This is different from Kim (2010) where the break parameters are fixed constants.

Now, we write each equation in a matrix form as

$$\underset{(T\times 1)}{Y_i} = \underset{(T\times 1)}{d_i} + \underset{(T\times 1)}{U_i} = \underset{(T\times 1)}{d_i} + \underset{(T\times r)}{F}\underset{(r\times 1)}{h_i} + \underset{(T\times 1)}{E_i}$$

where $Y_i = (y_{i1}, \ldots, y_{iT})'$, $d_i = (d_{1i}, \ldots, d_{Ti})'$ and $U_i = (u_{1i}, \ldots, u_{Ti})'$. The entire system is written as

$$Y = d + U = d + FH + E$$

where $Y = [Y_1, \ldots, Y_N]$, $d = [d_1, \ldots, d_N]$, $U = [U_1, \ldots, U_N]$, $H = [h_1, \ldots, h_N]$, $F = [F_1, \ldots, F_T]'$, and $E = [E_1, \ldots, E_N]$.

3 Estimation of the Common Break Date

The common break date estimate analyzed in Kim (2010) is obtained by simply minimizing the sum of squared residuals for all permissible break dates. Denote by T_b a generic break date and by $\lambda = T_b/T$ a generic break fraction. Now, we define regressors. $\iota = (1, \ldots, 1)'$, $\tau = (1, \ldots, T)'$, $C(T_b) = (C_1(T_b), \ldots, C_T(T_b))'$, and $B(T_b) = (B_1(T_b), \ldots, B_T(T_b))'$ where

$$C_t(T_b) = \begin{cases} 0 & \text{if } t \leq T_b \\ 1 & \text{if } t > T_b \end{cases} \quad \text{and} \quad B_t(T_b) = \begin{cases} 0 & \text{if } t \leq T_b \\ t - T_b & \text{if } t > T_b \end{cases}.$$

Also, define X_{T_b} to be the collection of these regressors, that is,

$$X_{T_b} = \begin{cases} [\iota, \tau, B(T_b)] & \text{Model I} \\ [\iota, \tau, C, B(T_b)] & \text{Model II} \\ [\iota, \tau, C(T_b)] & \text{Model III} \end{cases}$$

Then, the simple break date estimate $\hat{T}_0$ is the date that minimizes the sum of squared residuals:

$$\hat{T}_0 = \arg\min_{T_b} \; SSR(T_b) \quad \text{and} \quad \hat{\lambda} = \hat{T}_0/T \tag{2}$$

where $SSR(T_b) = tr\,[Y'(I - P_{T_b})Y]$ and $P_{T_b} = X_{T_b}(X'_{T_b}X_{T_b})^{-1}X'_{T_b}$.

One of the main results in Kim (2010) is that common factors with factor loadings correlated with the slope parameters, in the sense that $A_{H\dot{\gamma}} \neq 0$ and $A_{H\dot{\theta}} \neq 0$, slow down the rate of convergence of the break date/fraction estimate. In particular, the rate of convergence in the presence of such common factors is only as fast as the one achieved with one time series.

The main idea of this paper is to use the fact that the local breaks specified in Assumption 3 are large enough for the purpose of break date/fraction estimation but small enough not to interfere with the estimation of the common components. Hence, we first estimate the common components with ignoring the existing trend breaks, and then subtract these common components estimates from the original observations to eliminate the cross equation dependence. The break date is estimated from these observations less the common components estimates. The exact procedure is as follows.

1. Take a difference and subtract the individual sample mean of the data: $\Delta^* y_{ti} = y_{ti} - y_{t-1i} - (y_{Ti} - y_{1i})/(T-1)$ for $t = 2, \ldots, T$ and $i = i, \ldots, N$. Let the data matrix after first differencing and demeaning be $\Delta^* Y$. That is, $\Delta^* Y = [\Delta^* y_{ti}]$, which is $(T-1) \times N$.

2. Estimate the common factors: Compute $\hat{f} = [\hat{f}_2, \ldots, \hat{f}_T]'$, the principal component estimate for $f = [f_2, \ldots, f_T]'$ where $f_t = F_t - F_{t-1} - (F_T - F_1)/(T-1)$. This is $\sqrt{T-1}$ times the r eigenvectors corresponding to the first r largest eigenvalues of the $(T-1) \times (T-1)$ matrix $\Delta^* Y \Delta^* Y'$. Let the partial sums of $\hat{f}_t$ be

$$\hat{F}_t = \sum_{s=2}^{t} \hat{f}_s, \; t = 2, 3, \ldots \tag{3}$$

   Then, $\hat{F} = [\hat{F}_2, \ldots, \hat{F}_T]'$ is the estimate for the common factor matrix $F = [F_2, \ldots, F_T]'$.

3. Estimate the factor loadings: Under the normalization $\hat{f}'\hat{f}/(T-1) = I_r$, the estimated loading matrix is $\hat{H}' = (T-1)^{-1}\Delta^* Y' \hat{f}$.

4. Subtract $\hat{F}\hat{H}$, the common components estimate from $Y = [y_{ti}]$ for $t = 2, \ldots, T$ and $i = i, \ldots, N$, the original series without the first observations.

5. Estimate the break date by minimizing the sum of the squared residuals over the possible break dates:

$$\widetilde{T}_0 = \arg\min_{T_b} \; SSR(T_b) \quad \text{and} \quad \widetilde{\lambda} = \widetilde{T}_0/T \tag{4}$$

   where $SSR(T_b) = tr\left[\hat{Y}'(I - P_{T_b})\hat{Y}\right]$ and $\hat{Y} = Y - \hat{F}\hat{H}$.

Because the estimation procedure described above ignores the existing break, the consistency results established in Bai and Ng (2004) are not directly applicable. In particular, Step 1 does not eliminate all deterministic components because of the breaking trends. However, we show that the estimated factors are consistent for the true factors scaled by a nonsingular rotation matrix R, if a deterministic term $\Delta\delta_{t,T}$, which is some linear combination of the neglected breaking trends, is subtracted from them. This result does not require the break parameters to be local to zero. Let $C_{N,T} = \min\{\sqrt{N}, \sqrt{T}\}$.

Lemma 1 *Let $\Delta^* d_t = (\Delta^* d_{t1}, \ldots, \Delta^* d_{tN})'$ with $\Delta^* d_{ti} = d_{ti} - d_{t-1i} - (d_{Ti} - d_{1i})/(T-1)$ and $\hat{f}_t$ be the principal component estimate for f_t defined in Step 2 above. Suppose that Assumptions 1 and 2 hold. Then, there exist a square matrix R with rank r and a $r \times 1$ vector $\Delta\delta_{t,T}$, each element of which is some linear combination of $\Delta^* d_t$, such that as N, $T \to \infty$,*

$$\frac{1}{T}\sum_{t=2}^{T} \left\| (\hat{f}_t - \Delta\delta_{t,T}) - R' f_t \right\|^2 = O_p\left(C_{N,T}^{-2}\right).$$

One implication of the above lemma is that the common factor estimate $\hat{f}_t$ includes a deterministic term $\Delta\delta_{t,T}$, so does $\hat{F}_t$, the partial sum of $\hat{f}_t$. Define $\delta_T = [\delta_{2,T}, \ldots, \delta_{T,T}]'$ with $\delta_{t,T} = \sum_{j=2}^{t} \Delta\delta_{j,T}$. Then, we naturally write the new data $\hat{Y}$ as

$$\hat{Y} = d + U - \hat{F}\hat{H} = \hat{d} + \hat{U} \tag{5}$$

where

$$\begin{aligned} \hat{d} &= d - \delta_T \hat{H} \\ \hat{U} &= E + FH - (\hat{F} - \delta_T)\hat{H} \end{aligned}$$

The decomposition that will play a key role is:

$$\begin{aligned} & SSR(T_b) - SSR(T_0) \\ = \ & tr\left[\hat{d}'(P_{T_0} - P_{T_b})\hat{d}\right] + 2\ tr\left[\hat{d}'(P_{T_0} - P_{T_b})\hat{U}\right] + tr\left[\hat{U}'(P_{T_0} - P_{T_b})\hat{U}\right] \\ \equiv \ & (\hat{X}\hat{X}) + 2(\hat{X}\hat{U}) + (\hat{U}\hat{U}) \end{aligned} \tag{6}$$

The consistency and the rate of convergence of the common break date estimate can be shown by expressing the orders of magnitude of the three terms $(\hat{X}\hat{X})$, $(\hat{X}\hat{U})$ and $(\hat{U}\hat{U})$ in terms of $|T_b - T_0|$. Then, the usual argument for consistency first supposes that the break date estimate is not consistent at a certain rate. Then, the term $(\hat{X}\hat{X})$, that is always positive, becomes of strictly greater order of magnitude than the other terms, and thus the inequality

$$SSR(\widetilde{T}_0) - SSR(T_0) \leq 0 \tag{7}$$

cannot hold with probability one as the sample size grows. Because the inequality in (7) must be true by definition, the supposition is a contradiction and the consistency follows.

In the appendix, we show that

$$(P_{T_0} - P_{T_b})\hat{d} = (P_{T_0} - P_{T_b})dM_{\hat{H}'} \tag{8}$$

where $M_{\hat{H}'} = I - \hat{H}'(\hat{H}\hat{H}')^{-1}\hat{H}$. Then, it follows that

$$\begin{aligned} (\hat{X}\hat{U}) &= tr\left[\hat{d}'(P_{T_0} - P_{T_b})(E + FH - (\hat{F} - \delta_T)\hat{H})\right] \\ &= tr\left[d'(P_{T_0} - P_{T_b})EM_{\hat{H}'}\right] + tr\left[d'(P_{T_0} - P_{T_b})FHM_{\hat{H}'}\right] \end{aligned} \tag{9}$$

In the above equation, the second term would not exist, if there are no common factors. Hence, as long as the estimate for factor loadings $\hat{H}$ is consistent at a fast enough rate so that the second term is of smaller order than the first, the effect of the common components can be eliminated asymptotically.

When the break parameters are local to zero, the neglected break is indeed small enough so that the factors and loadings are consistently estimated. The relevant results are stated in the next lemma.

Lemma 2 *Let $\hat{f}_t$ be the principal component estimate for f_t defined in Step 2 above. Suppose that Assumptions 1, 2, and 3 hold. Then, as N, $T \to \infty$,*

$$\frac{1}{T}\sum_{t=2}^{T}\left\|\hat{f}_t - R'f_t\right\|^2 = O_p\left(C_{N,T}^{-2}\right),$$

and, if $N/T^2 \to 0$,

$$\frac{1}{\sqrt{N}}(\hat{H} - R^{-1}H)\dot{\gamma}' = o_p(1) \quad \textit{and} \quad \frac{1}{\sqrt{N}}(\hat{H} - R^{-1}H)\dot{\theta}' = o_p(1)$$

We also need consistency of $\hat{F}_t$, the partial sum of $\hat{f}_t$, for F_t. Bai and Ng (2004) show a similar result with no neglected trend break (See their Lemma 2 and the subsequent comment). However, this convergence result is not tight enough for our purpose, since it is established without making use of the fact that F_t and e_{ti} do not have a unit root. The following lemma states the convergence of $\hat{F}_t$ in the presence of neglected trend breaks with assuming F_t has no unit root.

Lemma 3 *Let $\hat{F}_t$ be defined in (3) and $\{a_{t,T}\}$ be an array of constants such that*

$$\max_{1\le t\le T}|a_{t,T}| < M < \infty$$

for all T. Suppose that Assumptions 1, 2, and 3 hold. Then, there exists a square matrix R with rank r and $\delta_{t,T}$, each element of which is some linear combination of a constant, t, and $d_t = (d_{t1}, \ldots, d_{tN})'$, such that as N, $T \to \infty$,

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T} a_{t,T}(\hat{F}_t - RF_t - \delta_{t,T}) = O_p\left(C_{N,T}^{-1}\right)$$

Based upon Lemmas 1, 2 and 3, we derive the orders of magnitude of $(\hat{X}\hat{X})$, $(\hat{X}\hat{U})$ and $(\hat{U}\hat{U})$.

Lemma 4 *Under Assumptions 1, 2, and 3, we have for all generic T_b:*

	$(\hat{X}\hat{X})$	$(\hat{X}\hat{U})$	$(\hat{U}\hat{U})$
Model I	$\lvert T_b - T_0\rvert^2 O(T)$	$\lvert T_b - T_0\rvert O_p(T^{1/2})$	$\lvert T_b - T_0\rvert O_p(T^{-1}N)$
Model II	$\lvert T_b - T_0\rvert^3 O(1)$	$\lvert T_b - T_0\rvert^{3/2} O_p(1)$	$\lvert T_b - T_0\rvert^{1/2} O_p(T^{-1/2}N)$
Model III	$\lvert T_b - T_0\rvert O(1)$	$\lvert T_b - T_0\rvert^{1/2} O_p(1)$	$\lvert T_b - T_0\rvert^{1/2} O_p(T^{-1/2}N)$

Now, we present the first main result of this paper in the next theorem. It states the rate of convergence of the break date estimate $\widetilde{T}_0$ for each model as well as that of the simple break date estimate $\hat{T}_0$ for comparison purpose.

Theorem 1 *Let the break date estimates $\hat{T}_0$ and $\widetilde{T}_0$ be defined in (2) and (4). Suppose that Assumptions 1, 2, and 3 hold. Then, as $(T, N) \to \infty$, we have the following results.*

(i) In Model I (Joint Broken Trend), if $N/T \to 0 < \rho < \infty$,

$$|\widetilde{T}_0 - T_0| = O_p(T^{-1/2}) \quad \text{and} \quad |\hat{T}_0 - T_0| = O_p(T^{-1/2}N^{1/2}).$$

(ii) In Model II (Local Disjoint Trend), if $N^2/T = O(1)$,

$$|\widetilde{T}_0 - T_0| = O_p(1) \quad \text{and} \quad |\hat{T}_0 - T_0| = O_p(N^{1/3}).$$

(iii) In Model III (Mean Shift), if $N^2/T = O(1)$,

$$|\widetilde{T}_0 - T_0| = O_p(1) \quad \text{and} \quad |\hat{T}_0 - T_0| = O_p(N).$$

Remark 1 *In Model I, the additional condition $N/T \to 0 < \rho < \infty$ can be relaxed to $N/T = O(1)$, if all break dates are the same.*

The rate of convergence of the common break date estimate $\widetilde{T}_0$ does not depend on the number of equations N, while that of the simple break date estimate $\hat{T}_0$ is decreasing in N. When there is strong cross equation dependence, the break date estimate $\hat{T}_0$ becomes more and more imprecise in the sense that the rate of convergence gets slower as more equations are added into the system. One intuitive explanation for this is that, because the break parameters and the factor loadings vary together, that is, $A_{H\dot{\gamma}} \neq 0$ and $A_{H\dot{\theta}} \neq 0$, the trend breaks are covered up by the common factors similarly in all equations. Hence, each additional equation does not deliver much information on the common break date whereas the break parameters shrink to zero making the break date estimation more difficult. On the other hand, subtraction of the consistent common components estimates from the original data essentially eliminates the strong cross equation correlation and makes each equation more informative about the break date, thereby allowing the break date estimate to keep its precision despite the shrinking break parameters. This is exactly the benefit of using $\widetilde{T}_0$ over $\hat{T}_0$. In the next theorem, we state the limiting distribution of $\widetilde{\lambda} = \widetilde{T}_0/T$.

Theorem 2 *Suppose that Assumptions 1, 2, and 3 hold. Define*

$$
\begin{aligned}
\bar{A}_{\theta\theta} &= \lim_N N^{-1}\dot{\theta} M_{H'}\dot{\theta}' = \dot{A}_{\theta\theta} - A'_{H\dot{\theta}} A^{-1}_{HH} A_{H\dot{\theta}} \\
\bar{A}_{\gamma\gamma} &= \lim_N N^{-1}\dot{\gamma} M_{H'}\dot{\gamma}' = \dot{A}_{\gamma\gamma} - A'_{H\dot{\gamma}} A^{-1}_{HH} A_{H\dot{\gamma}} \\
\bar{A}_{\gamma\theta} &= \lim_N N^{-1}\dot{\gamma} M'_{H'}\dot{\theta} = \dot{A}_{\gamma\theta} - A'_{H\dot{\gamma}} A^{-1}_{HH} A_{H\dot{\theta}} \\
\bar{S}_{\gamma\gamma} &= \lim_N N^{-1}\dot{\gamma} M_{H'} D\Sigma_\varepsilon D M_{H'}\dot{\gamma}' \\
&= \dot{S}_{\gamma\gamma} - A'_{H\dot{\gamma}} A^{-1}_{HH} S_{H\dot{\gamma}} - S'_{H\dot{\gamma}} A^{-1}_{HH} A_{H\dot{\gamma}} + A'_{H\dot{\gamma}} A^{-1}_{HH} S_{HH} A^{-1}_{HH} A_{H\dot{\gamma}} \\
Q_{\gamma\gamma\gamma\gamma} &= \dot{A}_{\gamma\gamma\gamma\gamma} - 2A'_{H\dot{\gamma}\dot{\gamma}\dot{\gamma}} A^{-1}_{HH} A_{H\dot{\gamma}} + A'_{H\dot{\gamma}} A^{-1}_{HH} A_{H\dot{\gamma}\dot{\gamma}H} A^{-1}_{HH} A_{H\dot{\gamma}}
\end{aligned}
$$

where $M_{H'} = I_N - H'(HH')^{-1}H$. *Then, as* $(T, N) \to \infty$, *we have the following results.*

(i) Model I (Joint Broken Trend): If $N/T \to 0 < \rho < \infty$,

$$
T^{3/2}(\widetilde{\lambda} - \lambda_0) \xrightarrow{d} N\left(0, \frac{4\bar{S}_{\gamma\gamma}}{(1-\lambda_0)\lambda_0 \bar{A}^2_{\gamma\gamma}} + \frac{\sigma_b^2 Q_{\gamma\gamma\gamma\gamma}}{\rho \bar{A}^2_{\gamma\gamma}}\right)
$$

(ii) Model II (Local Disjoint Trend) and Model III (Mean Shift): Let $\bar{\Xi}_1$ *and* $\bar{\Xi}_2$ *be as defined in Assumption 3. Then,* $N_1 = (N_{-m+1,1}, \ldots, N_{m,1})'$ *and* $N_2 = (N_{-m+1,2}, \ldots, N_{m,2})'$ *are multivariate normal such that* $N_1 = B_1 W$ *and* $N_2 = B_2 W$ *with* $W \sim N(0, I_{2m})$, $\bar{\Xi}_1 = B_1 B_1'$ *and* $\bar{\Xi}_2 = B_2 B_2'$.

Define the stochastic process $V^*(m)$ *to be such that* $V^*(0) = 0$, $V^*(m) = V_1(m)$ *for* $m < 0$ *and* $V^*(m) = V_2(m)$ *for* $m > 0$. *Then, if* $N^2/T \to 0$,

$$
T(\widetilde{\lambda} - \lambda_0) \xrightarrow{d} m_I^\infty = \arg\min_m V^*(m)
$$

where, for Model II,

$$
\begin{aligned}
V_1(m) &= \sum_{k=m+1}^{0} \left[\left(\bar{A}_{\theta\theta} + \sigma_b^2 \dot{A}_{\gamma\gamma}\right) + \bar{A}_{\gamma\gamma}k^2 + 2\bar{A}_{\gamma\theta}k\right][1 - 2P(\Delta T_i + 1 \le k)] \\
&\quad -2\sum_{k=m+1}^{0} (N_{1,k} + kN_{2,k}) \quad \text{for } m = -1, -2, \ldots \\
V_2(m) &= \sum_{k=1}^{m} \left[(\bar{A}_{\theta\theta} + \sigma_b^2 \dot{A}_{\gamma\gamma}) + \bar{A}_{\gamma\gamma}k^2 + 2\bar{A}_{\gamma\theta}k\right][1 - 2P(\Delta T_i \ge k)] \\
&\quad +2\sum_{k=1}^{m} (N_{1,k} + kN_{2,k}) \quad \text{for } m = 1, 2, \ldots
\end{aligned}
$$

and, for Model III,

$$V_1(m) = \sum_{k=m+1}^{0} \bar{A}_{\theta\theta} \left[1 - 2P(\Delta T_i + 1 \leq k)\right] - 2 \sum_{k=m+1}^{0} N_{1,k} \quad \textit{for } m = -1, -2, \ldots$$

$$V_2(m) = \sum_{k=1}^{m} \bar{A}_{\theta\theta} \left[1 - 2P(\Delta T_i \geq k)\right] + 2 \sum_{k=1}^{m} N_{1,k} \quad \textit{for } m = 1, 2, \ldots.$$

These limiting distributions not only involve various model parameters but also depend on the distribution of the individual break date. This is very unfortunate because the distribution of the individual break date cannot be estimated from the estimates of the individual break dates. The main reason is that the individual break date estimates are obtained from only one time series and they do not converge fast enough. Nevertheless, these limiting distributions can still be used with an assumption on the distribution of the individual break date, including the one that all break dates are identical.

Regression Coefficients The regression coefficients (μ_i, β_i, θ_i, γ_i) can be estimated by applying the least squares procedure to each equation based on the estimated individual break date, the date that minimizes the sum of squared residuals of each individual series. A standard result pertaining to non-trending variables is that the least squares estimate has the same limiting distribution whether one uses the true break date or the estimated one. This result obviously applies to Model III.

For Models I and II, Theorem 6 of Perron and Zhu (2005) shows that this type of invariance does not hold. More precisely, the least squares estimate for (μ_i, β_i, γ_i) in Model I has two different limiting distributions depending on whether the true or the estimated break date is used, although the rate of convergence remains the same in both cases. The exact expressions for these limiting distributions can be found in Theorem 6 of Perron and Zhu (2005). In Model II, the least squares estimate for (μ_i, β_i, γ_i) has the same limiting distribution in both cases. In fact, this limiting distribution is the same as the one obtained under Model I with the estimated break date. In contrast, the least squares estimate for θ_i is not consistent in Model II with the estimated break date while it is consistent at rate $\sqrt{T}$ with the true break date. An important consequence is that an asymptotically valid confidence interval for the individual break date cannot be formed since the limiting distribution of the individual break date estimate in Perron and Zhu (2005) depends on θ_i.

Now, suppose that the break dates are identical in all equations. Then, the regression coefficients (μ_i, β_i, θ_i, γ_i) can be estimated based on the common break date estimate $\widetilde{T}_0$ instead of the individual break date estimate. In this case, it can be shown in Model I that the limiting distribution of the least squares estimate for $(\mu_i, \beta_i, \gamma_i)$ using $\widetilde{T}_0$ becomes the same as the one under the true break date due to the faster rate of convergence of $\widetilde{T}_0$. In other words, the common break date estimate $\widetilde{T}_0$ can be treated as if it were the true break date at least for the purpose of the regression coefficients estimation.

Using Theorem 6 in Perron and Zhu (2005), it can be shown that the least squares estimate for θ_i obtained based on $\widetilde{T}_0$, say $\widetilde{\theta}_i$, is such that

$$\widetilde{\theta}_i - \theta_i = \gamma_i m_I^{\infty} + o_p(1),$$

which means that the inconsistency of $\widetilde{\theta}_i$ is due to a term that is multiplicative of the slope change parameter γ_i. It follows that $\widetilde{\theta}_i - \theta_i = O_p(N^{-1/2})$ under the local to zero assumption. Given that we require $N^2/T \to 0$, this should be viewed as considerably slower than the $\sqrt{T}$ rate obtained with the true break date. Also, because $\dot{\theta}_i = \sqrt{N}\theta_i$, $\bar{A}_{\theta\theta}$ cannot be consistently estimated, on which the limiting distribution of $\widetilde{T}_0$ depends. Therefore, the lack of asymptotically valid confidence intervals for T_0 in Model II remains unsolved even with a large panel of data.

Confidence Intervals for the Common Break Date When the break dates are the same in all equations or the distribution from which the break dates are drawn is known, the limiting distributions for Models I and III provided in Theorem 2 can be used to form a confidence interval for T_0. Note that the intercept and slope change parameters of $\hat{Y}$ in (5) are $\theta M_{\hat{H}'}$ and $\gamma M_{\hat{H}'}$ instead of θ and γ. Hence, $\bar{A}_{\theta\theta}$ and $\bar{A}_{\gamma\gamma}$ can be estimated relatively easily. Let $\widetilde{\gamma}$ and $\widetilde{\theta}$ be the least squares estimates of the slope and intercept change parameters of $\hat{Y}$ respectively in Models I and III. Then, we have

$$\widetilde{\gamma}\widetilde{\gamma}' = \frac{1}{N}\dot{\gamma}M_{\hat{H}'}\dot{\gamma}' + o_p(1) = \bar{A}_{\gamma\gamma} + o_p(1) \tag{10}$$

$$\widetilde{\theta}\widetilde{\theta}' = \frac{1}{N}\dot{\theta}M_{\hat{H}'}\dot{\theta}' + o_p(1) = \bar{A}_{\theta\theta} + o_p(1). \tag{11}$$

For Model I, $\bar{S}_{\gamma\gamma} = \lim_N N^{-1}\dot{\gamma}M_{H'}D\Sigma_\varepsilon DM_{H'}\dot{\gamma}'$ should be estimated, which involves the longrun variances of all N individual specific errors. It can be done by estimating N longrun variances separately, but a more convenient method is available when we realize $\bar{S}_{\gamma\gamma}$ is the longrun variance of the cross sectional sum of the error components weighted by the slope change parameters. From (5), the weighted cross sectional sum of the error components is given by

$$\hat{U}M_{\hat{H}'}\gamma' = EM_{\hat{H}'}\gamma' + FHM_{\hat{H}'}\gamma' = \frac{1}{\sqrt{N}}EM_{\hat{H}'}\dot{\gamma}' + \frac{1}{\sqrt{N}}FHM_{\hat{H}'}\dot{\gamma}'$$

The second term is irrelevant since $N^{-1/2}HM_{\hat{H}'}\dot{\gamma}' = o_p(1)$ as shown in (A.17). Note that the i^{th} element of the first term is

$$\frac{1}{\sqrt{N}}\sum_i e_{ti}\bar{\gamma}_i + o_p(1)$$

where $\bar{\gamma}_i$ is the i^{th} element of $(I - H'(HH')^{-1}H)\dot{\gamma}'$, and its longrun variance is $\bar{S}_{\gamma\gamma}$. In practice, conditional on $\widetilde{T}_0$, detrend $\hat{Y}$. Let $\widetilde{u}_{ti}$ be the corresponding residual. Then, compute the longrun variance of the weighted sum

$$\sum_{i=1}^{N}\widetilde{u}_{ti}\widetilde{\gamma}_i \tag{12}$$

where $\widetilde{\gamma}_i$ is the least squares estimate of the slope change parameter in the i^{th} column of $\hat{Y}$. This longrun variance estimate is a natural estimate for $\bar{S}_{\gamma\gamma}$.

For Model III, $\bar{\Xi}_1 = \lim_N N^{-1}\sum_{i=1}^{N}\bar{\theta}_i^2\Gamma_i$ should be estimated. Using a similar argument to Model I, it can be estimated as the covariance matrix of the weighted sum

$$\sum_{i=1}^{N}\widetilde{u}_{ti}\widetilde{\theta}_i \tag{13}$$

where $\widetilde{\theta}_i$ is the least squares estimate of the intercept change parameter in the i^{th} column of $\hat{Y}$. Upon the estimate of $\bar{\Xi}_1$, the process $V^*(m)$ can be simulated and the quantiles of m_I^{∞} can be found.

Large Trend Breaks Since the asymptotic results presented above depend on the local to zero specification for the break parameters, they may not be suitable when the trend breaks are large. However, it is important to note that Lemma 1 is valid regardless of the magnitudes of breaks, so are the expressions in (6), (8) and (9). The factor loadings estimate $\hat{H}$ is not consistent when the breaks are large. Then, we can see from (6), (8) and (9) that the subtraction of the common components estimates does not eliminate the cross sectional dependence but it does not alter the true break date either. In other words, the subtraction step yields another set of panel data which has a common trend break exactly at the same date as before but still has strong cross equation dependence. In that case, the common break date estimate $\widetilde{T}_0$ performs similarly to the simple estimate $\hat{T}_0$. The asymptotic properties of $\widetilde{T}_0$ should be analyzed in the presence of strong cross equation dependence, now. Kim (2010) provides the relevant results with assuming all break dates are the same. Suppose that

$$\begin{aligned}\widetilde{A}_{H\gamma} &= p\lim_{N,T}\frac{1}{N}HM_{\hat{H}'}\gamma' \\ \widetilde{A}_{\gamma\gamma} &= p\lim_{N,T}\frac{1}{N}\gamma M_{\hat{H}'}\gamma' \\ \widetilde{A}_{\theta\theta} &= p\lim_{N,T}\frac{1}{N}\theta M_{\hat{H}'}\theta'\end{aligned}$$

Then, for Model I,

$$T^{3/2}(\widetilde{\lambda}-\lambda_0)\xrightarrow{d}N\left(0,\frac{4}{(1-\lambda_0)\lambda_0 A_{\gamma\gamma}^2}\widetilde{A}'_{H\gamma}C(1)C(1)\widetilde{A}_{H\gamma}\right)$$

where $C(1)$ is the longrun variance of the common factors as defined in Assumption 2. Note that $\widetilde{\gamma}\widetilde{\gamma}'$ in (10) now gives

$$\widetilde{\gamma}\widetilde{\gamma}'\approx N\cdot\widetilde{A}_{\gamma\gamma}.$$

Also, the cross equation sum of the error components weighted by the slope change parameters is such that

$$\begin{aligned}\hat{U}M_{\hat{H}'}\gamma' &= EM_{\hat{H}'}\gamma'+FHM_{\hat{H}'}\gamma' \\ &= EM_{\hat{H}'}\gamma'+N\cdot F\widetilde{A}_{H\gamma}.\end{aligned}$$

Here the first term is irrelevant since it is only a sum of N independent terms and the longrun variance estimated from the weighted sum in (12) is asymptotically $N^2\cdot\widetilde{A}'_{H\gamma}C(1)C(1)\widetilde{A}_{H\gamma}$. Therefore, if a confidence interval is formed as illustrated for the local break case, it is still asymptotically valid even for large breaks[1]. Now suppose that

$$\begin{aligned}\widetilde{A}_{H\theta} &= p\lim_{N,T}\frac{1}{N}HM_{\hat{H}'}\theta' \\ \widetilde{A}_{\theta\theta} &= p\lim_{N,T}\frac{1}{N}\theta M_{\hat{H}'}\theta'\end{aligned}$$

and define the stochastic process $S^*(m)$ such that $S^*(0)=0$, $S^*(m)=S_1(m)$ for $m<0$ and $S^*(m)=S_2(m)$ for $m>0$. Then, for Model III,

$$T(\hat{\lambda}-\lambda_1)\xrightarrow{d}m_\infty=\arg\min_m S^*(m)$$

where

$$\begin{aligned}S_1(m) &= \widetilde{A}_{\theta\theta}\,|m|-2\sum_{k=m+1}^{0}F_t'\widetilde{A}_{H\theta},\ m=-1,-2,\ldots \\ S_2(m) &= \widetilde{A}_{\theta\theta}m+2\sum_{k=1}^{m}F_t'\widetilde{A}_{H\theta},\ m=1,2,\ldots.\end{aligned}$$

[1]Note the scaler N will cancel out in the expression of the asymptotic variance.

This result means that the limiting distribution of the break date estimate depends on the distribution of common factors. This is an important difference from the local break case where the limiting distribution depends on normal variates only. Nevertheless, a practical choice is to assume F_t is multivariate normal. In that case, the cross equation sum of the error components weighted by the intercept change parameter is such that

$$\hat{U}M_{\hat{H}'}\theta' = EM_{\hat{H}'}\theta' + N \cdot F\widetilde{A}_{H\theta}.$$

Here the first term is again irrelevant asymptotically and the covariance matrix estimated from the weighted sum in (13) asymptotically corresponds to that of $N \cdot F\widetilde{A}_{H\theta}$. Also, $\widetilde{\theta\theta}'$ in (11) is such that

$$\widetilde{\theta\theta}' \approx N \cdot \widetilde{A}_{\theta\theta}$$

Therefore, the procedure to simulate the process $V^*(m)$ will essentially simulate N times the process $S^*(m)$, and the confidence interval will remain asymptotically valid even for large trend breaks if the common factors are multivariate normal.

4 Monte Carlo Simulation

The first experiment is to compare the two break date estimates, $\widetilde{T}_0$ and $\hat{T}_0$ defined in (2) and (4). The data is generated according to the models described in Section 2:

$$\begin{aligned} y_{ti} &= d_{ti} + u_{ti}, \; (i = 1, \ldots, N \text{ and } t = 1, \ldots T) \\ u_{ti} &= h_i' F_t + e_{ti}. \end{aligned}$$

The time dimension T is set at 100, 200, 300, and 500, and the cross section dimension N is set at 20, 50, and 100. The common factor is such that $F_t = 0.6F_{t-1} + w_t$ with $w_t \sim iid\ N(0,1)$ and the factor loading h_i is independently drawn from $U(0,2)$. The pre-break intercepts and slopes are set at zero ($\mu_i = \beta_i = 0$, for all i), since the Monte Carlo results are exactly invariant to these parameter values. More specifically, the deterministic component d_{ti} is specified to be

$$d_{ti} = \begin{cases} \gamma_i B_{ti} & \text{Model I} \\ \theta_i C_{ti} + \gamma_i B_{ti} & \text{Model II} \\ \theta_i C_{ti} & \text{Model III.} \end{cases}$$

The mean break date T_0 is always $0.5T$. The individual break date $T_i = T_0 + \Delta T_i$ and ΔT_i is drawn from $N(0,2)$ and rounded to the nearest integer. In Model I, γ_i is set at 0.1, 0.3, 0.5, and 0.7 for all i. In Model II, θ_i and γ_i are the same and are set at 0.1, 0.4, 0.7, and 1.4 for all i.

In Model III, θ_i is set at 1.0, 1.4, 2.0, and 5.0 for all i. The individual specific error e_{ti} is mildly autocorrelated. $e_{ti} = \rho_i e_{t-1i} + \varepsilon_{ti}$, $\varepsilon_{ti} \sim iid\ N(0,1)$ where ρ_i is drawn from $U(0, 0.5)$. For each N value, one set of $\{(h_i, \Delta T_i, \rho_i),\ i = 1, \ldots, N\}$ is kept for all T values and Monte Carlo repetitions. In all experiments, the number of replications is 2,000. The root mean squared errors (RMSE) of $\widetilde{T}_0$ and $\hat{T}_0$ are reported in Table 1.

Before we discuss the results, we note that the break parameters are selected ex post so that we can cover from small to large breaks. By small breaks, we mean those breaks with which the RMSE of the simple break date estimate $\hat{T}_0$ is above 15% of the time span T so that the break date estimate may not be informative enough. By large breaks, we mean those breaks with which the RMSE of the simple break date estimate $\hat{T}_0$ is less than two so that the break date is estimated extremely precisely even without eliminating the cross sectional dependence.

A few observations are noteworthy. First, the RMSE of $\widetilde{T}_0$ is always smaller than that of $\hat{T}_0$, as expected from the asymptotic result in the previous section. The difference is especially pronounced for small breaks. The asymptotic result showing the superiority of $\widetilde{T}_0$ depends on the small break assumption, but this simulation result shows that $\widetilde{T}_0$ still provides meaningful improvement even with fairly large breaks. Second, the RMSE of $\widetilde{T}_0$ is clearly decreasing as N increases, while no such pattern appears for $\hat{T}_0$. This is because the fixed break parameters in the simulation actually correspond to larger breaks as N increases from the standpoint of the asymptotic framework. The rate of convergence of $\widetilde{T}_0$ is not dependent of N in Theorem 1, but the RMSE of $\widetilde{T}_0$ is decreasing as we provide larger breaks. On the other hand, the RMSE of $\hat{T}_0$ remains very similar as N increases, because the deteriorating performance of $\hat{T}_0$ due to its rate of convergence decreasing in N is compensated by larger breaks. Third, both the common break date estimates, $\widetilde{T}_0$ and $\hat{T}_0$ perform better in Model I than in II or III. This is a reflection of the faster rate of convergence of these estimates in Model I.

The second experiment is to see the finite sample performances of the confidence intervals formed following the illustrations in the previous section. The data is generated the same as in the first experiment except that the break dates are identical in all equations. For Model I, the longrun variance is estimated by applying a heteroskedasticity and autocorrelation consistent covariance matrix estimate with the Quadratic Spectral window where the bandwidth parameter is selected using Andrews's (1991) data dependent method with AR(1) approximation. For Model II, the covariance matrix is estimated assuming that the weighted sum of the errors has an AR(1) structure. For both Models I and III, the confidence interval is forced to have a length of at least two periods.

The coverage rates and the relative average lengths are reported in Table 2. The numbers in parenthesis stand for the ratio of the average lengths of the confidence intervals for $\hat{T}_0$ and $\widetilde{T}_0$. For example, when $\gamma_i = 0.1$, $T = 100$ and $N = 20$ in Model I, (2.70) means that the average length of the confidence interval for $\hat{T}_0$ is 2.7 times larger than that for $\widetilde{T}_0$. A few observations are noteworthy again. First, the coverage rates for $\widetilde{T}_0$ are always greater than their counterparts for $\hat{T}_0$, while the average lengths are far shorter, which shows a clear advantage of using $\widetilde{T}_0$. However, this does not mean that the distribution of $\widetilde{T}_0$ is better approximated than that of $\hat{T}_0$. In fact, the distributions of the two estimates are approximated using exactly the same tool. The better coverage rates are simply a consequence of more precision obtained through elimination of the cross sectional dependence. Second, the coverage rates are actually greater than the nominal 95% rate for large breaks. This is due to the fact that we forced the intervals to be at least two periods long. When the reported coverage rates are one in the table, the associated average length of confidence intervals are very close to two.

5 Empirical Illustration

A common trend break is estimated from the personal income data obtained from the Regional Economic Information System, Bureau of Economic Analysis, U.S. Department of Commerce. The data spans from 1960.I to 2009.I and covers 51 U.S. states. Hence, $T = 197$ and $N = 51$. We assume that the personal income follows Model I. In order to decide the number of common factors, we used IC_{p1}, IC_{p2}, and IC_{p3}, the three information criteria suggested by Bai and Ng (2002) with the maximum number of factors being 13. The first and third criteria selected 11 and 13 respectively while the second one selected 3. Hence, we are somewhat inclined towards large numbers of common factors, but we will report the estimated break date for a range of numbers of common factors. Table 3 below reports the results.

The estimated break date in the U.S. aggregate personal income is 1987.Q3 with the confidence interval being from 1985.Q2 to 1989.Q2. From the panel of state by state personal income, the simple common break date estimate $\hat{T}_0$ is 1985.Q4, which still in the mid 80's. In fact, the two confidence intervals include the other estimate. However, the picture is completely different when the common components are eliminated. When the number of specified common factors is seven or greater, the common break date estimate $\widetilde{T}_0$ is in the mid 70's, which corresponds to the date of productivity slowdown due to the oil shock. Also, note that the confidence interval are tighter when we specify large numbers of common factors.

Table 3. Common Trend Break in the U.S. Personal Income, 1960.I$\sim$2009.I

			Estimated Break Date	95% Confidence Interval
U.S. aggregate Personal Income			1987.Q3	[1985.Q2, 1989.Q2]
State by State Personal Income	$\hat{T}_0$		1985.Q4	[1984.Q1, 1987.Q3]
	$\widetilde{T}_0$	$r = 3$	1983.Q4	[1982.Q4, 1984.Q4]
		5	1982.Q3	[1981.Q2, 1983.Q4]
		7	1974.Q4	[1973.Q2, 1976.Q2]
		9	1974.Q3	[1973.Q1, 1976.Q1]
		11	1973.Q2	[1973.Q1, 1973.Q3]
		13	1975.Q3	[1975.Q2, 1975.Q4]

6 Conclusion

This paper analyzes the issues of estimating a common local break in time trends for large panels. A novel feature of this paper is that we model the break parameters to be local to zero. Then, the common components can be consistently estimated even if the broken trends are ignored in the estimation procedure. These common components estimates remove the cross sectional dependence when subtracted from the original observations. Hence the common break date estimate obtained after the subtraction achieves a faster rate of convergence than the one obtained without the subtraction. Another novel feature is that the break date in each equation is assumed to be a draw from a common distribution and thus varies across equations. This is a generalization from the identical break date assumption used in Kim (2010). Randomly drawn break date does not change the rate of convergence of the proposed common break date estimate. However, the limiting distribution of the common break date estimate depends on the individual break date distribution. This is an unattractive feature for practitioners because the individual break date distribution cannot be estimated from the estimates of the individual break dates due to their slow rate of convergence. Nevertheless, the limiting distribution of the common break date estimate can be used to form confidence intervals if the individual break date distribution is known.

References

[1] Andrews, D.W.K. (1991) Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, Econometrica 59, 817-858.

[2] Bai, J. (2003) Inferential Theory for Factor Models of Large Dimensions. Econometrica 71, 135-171.

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

[4] Bai, J. and S. Ng (2004) A Panic Attack on Unit Roots and Cointegration. Econometrica 72, 1127-1177.

[5] Hannan, E. J. (1970) Multiple Time Series. John Wiley and Sons, Inc.

[6] Kim, D. (2010) Estimating a Common Deterministic Time Trend Break in Large Panels with Cross Sectional Dependence. Working Paper, Department of Economics, University of Virginia.

[7] Perron, P. and X. Zhu (2005) Structural Breaks with Deterministic and Stochastic Trends. Journal of Econometrics 129, 65-119.

[8] Phillips, P. C. B. and H. R. Moon (1999) Linear Regression Limit Theory for Nonstationary Panel Data. Econometrica 67, 1057-1111.

<Appendix>

We assume there are $T+1$ observations ($t = 0, 1, \ldots, T$) for simplicity. Recall that $\Delta^* y_{ti} = y_{ti} - y_{t-1i} - T^{-1}(y_{Ti} - y_{0i})$ for $t = 1, \ldots, T$ and $i = i, \ldots, N$. Also, let $f_t = F_t - F_{t-1} - T^{-1}(F_T - F_0)$, $z_{ti} = e_{ti} - e_{t-1i} - T^{-1}(e_{Ti} - e_{0i})$ and $\Delta^* d_{ti} = C^*_{ti}\gamma_i$ for Model I, $D^*_{ti}\theta_i + C^*_{ti}\gamma_i$ for Model II and $D^*_{ti}\theta_i$ for Model III, where $D^*_{ti} = D_t(T_i) - T^{-1}$, $D_t(T_i) = 1$ if $t = T_i + 1$, 0, otherwise, and $C^*_{ti} = C_{ti} - 1 + \lambda_i$ with $\lambda_i = T_i/T$. Then we can write $\Delta^* y_{ti} = \Delta^* d_{ti} + f_t' h_i + z_{ti}$, and $\Delta^* Y = [\Delta^* y_{ti}]$ and $\Delta^* d = [\Delta^* d_{ti}]$, which are $T \times N$. $\mathcal{E}(\cdot)$ denotes the mathematical expectation.

Let V_{NT} be the $r \times r$ diagonal matrix of the first r largest eigenvalues of $(NT)^{-1}\Delta^* Y \Delta^* Y'$ in decreasing order and $R' = V_{NT}^{-1}\hat{f}'(fH + \Delta^* d)H'/(NT)$. Then, $\hat{f}'\hat{f} = O_p(T)$, $\hat{f}'f = O_p(T)$, $N^{-1}\hat{H}\hat{H}' = V_{N,T} = O_p(1)$, and $R = O_p(1)$. These results are rather straightforward and will be used without proofs. Furthermore, note that $f'zH' = \sum_{t=1}^T f_t z_t' H' = O_p(\sqrt{TN})$ and $f'z\dot{\gamma}' = \sum_{t=1}^T f_t z_t' \dot{\gamma}' = O_p(\sqrt{TN})$.

>From the equality that $\hat{f} = (NT)^{-1}\Delta^* Y \Delta^* Y' \hat{f} V_{NT}^{-1}$, we have the following key equality.

$$\hat{f}_t - R' f_t = V_{NT}^{-1}\frac{1}{NT}\begin{bmatrix} \hat{f}'\Delta^* d\Delta^* d_t + \hat{f}' f H \Delta^* d_t + \hat{f}' z \Delta^* d_t \\ +\hat{f}'\Delta^* d z_t + \hat{f}' f H z_t + \hat{f}' z H' f_t + \hat{f}' z z_t \end{bmatrix} \tag{A.1}$$

Proof of Lemma 1: Collect the first three terms in (A.1) and let

$$\begin{aligned} \Delta\delta_{t,T} &= \frac{1}{NT}V_{NT}^{-1}\left[\hat{f}'\Delta^* d\Delta^* d_t + \hat{f}' f H \Delta^* d_t + \hat{f}' z \Delta^* d_t\right] \\ &= \frac{1}{NT}V_{NT}^{-1}\hat{f}'\left[\Delta^* d + fH + z\right]\Delta^* d_t \\ &= \frac{1}{NT}V_{NT}^{-1}\hat{f}'\Delta^* Y \Delta^* d_t \\ &= \frac{1}{N}V_{NT}^{-1}\hat{H}\Delta^* d_t = (\hat{H}\hat{H}')^{-1}\hat{H}\Delta^* d_t \end{aligned} \tag{A.2}$$

Note that

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^T \left\|\hat{f}_t - R' f_t - \Delta\delta_{t,T}\right\|^2 &\leq \left\|V_{NT}^{-1}\right\|^2 \frac{1}{N^2 T^3}\sum_{t=1}^T \left\|\hat{f}'\Delta^* d z_t + \hat{f}' f H z_t + \hat{f}' z H' f_t + \hat{f}' z z_t\right\|^2 \\ &\leq 2\left\|V_{NT}^{-1}\right\|^2 \frac{1}{N^2 T^3}\sum_{t=1}^T \left[\left\|\hat{f}'\Delta^* d z_t\right\|^2 + \left\|\hat{f}' f H z_t + \hat{f}' z H' f_t + \hat{f}' z z_t\right\|^2\right] \\ &= 2\left\|V_{NT}^{-1}\right\|^2 \frac{1}{N^2 T^3}\sum_{t=1}^T \left\|\hat{f}'\Delta^* d z_t\right\|^2 + O_p(C_{N,T}^{-2}) \end{aligned}$$

where the last equality is shown in Bai and Ng (2002). Then, without loss of generality, for Model II,

$$\begin{aligned}
\left\| \hat{f}' \Delta^* dz_t \right\| &= \left\| \sum_{s=1}^{T} \hat{f}_s \sum_{i=1}^{N} \left(C_{si}^* \gamma_i + D_{si}^* \theta_i \right) z_{ti} \right\| \\
&\leq \left[\sum_{s=1}^{T} \left\| \hat{f}_s \right\|^2 \right]^{1/2} \left[\sum_{s=1}^{T} \left(\sum_{i=1}^{N} \left(C_{si}^* \gamma_i + D_{si}^* \theta_i \right) z_{ti} \right)^2 \right]^{1/2} \\
&= O_p(\sqrt{T}) O_p(\sqrt{TN}) \\
&= O_p(T\sqrt{N}) \qquad \text{(A.3)}
\end{aligned}$$

because

$$\mathcal{E}\left(\sum_{i=1}^{N} \left(C_{si}^* \gamma_i + D_{si}^* \theta_i \right) z_{ti}\right)^2 = \sum_{i=1}^{N} \left(C_{si}^* \gamma_i + D_{si}^* \theta_i \right)^2 \mathcal{E} z_{ti}^2 = O(N),$$

where γ_is and θ_is are treated as fixed constants. Therefore,

$$\frac{1}{N^2 T^3} \sum_{t=1}^{T} \left\| \hat{f}' \Delta^* dz_t \right\|^2 = O_p(N^{-1})$$

and the claim in the lemma follows.

Derivation of (8): We show this equation for Model II only. The other models are only a special case. From (A.2), $\Delta \delta_{t,T} = (\hat{H}\hat{H}')^{-1} \hat{H} \Delta^* d_t$, and

$$\begin{aligned}
\delta_{t,T} &= (\hat{H}\hat{H}')^{-1} \hat{H} \sum_{s=1}^{t} \Delta^* d_s = (\hat{H}\hat{H}')^{-1} \hat{H} \sum_{s=1}^{t} \begin{pmatrix} D_{s1}^* \theta_1 + C_{s1}^* \gamma_1 \\ \vdots \\ D_{si}^* \theta_i + C_{si}^* \gamma_i \\ \vdots \\ D_{sN}^* \theta_N + C_{sN}^* \gamma_N \end{pmatrix} \\
&= (\hat{H}\hat{H}')^{-1} \hat{H} \begin{pmatrix} \vdots \\ \left(C_{ti} - \frac{t}{T} \right) \theta_i + \left(B_{ti} - (1-\lambda_i) t \right) \gamma_i \\ \vdots \end{pmatrix}
\end{aligned}$$

since $\delta_{t,T} = \sum_{j=2}^{t} \Delta \delta_{j,T}$. Hence, for any generic T_b,

$$(I - P_{T_b}) \delta_T \hat{H} = (I - P_{T_b}) d \hat{H}' (\hat{H}\hat{H}')^{-1} \hat{H}$$

and (8) follows from noting that $P_{T_0} - P_{T_b} = (I - P_{T_b}) - (I - P_{T_0})$. ■

Proof of Lemma 2: Without loss of generality, consider only Model II.

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\|\Delta\delta_{t,T}\|^2 &= \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}V_{NT}^{-1}\hat{H}\Delta^* d_t\right\|^2 \\ &\leq \frac{1}{N^2}\left\|V_{NT}^{-1}\right\|^2\left\|\hat{H}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\|\Delta^* d_t\|^2 \\ &= \frac{1}{N^2}\left\|V_{NT}^{-1}\right\|^2\left\|\hat{H}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\Delta^* d_{ti}^2 \\ &\leq \frac{2}{N^3}\left\|V_{NT}^{-1}\right\|^2\left\|\hat{H}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}(D_{si}^*\dot{\theta}_i)^2 + \frac{2}{N^3}\left\|V_{NT}^{-1}\right\|^2\left\|\hat{H}\right\|^2\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}(C_{si}^*\dot{\gamma}_i)^2 \\ &= O_p(N^{-1})\end{aligned} \tag{A.4}$$

and therefore

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{f}_t - R'f_t\right\|^2 = O_p(C_{N,T}^{-2}).$$

For the second result, note that

$$\hat{H} - R^{-1}H = \frac{1}{T}\hat{f}'\left(fR - \hat{f} + \Delta\delta\right)R^{-1}H + \frac{1}{T}\hat{f}'z + \frac{1}{T}\hat{f}'\Delta^* d - \frac{1}{T}\hat{f}'\Delta\delta R^{-1}H,$$

and thus we show that

$$\begin{aligned}(\hat{H} - R^{-1}H)\dot{\gamma}' &= \frac{1}{T}\hat{f}'\left(fR - \hat{f} + \Delta\delta_T\right)R^{-1}H\dot{\gamma}' + \frac{1}{T}\hat{f}'z\dot{\gamma}' \\ &\quad + \frac{1}{T}\hat{f}'\Delta^* d\dot{\gamma}' - \frac{1}{T}\hat{f}'\Delta\delta_T R^{-1}H\dot{\gamma}' \\ &= O_p(NC_{N,T}^{-2}) + O_p(\sqrt{N}C_{N,T}^{-1}) + O_p(\sqrt{N}C_{N,T}^{-1}) + O_p(\sqrt{N}C_{N,T}^{-1})\end{aligned} \tag{A.5}$$

First note that the second term in (A.5) is such that

$$\begin{aligned}\left\|\frac{1}{NT}\hat{f}'z\dot{\gamma}'\right\| &\leq \left\|\frac{1}{NT}\left(\hat{f} - fR\right)'z\dot{\gamma}'\right\| + \left\|\frac{1}{NT}Rf'z\dot{\gamma}'\right\| \\ &\leq \left(\frac{1}{NT}\sum_{t=1}^{T}\left\|\hat{f}_t - R'f_t\right\|^2\right)^{1/2}\left(\frac{1}{NT}\sum_{t=1}^{T}\|\dot{\gamma}z_t\|^2\right)^{1/2} + \|R\|\left\|\frac{1}{NT}f'z\dot{\gamma}'\right\| \\ &= O_p(N^{-1/2}C_{N,T}^{-1}) + O_p(N^{-1/2}T^{-1/2})\end{aligned} \tag{A.6}$$

For the third term in (A.5),

$$\begin{aligned}\frac{1}{T}\hat{f}'\Delta^* d\dot{\gamma}' &= \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{f}_t C_{ti}^*\dot{\gamma}_i^2 + \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{f}_t D_{ti}^*\dot{\theta}_i\dot{\gamma}_i' \\ &= O_p(\sqrt{N}C_{N,T}^{-1})\end{aligned}$$

because

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t C_{ti}^* &= \frac{1}{T}\sum_{t=1}^{T} R' f_t C_{ti}^* + \frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - R' f_t) C_{ti}^* \\
&\leq O_p(T^{-1/2}) + \left(\frac{1}{T}\sum_{t=2}^{T}\left\|\hat{f}_t - R' f_t\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}(C_{ti}^*)^2\right)^2 \\
&= O_p(T^{-1/2}) + O_p(C_{N,T}^{-1}) = O_p(C_{N,T}^{-1}) \qquad \text{(A.7)}
\end{aligned}$$

and similarly

$$\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t D_{ti}^* = O_p(T^{-1/2}C_{N,T}^{-1}).$$

The result for the fourth term in (A.5) can be shown in a similar manner.

For the first term in (A.5), note that

$$\begin{aligned}
\left\|\frac{1}{T}\left(fR - \hat{f} + \Delta\delta_T\right)'\Delta\delta_T\right\| &\leq \left(\frac{1}{T}\sum_{t=2}^{T}\left\|\hat{f}_t - R' f_t - \Delta\delta_{t,T}\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\|\Delta\delta_{t,T}\|^2\right)^{1/2} \\
&= O_p(C_{N,T}^{-1}N^{-1/2}) \qquad \text{(A.8)}
\end{aligned}$$

from Lemma 1 and (A.4). Also, we have

$$\begin{aligned}
\frac{1}{T}\left(fR - \hat{f} + \Delta\delta_T\right)' f &= V_{NT}^{-1}\frac{1}{NT^2}\sum_{t=1}^{T}\left[\hat{f}'\Delta^* dz_t f_t' + \hat{f}' f H z_t f_t' + \hat{f}' z H' f_t f_t' + \hat{f}' z z_t f_t'\right] \\
&= O_p(C_{N,T}^{-2}) \qquad \text{(A.9)}
\end{aligned}$$

from Lemma B.2 of Bai (2003) and the fact that

$$\begin{aligned}
\left\|\frac{1}{NT^2}\sum_{t=1}^{T}\hat{f}'\Delta^* dz_t f_t'\right\| &\leq \frac{1}{NT^2}\sum_{t=1}^{T}\left\|\hat{f}'\Delta^* dz_t\right\|\|f_t'\| \\
&\leq \frac{1}{N}\left(\frac{1}{T^3}\sum_{t=1}^{T}\left\|\hat{f}'\Delta^* dz_t\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\|f_t'\|^2\right)^{1/2} \\
&= O_p(N^{-1})
\end{aligned}$$

which is true from (A.3) combined with the local to zero break parameters. Now, the first term in (A.5) is such that

$$\begin{aligned}
\frac{1}{T}\hat{f}'\left(fR - \hat{f} + \Delta\delta_T\right) &= \frac{1}{T}(fR + \Delta\delta_T)'\left(fR - \hat{f} + \Delta\delta_T\right) - \frac{1}{T}\left(fR - \hat{f} + \Delta\delta_T\right)'\left(fR - \hat{f} + \Delta\delta_T\right) \\
&= \frac{1}{T}(fR + \Delta\delta_T)'\left(fR - \hat{f} + \Delta\delta_T\right) + O_p(C_{N,T}^{-2}) \\
&= \frac{1}{T}R' f'\left(fR - \hat{f} + \Delta\delta_T\right) + O_p(C_{N,T}^{-1}N^{-1/2}) + O_p(C_{N,T}^{-2}) \\
&= O_p(C_{N,T}^{-2})
\end{aligned}$$

from Lemma 1, (A.8) and (A.9). Therefore, if $N/T^2 = o(1)$,

$$\frac{1}{\sqrt{N}}(\hat{H} - R^{-1}H)\dot{\gamma}' = o_p(1)$$

and similarly

$$\frac{1}{\sqrt{N}}(\hat{H} - R^{-1}H)\dot{\theta} = o_p(1)$$

■

Proof of Lemma 3: The partial sum of the fourth term in (A.1) is

$$\frac{1}{NT}\sum_{k=1}^{t}\hat{f}'\Delta^* dz_k = \left(\frac{1}{T}\hat{f}'\Delta^* d\right)\frac{1}{N}\left(\sum_{k=1}^{t} z_k\right) = \left(\frac{1}{T}\hat{f}'\Delta^* d\right)\frac{1}{N}\left(e_t - e_0 - (e_{Ti} - e_{0i})\frac{t}{T}\right).$$

Here, e_{0i} and $T^{-1}(e_{Ti} - e_{0i})t$ can be ignored since they are a constant term and a linear time trend. Then, from (A.7),

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} a_{t,T}\left(\frac{1}{T}\hat{f}'\Delta^* d\right)\frac{1}{N}e_t &= \frac{1}{N^{3/2}}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{s=1}^{T}\hat{f}_s\left(C_{si}^*\dot{\gamma}_i + D_{si}^*\dot{\theta}_i\right)\right]\frac{1}{\sqrt{T}}\sum_{t=1}^{T} a_{t,T}e_{ti} \\ &= O_p(N^{-1/2})O_p(C_{N,T}^{-1})\end{aligned}$$

Similarly, the partial sum of the fifth term in (A.1) less a constant term and a linear time trend is

$$\frac{1}{\sqrt{N}}\left(\frac{1}{T}\hat{f}'f\right)\frac{1}{\sqrt{N}}He_t$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} a_{t,T}\frac{1}{\sqrt{N}}\left(\frac{1}{T}\hat{f}'f\right)\frac{1}{\sqrt{N}}He_t = \frac{1}{\sqrt{N}}\left(\frac{1}{T}\hat{f}'f\right)\frac{1}{\sqrt{TN}}\sum_{t=1}^{T} a_{t,T}\sum_{i=1}^{N} h_i e_{ti} = O_p(N^{-1/2}).$$

The partial sum of the sixth term in (A.1) less a constant term and a linear time trend is

$$\left(\frac{1}{TN}\hat{f}'zH'\right)F_t.$$

Then, it follows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} a_{t,T}\left(\frac{1}{TN}\hat{f}'zH'\right)F_t = \left(\frac{1}{TN}\hat{f}'zH'\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} a_{t,T}F_t = O_p(C_{N,T}^{-2})$$

since $(NT)^{-1}\hat{f}'zH' = O_p(C_{N,T}^{-2})$ similarly to (A.6).

For the last term in (A.1), let $\gamma_N(s-k) = \mathcal{E}\Delta e_s'\Delta e_k/N$ and $R_N(s-k) = \mathcal{E}e_s'e_k/N$. Then

$$\gamma_N(s-k) = R_N(s-k) - R_N(s-k+1) - R_N(s-k-1) + R_N(s-k)$$

and

$$\sum_{k=1}^{t}\gamma_N(s-k)=R_N(s-t)-R_N(s-t-1)-R_N(s)+R_N(s-1)$$

We expand the last term in (A.1) into

$$\begin{aligned}
\frac{1}{TN}\hat{f}'zz_k &= \frac{1}{TN}\sum_{s=1}^{T}\hat{f}_s z_s' z_k \\
&= \frac{1}{TN}\sum_{s=1}^{T}\hat{f}_s\begin{pmatrix}\Delta e_s'\Delta e_k - T^{-1}\Delta e_s'(e_T-e_0)\\ -T^{-1}(e_T-e_0)'\Delta e_k + T^{-2}(e_T-e_0)'(e_T-e_0)\end{pmatrix} \\
&= \frac{1}{T}\sum_{s=1}^{T}\hat{f}_s\gamma_N(s-k)+\frac{1}{T}\sum_{s=1}^{T}\hat{f}_s\left(\frac{\Delta e_s'\Delta e_k}{N}-\gamma_N(s-k)\right)-\frac{1}{T^2N}\sum_{s=1}^{T}\hat{f}_s\Delta e_s'(e_T-e_0) \\
&\quad -\frac{1}{T^2N}\sum_{s=1}^{T}\hat{f}_s(e_T-e_0)'\Delta e_k + T^{-2}\frac{1}{TN}\sum_{s=1}^{T}\hat{f}_s(e_T-e_0)'(e_T-e_0). \qquad (A.10)
\end{aligned}$$

Then, the partial sum of the first part in (A.10) is

$$\begin{aligned}
\sum_{k=1}^{t}\frac{1}{T}\sum_{s=1}^{T}\hat{f}_s\gamma_N(s-k) &= \frac{1}{T}\sum_{s=1}^{T}\hat{f}_s\sum_{k=1}^{t}\gamma_N(s-k) \qquad (A.11)\\
&= \frac{1}{T}\sum_{s=1}^{T}\hat{f}_s\left(R_N(s-t)-R_N(s-t-1)-R_N(s)+R_N(s-1)\right)
\end{aligned}$$

The last two terms in (A.11) are constant over t and can be ignored. The first term in (A.11) is such that

$$\begin{aligned}
\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}a_{t,T}\frac{1}{T}\sum_{s=1}^{T}\hat{f}_sR_N(s-t)\right\| &= \left\|\frac{1}{\sqrt{T}}\frac{1}{T}\sum_{s=1}^{T}\hat{f}_s\sum_{t=1}^{T}a_{t,T}R_N(s-t)\right\| \\
&\le \frac{1}{\sqrt{T}}\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{f}_s\right\|\sum_{t=1}^{T}|a_{t,T}|\,|R_N(s-t)| \\
&\le \frac{1}{\sqrt{T}}\max_t|a_{t,T}|\sum_{h=-\infty}^{\infty}|R_N(h)|\,\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{f}_s\right\| = O_p(T^{-1/2})
\end{aligned}$$

The second term in (A.11) has the same order of magnitude.

Let $\kappa_i(p,q,r,s)$ be the fourth order cumulant of e_{si}, $\mathcal{E}e_{si}e_{ti} = r_i(s-t)$, $\xi_{i,N}(s,t) = e_{si}e_{ti} - r_i(s-t)$ and $\xi_N(s,k) = N^{-1/2}\sum_{i=1}^N \xi_{i,N}(s,t)$. Note that

$$\begin{aligned}
& \frac{1}{T}\mathcal{E}\left(\sum_{t=1}^T a_{t,T}\xi_N(s,t)\right)^2 \\
= & \frac{1}{NT}\mathcal{E}\left(\sum_{t=1}^T a_{t,T}\sum_{i=1}^N \xi_{i,N}(s,t)\right)^2 \\
= & \frac{1}{NT}\sum_{i=1}^N \mathcal{E}\left(\sum_{t=1}^T a_{t,T}\xi_{i,N}(s,t)\right)^2 \\
= & \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T a_{t,T}\sum_{u=1}^T a_{u,T}\mathcal{E}\left(\xi_{i,N}(s,t)\xi_{i,N}(s,u)\right) \\
= & \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T a_{t,T}\sum_{u=1}^T a_{u,T}\left(r_i(0)r_i(u-s) + r_i(s-t)r_i(u-t) + \kappa_i(0,0,s-t,u-t)\right) \\
< & \frac{(\max_t |a_{t,T}|)^2}{NT}\sum_{i=1}^N\sum_{t=1}^T\left[(|r_i(0)| + |r_i(s-t)|)\sum_{h=-\infty}^{\infty}|r_i(h)| + \sum_{d=-\infty}^{\infty}\sum_{c=-\infty}^{\infty}|\kappa_i(0,0,c,d)|\right] = O(1).
\end{aligned}$$

where the second equality follows from the independence of e_i across i and for the fourth equality see Hannan (1970) page 23.

The partial sum of the second term in (A.10) is

$$\begin{aligned}
& \sum_{k=1}^t \frac{1}{T}\sum_{s=1}^T \hat{f}_s\left(\frac{\Delta e_s'\Delta e_k}{N} - \gamma_N(s-k)\right) \\
= & \frac{1}{T}\sum_{s=1}^T \hat{f}_s\sum_{k=1}^t\left(\frac{\Delta e_s'\Delta e_k}{N} - \gamma_N(s-k)\right) \\
= & \frac{1}{\sqrt{N}}\frac{1}{T}\sum_{s=1}^T \hat{f}_s\left(\xi_N(s,t) - \xi_N(s-1,t) - \xi_N(s,0) + \xi_N(s-1,0)\right). \qquad \text{(A.12)}
\end{aligned}$$

In (A.12), the last two terms are constants and thus ignored. The first two terms in (A.12) are of the same order of magnitude.

$$\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^T a_{t,T}\frac{1}{\sqrt{N}}\frac{1}{T}\sum_{s=1}^T \hat{f}_s\xi_N(s,t) \\
= & \frac{1}{\sqrt{N}}\frac{1}{T^{3/2}}\sum_{s=1}^T \hat{f}_s\left(\sum_{t=1}^T a_{t,T}\xi_N(s,t)\right) \\
\leq & \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{s=1}^T ||\hat{f}_s||^2\right)^{1/2}\left(\frac{1}{T^2}\sum_{s=1}^T\left(\sum_{t=1}^T a_{t,T}\xi_N(s,t)\right)^2\right)^{1/2} = O_p(N^{-1/2}).
\end{aligned}$$

The partial sums of the third and fifth in (A.10) are linear in time and thus ignored. Lastly, the partial sum of the fourth term in (A.10) less a constant and a linear time trend is

$$\frac{1}{T^2N}\sum_{s=1}^{T}\hat{f}_s(e_T-e_0)'e_t$$

and thus

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}a_{t,T}\frac{1}{T^2N}\sum_{s=1}^{T}\hat{f}_s(e_T-e_0)'e_t=\frac{1}{T^2N}\sum_{s=1}^{T}\hat{f}_s(e_T-e_0)'\frac{1}{\sqrt{T}}\sum_{t=1}^{T}a_{t,T}e_t=O_p(T^{-1}).\ \blacksquare$$

Proof of Lemma 4: We prove Models I and II only, since Model III can be shown easily from Model II. We begin with introducing some more notations. Let $\Delta_T = diag\{T^{-1/2}, T^{-3/2}, T^{-3/2}\}$ for Model I and $diag\{T^{-1/2}, T^{-3/2}, T^{-1/2}, T^{-3/2}\}$ for Model II. Define $\widetilde{\iota}(T_b) = (\widetilde{\iota}_1(T_b), \dots \widetilde{\iota}_T(T_b))'$ such that $B(T_0) - B(T_b) = (T_b - T_0)\widetilde{\iota}(T_b)$, that is,

$$\text{if } T_b > T_0,\quad \widetilde{\iota}_t(T_b) \equiv \begin{cases} 0 & \text{if } 1 \le t \le T_0, \\ (t-T_0)/(T_b-T_0) & \text{if } T_0+1 \le t \le T_b, \\ 1 & \text{if } T_b+1 \le t \le T, \end{cases}$$

$$\text{if } T_b < T_0,\quad \widetilde{\iota}_t(T_b) \equiv \begin{cases} 0 & \text{if } 1 \le t \le T_b, \\ (t-T_b)/(T_0-T_b) & \text{if } T_b+1 \le t \le T_0, \\ 1 & \text{if } T_0+1 \le t \le T, \end{cases}$$

$$\text{if } T_b = T_0,\quad \widetilde{\iota}_t(T_b) \equiv \begin{cases} 0 & \text{if } 1 \le t \le T_0, \\ 1 & \text{if } T_0+1 \le t \le T. \end{cases}$$

Also, define $\alpha(T_b) = (\alpha_1(T_b), \dots, \alpha_T(T_b))'$ and $\kappa(T_b) = (\kappa_1(T_b), \dots, \kappa_T(T_b))'$ where

$$\text{if } T_b \ge T_0,\quad \alpha_t(T_b) = \begin{cases} 1 & \text{if } T_0+1 \le t \le T_b, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{if } T_0 > T_b,\quad \alpha_t(T_b) = \begin{cases} -1 & \text{if } T_b+1 \le t \le T_0, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$\text{if } T_b \ge T_0,\quad \kappa_t(T_b) = \begin{cases} t-T_0 & \text{if } T_0+1 \le t \le T_b, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{if } T_0 > T_b,\quad \kappa_t(T_b) = \begin{cases} -(t-T_0) & \text{if } T_b+1 \le t \le T_0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that

$$\begin{aligned} C_{ti} &= C_t(T_0) - [C_t(T_0) - C_{ti}] = C_t(T_0) - \alpha_t(T_i) \\ B_{ti} &= B_t(T_0) - [B_t(T_0) - B_{ti}] = B_t(T_0) - \Delta T_i \, \widetilde{\iota}_t(T_i) \\ \widetilde{\iota}_t(T_i) &= \frac{1}{\Delta T_i}\kappa_t(T_i) - C_t(T_i) \end{aligned} \quad (A.13)$$

Let Π_i be the regression coefficients for equation i, that is,

$$\Pi_i = \begin{cases} (\mu_i, \beta_i, \gamma_i)' & \text{Model I} \\ (\mu_i, \beta_i, \theta_i, \gamma_i)' & \text{Model II} \\ (\mu_i, \beta_i, \theta_i)' & \text{Model III} \end{cases}$$

Then, we write each equation in a matrix form as

$$\underset{(T\times 1)}{d_i} = \underset{(T\times 3 \text{ or } T\times 4)}{X_{T_0}} \underset{(3\times 1 \text{ or } 4\times 1)}{\Pi_i} - \underset{(T\times 1)}{Z_i} \quad (A.14)$$

where $Z_i = (z_{i1}, \ldots, z_{iT})'$ and $z_{it} = \Delta T_i \, \widetilde{\iota}_t(T_i)\gamma_i$ for Model I, $\alpha_t(T_i)\theta_i + \Delta T_i \, \widetilde{\iota}_t(T_i)\gamma_i$ for Model II and $\alpha_t(T_i)\theta_i$ for Model III, using (A.13).

The entire system is written as

$$d = X_{T_0}\Pi - Z$$

where $Z = [Z_1, \ldots, Z_N]$ and $\Pi = [\Pi_1, \ldots, \Pi_N]$.

First, consider Model I,

$$\begin{aligned} (\hat{X}\hat{X}) &= tr\left[M_{\hat{H}'} d'(P_{T_0} - P_{T_b})d\right] \\ &= tr\left[M_{\hat{H}'}\Pi'(X_{T_0} - X_{T_b})'(I - P_{T_b})(X_{T_0} - X_{T_b})\Pi\right] \\ &\quad -2tr\left[M_{\hat{H}'}\Pi'(X_{T_0} - X_{T_b})'(I - P_{T_b})Z\right] \\ &\quad +tr\left[M_{\hat{H}'}Z'(P_{T_0} - P_{T_b})Z\right] \\ &= (\hat{X}\hat{X})_{11} - 2(\hat{X}\hat{X})_{12} + (\hat{X}\hat{X})_{22} \\ &= |T_b - T_0|^2 O_p(T) + |T_b - T_0| O_p(TN^{-1/2}) + |T_b - T_0| O_p(1) \end{aligned} \quad (A.15)$$

To show (A.15), note

$$(\hat{X}\hat{X})_{11} = |T_b - T_0|^2 tr\left[\widetilde{\iota}(T_b)'(I - P_{T_b})\widetilde{\iota}(T_b)\gamma M_{\hat{H}'}\gamma'\right] = |T_b - T_0|^2 O_p(T)$$

because $\widetilde{\iota}(T_b)'(I - P_{T_b})\widetilde{\iota}(T_b) = O(T)$ from Perron and Zhu (2005) and

$$\gamma M_{\hat{H}'}\gamma' = \frac{1}{N}\dot{\gamma}(I - \hat{H}'(\hat{H}\hat{H}')\hat{H})\dot{\gamma}' \xrightarrow{p} \bar{A}_{\gamma\gamma} = \dot{A}_{\gamma\gamma} - A_{\dot{\gamma}H}A_{HH}^{-1}A_{H\dot{\gamma}}$$

Also,

$$\begin{aligned} (\hat{X}\hat{X})_{12} &= tr\left[M_{\hat{H}'}\Pi'(X_{T_0} - X_{T_b})'(I - P_{T_b})Z\right] \\ &= (T_b - T_0)tr\left[\gamma'\widetilde{\iota}(T_b)'(I - P_{T_b})Z\right] \\ &\quad -(T_b - T_0)tr\left[\hat{H}'(\hat{H}\hat{H}')\hat{H}\gamma'\widetilde{\iota}(T_b)'(I - P_{T_b})Z\right] \\ &= |T_b - T_0| O_p(TN^{-1/2}) \end{aligned}$$

because, for the first part,

$$tr\left[\gamma'\widetilde{\iota}(T_b)'(I-P_{T_b})Z\right]=\frac{1}{N}\sum_{i=1}^{N}\widetilde{\iota}(T_b)'(I-P_{T_b})\widetilde{\iota}(T_i)\dot{\gamma}_i^2\Delta T_i=O_p(TN^{-1/2})$$

which follow from the fact that ΔT_i is a iid draw from a distribution with finite variance, $\widetilde{\iota}(T_b)'(I-P_{T_b})\widetilde{\iota}(T_i)=O(T)$ uniformly in i and $\max\{\dot{\gamma}_i^2\}=O_p(1)$ from Assumption 3(iii). A similar argument applies to the second part. A useful decomposition of $(\hat{X}\hat{X})_{22}$ is

$$\begin{aligned}
& tr\left[M_{\hat{H}'}Z'(P_{T_0}-P_{T_b})Z\right] \\
= \ & tr\left[M_{\hat{H}'}Z'(X_{T_0}-X_{T_b})\Delta_T\left(\Delta_T X_{T_0}'X_{T_0}\Delta_T\right)^{-1}\Delta_T X_{T_0}'Z\right] \\
& +tr\left[M_{\hat{H}'}Z'X_{T_b}\Delta_T\left(\Delta_T X_{T_b}'X_{T_b}\Delta_T\right)^{-1}\Delta_T\left(X_{T_b}'X_{T_b}-X_{T_0}'X_{T_0}\right)\Delta_T\left(\Delta_T X_{T_0}'X_{T_0}\Delta_T\right)^{-1}\Delta_T X_{T_0}'Z\right] \\
& +tr\left[M_{\hat{H}'}Z'X_{T_b}\Delta_T\left(\Delta_T X_{T_b}'X_{T_b}\Delta_T\right)^{-1}\Delta_T(X_{T_0}-X_{T_b})'Z\right] \\
= \ & R_1+R_2+R_3
\end{aligned}$$

Then,write

$$\begin{aligned}
R_1 &= tr\left[M_{\hat{H}'}Z'(X_{T_0}-X_{T_b})\Delta_T\left(\Delta_T X_{T_0}'X_{T_0}\Delta_T\right)^{-1}\Delta_T X_{T_0}'Z\right] \\
&= tr\left[\left(\Delta_T X_{T_0}'X_{T_0}\Delta_T\right)^{-1}\Delta_T X_{T_0}'ZM_{\hat{H}'}Z'(X_{T_0}-X_{T_b})\Delta_T\right] \\
&= vec\left(\left(\Delta_T X_{T_0}'X_{T_0}\Delta_T\right)^{-1}\right)'vec\left(\Delta_T X_{T_0}'ZM_{\hat{H}'}Z'(X_{T_0}-X_{T_b})\Delta_T\right)
\end{aligned}$$

where $vec((\Delta_T X_{T_0}'X_{T_0}\Delta_T)^{-1})=O_p(1)$ from Lemma A.1 in Kim (2010).

$$\begin{aligned}
\Delta_T X_{T_0}'ZZ'(X_{T_0}-X_{T_b})\Delta_T &= \frac{1}{N}\sum_{i=1}^{N}\Delta T_i^2\dot{\gamma}_i^2\Delta_T X_{T_0}'\widetilde{\iota}(T_i)\widetilde{\iota}(T_i)'(X_{T_0}-X_{T_b})\Delta_T \\
&= |T_b-T_0|O_p(1)
\end{aligned}$$

because, uniformly in i,

$$\begin{aligned}
\Delta_T X_{T_0}'\widetilde{\iota}(T_i) &= O(T^{1/2}) \\
\Delta_T(X_{T_0}-X_{T_b})'\widetilde{\iota}(T_i) &= |T_b-T_0|O(T^{-1/2}).
\end{aligned}$$

Note that $\Delta_T\left(X_{T_b}'X_{T_b}-X_{T_0}'X_{T_0}\right)\Delta_T=|T_b-T_0|O_p(T^{-1})$. Then it follows that R_2 and R_3 are of the same order as R_1. Thus,

$$(\hat{X}\hat{X})_{22}=tr\left[M_{\hat{H}'}Z'(P_{T_0}-P_{T_b})Z\right]=|T_b-T_0|O(1).$$

Consider $(\hat{X}\hat{U})$. From (9),

$$\begin{aligned}
(\hat{X}\hat{U}) &= tr\left[d'(P_{T_0}-P_{T_b})UM_{\hat{H}'}\right] \\
&= tr\left[\Pi'(X_{T_0}-X_{T_b})'(I-P_{T_b})UM_{\hat{H}'}\right]-tr\left[Z'(P_{T_0}-P_{T_b})UM_{\hat{H}'}\right] \\
&= (\hat{X}\hat{U})_1-(\hat{X}\hat{U})_2
\end{aligned}$$

Here,

$$
\begin{aligned}
(\hat{X}\hat{U})_1 &= (T_b - T_0)\,\widetilde{\iota}(T_b)'(I - P_{T_b})FH(I - P_{\hat{H}'})\gamma' \\
&\quad +(T_b - T_0)\,\widetilde{\iota}(T_b)'(I - P_{T_b})E(I - P_{\hat{H}'})\gamma' \\
&= |T_b - T_0|\,o_p(T^{1/2}) + |T_b - T_0|\,O_p(T^{1/2}). \qquad \text{(A.16)}
\end{aligned}
$$

because $\widetilde{\iota}_b'(I - P_{T_b})F = O_p(T^{1/2})$ from Lemma A.4 in Kim (2010), $\widetilde{\iota}_b'(I - P_{T_b})E(I - P_{\hat{H}'})\dot{\gamma}' = O_p(T^{1/2}N^{1/2})$ and

$$
\begin{aligned}
R^{-1}H(I - P_{\hat{H}'})\gamma' &= (\hat{H}' - R^{-1}H)(I - P_{\hat{H}'})\gamma' \\
&= (\hat{H}' - R^{-1}H)\gamma' - (\hat{H}' - R^{-1}H)\hat{H}'(\hat{H}\hat{H}')^{-1}\hat{H}\gamma' \\
&= \frac{1}{\sqrt{N}}(\hat{H}' - R^{-1}H)\dot{\gamma}' - \frac{1}{\sqrt{N}}(\hat{H}' - R^{-1}H)\hat{H}'(\frac{1}{N}\hat{H}\hat{H}')^{-1}\frac{1}{N}\hat{H}\dot{\gamma}' \\
&= o_p(1). \qquad \text{(A.17)}
\end{aligned}
$$

from Lemma 2. $(\hat{X}\hat{U})_2$ is of strictly smaller order than $(\hat{X}\hat{U})_1$, although the details of the derivation are omitted.

Now, consider a decomposition of $(\hat{U}\hat{U})$:

$$
\begin{aligned}
& tr\left[\hat{U}'(P_{T_0} - P_{T_b})\hat{U}\right] \qquad \text{(A.18)} \\
= \; & tr\left[\hat{U}'(X_{T_0} - X_{T_b})\Delta_T\left(\Delta_T X_{T_0}' X_{T_0}\Delta_T\right)^{-1}\Delta_T X_{T_0}'\hat{U}\right] \\
& + tr\left[\hat{U}'X_{T_b}\Delta_T\left(\Delta_T X_{T_b}' X_{T_b}\Delta_T\right)^{-1}\Delta_T\left(X_{T_b}'X_{T_b} - X_{T_0}'X_{T_0}\right)\Delta_T\left(\Delta_T X_{T_0}'X_{T_0}\Delta_T\right)^{-1}\Delta_T X_{T_0}'\hat{U}\right] \\
& + tr\left[\hat{U}'X_{T_b}\Delta_T\left(\Delta_T X_{T_b}'X_{T_b}\Delta_T\right)^{-1}\Delta_T(X_{T_0} - X_{T_b})'\hat{U}\right] \\
= \; & r_1 + r_2 + r_3
\end{aligned}
$$

Write

$$
\begin{aligned}
r_1 &= tr\left[\hat{U}'(X_{T_0} - X_{T_b})\Delta_T\left(\Delta_T X_{T_0}'X_{T_0}\Delta_T\right)^{-1}\Delta_T X_{T_0}'\hat{U}\right] \\
&= tr\left[\left(\Delta_T X_{T_0}'X_{T_0}\Delta_T\right)^{-1}\Delta_T X_{T_0}'\hat{U}\hat{U}'(X_{T_0} - X_{T_b})\Delta_T\right] \\
&= vec\left(\left(\Delta_T X_{T_0}'X_{T_0}\Delta_T\right)^{-1}\right)' vec\left(\Delta_T X_{T_0}'\hat{U}\hat{U}'(X_{T_0} - X_{T_b})\Delta_T\right).
\end{aligned}
$$

Now, let $(\Delta FH) = FH - (\hat{F} - \delta_T)\hat{H}$, then

$$
\begin{aligned}
& vec\left(\Delta_T X_{T_0}'\hat{U}\hat{U}'(X_{T_0} - X_{T_b})\Delta_T\right) \\
= \; & vec\left(\Delta_T X_{T_0}'(\Delta FH)(\Delta FH)'(X_{T_0} - X_{T_b})\Delta_T\right) + vec\left(\Delta_T X_{T_0}'(\Delta FH)E'(X_{T_0} - X_{T_b})\Delta_T\right) \\
& + vec\left(\Delta_T X_{T_0}'E(\Delta FH)'(X_{T_0} - X_{T_b})\,\Delta_T\right) + vec\left(\Delta_T X_{T_0}'EE'(X_{T_0} - X_{T_b})\,\Delta_T\right) \\
= \; & r_{11} + r_{12} + r_{13} + r_{14}.
\end{aligned}
$$

>From Kim (2010), r_{12} and r_{13} are of strictly smaller order of magnitude than r_{14}. Also, $r_{14} = |T_b - T_0|\,O_p(T^{-1}N)$.

Now, realize that

$$\begin{aligned}(\Delta FH) &= FH - (\hat{F} - \delta_T)\hat{H} \\ &= (FR - (\hat{F} - \delta_T))R^{-1}H + (FR - (\hat{F} - \delta_T))(\hat{H} - R^{-1}H) - FR(\hat{H} - R^{-1}H),\end{aligned}$$

and note that $\Delta_T X'_{T_0}(FR-(\hat{F}-\delta_T)) = O_p(C_{N,T}^{-1})$, $\Delta_T(X_{T_0}-X_{T_b})'(FR-(\hat{F}-\delta_T)) = |T_b - T_0|\, O_p(C_{N,T}^{-1}T^{-1})$, and $H(\hat{H} - R^{-1}H)' = o_p(\sqrt{N})$, which are true from Lemmas 2 and 3, while $\Delta_T X'_{T_0}F = O_p(1)$ and $\Delta_T(X_{T_0} - X_{T_b})'F = |T_b - T_0|\, O_p(T^{-1})$ from Kim (2010). In addition, similarly to (A.5),

$$\frac{1}{N}(\hat{H} - R^{-1}H)(\hat{H} - R^{-1}H)' = O_p(C_{N,T}^{-2})$$

because, without loss of generality, for Model II,

$$\frac{1}{T^2}\hat{f}'\left(fR - \hat{f} + \Delta\delta_T\right)R^{-1}\left(\frac{1}{N}HH'\right)R^{-1\prime}\left(fR - \hat{f} + \Delta\delta_T\right)'\hat{f} = O_p(C_{N,T}^{-4})$$

$$\begin{aligned}\left\|\frac{1}{NT^2}\hat{f}'zz'\hat{f}\right\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{f}_s z_{si}\right)\left(\frac{1}{T}\sum_{s=1}^{T}\hat{f}_s z_{si}\right)'\right\| \\ &\leq \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}\hat{f}_s z_{si}\right\|^2 \\ &\leq \frac{2}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}(\hat{f}_s - R'f_s)z_{si}\right\|^2 + \frac{2}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}R'f_s z_{si}\right\|^2 \\ &= O_p(C_{N,T}^{-2})\end{aligned}$$

$$\begin{aligned}&\frac{1}{NT^2}\left\|\hat{f}'\Delta^* d\,(\Delta^* d)'\,\hat{f}\right\| \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t\left(C^*_{ti}\dot{\gamma}_i + D^*_{ti}\dot{\theta}_i\right)\right\|^2 = O_p(N^{-1}C_{N,T}^{-2})\end{aligned}$$

$$\begin{aligned}\frac{1}{T}\hat{f}'\Delta^* d\dot{\gamma}' &= \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{f}_t C^*_{ti}\dot{\gamma}_i^2 + \frac{1}{T\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{f}_t D^*_{ti}\dot{\theta}_i\dot{\gamma}_i' \\ &= O_p(\sqrt{N}C_{N,T}^{-1})\end{aligned}$$

and

$$\frac{1}{NT^2}\hat{f}'\Delta\delta_T R^{-1}HH'R^{-1\prime}\Delta\delta_T'\hat{f} = O_p(N^{-1}C_{N,T}^{-2}).$$

Then, we can show that r_{11} is of smaller order than r_{14}. Thus, $r_1 = |T_b - T_0|\, O_p(T^{-1}N)$. A similar argument shows that $r_2 = |T_b - T_0|\, O_p(T^{-1}N)$ and $r_3 = |T_b - T_0|\, O_p(T^{-1}N)$. Therefore, it follows that $(\hat{U}\hat{U}) = |T_b - T_0|\, O_p(T^{-1}N)$.

Now Consider Model II. The difference is that there is an extra regressor. Note that from (A.13) and (A.14)

$$\begin{aligned}
d_{ti} - \mu_i - \beta_i t &= C_t(T_0)\theta_i + B_t(T_0)\gamma_i - \alpha_t(T_i)\theta_i - \Delta T_i \widetilde{\iota}_t(T_i)\gamma_i \\
&= C_t(T_0)\theta_i + B_t(T_0)\gamma_i - \alpha_t(T_i)\theta_i - (\kappa_t(T_i) - \Delta T_i C_t(T_i))\gamma_i \\
&= C_t(T_0)(\theta_i + \Delta T_i\gamma_i) + B_t(T_0)\gamma_i - \alpha_t(T_i)(\theta_i + \Delta T_i\gamma_i) - \kappa_t(T_i)\gamma_i \\
&= C_t(T_0)\theta_i^a + B_t(T_0)\gamma_i - \alpha_t(T_i)\theta_i^a - \kappa_t(T_i)\gamma_i
\end{aligned}$$

where $\theta_i^a = (\theta_i + \Delta T_i\gamma_i)$. Now, consider a decomposition of $(\hat{X}\hat{X})$ similar to (A.15). Then,

$$\begin{aligned}
& (\hat{X}\hat{X})_{11} \qquad\qquad (A.19) \\
= \; & tr\left[M_{\hat{H}'}(\theta^{a\prime}\alpha(T_b)' + \gamma'\kappa(T_b)')(I - P_{T_b})(\alpha(T_b)\theta^a + \kappa(T_b)\gamma)M_{\hat{H}'}\right] \\
= \; & \kappa(T_b)'(I - P_{T_b})\kappa(T_b)\ (\gamma M_{\hat{H}'}\gamma') \\
& +2\kappa(T_b)'(I - P_{T_b})\alpha(T_b)\ (\theta^a M_{\hat{H}'}\gamma') \\
& +\alpha(T_b)'(I - P_{T_b})\alpha(T_b)\ (\theta^a M_{\hat{H}'}\theta^{a\prime})
\end{aligned}$$

where $\kappa'(I - P_{T_b})\kappa = |T_b - T_0|^3 O(1)$, $\kappa'(I - P_{T_b})\alpha = |T_b - T_0|^2 O(1)$, and $\alpha'(I - P_{T_b})\alpha = |T_b - T_0| O(1)$. The dominant term is $\kappa'(I - P_{T_b})\kappa$, and we can have

$$(\hat{X}\hat{X})_{11} = |T_b - T_0|^3 O(1).$$

Also,

$$\begin{aligned}
(\hat{X}\hat{X})_{12} &= tr\left[M_{\hat{H}'}\theta^{a\prime}\alpha(T_b)'(I - P_{T_b})Z\right] + tr\left[M_{\hat{H}'}\gamma'\kappa(T_b)'(I - P_{T_b})Z\right] \\
&= \sum_{i=1}^{N}\alpha(T_b)'(I - P_{T_b})(\kappa(T_i)\gamma_i + \alpha(T_i)\theta_i^a)\widetilde{\theta}_i^a \\
&\quad + \sum_{i=1}^{N}\kappa(T_b)'(I - P_{T_b})(\kappa(T_i)\gamma_i + \alpha(T_i)\theta_i^a)\widetilde{\gamma}_i
\end{aligned}$$

where $\widetilde{\theta}_i^a$ is the i^{th} element of $M_{\hat{H}'}\theta^{a\prime}$ and $\widetilde{\gamma}_i$ is the ith element of $M_{\hat{H}'}\gamma'$. $(\hat{X}\hat{X})_{12}$ cannot be of greater order of magnitude than $(\hat{X}\hat{X})_{11}$, because, for example,

$$\kappa(T_b)'(I - P_{T_b})\kappa(T_i) = \begin{cases} \kappa(T_b)'(I - P_{T_b})\kappa(T_b) & \text{if } T_i > T_b > T_0 \text{ or } T_i < T_b < T_0 \\ \kappa(T_i)'(I - P_{T_b})\kappa(T_i) & \text{if } T_b > T_i > T_0 \text{ or } T_b < T_i < T_0 \\ 0 & \text{otherwise} \end{cases}$$

and thus it is bounded by $\kappa(T_b)'(I - P_{T_b})\kappa(T_b)$. It can be further shown that $(\hat{X}\hat{X})_{22}$ is of smaller order of magnitude than $(\hat{X}\hat{X})_{11}$ and $(\hat{X}\hat{X})_{12}$, although we do not provide the details to conserve the space.

For $(\hat{X}\hat{U})$, first write

$$(\hat{X}\hat{U})_1 = tr\left[M_{\hat{H}'}(\theta^{a\prime}\alpha(T_0)+\gamma'\kappa(T_0)')(I-P_{T_b})U\right] \quad (A.20)$$
$$= tr\left[M_{\hat{H}'}(\theta^{a\prime}\alpha(T_0)+\gamma'\kappa(T_0)')(I-P_{T_b})(FH+E)\right]$$

It can be shown that

$$tr\left[M_{\hat{H}'}\theta^{a\prime}\alpha(T_0)'(I-P_{T_b})E\right] = |T_b-T_0|^{1/2}\,O_p(1)$$
$$tr\left[M_{\hat{H}'}\gamma'\kappa(T_0)'(I-P_{T_b})E\right] = |T_b-T_0|^{3/2}\,O_p(1),$$

using Lemma A.2 in Kim (2010). Furthermore, $(\hat{X}\hat{U})_2 = tr\left[Z'(P_{T_0}-P_{T_b})UM_{\hat{H}'}\right]$ is of smaller order than $(\hat{X}\hat{U})_1$ as in Model I. Hence $(\hat{X}\hat{U}) = |T_b-T_0|^{3/2}\,O_p(1)$.

For $(\hat{U}\hat{U})$, consider the decomposition in (A.18). Now, $\Delta_T(X_{T_0}-X_{T_b}) = [0,0,\alpha,|T_b-T_0|\,\widetilde{\iota}_b]$. Using an approach similar to Model I, we obtain that $(\hat{U}\hat{U}) = |T_b-T_0|^{1/2}\,O_p(T^{-1/2}N)$. ■

Proof of Theorem 1: Omitted since it is straightforward from Lemma 4. For a similar argument, see Kim (2010).

Proof of Theorem 2: (i) Consider Model I. Let $m_T = T^{1/2}(T_b-T_0)$ and $D(C)=\{T_b : |T_b-T_0| < CT^{-1/2}\}$ for a positive number C. On the set $D(C)$, $(\hat{X}\hat{X})_{11} = O_p(1)$, $(\hat{X}\hat{X})_{12} = O_p(1)$, $(\hat{X}\hat{X})_{22} = o_p(1)$, $(\hat{X}\hat{U})_1 = O_p(1)$, $(\hat{X}\hat{U})_2 = o_p(1)$ and $(\hat{U}\hat{U}) = o_p(1)$, if $T/N \to 0 < \rho < \infty$. From (A.15),

$$(\hat{X}\hat{X})_{11} = m_T^2\left(\frac{1}{T}\widetilde{\iota}(T_b)'(I-P_{T_b})\widetilde{\iota}(T_b)\right)\left(\frac{1}{N}\dot{\gamma}(I-\hat{H}'(\hat{H}\hat{H}')\hat{H})\dot{\gamma}'\right)$$
$$= m_T^2\frac{(1-\lambda_0)\lambda_0}{4}\bar{A}_{\gamma\gamma}+o_p(1)$$

and

$$(\hat{X}\hat{X})_{12} = (T_b-T_0)tr\left[(I-\hat{H}'(\hat{H}\hat{H}')\hat{H})\gamma'\widetilde{\iota}(T_b)'(I-P_{T_b})Z\right]$$
$$= m_T\frac{1}{T^{1/2}N}\sum_{i=1}^{N}\widetilde{\iota}(T_b)'(I-P_{T_b})\widetilde{\iota}(T_i)\Delta T_i(\dot{\gamma}_i^2-\gamma_i h_i'A_{HH}^{-1}A_H\dot{\gamma})+o_p(1)$$
$$= m_T\frac{1}{\sqrt{\rho}T}\widetilde{\iota}(T_b)'(I-P_{T_b})\widetilde{\iota}(T_b)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta T_i(\dot{\gamma}_i^2-\gamma_i h_i'A_{HH}^{-1}A_H\dot{\gamma})+o_p(1)$$

>From (A.16),

$$(\hat{X}\hat{U})_1 = (T_b-T_0)\widetilde{\iota}(T_b)'(I-P_{T_b})E\gamma'$$
$$= m_T\left(\frac{1}{\sqrt{TN}}\widetilde{\iota}(T_b)'(I-P_{T_b})E(I-P_{\hat{H}'})\dot{\gamma}'\right)+o_p(1).$$

Hence, on the set $D(C)$,

$$
\begin{aligned}
m_T^* &= \arg\min_{m_T \text{ on } D(C)} \left[(\hat{X}\hat{X})_{11} - 2(\hat{X}\hat{X})_{12} + 2(\hat{X}\hat{U})_1 + o_p(1)\right] \\
&= \arg\min_{m_T \text{ on } D(C)} \left[\begin{array}{c} m_T^2\left(\frac{1}{T}\widetilde{\iota}(T_b)'(I-P_{T_b})\widetilde{\iota}(T_b)\right)\bar{A}_{\gamma\gamma} \\ -2m_T\frac{1}{\sqrt{\rho}}\frac{1}{T}\widetilde{\iota}(T_b)'(I-P_{T_b})\widetilde{\iota}(T_b)\frac{1}{\sqrt{N}}\sum_{i=1}^N \Delta T_i(\dot{\gamma}_i^2 - \dot{\gamma}_i h_i' A_{HH}^{-1} A_H \dot{\gamma}) \\ +2m_T\left(\frac{1}{\sqrt{TN}}\widetilde{\iota}(T_b)'(I-P_{T_b})E(I-P_{\hat{H}'})\dot{\gamma}'\right) + o_p(1) \end{array}\right] \\
&= -\left(\frac{1}{T}\widetilde{\iota}(T_b)'(I-P_{T_b})\widetilde{\iota}(T_b)\bar{A}_{\gamma\gamma}\right)^{-1}\left(\frac{1}{\sqrt{TN}}\widetilde{\iota}(T_b)'(I-P_{T_b})E(I-P_{\hat{H}'})\dot{\gamma}'\right) \\
&\quad + \frac{1}{\sqrt{\rho}\bar{A}_{\gamma\gamma}}\frac{1}{\sqrt{N}}\sum_{i=1}^N \Delta T_i(\dot{\gamma}_i^2 - \gamma_i h_i' A_{HH}^{-1} A_H \dot{\gamma}) + o_p(1).
\end{aligned}
$$

For the first term, it follows from the Joint Limit CLT (Theorem 3 in Phillips and Moon, 1999) that

$$
\frac{1}{\sqrt{TN}}\widetilde{\iota}(T_b)'(I-P_{T_b})E(I-P_{\hat{H}'})\dot{\gamma}' \xrightarrow{d} N\left(0, \frac{(1-\lambda_0)\lambda_0}{4}\bar{S}_{\gamma\gamma}\right)
$$

as N and T increase to infinity. For the exact argument for this Joint Limit convergence, see Kim (2010). The second term follows Lindeberg-Feller CLT

$$
\frac{1}{\sqrt{\rho}\bar{A}_{\gamma\gamma}}\frac{1}{\sqrt{N}}\sum_{i=1}^N \Delta T_i(\dot{\gamma}_i^2 - \dot{\gamma}_i h_i' A_{HH}^{-1} A_H \dot{\gamma}) \xrightarrow{d} N\left(0, \frac{\sigma_b^2 Q_{\gamma\gamma\gamma\gamma}}{\rho \bar{A}_{\gamma\gamma}^2}\right)
$$

Since these two terms are independent, we have

$$
m_T^* \xrightarrow{d} N\left(0, \frac{4\bar{S}_{\gamma\gamma}}{(1-\lambda_0)\lambda_0\bar{A}_{\gamma\gamma}^2} + \frac{\sigma_b^2 Q_{\gamma\gamma\gamma\gamma}}{\rho\bar{A}_{\gamma\gamma}^2}\right).
$$

This is arbitrarily close to the limiting distribution of $T^{3/2}(\widetilde{\lambda}-\lambda_0)$ since C can be any large number.

For (ii), Consider Model II. Let $m_T = (T_b - T_0)$ and $D(C) = \{T_b : |T_b - T_0| < C\}$ for a positive number C. On the set $D(C)$, from (A.19) and (A.20),

$$
\begin{aligned}
(\hat{X}\hat{X})_{11} &= \kappa(T_b)'(I-P_{T_b})\kappa(T_b)\ (\gamma(I-\hat{H}'(\hat{H}\hat{H}')^{-1}\hat{H})\gamma') \\
&\quad +2\kappa(T_b)'(I-P_{T_b})\alpha(T_b)\ (\theta^a(I-\hat{H}'(\hat{H}\hat{H}')^{-1}\hat{H})\gamma') \\
&\quad +\alpha(T_b)'(I-P_{T_b})\alpha(T_b)\ (\theta^a(I-\hat{H}'(\hat{H}\hat{H}')^{-1}\hat{H})\theta^{a\prime}) \\
&= \kappa(T_b)'\kappa(T_b)\ (\gamma(I-\hat{H}'(\hat{H}\hat{H}')^{-1}\hat{H})\gamma') \\
&\quad +2\kappa(T_b)'\alpha(T_b)\ (\theta^a(I-\hat{H}'(\hat{H}\hat{H}')^{-1}\hat{H})\gamma') \\
&\quad +\alpha(T_b)'\alpha(T_b)\ (\theta^a(I-\hat{H}'(\hat{H}\hat{H}')^{-1}\hat{H})\theta^{a\prime}) + o_p(1) \\
&= \begin{cases} \sum_{t=T_b+1}^{T_0}\left(\bar{A}_{\theta\theta} + \sigma_b^2 A_{\gamma\gamma} + \bar{A}_{\gamma\gamma}(t-T_0)^2 + 2\bar{A}_{\gamma\theta}(t-T_0)\right) + o_p(1) & \text{if } T_b < T_0 \\ \sum_{t=T_0+1}^{T_b}\left(\bar{A}_{\theta\theta} + \sigma_b^2 A_{\gamma\gamma} + \bar{A}_{\gamma\gamma}(t-T_0)^2 + 2\bar{A}_{\gamma\theta}(t-T_0)\right) + o_p(1) & \text{if } T_b > T_0 \end{cases}
\end{aligned}
$$

because

$$\begin{aligned}
& \theta^a(I-\hat{H}'(\hat{H}\hat{H}')^{-1}\hat{H})\gamma' \\
= \;& \bar{A}_{\gamma\theta}+\frac{1}{N}\sum_{i=1}^{N}\Delta T_i\dot{\gamma}_i^2-\frac{1}{N}\sum_{i=1}^{N}\Delta T_i\dot{\gamma}_i\hat{h}_i'\left(\frac{1}{N}\sum_{i=1}^{N}\hat{h}_i\hat{h}_i'\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\hat{h}_i\dot{\gamma}_i+o_p(1) \\
= \;& \bar{A}_{\gamma\theta}+o_p(1)
\end{aligned}$$

and similarly

$$\begin{aligned}
& \theta^a(I-\hat{H}'(\hat{H}\hat{H}')^{-1}\hat{H})\theta^{a\prime} \\
= \;& \bar{A}_{\theta\theta}+\frac{1}{N}\sum_{i=1}^{N}(\Delta T_i)^2\dot{\gamma}_i^2-\frac{1}{N}\sum_{i=1}^{N}\Delta T_i\dot{\gamma}_i\hat{h}_i'\left(\frac{1}{N}\sum_{i=1}^{N}\hat{h}_i\hat{h}_i'\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\hat{h}_i\dot{\gamma}_i\Delta T_i+o_p(1) \\
= \;& \bar{A}_{\theta\theta}+\sigma_b^2A_{\gamma\gamma}+o_p(1).
\end{aligned}$$

For $(\hat{X}\hat{X})_{12}$,

$$\begin{aligned}
(\hat{X}\hat{X})_{12} &= tr\left[M_{\hat{H}'}\theta^{a\prime}\alpha(T_b)'(I-P_{T_b})Z\right]+tr\left[M_{\hat{H}'}\gamma'\kappa(T_b)'(I-P_{T_b})Z\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\alpha(T_b)'\alpha(T_i)(\dot{\theta}_i+\Delta T_i\dot{\gamma}_i)(\dot{\theta}_i-h_i'A_{HH}^{-1}A_{H\dot{\theta}}+\Delta T_i\dot{\gamma}_i) \\
&\quad +\frac{1}{N}\sum_{i=1}^{N}\alpha(T_b)'\kappa(T_i)\dot{\gamma}_i(\dot{\theta}_i-h_i'A_{HH}^{-1}A_{H\dot{\theta}}) \\
&\quad +\frac{1}{N}\sum_{i=1}^{N}\kappa(T_b)'\kappa(T_i)\dot{\gamma}_i(\dot{\gamma}_i-h_i'A_{HH}^{-1}A_{H\dot{\gamma}}) \\
&\quad +\frac{1}{N}\sum_{i=1}^{N}\kappa(T_b)'\alpha(T_i)(\dot{\theta}_i+\Delta T_i\dot{\gamma}_i)(\dot{\gamma}_i-h_i'A_{HH}^{-1}A_{H\dot{\gamma}}) \\
&\quad +o_p(1)
\end{aligned}$$

Now, note that

$$\begin{aligned}
\mathcal{E}\alpha_t(T_i) &= \begin{cases} P(\Delta T_i\geq t-T_0) & \text{if } t\geq T_0+1 \\ -P(\Delta T_i+1\leq t-T_0) & \text{if } t\leq T_0 \end{cases} \\
\mathcal{E}\kappa_t(T_i) &= \begin{cases} (t-T_0)P(\Delta T_i\geq t-T_0) & \text{if } t\geq T_0+1 \\ (T_0-t)P(\Delta T_i+1\leq t-T_0) & \text{if } t\leq T_0 \end{cases}
\end{aligned}$$

Hence,

$$
\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{N}\alpha(T_b)'\alpha(T_i)(\dot\theta_i+\Delta T_i\dot\gamma_i)(\dot\theta_i-h_i'A_{HH}^{-1}A_{H\dot\theta}+\Delta T_i\dot\gamma_i) \\
= & \frac{1}{N}\sum_{i=1}^{N}\alpha(T_b)'\alpha(T_i)\dot\theta_i(\dot\theta_i-h_i'A_{HH}^{-1}A_{H\dot\theta}+\Delta T_i\dot\gamma_i) \\
& +\frac{1}{N}\sum_{i=1}^{N}\alpha(T_b)'\alpha(T_i)\Delta T_i\dot\gamma_i(\dot\theta_i-h_i'A_{HH}^{-1}A_{H\dot\theta}+\Delta T_i\dot\gamma_i) \\
= & \alpha(T_b)'\mathcal{E}\alpha(T_i)\frac{1}{N}\sum_{i=1}^{N}\left[\dot\theta_i(\dot\theta_i-h_i'A_{HH}^{-1}A_{H\dot\theta})+(\Delta T_i\dot\gamma_i)^2\right]+o_p(1) \\
= & \begin{cases} (\bar A_{\theta\theta}+\sigma_b^2A_{\gamma\gamma})\sum_{t=T_0+1}^{T_b}P(\Delta T_i\geq t-T_0)+o_p(1) & \text{if } T_b\geq T_0+1 \\ (\bar A_{\theta\theta}+\sigma_b^2A_{\gamma\gamma})\sum_{t=T_b+1}^{T_0}P(\Delta T_i+1\leq t-T_0)+o_p(1) & \text{if } T_b+1\leq T_0 \end{cases}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{N}\alpha(T_b)'\kappa(T_i)\dot\gamma_i(\dot\theta_i-h_i'A_{HH}^{-1}A_{H\dot\theta}) \\
& \text{and } \frac{1}{N}\sum_{i=1}^{N}\kappa(T_b)'\alpha(T_i)(\dot\theta_i+\Delta T_i\dot\gamma_i)(\dot\gamma_i-h_i'A_{HH}^{-1}A_{H\dot\gamma}) \\
= & \begin{cases} \bar A_{\gamma\theta}\sum_{t=T_0+1}^{T_b}(t-T_0)P(\Delta T_i\geq t-T_0)+o_p(1) & \text{if } T_b\geq T_0+1 \\ \bar A_{\gamma\theta}\sum_{t=T_b+1}^{T_0}(t-T_0)P(\Delta T_i+1\leq t-T_0)+o_p(1) & \text{if } T_b+1\leq T_0 \end{cases}
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{N}\kappa(T_b)'\kappa(T_i)\dot\gamma_i(\dot\gamma_i-h_i'A_{HH}^{-1}A_{H\dot\gamma}) \\
= & \begin{cases} \bar A_{\gamma\gamma}\sum_{t=T_0+1}^{T_b}(t-T_0)^2P(\Delta T_i\geq t-T_0)+o_p(1) & \text{if } T_b\geq T_0+1 \\ \bar A_{\gamma\gamma}\sum_{t=T_b+1}^{T_0}(t-T_0)^2P(\Delta T_i+1\leq t-T_0)+o_p(1) & \text{if } T_b+1\leq T_0 \end{cases}
\end{aligned}
$$

Now,

$$
\begin{aligned}
(\hat X\hat U)_1 &= tr\left[(I-\hat H'(\hat H\hat H')^{-1}\hat H)(\theta^{a\prime}\alpha(T_b)'+\gamma'\kappa(T_b)')(I-P_{T_b})(\bar F\bar H+U)\right] \\
&= tr\left[(I-\hat H'(\hat H\hat H')^{-1}\hat H)\theta^{a\prime}\alpha(T_b)'E\right]+tr\left[(I-\hat H'(\hat H\hat H')^{-1}\hat H)\gamma'\kappa(T_b)'E\right]+o_p(1) \\
&= \begin{cases} -\sum_{t=T_b+1}^{T_0}\begin{bmatrix} \left(N^{-1/2}\sum_{i=1}^{N}e_{ti}(\bar\theta_i+\Delta T_i\dot\gamma_i)\right) \\ +(t-T_0)\left(N^{-1/2}\sum_{i=1}^{N}e_{ti}\bar\gamma_i\right)\end{bmatrix}+o_p(1) & \text{if } T_b<T_0 \\ \sum_{t=T_0+1}^{T_b}\begin{bmatrix} \left(N^{-1/2}\sum_{i=1}^{N}e_{ti}(\bar\theta_i+\Delta T_i\dot\gamma_i)\right) \\ +(t-T_0)\left(N^{-1/2}\sum_{i=1}^{N}e_{ti}\bar\gamma_i\right)\end{bmatrix}+o_p(1) & \text{if } T_b>T_0 \end{cases}
\end{aligned}
$$

where $\bar{\theta}_i$ is the ith element of $(I - H'(HH')^{-1}H)\dot{\theta}'$ and $\bar{\gamma}_i$ is the ith element of $(I - H'(HH')^{-1}H)\dot{\gamma}'$. Furthermore, $(UU) = o_p(1)$ if $N^2/T \to 0$.

Let $\bar{\Xi}_1 = \lim_N N^{-1} \sum_{i=1}^N (\bar{\theta}_i^2 + \sigma_b^2 \dot{\gamma}_i^2)\Gamma_i$ and $\bar{\Xi}_2 = \lim_N N^{-1} \sum_{i=1}^N \bar{\gamma}_i^2 \Gamma_i$ where Γ_i is $2C \times 2C$ matrix whose (p,q) element is the autocovariance of e_{ti} at $p-q$ lag. Also, define $N_1 = B_1 W$ and $N_2 = B_2 W$ where $\bar{\Xi}_1 = B_1 B_1'$, $\bar{\Xi}_2 = B_2 B_2'$ and $W \sim N(0, I_{2C})$.

Then, from the Lindeberg-Feller CLT and Cramer-Wold device,

$$N^{-1/2} \sum_{i=1}^N \begin{pmatrix} e_{T_b - C+1,i} \\ \vdots \\ e_{T_b+C,i} \end{pmatrix} (\bar{\theta}_i + \Delta T_i \dot{\gamma}_i) \xrightarrow{d} N_1$$

and

$$N^{-1/2} \sum_{i=1}^N \begin{pmatrix} e_{T_b - C+1,i} \\ \vdots \\ e_{T_b+C,i} \end{pmatrix} \bar{\gamma}_i \xrightarrow{d} N_2.$$

Therefore, on the set $D(C)$, if $N^2/T \to 0$,

$$\begin{aligned} m_T^* &= \arg \min_{m_T \text{ on } D(C)} \left[(\hat{X}\hat{X})_{11} - 2(\hat{X}\hat{X})_{12} + 2(\hat{X}\hat{U})_1 + o_p(1) \right] \\ &= \arg \min_{m_T \text{ on } D(C)} \left[V^*(m) + o_p(1) \right] \end{aligned}$$

where the stochastic process $V^*(m)$ is such that $V^*(0) = 0$, $V^*(m) = V_1(m)$ for $m < 0$ and $V^*(m) = V_2(m)$ for $m > 0$, with

$$\begin{aligned} V_1(m) &= \sum_{k=m+1}^{0} \left[\left(\bar{A}_{\theta\theta} + \sigma_b^2 \dot{A}_{\gamma\gamma} \right) + \bar{A}_{\gamma\gamma} k^2 + 2\bar{A}_{\gamma\theta} k \right] [1 - 2P(\Delta T_i + 1 \le k)] \\ &\quad -2 \sum_{k=m+1}^{0} (N_{1,k} + k N_{2,k}) \quad \text{for } m = -1, -2, \ldots \\ V_2(m) &= \sum_{k=1}^{m} \left[(\bar{A}_{\theta\theta} + \sigma_b^2 \dot{A}_{\gamma\gamma}) + \bar{A}_{\gamma\gamma} k^2 + 2\bar{A}_{\gamma\theta} k \right] [1 - 2P(\Delta T_i \ge k)] \\ &\quad +2 \sum_{k=1}^{m} (N_{1,k} + k N_{2,k}) \quad \text{for } m = 1, 2, \ldots \end{aligned}$$

The limiting distribution of m_T^* is arbitrarily close to that of $T(\widetilde{\lambda} - \lambda_0)$ for large C. ■

Table 1. Root Mean Squared Errors of $\hat{T}_0$ and $\widetilde{T}_0$

a. Model I

γ_i	T		$N=20$	50	100	γ_i	T		$N=20$	50	100
0.1	100	$\hat{T}_0$	15.71	15.01	15.49	0.5	100	$\hat{T}_0$	2.18	2.03	2.13
		$\widetilde{T}_0$	*3.56*	*2.12*	*1.48*			$\widetilde{T}_0$	*0.67*	*0.63*	*0.33*
	200	$\hat{T}_0$	8.64	8.26	8.72		200	$\hat{T}_0$	1.46	1.40	1.43
		$\widetilde{T}_0$	*1.92*	*1.31*	*0.87*			$\widetilde{T}_0$	*0.46*	*0.62*	*0.19*
	300	$\hat{T}_0$	6.18	5.95	6.28		300	$\hat{T}_0$	1.21	1.17	1.18
		$\widetilde{T}_0$	*1.55*	*1.07*	*0.74*			$\widetilde{T}_0$	*0.35*	*0.55*	*0.11*
	500	$\hat{T}_0$	4.57	4.58	4.56		500	$\hat{T}_0$	0.95	0.92	0.90
		$\widetilde{T}_0$	*1.16*	*0.86*	*0.60*			$\widetilde{T}_0$	*0.21*	*0.51*	*0.05*
0.3	100	$\hat{T}_0$	3.83	3.60	3.85	0.7	100	$\hat{T}_0$	1.55	1.47	1.46
		$\widetilde{T}_0$	*1.03*	*0.78*	*0.54*			$\widetilde{T}_0$	*0.52*	*0.63*	*0.25*
	200	$\hat{T}_0$	2.47	2.40	2.39		200	$\hat{T}_0$	1.05	1.02	1.01
		$\widetilde{T}_0$	*0.69*	*0.66*	*0.34*			$\widetilde{T}_0$	*0.32*	*0.58*	*0.14*
	300	$\hat{T}_0$	1.96	1.86	1.95		300	$\hat{T}_0$	0.88	0.84	0.87
		$\widetilde{T}_0$	*0.61*	*0.60*	*0.30*			$\widetilde{T}_0$	*0.22*	*0.54*	*0.08*
	500	$\hat{T}_0$	1.52	1.51	1.47		500	$\hat{T}_0$	0.71	0.70	0.69
		$\widetilde{T}_0$	*0.45*	*0.59*	*0.20*			$\widetilde{T}_0$	*0.09*	*0.51*	*0.00*

1. T_i, the true break date in each equation is drawn from $N(0.5T, 2)$ and rounded to the nearest interger. 2. The break date estimators, $\hat{T}_0$ and $\widetilde{T}_0$ are defined in (2) and (4). 3. DGP: $y_{ti} = \gamma_i B_{ti} + h_i' F_t + e_{ti}$, $F_t = 0.6F_{t-1} + w_t$ with $w_t \sim iid\ N(0,1)$, h_i is drawn from $U(0,2)$, $e_{ti} = \rho_i e_{t-1i} + \varepsilon_{ti}$ with $\varepsilon_{ti} \sim iid\ N(0,1)$, and ρ_i is drawn from $U(0, 0.5)$. 4. For each N value, one set of $\{(h_i, T_i, \rho_i),\ i = 1, \ldots, N\}$ is kept for all T values and Monte Carlo repetitions. 5. The number of replications is 2,000.

Table 1. Root Mean Squared Errors of $\hat{T}_0$ and $\widetilde{T}_0$ (continued)

b. Model II

θ_i and γ_i	T		$N=20$	50	100	θ_i and γ_i	T		$N=20$	50	100
0.1	100	$\hat{T}_0$	18.28	18.12	18.84	0.7	100	$\hat{T}_0$	3.11	3.12	3.04
		$\widetilde{T}_0$	*7.60*	*5.88*	*4.75*			$\widetilde{T}_0$	*1.47*	*1.31*	*1.19*
	200	$\hat{T}_0$	16.51	16.56	16.32		200	$\hat{T}_0$	2.89	2.98	2.89
		$\widetilde{T}_0$	*6.24*	*5.03*	*4.10*			$\widetilde{T}_0$	*1.42*	*1.29*	*1.15*
	300	$\hat{T}_0$	15.30	14.77	14.91		300	$\hat{T}_0$	2.75	2.99	2.90
		$\widetilde{T}_0$	*5.68*	*4.65*	*3.80*			$\widetilde{T}_0$	*1.42*	*1.29*	*1.14*
	500	$\hat{T}_0$	14.31	13.85	14.13		500	$\hat{T}_0$	2.75	2.88	2.94
		$\widetilde{T}_0$	*5.52*	*4.31*	*3.44*			$\widetilde{T}_0$	*1.38*	*1.25*	*1.14*
0.4	100	$\hat{T}_0$	5.21	5.11	5.29	1.4	100	$\hat{T}_0$	1.59	1.67	1.70
		$\widetilde{T}_0$	*2.13*	*1.78*	*1.54*			$\widetilde{T}_0$	*1.13*	*1.04*	*1.01*
	200	$\hat{T}_0$	4.89	4.72	4.69		200	$\hat{T}_0$	1.59	1.63	1.65
		$\widetilde{T}_0$	*1.95*	*1.65*	*1.42*			$\widetilde{T}_0$	*1.15*	*1.05*	*1.00*
	300	$\hat{T}_0$	4.57	4.67	4.54		300	$\hat{T}_0$	1.59	1.67	1.59
		$\widetilde{T}_0$	*1.87*	*1.60*	*1.37*			$\widetilde{T}_0$	*1.14*	*1.05*	*1.01*
	500	$\hat{T}_0$	4.52	4.70	4.62		500	$\hat{T}_0$	1.53	1.63	1.59
		$\widetilde{T}_0$	*1.80*	*1.52*	*1.36*			$\widetilde{T}_0$	*1.13*	*1.05*	*1.00*

1. T_i, the true break date in each equation is drawn from $N(0.5T, 2)$ and rounded to the nearest interger. 2. The break date estimators, $\hat{T}_0$ and $\widetilde{T}_0$ are defined in (2) and (4). 3. DGP: $y_{ti} = \theta_i C_{ti} + \gamma_i B_{ti} + h_i' F_t + e_{ti}$, $F_t = 0.6F_{t-1} + w_t$ with $w_t \sim iid\ N(0,1)$, h_i is drawn from $U(0,2)$, $e_{ti} = \rho_i e_{t-1i} + \varepsilon_{ti}$ with $\varepsilon_{ti} \sim iid\ N(0,1)$, and ρ_i is drawn from $U(0,0.5)$. 4. For each N value, one set of $\{(h_i, T_i, \rho_i),\ i = 1, \ldots, N\}$ is kept for all T values and Monte Carlo repetitions. 5. The number of replications is 2,000.

Table 1. Root Mean Squared Errors of $\hat{T}_0$ and $\widetilde{T}_0$ (continued)

c. Model III

θ_i	T		$N=20$	50	100	θ_i	T		$N=20$	50	100
1.0	100	$\hat{T}_0$	25.61	25.79	26.06	2.0	100	$\hat{T}_0$	21.41	20.66	20.60
		$\widetilde{T}_0$	*17.30*	*11.34*	*6.40*			$\widetilde{T}_0$	*2.45*	*1.25*	*0.93*
	200	$\hat{T}_0$	51.23	52.13	51.03		200	$\hat{T}_0$	32.58	32.78	32.44
		$\widetilde{T}_0$	*19.82*	*4.65*	*1.43*			$\widetilde{T}_0$	*1.37*	*1.08*	*0.82*
	300	$\hat{T}_0$	74.98	73.64	75.56		300	$\hat{T}_0$	37.04	35.08	37.30
		$\widetilde{T}_0$	*11.05*	*1.92*	*1.28*			$\widetilde{T}_0$	*1.29*	*1.02*	*0.79*
	500	$\hat{T}_0$	115.69	114.18	112.81		500	$\hat{T}_0$	33.80	31.79	32.74
		$\widetilde{T}_0$	*3.16*	*1.66*	*1.13*			$\widetilde{T}_0$	*1.24*	*1.02*	*0.75*
1.4	100	$\hat{T}_0$	24.35	23.94	24.32	5.0	100	$\hat{T}_0$	3.72	3.28	3.45
		$\widetilde{T}_0$	*9.88*	*3.54*	*1.18*			$\widetilde{T}_0$	*0.99*	*1.02*	*0.83*
	200	$\hat{T}_0$	46.22	44.91	44.65		200	$\hat{T}_0$	1.46	1.32	1.26
		$\widetilde{T}_0$	*4.51*	*1.31*	*0.98*			$\widetilde{T}_0$	*0.93*	*0.89*	*0.72*
	300	$\hat{T}_0$	61.14	59.89	60.89		300	$\hat{T}_0$	1.38	1.34	1.33
		$\widetilde{T}_0$	*1.92*	*1.23*	*0.94*			$\widetilde{T}_0$	*0.89*	*0.87*	*0.66*
	500	$\hat{T}_0$	79.89	73.80	79.33		500	$\hat{T}_0$	1.37	1.30	1.21
		$\widetilde{T}_0$	*1.85*	*1.17*	*0.90*			$\widetilde{T}_0$	*0.86*	*0.87*	*0.63*

1. T_i, the true break date in each equation is drawn from $N(0.5T, 2)$ and rounded to the nearest interger. 2. The break date estimators, $\hat{T}_0$ and $\widetilde{T}_0$ are defined in (2) and (4). 3. DGP: $y_{ti} = \theta_i C_{ti} + h_i' F_t + e_{ti}$, $F_t = 0.6F_{t-1} + w_t$ with $w_t \sim iid\ N(0,1)$, h_i is drawn from $U(0,2)$, $e_{ti} = \rho_i e_{t-1i} + \varepsilon_{ti}$ with $\varepsilon_{ti} \sim iid\ N(0,1)$, and ρ_i is drawn from $U(0, 0.5)$. 4. For each N value, one set of $\{(h_i, T_i, \rho_i),\ i = 1, \ldots, N\}$ is kept for all T values and Monte Carlo repetitions. 5. The number of replications is 2,000.

Table 2. Coverage Rates of Confidence Intervals for $\hat{T}_0$ and $\widetilde{T}_0$

a. Model I

γ_i	T		$N=20$	50	100	γ_i	T		$N=20$	50	100
0.1	100	$\hat{T}_0$	0.66	0.65	0.64	0.5	100	$\hat{T}_0$	0.92	0.92	0.92
		$\widetilde{T}_0$	*0.87*	*0.92*	*0.95*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(2.70)	(3.77)	(5.25)				(2.93)	(3.20)	(3.23)
	200	$\hat{T}_0$	0.84	0.84	0.84		200	$\hat{T}_0$	0.97	0.98	0.98
		$\widetilde{T}_0$	*0.97*	*0.98*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(3.24)	(4.73)	(5.65)				(2.73)	(2.70)	(2.75)
	300	$\hat{T}_0$	0.91	0.91	0.89		300	$\hat{T}_0$	0.98	0.98	0.98
		$\widetilde{T}_0$	*0.98*	*0.99*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(3.27)	(4.86)	(5.20)				(2.30)	(2.26)	(2.32)
	500	$\hat{T}_0$	0.94	0.94	0.94		500	$\hat{T}_0$	1.00	1.00	0.99
		$\widetilde{T}_0$	*0.99*	*1.00*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(3.22)	(4.00)	(7.63)				(2.01)	(2.00)	(2.01)
0.3	100	$\hat{T}_0$	0.86	0.86	0.85	0.7	100	$\hat{T}_0$	0.96	0.95	0.95
		$\widetilde{T}_0$	*0.99*	*0.99*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(2.53)	(4.27)	(4.99)				(2.42)	(2.38)	(2.44)
	200	$\hat{T}_0$	0.94	0.94	0.94		200	$\hat{T}_0$	0.99	0.99	0.99
		$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(2.68)	(4.08)	(4.18)				(2.06)	(2.04)	(2.05)
	300	$\hat{T}_0$	0.97	0.96	0.96		300	$\hat{T}_0$	1.00	1.00	1.00
		$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(3.46)	(3.60)	(3.67)				(1.99)	(1.99)	(1.99)
	500	$\hat{T}_0$	0.98	0.98	0.99		500	$\hat{T}_0$	1.00	1.00	1.00
		$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(3.05)	(3.01)	(3.05)				(1.79)	(1.74)	(1.80)

1. T_i, the true break date in each equation is $0.5T$. 2. The break date estimators, $\hat{T}_0$ and $\widetilde{T}_0$ are defined in (2) and (4). 3. The numbers in parenthesis are the average length of the confidence interval for $\hat{T}_0$ over that for $\widetilde{T}_0$. 4. DGP: $y_{ti} = \gamma_i B_{ti} + h_i' F_t + e_{ti}$, $F_t = 0.6F_{t-1} + w_t$ with $w_t \sim iid\ N(0,1)$, h_i is drawn from $U(0,2)$, $e_{ti} = \rho_i e_{t-1i} + \varepsilon_{ti}$ with $\varepsilon_{ti} \sim iid\ N(0,1)$, and ρ_i is drawn from $U(0,0.5)$. 5. For each N value, one set of $\{(h_i,\rho_i),\ i=1,\ldots,N\}$ is kept for all T values and Monte Carlo repetitions. 6. The number of replications is 2,000.

Table 2. Coverage Rates of Confidence Intervals for $\hat{T}_0$ and $\widetilde{T}_0$ (continued)

b. Model III

θ_i	T		$N=20$	50	100	θ_i	T		$N=20$	50	100
1.0	100	$\hat{T}_0$	0.29	0.31	0.30	2.0	100	$\hat{T}_0$	0.52	0.52	0.52
		$\widetilde{T}_0$	*0.42*	*0.71*	*0.91*			$\widetilde{T}_0$	*0.98*	*1.00*	*1.00*
			(6.69)	(8.74)	(8.55)				(6.86)	(6.94)	(6.74)
	200	$\hat{T}_0$	0.31	0.31	0.31		200	$\hat{T}_0$	0.66	0.70	0.68
		$\widetilde{T}_0$	*0.72*	*0.89*	*0.98*			$\widetilde{T}_0$	*0.99*	*1.00*	*1.00*
			(5.89)	(11.84)	(11.98)				(9.14)	(8.95)	(9.22)
	300	$\hat{T}_0$	0.33	0.34	0.32		300	$\hat{T}_0$	0.77	0.78	0.77
		$\widetilde{T}_0$	*0.83*	*0.90*	*0.99*			$\widetilde{T}_0$	*0.99*	*1.00*	*1.00*
			(5.24)	(12.84)	(12.99)				(9.97)	(9.73)	(10.05)
	500	$\hat{T}_0$	0.38	0.38	0.38		500	$\hat{T}_0$	0.87	0.88	0.88
		$\widetilde{T}_0$	*0.90*	*0.95*	*0.99*			$\widetilde{T}_0$	*0.99*	*1.00*	*1.00*
			(4.77)	(13.50)	(13.55)				(10.57)	(10.41)	(10.63)
1.4	100	$\hat{T}_0$	0.39	0.39	0.40	5.0	100	$\hat{T}_0$	0.98	0.98	0.97
		$\widetilde{T}_0$	*0.79*	*0.97*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(7.62)	(8.22)	(8.07)				(1.08)	(1.05)	(1.08)
	200	$\hat{T}_0$	0.44	0.47	0.44		200	$\hat{T}_0$	0.98	0.98	0.98
		$\widetilde{T}_0$	*0.91*	*0.99*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(10.00)	(11.12)	(11.31)				(1.06)	(1.04)	(1.05)
	300	$\hat{T}_0$	0.53	0.53	0.52		300	$\hat{T}_0$	0.98	0.98	0.98
		$\widetilde{T}_0$	*0.93*	*1.00*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(10.26)	(12.25)	(12.28)				(1.04)	(1.03)	(1.04)
	500	$\hat{T}_0$	0.63	0.63	0.64		500	$\hat{T}_0$	0.98	0.98	0.97
		$\widetilde{T}_0$	*0.95*	*1.00*	*1.00*			$\widetilde{T}_0$	*1.00*	*1.00*	*1.00*
			(10.62)	(13.00)	(13.05)				(1.03)	(1.01)	(1.02)

1. T_i, the true break date in each equation is $0.5T$. 2. The break date estimators, $\hat{T}_0$ and $\widetilde{T}_0$ are defined in (2) and (4). 3. The numbers in parenthesis are the average length of the confidence interval for $\hat{T}_0$ over that for $\widetilde{T}_0$. 4. DGP: $y_{ti} = \theta_i C_{ti} + h_i' F_t + e_{ti}$, $F_t = 0.6F_{t-1} + w_t$ with $w_t \sim iid\ N(0,1)$, h_i is drawn from $U(0,2)$, $e_{ti} = \rho_i e_{t-1i} + \varepsilon_{ti}$ with $\varepsilon_{ti} \sim iid\ N(0,1)$, and ρ_i is drawn from $U(0,0.5)$. 5. For each N value, one set of $\{(h_i, \rho_i),\ i = 1, \ldots, N\}$ is kept for all T values and Monte Carlo repetitions. 6. The number of replications is 2,000.