Journal of Econometrics 157 (2010) 78–92

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Common breaks in means and variances for panel data

Jushan Bai *

Department of Economics, Columbia University, 420 West 118th Street, New York, NY 10027, United States
SEM, Tsinghua University, and CEMA, Central University of Economics and Finance, Beijing, China

ARTICLE INFO

Article history:
Available online 30 October 2009

JEL classification:
C23
C33
C51

Keywords:
Change points
Panel data
Structural breaks
Common breaks

ABSTRACT

This paper establishes the consistency of the estimated common break point in panel data. Consistency is obtainable even when a regime contains a single observation, making it possible to quickly identify the onset of a new regime. We also propose a new framework for developing the limiting distribution for the estimated break point, and show how to construct confidence intervals. The least squares method is used for estimating breaks in means and the quasi-maximum likelihood (QML) method is used to estimate breaks in means and in variances. QML is shown to be more efficient than the least squares even if there is no change in the variances.

1. Introduction

This paper studies the problem of structural changes for panel data, in which there are N series (variables), and each series has T observations. It is assumed that a change has taken place in each series at an unknown common point, referred to as the common break point. This paper focuses on the statistical properties of the estimated break point when N is large.

Common breaks in panel data are wide spread phenomena. For example, a credit crunch or debt crisis may affect every company's stock returns, and an oil price shock may impact every country's output. A tax policy change may alter each firm's investment incentive. A fad or fashion can influence a large section of the society. Likewise, an emergence of new technology, a discovery of a new medicine, and an enaction of new governmental program have their own consequences on people or other entities. While it may be difficult to identify a break point with a single series, it should be, intuitively, much easier to locate the common break point using a number of series together. In this paper, we explore the panel data approach to the estimation of break point.

In comparison to the vast literature on change point for univariate series, the corresponding literature for panel data is quite small. Joseph and Wolfson (1992, 1993) are the early researchers who laid the groundwork in this area. They proposed a random break model in which each series has its own break point; across the N series, the break points are assumed to be independent and identically distributed (iid). They showed that the common distribution of the iid break points can be consistently estimated. Under the random break model, the likelihood function is similar to that of mixture distributions and the maximum likelihood estimators are obtained via EM algorithm. The present paper departs from theirs in several directions.

First, motivated by some concrete economic problems, we focus on the common break situation, in which all series have the same break point. Theoretically, common break is a more restrictive assumption than the random breaks of Joseph and Wolfson (1993). Nevertheless, when break points are indeed common, as a result of common shocks or policy shift affecting every individual, imposing the constraint gives a more precise estimation. Computationally, common break model is much simpler. Furthermore, even if each series has its own break point, the common break method can be considered as estimating the mean of the random break points, which can be useful.

Second, the maximum likelihood estimation of Joseph and Wolfson (1993) assumes that the pre-break data for all series are sampled from a single distribution F_1 and the post-break data from another single distribution F_2. For panel data, this appears to be a restricted assumption. In the case of a break in a time series mean, this would imply that the pre-break mean μ_{i1} and the post-break mean μ_{i2} for the ith series do not depend on i, and, of course, neither does the magnitude of break $\mu_{i2} - \mu_{i1}$. The present paper allows heterogeneous means, which are important for some practical applications. For example, the effect of an oil price shock on economic growth varies from country to country, depending on whether an economy is oil importing or exporting as well as on an economy's extent of oil consumption. The magnitude of change

* Corresponding address: Department of Economics, Columbia University, 420 West 118th Street, New York, NY 10027, United States. Tel.: +1 (212) 998 8960; fax: +1 (212) 995 4186.
E-mail address: jb3064@columbia.edu.

0304-4076/$ – see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.jeconom.2009.10.020

in the mean growth rate $\mu_{i2} - \mu_{i1}$ depends on i and it can be positive for some countries and negative for some others. From the statistical point of view, the heterogeneous values for (μ_{i1}, μ_{i2}) create a truly incidental parameter problem for large N panel data. Nevertheless, even with the presence of incidental parameters, we establish consistency of the break point estimator. In addition, we consider least squares estimation, requiring no distributional assumptions except existence of some moments.

Third, in terms of the theoretical proofs, Joseph and Wolfson (1993) provided an indirect proof by verifying the five assumptions of Kiefer and Wolfowitz (1956). In this paper, under the common break setup, we provide an elementary, self-contained, and yet rigorous proof for consistency, without assuming a compact parameter space as required by Kiefer and Wolfowitz.

Fourth, Joseph and Wolfson (1993) considered the case of fixed T. We consider both fixed T and T going to infinity. When $T \to \infty$, and if the common break point k_0 is allowed to take on any integer value in $[1, T-1]$, it becomes more difficult to locate the break point, in this case, we show that consistency requires $T/N \to 0$. If the conventional assumption that $k_0 = [T\tau_0]$ $(0 < \tau_0 < 1)$ is imposed, $T/N \to 0$ is not needed, and $N \to \infty$ is sufficient. A large T helps identity the break point, and indeed, we show that the break point can be consistently estimated even with small (converging to zero) magnitude of breaks. Whether or not k_0 is restricted, an unbounded T requires a separate argument for consistency. This is due to, among other things, the following fact: the maximum of any fixed number of random variables is stochastically bounded ($O_p(1)$), but the maximum of unbounded number of random variables diverges, in general, as the number of random variables going to infinity.

Finally, we propose a new framework to formulate the limiting distribution. This new framework is suitable under the common break assumption, and it overcomes a number of limitations associated with univariate data. Based on the limiting distribution, we show how to construct confidence intervals for the true common break point.

The random break model of Joseph and Wolfson (1993) is extended to autoregressive models by Joseph et al. (1996). A Bayesian framework was considered by Joseph et al. (1997). Application oriented Bayesian models (medical) were studied by Skates et al. (2001), and Jackson and Sharples (2004). More recently, Emerson and Kao (2001, 2002) and Wachter and Tzavalis (2004a,b) developed test statistics for break points in panel data. In this paper, we provide a systematic treatment of panel data change point problem from the perspective of classical inference. The focus is on the estimation of the break point given its existence.

For panel data, the number of series N can be much larger than the number of observations T. For microeconomic data, N usually represents the number of firms or the number of individuals, and T is the number of years. For example, the Panel Study of Income Dynamics (PSID) data have thousands of families, but the number of observations over time is about 40. For macroeconomic data, N and T are usually comparable in size, e.g., the number of countries and number of years with complete GDP data. Therefore, in developing the asymptotic theory, the limit is taken as the number of series goes to infinity. The number of observations T is fixed or going to infinity as well.

Panel data approach to the change point problem offers an interesting and unique perspective that is not shared by a univariate or fixed N approach. In a univariate case, the break point cannot be consistently estimated, no matter how large is the sample. We show that in panel data it is possible to obtain consistent estimates, as the number of series going to infinity. In a univariate setting, it is impossible to identify the break point when a regime has a single observation, because the change can easily be mistaken as an unusual realization of the disturbance term. With the assistance of the panel data, we show that consistency is attainable even when a regime has a single observation. This property is especially useful when the objective is to locate as quickly as possible the onset of a new regime or the turning point, without the need of waiting for many observations from the new regime.

2. The model

To highlight the main idea, we shall consider a simple mean shift model.

$$\begin{aligned} Y_{it} &= \mu_{i1} + e_{it}, \quad t = 1, 2, \ldots, k_0 \\ Y_{it} &= \mu_{i2} + e_{it}, \quad t = k_0 + 1, \ldots, T \\ & \quad i = 1, 2, \ldots, N \end{aligned} \tag{1}$$

where $E(e_{it}) = 0$ for all i and t. In this model, each series has a break at k_0, where k_0 is unknown. The pre-break mean of Y_{it} is μ_{i1} and post-break mean is μ_{i2}. The difference $\mu_{i2} - \mu_{i1}$ represents the magnitude of break, which can be either random or nonrandom, and is assumed to be independent of error process e_{it}. In addition, $E(\mu_{i2} - \mu_{i1})^2 \le M$ for all i. We refer to T as the number of observations or sample size, and refer to N as the number of variables or the number of series. For panel data, N is usually large relative to T. This common break model in panel data is called "fixed τ model" by Joseph and Wolfson (1992). Note that $k_0 = T$ implies no break in the sample. Thus the maximum value for k_0 is $T - 1$, assuming existence of a break.

We consider two cases with respect to the size T; the first case corresponds to a fixed T, and k_0 can take on any value from 1 to $T-1$ without restriction. The second case assumes T grows without bound. For the latter, the break point k_0 is assumed to be bounded away from 1 and T such that

$$k_0 = [T\tau_0]$$

with $\tau_0 \in (0, 1)$ so that k_0 is a positive fraction of the total sample size, where $[x]$ represents the integer part of x. This is a conventional assumption in the change point literature, see, e.g., Csörgő and Horváth (1997). In terms of consistency, this assumption is removable in the panel data setting, as long as $T/N \to 0$, see Remark 3 in Appendix for a proof. However, for our second consistency theorem and for the limiting distribution, this condition is needed and thus it is maintained throughout. To avoid technical niceties, we simply assume $k_0 = T\tau_0$. Again, for fixed T, no such a restriction is necessary.

We assume the error process e_{it} is stationarity in the time dimension. More specifically,

Assumption 1. $e_{it} = \sum_{j=0}^{\infty} a_{ij}\varepsilon_{i,t-j}$, $\varepsilon_{it} \sim (0, \sigma_{i\varepsilon}^2)$ are iid over t; $\sum_j j|a_{ij}| \le M$ for all i. In addition, e_{it} are independent over i.

Let $\sigma_i^2 = E(e_{it}^2) = \sigma_{i\varepsilon}^2(\sum_j a_{ij}^2)$. We assume there is no change in the variance. If a shift in the variance also occurs, it is more efficient to estimate simultaneously the break in variance and the break in mean. When a break in variance exists but is ignored, the estimated break in mean, while consistent, will have a nonsymmetric limiting distribution. Bai (1997b) discusses nonsymmetric distribution due to nonconstant variance or nonstationary regressors. Changes in variance will be analyzed in Section 5.

Assumption 1 assumes e_{it} are cross-sectionally independent. This assumption can be relaxed without affecting the consistency result. Cross-sectional independence is mainly used for the term $N^{-1/2}\sum_{i=1}^{N} e_{it}$ to be $O_p(1)$ (for consistency) and to have the central limit theorem (for limiting distribution). Clearly, central limit theorem still holds under cross-sectional correlation, provided that the correlation is weak.

With respect to the size of breaks, it seems reasonable to assume a positive limit for $\liminf_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2 > 0$. This would be the case if the magnitude of breaks $\mu_{i2} - \mu_{i1}$ are iid random variables with positive variance. However, we shall assume a weaker condition:

Assumption 2.

$$\lim_{N\to\infty} N^{-1/2} \sum_{i=1}^{N} (\mu_{i2} - \mu_{i1})^2 = \infty. \tag{2}$$

The sum is divided by $N^{1/2}$ instead of N. The condition does not require every series to have a break.

3. Least squares estimation of the break point

For a given k such that $1 \le k \le T-1$, define

$$\bar{Y}_{i1} = \frac{1}{k}\sum_{t=1}^{k} Y_{it}$$

$$\bar{Y}_{i2} = \frac{1}{T-k}\sum_{t=k+1}^{T} Y_{it}$$

so that $\bar{Y}_{i1}$ and $\bar{Y}_{i2}$ are estimators for μ_{i1} and μ_{i2}, respectively. Their dependence on k is suppressed for notational simplicity. They are biased estimators unless $k = k_0$. This biased property when $k \neq k_0$ in fact helps identity the true break point. Define the sum of squared residuals for the ith equation as

$$S_{iT}(k) = \sum_{t=1}^{k} (Y_{it} - \bar{Y}_{i1})^2 + \sum_{t=k+1}^{T} (Y_{it} - \bar{Y}_{i2})^2$$

$k = 1, 2, \ldots, T-1$. We define $S_{iT}(k) = \sum_{t=1}^{T} (Y_{it} - \bar{Y}_i)^2$ for $k = T$, where $\bar{Y}_i$ is the whole sample mean. In this way, $S_{iT}(k)$ is defined for every $k = 1, \ldots, T$. The total sum of squared residuals across all equations is defined as

$$SSR(k) = \sum_{i=1}^{N} S_{iT}(k).$$

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

$$\hat{k} = \underset{1\le k\le T-1}{\operatorname{argmin}}\, SSR(k).$$

This estimator is straightforward to compute.

An alternative procedure is to estimate the break point series by series and then take the average over the N estimates. This method is not recommended because it does not guarantee consistency. If a portion of series are not subject to breaks, the estimated break points for these series can take on any value. Averaging those values will not give the correct answer since they do not necessarily fluctuates around k_0, particularly when k_0 is near the boundary.

When $N = 1$ (a univariate series), it is well known that

$$\hat{k} = k_0 + O_p(1) \tag{3}$$

so that the difference between $\hat{k}$ and true break point k_0 is stochastically bounded. For a fixed T, such a statement is not helpful because T is bounded and $\hat{k} - k_0$ is always bounded. When $T \to \infty$, the statement of (3) is quite strong. It implies that

$$\hat{\tau} = \tau_0 + O_p(T^{-1})$$

where $\hat{\tau} = \hat{k}/T$. So in terms of the fraction of the sample size, $\hat{\tau}$ is T-consistent for τ_0. Nevertheless, $\hat{k}$ itself is not consistent for k_0 in univariate framework.

For panel data, however, much stronger statements can be made. We shall prove the following result:

Theorem 3.1. *For model (1), assume Assumptions 1 and 2 hold. Then under either fixed T or unbounded T,*

$$\lim_{N\to\infty} P(\hat{k} = k_0) = 1.$$

To access the theoretical result numerically, we report some Monte Carlo simulations. Attention is given to the effect of N on the precision of estimated $\hat{k}$. We generate data based on model (1), with $T = 10$, $k_0 = 5$, and $N = 1, 10, 20, 100$. The actual levels of μ_{i2} and μ_{i1} are not essential, it is the difference $\mu_{i2} - \mu_{i1}$ that matters. This difference is generated as having a uniform distribution on the interval $(-2, 2)$,

$$\mu_{i2} - \mu_{i1} \sim U(-2, 2), \quad i = 1, 2, \ldots, N.$$

The disturbances e_{it} are iid standard normal $N(0, 1)$.

Fig. 1 displays four histograms for the estimated break points, with each corresponding to a different value of N. It is clear that the precision of $\hat{k}$ improves markedly as N increases. With a single series ($N = 1$), it is difficult to estimate the break point. With $N = 100$, the break point is estimated precisely (100% of precision).

More importantly, even when a regime has a single observation, the break point can still be estimated precisely. For example, with $k_0 = 9$, the second regime has a single observation given $T = 10$. Fig. 2 shows the histograms of $\hat{k}$ when $k_0 = 9$. The assertion that precise estimation is obtained as N increases holds for every k_0 in between 1 and $T-1$. This conforms with Theorem 3.1 since the theorem does not impose any restriction on k_0.

Theorem 3.1 states that the estimated break point $\hat{k}$ is consistent for k_0 under condition (2), i.e., $N^{-1/2}\sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2$ grows unbounded as N increases. The consistency holds whether T is bounded or unbounded. We now show that condition (2) can be weakened to

$$\lim_{N\to\infty} \sum_{i=1}^{N} (\mu_{i2} - \mu_{i1})^2 = \infty. \tag{4}$$

It is clear that (2) implies (4); the latter does not require a rate at which the sum diverges to infinity. As a cost to the weaker condition, we have to assume T is larger than N such that

$$\frac{\log(\log(T))}{T} N \to 0 \tag{5}$$

as T and N going to infinity. This restricts the relative rate at which T and N diverge. Intuitively, this means that with more observations, one can detect faint signals.

Theorem 3.2. *Under Assumption 1, (4) and (5), we have*

$$\lim_{N,T\to\infty} P(\hat{k} = k_0) = 1.$$

In summary, under Assumption 2, consistency is possible for either fixed T or large T. Under (4), T must be large to obtain consistency.

4. Limiting distribution

In the previous section, we showed that $P(\hat{k} = k_0) \to 1$, implying a degenerate limiting distribution for $\hat{k}$. In practice, either due to too small magnitude of breaks or due to the finiteness of N, we cannot expect $\hat{k}$ to coincide with k_0. It is therefore of interest to study the distribution of $\hat{k}$. Limiting distributions can be used to construct confidence intervals for the true break point.

In the univariate case, there are two frameworks for developing the limiting distribution of $\hat{k}$. For concreteness, consider a univariate series

$$Y_t = \mu_1 + e_t, \quad t = 1, 2, \ldots, k_0$$
$$Y_t = \mu_2 + e_t, \quad t = k_0 + 1, \ldots, T.$$

The first framework assumes the magnitude of break $|\mu_2 - \mu_1|$ is fixed and the second assumes $|\mu_2 - \mu_1|$ depends on T such that

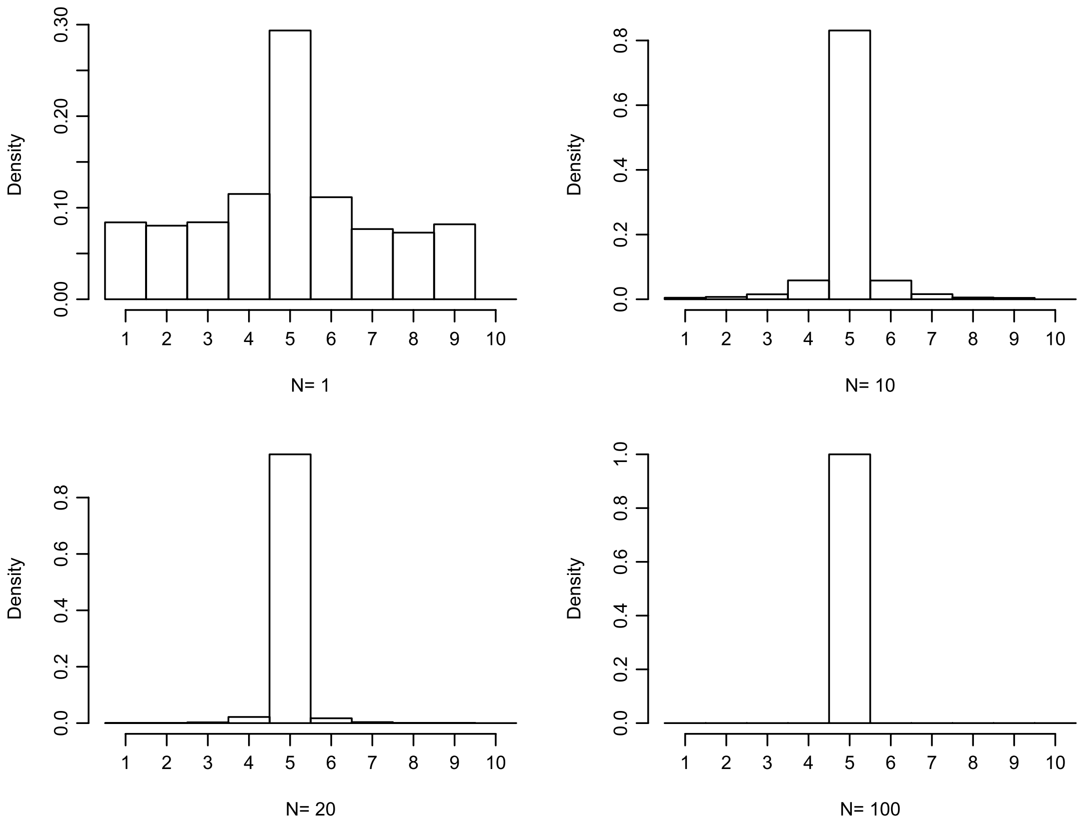

Fig. 1. Histogram of the estimated break points ($T = 10, k_0 = 5$).

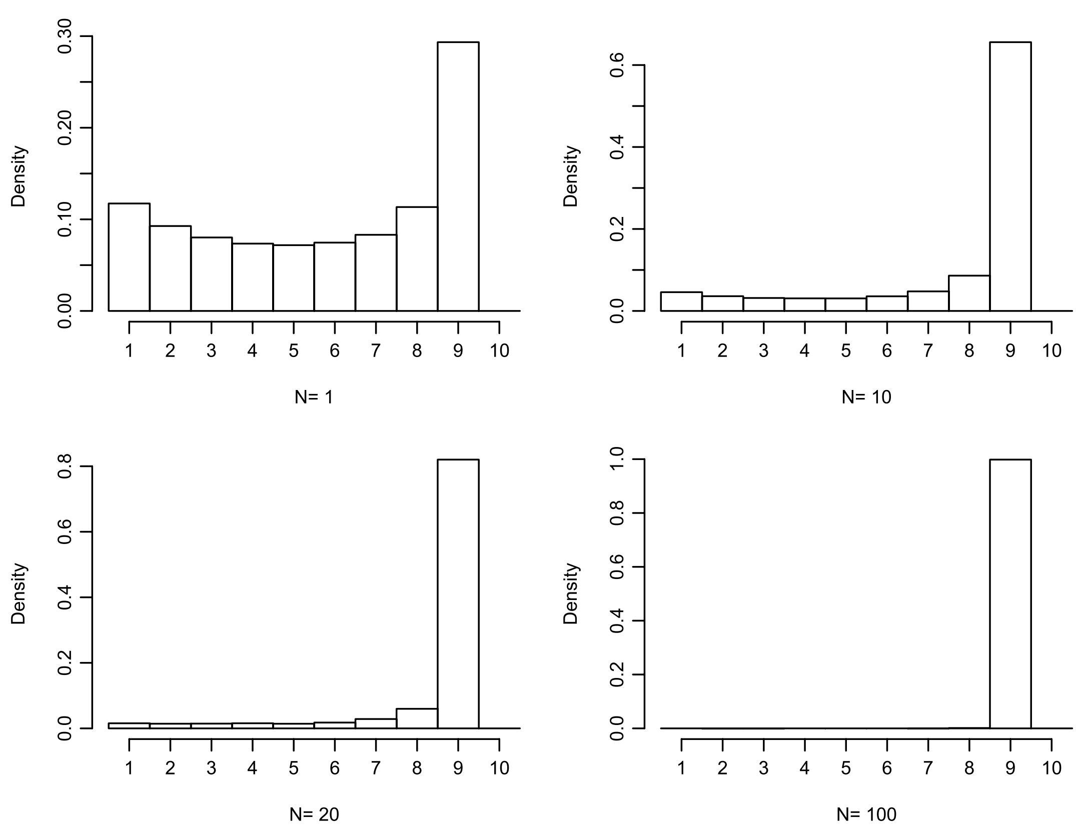

Fig. 2. Histogram of the estimated break points ($T = 10, k_0 = 9$).

$v_T = |\mu_2 - \mu_1|$ converges to zero as $T \to \infty$. Under fixed $\mu_2 - \mu_1$, the limiting distribution is given by

$$\hat{k} - k_0 \xrightarrow{d} \text{argmax}_\ell V(\ell) \tag{6}$$

where $V(0) = 0$ and

$$V(\ell) = (\mu_2 - \mu_1)^2|\ell| - 2(\mu_2 - \mu_1)\sum_{s=-\ell+1}^{0} e_s, \quad \ell = -1, -2, \ldots.$$

$$V(\ell) = (\mu_2 - \mu_1)^2\ell - 2(\mu_2 - \mu_1)\sum_{s=1}^{\ell} e_s, \quad \ell = 1, 2, \ldots$$

see Bhattacharya (1994). The above assumes e_t is strictly stationary, although independence is not necessary. If e_t is not strictly stationary, then $\sum_{s=1}^{\ell} e_s$ must be replaced by $\sum_{s=k_0+1}^{k_0+\ell} e_s$ and $\sum_{s=-\ell+1}^{0} e_s$ replaced by $\sum_{s=k_0-\ell+1}^{k_0} e_s$, see Bai (1997a,b).

In addition to the strict stationarity requirement, another restricted feature of the distribution is its dependence on the underlying distribution of e_t. Different distributions for the disturbances e_t imply different distributions for $\hat{k} - k_0$. For the limiting distribution to be useful, one must know the distribution of e_t.

The second framework assumes $v_T = \mu_2 - \mu_1 \to 0$ and $Tv_T^2 \to \infty$. The limiting distribution is

$$(v_T^2/\sigma^2)(\hat{k} - k_0) \xrightarrow{d} \underset{r}{\text{argmin}}[-|r| + 2W(r)] \tag{7}$$

where σ^2 is the long run variance of e_t, that is, the spectral density at frequency zero multiplied by 2π; $W(r)$ is a two-sided Brownian motion on the real line. More specifically $W(0) = 0$ and $W(r) = W_1(r)$ for $r > 0$, and $W(r) = W_2(-r)$ for $r < 0$, where W_1 and W_2 are two independent Brownian motions on $[0, \infty)$; See Picard (1985), Yao (1987), or Bai (1994) for e_t being a linear process.

The second framework implies that $\hat{k} - k_0$ diverges to infinity as $T \to \infty$ because v_T^2 goes to zero and the product $v_T^2(\hat{k} - k_0)$ has a non-degenerate distribution.

These limitations associated with a univariate series can be overcome under panel data. Below we introduce a new framework for developing the limiting theory. Under the new framework, we show that the limiting distribution does not depend on the underlying distribution of e_{it}, nor on the strict stationarity of e_{it}, nor on the divergence of $\hat{k} - k_0$.

We shall assume T is unbounded, as in the previous section. Thus limits are taken as N and T both go to infinity. For fixed T, the limiting distribution will depend on the exact location of k_0, and it also has more complicated expression. We therefore focus on the case of unbounded T in the limit.

Recall that as long as $\lambda_N = \sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2 \to \infty$, we have $P(\hat{k} = k_0) \to 1$, implying a degenerate distribution. To obtain a non-degenerate distribution, we assume $\lim_{N\to\infty}\sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2 = \lambda > 0$ for some $\lambda < \infty$. Roughly speaking, this implies that the magnitude of break in each equation is small. For fixed N, the sum λ_N is always bounded. So a bounded limit provides a good approximation for panel data with not too large N.

To derive the limiting distribution, we make the following specific assumption:

$$\mu_{i2} - \mu_{i1} = N^{-1/2}\Delta_i, \text{ with } \lim_{N\to\infty}\sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2 = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Delta_i^2 = \lambda \tag{8}$$

where Δ_i is uniformly bounded, or in the case of stochastic Δ_i, its variance is uniformly bounded. In addition, we assume

$$\lim_{N\to\infty}\sum_{i=1}^{N}[(\mu_{i2} - \mu_{i1})^2\sigma_i^2] = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\Delta_i^2\sigma_i^2 = \phi. \tag{9}$$

Note that $\phi = \sigma^2\lambda$ when $\sigma_i^2 = \sigma^2$ for all i.

Lemma 4.1. *Under Assumption 1, (5) and (8), we have*

$$\hat{k} - k_0 = O_p(1).$$

The $O_p(1)$ in the lemma is genuine in the sense that it is not $o_p(1)$. Thus, due to small magnitude of breaks, $\hat{k}$ does not collapse to k_0, leading to a non-degenerate distribution. Nevertheless, this lemma says that the break point can be estimated well because $\hat{\tau} = \hat{k}/T$ is still T consistent for τ_0.

Unlike the univariate case, even though $\hat{k} - k_0 = O_p(1)$, the limiting distribution will not depend on the underlying distribution of e_{it}. We state the limiting distribution in the following theorem.

Theorem 4.2. *Assume e_{it} are uncorrelated over t. Under conditions (5), (8) and (9), as $N, T \to \infty$,*

$$\hat{k} - k_0 \xrightarrow{d} \underset{\ell}{\text{argmin}}\left[\,|\ell|\,\lambda + 2\sqrt{\phi}\;W(\ell)\,\right] \tag{10}$$

where $W(0) = 0$ and

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

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

and Z_s, $s = \ldots -2, -1, 0, 1, 2, \ldots$ are iid standard normal random variables.

The key distinction between the limiting distribution in Theorem 4.2 and the limiting distribution in (6) is that the random variables Z_t are standard normal random variables instead of the disturbances of the regression. The normality of Z_s is due to the central limit theorem applied to cross-sectional regression errors. In the above theorem, the e_{it} are assumed to be uncorrelated over t. Under serial correlation for e_{it}, the theorem continues to hold, except that the normal random variables Z_s will be correlated. In order to derive or simulate the probability distribution for the limiting random variable, the correlation coefficients $\rho_h = E(Z_sZ_{s+h})$ are needed. For fixed T, it may be difficult to estimate those coefficients. But for simple parametric processes such that $e_{it} = \rho e_{it-1} + \eta_{it}$, panel data techniques as described in Arellano (2003) and Hsiao (2003) can be used to consistently estimate ρ even for fixed T. However, under fixed T, the two-sided random walk (defined on a bounded set) will be nonsymmetric implying a nonsymmetric limiting distribution for $\hat{k}$.

For given λ and ϕ, we can easily simulate the distribution on the right hand side (10). From the simulated distribution, confidence intervals on k_0 can also be constructed. Also note that λ and ϕ can be estimated consistently. In what follows, we shall provide a more practical way of constructing confidence intervals. We only need to simulate a standard distribution for $\lambda = 1$ and $\phi = 1$. For all other values of λ and ϕ, a simple transformation is sufficient. Given the critical values reported below, no further simulation is needed for applied researchers.

The $W(\ell)$ in the limit is a two-sided and discrete-time Gaussian random walk, quite analogous to the continuous-time Gaussian random walk (Brownian motion) in (7). For a moment, assume that

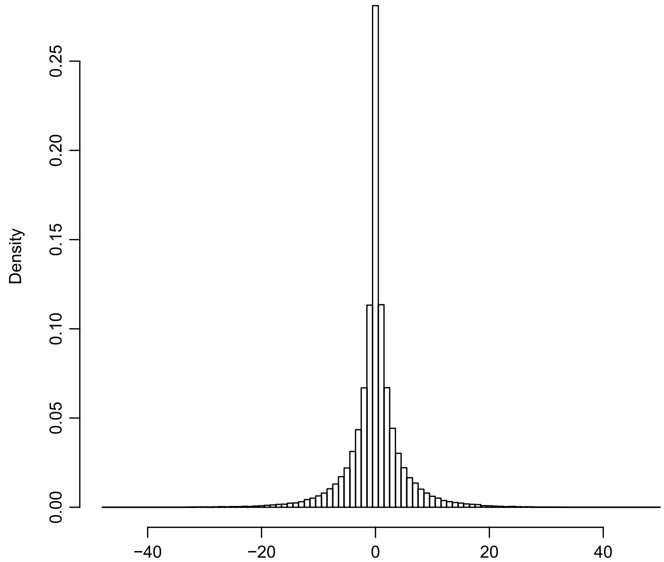

Fig. 3. The distribution of $\text{argmin}_\ell[|\ell| + 2W(\ell)]$, where $W(\ell)$ is a two-sided Gaussian random walk ($\ell = \ldots, -2, -1, 0, 1, 2, \ldots$). The distribution is obtained from simulation via 100,000 repetitions.

ℓ takes on continuous values and $W(\ell)$ is a Brownian motion, then by a simple change in variable, the right hand side of (10) is equal in distribution to $(\phi/\lambda^2)\text{argmin}_\ell[|\ell| + 2W(\ell)]$. Then we can rewrite (10) as

$$(\lambda^2/\phi)(\hat{k} - k_0) \xrightarrow{d} \underset{\ell}{\text{argmin}}\Big[\,|\ell| + 2W(\ell)\,\Big]. \tag{11}$$

The above can be rewritten as

$$A_N(\hat{k} - k_0) \xrightarrow{d} \underset{\ell}{\text{argmin}}\Big[\,|\ell| + 2W(\ell)\,\Big] \tag{12}$$

where

$$A_N = \frac{\left[\sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2\right]^2}{\sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2\sigma_i^2}$$

because $A_N \to \lambda^2/\phi$ in view of (8) and (9). In the special case that $\sigma_i^2 = \sigma^2$ for all i, A_N is simplified to $A_N = \sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2/\sigma^2$.

However, ℓ only takes on integer values, and $W(\ell)$ is a random walk, the change in variable argument does not hold exactly so that (12) is only an approximation. Nevertheless, the approximation holds well because a Gaussian random walk and a Brownian motion (evaluated at integer time) have the same distribution. The quality of the approximation is also confirmed by Monte Carlo simulations.

Let $\ell^* = \text{argmin}_\ell\{|\ell| + 2W(\ell)\}$. The distribution of ℓ^* is free from unknown parameters and is easily simulated, see Fig. 3, which is obtained from 100,000 repetitions. This discrete density function quite resembles the continuous density function of the random variable in (7). This is not surprising because, as mentioned earlier, a Gaussian random walk and a Brownian motion (evaluated at integer time) have the same distribution. From the simulated distribution, it is found that

$$P(|\ell^*| \le 7) \simeq 0.90$$
$$P(|\ell^*| \le 11) \simeq 0.95$$
$$P(|\ell^*| \le 20) \simeq 0.99.$$

Using (12), the 90% confidence interval for k_0 is constructed as

$$[\hat{k} - \text{floor}(7/\hat{A}_N),\ \hat{k} + \text{ceiling}(7/\hat{A}_N)] \tag{13}$$

Table 1
Coverage rate and the length of confidence intervals ($T = 100$).

Distribution of e_{it}	N	Coverage rate			Median length of CI		
		90%	95%	99%	90%	95%	99%
$N(0,1)$	1	0.635	0.719	0.812	25	39	70
	5	0.829	0.886	0.954	9	13	23
	10	0.900	0.932	0.979	5	7	13
	15	0.937	0.968	0.989	5	7	9
	20	0.949	0.983	0.994	3	5	7
$\chi^2_{(5)}$	1	0.647	0.722	0.815	25	39	69
	5	0.824	0.885	0.944	9	13	23
	10	0.905	0.941	0.979	5	9	13
	15	0.933	0.963	0.989	5	5	9
	20	0.942	0.959	0.986	3	5	7
$t_{(5)}$	1	0.646	0.713	0.800	25	39	69
	5	0.830	0.883	0.927	9	13	23
	10	0.904	0.947	0.978	5	7	13
	15	0.939	0.967	0.989	5	5	9
	20	0.935	0.968	0.988	3	5	7

where floor(x) is the largest integer smaller than or equal to x, and ceiling(x) is the smallest integer larger than or equal to x; $\hat{A}_N$ is an estimate of A_N, obtained by replacing unknown parameters by their estimated values. We prove in Appendix that

$$\hat{A}_N = A_N + o_p(1) \tag{14}$$

thus replacing A_N by $\hat{A}_N$ has no asymptotic effect on the confidence intervals. Replacing the value 7 in the confidence interval (13) by 11 and 20 will give rise to the 95% and 99% confidence intervals, respectively.

Monte Carlo simulation. To examine the coverage probability of the confidence intervals based on (13), we simulate observations from model (1) with $T = 100$, $k_0 = 50$, and a number of different values of N. The disturbances e_{it} are generated as iid for all i and t with three different distributions, namely, normal $N(0, 1)$, chi-square $\chi^2_{(5)}$, and the student t distribution $t_{(5)}$. The last two distributions allows us to see the effect of skewness and heavy-tailedness on the coverage probability.

The estimator $\hat{k}$ is invariant to the actual levels of μ_{i1} and μ_{i2}; only the difference $\mu_{i2} - \mu_{i1}$ matters. We generate the magnitude of break as uniform random variables such that

$$\mu_{i2} - \mu_{i1} \sim \sigma \cdot U(-1, 1), \quad i = 1, 2, \ldots, N$$

where σ is the standard deviation of the disturbances, i.e., $\sigma^2 = E(e_{it}^2)$. These are small levels of breaks relative to the error variance. From a single series, it would be very difficult to accurately estimate the break point for this magnitude of break.

To construct confidence intervals, we estimate A_N as follows. In each repetition, we estimate the break point $\hat{k}$ first. We then estimate μ_{i1} and μ_{i2} based on the pre-break and post-break sample means using $\hat{k}$ as the break point. Then A_N is estimated by $\hat{A}_N = \sum_{i=1}^{N}(\hat{\mu}_{i2} - \hat{\mu}_{i1})^2/\hat{\sigma}^2$, where $\hat{\sigma}^2$ is the estimated error variance defined as $\hat{\sigma}^2 = \sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2/(NT - 2N)$.

Table 1 presents the coverage rates (frequencies) from 5000 repetitions. The first column lists the distribution types, the second column gives the values on $N = 1, 5, 10, 15, 20$. The next three columns provide the coverage rates based on the confidence intervals in (13) (with 7 replaced by 11 and 20 for the 95% and 99% confidence intervals). The last three columns are the median length of the confidence intervals.

Upon inspecting the table, three conclusions can be drawn. First, as N increases, the length of the confidence interval becomes tighter, reflecting more precise estimation of k_0 (see the last three columns). Even with narrower confidence intervals, the coverage rate increases as N gets larger. The interval is widest for $N = 1$ with poor coverage rate. Due to the use of "floor" and "ceiling" functions, the shortest confidence interval (by construction) would contain three integers $(\hat{k}-1, \hat{k}, \hat{k}+1)$. If we round $7/\hat{A}_N$ to the nearest integer instead of using the floor and ceiling functions, then the shortest confidence interval would contain a single element $\{\hat{k}\}$. This would be the case under large magnitude of breaks (more precisely, when $7/\hat{A}_N$ is less than 0.5). We did not consider such intervals. Second, as N increases, say no smaller than 15, the coverage rate in general exceeds the nominal rate. This is consistent with the theory which predicts $\hat{k}-k_0$ converges to zero under fixed magnitude of breaks. While ideally, one may want to see the actual coverage rates being close to their nominal counterparts, it might be desirable to have the coverage rate higher than the theoretical one. It is more assuring to have conservative confidence intervals but not at the expense of widening the intervals. In our case, as N increases (not reported) the coverage rates will increase to 1. Finally, skewness and heavy-tailedness do not seem to adversely affect the confidence intervals, as shown in the second and third panel of Table 1, as predicted by the theory.

Determining which series had undergone a change. Our assumptions do not require each series to have a break. It is therefore of interest to determine which series had in fact undergone a structural break. For this purpose, the standard Chow test (Chow, 1960) computed at the estimated common break and the associated chi-square critical values are applicable, provided that the test is performed series by series. To see this, consider testing whether there is a break in the first series. Under large N, the estimated common break can be viewed as exogenous given, i.e., independent of the first series since the contribution from the first series to the estimated break point is asymptotically negligible as $N \to \infty$. Because for each exogenous given break point, the Chow test has a chi-square limiting distribution, the usual chi-square critical values are usable. Alternatively, one can estimate the common break point with the rest of $N-1$ series, and compute the Chow test for the first series by splitting the sample at the estimated break point. The predictive Chow test can be used if one of the regimes does not contain enough observations, a relevant case for multiple regressions.

5. Extension: Common breaks in variances

5.1. Model and estimation

Consider

$$\begin{aligned} Y_{it} &= \mu_{i1} + \sigma_{i1}\,\eta_{it}, \quad t = 1, 2, \ldots, k_0 \\ Y_{it} &= \mu_{i2} + \sigma_{i2}\,\eta_{it}, \quad t = k_0+1, \ldots, T \\ &\quad i = 1, 2, \ldots, N \end{aligned} \tag{15}$$

where $E(\eta_{it}) = 0$, $var(\eta_{it}) = 1$, and for each i, η_{it} is a linear process such that $\eta_{it} = \sum_{j=0}^{\infty} b_{ij}\varepsilon_{i,t-j}$, with ε_{it} iid (0, 1), $\sum_{j=1}^{\infty} j|b_{ij}| < \infty$, and $var(\eta_{it}) = \sum_{j=0}^{\infty} b_{ij}^2 = 1$. In addition, $E(\varepsilon_{it}^4)$ is bounded, and $k_0 = [T\tau_0]$, $\tau_0 \in (0, 1)$.

We assume either $\mu_{i1} \neq \mu_{i2}$ or $\sigma_{i1} \neq \sigma_{i2}$. For ease of analysis, we shall assume that σ_{i1} and σ_{i2} are uniformly bounded above and bounded away from zero. That is, there exist $a > 0$ and $A < \infty$ such that $a \le \sigma_{ik}^2 \le A$ $(k = 1, 2)$. To estimate the break point, we use the quasi-maximum likelihood method (QML) by treating η_{it} as if they were iid normal $N(0, 1)$. Let

$$\hat{\sigma}_{i1}^2(k) = \frac{1}{k}\sum_{t=1}^{k}(Y_{it}-\bar{Y}_{i1})^2, \qquad \hat{\sigma}_{i2}^2(k) = \frac{1}{T-k}\sum_{t=k+1}^{T}(Y_{it}-\bar{Y}_{i2})^2.$$

The QML objective function for series i is equal to

$$k\,\log\hat{\sigma}_{i1}^2(k) + (T-k)\,\log\hat{\sigma}_{i2}^2(k)$$

multiplied by $-1/2$. For a single series, the break point is estimated by minimizing the above objective function, as discussed in Bai (2000, p. 335). For N independent series, the objective function becomes

$$U_{NT}(k) = k\sum_{i=1}^{N}\log\hat{\sigma}_{i1}^2(k) + (T-k)\sum_{i=1}^{N}\log\hat{\sigma}_{i2}^2(k). \tag{16}$$

The break point estimator is defined as $\hat{k} = \mathrm{argmin}_k U_{NT}(k)$. To study the property of $\hat{k}$, we need to make an assumption about the magnitude of breaks. In case of no breaks in means, consistent estimation of the break point relies on breaks in variances. Let

$$f(x) = x - 1 - \log(x), \quad x > 0.$$

Function $f(x)$ is monotonically decreasing on (0, 1] and monotonically increasing on $[1, \infty)$, thereby achieving its unique minimum at $x = 1$ with $f(1) = 0$ and $f(x) > 0$ for $x \neq 1$. Note that $\sigma_{i1}^2 \neq \sigma_{i2}^2$ if and only if $f(\sigma_{i1}^2/\sigma_{i2}^2) > 0$. The required condition for consistent estimation of the break point is (in the absence of mean breaks)

$$\lim_{N\to\infty}\sum_{i=1}^{N} f(\sigma_{i1}^2/\sigma_{i2}^2) = \infty. \tag{17}$$

Analogous to (4), this condition does not require a rate of divergence. If it is assumed that $|\sigma_{i1}^2/\sigma_{i2}^2 - 1| > c$ for all i, then $\sum_{i=1}^{N} f(\sigma_{i1}^2/\sigma_{i2}^2) \ge c_1 N$ (for some $c_1 > 0$) so that it increases at rate N. Condition (17) is much weaker. This weak condition, however, requires the sample size T to go to infinity for asymptotic theory. Unless normality is assumed, the fixed T framework appears to be very complicated even assuming $\sum_{i=1}^{N} f(\sigma_{i1}^2/\sigma_{i2}^2)$ to be of $O(N)$. Thus we leave the fixed T analysis as an open research problem, and assume T is larger than N such that (5) hold.

Consistent estimation of the break point requires either breaks in variance as described by (17) or breaks in mean as described by (4), i.e.,

$$\lim_{N\to\infty}\sum_{i=1}^{N}(\mu_{i1}-\mu_{i2})^2 = \infty. \tag{18}$$

Now we are ready to state the main result

Theorem 5.1. *Assume that model (15) and assumption (5) hold. Then under either (17) or (18), we have, as $N, T \to \infty$,*

$$P(\hat{k} = k_0) \to 1.$$

Therefore, in presence of either mean breaks or variance breaks (or both), the break point is consistently estimable. In presence of breaks in variance only, consistent estimation of the break point is possible if QML method is used. The least squares method in previous sections is not designed to estimate breaks in variance, and will not give consistent estimation unless mean breaks exist.

5.2. Limiting distribution

The usefulness of limiting distributions is explained in early sections. That motivation remains pertinent here and will not be repeated. To derive the limiting distribution, we use the framework in which the magnitude of breaks is small. For breaks in variance, we assume $d_i = (\sigma_{i1}^2/\sigma_{i2}^2) - 1 \to 0$ so that σ_{i1} and σ_{i2} collapse to a common value as $N \to \infty$. We shall denote this common value by σ_i^2. Theorem 5.1 shows that as long as $\sum_{i=1}^{N} f(\sigma_{i1}^2/\sigma_{i2}^2) \to \infty$, a degenerate limiting distribution is assured. To obtain a non-degenerate limiting distribution, this limit must be bounded. We assume $d_i = \frac{1}{\sqrt{N}}\delta_i$ with δ_i satisfying

$$\frac{1}{N}\sum_{i=1}^{N}\delta_i^2 \to \omega, \quad \text{say}$$

for some $\omega > 0$. Then $f(\sigma_{i1}^2/\sigma_{i2}^2) = f(1+d_i) = 2^{-1}d_i^2 + O(d_i^3) = 2^{-1}N^{-1}\delta_i^2 + O(N^{-3/2})$. This implies

$$\sum_{i=1}^{N} f(\sigma_{i1}^2/\sigma_{i2}^2) = \frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2 + o(1) \to \frac{1}{2}\omega.$$

Similarly, we let $\mu_{i1} - \mu_{i2} = N^{-1/2}\Delta_i$ such that

$$\sum_{i=1}^{N}(\mu_{i1} - \mu_{i2})^2/\sigma_i^2 = \frac{1}{N}\sum_{i=1}^{N}\Delta_i^2/\sigma_i^2 \longrightarrow \tau > 0, \quad \text{say.}$$

Finally, let κ be the fourth cumulant of η_{it}. Having defined the parameters ω, τ, and κ, we are ready to state the limiting distribution.

Theorem 5.2. *For small magnitudes of breaks described by τ and ω, if each of η_{it} and $\eta_{it}^2 - 1$ is a serially uncorrelated sequence, then*

$$\hat{k} - k_0 \xrightarrow{d} \underset{\ell}{\text{argmin}}\Big[\,|\ell|(\tau + 2^{-1}\omega) + 2\sqrt{\tau}\,W_1(\ell) + \sqrt{(\kappa+2)\omega}\,W_2(\ell)\Big].$$

where both $W_1(\ell)$ and $W_2(\ell)$ are two-sided Gaussian random walks, having the same form as the $W(\ell)$ defined in Theorem 4.2; $\ell = \ldots, -2, -1, 0, 1, 2, \ldots$.

Two special cases are of interest. In the absence of breaks in variance, $\omega = 0$, we have

Corollary 5.3. *Under assumptions of Theorem 5.2 and no breaks in variance,*

$$\hat{k} - k_0 \xrightarrow{d} \underset{\ell}{\text{argmin}}\Big[\,|\ell|\,\tau + 2\sqrt{\tau}\,W_1(\ell)\Big].$$

Analogous to (11), a variable transformation will lead to $\tau(\hat{k} - k_0) \xrightarrow{d} \text{argmin}_\ell\Big[\,|\ell| + 2W_1(\ell)\Big] =^d \ell^*$, where "$=^d$" means equal in distribution, and ℓ^* is defined earlier. Again, this result holds only approximately since W is not a continuous-time Brownian motion. In view that τ is the limit of $B_N = \sum_{i=1}^N(\mu_{i1} - \mu_{i2})^2/\sigma_i^2$, we have

$$B_N(\hat{k} - k_0) \xrightarrow{d} \ell^*.$$

The limit remains valid when B_N is estimated by replacing μ_{i1}, μ_{i2}, σ_i^2 by their estimates. Recall that the least squares estimator has the same limit except that B_N is replaced by A_N, see (12). By the Cauchy–Schwarz inequality, $A_N \le B_N$. This implies that the QML is more efficient than the least squares. In the absence of cross-sectional heteroskedasticity (i.e., $\sigma_i^2 = \sigma^2$ for all i), the least squares method and QML are equivalent. With heteroskedasticity, efficiency gain arises from the fact that QML is asymptotically using the criterion $\sum_{i=1}^N \sigma_i^{-2}S_{iT}(k)$, which is the GLS criterion. See the proof in the Appendix. Thus QML down weights more noisy series. Clearly, the GLS criterion itself is infeasible since σ_i^2 is not observable. But a feasible GLS, a two-step procedure, is readily available. The first step uses the least squares method to get $\hat{k}$. From this $\hat{k}$, we estimate the variance $\hat{\sigma}_i^2 = (T-2)^{-1}S_{iT}(\hat{k})$. We then reestimate the break point with the feasible GLS criterion $\sum_{i=1}^N \hat{\sigma}_i^{-2}S_{iT}(k)$. It can be shown that the feasible GLS is asymptotically equivalent to QML.

It is interesting to note that when changes in variances are also present, if the following weighted least squares (WLS) criterion is used to estimate the break in means, $WLS(k) = \frac{1}{\sigma_{i1}^2}\sum_{t=1}^k(Y_{it} - \bar{Y}_{i1})^2 + \frac{1}{\sigma_{i2}^2}\sum_{t=k+1}^T(Y_{it} - \bar{Y}_{i2})^2$, where σ_{i1} and σ_{i2} are assumed to be known for simplicity, then the estimator for k_0 is not necessarily consistent, while both the least squares and QML will be consistent. Consider, for example, $\mu_{i2} = 2\mu_{i1}$ and $\sigma_{i2} = 2\sigma_{i1}$. WLS is equivalent to dividing the subsample $[k, T]$ by 2. Upon the division, there is no break in either the mean or variance, at least evaluated at $k = k_0$. For this particular example, QML will work well since breaks exist in both means and variances.

Another special case of Theorem 5.2 is that there are no breaks in mean but only breaks in variance. This gives $\tau = 0$, and we have

Corollary 5.4. *Under the assumptions of Theorem 5.2 and absence of mean breaks,*

$$\hat{k} - k_0 \xrightarrow{d} \underset{\ell}{\text{argmin}}\Big[\,|\ell|\,2^{-1}\omega + \sqrt{(\kappa_2+2)\omega}\,W_2(\ell)\Big].$$

Again a variable transformation shows that $\frac{\omega}{\kappa+2}(\hat{k} - k_0) \xrightarrow{d} \text{argmin}_\ell[\,|\ell| + 2W_2(\ell)] =^d \ell^*$. In view ω is the limit of $2\sum_{i=1}^N f(\sigma_{i1}^2/\sigma_{i2}^2)$, we have

$$\left[\frac{2}{\kappa+2}\sum_{i=1}^N f(\sigma_{i1}^2/\sigma_{i2}^2)\right](\hat{k} - k_0) \xrightarrow{d} \ell^*.$$

It can be shown that the same limit is obtained when κ and σ_{ik} are replaced by their consistent estimates $\hat{\kappa}$ and $\hat{\sigma}_{ik}^2$ $(k = 1, 2)$. This result can be used to construct confidence intervals. Under normality of η_{it}, the fourth cumulant $\kappa = 0$. In addition, with small changes, $\sum_{i=1}^N f(\sigma_{i1}^2/\sigma_{i2}^2) = 2^{-1}\sum_{i=1}^N d_i^2 + o(1)$ with $d_i = \sigma_{i1}^2/\sigma_{i2}^2 - 1$, so that $2^{-1}\sum_{i=1}^N d_i^2(\hat{k} - k_0)$ converges to the same limiting distribution. This expression is analogous to Corollary 7 of Bai (2000) for the case of $N = 1$.

Remark 1. The two Gaussian random walks W_1 and W_2 in Theorem 5.2 will be independent if $E(\eta_{it}^3) = 0$. They will also be independent if the magnitude of breaks in mean and the magnitude of breaks in variance are orthogonal across i in the sense that $\sum_{i=1}^N[(\mu_{i1} - \mu_{i2})/\sigma_i][(\sigma_{i1}^2 - \sigma_{i2}^2)/\sigma_i^2] = N^{-1}\sum_{i=1}^N(\Delta_i/\sigma_i)\delta_i \to 0$. This follows because the normal random variables that make up the two random walks are the limits of $N^{-1/2}\sum_{i=1}^N(\Delta_i/\sigma_i)\eta_{it}$ and $N^{-1/2}\sum_{i=1}^N\delta_i(\eta_{it}^2 - 1)$, respectively. The limits are independent if $E(\eta_{it}^3) = 0$ or $N^{-1}\sum_{i=1}^N(\Delta_i/\sigma_i)\delta_i \to 0$; see further details in Appendix.

Remark 2. When W_1 and W_2 are independent, $2\sqrt{\tau}\,W_1(\ell) + \sqrt{(\kappa+2)\omega}\,W_2(\ell)$ has the same distribution as $[4\tau + (\kappa + 2)\omega]^{1/2}W(\ell) = 2[\tau + (2+\kappa)\omega/4]^{1/2}W(\ell)$, where $W(\ell)$ is a two-sided Gaussian random walk. Therefore, we can rewrite the limiting distribution as

$$\hat{k} - k_0 \xrightarrow{d} \underset{\ell}{\text{argmin}}\Big[\,|\ell|(\tau + 2^{-1}\omega) + 2[\tau + (2+\kappa)\omega/4]^{1/2}W(\ell)\Big].$$

To allow dependence between W_1 and W_2. Let $\mu_3 = E(\eta_{it}^3)$ and assume $\sum_{i=1}^N[(\mu_{i1}-\mu_{i2})/\sigma_i][(\sigma_{i1}^2-\sigma_{i2}^2)/\sigma_i^2] = N^{-1}\sum_{i=1}^N(\Delta_i/\sigma_i)\delta_i \to \pi$. We show in Appendix that

$$\hat{k} - k_0 \xrightarrow{d} \underset{\ell}{\text{argmin}}\Big[\,|\ell|(\tau + 2^{-1}\omega) + 2[\tau + (2+\kappa)\omega/4 + \mu_3\pi]^{1/2}W(\ell)\Big]. \quad (19)$$

A change in variable implies the following approximation

$$\frac{(\tau + 2^{-1}\omega)^2}{\tau + (2+\kappa)\omega/4 + \mu_3\pi}(\hat{k} - k_0) \xrightarrow{d} \ell^*.$$

The case of $\mu_3 = 0$ has a similar form as Corollary 3 in Bai (2000) with fixed N but allowing correlations across i for η_{it} (VAR system). But there, the Gaussian random walk is replaced by a continuous-time Brownian motion.

6. Discussion: Multiple breaks

Using panel data is particularly suitable for estimating multiple breaks. With multiple breaks, regime span is short, making it difficult to identity the breaks with a single series. As demonstrated in previous sections, with panel data, a break point can be consistently estimated even if a regime contains a single observation. In this section, we discuss how to estimate multiple common break points. Consider the model with m breaks occurring at $k_1^0, k_2^0, \ldots, k_m^0$:

$$
\begin{aligned}
Y_{it} &= \mu_{i1} + e_{it}, \quad t = 1, 2, \ldots, k_1^0, \\
Y_{it} &= \mu_{i2} + e_{it}, \quad t = k_1^0 + 1, \ldots, k_2^0, \\
&\vdots \\
Y_{it} &= \mu_{i,m+1} + e_{it}, \quad t = k_m^0 + 1, \ldots, T.
\end{aligned}
$$

There are two estimating procedures: simultaneous approach and one-at-a-time approach. The former estimates all the break points simultaneously. This method is discussed in Bai and Perron (1998, 2003). Here we elaborate the one-at-a-time approach of Bai (1997a). This method is computationally simple. A researcher with a least squares routine is able to estimate multiple breaks one by one. The number of least squares required is a linear function of the sample size T.

Suppose that the number of breaks, m ($m > 1$), is given. Despite multiple breaks, the one-at-a-time approach proceeds as if there were just a single break. Thus the objective function is identical to the $SSR(k)$ defined in Section 3. Let $\hat{k}_1$ be the point that minimizes $SSR(k)$. This $\hat{k}_1$ is not necessarily estimating k_1^0. However, it is consistent for one of true break points. Which one it is being estimated depends on which true break point gives the largest reduction in the sum of squared residuals. This in turn depends on the magnitudes of breaks and regime spans, see Bai (1997a). Once the first break is obtained, we split the sample at the estimated break point, resulting in two subsamples. We then estimate a single break point in each of the subsamples, but only one of them is retained as our second estimator. The one that gives a larger reduction in the sum of squared residuals is kept. Denote the resulting estimator as $\hat{k}_2$. If $\hat{k}_2 < \hat{k}_1$ (the retained estimator is from the first subsample) we relabel them (switching the subscripts) so that $\hat{k}_1 < \hat{k}_2$. If $m = 2$, the procedure is stopped, and it can be shown that $\hat{k}_j$ is consistent for k_j^0 ($j = 1, 2$). If $m > 2$, we need to estimate the third break point from the resulting three subsamples separated by the break points $(\hat{k}_1, \hat{k}_2)$. Again, a single break point is estimated for each of the subsamples, and the one that achieves the largest reduction in the sum of squared residuals is retained as our estimate for the third break point. The procedure is repeated until m breaks are obtained.

In practice, the number of breaks, m, is unknown. In the above one-at-a-time approach, a test for existence of break point can be applied to each subsample before estimating a break point. Alternatively, one may use information criteria such as AIC and BIC to determine the number of breaks, see Bai and Perron (2003) and the references therein. However, for panel data, the corresponding criteria are not well understood. Our preliminary analysis shows that, under fixed T, the AIC criterion is consistent and the BIC is not, contrary to conventional wisdom. The latter is consistent if T is large. Further investigation is called for.

7. Concluding remarks

This paper develops some statistical theory for common breaks in panel data. Some unique properties for the estimated break point are derived. For example, a break point can be consistently estimated even when a regime contains a single observation. This property is appealing when the objective is to locate as quickly as possible the onset of a new regime or the turning point, without the need of waiting for many observations from the new regime. Asymptotic theory is developed as $N \to \infty$ with T either fixed or growing to infinity. We also propose a new framework for developing the limiting distribution for the estimated break point and we show how to use the limiting distribution to construct confidence intervals. Both the least squares method and the quasi-maximum likelihood method (QML) are studied. The QML method can consistently estimate the break points occurring in either means or variances. In general, QML is more efficient than the least squares method even in the absence of breaks in variances.

Acknowledgements

The author thanks a co-editor (Qi Li) and three anonymous referees for their constructive comments. This paper was presented at the International Conference on Breaks and Persistence in Econometrics, London, 2006. Partial financial support from the NSF (grant number SES-0551275) is acknowledged.

Appendix

We first establish a number of preliminary results, which are concerned with properties of the sum of squared residuals as k varies in $[1, T]$. Define

$$U_{NT}(k) = \frac{1}{NT} SSR(k) = \frac{1}{NT} \sum_{i=1}^{N} S_{iT}(k).$$

Lemma A.1. *For bounded or unbounded T,*

$$\sup_{1 \le k \le T} |U_{NT}(k) - EU_{NT}(k)| = O_p\left(\frac{1}{\sqrt{NT}}\right).$$

This lemma says that the deviation of $U_{NT}(k)$ from its expected value is small uniformly in k. We next argue that the expected value of $U_{NT}(k)$ attains its unique minimum at k_0.

Proof of Lemma A.1. Due to symmetry, it is sufficient to consider the case of $k \le k_0$. For $k \le k_0$,

$$\bar{Y}_{i1} = \mu_{i1} + \frac{1}{k} \sum_{t=1}^{k} e_{it},$$

$$
\begin{aligned}
\bar{Y}_{i2} &= \mu_{i1} + \frac{T - k_0}{T - k}(\mu_{i2} - \mu_{i1}) + \frac{1}{T - k} \sum_{t=k+1}^{T} e_{it} \\
&= \frac{k_0 - k}{T - k}(\mu_{i1} - \mu_{i2}) + \mu_{i2} + \frac{1}{T - k} \sum_{t=k+1}^{T} e_{it}.
\end{aligned}
$$

Each of the two alternative expressions for $\bar{Y}_{i2}$ will be used later. Introduce

$$\bar{e}_{i1} = \frac{1}{k} \sum_{t=1}^{k} e_{it}, \qquad \bar{e}_{i2} = \frac{1}{T - k} \sum_{t=k+1}^{T} e_{it}$$

their dependence on k is suppressed. Furthermore, let

$$a_{ik} = \frac{T - k_0}{T - k}(\mu_{i2} - \mu_{i1}), \qquad b_{ik} = \frac{k_0 - k}{T - k}(\mu_{i1} - \mu_{i2}).$$

It follows that

$$
\begin{aligned}
\bar{Y}_{i1} &= \mu_{i1} + \bar{e}_{i1} \\
\bar{Y}_{i2} &= \mu_{i1} + a_{ik} + \bar{e}_{i2} \\
&= \mu_{i2} + b_{ik} + \bar{e}_{i2}.
\end{aligned}
$$

By the definition of $S_{iT}(k)$, we have

$$S_{iT}(k) = \sum_{t=1}^{k}(e_{it} - \bar{e}_{i1})^2 + \sum_{t=k+1}^{k_0}(e_{it} - a_{ik} - \bar{e}_{i2})^2 + \sum_{t=k_0+1}^{T}(e_{it} - b_{ik} - \bar{e}_{i2})^2.$$

Note that for $k = k_0$, the sum $\sum_{t=k+1}^{k_0}$ is defined to be zero. Expanding and combining terms, we have

$$S_{iT}(k) = \sum_{t=1}^{k}(e_{it} - \bar{e}_{i1})^2 + \sum_{t=k+1}^{T}(e_{it} - \bar{e}_{i2})^2 + (k_0 - k)a_{ik}^2 + (T - k_0)b_{ik}^2 - 2a_{ik}\sum_{t=k+1}^{k_0}(e_{it} - \bar{e}_{i2}) - 2b_{ik}\sum_{t=k_0+1}^{T}(e_{it} - \bar{e}_{i2}).$$

The first two expressions on the right can be written as

$$\sum_{t=1}^{k}(e_{it} - \bar{e}_{i1})^2 + \sum_{t=k+1}^{T}(e_{it} - \bar{e}_{i2})^2 = \sum_{t=1}^{T}e_{it}^2 - k\bar{e}_{i1}^2 - (T-k)\bar{e}_{i2}^2.$$

Combining terms and summing over i, we obtain

$$SSR(k) = (k_0 - k)\sum_{i=1}^{N}a_{ik}^2 + (T-k)\sum_{i=1}^{N}b_{ik}^2 + \sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2 - \sum_{i=1}^{N}k\bar{e}_{i1}^2 - \sum_{i=1}^{N}(T-k)\bar{e}_{i2}^2 - 2\sum_{i=1}^{N}a_{ik}\sum_{t=k+1}^{k_0}(e_{it} - \bar{e}_{i2}) - 2\sum_{i=1}^{N}b_{ik}\sum_{t=k_0+1}^{T}(e_{it} - \bar{e}_{i2}). \quad (20)$$

Diving $SSR(k)$ by NT, and then subtracting its expected value, we obtain

$$U_{NT}(k) - EU_{NT}(k) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(e_{it}^2 - Ee_{it}^2\right) - \frac{1}{NT}\sum_{i=1}^{N}k[\bar{e}_{i1}^2 - E\bar{e}_{i1}^2] - \frac{1}{NT}\sum_{i=1}^{N}(T-k)[\bar{e}_{i2}^2 - E\bar{e}_{i2}^2] - \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=k+1}^{k_0}a_{ik}(e_{it} - \bar{e}_{i2}) - \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=k_0+1}^{T}b_{ik}(e_{it} - \bar{e}_{i2}).$$

From $|a_{ik}| \le |\mu_{i2} - \mu_{i1}|$ and $|b_{ik}| \le (k_0/(T-k_0)|\mu_{i2} - \mu_{i1}|)$, and a_{ik} and b_{ik} are independent of e_{it}, the last two terms are each $O_P([NT]^{-1/2})$. The first term on the right hand side is $O_P([NT]^{-1/2})$ by assumptions on e_{it}. It remains to consider the middle terms. Now

$$(NT)^{-1}\sum_{i=1}^{N}k[\bar{e}_{i1}^2 - E(\bar{e}_{i1}^2)] = T^{-1}N^{-1/2}\eta_k$$

where

$$\eta_k = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big[\Big(\frac{1}{\sqrt{k}}\sum_{t=1}^{k}e_{it}\Big)^2 - E\Big(\frac{1}{\sqrt{k}}\sum_{t=1}^{k}e_{it}\Big)^2\Big].$$

The assumption on e_{it} implies that $E(\eta_k)^2 \le M$ for all k. Thus $\max_{1\le k\le T}|\eta_k|$ is bounded by $O_p(T^{1/2})$. This implies that $N^{-1/2}T^{-1}\eta_k$ is bounded by $O_p([NT]^{-1/2})$ uniformly in k. Similarly, $(NT)^{-1}\sum_{i=1}^{N}(T-k)[\bar{e}_{i2}^2 - E\bar{e}_{i2}^2] = O_p([NT]^{-1/2})$ uniformly in k. This completes the proof of Lemma A.1. □

Lemma A.2. *For all $k \in [1, T]$, the expected value of $U_{NT}(k)$ satisfies*

$$EU_{NT}(k) - EU_{NT}(k_0) \ge \lambda_N C|k - k_0|/(NT)$$

where $\lambda_N = \sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2$ and for some $C > 0$.

Proof of Lemma A.2. Again, by symmetry, it is sufficient to consider $k \le k_0$. Note $k\bar{e}_{i1}^2 = k^{-1}(\sum_{t=1}^{k}e_{it})^2$ and $(T-k)\bar{e}_{i2}^2 = (\sum_{t=k+1}^{T}e_{it})^2/(T-k)$. In addition, $a_{ik} = b_{ik} = 0$ for $k = k_0$. Thus from (20), we have

$$U_{NT}(k) - U_{NT}(k_0) = (k_0 - k)\frac{1}{NT}\sum_{i=1}^{N}a_{ik}^2 + (T-k_0)\frac{1}{NT}\sum_{i=1}^{N}b_{ik}^2 - \frac{1}{NT}\sum_{i=1}^{N}\left[\frac{1}{k}\Big(\sum_{t=1}^{k}e_{it}\Big)^2 - \frac{1}{k_0}\Big(\sum_{t=1}^{k_0}e_{it}\Big)^2\right] - \frac{1}{NT}\sum_{i=1}^{N}\left[\frac{1}{T-k}\Big(\sum_{t=k+1}^{T}e_{it}\Big)^2 - \frac{1}{T-k_0}\Big(\sum_{t=k_0+1}^{T}e_{it}\Big)^2\right] - \frac{2}{NT}\sum_{i=1}^{N}a_{ik}\sum_{t=k+1}^{k_0}e_{it} - \frac{2}{NT}\sum_{i=1}^{N}b_{ik}\sum_{t=k_0+1}^{T}e_{it} + \frac{2}{NT}\sum_{i=1}^{N}[(k_0-k)a_{ik} + (T-k_0)b_{ik}]\bar{e}_{i2}.$$

All terms on the right hand side have zero mean except the first two terms, which are nonnegative. Thus

$$E[U_{NT}(k)] - E[U_{NT}(k_0)] \ge (k_0 - k)\frac{1}{NT}\sum_{i=1}^{N}a_{ik}^2 = (k_0 - k)\left(\frac{T-k_0}{T-k}\right)^2\frac{1}{NT}\sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2.$$

For the case of $k > k_0$, the above inequality holds with $(T-k_0)/(T-k)$ replaced by k_0/k. Note that $(T-k_0)/(T-k) \ge (1-k_0/T) = 1-\tau_0$ for $k \le k_0$ and $k_0/k \ge k_0/T = \tau_0$ for $k > k_0$, Thus, the lemma is proved for $C = \min[(1-\tau_0)^2, \tau_0^2]$. □

Remark 3. When T is fixed, no restriction on k_0 is imposed so C can take on the minimum value $1/T$ when $k_0 = 1$ or $k_0 = T - 1$. Since T is fixed, this minimum value is still a fixed positive number (limits are taken as $N \to \infty$ only). For the case of $T \to \infty$, in view of $k_0 = [T\tau_0]$ or simply $k_0 = T\tau_0$ (to avoid the niceties), $C = \min[(1-\tau_0)^2, \tau_0^2]$ is a fixed number. In either case, C can be treated as a fixed constant. Even for the case $T \to \infty$, we can remove the requirement of $k_0 = T\tau_0$ so that a regime can have a single observation, and C can take on the value $1/T$. In this case, we would require condition $N^{-1}\sum_{i=1}^{N}(\mu_{i1} - \mu_{i2})^2 \ge c > 0$ instead of (2) to obtain $|\hat{k} - k_0| \to 0$. This simply follows from (21) with $C = 1/T$, we have

$$|\hat{k} - k_0| \le \left(\frac{T}{N}\right)^{3/2}\left[\frac{1}{N}\sum_{i=1}^{N}(\mu_{i2} - \mu_{i1}^2)\right]^{-1}O_p(1) \to 0$$

if $T/N \to 0$. In summary, even if $T \to \infty$, we can still allow a regime to contain a single observation provided that N is much larger than T and $N^{-1}\sum_{i=1}^{N}(\mu_{i1} - \mu_{i2})^2 \ge c > 0$.

By Lemma A.1, $U_{NT}(k)$ is uniformly closed to its expected value, and by Lemma A.2, the expected value has a unique minimum at k_0, it is thus reasonable to expect that $U_{NT}(k)$ achieves its minimum close to k_0. This is the essence for the argument of Theorem 3.1. We now provide a formal proof of the theorem.

Proof of Theorem 3.1. First note that $\hat{k} = \text{argmin}_k SSR(k) = \text{argmin}_k U_{NT}(k)$. Adding and subtracting,

$$\begin{aligned}U_{NT}(k) - U_{NT}(k_0) &= U_{NT}(k) - EU_{NT}(k) - [U_{NT}(k_0) - EU_{NT}(k_0)] \\ &\quad + EU_{NT}(k) - EU_{NT}(k_0) \\ &\geq -2 \sup_{1\leq j\leq T} |U_{NT}(j) - EU_{NT}(j)| + EU_{NT}(k) - EU_{NT}(k_0) \\ &\geq -2 \sup_{1\leq j\leq T} |U_{NT}(j) - EU_{NT}(j)| + \lambda_N C|k-k_0|/(NT).\end{aligned}$$

The last inequality follows from Lemma A.2. The above inequality holds for each $k \in [1, T]$ and particularly, it holds for $\hat{k}$. Noting that $U_N(\hat{k}) - U_{NT}(k_0) \leq 0$, we have

$$\begin{aligned}|\hat{k} - k_0| &\leq 2\lambda_N^{-1} C^{-1}(NT) \sup_{1\leq j\leq T} |U_{NT}(j) - EU_{NT}(j)| \\ &= 2C^{-1}(NT)^{1/2}\Big[\sum_{i=1}^N (\mu_{i2}-\mu_{i1})^2\Big]^{-1} O_p(1) \qquad (21)\end{aligned}$$

the last expression follows from Lemma A.1.

Case 1: *T is fixed*. For this case, $|\hat{k} - k_0| \xrightarrow{p} 0$ follows immediately from (21) and $N^{1/2}[\sum_{i=1}^N (\mu_{i2} - \mu_{i1}^2)]^{-1} \to 0$, as $N \to \infty$ by assumption (2).

Case 2: *T is unbounded.* From (21),

$$|\hat{k} - k_0| = o_p(\sqrt{T}).$$

It follows that for any given $\delta > 0$, $|\hat{k} - k_0| \leq \delta T$ for all large T, with probability tending to 1. Because $k_0 = [T\tau_0]$ and $0 < \tau_0 < 1$, there exists an $\eta > 0$ such that

$$P(\hat{k} \in [\eta T, (1-\eta)T]) \to 1$$

as T tends to infinity. That is

$$P(\hat{k} \in D) \to 1$$

where $D = \{k;\ \eta T \leq k \leq (1-\eta)T\}$.

Define the set $D(k_0) = D \setminus \{k_0\}$ so that $D(k_0)$ excludes k_0 from D. Then

$$\begin{aligned}P(\hat{k} \neq k_0) &= P(\hat{k} \notin D) + P(\hat{k} \in D, \hat{k} \neq k_0) \\ &= P(\hat{k} \notin D) + P(\hat{k} \in D(k_0)).\end{aligned}$$

By definition, $U_{NT}(\hat{k}) \leq U_{NT}(k_0)$. So a necessary condition for $\hat{k} \in D(k_0)$ is $\min_{k\in D(k_0)} U_{NT}(k) - U_{NT}(k_0) \leq 0$. But this event has a probability tending to zero by Lemma A.3, which implies that $P(\hat{k} \in D(k_0)) \to 0$. Thus

$$P(\hat{k} \neq k_0) = P(\hat{k} \notin D) + P(\hat{k} \in D(k_0)) \to 0.$$

This completes the proof of Theorem 3.1. □

Lemma A.3. *Let $D(k_0)$ be as defined in the proof of Theorem 3.1, as $N, T \to \infty$,*

$$P\Big(\min_{k\in D(k_0)} U_{NT}(k) - U_{NT}(k_0) \leq 0\Big) \to 0.$$

This lemma says $U_{NT}(k)$ cannot achieve its minimum when $k \neq k_0$.

Proof of Lemma A.3. From $U_{NT}(k) = SSR(k)/(NT)$, the lemma is equivalent to

$$P\Big(\min_{k\in D, k\neq k_0}\Big[SSR(k) - SSR(k_0)\Big] > 0\Big) \to 1.$$

It is again sufficient to consider $k \leq k_0$ due to symmetry. Multiplying NT on each side of $U_{NT}(k) - U_{NT}(k_0)$ (see the proof of the previous lemma), we have

$$\begin{aligned}SSR(k) - SSR(k_0) &= (k_0 - k)\sum_{i=1}^N a_{ik}^2 + (T-k_0)\sum_{i=1}^N b_{ik}^2 \\ &\quad - \sum_{i=1}^N \left[\frac{1}{k}\Big(\sum_{t=1}^k e_{it}\Big)^2 - \frac{1}{k_0}\Big(\sum_{t=1}^{k_0} e_{it}\Big)^2\right] \\ &\quad - \sum_{i=1}^N \left[\frac{1}{T-k}\Big(\sum_{t=k+1}^T e_{it}\Big)^2 - \frac{1}{T-k_0}\Big(\sum_{t=k_0+1}^T e_{it}\Big)^2\right] \\ &\quad - 2\sum_{i=1}^N a_{ik}\sum_{t=k+1}^{k_0} e_{it} - 2\sum_{i=1}^N b_{ik}\sum_{t=k_0+1}^T e_{it} \\ &\quad + 2\sum_{i=1}^N [(k_0-k)a_{ik} + (T-k_0)b_{ik}]\bar{e}_{i2}. \qquad (22)\end{aligned}$$

Consider each term on the right hand side. For the first term,

$$(k_0 - k)\sum_{i=1}^N a_{ik}^2 = (k_0-k)\Big(\frac{T-k_0}{T-k}\Big)^2 \sum_{i=1}^N (\mu_{i2}-\mu_{i1})^2,$$

and $(T-k_0)/(T-k) \geq (T-k_0)/T = 1 - \tau_0 > 0$ for $k \leq k_0$. Thus the first term is no smaller than $(k_0 - k)(1-\tau_0)^2\sqrt{N}d_N$, where $d_N = N^{-1/2}\sum_{i=1}^N (\mu_{i2}-\mu_{i1})^2$. By (2), $d_N \to \infty$. We next show all other terms are either nonnegative and can be ignored or dominated by the first term with the largest term being of $(k_0 - k)\sqrt{N}O_p(1)$ so that $SSR(k) - SSR(k_0) \to +\infty$.

The second term on the right is nonnegative and can be ignored. For $k \in D$, k satisfies $k \geq \eta T$. This is, k must be a positive fraction of the sample size. Therefore, $\frac{1}{k}(\sum_{t=1}^k e_{it})^2 = O_p(1)$ uniformly in $k \in D$ by the invariance principle. From $E\frac{1}{k}(\sum_{t=1}^k e_{it})^2 = \sigma_i^2$ for all k, adding and subtracting terms, the third term can be rewritten as

$$\sqrt{N}\Big\{N^{-1/2}\sum_{i=1}^N\Big[\frac{1}{k}\Big(\sum_{t=1}^k e_{it}\Big)^2 - \sigma_i^2\Big] - N^{-1/2} \times \sum_{i=1}^N \Big[\frac{1}{k_0}\Big(\sum_{t=1}^{k_0} e_{it}\Big)^2 - \sigma_i^2\Big]\Big\}.$$

The two expressions inside the braces are each $O_p(1)$. Thus the third term is of $\sqrt{N}O_p(1)$. Similarly, the fourth term on the right of hand side is $\sqrt{N}O_p(1)$. Next,

$$2\sum_{i=1}^N a_{ik}\sum_{t=k+1}^{k_0} e_{it} = 2(k_0-k)\sqrt{N}\Big[\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\Big(N^{-1/2}\sum_{i=1}^N e_{it}\Big)\Big]$$

which is $(k_0 - k)\sqrt{N}O_p(1)$, where $O_p(1)$ is uniform in $k \leq k_0 - 1$. The sixth term

$$2\sum_{i=1}^N b_{ik}\sum_{t=k_0+1}^T e_{it} = 2(k_0-k)\sqrt{N}\frac{T-k_0}{T-k} \times \left[\frac{1}{T-k_0}\frac{1}{\sqrt{N}}\sum_{t=k_0+1}^T\sum_{i=1}^N (u_{i2}-\mu_{i1})e_{it}\right]$$

which is $(k_0 - k)\sqrt{N}O_p(T^{-1/2})$ in view of $(T-k_0)/(T-k) \leq 1$ for $k \leq k_0$. The last term is equal to, noting that $(k_0 - k)a_{ik} = (T-k_0)b_{ik}$,

$$4\sum_{i=1}^N (k_0-k)a_{ik}\bar{e}_{i2} = 4(k_0-k)\sqrt{N}\frac{T-k_0}{T-k} \times \left[\frac{1}{T-k}\frac{1}{\sqrt{N}}\sum_{t=k+1}^T\sum_{i=1}^N (\mu_{i2}-\mu_{i1})e_{it}\right]$$

which is $(k_0-k)\sqrt{N}O_p(T^{-1/2})$ uniformly in $k \le k_0$. In summary, all terms are dominated by the first term, which diverges to positive infinity for $k \ne k_0$. This completes the proof of Lemma A.3. □

Lemma A.4. *Under assumptions of Theorem 3.2, Lemma A.3 still holds. That is,*

$$P\Big(\min_{k\in D(k_0)} U_{NT}(k) - U_{NT}(k_0) \le 0\Big) \to 0,$$

where $D(k_0)$ is defined in the proof of Theorem 3.1.

Proof of Lemma A.4. The proof is similar to but more delicate than that of Lemma A.3. The first term on the right hand side of (22) is

$$(k_0-k)\left(\frac{T-k_0}{T-k}\right)^2 \sum_{i=1}^N (\mu_{i2}-\mu_{i1})^2 \ge (k_0-k)(1-\tau_0)^2\lambda_N \quad (23)$$

where $\lambda_N = \sum_{i=1}^N (\mu_{i1}-\mu_{i2})^2 \to \infty$ by assumption. We next argue that all other terms are either dominated by the first term or are nonnegative and can be ignored. The second term is nonnegative and is thus ignored. For the third term, rewrite

$$\frac{1}{k}\Big(\sum_{t=1}^k e_{it}\Big)^2 - \frac{1}{k_0}\Big(\sum_{t=1}^{k_0} e_{it}\Big)^2$$
$$= \frac{k_0-k}{k_0}\frac{1}{k}\Big(\sum_{t=1}^k e_{it}\Big)^2 - 2\frac{1}{k_0}\Big(\sum_{t=1}^k e_{it}\Big)\Big(\sum_{t=k+1}^{k_0} e_{it}\Big) - \frac{k_0-k}{k_0}\frac{1}{k_0-k}\Big(\sum_{t=k+1}^{k_0} e_{it}\Big)^2.$$

Summing over i, we have

$$\sum_{i=1}^N \left[\frac{1}{k}\left(\sum_{t=1}^k e_{it}\right)^2 - \frac{1}{k_0}\left(\sum_{t=1}^{k_0} e_{it}\right)^2\right]$$
$$= (k_0-k)\left\{\frac{N}{k_0}\frac{1}{N}\sum_{i=1}^N\left(\frac{1}{\sqrt{k}}\sum_{t=1}^k e_{it}\right)^2 - 2\frac{\sqrt{N}}{\sqrt{k_0}}\frac{1}{\sqrt{N}}\sum_{i=1}^N\left[\left(\frac{1}{\sqrt{k_0}}\sum_{t=1}^k e_{it}\right)\left(\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} e_{it}\right)\right] - \frac{N}{k_0}\frac{1}{N}\sum_{i=1}^N\left(\frac{1}{\sqrt{k_0-k}}\sum_{t=k+1}^{k_0} e_{it}\right)^2\right\}. \quad (24)$$

We shall argue each expression inside the braces is $o_p(1)$. For $k \in D$, $k \ge T\eta$ and so the first expression has uniformly bounded (in k) summands, thus it can be written as $(N/k_0)O_p(1) = o_p(1)$ because $N/[T\tau_0] \to 0$. For the second expression in the braces, the summand is uniformly bounded in k, and it has zero mean if e_{it} are uncorrelated over t. For serially correlated e_{it}, let $\xi_{it,k} = \sum_{t=1}^k e_{it}\sum_{t=k+1}^{k_0} e_{it}$. Lemma 11 of Bai (1997a) shows that $\gamma_{ik} = E(\xi_{it,k})$ are uniformly bounded in k. Also, γ_{ik} are uniformly bounded in i by Assumption 1. Thus, there exists γ such that $|\gamma_{ik}| \le \gamma$. Adding and subtracting terms,

$$N^{-1/2}\sum_{i=1}^N \frac{1}{\sqrt{k_0}}\frac{1}{k_0-k}\xi_{it,k} = N^{-1/2}\sum_{i=1}^N \frac{1}{\sqrt{k_0}}\frac{1}{k_0-k}(\xi_{it,k}-\gamma_{ik}) + N^{-1/2}\sum_{i=1}^N \frac{1}{\sqrt{k_0}}\frac{1}{k_0-k}\gamma_{ik}.$$

The first term on the right hand side is $O_p(1)$ because the summands are all $O_p(1)$ random variables and have zero mean. The second term is bounded by $\gamma\sqrt{N/k_0}$ because $|\gamma_{ik}| \le \gamma$ and $\frac{1}{k_0-k} \le 1$ for $k_0 \ne k$. In summary, the second expression inside the braces of (23) is $O_p(\sqrt{N/k_0}) + O_p(N/k_0)$, which is $o_p(1)$.

For the last expression inside the braces, the summand is uniformly bounded by $2\log(\log(T))$ (as k varies from 1 to k_0) due to the law of iterated logarithm, so the last expression is $O_p(\log(\log(T)))N/T\tau_0 = o_p(1)$ by (5). Therefore (24) is $(k_0-k)o_p(1)$, and thus is dominated by the first term in (23). Similarly, the fourth term is $(k_0-k)o_p(1)$. Next consider the fifth term.

$$\sum_{i=1}^N a_{ik}\sum_{t=k+1}^{k_0} e_{it} = (k_0-k)\left(\frac{T-k_0}{T-k}\right)\frac{1}{k_0-k}\sum_{t=k+1}^{k_0} \times \left[\sum_{i=1}^N(\mu_{i2}-\mu_{i1})e_{it}\right]. \quad (25)$$

Note $(T-k_0)/(T-k) \le 1$. Denote $\eta_t = \sum_{i=1}^N (\mu_{i2}-\mu_{i1})e_{it}$, by the Hajek–Renyi inequality (see Bai, 1994), for some $C < \infty$, and for any $\delta > 0$,

$$P\Big(\sup_{k\le k_0-1}\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}\eta_t > \delta\lambda_N\Big) < \frac{Var(\eta_t)}{\delta^2\lambda_N^2}C$$
$$= \frac{\sum_{i=1}^N(\mu_{i2}-\mu_{i1})^2\sigma_i^2}{\delta\lambda_N^2}C \le \frac{\bar\sigma^2\lambda_N}{\delta^2\lambda_N^2}C = \frac{\bar\sigma^2}{\delta^2\lambda_N}C \to 0 \quad (26)$$

because $\lambda_N \to \infty$ by assumption (4), where $\bar\sigma^2 = \max_i \sigma_i^2 < \infty$. Thus the fifth term is dominated by $(k_0-k)\delta\lambda_N$ for arbitrary small δ, thus is dominated by the first term, see (23).

The remaining three terms are all smaller magnitude. They are in the order of $(k_0-k)\lambda_N O_p(T^{-1/2})$, thus dominated by the first term. This completes the proof. □

Proof of Theorem 3.2. From (21), we have

$$|\hat k - k_0| \le 2C^{-1}(NT)^{1/2}\Big[\sum_{i=1}^N(\mu_{i2}-\mu_{i1})^2\Big]^{-1}O_p(1).$$

Divide T on each side,

$$\left|\frac{\hat k}{T}-\frac{k_0}{T}\right| \le 2C^{-1}\left(\frac{N}{T}\right)^{1/2}\Big[\sum_{i=1}^N(\mu_{i2}-\mu_{i1})^2\Big]^{-1}O_p(1) \xrightarrow{p} 0 \quad (27)$$

either because $(N/T) \to 0$ by assumption (5) or because $\lambda_N = \sum_{i=1}^N(\mu_{i1}-\mu_{i2})^2$ goes to infinity by assumption (4). Thus for the same set D defined in the proof of Theorem 3.1, we have

$$P(\hat k \in D) \to 1.$$

The rest of proof is the same as that of Theorem 3.1, except that we invoke Lemma A.4 instead of Lemma A.3. □

Proof of Lemma 4.1. The proof is similar to the that of Theorem 3.2. By (27), $|\hat\tau - \tau_0| \xrightarrow{p} 0$, thus for the same set D as in the proof of Theorem 3.2,

$$P(\hat k \in D) \to 1.$$

Lemma 4.1 is equivalent to the statement that for every $\epsilon > 0$, there exists $M > 0$ such that

$$P(|\hat k - k_0| > M) < \epsilon.$$

Thus it is sufficient to show for any $\delta > 0$, there exists an $M < \infty$ such that

$$P\Big(\min_{k\in D, |k-k_0|>M} SSR(k) - SSR(k_0) \le 0\Big) < \epsilon$$

for all large N, T satisfying (5). We show this by showing $SSR(k) - SSR(k_0) > 0$ for $k \in D$ and $|k - k_0| > M$. Again, we focus on the case of $k < k_0$. When $k < k_0$, $SSR(k) - SSR(k_0)$ is given by (22). Similar to the previous lemmas, we show that the first term on the right hand side of (22) dominates all other terms for $k < k_0 - M$ and for large M. The first term is no smaller than, see (23)

$$(k_0 - k)(1 - \tau_0)^2\lambda_N$$

which is $O(1)$ by not $o(1)$ because $\lambda_N \to \lambda > 0$ by assumption. The second term is nonnegative and can be ignored. The third and the fourth term are shown to be $(k_0-k)o_p(1)$ in the proof of Lemma A.4 and are thus dominated by the first term. Consider the fifth term, which is given by (25). Using similar argument as in (26), for any given small $\delta > 0$, by the Hajek and Renyi inequality, for some $C < \infty$,

$$P\Big(\sup_{k\le k_0-M} \frac{1}{k_0-k}\sum_{t=k+1}^{k_0} \eta_t > \delta\lambda_N\Big) < \frac{Var(\eta_t)}{M\delta^2\lambda_N^2}C$$
$$= \frac{\sum_{i=1}^{N}(u_{i2}-\mu_{i1})^2\sigma_i^2}{M\delta\lambda_N^2}C \le \frac{\bar\sigma^2\lambda_N}{M\delta^2\lambda_N^2}C = \frac{\bar\sigma^2}{M\delta^2\lambda_N}C < \epsilon$$

as long as M is large. Thus the fifth term is dominated by $(k_0 - k)\delta\lambda_N$, thus dominated by the first term. The remaining terms are all $(k_0-k)\lambda_N O_p(T^{-1/2})$ and thus dominated by the first term. Since the first term is positive and dominates all other terms, $SSR(k) - SSR(k_0)$ is positive. This complete the proof of Lemma 4.1. □

Proof of Theorem 4.2. We examine the behavior of $SSR(k) - SSR(k_0)$ for $|k - k_0|$ on bounded set because Lemma 4.1 shows that $|\hat k - k_0|$ is $O_p(1)$. For the terms on the right hand side of (22), The previous lemma shows that the first and the fifth terms are $(k_0 - k)O_p(1)$ and all other terms are $(k_0 - k)o_p(1)$. Now the first term is

$$(k_0 - k)\sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2$$

plus an negligible term because $(T - k_0)/(T - k) \to 1$ for $k - k_0 = O_p(1)$. Similarly, the fifth term is equal to the following plus an negligible term

$$-2\sum_{t=k+1}^{k_0}\Big[\sum_{i=1}^{N}(\mu_{i2}-\mu_{i1})e_{it}\Big] = -2\sum_{t=k+1}^{k_0} N^{-1/2}\sum_{i=1}^{N}\Delta_i e_{it}$$
$$= -2\sqrt{\phi_N}\sum_{t=k+1}^{k_0}\Big(N^{-1/2}\sum_{i=1}^{N} w_i e_{it}\Big)$$

where

$$w_i = \frac{\Delta_i}{\sqrt{\phi_N}} = \frac{\Delta_i}{\left(\frac{1}{N}\sum_{j=1}^{N}\Delta_j^2\sigma_j^2\right)^{1/2}}.$$

By assumption (9), $\phi_N = \frac{1}{N}\sum_{j=1}^{N}\Delta_j^2\sigma_j^2 \to \phi$, and $N^{-1/2}\sum_{i=1}^{N} w_i e_{it} \to Z_t$, where Z_t is $N(0, 1)$. This follows from the central limit theorem as $N \to \infty$ and $Var(N^{-1/2}\sum_{i=1}^{N} w_i e_{it}) = 1$. Thus the limit of the fifth term is $2\phi^{1/2}\sum_{t=k+1}^{k_0} Z_t$ (note Z_t and $-Z_t$ have the same distribution). In summary, for $k \le k_0$,

$$SSR(k) - SSR(k_0) \xrightarrow{d} (k_0 - k)\lambda + 2\phi^{1/2}\sum_{t=k+1}^{k_0} Z_t.$$

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

$$SSR(k) - SSR(k_0) \xrightarrow{d} (k - k_0)\lambda + 2\phi^{1/2}\sum_{t=k_0+1}^{k} Z_t.$$

This is equivalent to the statement of Theorem 4.2. □

Proof of (14). Consider the numerator of A_N. We shall show

$$\sum_{i=1}^{N}(\hat\mu_{i2} - \hat\mu_{i1})^2 = \sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2 + o_p(1). \quad (28)$$

From $\hat k = k_0 + O_p(1)$, it is easy to show that the asymptotic behavior of $\hat\mu_{i2} - \hat\mu_{i1}$ is not affected by assuming $\hat k$ is equal to k_0. With $\hat k$ being k_0, we have

$$\hat\mu_{i2} - \hat\mu_{i1} = \mu_{i2} - \mu_{i1} + \frac{1}{T-k_0}\sum_{t=k_0+1}^{T} e_{it} - \frac{1}{k_0}\sum_{t=1}^{k_0} e_{it}.$$

Thus

$$\sum_{i=1}^{N}(\hat\mu_{i2} - \hat\mu_{i1})^2 = \sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2 + \sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})\xi_i + \sum_{i=1}^{N}\xi_i^2$$

where $\xi_i = \frac{1}{T-k_0}\sum_{t=k_0+1}^{T} e_{it} - \frac{1}{k_0}\sum_{t=1}^{k_0} e_{it}$. From $k_0 = [T\tau_0]$, we have $\xi_i = O_p(T^{-1/2})$ and $\xi_i^2 = O_p(T^{-1})$ for each i. It follows that $\sum_{i=1}^{N}\xi_i^2 = (N/T)\frac{1}{N}\sum_{i=1}^{N}(T\xi_i^2) = O_p(N/T) = o_p(1)$ by condition (5). Next, by (8)

$$\sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})\xi_i = T^{-1/2}\frac{1}{\sqrt N}\sum_{i=1}^{N}\Delta_i(\sqrt{T}\xi_i)$$
$$= O_p(T^{-1/2}) = o_p(1).$$

This proves (28). Similarly, for the denominator of A_N, we can show

$$\sum_{i=1}^{N}(\hat\mu_{i2} - \hat\mu_{i1})^2\hat\sigma_i^2 = \sum_{i=1}^{N}(\mu_{i2} - \mu_{i1})^2\sigma_i^2 + o_p(1).$$

Combining results, we obtain $\hat A_N = A_N + o_p(1)$. □

Proof of Theorem 5.1. Subtract from the objective function (16) of the following term

$$k_0\sum_{i=1}^{N}\log\sigma_{i1}^2 + (T - k_0)\sum_{i=1}^{N}\log\sigma_{i2}^2$$

we obtain a "centered" objective function

$$k\sum_{i=1}^{N}\log\hat\sigma_{i1}^2(k) + (T-k)\sum_{i=1}^{N}\log\hat\sigma_{i2}^2(k)$$
$$-k_0\sum_{i=1}^{N}\log\sigma_{i1}^2 - (T-k_0)\sum_{i=1}^{N}\log\sigma_{i2}^2. \quad (29)$$

This does not alter $\hat k$ since the subtracted term does not depend on k. When context is clear, we simply write $\hat\sigma_{i1}^2$ for $\hat\sigma_{i1}^2(k)$ and $\hat\sigma_{i2}^2$ for $\hat\sigma_{i2}^2(k)$. From

$$k\log\hat\sigma_{i1}^2 - k_0\log\sigma_{i1}^2 = k\log\left(\frac{\hat\sigma_{i1}^2}{\sigma_{i1}^2}\right) + (k - k_0)\log\sigma_{i1}^2, \quad \text{and}$$

$$(T-k)\log\hat{\sigma}_{i2}^2-(T-k_0)\log\sigma_{i2}^2=(T-k)\log\left(\frac{\hat{\sigma}_{i2}^2}{\sigma_{i2}^2}\right)$$
$$-(k-k_0)\log\sigma_{i2}^2$$

the centered objective function can be rewritten as

$$\sum_{i=1}^N\left[k\log\left(\frac{\hat{\sigma}_{i1}^2}{\sigma_{i1}^2}\right)+(T-k)\log\left(\frac{\hat{\sigma}_{i2}^2}{\sigma_{i2}^2}\right)-(k_0-k)\log\left(\frac{\sigma_{i1}^2}{\sigma_{i2}^2}\right)\right].$$

For any given N, it is known that $\hat{k}=k_0+O_p(1)$, see Bai et al. (1998) and Bai (2000). This implies that $\hat{\sigma}_{i1}^2=\sigma_{i1}^2+O_p(T^{-1/2})$. From $\log(1+x)=x+O(x^2)$ for $x=o(1)$, we have

$$\log\left(\frac{\hat{\sigma}_{i1}^2}{\sigma_{i1}^2}\right)=\frac{\hat{\sigma}_{i1}^2-\sigma_{i1}^2}{\sigma_{i1}^2}+O_p(T^{-1}).$$

The first term on the right hand side is $O_p(T^{-1/2})$, thereby dominates the last term. Thus we can ignore the $O_p(T^{-1})$ term and write $\log(\hat{\sigma}_{i1}^2/\sigma_{i1}^2)\approx\hat{\sigma}_{i1}^2/\sigma_{i1}^2-1$. Similarly $\log(\hat{\sigma}_{i2}^2/\sigma_{i2}^2)\approx\hat{\sigma}_{i2}^2/\sigma_{i2}^2-1$. Thus the objective function can be approximated by

$$\sum_{i=1}^N\left[k\hat{\sigma}_{i1}^2/\sigma_{i1}^2+(T-k)\hat{\sigma}_{i2}^2/\sigma_{i2}^2-T-(k_0-k)\log(\sigma_{i1}^2/\sigma_{i2}^2)\right].\quad(30)$$

It is noted in passing that if $\sigma_{i1}=\sigma_{i2}$ for all i, the last term is zero; the remaining term coincides with the generalized least squares objective function $\sum_{i=1}^N\sigma_i^{-2}S_{iT}(k)$. This means that QML is asymptotically equivalent to GLS in absence of breaks in variance.

Again by symmetry, we only consider the case of $k<k_0$, and show that the objective function cannot achieve its minimum for $k<k_0$. We first prove the theorem for the simple case in which there are no mean parameters thus no mean breaks. That is, we first consider the model $Y_{it}=\sigma_{i1}\eta_{it}$ for $t\le k_0$ and $Y_{2t}=\sigma_{2t}\eta_{it}$ for $t>k_0$. For $k<k_0$. Then $\hat{\sigma}_{i1}^2(k)=\sigma_{i1}^2\frac{1}{k}\sum_{t=1}^k\eta_{it}^2$, and $\hat{\sigma}_{i2}^2(k)=\frac{1}{T-k}\sigma_{i1}^2\sum_{t=k+1}^{k_0}\eta_{it}^2+\frac{1}{T-k}\sigma_{i2}^2\sum_{t=k_0+1}^T\eta_{it}^2$. It follows that

$$\sum_{i=1}^N\left[k\hat{\sigma}_{i1}^2/\sigma_{i1}^2+(T-k)\hat{\sigma}_{i2}^2/\sigma_{i2}^2\right]$$
$$=\sum_{i=1}^N\left[\sum_{t=1}^k\eta_{it}^2+\frac{\sigma_{i1}^2}{\sigma_{i2}^2}\sum_{t=k+1}^{k_0}\eta_{it}^2+\sum_{t=k_0+1}^T\eta_{it}^2\right].$$

Subtract from the objective function the term $\sum_{i=1}^N\sum_{t=1}^T\eta_{it}^2$ so that (30) becomes

$$V(k)=\sum_{i=1}^N d_i\sum_{t=k+1}^{k_0}(\eta_{it}^2-1)+(k_0-k)\sum_{i=1}^N f(1+d_i)\quad(31)$$

where $d_i=\sigma_{i1}^2/\sigma_{i2}^2-1$. Note that $f(1+d_i)=f(\sigma_{i1}^2/\sigma_{i2}^2)$ with $f(x)=x-1-\log(x)$. Clearly, $V(k_0)=0$. We shall show that for any $k<k_0$, the above objective function diverges to infinity as $N\to\infty$ so that its minimum cannot be achieved at a point other than k_0. Because the second term on the right hand side diverges to infinity by assumption, it is sufficient to show the first term is dominated by the second term. Divide $V(k)$ by k_0-k,

$$\frac{V(k)}{k_0-k}=\sum_{i=1}^N d_i\xi_{ik}+\sum_{i=1}^N f(1+d_i)=G(k),\quad\text{say}$$

where $\xi_{ik}=\frac{1}{k_0-k}\sum_{t=k+1}^{k_0}(\eta_{it}^2-1)$. The variables ξ_{ik} have zero mean and are independent over i. In addition, $var(\xi_{ik})\le M$ for all i and k, for some $M<\infty$. By assumption, $\sum_{i=1}^N f(1+d_i)\to\infty$. To show $G(k)\to\infty$ with probability tending to 1, it is sufficient to show that the first term of $G(k)$ is dominated by the second, i.e.,

$$\frac{\sum_{i=1}^N d_i\xi_{ik}}{\sum_{i=1}^N f(1+d_i)}\xrightarrow{p}0.$$

But the variance of above is bounded by

$$M\frac{\sum_{i=1}^N d_i^2}{\left[\sum_{i=1}^N f(1+d_i)\right]^2}.\quad(32)$$

It suffices to show this variance goes to zero. Consider the case in which d_i is small for all i so that $f(1+d_i)=\frac{1}{2}d_i^2+o(d_i^3)$. Write $a_N=\sum_{i=1}^N f(1+d_i)$. Thus

$$\frac{\sum_{i=1}^N d_i^2}{\left[\sum_{i=1}^N f(1+d_i)\right]^2}=\frac{\sum_{i=1}^N d_i^2}{a_N\sum_{i=1}^N[2^{-1}d_i^2+o(d_i^3)]}=\frac{1}{a_N}O(1)\to 0$$

because $a_N\to\infty$ by assumption. Next consider the case in which all d_i are large such that $|d_i|>c$. We shall assume d_i is bounded above such that $|d_i|\le C$ (this in fact is not necessary, but otherwise, technical niceties are needed and they do not provide any further insight). The numerator of (32) is $O(N)$, and the denominator is $O(N^2)$ because $f(1+d_i)>c_1$ for all i. It follows that (32) goes to zero. Next suppose that some d_i are small and some are large. Choose $c\in(0,1)$ such that for $|d_i|\le c$, the expansion $f(1+d_i)=2^{-1}d_i^2+o(d_i^3)\ge 3^{-1}d_i^2$ holds. Let $A=\{i;|d_i|\le c\}$ and $B=\{i;|d_i|>c\}$. Let N_2 be the number of elements in B. For $i\in B$, $f(1+d_i)>c_1$ for some $c_1>0$. Then

$$\frac{\sum_{i=1}^N d_i^2}{\left[\sum_{i=1}^N f(1+d_i)\right]^2}\le\frac{1}{a_N}\left[\frac{\sum_{i\in A}d_i^2+N_2C}{3^{-1}\sum_{i\in A}d_i^2+N_2c_1}\right]=\frac{1}{a_N}O(1)\to 0.$$

In the above, the term inside the brackets is $O(1)$. This proves the theorem for the case of no mean parameters.

Next consider the existence of breaks in both mean and variance parameters. From the definition of $\hat{\sigma}_{ik}^2(k)(k=1,2)$, the objective function (30) is equal to (ignore the term $-T$),

$$\sum_{i=1}^N\left[\frac{1}{\sigma_{i1}^2}\sum_{t=1}^k(Y_{it}-\bar{Y}_{i1})^2+\frac{1}{\sigma_{i2}^2}\sum_{t=k+1}^T(Y_{it}-\bar{Y}_{i2})^2\right.$$
$$\left.-(k_0-k)\log(\sigma_{i1}^2/\sigma_{i2}^2)\right].\quad(33)$$

Again, it is sufficient to consider the case of $k<k_0$ by symmetry. In view of $\hat{k}=k_0+O_p(1)$, we only need to consider those k such that $|k-k_0|$ is bounded. Together with $k_0=[T\tau_0]$, $\tau_0\in(0,1)$, we have $\bar{Y}_{i1}=u_{i1}+O_p(T^{-1/2})$ and $\bar{Y}_{i2}=u_{i2}+O_p(T^{-1/2})$. Thus

$$\begin{aligned}Y_{it}-\bar{Y}_{i1}&=Y_{it}-\mu_{i1}+(\mu_{i1}-\bar{Y}_{i1})=\sigma_{i1}\eta_{it}-(\bar{Y}_{i1}-\mu_{i1})\\&=\sigma_{i1}\eta_{it}+O_p(T^{-1/2}),\quad\text{for } t\le k;\end{aligned}$$

$$\begin{aligned}Y_{it}-\bar{Y}_{i2}&=Y_{it}-\mu_{i1}+(\mu_{i1}-\mu_{i2})-(\bar{Y}_{i2}-\mu_{i2})\\&=\sigma_{i1}\eta_{it}+(\mu_{i1}-\mu_{i2})-(\bar{Y}_{i2}-\mu_{i2})\\&=\sigma_{i1}\eta_{it}+(\mu_{i1}-\mu_{i2})+O_p(T^{-1/2}),\quad\text{for } k+1\le t\le k_0;\end{aligned}$$

$$\begin{aligned}Y_{it}-\bar{Y}_{i2}&=Y_{it}-\mu_{i2}-(\bar{Y}_{i2}-\mu_{i2})=\sigma_{i2}\eta_{it}-(\bar{Y}_{i2}-\mu_{i2})\\&=\sigma_{i2}\eta_{it}+O_p(T^{-1/2}),\quad\text{for } t>k_0.\end{aligned}$$

Plug in these expressions into (33), and ignore expressions involving $O_p(T^{-1})$ because they are dominated by the remaining terms, we can rewrite (33) as

$$\sum_{i=1}^{N}\Bigg[\sum_{t=1}^{k}\eta_{i1}^2+\frac{1}{\sigma_{i2}^2}\sum_{t=k+1}^{k_0}[\sigma_{i1}\eta_{it}+\mu_{i1}-\mu_{i2}]^2 + \sum_{t=k_0+1}^{T}\eta_{it}^2-(k_0-k)\log(\sigma_{i1}^2/\sigma_{i2}^2)\Bigg]. \quad (34)$$

Expanding the middle term, and subtracting $\sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{it}^2$, the above can be rewritten as

$$\sum_{i=1}^{N}\Bigg[\left(\frac{\sigma_{i1}^2}{\sigma_{i2}^2}-1\right)\sum_{t=k+1}^{k_0}\eta_{it}^2+2\frac{\sigma_{i1}}{\sigma_{i2}^2}(\mu_{i1}-\mu_{i2})\sum_{t=k+1}^{k_0}\eta_{it} + \frac{1}{\sigma_{i2}^2}(k_0-k)(\mu_{i1}-\mu_{i2})^2-(k_0-k)\log(\sigma_{i1}^2/\sigma_{i2}^2)\Bigg]$$

$$= 2\sum_{i=1}^{N}\frac{\sigma_{i1}}{\sigma_{i2}}a_i\sum_{t=k+1}^{k_0}\eta_{it}+(k_0-k)\sum_{i=1}^{N}a_i^2 \quad (35)$$

$$+\sum_{i=1}^{N}d_i\sum_{t=k+1}^{k_0}(\eta_{it}^2-1)+(k_0-k)\sum_{i=1}^{N}f(1+d_i) \quad (36)$$

where $a_i=(\mu_{i1}-\mu_{i2})/\sigma_{i2}$ and $d_i=(\sigma_{i1}^2/\sigma_{i2}^2)-1$. Eq. (35) is due to breaks in mean, and Eq. (36) is due to breaks in variance. The latter is already analyzed, see (31). The same argument will show that the first term in (35) is dominated by the second term, for $k<k_0$ and under (18). This means that for $k<k_0$, (35) diverges to infinity, thus the objective function cannot achieve its minimum. This completes the proof of Theorem 5.1. □

Proof of Theorem 5.2. We need to derive the limits of (35) and (36). By the definition of a_i, $\sum_{i=1}^{N}a_i^2=\sum_{i=1}^{N}(\mu_{i1}-\mu_{i2})^2/\sigma_{i2}^2=\frac{1}{N}\sum_{i=1}^{N}\Delta_i^2/\sigma_{i2}^2\to\tau$. By the definition of d_i, $d_i=(\sigma_{i1}/\sigma_{i2}-1)(\sigma_{i1}/\sigma_{i2}+1)$, we have $\sigma_{i1}/\sigma_{i2}=1+d_i/(1+\sigma_{i1}/\sigma_{i2})=1+O(N^{-1/2})$ because $d_i=O(N^{-1/2})$. Thus $(\sigma_{i1}/\sigma_{i2})a_i=a_i+a_iO(N^{-1/2})=a_i+O(N^{-1})$. It follows that

$$\sum_{i=1}^{N}\frac{\sigma_{i1}}{\sigma_{i2}}a_i\,\eta_{it}=\sum_{i=1}^{N}a_i\,\eta_{it}+o_p(1) = N^{-1/2}\sum_{i=1}^{N}(\Delta_i/\sigma_{i2})\eta_{it}\xrightarrow{d}\sqrt{\tau}Z_t$$

where Z_t is $N(0,1)$ by the central limit theory. We have used the fact that η_{it} are zero mean and unit variance random variables and are independent over i, and $N^{-1}\sum_{i=1}^{N}\Delta_i^2/\sigma_{i2}\to\tau$. Thus (35) converges in distribution for each fixed k to $2\sqrt{\tau}\sum_{t=k+1}^{k_0}Z_t+(k_0-k)\tau$, for $k\le k_0$. Similarly, for $k>k_0$ and by symmetry, the limit is $2\sqrt{\tau}\sum_{t=k_0+1}^{k}Z_t+(k-k_0)\tau$. Renaming $k-k_0=\ell$, the limit can be written as $2\sqrt{\tau}W_1(\ell)+|\ell|\tau$, where $W_1(\ell)$ is a two-sided Gaussian random walk.

To derive the limit of (36) notice that for small d_i, $f(1+d_i)=2^{-1}d_i^2+O(d_i^3)=2^{-1}N^{-1}\delta_i^2+O(N^{-3/2})$. Thus $\sum_{i=1}^{N}f(1+d_i)=2^{-1}N^{-1}\sum_{i=1}^{N}\delta_i^2+O(N^{-1/2})\to2^{-1}\omega$ by assumption. Next, $\sum_{i=1}^{N}d_i(\eta_{it}^2-1)=N^{-1/2}\sum_{i=1}^{N}\delta_i(\eta_{it}^2-1)$. The variance of $\delta_i(\eta_{it}^2-1)$ is equal to $\delta_i^2(\kappa+2)$. By the central limit theorem, $N^{-1/2}\sum_{i=1}^{N}\delta_i(\eta_{it}^2-1)\xrightarrow{d}\sqrt{(\kappa+2)\omega}Z_t^*$, where Z_t^* is $N(0,1)$. Thus for $k\le k_0$, the limit of (36) is $\sqrt{(\kappa+2)\omega}\sum_{t=k+1}^{k_0}Z_t^*+(k_0-k)2^{-1}\omega$. Combining with the case of $k>k_0$, the limit can be rewritten as $\sqrt{(\kappa+2)\omega}W_2(\ell)+|\ell|2^{-1}\omega$, where $\ell=k-k_0$, and $W_2(\ell)$ is a two-sided Gaussian random walk. Adding up the limits of (35) and (36) leads to Theorem 5.2.

Let $Z_{Nt}=N^{-1/2}\sum_{i=1}^{N}(\Delta_i/\sigma_{i2})\eta_{it}$ and $Z_{Nt}^*=N^{-1/2}\sum_{i=1}^{N}\delta_i(\eta_{it}^2-1)$. The above analysis shows $Z_{Nt}\xrightarrow{d}\sqrt{\tau}Z_t$ and $Z_{Nt}^*\xrightarrow{d}[(\kappa+2)\omega]^{1/2}Z_t^*$. The variables Z_{Nt} and Z_{Nt}^* are asymptotically independent if $E(\eta_{it}^3)=0$, or $N^{-1}\sum_{i=1}^{N}(\Delta_i/\sigma_{i2})\delta_i\to0$. This follows from $E(Z_{Nt}Z_{Nt}^*)\to0$ in either case, as $N\to\infty$. This verifies the claim in Remark 1. To prove (19) in Remark 2, let μ_3 and π be defined there. Then $(Z_{Nt},Z_{Nt}^*)'$ converges jointly to a bivariate normal vector with covariance matrix

$$\begin{bmatrix}\tau & \mu_3\pi\\ \mu_3\pi & (\kappa+2)\omega\end{bmatrix}.$$

Thus $2Z_{Nt}+Z_{Nt}^*\xrightarrow{d}2[\tau+(\kappa+2)\omega/4+\mu_3\pi]^{1/2}Z_t^\dagger$, where $Z_t^\dagger$ is $N(0,1)$. The nonrandom component $\sum_{i=1}^{N}a_i^2+\sum_{i=1}^{N}f(1+d_i)$ in (35) and (36) has a limit $\tau+2^{-1}\omega$, as argued earlier. Combining the results we obtain (19). □

References

Arellano, M., 2003. Panel Data Econometrics. Oxford University Process, New York.

Bai, J., 1994. Least squares estimation of a shift in linear processes. Journal of Time Series Analysis 15, 453–472.

Bai, J., 1997a. Estimating multiple breaks one at a time. Econometric Theory 13, 315–352.

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

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

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

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change. Journal of Applied Econometrics 18, 1–22.

Bai, J., Lumsdaine, R., Stock, J., 1998. Testing for and dating common breaks in multivariate time series. Review of Economc Studies 65, 395–432.

Bhattacharya, P.K., 1994. Some aspects of change-point problems. In: Carlstein, E., Müller, H.G., Siegmund, D. (Eds.), Change-point Problems. Institute of Mathematical Statistics, CA, pp. 28–56.

Chow, G., 1960. Tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 591–605.

Csörgő, M., Horváth, L., 1997. Limit Theory in Change-Point Analysis. Wiley, New York.

Emerson, J., Kao, C., 2001. Testing for structural change of a time trend regression in panel data: Part I. Journal of Propagations in Probability and Statistics 2, 57–75.

Emerson, J., Kao, C., 2002. Testing for structural change of a time trend regression in panel data: Part II. Journal of Propagations in Probability and Statistics 2, 207–250.

Hsiao, C., 2003. Analysis of Panel Data. Cambridge University Press, New York.

Jackson, C.H., Sharples, L.D., 2004. Models for longitudinal data with censored changepoints. Applied Statistics 53, 149–162.

Joseph, L., Vandal, A., Wolfson, D.B., 1996. Estimation in the multi-path change-point problem for correlated data. Canadian Journal of Statistics 24, 37–53.

Joseph, L., Wolfson, D.B., 1992. Estimation in multi-path change-point problems. Communications in Statistics, Theory and Methods 21, 897–913.

Joseph, L., Wolfson, D.B., 1993. Maximum likelihood estimation in the multi-path changepoint problem. Annals of the Institute of Statistical Mathematics 45, 511–530.

Joseph, L., Wolfson, D.B., Berger, R.D., Lyle, R.M., 1997. Analysis of panel data with change-points. Statistica Sinica 7, 687–703.

Kiefer, J., Wolfowitz, J., 1956. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Annals of Mathematical Statistics 27, 887–906.

Picard, D., 1985. Testing and Estimating change-points in time series. Advances in Applied Probability 17, 841–867.

Skates, S.J., Pauler, D.K., Jacobs, I.J., 2001. Screening based on the risk of cancer calculation from Bayesian hierarchical changepoint and mixture models of longitudinal markers. Journal of the American Statistical Association 96, 429–439.

Wachter, S.D., Tzavalis, E., (2004a). Detection of structural breaks in linear dynamic panel data models. Working paper No. 505, Queen Mary, University of London.

Wachter, S.D., Tzavalis, E., (2004b). Monte Carlo comparison of model and moment selection and classical inference approaches to break detection in panel data models. Unpublished manuscript, Queen Mary, University of London.

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