Estimating a Common Deterministic Time Trend Break in Large Panels with Cross Sectional Dependence

Dukpa Kim*

University of Virginia

October 17, 2009

Abstract

This paper develops an estimation procedure for a common deterministic time trend break in large panels with cross-sectional dependence. The deterministic trend is assumed to have a change in the slope or in both the slope and intercept. The date of the change is common for all equations. The estimation method is simply minimizing the sum of squared residuals (SSR) for all possible break dates. The consistency and the asymptotic distribution of the proposed break date estimate are shown under various assumptions on the error process. In the time dimension, the error process in each equation is assumed to be either a stationary linear process or a unit root process. In the cross sectional span, the errors are assumed to be either independent across equations or strongly correlated through the common factor structure. The amount of serial correlation in each equation and the cross sectional correlation are important factors that decide the statistical properties of the proposed break date estimate. Strong correlation in both the time and cross sectional span slows down the rate of convergence of the break date estimate. The rate of convergence is faster in the case of stationary error process than in the presence of a unit root. In the absence of the common components, the rate of convergence of the break date estimate depends on the number of equations and thus is faster than the univariate case. When the common factors exist and the factor loadings are correlated with the slope change parameters, the rate of convergence reduces to the univariate case. The limiting distribution of the break date estimate is normal if a break is allowed to occur only in the slopes irrespective of the existence of common components and a unit root. When a change is allowed in both the slope and intercept, the limiting distribution is highly non-standard but analogous to the univariate case.

JEL Classification Number: **C33**
Keywords: structural break, deterministic trend, panel data

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

1 Introduction

While the issues related to structural breaks have drawn a lot of attention in both econometrics and statistics, relatively small attention has been paid to common breaks in large panels. The extension of structural break models to the panel data setup is important because a structural break is regarded as an exogenous shock with permanent effects on the distributions of economic variables and such a shock is likely to have impacts on many economic variables simultaneously. However, the econometric tools currently available are mostly designed for univariate models, for example Bai (1997) and Bai and Perron (1998), or multivariate models where the number of equations are fixed, for example Bai et. al (1998).

In this paper, we develop an estimation procedure for a common deterministic time trend break in large panels with cross sectional dependence. 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 or in both the slope and intercept. The date of the change is common for all equations. The estimation method is simply minimizing the sum of squared residuals (SSR) for all possible break dates and choosing the date with the minimum SSR.

The statistical properties of our break date estimate heavily depends on the specification of the error process. Especially the amount of serial correlation in each equation and the cross sectional correlation are important factors. In order to be able to allow for a strong form of cross equation correlation, we assume that the error process has the common factor structure analyzed in Bai and Ng (2004). That is, the error process is the sum of an individual specific error and the common factors scaled by an individual loading vector. Also, for the serial correlation within each equation, both the common factors and the individual specific errors are assumed to be either a stationary linear process or a unit root process.

Under various combination of the aforementioned specifications, we establish the consistency of the proposed break date estimate and derive its limiting distribution. The employed asymptotic framework is the joint limit theory analyzed in Phillips and Moon (1999) where both the time span and the number of equations grow without assuming any particular growth path. The joint limit theory is often preferred to the sequential limit theory where the time span tends to infinity first and the cross sectional span follows, since the results obtained under the former are more robust to the relative size of the time and cross sectional span.

We show that strong correlation in both the time and cross sectional span slows down the rate of convergence of the break date estimate. The rate of convergence is slower in the presence of a unit root than the case of stationary error process. This finding is in paralell with the univariate results shown by Perron and Zhu (2005). In the absence of the common components, we show that the rate of convergence of the break date estimate depends on the number of equations and thus is faster than that in the univariate case. This is exactly the benefit of having a panel of data. On the other hand, when the common factors exist and the factor loadings are correlated with the slope change parameters, the rate of convergence reduces to the univariate case and the efficiency gain in terms of the rate of convergence disappears.

The limiting distribution of the break date estimate is normal if a break is assumed to occur only in the slopes irrespective of the existence of common components or a unit root. When a change is allowed in both the slope and intercept, the limiting distribution is highly non-standard but analogous to the univariate case of Perron and Zhu (2005). When there are common components, the limiting distribution of the break date estimate depends on the longrun variance of the common factors. We discuss a simply method to estimate the longrun variance of the common factors which does not require the estimation of the common factors themselves.

There are two important papers related to this paper. Perron and Zhu (2005) analyzed the deterministic trend break in the univariate models. Indeed, our models are panel extensions of Perron and Zhu (2005). Bai (2008) analyzed a common mean shift in a panel data model. Bai's (2008) work differs from ours in that it concerns only mean shift without a linear time trend and the errors are only stationary without common components. Also, Bai (2008) analyzes the variance break but we do not.

The rest of the paper is organized as the following. Section 2 introduces the models and assumptions. Section 3 presents the main theoretical results. Section 4 demonstrates the validity of the asymptotic results via Monte Carlo experiments. Section 5 shows an empirical illustration. Section 6 contains concluding remarks. All mathematical derivations and proofs are collected in the appendix.

2 Model and Assumptions

We consider a panel data model where 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 two cases:

$$d_{ti} = \begin{cases} \mu_i + \beta_i t + \gamma_i B_t & \text{Model I} \\ \mu_i + \beta_i t + \theta_i C_t + \gamma_i B_t & \text{Model II} \end{cases}$$

where

$$C_t = \begin{cases} 0 & \text{if } t \leq T_1 \\ 1 & \text{if } t > T_1 \end{cases}, \text{ and } B_t = \begin{cases} 0 & \text{if } t \leq T_1 \\ t - T_1 & \text{if } t > T_1 \end{cases}.$$

The regression coefficients are not restricted to be common across equations, and instead of pooling the data we apply the ordinary least squares procedure to each equation. In terms of Perron and Zhu's (2005) terminology, Models I and II are respectively referred to as the joint broken trend and local disjoint trend model. Perron and Zhu (2005) analyzed another model, namely, the global disjoint trend model which uses B_t^{dj} instead of B_t, where $B_t^{dj} = 0$, if $t \leq T_1$ and t, otherwise. We exclude this model from our analysis because in such a model the estimate for the break date is already consistent at an extremely fast rate only with a univariate time series and there doesn't seem to be much efficiency gain with practical importance due to the panel structure.

Assumption 1 *The true break date T_1 is unknown and the break fraction $\lambda_1 = T_1/T \in [\pi, 1-\pi]$, $\pi \in (0, 1/2)$ is fixed for all T.*

Note that the break date T_1 is common for all equations, and the break fraction λ_1 is fixed as the sample size grows. The trimming of the location of λ_1 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.

Now, let $\iota = (1, \ldots, 1)'$, $\tau = (1, \ldots, T)'$, $C = (C_1, \ldots, C_T)'$, $B = (B_1, \ldots, B_T)'$, and $U_i = (u_{i1}, \ldots, u_{iT})'$. Also, $X_{T_1} = [\iota, \tau, B]$ and $\Pi_i = (\mu_i, \beta_i, \gamma_i)'$ for Model I, and $X_{T_1} = [\iota, \tau, C, B]$ and $\Pi_i = (\mu_i, \beta_i, \theta_i, \gamma_i)'$ for Model II. Then, we write each equation in a matrix form as

$$\underset{(T\times 1)}{Y_i} = \underset{(T\times 3 \text{ or } T\times 4)}{X_{T_1}} \underset{(3\times 1 \text{ or } 4\times 1)}{\Pi_i} + \underset{(T\times 1)}{U_i} \tag{1}$$

and the N equations as

$$Y = X_{T_1}\Pi + U \tag{2}$$

where $Y = [Y_1, \ldots, Y_N]$, $\Pi = [\Pi_1, \ldots, \Pi_N]$, and $U = [U_1, \ldots, U_N]$. Also, define row vectors $\mu = (\mu_1, \ldots, \mu_N)$, $\beta = (\beta_1, \ldots, \beta_N)$, $\theta = (\theta_1, \ldots, \theta_N)$ and $\gamma = (\gamma_1, \ldots, \gamma_N)$. Then, an alternative expression for Π is $[\mu', \beta', \gamma']'$ for Model I and $[\mu', \beta', \theta', \gamma']'$ for Model II.

Denote by T_b a generic break date and by $\lambda = T_b/T$ a generic break fraction. X_{T_b} is defined analogously to X_{T_1} and has a break date T_b. Then, the sum of squared residuals for each T_b is given by

$$SSR(T_b) = tr\left[Y'(I - P_{T_b})Y\right]$$

where $P_{T_b} = X_{T_b}(X_{T_b}'X_{T_b})^{-1}X_{T_b}'$, and the estimated break date $\hat{T}_1$ is the date that minimizes the sum of squared residuals:

$$\hat{T}_1 = \arg\min_{T_b} SSR(T_b) \quad \text{with} \quad \hat{\lambda} = \hat{T}_1/T$$

Remark 1 *We point out an important identification issue in Model II. In order to emphasize the break date write $C_t = C_t(T_1)$ and $B_t = B_t(T_1)$. Also, let $D_t(T_1)$ be a one time dummy variable (i.e., $D_t(T_1) = 1$, if $t = T_1 + 1$, and 0, otherwise) Then, note that*

$$\theta C_t(T_1) + \gamma B_t(T_1) = (\theta + \gamma)D_t(T_1) + (\theta + \gamma)C_t(T_1 + 1) + \gamma B_t(T_1 + 1)$$

or

$$\theta C_t(T_1) + \gamma B_t(T_1) = -\theta D_t(T_1 - 1) + (\theta - \gamma)C_t(T_1 - 1) + \gamma B_t(T_1 - 1)$$

This means that, if $\theta = -\gamma$ or 0, there are two break dates (T_1 and $T_1 + 1$ or $T_1 - 1$ and T_1) that can generate exactly the same time trend. This feature is of less concern if we are concerned only with the consistency of the break fraction $\hat{\lambda}_1$. However, with a panel of data we will establish the consistency of break date estimates in the next section, and there it should be understood that the model is assumed not to have any identification issue of this kind.

Remark 2 *An alternative way of locating the break date may be to use the cross-sectional average. If the law of large numbers (LLN) is applicable to u_{ti} across i, then for large N, the cross-sectional average will contain only the average of the deterministic part, and the break point will be even visible. However, this method requires that the averages of the break parameters θ_i and γ_i have a non-zero limit, and we do not analyze this method in this paper. See Assumption 4 below where the averages of the break parameters θ_i and γ_i are allowed to be zero, and Assumption 5 where the LLN doesn't have to apply.*

The statistical properties of $\hat{T}_1$ depend on both the cross sectional and serial correlation of u_{ti}, and we assume that the error component u_{ti} consists of two terms:

$$u_{ti} = h_i' F_t + e_{ti} \tag{3}$$

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. Under the decomposition in (3), we can write (1) and (2) as

$$Y_i = \underset{(T\times 3 \text{ or } T\times 4)}{X_{T_1}} \underset{(3\times 1 \text{ or } 4\times 1)}{\Pi_i} + \underset{(T\times r)}{F} \underset{(r\times 1)}{h_i} + \underset{(T\times 1)}{E_i}$$

and

$$Y = X_{T_1}\Pi + FH + E$$

where $H = [h_1, \ldots, h_N]$, $F = [F_1, \ldots, F_T]'$, and $E = [E_1, \ldots, E_N]$.

As a matter of notation, L denotes a lag operator. For a matrix A, $\|A\| = (tr(A'A))^{1/2}$ with $tr(\cdot)$ denoting the trace of A. We also use $\rightarrow$ to denote the convergence of non-random elements,

$\xrightarrow{p}$ convergence in probability, $\xrightarrow{d}$ convergence in distribution, and $\Rightarrow$ weak convergence in the space $D[0,1]$ under the Skorohod metric. 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 $w_t \sim iid(0, I_r)$ 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 $d_{i0} = 0$, $\sum_{j=0}^{\infty} j\,|d_{ij}| < M$ 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 $\sigma_i^2 < M$.
(iii) F_t and e_{ti} are independent.

Under Assumption 2, the factors and individual specific errors are assumed to be stationary linear processes with one summable coefficients. This linear process assumption makes it particularly convenient to apply the Functional Central Limit Theorem (See Phillips and Solo, 1992) and the Joint Limit Central Limit Theorem (See Phillips and Moon, 1999), which are two of the main tools used to derive the asymptotic distribution of the break date estimate. The above assumption also implies that both the long-run and short-run variances of the individual specific errors are bounded uniformly in i and there is no one series which can dominate all other series. Also, the primitive shocks w_t and ε_{ti} are iid over t and thus F_t and e_{ti} are strictly stationary processes. We show that in the next section that this strict stationarity is an important feature in deriving the asymptotic distributions of the break date estimate in Model II.

An alternative assumption to be used is that the common factors and the individual specific shocks are integrated of order one.

Assumption 3 *(i) $(1-L)F_t$ satisfies Assumption 2(i) and $F_0 = 0$.*
(ii) $(1-L)e_{ti}$ satisfies Assumption 2(ii) and $e_{0i} = 0$.
(iii) F_t and e_{ti} are independent.

In Section 3, we present our asymptotic results with alternating Assumptions 2 and 3. We make the following assumptions for the break parameters and the factor loadings.

Assumption 4 *$N^{-1}\theta\theta' \to A_{\theta\theta} \neq 0$, $N^{-1}\theta\gamma' \to A_{\theta\gamma} \neq 0$, $N^{-1}\gamma\gamma' \to A_{\gamma\gamma} \neq 0$, and $N^{-1}\gamma D \Sigma_\varepsilon D\gamma' \to S_{\gamma\gamma} \neq 0$, where $\Sigma_\varepsilon = diag\{\sigma_1^2, \ldots, \sigma_N^2\}$ and $D = diag\{d_1(1), \ldots, d_N(1)\}$. $\max\{\gamma_1^2, \ldots, \gamma_N^2\} = O(1)$.*

Assumption 5 *h_i and γ_i are such that $N^{-1}H\gamma' \to A_{H\gamma}$, $N^{-1}HH' \to A_{HH}$, and $N^{-1}H\theta' \to A_{H\theta}$, where $A_{H\gamma}$, A_{HH}, and $A_{H\theta}$ are some fixed matrices.*

In the next section, we show that the limiting (uncentered) cross moment $A_{H\gamma}$ between the slope parameter γ_i and the factor loading h_i plays an important role for both the rate of convergence and the asymptotic distribution of the break date estimate. When $A_{H\gamma} \neq 0$, the results are similar to those of the univariate case obtained by Perron and Zhu (2005). On the other hand, when $A_{H\gamma} = 0$, the break date estimate is consistent at a faster rate than the univariate case.

3 Asymptotic Results

First we provide consistency of the break date $\hat{T}_1$ or break fraction $\hat{\lambda}_1$. The notation $(T, N) \to \infty$ means that both T and N jointly go to infinity, and we do not assume that they follow a particular path or grow sequentially. The following theorem pertains to the case of stationary error processes.

Theorem 1 *(Stationary Errors) Suppose that Assumptions 1, 2, 4 and 5 hold. Then, we have the following results.*

A. Model I (Joint Broken Trend)

(i) As $N \to \infty$ *with* T *fixed,* $|\hat{T}_1 - T_1| = o_p(1)$*, if* $h_i = 0$ *and* $d_i(L) = 1$ *for all* i*.*

(ii) As $(T, N) \to \infty$*,* $|\hat{T}_1 - T_1| = O_p(T^{-1/2}N^{-1/2})$*, if* $h_i = 0$ *for all* i*.*

(iii) As $(T, N) \to \infty$*,* $|\hat{T}_1 - T_1| = O_p(T^{-1/2})$*, if* $A_{H\gamma} \neq 0$*.*

B. Model II (Local Disjoint Trend)

(iv) As $N \to \infty$ *with* T *fixed,* $|\hat{T}_1 - T_1| = o_p(1)$*, if* $h_i = 0$ *and* $d_i(L) = 1$ *for all* i*.*

(v) As $(T, N) \to \infty$*,* $|\hat{T}_1 - T_1| = o_p(1)$*, if* $h_i = 0$ *for all* i*.*

(vi) As $(T, N) \to \infty$*,* $|\hat{T}_1 - T_1| = O_p(1)$*, if* $A_{H\gamma} \neq 0$*.*

As shown in (i) and (iv), the consistency of the break date estimate is established in both models even with T fixed when there is neither common component nor serial correlation in u_{ti}. When there are common components or serial correlation, both N and T should grow in order to have the consistency of the break date estimate. In the next section, we provide a Monte Carlo simulation result where the distribution of the break date estimate concentrates on a wrong break date as N grows when there is serial correlation in each equation without any common component.

For the case with serial correlation but without common component ($h_i = 0$ for all i), see (ii) and (v). We show the consistency of the break date estimate for both models. This result has more significance in Model II, because the break date estimate in Model II is not consistent in the univariate model and now a large panel makes it consistent. On the other hand, the break date estimate in Model I is anyway consistent even in the univariate model. The rate of convergence is $\sqrt{TN}$ in Model I, but it does depend on the relative magnitude of N and T in Model II and we simply denote it as $o_p(1)$.

Now, see the results (iii) and (vi). When there are common factors with loadings that are asymptotically non-negligible in the sense that $A_{H\gamma} \neq 0$, the convergence rate of the break date estimate reduces and is the same as that of the univariate model in Perron and Zhu (2005). Especially, the rate of convergence does not depend on the number of equation N and there is no efficiency gain from panel data. This is a natural consequence of strong cross sectional dependence generated by the common components.

The next theorem states the results when u_{ti} is integrated.

Theorem 2 *(Integrated Errors) Suppose that Assumptions 1, 3, 4 and 5 hold. Then we have the following results.*

A. Model I (Joint Broken Trend)

(i) As $(T, N) \to \infty$, $|\hat{\lambda}_1 - \lambda_1| = O_p(T^{-1/2}N^{-1/2})$, if $h_i = 0$ for all i.

(ii) As $(T, N) \to \infty$, $|\hat{\lambda}_1 - \lambda_1| = O_p(T^{-1/2})$, if $A_{H\gamma} \neq 0$.

B. Model II (Local Disjoint Trend)

(iii) As $(T, N) \to \infty$, $|\hat{\lambda}_1 - \lambda_1| = o_p(T^{-1/2})$, if $h_i = 0$ for all i.

(iv) As $(T, N) \to \infty$, $|\hat{\lambda}_1 - \lambda_1| = O_p(T^{-1/2})$, if $A_{H\gamma} \neq 0$.

When the error terms are integrated, the break date estimate $\hat{T}_1$ in both models is not consistent whether there are common components or not. This is the reason why we present the results in terms of the break fraction estimate $\hat{\lambda}_1 = \hat{T}_1/T$ which is consistent at least at rate $T^{1/2}$. When there is no common component, the rate of convergence of $\hat{\lambda}_1$ is $\sqrt{TN}$ in Model I but not available in Model II. In both models, but $\hat{\lambda}_1$ is consistent at rate $T^{1/2}$ in the presence of the common components and this rate is the same as in the univariate case.

The next theorem states the limiting distribution of the break fraction estimate.

Theorem 3 *(Stationary Errors) Suppose that Assumptions 1, 2, 4 and 5 hold. As $(T, N) \to \infty$, we have the following results:*

A. Model I (Joint Broken Trend)

(i) If $N/T^3 \to 0$ and $h_i = 0$ for all i,

$$T^{3/2}N^{1/2}(\hat{\lambda} - \lambda_1) \xrightarrow{d} N\left(0, \frac{4}{(1-\lambda_1)\lambda_1 A_{\gamma\gamma}^2} S_{\gamma\gamma}\right)$$

(ii) If $A_{H\gamma} \neq 0$,

$$T^{3/2}(\hat{\lambda} - \lambda_1) \xrightarrow{d} N\left(0, \frac{4}{(1-\lambda_1)\lambda_1 A_{\gamma\gamma}^2} A'_{H\gamma}C(1)C(1)'A_{H\gamma}\right)$$

B. Model II (Local Disjoint Trend)

(iii) If $A_{H\theta} \neq 0$ and $A_{H\gamma} \neq 0$,

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

where the stochastic process $S^*(m)$ is such that $S^*(0) = 0$, $S^*(m) = S_1(m)$ for $m < 0$ and $S^*(m) = S_2(m)$ for $m > 0$, with

$$\begin{aligned} S_1(m) &= \sum_{k=m+1}^{0} \left(A_{\theta\theta} + A_{\gamma\gamma}k^2 + 2A_{\gamma\theta}k\right) - 2\sum_{k=m+1}^{0} F_t' \left(A_{H\theta} + kA_{H\gamma}\right), \; m = -1, -2, \ldots \\ S_2(m) &= \sum_{k=1}^{m} \left(A_{\theta\theta} + A_{\gamma\gamma}k^2 + 2A_{\gamma\theta}k\right) + 2\sum_{k=1}^{m} F_t' \left(A_{H\theta} + kA_{H\gamma}\right), \; m = 1, 2, \ldots \end{aligned}$$

In Model I, the distribution of the break fraction estimate is approximated by a normal distribution irrespective of the presence of the common components. In both cases, the variance of the limiting distribution depends on the location of the true break fraction λ_1 and the break parameters γ_i's. The closer λ_1 is to the mid point of the time span and the larger the slope changes are, the smaller the asymptotic variance is. It also depends on the long-run variance of the individual specific shocks when there is no common components (recall from Assumption 4 that $A_{\gamma\gamma} = \lim N^{-1}\gamma\gamma'$, and $S_{\gamma\gamma} = \lim N^{-1}\gamma D\Sigma_\varepsilon D\gamma' = \lim N^{-1}\sum d_i^2(1)\sigma_i^2\gamma_i^2$). With common components, it depends on the long-run variance of the common components but not on the individual specific variance parameters. Also, note that the result in (i) is derived with an additional assumption that $N/T^3 \to 0$. This additional condition is not restrictive in most time series panels where N and T are of similar size or T is much larger.

In Model II, the limiting distribution of the break fraction estimate is derived in the presence of common components using the strict stationarity of F_t given in Assumption 2. It is highly non-standard but analogous to the univariate case. It depends on the exact distribution of F_t and is of little use in practice unless an assumption on the distribution of F_t is made.

In order to form a confidence interval from the above results, various parameters other than the break fraction should be estimated. The slope parameter γ can be consistently estimated via the least squares estimate conditional on the estimated break date. When there is no common factor, the long-run variances of the individual specific errors can be consistently estimated from the estimated residuals $\hat{u}_{ti} = y_{ti} - \hat{d}_{ti}$ by applying a standard long-run variance estimator to each equation. When there are common components, the long-run variance of the common factors needs to be estimated, but the factors or the loadings don't have to be estimated. The cross-sectional average of the estimated residuals multiplied by the slope parameter estimates is, for a large sample,

$$\frac{1}{N}\sum_{i=1}^{N} \hat{u}_{ti}\hat{\gamma}_i \approx \frac{1}{N}\sum_{i=1}^{N} (F_t'h_i + +e_{ti})\gamma_i \approx F_t'A_{H\gamma} + O_p(N^{-1/2}) \tag{4}$$

Hence, $A'_{H\gamma}C(1)C(1)'A_{H\gamma}$ can be consistently estimated using any standard long-run variance estimator applied to this cross-sectional average of the estimated residuals.

When there is no common component ($h_i = 0$ for all i) in Model II, thus the limiting distribution of the break fraction estmate depends on the relative magnitude of N and T and is of a complicated form. We take an alternative approach that yields a non-degenerate limiting distribution of the break fraction estimate. That is to assume the break parameters are of an order of $N^{-1/2}$. To be more precise, we make the following assumption.

Assumption 6 *(i)* $\gamma = N^{-12}\dot{\gamma}$, $\theta = N^{-1/2}\dot{\theta}$, $\theta\theta' \to \dot{A}_{\theta\theta}$, $\theta\gamma' \to \dot{A}_{\theta\gamma}$, $\gamma\gamma' \to \dot{A}_{\gamma\gamma} \neq 0$, *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* $D = diag\{d_1(1), \ldots, d_N(1)\}$.

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

(iii) $N^{-1}\sum_{i=1}^{N}\sigma_i^2\dot{\theta}_i^2 \to \bar{\sigma}_\theta^2$ *and* $N^{-1}\sum_{i=1}^{N}\sigma_i^2\dot{\gamma}_i^2 \to \bar{\sigma}_\gamma^2$.

The limiting distribution of the break date estimate under this alternative assumption is given in the next theorem.

Theorem 4 *(Stationary Errors) Suppose that Assumptions 1, 2, 5 and 6 hold and that* $h_i = 0$ *and* $d_i(L) = 1$ *for all* i. *Let* $W = (W_1, \ldots, W_m)'$ *with* $W \sim N(0, I_m)$. *Then, in Model II, we have as* $(T,\, N) \to \infty$,

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

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*

$$\begin{aligned} V_1(m) &= \sum_{k=m+1}^{0}\left(\dot{A}_{\theta\theta} + \dot{A}_{\gamma\gamma}k^2 + 2\dot{A}_{\gamma\theta}k\right) - 2\sum_{k=m+1}^{0}\left(\bar{\sigma}_\theta + k\bar{\sigma}_\gamma\right)W_k,\ m = -1, -2, \ldots \\ V_2(m) &= \sum_{k=1}^{m}\left(\dot{A}_{\theta\theta} + \dot{A}_{\gamma\gamma}k^2 + 2\dot{A}_{\gamma\theta}k\right) + 2\sum_{k=1}^{m}\left(\bar{\sigma}_\theta + k\bar{\sigma}_\gamma\right)W_k,\ m = 1, 2, \ldots \end{aligned}$$

Note that the above limiting distribution is obtained under an additional assumption that u_{ti} is iid over t ($d_i(L) = 1$). If this assumption does not hold, the limiting distribution of the break fraction estimate is of a more complex form depending on the autocovariance functions of the individual specific errors. The form of the process $V^*(m)$ resembles the process $S^*(m)$ in the previous theorem. However, an important difference is that $V^*(m)$ is defined in terms of a standard multivariate normal variate and it does not depend on the exact distribution of the error process.

The next theorem states the limiting distribution of the break fraction estimate when the error terms are integrated.

Theorem 5 *(Integrated Errors) Suppose that Assumptions 1, 3, 4 and 5 hold. As $(T,\, N) \to \infty$, we have the following results:*

A. Model I (Joint Broken Trend)

(i) If $N/T \to 0 \leq \kappa < \infty$ and $h_i = 0$ for all i,

$$\sqrt{TN}(\hat{\lambda} - \lambda_1) \xrightarrow{d} N\left(\sqrt{\kappa}\frac{\bar{\sigma}_d^2}{A_{\gamma\gamma}}\frac{1-2\lambda_1}{5(1-\lambda_1)\lambda_1}, \frac{2}{15A_{\gamma\gamma}^2}S_{\gamma\gamma}\right)$$

(ii) If $A_{H\gamma} \neq 0$,

$$\sqrt{T}(\hat{\lambda} - \lambda_1) \xrightarrow{d} N\left(0, \frac{2}{15A_{\gamma\gamma}^2}A_{H\gamma}'C(1)C(1)'A_{H\gamma}\right)$$

B. Model II (Local Disjoint Trend)

(iii) Let $\xi_{11} = \left[\int_0^1 W_F(r)dr, \int_0^1 rW_F(r)dr, \int_{\lambda_1}^1 W_F(r)dr, \int_{\lambda_1}^1 (r-\lambda_1)W_F(r)dr\right]'$, $\xi_{21} = \left[0, 0, W_F(\lambda_1), \int_{\lambda_1}^1 W_F(r)dr\right]'$, $\xi_3 = \int_0^{\lambda_1}\frac{(3r-2r\lambda_1)}{\lambda_1^2}dW_F(r)'$, and $\xi_4 = \int_{\lambda_1}^1 \frac{(r-1)(3r-2\lambda_1-1)}{(1-\lambda_1)^2}dW_F(r)'$, where $W_F(r)$ be an r-dimensional standard Wiener process. Then,

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

where the stochastic process $Z^*(m)$ is defined such that $Z^*(0) = 0$, $Z^*(m) = Z(m) + m^2\xi_4 C(1)'A_{H\gamma}$ for $m < 0$ and $Z^*(m) = Z(m) + m^2\xi_3 C(1)'A_{H\gamma}$ for $m > 0$, and

$$Z(m) = A_{\gamma\gamma}\frac{|m|^3}{3} + m\left(tr\begin{bmatrix} 2\Omega_1\left(\xi_{11}C(1)'A_{HH}C(1)\xi_{21}'\right) \\ -\Omega_1\Sigma_f\Omega_1\left(\xi_{11}C(1)'A_{HH}C(1)\xi_{11}'\right)\end{bmatrix} + \bar{\sigma}_d^2\frac{2\left(2\lambda_1^2 - 46\lambda_1 + 45\right)}{15\lambda_1}\right)$$

with $\bar{\sigma}_d^2 = \lim_N N^{-1}\sum_{i=1}^N d_i(1)^2\sigma_i^2$,

$$\Sigma_f = \begin{pmatrix} 0 & 0 & 1 & 1-\lambda_1 \\ & 0 & \lambda_1 & (1-\lambda_1^2)/2 \\ & & 1 & 1-\lambda_1 \\ & & & (1-\lambda_1)^2 \end{pmatrix} \text{ and } \Omega_1 = \begin{pmatrix} \frac{4}{\lambda_1} & -\frac{6}{\lambda_1^2} & \frac{2}{\lambda_1} & \frac{6}{\lambda_1^2} \\ & \frac{12}{\lambda_1^3} & -\frac{6}{\lambda_1^2} & -\frac{12}{\lambda_1^3} \\ & & \frac{4}{\lambda_1(1-\lambda_1)} & 6\frac{1-2\lambda_1}{\lambda_1^2(1-\lambda_1)^2} \\ & & & 12\frac{3\lambda_1^2-3\lambda_1+1}{\lambda_1^3(1-\lambda_1)^3} \end{pmatrix}$$

The results are similar to the stationary error case. However, some of the important differences are worth mentioning. The rate of convergence is slower than the stationary error case as already pointed out in Theorem 2. The variance of the limiting distribution does not depend on the location of the true break fraction λ_1. When there is no common component in Model I, the limiting distribution is obtained under the assumption that $N/T \to 0 \leq \kappa < \infty$, which implies that T should not be too small relative to N, and it has non-zero mean that is $O(\sqrt{\kappa})$. This bias is unlikely to be negligible unless T is extremely large relative to N. For example, when T is five times larger than N, $\sqrt{\kappa} \simeq 0.45$. Also, note that this bias does not exist when $\lambda_1 = 0.5$. See Section 4 below for some Monte Carlo simulation results.

In Model II, the limiting distribution is highly non-standard again, but the process $Z^*(m)$ is defined in terms of a standard Wiener process. Hence, no distributional assumption is required on F_t to simulate $Z^*(m)$. When there is no common component,

$$Z^*(m) = A_{\gamma\gamma}\frac{|m|^3}{3} + m\bar{\sigma}_d^2 \frac{2\left(2\lambda_1^2 - 46\lambda_1 + 45\right)}{15\lambda_1}$$

and thus $\sqrt{T}(\hat{\lambda} - \lambda_1) \xrightarrow{p} 0$, as claimed in Theorem 2(iii). A non-degenerate limiting distribution is not available in this case.

In order to use these limiting distributions, a number of parameters should be estimated. The limiting variance and bias of the limiting distribution in Model I can be estimated from the first difference of the estimated residuals using a method discussed in the stationary case. For the limiting distribution for Model II, the process $Z^*(m)$ should be simulated and it requires not only the parameters associated with the common components but also those associated with the individual specific errors. For that matter, the common components and the individual specific errors need to be estimated separately. See Bai and Ng (2002, 2004) and Bai (2003) for the estimation of the common factors. It is well known that the factors and the factor loadings can be consistently estimated only up to a non-singular rotation matrix. That is, the principal component estimators are consistent only for $h_i'R^{-1}$ and RF_t for some non-singular matrix R. However, this doesn't pose any difficulty in simulating the limiting distribution of Model II. As long as γ, $h_i'R^{-1}$ and RF_t are identified, so are $R'^{-1}A_{HH}R^{-1}$, $R'^{-1}A_{H\gamma}$ and $RC(1)$, which are enough to identify $C(1)'A_{HH}C(1)$ and $C(1)'A_{H\gamma}$. Also see Bai and Carrion-i-Silvestre (2009) for the estimation of the longrun variances of the individual specific errors.

4 Monte Carlo Simulations

In this section, we demonstrate the appropriateness of the asymptotic results presented in the previous section via Monte Carlo experiments. In all experiments, the number of replications is 2,000. 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 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. Hence,

$$d_{ti} = \begin{cases} \gamma_i B_t & \text{Model I} \\ \theta_i C_t + \gamma_i B_t & \text{Model II} \end{cases}$$

Figure 1 illustrates the consistency of the break date estimate in the absence of the common component (i.e., $h_i = 0$, for all i) when T is fixed. T is 10 and the true break date T_1 is 3. The number of cross sectional units N is set at 1, 100, 300, and 500. The first two rows correspond to Model I and the other two to Model II. γ_i is drawn from $U(0,1)$ for both models and θ_i is drawn from $U(0.2, 0.7)$ for Model II. In the first and third rows of panels, the errors are *iid* $N(0,1)$. The bar graph in each panel depicts the relative frequencies of the break date estimates. Note that in both models the relative frequency of the true break date approaches one as N increases. On the other hand, in the second and fourth rows, autocorrelated errors are used with the autoregressive parameter 0.8. That is, $e_{ti} = 0.8e_{ti} + \varepsilon_{ti}$ with $\varepsilon_{ti} \sim iid\ N(0,1)$. As N increases, the probability mass function concentrates at a wrong break date; it concentrates on the mid-point in Model I and on the fourth and fifth dates with almost equal probabilities in Model II.

Figures 2 and 3 show the estimated probability density of the break date estimate around the true break date ($\hat{T}_1 - T_1$) as both T and N increase. γ_i is drawn from $U(0, 0.3)$ for both models and θ_i is drawn from $U(0.1, 0.4)$ for Model II. Each of the four columns corresponds to T of 100, 200, 300, and 500 respectively. The true break date T_1 is always $0.3T$. There are four lines displayed in each panel and each of them corresponds to N of 1, 10, 50, and 100 respectively. The panels in the first row show the case with *iid* errors where $u_{ti} = \varepsilon_{ti} \sim iid\ N(0,1)$. The panels in the second row show the case with autocorrelated errors where $u_{ti} = e_{ti}$, $e_{ti} = \rho_i e_{t-1i} + \varepsilon_{ti}$, $\varepsilon_{ti} \sim iid\ N(0,1)$ and ρ_i is drawn from $U(0.4, 0.7)$. The panels in the third row show the case with a common factor, where e_{ti} is the same as the one in the second rows, $F_t = 0.6F_{t-1} + w_t$ with $w_t \sim iid\ N(0,1)$ and h_i is 0.5 for $N = 1$ and drawn from $U(0,1)$ for the other values of N.

The first observation from Figure 2 is that the displayed densities are all bell shaped as expected from the asymptotic normality of the estimated break date given in Theorem 3. Also, the densities concentrate around the origin as T and N increase when there is no common component. When there is a common factor, for any given value of N, the empirical densities look more concentrated around the origin as T increases. However, for any given T, the empirical densities are almost on top of each other for all values of N except for 1. This is due to the fact that the rate of convergence of the break date estimate doesn't depend on N in the presence of common components.

The picture is somewhat different for Model II, though some important features are shared with Model I. In Figure 3, the densities show bimodality especially for large N. This observation is related to the non-standard asymptotic distribution of the break date estimate given in Theorem 3. In the absence of a common component, the empirical densities concentrate around the origin as T and N increase. When there is a common component, the densities do not concentrate as T, N or both increase, since the break date estimate is inconsistent in Model II.

Figures 4 and 5 correspond to the case with integrated errors. Each panel displays the empirical probability densities of $T^{-1/2}(\hat{T}_1 - T_1)$. The results in the first row are obtained using integrated individual specific errors and no common components. Thus, $u_{ti} = e_{ti}$, $e_{ti} = e_{t-1i} + \varepsilon_{ti}$, $\varepsilon_{ti} \sim iid$ $N(0,1)$. In the second row, the common factor is integrated so that $F_t = F_{t-1} + w_t$ with $w_t \sim iid$ $N(0,1)$ and h_i is 0.5 for $N = 1$ and drawn from $U(0,1)$ for the other values of N, while $e_{ti} = \varepsilon_{ti}$ and $\varepsilon_{ti} \sim iid\ N(0,1)$. In the third row, both the common factor and the individual specific errors are integrated. The slope change parameter γ_i and the intercept change parameter θ_i are the same as in Figures 2 and 3.

In Figure 4, the densities are again bell shaped. In the first row, a distinct feature is that the densities not only concentrate around the mode but also shift to the right as N increases for any given T. This shift is the bias term provided in Theorem 5(i). In the second row, where an integrated common factor exists but the individual specific shocks are *iid* over T, the densities are centered at the origin though some asymmetry is observed especially for small T. Furthermore, because of the existence of the common component, the densities do not concentrate as N increases. In the third row, where both the common factor and individual specific shocks are integrated, the results are intermediate of the two previous cases.

The densities displayed in Figure 5 are somewhat similar to those in Figure 4. The densities concentrate around the mode as N increases for any given T without common components. When there is a common component, the densities do not concentrate. Also, the bimodality appearing in the stationary case is not present.

5 Application

As an empirical illustration, we estimate a common slope change in disaggregate level price indices. The data is the log of the seasonally adjusted quarterly price indices for Gross Domestic Products of the United States. It is obtained from the Bureau of Economic Analysis website[1]. The data has 34 individual series with each of them spanning from 1984:Q1 to 2009.Q2 (hence 102 observations in each series). Price indices are often analyzed in first difference form because they would then become inflation rates. Empirical studies on the statistical properties of the inflation series are rich, but the results are somewhat conflicting depending on the time span and the employed statistical

[1] Table 1.5.4. Price Indexes for Gross Domestic Product, Expanded Detail.

methods. For example, Pivetta and Reis (2003) find inflation rate very persistent which implies that price index series are integrated twice (I(2)) or being close to be so. On the other hand, Clark (2006) finds that the average persistence in disaggregate inflation rates is below aggregate persistence and that even further reduction in persistence is observed when a mean shift is taken into account. Clark (2006) used one common break date 1993.Q1 which he obtained by estimating a break date from each series and simply averaging them. We take Clark's (2006) view and assume that price indices are integrated processes (I(1)) around a broken deterministic trend. In other words, we assume that price indices are generated by Model I with integrated common components and error processes. The main results are presented in Table 1. For the estimation of the longrun variance of the common factors, we applied the Quadratic Spectral window and selected the bandwidth parameter using Andrews's (1991) data dependent method with AR(1) approximation.

Table 1. Estimated Common Break Date in Disaggregate Price Indices

	$\hat{T}_1$	95% C. I.
Common break date in the disaggregate GDP price indices	94.Q2	[92.Q1, 96.Q3]
Break date in the aggregate GDP price index	91.Q3	[89.Q2, 93.Q4]

The estimated date of the common slope change in the 34 disaggregate level price indices is 1994.Q2 with the 95% confidence interval ranging from 1992.Q1 to 1996.Q3. On the other hand, the estimated date of the slope change in the univariate aggregate GDP price index is 1991.Q3 with the 95% confidence interval ranging from 1989.Q2 to 1993.Q4. Although the two confidence intervals overlap somewhat, the two point estimates differ by 11 quarters, which is not ignorable. The two confidence intervals have exactly the same length of 4 years and 2 quarters. This is related to the fact that the rate of convergence of the common break date in the existence of the common components is the same as the univariate case.

In Table 2, we report the break date from each of the 34 series. In (i) Goods, (ii) Services and (iv) Fixed Investments, all but one break date estimates are in the 1990s. On the other hand, we observe a handful of break dates estimates in the 2000s especially in (iii) Net Exports and (v) Government Consumption Expenditures and Gross Investment. However, these are often associated with a very wide confidence interval. The mean and median of these break dates are 1995.Q1 and 1992.Q4 respectively, and the mean of the lengths of the confidence intervals is 7 years and 2 quarters.

Table 2.Estimated Individual Break Dates

(i) Goods

		$\hat{T}_1$	95% C. I.
Durable	Motor Vehicles and parts	97.Q1	[95.Q3, 98.Q3]
	Furnishings and durable household equip.	96.Q2	[94.Q4, 97.Q4]
	Recreational goods and vehicles	93.Q3	[92.Q2, 94.Q4]
	Other durable goods	92.Q3	[90.Q4, 94.Q2]
	Food/beverages purch'd for off-premises cons.	90.Q4	[86.Q1, 95.Q3]
Nondurable	Clothing and footwear	92.Q2	[91.Q1, 93.Q3]
	Gasoline and other energy goods	99.Q2	[92.Q2, 06.Q2]
	Other nondurable goods	92.Q2	[91.Q4, 93.Q3]

(ii) Services

		$\hat{T}_1$	95% C. I.
Household	Housing and utilities	89.Q1	[86.Q4, 91.Q2]
	Health care	93.Q1	[92.Q2, 93.Q4]
	Transportation serv.	91.Q4	[88.Q1, 95.Q3]
	Recreation serv.	91.Q4	[90.Q1, 93.Q3]
	Food serv. and accommd.	90.Q2	[88.Q3, 92.Q1]
	Financial serv. and insur.	92.Q4	[88.Q3, 97.Q1]
	Other serv.	93.Q2	[89.Q4, 96.Q4]
Nonprofit institutions	Gross output of nonprft inst.	91.Q2	[89.Q3, 93.Q1]

(iii) Net Exports

		$\hat{T}_1$	95% C. I.
Exports	Goods	03.Q2	[99.Q1, 07.Q3]
	Services	91.Q4	[85.Q3, 98.Q1]
Imports	Goods	03.Q3	[98.Q4, 08.Q2]
	Services	03.Q3	[94.Q2, 12.Q4]

(iv) Fixed Investments

		$\hat{T}_1$	95% C. I.
Nonresidential	Structures	01.Q3	[00.Q2, 02.Q4]
	Computers/peri'l equip.	91.Q1	[86.Q1, 96.Q1]
	Software	97..Q1	[94.Q1, 00.Q1]
	Other	93.Q1	[91.Q3, 94.Q3]
	Industrial equip.	91.Q4	[89.Q4, 93.Q4]
	Transportation equip.	92.Q2	[88.Q1, 96.Q3]
	Other equip.	92.Q1	[89.Q1, 95.Q1]
Residential		98.Q3	[89.Q2, 07.Q4]

(v) Government Consumption Expenditures and Gross Investment

		$\hat{T}_1$	95% C. I.
Federal - defense	Cons. expend.	01.Q1	[98.Q2, 04.Q2]
	Gross Investment	87.Q1	[85.Q4, 88.Q2]
Federal - nondefense	Cons. expend.	01.Q4	[80.Q2, 23.Q2]
	Gross Investment	91.Q1	[88.Q1, 94.Q1]
State and local	Cons. expend.	02.Q4	[99.Q1, 06.Q3]
	Gross Investment	03.Q4	[02.Q4, 04.Q4]

6 Conclusion

We showed how to estimate a common deterministic trend break in large panels through minimizing the sum of squared residuals. The statistical properties of the proposed break date estimate depends on a number of model assumptions. The form of the break, that is, whether it is only a slope change or a slope change combined with an intercept shift, affects the rate of convergence and the form of the limiting distribution of the break date estimate. The amount of correlation in both the time and cross sectional span affects the rate of convergence. Strong correlation in any of the time and cross sectional span results in a slower rate of convergence. Especially the strong cross sectional dependence generated by the common factor structure of the error processes makes the rate of convergence the same as the univariate case and thus eliminates the benefit of the panel data.

References

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

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

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

[4] Bai, J. (2008) Common Breaks in Means and Variances for Panel Data. forthcoming in Journal of Econometrics.

[5] Bai, J. and J.L. Carrion-i-Silvestre (2009) Structural Changes, Common Stochastic Trends, and Unit Roots in Panel Data. Review of Economic Studies 76, 471-501.

[6] Bai, J., R.L. Lumsdaine and J.H. Stock (1998) Testing for and Dating Breaks in Multivariate Time Series. Review of Economic Studies 65, 395-432.

[7] Bai, J. and P. Perron (1998) Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica 66, 47-78.

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

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

[10] Clark, T.E. (2006) Disaggregate Evidence on the Persistence of Consumer Price Inflation. Journal of Applied Econometrics 21, 563-587.

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

[12] Phillips and Moon (1999) Linear Regression Limit Theory for Nonstationary Panel Data. Econometrica 67, 1057-1111.

[13] Phillips and Solo (1992) Asymptotics for Linear Process, Annals of Statistics 20(2), 971-1001.

[14] Pivetta, F. and R. Reis (2007) The Persistence of Inflation in the United States. Journal of Economic Dynamics and Control 31(4), 1326-1358

<**Appendix**>

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. Also, $\widetilde{\iota}_b = (\widetilde{\iota}_b(1), \ldots \widetilde{\iota}_b(T))'$, $\iota_b = (\iota_b(1), \ldots \iota_b(T))'$, $\alpha = (\alpha_1, \ldots, \alpha_T)'$ and $\kappa = (\kappa_1, \ldots, \kappa_T)'$ where

$$\text{if } T_b > T_1, \quad \widetilde{\iota}_b(t) \equiv \begin{cases} 0 & \text{if } 1 \leq t \leq T_1, \\ (t - T_1)/(T_b - T_1) & \text{if } T_1 + 1 \leq t \leq T_b, \\ 1 & \text{if } T_b + 1 \leq t \leq T, \end{cases}$$

$$\text{if } T_b < T_1, \quad \widetilde{\iota}_b(t) \equiv \begin{cases} 0 & \text{if } 1 \leq t \leq T_b, \\ -(t - T_b)/(T_1 - T_b) & \text{if } T_b + 1 \leq t \leq T_1, \\ -1 & \text{if } T_1 + 1 \leq t \leq T, \end{cases}$$

$$\text{if } T_b = T_1, \quad \widetilde{\iota}_b(t) = \iota_b(t) \equiv \begin{cases} 0 & \text{if } 1 \leq t \leq T_1, \\ 1 & \text{if } T_1 + 1 \leq t \leq T. \end{cases}$$

$$\text{if } T_b \geq T_1, \quad \alpha_t = \begin{cases} 1 & \text{if } T_1 + 1 \leq t \leq T_b, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{if } T_1 > T_b, \quad \alpha_t = \begin{cases} -1 & \text{if } T_b + 1 \leq t \leq T_1, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$\text{if } T_b \geq T_1, \quad \kappa_t = \begin{cases} t - T_1 & \text{if } T_1 + 1 \leq t \leq T_b, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{if } T_1 > T_b, \quad \kappa_t = \begin{cases} -(t - T_1) & \text{if } T_b + 1 \leq t \leq T_1, \\ 0 & \text{otherwise.} \end{cases}$$

The following results pertaining to the deterministic terms are taken from Perron and Zhu (2005) and we will use them without proofs.

Lemma A.1 *(i)* $\Delta_T(X'_{T_b}X_{T_b} - X'_{T_1}X_{T_1})\Delta_T = |T_b - T_1|\, O(T^{-1})$ *(ii)* $\widetilde{\iota}'_b\Delta_T X_{T_b} = O(T^{1/2})$ *(iii)* $\alpha'\Delta_T X_{T_b} = |T_b - T_1|\, O(T^{-1/2})$ *(iv)* $\kappa'\Delta_T X_{T_b} = |T_b - T_1|^2\, O(T^{-1/2})$ *(v)* $\widetilde{\iota}'_b(I - P_{T_b})\widetilde{\iota}_b = O(T)$ *(vi)*

$\Delta_T X_{T_1}' X_{T_1} \Delta_T = \Sigma_a + o_p(1)$, *where the symmetric matrix* Σ_a *is given by*

$$\Sigma_a = \begin{bmatrix} 1 & \frac{1}{2} & \frac{(1-\lambda_1)^2}{2} \\ & \frac{1}{3} & \frac{(1-\lambda_1)^2(\lambda_1+2)}{6} \\ & & \frac{(1-\lambda_1)^3}{3} \end{bmatrix} \quad and \quad \begin{bmatrix} 1 & \frac{1}{2} & 1-\lambda_1 & \frac{(1-\lambda_1)^2}{2} \\ & \frac{1}{3} & \frac{1-\lambda_1^2}{2} & \frac{(1-\lambda_1)^2(\lambda_1+2)}{6} \\ & & 1-\lambda_1 & \frac{(1-\lambda_1)^2}{2} \\ & & & \frac{(1-\lambda_1)^3}{3} \end{bmatrix}$$

for Model I and II respectively.

We prove the following results for the stationary error processes.

Lemma A.2 *Under Assumption 2, we have for all* T_b*, as* $T, N \to \infty$*:*

(i) $\Delta_T X_{T_b}' E\gamma' = O_p(N^{1/2})$, $\widetilde{\iota}_b' E\gamma' = O_p(T^{1/2}N^{1/2})$

(ii) $\Delta_T X_{T_b}' EH' = O_p(N^{1/2})$, $\widetilde{\iota}_b' EH' = O_p(T^{1/2}N^{1/2})$

(iii) In Model I, $\widetilde{\iota}_b'(I - P_{T_b})E\gamma' = O_p(T^{1/2}N^{1/2})$, $\widetilde{\iota}_b'(I - P_{T_b})EH' = O_p(T^{1/2}N^{1/2})$

(iv) $\alpha' F = |T_b - T_1|^{1/2} O_p(1)$, $\alpha' E\theta' = |T_b - T_1|^{1/2} O_p(N^{1/2})$, $\alpha' EH' = |T_b - T_1|^{1/2} O_p(N^{1/2})$

(v) $\kappa' F = |T_b - T_1|^{3/2} O_p(1)$, $\kappa' E\gamma' = |T_b - T_1|^{3/2} O_p(N^{1/2})$, $\kappa' EH' = |T_b - T_1|^{3/2} O_p(N^{1/2})$

(vi) In Model II, $\alpha'(I - P_{T_b})E\theta' = |T_b - T_1|^{1/2} O_p(N^{1/2})$, $\alpha'(I - P_{T_b})EH' = |T_b - T_1|^{1/2} O_p(N^{1/2})$, $\kappa'(I - P_{T_b})E\gamma' = |T_b - T_1|^{3/2} O_p(N^{1/2})$, $\kappa'(I - P_{T_b})EH' = |T_b - T_1|^{3/2} O_p(N^{1/2})$

Proof: (i) From the definition of Δ_T and X_{T_b}, we need to consider $T^{-1/2}\iota' E\gamma'$, $T^{-1/2}C' E\gamma'$, $T^{-3/2}\tau' E\gamma'$, and $T^{-3/2}B' E\gamma'$. Let $r_i(h)$ be the autocovariance function of e_{ti}. First note that

$$\begin{aligned} Var\left(T^{-1/2}\iota' E\gamma'\right) &= \frac{1}{T} Var(\sum_{i=1}^{N} \iota' E_i \gamma_i) = \frac{1}{T} \sum_{i=1}^{N} \gamma_i^2 Var\left(\iota' E_i\right) \\ &= \sum_{i=1}^{N} \gamma_i^2 \sum_{h=-T+1}^{T-1} \left(1 - \frac{|h|}{T}\right) r_i(h) \\ &\leq \sum_{i=1}^{N} \gamma_i^2 \sum_{h=-\infty}^{\infty} |r_i(h)| = \sum_{i=1}^{N} \gamma_i^2 \sigma_i^2 \sum_{h=-\infty}^{\infty} \left| \sum_{k=0}^{\infty} d_{i,k+|h|} d_{i,k} \right| \\ &\leq \sum_{i=1}^{N} \gamma_i^2 \sigma_i^2 \sum_{k=0}^{\infty} |d_{i,k}| \sum_{h=-\infty}^{\infty} \left| d_{i,k+|h|} \right| \leq 2M^3 \sum_{i=1}^{N} \gamma_i^2 = O(N) \end{aligned}$$

Since the elements in C, $T^{-1}\tau$, $T^{-1}B$, and $\widetilde{\iota}_b$ are no greater than those in ι, the results related to these terms can be proved analogously.

(ii) Completely analogous to (i).

(iii) Note that $|\widetilde{\iota}_b'(I - P_{T_b})E\gamma'| \leq |\widetilde{\iota}_b' E\gamma'| + |\widetilde{\iota}_b' P_{T_b} E\gamma'|$, where $\widetilde{\iota}_b' E\gamma' = O(T^{1/2}N^{1/2})$ from (i), and

$$\widetilde{\iota}_b' P_{T_b} E\gamma' = \widetilde{\iota}_b' X_{T_b} \Delta_T (\Delta_T X_{T_b}' X_{T_b} \Delta_T)^{-1} \Delta_T X_{T_b}' E\gamma' = O(T^{1/2}) O(1) O_p(N^{1/2})$$

Hence, the result follows.

(iv)

$$
\begin{aligned}
Var\left(|T_b - T_1|^{-1/2}\alpha' E\gamma'\right) &= \frac{1}{|T_b - T_1|} Var(\sum_{i=1}^{N} \alpha' E_i \gamma_i) = \frac{1}{|T_b - T_1|} \sum_{i=1}^{N} \gamma_i^2 Var\left(\alpha' E_i\right) \\
&= \frac{1}{|T_b - T_1|} \sum_{i=1}^{N} \gamma_i^2 \sum_{h=-T+1}^{T-1} \left(\sum_{j=1}^{T-|h|} \alpha_j \alpha_{j+|h|} \right) r_i(h) \\
&\leq \sum_{i=1}^{N} \gamma_i^2 \sum_{h=-\infty}^{\infty} |r_i(h)| \leq 2M^3 \sum_{i=1}^{N} \gamma_i^2 = O(N)
\end{aligned}
$$

(v) Recall that all elements of $|T_b - T_1|^{-1}\kappa$ are bounded by one and only $|T_b - T_1|$ of them are non-zero. The rest of the proof is completely analogous to (iv).

(vi)

$$
\begin{aligned}
\alpha'(I - P_{T_b})E\theta' &= \alpha' E\theta' - \alpha' X_{T_b} \Delta_T (\Delta_T X'_{T_b} X_{T_b} \Delta_T)^{-1} \Delta_T X'_{T_b} E\theta' \\
&= |T_b - T_1|^{1/2} O_p(N^{1/2}) + |T_b - T_1| O(T^{-1/2}) O(1) O_p(N^{1/2})
\end{aligned}
$$

The second term is always of a smaller order of magnitude than the first term because $|T_b - T_1|^{1/2}/T^{1/2}$ is at most $O_p(1)$ and can even be $o_p(1)$ if $|T_b - T_1| = o_p(T)$. A similar argument holds for the other terms. ■

We prove the following results for the integrated error processes.

Lemma A.3 *Under Assumption 3, we have for all T_b, as $T, N \to \infty$:*

(i) $\Delta_T X'_{T_b} E\gamma' = O_p(TN^{1/2})$ and $\tilde{\iota}'_b E\gamma' = O_p(T^{3/2}N^{1/2})$

(ii) $\Delta_T X'_{T_b} EH' = O_p(TN^{1/2})$ and $\tilde{\iota}'_b EH' = O_p(T^{3/2}N^{1/2})$

(iii) In Model I, $\tilde{\iota}'_b(I - P_{T_b})E\gamma' = O_p(T^{3/2}N^{1/2})$ and $\tilde{\iota}'_b(I - P_{T_b})EH' = O_p(T^{3/2}N^{1/2})$

(iv) $\alpha' F = |T_b - T_1| O_p(T^{1/2})$, $\alpha' E\theta' = |T_b - T_1| O_p(T^{1/2}N^{1/2})$ and $\alpha' EH' = |T_b - T_1| O_p(T^{1/2}N^{1/2})$

(v) $\kappa' F = |T_b - T_1|^2 O_p(T^{1/2})$, $\kappa' E\gamma' = |T_b - T_1|^2 O_p(T^{1/2}N^{1/2})$ and $\kappa' EH' = |T_b - T_1|^2 O_p(T^{1/2}N^{1/2})$

(vi) In Model II, $\alpha'(I - P_{T_b})E\theta' = |T_b - T_1| O_p(T^{1/2}N^{1/2})$, $\alpha'(I - P_{T_b})EH' = |T_b - T_1| O_p(T^{1/2}N^{1/2})$, $\kappa'(I - P_{T_b})E\gamma' = |T_b - T_1|^2 O_p(T^{1/2}N^{1/2})$ and $\kappa'(I - P_{T_b})EH' = |T_b - T_1|^2 O_p(T^{1/2}N^{1/2})$

Proof: For (i), (ii) and (iii), we show for $T^{-1/2}\iota' E\gamma'$, the first element of $\Delta_T X'_{T_b} E\gamma'$ only. The rest of results can be shown analogously. Let $\xi_{ti} = (1 - L)e_{ti}$ and its autocovariance function $r_i(h)$. First note that $E_i = G\xi_i$ where $\xi_i = (\xi_{1i}, \ldots, \xi_{Ti})'$ and G is a lower triangular matrix of ones. Let

$a_T = (a_{1,T}, \ldots, a_{T,T})' = T^{-1}G'\iota$. Then, each element of a_T is bounded by 1.

$$
\begin{aligned}
Var\left(T^{-3/2}\iota' E\gamma'\right) &= \frac{1}{T^3} Var(\sum_{i=1}^{N} \iota' E_i \gamma_i) = \frac{1}{T^3} \sum_{i=1}^{N} \gamma_i^2 Var\left(\iota' E_i\right) \\
&= \frac{1}{T^3} \sum_{i=1}^{N} \gamma_i^2 Var\left(\iota' G\xi_i\right) = \frac{1}{T} \sum_{i=1}^{N} \gamma_i^2 Var\left(a_T' \xi_i\right) \\
&= \frac{1}{T} \sum_{i=1}^{N} \gamma_i^2 \sum_{h=-T+1}^{T-1} \left(\sum_{j=1}^{T-|h|} a_{j,T} a_{j+|h|,T} \right) r_i(h) \\
&\leq \sum_{i=1}^{N} \gamma_i^2 \sum_{h=-\infty}^{\infty} |r_i(h)| \leq 2M^3 \sum_{i=1}^{N} \gamma_i^2 = O(N)
\end{aligned}
$$

(iv) Let $b_T = (b_{1,T}, \ldots, b_{T,T})' = |T_b - T_1|^{-1} G'\alpha$. Then, b_T has at most the bigger of T_1 and T_b non-zero positive elements and all of them are bounded by one.

$$
\begin{aligned}
Var\left(|T_b - T_1|^{-1} T^{-1/2} \alpha' E\gamma'\right) &= \frac{1}{|T_b - T_1|^2 T} Var(\sum_{i=1}^{N} \alpha' G\xi_i \gamma_i) = \frac{1}{T} \sum_{i=1}^{N} \gamma_i^2 Var\left(b_T' \xi_i\right) \\
&= \frac{1}{T} \sum_{i=1}^{N} \gamma_i^2 \sum_{h=-T+1}^{T-1} \left(\sum_{j=1}^{T-|h|} b_{j,T} b_{j+|h|,T} \right) r_i(h) \\
&\leq \sum_{i=1}^{N} \gamma_i^2 \sum_{h=-\infty}^{\infty} |r_i(h)| \leq 2M^3 \sum_{i=1}^{N} \gamma_i^2 = O(N)
\end{aligned}
$$

(v) and (vi) follow from a similar argument. ■

As the limit counter part of $\widetilde{\iota}_b(t)$, we define a continuous function $f_{\widetilde{\iota}_b}(r)$ over $[0, 1]$:

$$
\begin{aligned}
\text{if } \lambda &> \lambda_1, \quad f_{\widetilde{\iota}_b}(r) = \begin{cases} 0 & \text{if } 0 \leq r \leq \lambda_1, \\ (r - \lambda_1)/(\lambda - \lambda_1) & \text{if } \lambda_1 \leq r < \lambda, \\ 1 & \text{if } \lambda \leq r \leq 1, \end{cases} \\
\text{if } \lambda_1 &> \lambda, \quad f_{\widetilde{\iota}_b}(r) = \begin{cases} 0 & \text{if } 0 \leq r \leq \lambda, \\ -(r - \lambda)/(\lambda_1 - \lambda) & \text{if } \lambda \leq r < \lambda_1, \\ -1 & \text{if } \lambda_1 \leq r \leq 1, \end{cases} \\
\text{if } \lambda &= \lambda_1, \quad f_{\widetilde{\iota}_b}(r) = f_{\iota_b}(r) = \begin{cases} 0 & \text{if } 0 \leq r < \lambda_1, \\ 1 & \text{if } \lambda_1 \leq r \leq 1. \end{cases}
\end{aligned}
$$

Also, let $f(r, \lambda_1) = (1, r, (r - \lambda_1)^+)'$, $f_2(r, \lambda_1) = (1, r, 1(r \geq \lambda_1), (r - \lambda_1)^+)'$, $g(r, \lambda, \lambda_1) = (0, 0, (\lambda - \lambda_1) f_{\widetilde{\iota}_b}(r))'$, and $f^*_{\widetilde{\iota}_b}(r)$ be the projection residual of the least squares regression of $f_{\widetilde{\iota}_b}(r)$ on $f(r, \lambda_1)$.

Lemma A.4 *As $T \to \infty$, we have the following results under Assumption 2:*

In Model I,

(i) $T^{-1/2}\tilde{\iota}_b'(I - P_{T_b})F \Rightarrow \xi_F'$ *where* $\xi_F \sim N(0, \lambda_1(1-\lambda_1)C(1)C(1)'/4)$

(ii) $\Delta_T X_{T_1}' F \Rightarrow \int f(r, \lambda_1) dW_F(r)'C(1)'$

(iii) $\Delta_T (X_{T_1} - X_{T_b})' F \Rightarrow \int_0^1 g(r, \lambda, \lambda_1) dW_F(r)'C(1)'$

In Model II,

(iv) $\Delta_T X_{T_1}' F \Rightarrow \int f_2(r, \lambda_1) dW_F(r)'C(1)'$

Lemma A.5 *As $T \to \infty$, we have the following results under Assumption 3:*

In Model I,

(i) $T^{-3/2}\tilde{\iota}_b'(I - P_{T_b})F \Rightarrow \int_0^1 f_{\tilde{\iota}_b}^*(r) W_F(r)' dr C(1)'$

(ii) $T^{-1}\Delta_T X_{T_1}' F \Rightarrow \int f(r, \lambda_1) W_F(r)' dr C(1)'$

(iii) $T^{-1}\Delta_T (X_{T_1} - X_{T_b})' F \Rightarrow \int_0^1 g(r, \lambda, \lambda_1) W_F(r)' dr C(1)'$

In Model II,

(iv) $T^{-1}\Delta_T X_{T_1}' F \Rightarrow \int f_2(r, \lambda_1) W_F(r)' dr C(1)'$

The proofs for Lemma A.4 and Lemma A.5 can be found in Perron and Zhu (2005).

Now note that we have by definition

$$SSR(\hat{T}_1) - SSR(T_1) \leq 0 \tag{A.1}$$

and, as in Perron and Zhu (2005), we can write:

$$\begin{aligned} SSR(\hat{T}_1) - SSR(T_1) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})(X_{T_1} - X_{T_b})\Pi\right] \\ &\quad +2\ tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})U\right] \\ &\quad +tr\left[U'(P_{T_1} - P_{T_b})U\right] \\ &\equiv (XX) + 2(XU) + (UU) \end{aligned} \tag{A.2}$$

This decomposition will be used repeatedly below.

A.1 Proof of Theorem 1

Theorem 1 follows from Lemma A.6 below. Consider Model I with $h_i = 0$ and $d_i(L) = 1$ for all i. Suppose that $|\hat{T}_1 - T_1|$ is not $o_p(1)$ as N increases with T fixed. Note that (XX) is strictly positive and its order of magnitude is in its strict sense as shown in the proof of Lemma A.6. Then (XX) is of strictly greater order of magnitude than the other two terms, and as N increases, (A.2) cannot be negative with probability one. This is contradictory to the fact that $\hat{T}_1$ minimizes SSR among all possible break dates including the true one, and it follows that $|\hat{T}_1 - T_1|$ is $o_p(1)$.

When both T and N increase and $h_i = 0$ for all i, suppose that $|\hat{T}_1 - T_1|$ is $B_T O_p(T^{-1/2}N^{-1/2})$ where B_T is any sequence of numbers increasing in T. Then (XX) is the dominating term and the same contradiction argument holds, and $|\hat{T}_1 - T_1| = o_p(T^{-1/2})$. Hence $|\hat{T}_1 - T_1|$ is at most $O_p(T^{-1/2}N^{-1/2})$.

When $A_{H\gamma} \neq 0$, the contradiction arguments holds whenever $|\hat{T}_1 - T_1|$ is $B_T O_p(T^{-1/2})$. Hence, $|\hat{T}_1 - T_1|$ is at most $O_p(T^{-1/2})$. The proof for Model II is completely analogous.

Lemma A.6 *Under Assumptions 1, 2, 4 and 5, we have for all generic T_b:*
(i) In Model I,

$$\begin{aligned}
(XX) &= |T_b - T_1|^2 O(TN) \\
(XU) &= \begin{cases} |T_b - T_1| O_p(T^{1/2}N^{1/2}) & \text{if } h_i = 0 \text{ for all } i \\ |T_b - T_1| O_p(T^{1/2}N) & \text{if } A_{H\gamma} \neq 0 \end{cases} \\
(UU) &= \begin{cases} |T_b - T_1| O_p(T^{-1}N^{1/2}) & \text{if } h_i = 0 \text{ and } d_i(L) = 1 \text{ for all } i \\ |T_b - T_1| O_p(T^{-1}N) \end{cases}
\end{aligned}$$

(ii) In Model II,

$$\begin{aligned}
(XX) &= |T_b - T_1|^3 O(N) + |T_b - T_1| O(N) \\
(XU) &= \begin{cases} |T_b - T_1|^{3/2} O_p(N^{1/2}) + |T_b - T_1|^{1/2} O_p(N^{1/2}) & \text{if } h_i = 0 \text{ for all } i \\ |T_b - T_1|^{3/2} O_p(N) + |T_b - T_1|^{1/2} O_p(N) & \text{if } A_{H\gamma} \neq 0 \end{cases} \\
(UU) &= \begin{cases} |T_b - T_1|^{1/2} O_p(T^{-1/2}N^{1/2}) & \text{if } h_i = 0 \text{ and } d_i(L) = 1 \text{ for all } i \\ |T_b - T_1|^{1/2} O_p(T^{-1/2}N) \end{cases}
\end{aligned}$$

Proof: (i) For Model I,

$$\begin{aligned}
(XX) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})(X_{T_1} - X_{T_b})\Pi\right] \\
&= |T_b - T_1|^2 \ tr\left[\gamma'\widetilde{\iota}_b'(I - P_{T_b})\widetilde{\iota}_b\gamma\right] \\
&= |T_b - T_1|^2 \ \widetilde{\iota}_b'(I - P_{T_b})\widetilde{\iota}_b \ (\gamma\gamma') \\
&= |T_b - T_1|^2 O(NT)
\end{aligned}$$

where the last equality holds from Assumption 4 and Perron and Zhu's (2005) Lemma 1. Also note that $\lim_T T^{-1}\widetilde{\iota}_b'(I - P_{T_b})\widetilde{\iota}_b > 0$ as shown in Perron and Zhu and thus (XX) is $|T_b - T_1|^2 O(NT)$ in its strict sense, meaning that it's not $o(|T_b - T_1|^2 TN)$.

From Lemmas A.2 and A.4, and Assumption 5, it follows that

$$\begin{aligned}
(XU) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})U\right] \\
&= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})FH\right] + tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})E\right] \\
&= |T_b - T_1| \widetilde{\iota}_b'(I - P_{T_b})FH\gamma' + |T_b - T_1| \widetilde{\iota}_b'(I - P_{T_b})E\gamma' \\
&= |T_b - T_1| O_p(T^{1/2}N) + |T_b - T_1| O_p(T^{1/2}N^{1/2}).
\end{aligned} \tag{A.3}$$

Hence, $(XU) = |T_b - T_1| O_p(T^{1/2}N^{1/2})$ if $h_i = 0$, $\forall i$, and $|T_b - T_1| O_p(T^{1/2}N)$, if $A_{H\gamma} \neq 0$.

Now, consider a decomposition of (UU):

$$
\begin{aligned}
& tr\left[U'(P_{T_1}-P_{T_b})U\right] && \text{(A.4)}\\
= \;& tr\left[U'(X_{T_1}-X_{T_b})\Delta_T\left(\Delta_T X_{T_1}'X_{T_1}\Delta_T\right)^{-1}\Delta_T X_{T_1}'U\right] \\
& +tr\left[U'X_{T_b}\Delta_T\left(\Delta_T X_{T_b}'X_{T_b}\Delta_T\right)^{-1}\Delta_T\left(X_{T_b}'X_{T_b}-X_{T_1}'X_{T_1}\right)\Delta_T\left(\Delta_T X_{T_1}'X_{T_1}\Delta_T\right)^{-1}\Delta_T X_{T_1}'U\right] \\
& +tr\left[U'X_{T_b}\Delta_T\left(\Delta_T X_{T_b}'X_{T_b}\Delta_T\right)^{-1}\Delta_T(X_{T_1}-X_{T_b})'U\right] \\
= \;& R_1+R_2+R_3
\end{aligned}
$$

Write

$$
\begin{aligned}
R_1 &= tr\left[U'(X_{T_1}-X_{T_b})\Delta_T\left(\Delta_T X_{T_1}'X_{T_1}\Delta_T\right)^{-1}\Delta_T X_{T_1}'U\right] \\
&= tr\left[\left(\Delta_T X_{T_1}'X_{T_1}\Delta_T\right)^{-1}\Delta_T X_{T_1}'UU'(X_{T_1}-X_{T_b})\Delta_T\right] \\
&= vec\left(\left(\Delta_T X_{T_1}'X_{T_1}\Delta_T\right)^{-1}\right)' vec\left(\Delta_T X_{T_1}'UU'(X_{T_1}-X_{T_b})\Delta_T\right)
\end{aligned}
$$

where $vec((\Delta_T X_{T_1}'X_{T_1}\Delta_T)^{-1})=O_p(1)$ from Lemma A.1. Now,

$$
\begin{aligned}
& vec\left(\Delta_T X_{T_1}'UU'(X_{T_1}-X_{T_b})\Delta_T\right) \\
= \;& vec\left(\Delta_T X_{T_1}'FHH'F'(X_{T_1}-X_{T_b})\Delta_T\right)+vec\left(\Delta_T X_{T_1}'FHE'(X_{T_1}-X_{T_b})\Delta_T\right) \\
& +vec\left(\Delta_T X_{T_1}'EH'F'(X_{T_1}-X_{T_b})\Delta_T\right)+vec\left(\Delta_T X_{T_1}'EE'(X_{T_1}-X_{T_b})\Delta_T\right) \\
= \;& r_{11}+r_{12}+r_{13}+r_{14}.
\end{aligned}
$$

If $h_i=0$ for all i, r_{11}, r_{12}, and r_{13} are zeros. If not, $HH'=O(N)$ from Assumption 5. Also, from Lemmas A.2 and A.4, $\Delta_T X_{T_1}'F=O_p(1)$ and $\Delta_T X_{T_1}'EH'=O_p(N^{1/2})$. The first two columns of $\Delta_T(X_{T_1}-X_{T_b})$ are zeros and we only need to consider the third column, which is $|T_b-T_1|\widetilde{\iota}_b$. From Lemmas A.2 and A.4, $\Delta_T(X_{T_1}-X_{T_b})'F=|T_b-T_1|\,O_p(T^{-1})$ and $\Delta_T(X_{T_1}-X_{T_b})'EH=|T_b-T_1|\,O_p(T^{-1}N^{1/2})$. Hence, $r_{11}=|T_b-T_1|\,O_p(T^{-1}N)$, $r_{12}=|T_b-T_1|\,O_p(T^{-1}N^{1/2})$, and $r_{13}=|T_b-T_1|\,O_p(N^{1/2})$. For the term r_{14},

$$
\Delta_T X_{T_1}'EE'(X_{T_1}-X_{T_b})\Delta_T=\sum_{i=1}^{N}\Delta_T X_{T_1}'E_iE_i'(X_{T_1}-X_{T_b})\Delta_T
$$

where $\Delta_T X_{T_1}'E_iE_i'(X_{T_1}-X_{T_b}\Delta_T$ has a non-zero mean and is independent across i. Hence $r_{14}=|T_b-T_1|\,O_p(T^{-1}N)$. By collecting these terms, $R_1=|T_b-T_1|\,O_p(T^{-1}N)$. A similar argument shows that , $R_2=|T_b-T_1|\,O_p(T^{-1}N)$ and $R_3=|T_b-T_1|\,O_p(T^{-1}N)$. Therefore, it follows that $(UU)=|T_b-T_1|\,O_p(T^{-1}N)$. This order of magnitude is an upper bound for the general case. However, suppose that $h_i=0$ and $d_i(1)=1$ for all i. Then,

$$
\begin{aligned}
tr\left[U'(P_{T_1}-P_{T_b})U\right] &= tr\left[(P_{T_1}-P_{T_b})\sum_{i=1}^{N}E_iE_i'\right] \\
&= \sum_{i=1}^{N}tr\left[(P_{T_1}-P_{T_b})E_iE_i'\right]=\sum_{i=1}^{N}E_i'(P_{T_1}-P_{T_b})E_i
\end{aligned}
$$

and $\mathcal{E}tr\left[(P_{T_1} - P_{T_b})E_iE_i'\right] = 0$. Hence, $(UU) = |T_b - T_1|\, O_p(T^{-1}N^{1/2})$.

(ii) Consider Model II. The difference is that there is an extra regressor.

$$\begin{aligned}
(XX) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})(X_{T_1} - X_{T_b})\Pi\right] \\
&= tr\left[(\theta'\alpha' + \gamma'\kappa')(I - P_{T_b})(\alpha\theta + \kappa\gamma)\right] \\
&= \kappa'(I - P_{T_b})\kappa\ (\gamma\gamma') + 2\kappa'(I - P_{T_b})\alpha\ (\theta\gamma') + \alpha'(I - P_{T_b})\alpha\ (\theta\theta')
\end{aligned}$$

where $\kappa'(I-P_{T_b})\kappa = |T_b - T_1|^3\, O(1)$, $\kappa'(I-P_{T_b})\alpha = |T_b - T_1|^2\, O(1)$, and $\alpha'(I-P_{T_b})\alpha = |T_b - T_1|\, O(1)$. Depending on the order of magnitude of $|T_b - T_1|$, the dominant term is either $\kappa'(I - P_{T_b})\kappa$ or $\alpha'(I - P_{T_b})\alpha$. Hence we can write

$$(XX) = |T_b - T_1|^3\, O(N) + |T_b - T_1|\, O(N)$$

Also, note that $\lim |T_b - T_1|^{-3}\, \kappa'(I - P_{T_b})\kappa > 0$ and $\lim |T_b - T_1|^{-1}\, \alpha'(I - P_{T_b})\alpha > 0$ as $|T_b - T_1| \to \infty$, and thus (XX) is $|T_b - T_1|^3\, O(N)$ or $|T_b - T_1|\, O(N)$ in its strict sense.

For (XU), write

$$\begin{aligned}
(XU) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})U\right] \\
&= tr\left[(\theta'\alpha' + \gamma'\kappa')(I - P_{T_b})U\right]
\end{aligned}$$

It follows from Lemma A.2 that

$$\begin{aligned}
tr\left[\theta'\alpha'(I - P_{T_b})U\right] &= tr\left[\theta'\alpha'(I - P_{T_b})FH\right] + tr\left[\theta'\alpha'(I - P_{T_b})E\right] \\
&= |T_b - T_1|^{1/2}\, O_p(N) + |T_b - T_1|^{1/2}\, O_p(N^{1/2})
\end{aligned}$$

and

$$\begin{aligned}
tr\left[\gamma'\kappa'(I - P_{T_b})U\right] &= tr\left[\gamma'\kappa'(I - P_{T_b})FH\right] + tr\left[\gamma'\kappa'(I - P_{T_b})E\right] \\
&= |T_b - T_1|^{3/2}\, O_p(N) + |T_b - T_1|^{3/2}\, O_p(N^{1/2})
\end{aligned}$$

Hence $(XU) = |T_b - T_1|^{3/2}\, O_p(N^{1/2}) + |T_b - T_1|^{1/2}\, O_p(N^{1/2})$, if $h_i = 0$ for all i, $|T_b - T_1|^{3/2}\, O_p(N) + |T_b - T_1|^{1/2}\, O_p(N)$, otherwise.

For (UU), consider the decomposition in (A.4). Now, $\Delta_T(X_{T_1} - X_{T_b}) = [0, 0, \alpha, |T_b - T_1|\widetilde{\iota}_b]$, and it follows that $\Delta_T(X_{T_1} - X_{T_b})'F = |T_b - T_1|^{1/2}\, O_p(T^{-1/2})$ and $\Delta_T(X_{T_1} - X_{T_b})'EH = |T_b - T_1|^{1/2}\, O_p(T^{-1/2}N^{1/2})$. Using an approach similar to Model I, we obtain that $(UU) = |T_b - T_1|^{1/2}\, O_p(T^{-1/2}N^{1/2})$ if $h_i = 0$ and $d_i(L) = 1$ for all i, $|T_b - T_1|^{1/2}\, O_p(T^{-1/2}N)$, otherwise. ■

A.2 Proof of Theorem 2

Theorem 2 follows from Lemma A.7 below using the same argument as in the proof of Theorem 1.

Lemma A.7 *Under Assumptions 1, 3, 4 and 5, we have for all all generic T_b:*
(i) In Model I,

$$\begin{aligned}(XX) &= |T_b - T_1|^2 O(TN) \\ (XU) &= \begin{cases} |T_b - T_1| O_p(T^{3/2}N^{1/2}) & \text{if } h_i = 0 \text{ for all } i \\ |T_b - T_1| O_p(T^{3/2}N) & \text{if } A_{H\gamma} \neq 0 \end{cases} \\ (UU) &= |T_b - T_1| O_p(TN)\end{aligned}$$

(ii) In Model II,

$$\begin{aligned}(XX) &= |T_b - T_1|^3 O(N) \\ (XU) &= \begin{cases} |T_b - T_1|^2 O_p(T^{1/2}N^{1/2}) & \text{if } h_i = 0 \text{ for all } i \\ |T_b - T_1|^2 O_p(T^{1/2}N) & \text{if } A_{H\gamma} \neq 0 \end{cases} \\ (UU) &= |T_b - T_1| O_p(TN)\end{aligned}$$

Proof: (i) (XX) is the same as in Lemma A.6. The rest of the results follow from Lemma A.3 using an argument similar to the proof of Lemma A.6. ■

A.3 Proof of Theorem 3

(i) Consider Model I with $h_i = 0$ for all i. Let $m_T = N^{1/2}T^{1/2}(T_b - T_1)$ and $D(C) = \{T_b : |T_b - T_1| < CN^{-1/2}T^{-1/2}\}$ for a positive number C. On the set $D(C)$, $(XX) = O_p(1), (XU) = O_p(1)$ and $(UU) = O_p(T^{-3/2}N^{1/2})$. Here the term (UU) is asymptotically negligible if $N/T^3 \to 0$, and

$$\begin{aligned}(XX) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})(X_{T_1} - X_{T_b})\Pi\right] \\ &= (T_b - T_1)^2 \widetilde{\iota}_b'(I - P_{T_b})\widetilde{\iota}_b\, \gamma\gamma' \\ &= m_T^2 \left(\frac{1}{T}\widetilde{\iota}_b'(I - P_{T_b})\widetilde{\iota}_b\right)\left(\frac{1}{N}\gamma\gamma'\right)\end{aligned}$$

and

$$\begin{aligned}(XU) &= (T_b - T_1)\widetilde{\iota}_b'(I - P_{T_b})E\gamma' \\ &= m_T \left(\frac{1}{\sqrt{TN}}\widetilde{\iota}_b'(I - P_{T_b})E\gamma'\right) + o_p(1)\end{aligned}$$

Hence, on the set $D(C)$

$$\begin{aligned}m_T^* &= \arg\min_{m_T \text{ on } D(C)} [(XX) + 2(XU) + o_p(1)] \\ &= \arg\min_{m_T \text{ on } D(C)} \left[m_T^2 \left(\frac{1}{T}\widetilde{\iota}_b'(I - P_{T_b})\widetilde{\iota}_b\right)\left(\frac{1}{N}\gamma\gamma'\right) + 2m_T\left(\frac{1}{\sqrt{TN}}\widetilde{\iota}_b'(I - P_{T_b})E\gamma'\right) + o_p(1)\right] \\ &= -\left(\frac{1}{T}\widetilde{\iota}_b'(I - P_{T_b})\widetilde{\iota}_b\right)^{-1}\left(\frac{1}{N}\gamma\gamma'\right)^{-1}\left(\frac{1}{\sqrt{TN}}\widetilde{\iota}_b'(I - P_{T_b})E\gamma'\right)\end{aligned}$$

Note that $\lim_T T^{-1}\widetilde{\iota}_b'(I-P_{T_b})\widetilde{\iota}_b = (1-\lambda_1)\lambda_1/4$ and $\lim_N N^{-1}\gamma\gamma' = A_{\gamma\gamma}$. Recall that $e_{it} = d_i(1)\varepsilon_{it} + \widetilde{e}_{it-1} - \widetilde{e}_{it}$ with $\widetilde{e}_{it} = \sum_{k=0}^{\infty}\bar{d}_{ik}\varepsilon_{it-k}$ and $\bar{d}_{ik} = \sum_{i=k+1}^{\infty} d_i$. Let $\varepsilon_{it} = \sigma_i\eta_{it}$ with $\eta_{it} \sim iid(0,1)$, $\eta_i = (\eta_{i1},\ldots,\eta_{iT})'$ and $\Delta\widetilde{e}_i = (\widetilde{e}_{i0}-\widetilde{e}_{i1},\ldots,\widetilde{e}_{iT-1}-\widetilde{e}_{iT})'$ Then,

$$\begin{aligned}\frac{1}{\sqrt{TN}}\widetilde{\iota}_b'(I-P_{T_b})E\gamma' &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\widetilde{\iota}_b'(I-P_{T_b})E_i\gamma_i \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}C_iQ_{i,T} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}R_{i,T}\end{aligned}$$

where

$$\begin{aligned}C_i &= d_i(1)\sigma_i\gamma_i \\ Q_{i,T} &= \frac{1}{\sqrt{T}}\widetilde{\iota}_b'(I-P_{T_b})\eta_i \\ R_{i,T} &= \frac{1}{\sqrt{T}}\widetilde{\iota}_b'(I-P_{T_b})\Delta\widetilde{e}_i\gamma_i\end{aligned}$$

$Q_{i,T}$ is $iid(0, T^{-1}\widetilde{\iota}_b'(I-P_{T_b})\widetilde{\iota}_b)$ across i for all T. Since $Q_{i,T} = T^{-1/2}\widetilde{\iota}_b'(I-P_{T_b})\eta_i \Rightarrow Q_i \sim N(0,\lambda_1(1-\lambda_1)/4)$, $Q_{i,T}^2$ is convergent in distribution to Q_i^2 by the continuous mapping theorem. Also, $\mathcal{E}(Q_{i,T}^2) = T^{-1}\widetilde{\iota}_b'(I-P_{T_b})\widetilde{\iota}_b \to \mathcal{E}(Q_i^2) = (1-\lambda_1)\lambda_1/4$. This shows that $Q_{i,T}^2$ is uniformly integrable in T. Assumptions 2 and 4 imply that

$$\max_i \frac{C_i^2}{\sum_i C_i^2} = \frac{\max_i d_i^2(1)\sigma_i^2\gamma_i^2}{\gamma D\Sigma_\varepsilon D\gamma'} \le N^{-1}\frac{M^3\max_i\gamma_i^2}{N^{-1}\gamma D\Sigma_\varepsilon D\gamma'} = O(N^{-1})$$

Therefore, from the Joint Limit CLT (Theorem 3 in Phillips and Moon, 1999) we prove

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}C_iQ_{i,T} \xrightarrow{d} N\left(0, \frac{(1-\lambda_1)\lambda_1}{4}S_{\gamma\gamma}\right)$$

as $(N,T)\to\infty$. It remains to show that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}R_{i,T} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\widetilde{\iota}_b'\Delta\widetilde{e}_i\gamma_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\widetilde{\iota}_b'P_{T_b}\Delta\widetilde{e}_i\gamma_i = o_p(1) \tag{A.5}$$

First, $\widetilde{\iota}_b'\Delta\widetilde{e}_i/\sqrt{T} = |T_b-T_1|^{-1}(\widetilde{e}_{iT_1}+\cdots+\widetilde{e}_{iT_b-1})/\sqrt{T} - \widetilde{e}_{iT}/\sqrt{T}$, if $T_b > T_1$, $-|T_b-T_1|^{-1}(\widetilde{e}_{iT_b}+\cdots+\widetilde{e}_{iT_1-1})/\sqrt{T} + \widetilde{e}_{iT}/\sqrt{T}$, if $T_b < T_1$ and $(\widetilde{e}_{iT_1}-\widetilde{e}_{iT})/\sqrt{T}$ if $T_1 = T_b$. Since $|T_b - T_1| = o(1)$ on the set $D(C)$, we can write

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\widetilde{\iota}_b'\Delta\widetilde{e}_i\gamma_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\widetilde{e}_{iT_1}\gamma_i - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\widetilde{e}_{iT}\gamma_i + o_p(1).$$

From Assumptions 2 and 4

$$Var(\widetilde{e}_{iT_1}\gamma_i) = \sigma_i^2\gamma_i^2\sum_{k=0}^{\infty}\bar{d}_{ik}^2 \le \sigma_i^2\gamma_i^2\left(\sum_{k=0}^{\infty}\left|\bar{d}_{ik}\right|\right)^2 \le \max_i\left\{\gamma_1^2,\ldots,\gamma_N^2\right\}M^3$$

and thus

$$Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\widetilde{e}_{iT_1}\gamma_i\right)=\frac{1}{TN}\sum_{i=1}^{N}Var\left(\widetilde{e}_{iT_1}\gamma_i\right)\leq\frac{1}{T}\max_i\left\{\gamma_1^2,\ldots,\gamma_N^2\right\}M^3=o(1)$$

The same argument applies to $(NT)^{-1/2}\sum_{i=1}^{N}\widetilde{e}_{iT}\gamma_i$ and we show

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\widetilde{\iota}_b'\Delta\widetilde{e}_i\gamma_i=o_p(1).$$

For the second term in (A.5),

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\widetilde{\iota}_b'P_{T_b}\Delta\widetilde{e}_i=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\widetilde{\iota}_b'X_{T_b}\Delta_T\left(\Delta_T X_{Tb}'X_{Tb}\Delta_T\right)^{-1}\Delta_T X_{T_b}'\Delta\widetilde{e}_i=O(1)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta_T X_{T_b}'\Delta\widetilde{e}_i$$

and similarly as before

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Delta_T X_{T_b}'\Delta\widetilde{e}_i=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}\begin{pmatrix}\widetilde{e}_{i0}-\widetilde{e}_{iT}\\ T^{-1}(\widetilde{e}_{i0}+\cdots+\widetilde{e}_{iT})/-\widetilde{e}_{iT}\\ T^{-1}(\widetilde{e}_{iT_b+1}+\cdots+\widetilde{e}_{iT})/-\widetilde{e}_{iT}\end{pmatrix}=o_p(1).$$

Therefore, as $(N,T)\to\infty$

$$m_T^*\Rightarrow N\left(0,\frac{4}{(1-\lambda_1)\lambda_1 A_{\gamma\gamma}^2}S_{\gamma\gamma}\right)$$

Since $N^{1/2}T^{1/2}(\hat{T}_1-T_1)=O_p(1)$, we can choose C large enough so that the probability that $\hat{T}_1\in D(C)$ is arbitrarily close to one, and the statement in the Theorem follows.

For (ii), let $m_T=\sqrt{T}(T_b-T_1)$ and $D(C)=\{T_b:|T_b-T_1|<CT^{-1/2}\}$ for a positive number C. On the set $D(C)$, $(UU)/N$ is asymptotically negligible.

$$\frac{(XX)}{N}=m_T^2\left(\frac{1}{T}\widetilde{\iota}_b'(I-P_{T_b})\widetilde{\iota}_b\right)\left(\frac{1}{N}\gamma\gamma'\right)$$

$$\begin{aligned}\frac{(XU)}{N}&=tr\left[\Pi'(X_{T_1}-X_{T_b})'(I-P_{T_b})U\right]/N\\&=(T_b-T_1)\widetilde{\iota}_b'(I-P_{T_b})FH\gamma'/N+o_p(1)\\&=m_T\xi_F A_{H\gamma}+o_p(1)\end{aligned}$$

where $\xi_F=^d N(0,\lambda_1(1-\lambda_1)C(1)C(1)'/4)=O_p(1)$. Hence, using the same argument as before,

$$m_T^*\Rightarrow-\frac{4}{(1-\lambda_1)\lambda_1}A_{\gamma\gamma}^{-1}\xi_F A_{H\gamma}\sim N\left(0,\frac{4}{(1-\lambda_1)\lambda_1}A_{\gamma\gamma}^{-2}A_{H\gamma}'C(1)C(1)'A_{H\gamma}\right)$$

For (iii), consider Model II with $A_{H\gamma} \neq 0$. From Theorem 1, $|\hat{T}_1 - T_1| = O_p(1)$. Let $m_T = (T_b - T_1)$ and $D(C) = \{T_b : |T_b - T_1| < C\}$ for a positive number C. On the set $D(C)$,

$$
\begin{aligned}
(XX)/N &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})(X_{T_1} - X_{T_b})\Pi\right]/N \\
&= tr\left[\Pi'(X_{T_1} - X_{T_b})'(X_{T_1} - X_{T_b})\Pi\right]/N + o_p(1) \\
&= tr\left[(\alpha,\kappa)'(\alpha,\kappa)\begin{pmatrix}\theta\theta' & \theta\gamma' \\ \gamma\theta' & \gamma\gamma'\end{pmatrix}/N\right] + o_p(1) \\
&= tr\left[(\alpha,\kappa)'(\alpha,\kappa)\begin{pmatrix}A_{\theta\theta} & A_{\theta\gamma} \\ A_{\gamma\theta} & A_{\gamma\gamma}\end{pmatrix}\right] + o_p(1) \\
&= A_{\theta\theta}\alpha'\alpha + A_{\gamma\gamma}\kappa'\kappa + 2A_{\gamma\theta}\kappa'\alpha + o_p(1) \\
&= \begin{cases} \sum_{t=T_b+1}^{T_1}\left(A_{\theta\theta} + A_{\gamma\gamma}(t-T_1)^2 + 2A_{\gamma\theta}(t-T_1)\right) + o_p(1) & \text{if } T_b < T_1 \\ \sum_{t=T_1+1}^{T_b}\left(A_{\theta\theta} + A_{\gamma\gamma}(t-T_1)^2 + 2A_{\gamma\theta}(t-T_1)\right) + o_p(1) & \text{if } T_b > T_1 \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
(XU)/N &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})U\right]/N \\
&= tr\left[(\theta'\alpha' + \gamma'\kappa')(I - P_{T_b})U\right]/N \\
&= tr\left[(\theta'\alpha' + \gamma'\kappa')U\right]/N + o_p(1) \\
&= tr\left[(\alpha,\kappa)'FH(\theta',\gamma')\right]/N + o_p(1) \\
&= (\alpha,\kappa)'F(A_{H\theta}, A_{H\gamma}) + o_p(1) \\
&= \begin{cases} -\sum_{t=T_b+1}^{T_1}(F_t'A_{H\theta} + (t-T_1)F_t'A_{H\gamma}) + o_p(1) & \text{if } T_b < T_1 \\ \sum_{t=T_1+1}^{T_b}(F_t'A_{H\theta} + (t-T_1)F_t'A_{H\gamma}) + o_p(1) & \text{if } T_b > T_1 \end{cases}
\end{aligned}
$$

and $(UU)/N = o_p(1)$. Since F_t is strictly stationary under Assumption 2, define $S^*(m)$ such that $S^*(0) = 0$, $S^*(m) = S_1(m)$ for $m < 0$ and $S^*(m) = S_2(m)$ for $m > 0$, with

$$
\begin{aligned}
S_1(m) &= \sum_{k=m+1}^{0}\left(A_{\theta\theta} + A_{\gamma\gamma}k^2 + 2A_{\gamma\theta}k\right) - 2\sum_{k=m+1}^{0}\left(F_t'A_{H\theta} + kF_t'A_{H\gamma}\right), \; m = -1, -2, \ldots \\
S_2(m) &= \sum_{k=1}^{m}\left(A_{\theta\theta} + A_{\gamma\gamma}k^2 + 2A_{\gamma\theta}k\right) + 2\sum_{k=1}^{m}F_t'\left(A_{H\theta} + kA_{H\gamma}\right), \; m = 1, 2, \ldots
\end{aligned}
$$

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

$$
\begin{aligned}
m_T^* &= \arg\min_{m_T \text{ on } D(C)}\left[\{(XX) + 2(XU)\}/(T^2N) + o_p(1)\right] \\
&= \arg\min_{m_T \text{ on } D(C)}\left[S^*(m) + o_p(1)\right]
\end{aligned}
$$

The limiting distribution of m_T^* is arbitrarily close to that of $|\hat{T}_1 - T_1|$ for large C. ■

A.4 Proof of Theorem 4

Consider Model II under Assumption 6 with $h_i = 0$ and $d_i(L) = 1$ for all i. Lemma A.6 implies that $|\hat{T}_1 - T_1| = O_p(1)$. Let $m_T = (T_b - T_1)$ and $D(C) = \{T_b : |T_b - T_1| < C\}$ for a positive number C. On the set $D(C)$,

$$
\begin{aligned}
(XX) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})(X_{T_1} - X_{T_b})\Pi\right] \\
&= tr\left[\Pi'(X_{T_1} - X_{T_b})'(X_{T_1} - X_{T_b})\Pi\right] + o_p(1) \\
&= tr\left[(\alpha,\kappa)'(\alpha,\kappa)\begin{pmatrix} \theta\theta' & \theta\gamma' \\ \gamma\theta' & \gamma\gamma' \end{pmatrix}\right] + o_p(1) \\
&= tr\left[(\alpha,\kappa)'(\alpha,\kappa)\begin{pmatrix} \dot{A}_{\theta\theta} & \dot{A}_{\theta\gamma} \\ \dot{A}_{\gamma\theta} & \dot{A}_{\gamma\gamma} \end{pmatrix}\right] + o_p(1) \\
&= \begin{cases} \sum_{t=T_b+1}^{T_1}\left(\dot{A}_{\theta\theta} + \dot{A}_{\gamma\gamma}(t-T_1)^2 + 2\dot{A}_{\gamma\theta}(t-T_1)\right) + o_p(1) & \text{if } T_b < T_1 \\ \sum_{t=T_1+1}^{T_b}\left(\dot{A}_{\theta\theta} + \dot{A}_{\gamma\gamma}(t-T_1)^2 + 2\dot{A}_{\gamma\theta}(t-T_1)\right) + o_p(1) & \text{if } T_b > T_1 \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
(XU) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})U\right] \\
&= tr\left[(\theta'\alpha' + \gamma'\kappa')(I - P_{T_b})U\right] \\
&= tr\left[(\theta'\alpha' + \gamma'\kappa')U\right] + o_p(1) \\
&= \alpha' E\theta' + \kappa' E\gamma' + o_p(1) \\
&= \begin{cases} -\sum_{t=T_b+1}^{T_1}\left(\left(N^{-1/2}\sum_{i=1}^N e_{it}\dot{\theta}_i\right) + (t-T_1)\left(N^{-1/2}\sum_{i=1}^N e_{it}\dot{\gamma}_i\right)\right) + o_p(1) & \text{if } T_b < T_1 \\ \sum_{t=T_1+1}^{T_b}\left(\left(N^{-1/2}\sum_{i=1}^N e_{it}\dot{\theta}_i\right) + (t-T_1)\left(N^{-1/2}\sum_{i=1}^N e_{it}\dot{\gamma}_i\right)\right) + o_p(1) & \text{if } T_b > T_1 \end{cases}
\end{aligned}
$$

and $(UU) = o_p(1)$ if $N/T \to 0$. $(e_{it}\dot{\theta}_i)^2$ is independent across both i and t. Let $s_N^2 = \sum_{i=1}^N \mathcal{E}(e_{it}\dot{\theta}_i)^2 = \sum_{i=1}^N \sigma_i^2\dot{\theta}_i^2$, then the Lindeberg condition is satisfied because

$$
\begin{aligned}
\frac{1}{s_N^2}\sum_{i=1}^N \int_{\{(e_{it}\dot{\theta}_i)^2 > \varepsilon s_N^2\}} (e_{it}\dot{\theta}_i)^2 dP &\le \frac{1}{N^{-1}s_N^2}\max_i\left\{\int_{\{(e_{it}\dot{\theta}_i)^2 > \varepsilon s_N^2\}} (e_{it}\dot{\theta}_i)^2 dP\right\} \\
&\le \frac{M\max_i\left\{\dot{\theta}_i^2\right\}}{N^{-1}s_N^2}\left\{\int_{\left\{\eta_{it}^2 > \frac{\varepsilon s_N^2}{M\max_i\{\dot{\theta}_i^2\}}\right\}} \eta_{1t}^2 dP\right\} \to 0 \text{ as } N \to \infty
\end{aligned}
$$

where $\eta_{it} = e_{it}/\sigma_i \sim iid(0,1)$.

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

$$
N^{-1/2}\sum_{i=1}^N \begin{pmatrix} e_{iT_b-C} \\ \vdots \\ e_{iT_b+C} \end{pmatrix} \dot{\theta}_i \xrightarrow{d} N(0, \bar{\sigma}_\theta^2 I_{2C+1})
$$

where $\bar{\sigma}_\theta^2 = \lim \sum_{i=1}^N \sigma_i^2 \dot{\theta}_i^2$. Similarly,

$$N^{-1/2} \sum_{i=1}^N \begin{pmatrix} e_{iT_b - C} \\ \vdots \\ e_{iT_b + C} \end{pmatrix} \dot{\gamma}_i \xrightarrow{d} N\left(0, \bar{\sigma}_\gamma^2 I_{2C+1}\right)$$

where $\bar{\sigma}_\gamma^2 = \lim \sum_{i=1}^N \sigma_i^2 \dot{\gamma}_i^2$. Using this result,

$$(XU) \xrightarrow{d} \begin{cases} -\sum_{t=T_b+1}^{T_1} \left(\bar{\sigma}_\theta W(t-T_b) + (t-T_1)\bar{\sigma}_\gamma W(t-T_b)\right) & \text{if } T_b < T_1 \\ \sum_{t=T_1+1}^{T_b} \left(\bar{\sigma}_\theta W(t-T_1) + (t-T_1)\bar{\sigma}_\gamma W(t-T_1)\right) & \text{if } T_b > T_1 \end{cases}$$

where $W(s)$ iid $N(0,1)$. Finally, on the set $D(C)$, if $N/T \to 0$,

$$\begin{aligned} m_T^* &= \arg \min_{m_T \text{ on } D(C)} \left[\{(XX) + 2(XU)\} + o_p(1)\right] \\ &= \arg \min_{m_T \text{ on } D(C)} \left[V^*(m) + o_p(1)\right] \end{aligned}$$

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

$$\begin{aligned} V_1(m) &= \sum_{k=m+1}^{0} \left(\dot{A}_{\theta\theta} + \dot{A}_{\gamma\gamma} k^2 + 2\dot{A}_{\gamma\theta} k\right) - 2 \sum_{k=m+1}^{0} \left(\bar{\sigma}_\theta W(k) + k\bar{\sigma}_\gamma W(k)\right), \, m = -1, -2, \ldots \\ V_2(m) &= \sum_{k=1}^{m} \left(\dot{A}_{\theta\theta} + \dot{A}_{\gamma\gamma} k^2 + 2\dot{A}_{\gamma\theta} k\right) + 2 \sum_{k=1}^{m} \left(\bar{\sigma}_\theta W(k) + k\bar{\sigma}_\gamma W(k)\right), \, m = 1, 2, \ldots \end{aligned}$$

The limiting distribution of m_T^* is arbitrarily close to that of m_T^* for large C. ■

A.5 Proof of Theorem 5

Consider first (ii), Model I with $A_{H\gamma} \neq 0$. Let $m_T = (T_b - T_1)/\sqrt{T}$ and $D(C) = \{T_b : |T_b - T_1| < C\sqrt{T}\}$ for a positive number C. On the set $D(C)$, $(XX) = O_p(T^2 N), (XU) = O_p(T^2 N)$ and $(UU) = O_p(T^{3/2} N)$, and we can ignore the term (UU), since it is of strictly smaller order of magnitude. Then,

$$m_T^* = \arg \min_{m_T \text{ on } D(C)} \left[\{(XX) + 2(XU)\}/(T^2 N) + o_p(1)\right]$$

Also, note that

$$\begin{aligned} (XX)/(T^2 N) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})(X_{T_1} - X_{T_b})\Pi\right]/(T^2 N) \\ &= (T_b - T_1)^2 \tilde{\iota}_b'(I - P_{T_b})\tilde{\iota}_b \gamma\gamma'/(T^2 N) \\ &= m_T^2 \, T^{-1} \tilde{\iota}_b'(I - P_{T_b})\tilde{\iota}_b A_{\gamma\gamma} + o_p(1) \\ &= m_T^2 \left[\frac{(1-\lambda_1)\lambda_1}{4}\right] A_{\gamma\gamma} + o_p(1) \end{aligned}$$

and

$$\begin{aligned}(XU)/(T^2N) &= tr\left[\Pi'(X_{T_1}-X_{T_b})'(I-P_{T_b})U\right]/(T^2N)\\ &= (T_b-T_1)\widetilde{\iota}_b'(I-P_{T_b})FH\gamma'/(T^2N)+o_p(1)\\ &= m_T\int_{\lambda_1}^1 W_F^*(r)'drC(1)'A_{H\gamma}+o_p(1)\end{aligned}$$

where $W_F^*(r)$ is the residual function from a continuous time least-squares regression of $W_F(r)$ on $f(r,\lambda_1)$. Hence, on the set $D(C)$,

$$m_T^* = \arg\min_{m_T \text{ on } D(C)}\left[m_T^2\left[\frac{(1-\lambda_1)\lambda_1}{4}\right]A_{\gamma\gamma}+2m_T\int_{\lambda_1}^1 W_F^*(r)'drC(1)'A_{H\gamma}+o_p(1)\right]$$

and

$$\begin{aligned}m_T^* &\Rightarrow -\frac{4}{(1-\lambda_1)\lambda_1}A_{\gamma\gamma}^{-1}\int_{\lambda_1}^1 W_F^*(r)'drC(1)'A_{H\gamma}\\ &\sim N\left(0,\frac{2}{15A_{\gamma\gamma}^2}A_{H\gamma}'C(1)C(1)'A_{H\gamma}\right)\end{aligned}$$

using the fact that

$$\int_{\lambda_1}^1 W_F^*(r)'drC(1)'A_{H\gamma}\sim N\left(0,\frac{\lambda_1^2(1-\lambda_1)^2}{120}A_{H\gamma}'C(1)C(1)'A_{H\gamma}\right)$$

Since $\sqrt{T}(\hat{\lambda}-\lambda_1)$ is arbitrarily close to m_T^*, the assertion in the theorem follows.

For (i), let $m_T = N^{1/2}(T_b-T_1)/\sqrt{T}$ and $D(C)=\{T_b : |T_b-T_1|<CN^{-1/2}\sqrt{T}\}$ for a positive number C. On the set $D(C)$, $(XX)=O_p(T^2), (XU)=O_p(T^2)$ and $(UU)=O_p(T^{3/2}N^{1/2})$. Here the term (UU) is $O_p(T^2)$ if $N/T\to\kappa\geq 0$. Hence,

$$m_T^* = \arg\min_{m_T \text{ on } D(C)}\{(XX)+2(XU)+(UU)\}/T^2$$

The first term is such that

$$(XX)/T^2 = m_T^2\left[\frac{(1-\lambda_1)\lambda_1}{4}\right]A_{\gamma\gamma}+o_p(1)$$

The second term is

$$\begin{aligned}(XU)/T^2 &= (T_b-T_1)\widetilde{\iota}_b'(I-P_{T_b})E\gamma'/T^2\\ &= m_T\frac{1}{\sqrt{N}}\sum_{i=1}^N\frac{1}{T^{3/2}}\widetilde{\iota}_b'(I-P_{T_b})E_i\gamma_i\end{aligned}$$

From the Joint Limit CLT,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N\frac{1}{T^{3/2}}\widetilde{\iota}_b'(I-P_{T_b})E_i\gamma_i\xrightarrow{d}N\left(0,\frac{\lambda_1^2(1-\lambda_1)^2}{120}S_{\gamma\gamma}\right)\tag{A.6}$$

This result follows similarly to Theorem 3 (i). Recall that $(1-L)e_{it} = d_i(1)\varepsilon_{it} + \widetilde{e}_{it-1} - \widetilde{e}_{it}$ with $\widetilde{e}_{it} = \sum_{k=0}^{\infty} \bar{d}_{ik}\varepsilon_{it-k}$ and $\bar{d}_{ik} = \sum_{i=k+1}^{\infty} d_i$, and thus $e_{it} = d_i(1)\sum_{j=1}^{t}\varepsilon_{ij} + \widetilde{e}_{i0} - \widetilde{e}_{it}$. Let $\varepsilon_{it} = \sigma_i\eta_{it}$ with $\eta_{it} \sim iid(0,1)$, $\eta_i = (\eta_{i1},\ldots,\eta_{iT})'$, and G be a $T\times T$ lower triangular matrix of ones. Then,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{T^{3/2}}\widetilde{\iota}_b'(I-P_{T_b})E_i\gamma_i = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} C_iQ_{i,T} + o_p(1)$$

where

$$\begin{aligned} C_i &= d_i(1)\sigma_i\gamma_i \\ Q_{i,T} &= \frac{1}{T^{3/2}}\widetilde{\iota}_b'(I-P_{T_b})G\eta_i \end{aligned}$$

$Q_{i,T}$ is $iid(0, T^{-3}\widetilde{\iota}_b'(I-P_{T_b})GG'(I-P_{T_b})\widetilde{\iota}_b)$ across i for all T. Since $Q_{i,T} = T^{-3/2}\widetilde{\iota}_b'(I-P_{T_b})\eta_i \Rightarrow Q_i \sim N(0, \lambda_1^2(1-\lambda_1)^2/120)$, $Q_{i,T}^2$ is convergent in distribution to Q_i^2 by the continuous mapping theorem. Also, $\mathcal{E}(Q_{i,T}^2) = T^{-3}\widetilde{\iota}_b'(I-P_{T_b})GG'(I-P_{T_b})\widetilde{\iota}_b \to \mathcal{E}(Q_i^2) = \lambda_1^2(1-\lambda_1)^2/120$. This shows that $Q_{i,T}^2$ is uniformly integrable in T. Assumptions 3 and 4 imply that

$$\max_i \frac{C_i^2}{\sum_i C_i^2} = O(N^{-1})$$

This proves (A.6) from the Joint Limit CLT (Theorem 3 in Phillips and Moon, 1999).

For (UU), consider the decomposition in (A.4).

$$\begin{aligned} & tr\left[U'(P_{T_1}-P_{T_b})U\right] \\ = & \; tr\left[U'(X_{T_1}-X_{T_b})\Delta_T\left(\Delta_T X_{T_1}'X_{T_1}\Delta_T\right)^{-1}\Delta_T X_{T_1}'U\right] \\ & +tr\left[U'X_{T_b}\Delta_T\left(\Delta_T X_{T_b}'X_{T_b}\Delta_T\right)^{-1}\Delta_T\left(X_{T_b}'X_{T_b}-X_{T_1}'X_{T_1}\right)\Delta_T\left(\Delta_T X_{T_1}'X_{T_1}\Delta_T\right)^{-1}\Delta_T X_{T_1}'U\right] \\ & +tr\left[U'X_{T_b}\Delta_T\left(\Delta_T X_{T_b}'X_{T_b}\Delta_T\right)^{-1}\Delta_T(X_{T_1}-X_{T_b})'U\right] \\ = & \; R_1+R_2+R_3 \end{aligned}$$

Note that, as shown in PZ,

$$\Delta_T\left(X_{T_b}'X_{T_b}-X_{T_1}'X_{T_1}\right)\Delta_T = -(T_b-T_1)T^{-1}\Sigma_f + o(1)$$

with

$$\Sigma_f = \begin{bmatrix} 0 & 0 & 1-\lambda_1 \\ & 0 & (1-\lambda_1^2)/2 \\ & & (1-\lambda_1)^2 \end{bmatrix}$$

$$\Delta_T X_{T_1}'UU'(X_{T_1}-X_{T_b})\Delta_T = \Delta_T X_{T_1}'EE'(X_{T_1}-X_{T_b})\Delta_T$$

where

$$
\begin{aligned}
\frac{1}{(T_b-T_1)TN}\Delta_T X_{T_1}' E E'(X_{T_1}-X_{T_b})\Delta_T &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(T_b-T_1)T}\Delta_T X_{T_1}' E_i E_i'(X_{T_1}-X_{T_b})\Delta_T \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{d_i(1)^2\sigma_i^2}{(T_b-T_1)T}\Delta_T X_{T_1}' G\eta_i\eta_i' G'(X_{T_1}-X_{T_b})\Delta_T + o_p(1) \\
&= \frac{1}{N}\sum_{i=1}^{N} C_i Q_{i,T} + o_p(1) \qquad \text{(A.7)}
\end{aligned}
$$

where

$$
\begin{aligned}
C_i &= d_i(1)^2\sigma_i^2 \\
Q_{i,T} &= \frac{1}{(T_b-T_1)T}\Delta_T X_{T_1}' G\eta_i\eta_i' G'(X_{T_1}-X_{T_b})\Delta_T
\end{aligned}
$$

$Q_{i,T}$ is iid over i for all T, and

$$
Q_{i,T} \Rightarrow Q_i = \int_0^1 f(r,\lambda_1)W_i(r)dr\left(0,0,\int_{\lambda_1}^1 W_i(r)dr\right)
$$

as $T\to\infty$. If $\|Q_{i,T}\|$ is uniformly integrable, we can apply the Joint Probability Limits Theorem (Corollary 1, Phillips and Moon, 1999) to (A.7) and thus as $(T,N)\to\infty$

$$
\frac{1}{N}\sum_{i=1}^{N} C_i Q_{i,T} \xrightarrow{p} \bar{\sigma}_d^2 \mathcal{E} Q_i = \bar{\sigma}_d^2 \Sigma_e
$$

where $\bar{\sigma}_d^2 = \lim_N N^{-1}\sum_{i=1}^N d_i(1)^2\sigma_i^2$ and

$$
\mathcal{E}Q_i = \Sigma_e = \begin{pmatrix} 0 & 0 & \frac{2-3\lambda_1^2+\lambda_1^3}{6} \\ 0 & 0 & \frac{5+\lambda_1^4-6\lambda_1^2}{24} \\ 0 & 0 & \frac{-7\lambda_1^4+16\lambda_1^3-6\lambda_1^2-8\lambda_1+5}{24} \end{pmatrix}
$$

To show the uniform integrability of $\|Q_{i,T}\|$, define

$$
\begin{aligned}
L_{1,i,T} &= (T_b-T_1)^{-1}\Delta_T(X_{T_1}-X_{T_b})' G\eta_i \\
L_{2,i,T} &= T^{-1}\Delta_T X_{T_1}' G\eta_i
\end{aligned}
$$

then, $Q_{i,T} = L_{2,i,T}L_{1,i,T}'$. Note that $||L_{j,i,T}||^2$ $j=1,2$ are uniformly integrable in T because $||L_{j,i,T}||^2 \Rightarrow ||L_{ji}||^2$ and $\mathcal{E}||L_{j,i,T}||^2 \to \mathcal{E}||L_{ji}||^2$ as $T\to\infty$ where

$$
L_{1i} = \begin{bmatrix} 0 \\ 0 \\ \int_{\lambda_1}^1 W_i(r)'dr \end{bmatrix} \text{ and } L_{2i} = \int_0^1 f(r,\lambda_1)W_i(r)dr.
$$

This result also implies that $\sup_T \mathcal{E}\|Q_{j,i,T}\| < \infty$. Now, let the event $A(M) = \{\|Q_{i,T}\| > M\}$. Then,

$$\begin{aligned}
\sup_T \mathcal{E}\|Q_{i,T}\| 1_{A(M)} &\leq \sup_T \left(\mathcal{E}\|L_{1,i,T}\|^2 1_{A(M)}\right)^{1/2} \sup_T \left(\mathcal{E}\|L_{2,i,T}\|^2 1_{A(M)}\right)^{1/2} \\
&\to 0 \text{ as } M \to \infty
\end{aligned}$$

because, for $j = 1, 2$

$$\begin{aligned}
\sup_T \left(\mathcal{E}\|L_{j,i,T}\|^2 1_{A(M)}\right)^{1/2} &\leq \sup_T \left(\mathcal{E}\|L_{j,i,T}\|^2 1_{A(M)} 1_{B(\sqrt{M})} + \mathcal{E}\|L_{j,i,T}\|^2 1_{A(M)} 1_{B^c(\sqrt{M})}\right)^{1/2} \\
&\leq \left(\sup_T \mathcal{E}\|L_{j,i,T}\|^2 1_{A(M)} 1_{B(\sqrt{M})} + \sup_T \mathcal{E}\|L_{j,i,T}\|^2 1_{A(M)} 1_{B^c(\sqrt{M})}\right)^{1/2} \\
&\leq \left(\sup_T \mathcal{E}\|L_{j,i,T}\|^2 1_{B(\sqrt{M})} + \sup_T \sqrt{M} P(A(M))\right)^{1/2} \\
&\leq \left(\sup_T \mathcal{E}\|L_{j,i,T}\|^2 1_{B(\sqrt{M})} + \sqrt{M} \frac{\sup_T \mathcal{E}\|Q_{j,i,T}\|}{M}\right)^{1/2} \\
&\to 0 \text{ as } M \to \infty
\end{aligned}$$

where the event $B(\sqrt{M}) = \left\{\|L_{j,i,T}\|^2 > \sqrt{M}\right\}$.

Similarly as before,

$$\frac{1}{T^2 N} \Delta_T X'_{T_1} E E' X_{T_b} \Delta_T \xrightarrow{p} \bar{\sigma}_d^2 \mathcal{E} \int_0^1 f(r, \lambda_1) W_i(r) dr \int_0^1 f(r, \lambda_1)' W_i(r) dr = \bar{\sigma}_d^2 \Sigma_e$$

where

$$\begin{aligned}
\Sigma_d &= \mathcal{E} \int_0^1 f_2(r, \lambda_1) W_i(r) dr \int_0^1 f_2(r, \lambda_1)' W_i(r) dr \\
&= \begin{pmatrix} \frac{1}{3} & \frac{5}{24} & \frac{-\lambda_1^4 + 4\lambda_1^3 - 8\lambda_1 + 5}{24} \\ & \frac{2}{15} & \frac{-\lambda_1^5 + 10\lambda_1^3 - 25\lambda_1 + 16}{120} \\ & & \frac{7\lambda_1^5 - 20\lambda_1^4 + 10\lambda_1^3 + 20\lambda_1^2 - 25\lambda_1 + 8}{60} \end{pmatrix}
\end{aligned}$$

Collecting terms shows that, as $(T, N) \to \infty$,

$$\begin{aligned}
(UU)/T^2 &= m_T \sqrt{\kappa} \bar{\sigma}_d^2 tr\left[2\Sigma_a^{-1}\Sigma_e - \Sigma_a^{-1}\Sigma_f \Sigma_a^{-1}\Sigma_d\right] + o_p(1) \\
&= m_T \sqrt{\kappa} \bar{\sigma}_d^2 \frac{(2\lambda_1 - 1)}{10}
\end{aligned}$$

Therefore,

$$\begin{aligned}
m_T^* &= \arg \min_{m_T \text{ on } D(C)} \left\{(XX) + 2(XU) + (UU)\right\}/T^2 \\
&= \arg \min_{m_T \text{ on } D(C)} \left\{m_T^2 \left[\frac{(1-\lambda_1)\lambda_1}{4}\right] A_{\gamma\gamma} + 2m_T N \left(\sqrt{\kappa}\bar{\sigma}_d^2 \frac{(2\lambda_1 - 1)}{20}, \frac{\lambda_1^2(1-\lambda_1)^2}{120} S_{\gamma\gamma}\right) + o_p(1)\right\} \\
&\Rightarrow
\end{aligned}$$

$$
\begin{aligned}
m_T^* &\Rightarrow -\frac{4}{(1-\lambda_1)\lambda_1} A_{\gamma\gamma}^{-1}\; N\left(\sqrt{\kappa}\bar{\sigma}_d^2 \frac{(2\lambda_1 - 1)}{20}, \frac{\lambda_1^2(1-\lambda_1)^2}{120} S_{\gamma\gamma}\right) \\
&\sim N\left(\sqrt{\kappa}\frac{\bar{\sigma}_d^2}{A_{\gamma\gamma}} \frac{1-2\lambda_1}{5(1-\lambda_1)\lambda_1}, \frac{2}{15A_{\gamma\gamma}^2} S_{\gamma\gamma}\right)
\end{aligned}
$$

For (iii), consider Model II when $A_{H\gamma} \neq 0$. Let $m_T = (T_b - T_1)/\sqrt{T}$ and $D(C) = \{T_b : |T_b - T_1| < C\sqrt{T}\}$ for positive number C. On the set $D(C)$, $(XX) = O_p(T^{3/2}N), (XU) = O_p(T^{3/2}N)$ and $(UU) = O_p(T^{3/2}N)$. Hence no term is negligible.

$$
m_T^* = \arg \min_{m_T \text{ on } D(C)} \{(XX) + 2(XU) + (UU)\} / (T^{3/2}N)
$$

Note that with κ as defined in the proof of Lemma A.7

$$
\begin{aligned}
(XX)/(T^{3/2}N) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})(X_{T_1} - X_{T_b})\Pi\right]/(T^{3/2}N) \\
&= \kappa'(I - P_{T_b})\kappa\gamma\gamma'/(T^{3/2}N) + o_p(1) \\
&= (\kappa'\kappa/T^{3/2})(\gamma\gamma'/N) + o_p(1) \\
&= \frac{|T_b - T_1|^3}{3T^{3/2}}(\gamma\gamma'/N) + o_p(1) = \frac{|m_T|^3}{3}(\gamma\gamma'/N) + o_p(1)
\end{aligned}
$$

using the fact that $\kappa' X_{T_1}\Delta_T = |T_b - T_1|^2\, O_p(T^{-1/2})$. Under Assumption 5,

$$
\begin{aligned}
(XU)/(T^{3/2}N) &= tr\left[\Pi'(X_{T_1} - X_{T_b})'(I - P_{T_b})U\right]/(T^{3/2}N) \\
&= \kappa'(I - P_{T_b})U\gamma'/(T^{3/2}N) + o_p(1)
\end{aligned}
$$

where

$$
\begin{aligned}
\kappa' U\gamma'/(T^{3/2}N) &= T^{-3/2}\kappa' F(H\gamma'/N) + o_p(1) \\
&= \begin{cases} \frac{m_T^2}{2} W_F(\lambda_1)C(1)'(H\gamma'/N) + o_p(1) & \text{if } T_b \geq T_1, \\ -\frac{m_T^2}{2} W_F(\lambda_1)C(1)'(H\gamma'/N) + o_p(1) & \text{otherwise.} \end{cases}
\end{aligned}
$$

and

$$
T^{-1/2}\kappa' X_{T_1}\Delta_T = \begin{cases} \frac{m_T^2}{2}(1, \lambda_1, 0, 0), & \text{if } T_b \geq T_1, \\ \frac{m_T^2}{2}(1, \lambda_1, 1, 0), & \text{otherwise.} \end{cases}
$$

Combining these, we obtain

$$
\begin{aligned}
& (XU)/(T^{3/2}N) \\
= & \begin{cases} \frac{m_T^2}{2}\left[W_F(\lambda_1) - (1,\lambda_1,0,0)\Omega_1^{-1}\int_0^1 f_2(r,\lambda)W_F(r)'dr\right]C(1)'A_{H\gamma} + o_p(1) & \text{if } T_b \geq T_1, \\ \frac{m_T^2}{2}\left[-W_F(\lambda_1) - (1,\lambda_1,1,0)\Omega_1^{-1}\int_0^1 f_2(r,\lambda)W_F(r)'dr\right]C(1)'A_{H\gamma} + o_p(1) & \text{otherwise.} \end{cases} \\
= & \begin{cases} \frac{m_T^2}{2}\int_0^{\lambda_1}\frac{(3r-2r\lambda_1)}{\lambda_1^2}dW_F(r)'C(1)'A_{H\gamma} + o_p(1) & \text{if } T_b \geq T_1, \\ \frac{m_T^2}{2}\int_{\lambda_1}^1\frac{(r-1)(3r-2\lambda_1-1)}{(1-\lambda_1)^2}dW_F(r)'C(1)'A_{H\gamma} + o_p(1) & \text{otherwise.} \end{cases} \\
= & \begin{cases} \frac{m_T^2}{2}\xi_3 C(1)'A_{H\gamma} + o_p(1) & \text{if } T_b \geq T_1, \\ \frac{m_T^2}{2}\xi_4 C(1)'A_{H\gamma} + o_p(1) & \text{otherwise.} \end{cases}
\end{aligned}
$$

For (UU), consider the decomposition in (A.4) again. Note that

$$\Delta_T\left(X_{T_b}'X_{T_b} - X_{T_1}'X_{T_1}\right)\Delta_T = -(T_b - T_1)T^{-1}\Sigma_f$$

with

$$\Sigma_f = \begin{bmatrix} 0 & 0 & 1 & 1-\lambda_1 \\ & 0 & \lambda_1 & (1-\lambda_1^2)/2 \\ & & 1 & 1-\lambda_1 \\ & & & (1-\lambda_1)^2 \end{bmatrix}$$

From Lemma A.3,

$$
\begin{aligned}
& \Delta_T X_{T_1}'UU'(X_{T_1} - X_{T_b})\Delta_T \\
= & \Delta_T X_{T_1}'\left(FH + E\right)\left(FH + E\right)'(X_{T_1} - X_{T_b})\Delta_T \\
= & \Delta_T X_{T_1}'FHH'F'(X_{T_1} - X_{T_b})\Delta_T + \Delta_T X_{T_1}'EE'(X_{T_1} - X_{T_b})\Delta_T + o_p(T^{3/2}N)
\end{aligned}
$$

where

$$\frac{1}{T}\Delta_T X_{T_b}'FH \Rightarrow \int_0^1 f_2(r,\lambda)W_F(r)'drC(1)'H \equiv \xi_{11}C(1)'H$$

with $f_2(r,\lambda) = (1, r, 1(r \geq \lambda_1), (r-\lambda)^+)'$, and

$$\frac{1}{(T_b - T_1)}\Delta_T(X_{T_1} - X_{T_b})'FH \Rightarrow \begin{bmatrix} 0 \\ 0 \\ W_F(\lambda_1)' \\ \int_{\lambda_1}^1 W_F(r)'dr \end{bmatrix} C(1)'H \equiv \xi_{21}C(1)'H$$

Also,

$$
\begin{aligned}
\frac{1}{(T_b - T_1)TN}\Delta_T X_{T_1}' EE'(X_{T_1} - X_{T_b})\Delta_T &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(T_b-T_1)T}\Delta_T X_{T_1}' E_i E_i'(X_{T_1} - X_{T_b})\Delta_T \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{d_i(1)^2\sigma_i^2}{(T_b-T_1)T}\Delta_T X_{T_1}' G\eta_i\eta_i' G'(X_{T_1} - X_{T_b})\Delta_T + o_p(1) \\
&= \frac{1}{N}\sum_{i=1}^{N} C_i Q_{i,T} + o_p(1) \qquad \text{(A.8)}
\end{aligned}
$$

where

$$
\begin{aligned}
C_i &= d_i(1)^2\sigma_i^2 \\
Q_{i,T} &= \frac{1}{(T_b-T_1)T}\Delta_T X_{T_1}' G\eta_i\eta_i' G'(X_{T_1} - X_{T_b})\Delta_T
\end{aligned}
$$

$Q_{i,T}$ is iid over i for all T, and

$$
Q_{i,T} \Rightarrow Q_i = \int_0^1 f_2(r,\lambda_1)W_i(r)dr\left(0, 0, W_i(\lambda_1), \int_{\lambda_1}^1 W_i(r)dr\right)
$$

as $T \to \infty$. $\|Q_{i,T}\|$ is uniformly integrable, we can apply the Joint Probability Limits Theorem (Corollary 1, Phillips and Moon, 1999) to (A.8) and thus as $(T, N) \to \infty$

$$
\frac{1}{N}\sum_{i=1}^{N} C_i Q_{i,T} \xrightarrow{p} \bar{\sigma}_d^2 \mathcal{E} Q_i = \bar{\sigma}_d^2 \Sigma_e
$$

where $\bar{\sigma}_d^2 = \lim_N N^{-1}\sum_{i=1}^{N} d_i(1)^2\sigma_i^2$ and

$$
\mathcal{E}Q_i = \Sigma_e = \begin{pmatrix}
0 & 0 & \frac{2\lambda_1-\lambda_1^2}{2} & \frac{2-3\lambda_1^2+\lambda_1^3}{6} \\
0 & 0 & \frac{-\lambda_1^3+3\lambda_1^2}{6} & \frac{5+\lambda_1^4-6\lambda_1^2}{24} \\
0 & 0 & \lambda_1(1-\lambda_1) & \frac{1-3\lambda_1^2+2\lambda_1^3}{3} \\
0 & 0 & \frac{\lambda_1-2\lambda_1^2+\lambda_1^3}{2} & \frac{-7\lambda_1^4+16\lambda_1^3-6\lambda_1^2-8\lambda_1+5}{24}
\end{pmatrix}
$$

Also,

$$
\frac{1}{(T_b-T_1)TN}\Delta_T X_{T_1}' EE' X_{T_b}\Delta_T \xrightarrow{p} \bar{\sigma}_d^2 \mathcal{E}\int_0^1 f_2(r,\lambda_1)W_i(r)dr\int_0^1 f_2(r,\lambda_1)' W_i(r)dr = \bar{\sigma}_d^2\Sigma_e
$$

where

$$
\begin{aligned}
\Sigma_d &= \mathcal{E}\int_0^1 f_2(r,\lambda_1)W_i(r)dr\int_0^1 f_2(r,\lambda_1)' W_i(r)dr \\
&= \begin{pmatrix}
\frac{1}{3} & \frac{5}{24} & \frac{2-3\lambda_1^2+\lambda_1^3}{6} & \frac{-\lambda_1^4+4\lambda_1^3-8\lambda_1+5}{24} \\
 & \frac{2}{15} & \frac{5+\lambda_1^4-6\lambda_1^2}{24} & \frac{-\lambda_1^5+10\lambda_1^3-25\lambda_1+16}{120} \\
 & & \frac{1-3\lambda_1^2+2\lambda_1^3}{3} & \frac{-7\lambda_1^4+16\lambda_1^3-6\lambda_1^2-8\lambda_1+5}{24} \\
 & & & \frac{7\lambda_1^5-20\lambda_1^4+10\lambda_1^3+20\lambda_1^2-25\lambda_1+8}{60}
\end{pmatrix}
\end{aligned}
$$

Collecting terms shows that, as $(T,N)\to\infty$,

$$\begin{aligned}
(UU)/(T^{3/2}N) &= m_T tr\left[\begin{array}{c} 2\Omega_1\left(\xi_{11}C(1)'A_{HH}C(1)\xi_{21}'+\bar{\sigma}_d^2\Sigma_e\right) \\ -\Omega_1\Sigma_f\Omega_1\left(\xi_{11}C(1)'A_{HH}C(1)\xi_{11}'+\bar{\sigma}_d^2\Sigma_d\right) \end{array}\right]+o_p(1) \\
&= m_T tr\left[2\Omega_1\left(\xi_{11}C(1)'A_{HH}C(1)\xi_{21}'\right)-\Omega_1\Sigma_f\Omega_1\left(\xi_{11}C(1)'A_{HH}C(1)\xi_{11}'\right)\right] \\
&\quad +m_T\bar{\sigma}_d^2\frac{2\left(2\lambda_1^2-46\lambda_1+45\right)}{15\lambda_1}+o_p(1)
\end{aligned}$$

Therefore, the result in the Theorem follows.■

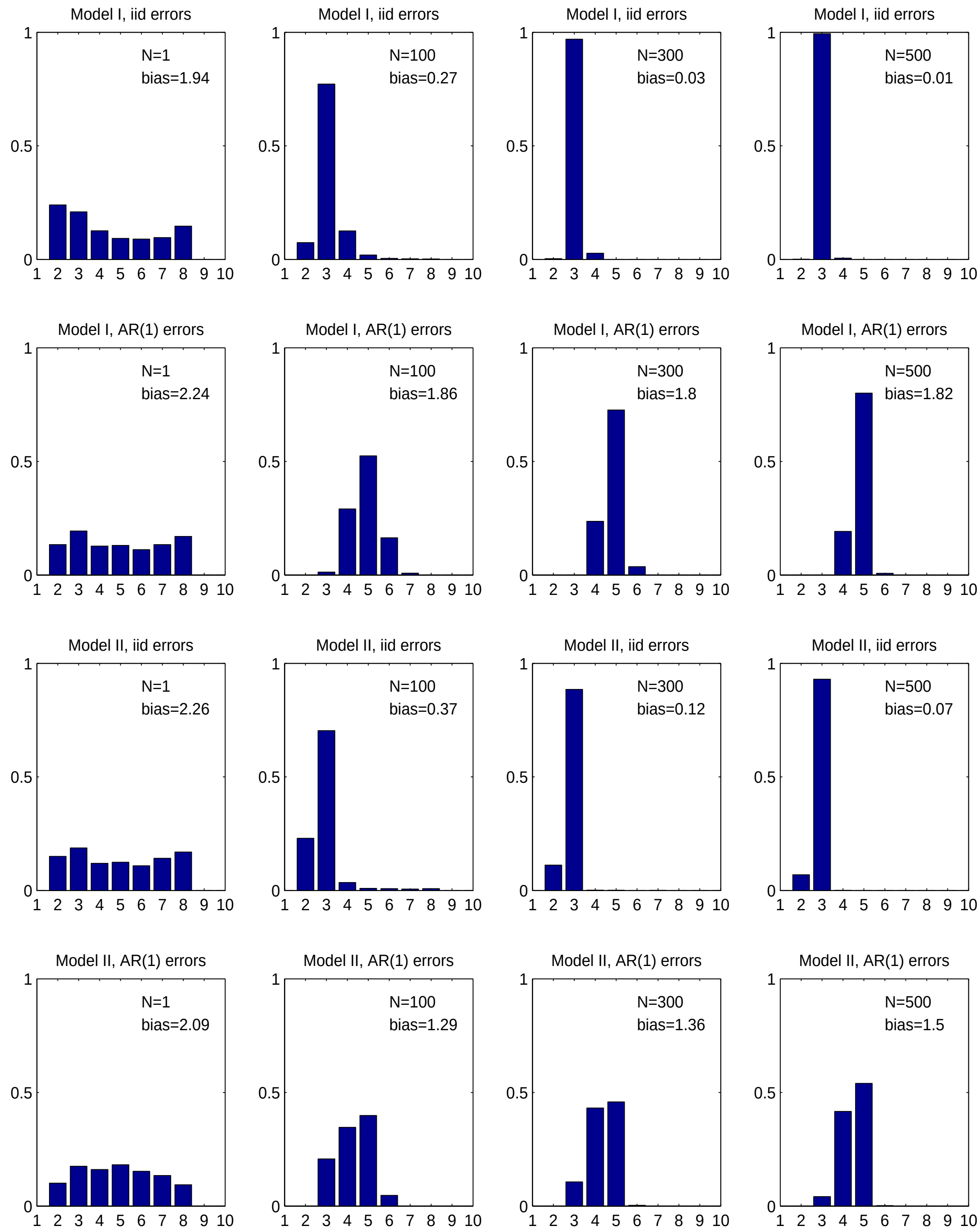

Figure 1. Consistency of the Break Date Estimate, Histogram of $\hat{T}_1$ with $T_1 = 3$.

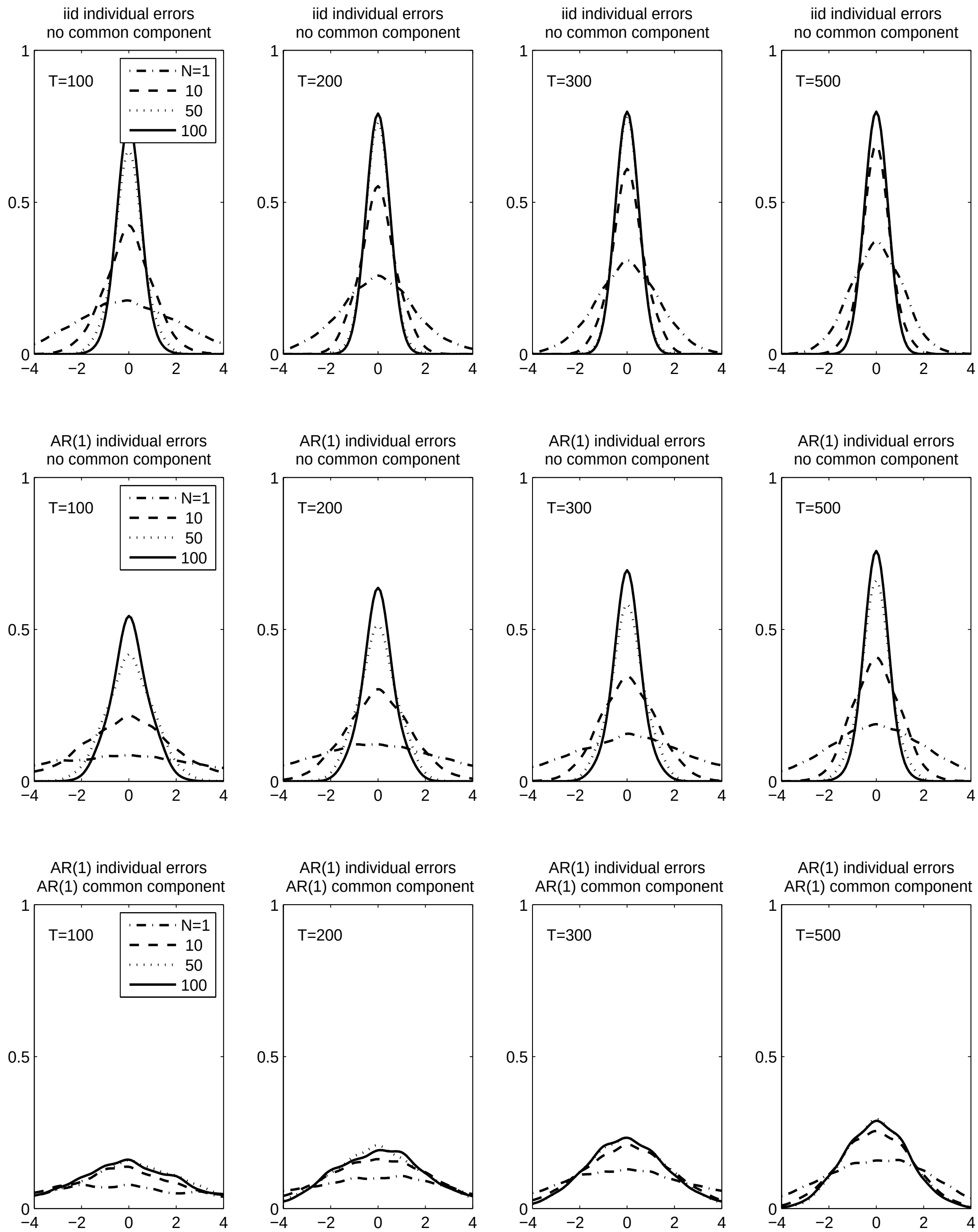

Figure 2. Estimated Density of $\hat{T}_1 - T_1$, Model I with Stationary Errors

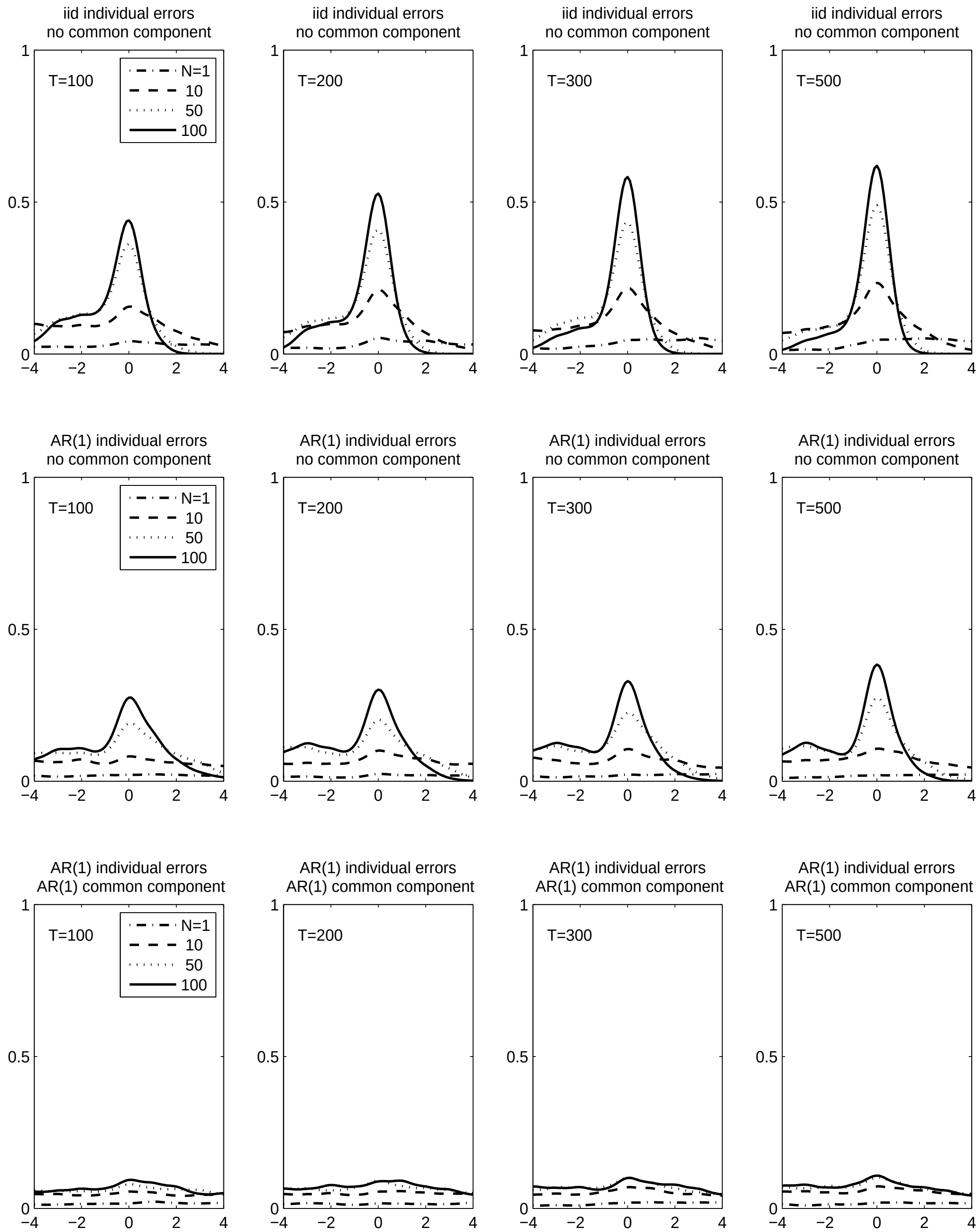

Figure 3. Estimated Density of $\hat{T}_1 - T_1$, Model II with Stationary Errors

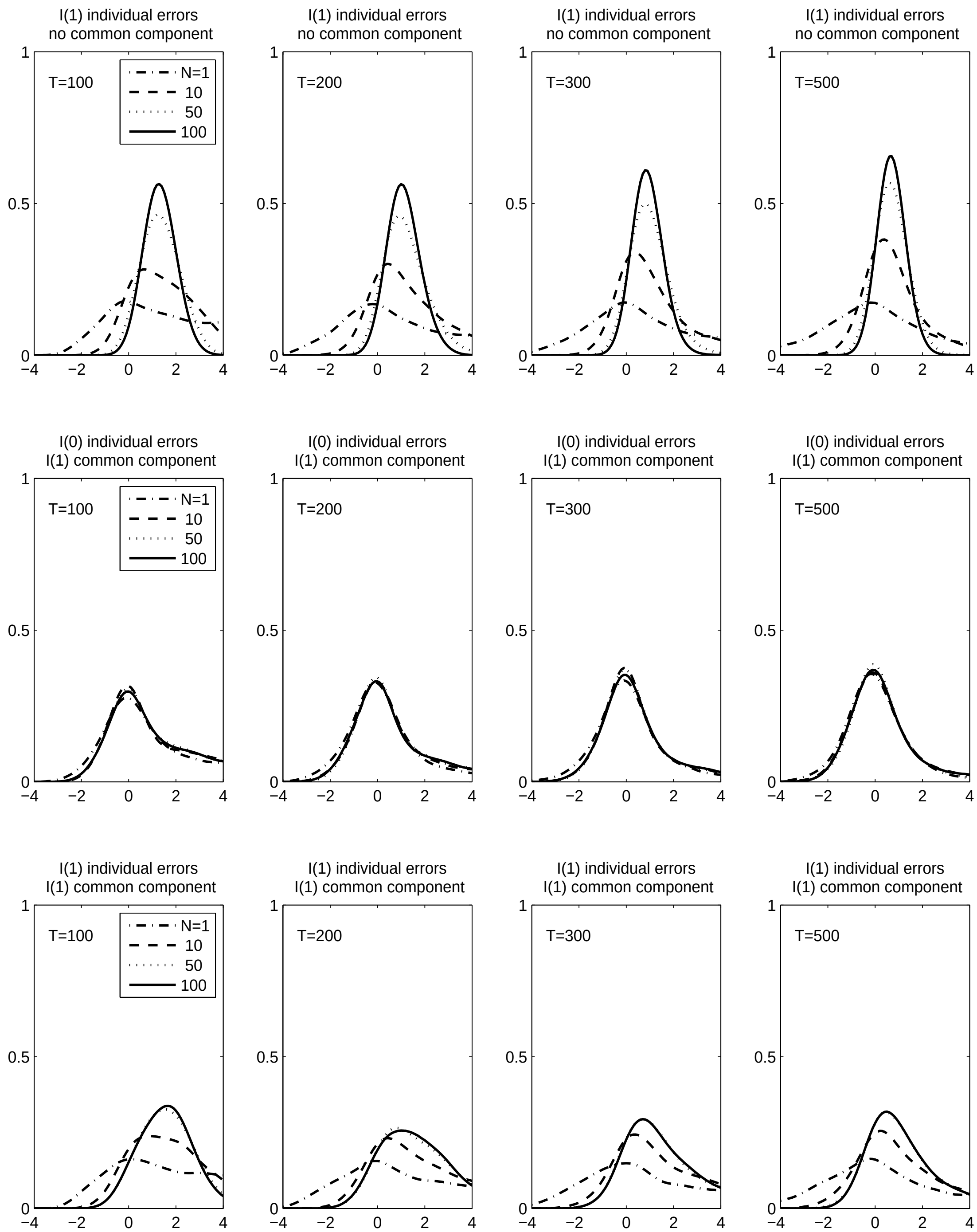

Figure 4. Estimated Density of $(\hat{T}_1 - T_1)/\sqrt{T}$, Model I with Integrated Errors

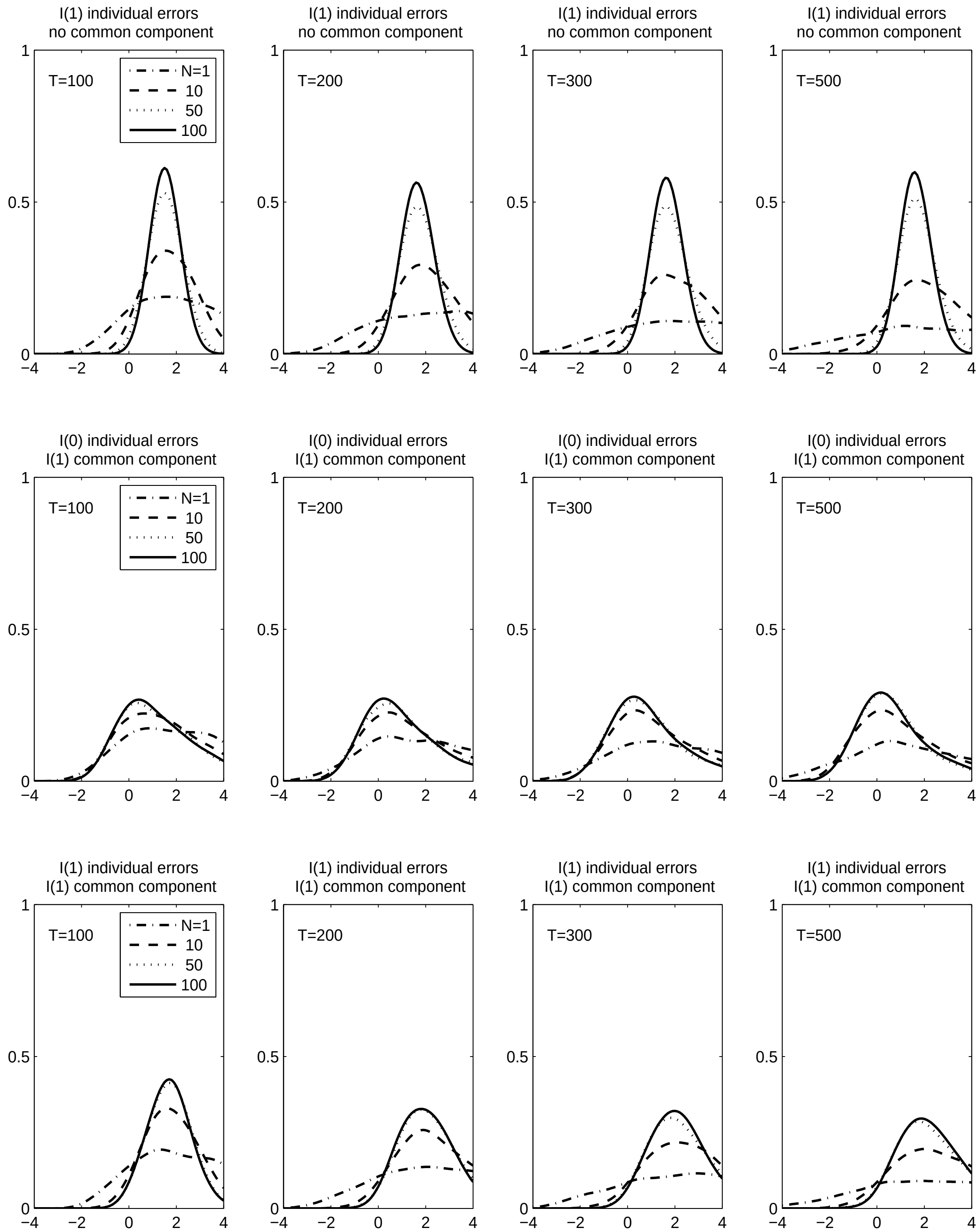

Figure 5. Estimated Density of $(\hat{T}_1 - T_1)/\sqrt{T}$, Model II with Integrated Errors