Break Point Estimation in Fixed Effects Panel Data

Otilia Boldea*, Zhuojiong Gan† and Bettina Drepper ‡

February 15, 2016

Abstract

In this paper, we propose estimating multiple break-points in panel data with fixed effects by ordinary least-squares. We show that, despite the endogeneity bias of the slope estimators, the break-point estimators are consistent as $N \to \infty$ provided a reasonable time homogeneity condition holds. We also propose an information criterion that selects the true number of breaks with probability one in the limit.

In addition to multiple breaks, we allow for time dependence in the errors. These two features may decrease the asymptotic efficiency of the usual fixed-effects slope estimators relative to other estimators. We study two other slope estimators and show that their relative efficiency may be higher than of the usual fixed-effects estimator even for iid data. We illustrate our findings via a simulation study. [1]

1 Introduction

This paper proposes estimating multiple break-points in panel data models with fixed effects via ordinary least-squares (OLS). Typically, OLS slope estimators are inconsistent due to the unobserved, individual heterogeneity bias. However, we show that as long as a reasonable time-homogeneity condition holds, this bias has no impact on the consistent estimation of break-points.

There is a growing literature on estimation of common break-points in panel data models. Breaks in panels may be due to financial crises, policy changes, housing bubbles, technological changes, to mention just a few causes.

*Tilburg University, Department of Econometrics and Operation Research and CentER, E-mail:o.boldea@uvt.nl.

†Corresponding author, Southwestern University of Finance and Economics, Chengdu, China, E-mail: ganzj@swufe.edu.cn.

‡Tilburg University, Department of Econometrics and Operation Research and CentER, E-mail: b.drepper@uvt.nl.

[1]Empirical application in progress.

Regarding break-point detection, Emerson and Kao (2001, 2002) and de Wachter and Tzavalis (2012) propose tests for the presence of breaks in (dynamic) panel data models and derive their asymptotic distribution. Regarding break-point estimation, there are different methods that were recently proposed. Using a panel data model with no covariates, Bai (2010) shows that the common break can be treated as a mean-shift and estimated via OLS. Feng, Kao and Lazarová (2009) and Baltagi, Kao and Liu (2015) propose OLS estimation of one common break in the slope parameters of a panel data regression model without individual specific effects. In the presence of fixed effects, they propose first-differencing (FD) the data prior to break point estimation; Baltagi, Kao and Liu (2015) show that the break-point estimator on the first-differenced data is consistent for $N, T \rightarrow \infty$. Adaptive group-fused LASSO (AGFL) estimators of the break-points in the first-differenced data are proposed in Qian and Su (2016). They derive conditions under which the adaptive group fused LASSO delivers the true number and location of break-points with probability one in the limit, with $N \rightarrow \infty$ and T fixed or $T \rightarrow \infty$. Qian and Su (2016) also derive the asymptotic properties of the post-LASSO first-difference parameter estimators with $N \rightarrow \infty$ and T fixed or $T \rightarrow \infty$. All these studies assume no cross section dependence and homogeneous panels. For cross-section dependence in the form of interactive fixed effects (heterogeneous panels), Li, Qian and Su (2015) propose estimating the number and location of break-points via a penalized principal component estimation with AGFL, and show that their method detects the true number and location of break-points with probability tending to one as $N \rightarrow \infty$. Baltagi, Feng and Kao (2016) allow for heterogeneous (random coefficient) slopes with breaks and propose (common-correlated effects) CCE estimation of the multiple break-points. They show that this method can also consistently estimate break-points, as $N, T \rightarrow \infty$.

All these papers start from the premise that since the OLS slope estimators are inconsistent due to fixed effects, these fixed effects need to be removed before break-point estimation. In contrast to these studies, our paper shows that the fixed effects need not be removed for consistency of break-point estimators.

Thus, *our first contribution* is to show that the OLS estimators of the break-points in homogeneous panels with fixed effects, like their AFGL or FD counterparts, are also consistent with $N \rightarrow \infty$ and T fixed, as long as a reasonable time homogeneity condition holds, known as *the Mundlak assumption*.[2] The intuition is that even though the OLS slope estimators at each

[2]This time homogeneity implies that we cannot allow for lagged dependent variables. But we do allow for the regressors and the errors to be weakly time-dependent. There are many examples where the time homogeneity

candidate break-point partition are inconsistent, the overall asymptotic bias in the OLS objective function is the same regardless of the break-point partition considered. Therefore, the endogeneity bias is irrelevant for minimizing the OLS criterion over different break partitions, yielding consistent estimators of the true break partition. A similar intuition can be found in Perron and Yamamoto (2013) for time series, although their endogeneity source is different.

Unlike current methods, our method can also detect a break-point in the first time period, and should be more accurate in finite samples because it doesn't remove sample information. To select the number of breaks, we propose an information criterion similar to BIC; we show that it correctly detects the number of break-points with probability tending to one as $N \to \infty$ and T is fixed.

Our second contribution is to show that the presence of breaks alters the usual intuition that, for example, the fixed-effects slope estimators are the most efficient when the data in the level equation is weakly dependent. To our best knowledge, the efficiency of slope estimators has not been analyzed before in the context of breaks. For example, Qian and Su (2016) and Baltagi, Feng and Kao (2015) both propose FD estimators, but these are known to be inefficient when the errors in the level equation are not unit root.

We assume weak time dependence in the level data and analyze three estimators and their asymptotic distributions. All these estimators are consistent, since they are all based on removing the fixed effects. But the way they are removed changes the underlying moment conditions and therefore the asymptotic variance of these estimators. The first one is the conventional fixed-effects (FE) estimator, obtained by demeaning the data before and after the break, and estimating the slopes in each subsample with demeaned data. The other two estimators are new: one of them (IV) uses the demeaned regressors as instruments in the level equation, and the other (FFE) first demeans the data over the full sample rather than subsamples, then estimates the slopes before and after the break jointly in this transformed equation. We show that with general weak dependence patterns, they cannot be further compared; in this case, stacking the moment conditions from the three estimators might be best. With iid mean-zero regressors and idiosyncratic errors (the most common setting), the FFE estimator is the most efficient, and not

condition holds, or more precisely when the impact of the fixed effects on the regressors and dependent variables does not diminish over time, especially when the number of periods is small. For example, in modelling the decision to buy a house, the initial impression of a neighborhood may influence both the subjective valuation of the house and the decision to buy a house in that neighborhood in the same way over different periods. In a growth regression, the initial quality of institutions can have the same persistent impact on both corruption and growth for many periods.

the FE estimator, because it uses more information from adjacent subsamples. If we also have random effects instead of fixed effects, both the FFE and IV estimators are more efficient than the FE estimator.

The rest of the paper is organized as follows. Section 2 introduces the model with individual fixed effects and a common break. Section 3 covers the properties of the OLS break point estimators of multiple breaks. Section 4 proposes an consistent information criterion to select the number of breaks. Section 5 derives the asymptotic propoerties of both the OLS slope estimators and the three estimators proposed. It also compares their asymptotic efficiency in special cases. Section 6 studies the finite-sample properties of the break point and slope estimators. Section 7 concludes. All the proofs are relegated to the Appendix.

Notation: Matrices and vectors are denoted with bold symbolds, and scalars are not. Define for a scalar S, the generalized vec operator $\mathbf{vec}_{1:S}(\boldsymbol{A}_s) \overset{\text{def}}{=} (\boldsymbol{A}'_1, \ldots, \boldsymbol{A}'_S)'$, stacking in order the matrices $\boldsymbol{A}_s, (s = 1, \ldots, S)$, which have the same number of columns. If S is the number of breaks, let $T_0 = 0$, $T_{S+1} = T$, with T the sample size, and let $\boldsymbol{T}_S = (T_0, \mathbf{vec}_{1:m}(T_s), T_{S+1})$ be **a sample partition** of interval $[1, T]$ into subsamples defined by $0 < T_1 < T_2 \ldots < T_S < T$, and define the intervals $I_s = [T_{s-1}+1, T_s]$ for $s = 1, \ldots, S+1$. Let $\boldsymbol{X} = \mathbf{vec}_{1:T}(\boldsymbol{X}_t)$ be the $NT \times p$ matrix that stacks $\boldsymbol{X}_t = \mathbf{vec}_{1:N}(\boldsymbol{x}'_{it})$ in order. Call $\widetilde{\boldsymbol{X}} = \mathbf{diag}(\mathbf{vec}_{1:T_1}(\boldsymbol{X}_t), \mathbf{vec}_{T_1+1:T_2}(\boldsymbol{X}_t), \ldots, \mathbf{vec}_{T_m+1:T_{m+1}}(\boldsymbol{X}_t))$ **the diagonal partition** of $\boldsymbol{X}$ at $\boldsymbol{T}_S$, with $X_1, \ldots, X_T$ on the diagonal and the rest of the elements zero. For any random vector z, denote by $|z|$ the Euclidean norm, and by $\|z\|_p = (\mathbb{E}|z|^p)^{1/p}$ the L_p norm.

2 Model

Assume that the true model is piecewise-linear with m^0 breaks:

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta}^0_j + c_i + \varepsilon_{it}, \qquad t \in I^0_j, \qquad j = 1, \ldots, m^0 + 1. \tag{1}$$

In (1), i are cross-sections, t for time, $i = 1, \ldots, N$ with $N \to \infty$, and $t = 1, \ldots, T$, with T fixed. Also, y_{it} is a scalar observed dependent variable; $\boldsymbol{x}_{it}$ is a $p \times 1$ observed vector of regressors that can be constant over time but vary across individuals; m^0 is the true unknown number of breakpoints with $1 \leq m^0 \leq T - 1$, $T^0_j, (j = 1, \ldots, m^0)$ are the true unknown break points belonging to the sample partition $\boldsymbol{T}^0_{m^0} = (T^0_0, \mathbf{vec}_{1:m^0}(T^0_j), T^0_{m^0+1})$, where $T^0_0 = 0$ and $T^0_{m^0+1} = T$. Also, $\boldsymbol{\beta}^0_j$

are unknown $p \times 1$ parameters of interest in intervals $I_j^0, (j = 1, \ldots, m^0 + 1)$, where $\boldsymbol{\beta}_j^0 \neq \boldsymbol{\beta}_s^0$ for all $s \neq j$, and ϵ_{it} are unobserved mean-zero idiosyncratic errors, uncorrelated with the unobserved individual specific effects c_i. We assume that c_i are mean-zero fixed effects ($\mathbb{E}(c_i) = 0$ and $\mathbb{E}(c_i \boldsymbol{x}_{it}) = \boldsymbol{a}_i$). Random effects are a special case, when $\boldsymbol{a}_i = 0$ for all i. As equation (1) suggests, we do not include time-fixed effects in our model. If they would be included as time-dummies, they would be equivalent to breaks in the intercept at each time-period.

Consistently estimating the number of breaks m^0 is important and discussed in Section 4. For now, assume that the number of breaks m^0 is known, and we are interested in estimating the unknown break-points $\boldsymbol{T}_{m^0}^0$ and the parameters $\boldsymbol{\beta}_j^0$.

Contrary to other papers, we propose estimating the true partition $\boldsymbol{T}_{m^0}^0$ by pooled least-squares (OLS). We show that the corresponding OLS estimator $\widehat{\boldsymbol{T}}_{m^0}$ is consistent when $N \to \infty$, for any fixed T. The key assumption is Assumption 1(iv) below. It states that the limiting cross-section average covariance between the regressors and the individual effects is constant over time, also known as the *Mundlak assumption.* This assumption is trivially satisfied for random effects, because in this case the covariance is zero for each time period. This assumption is also reasonable for certain fixed effects settings, because there are many instances where the initial impact of the fixed effects on the regressors is the same (initial impression, wealth, knowledge, etc.) even when the regressors evolve over time. This is especially the case when T is small. We do not allow for dynamics in the form of lagged dependent variables, because then, by definition of a dynamic panel with fixed effects, the covariance between these and the individual effects changes over time, violating Assumption 1(iv). However, we do allow for dynamics in the errors ϵ_{it} and regressors $\boldsymbol{x}_{it}$.

To describe the OLS break-point estimator $\widehat{\boldsymbol{T}}_{m^0}$, write the model in a more compact form by letting $u_{it} = c_i + \epsilon_{it}$, $\boldsymbol{u} = \mathbf{vec}_{1:T}(\mathbf{vec}_{1:N}(u_{it}))$, $\boldsymbol{\beta}^0 = \mathbf{vec}_{1:m^0+1}(\boldsymbol{\beta}_j^0)$, $\boldsymbol{y} = \mathbf{vec}_{1:T}(\mathbf{vec}_{1:N}(y_{it}))$, and $\widetilde{\boldsymbol{X}}^0$ the diagonal partition of $\boldsymbol{X}$ at the true partition $\boldsymbol{T}_{m^0}^0$. Then, we can write model (1) as:

$$\boldsymbol{y} = \widetilde{\boldsymbol{X}}^0 \boldsymbol{\beta}^0 + \boldsymbol{u}. \tag{2}$$

We estimate (2) by minimizing the sum of squared residuals over all sample partitions $\boldsymbol{T}_{m^0}$, or regressing $\boldsymbol{y}$ on $\widetilde{\boldsymbol{X}}$:

$$\min_{\boldsymbol{T}_{m^0}} S_{NT}(\boldsymbol{T}_{m^0}) \stackrel{\text{def}}{=} \min_{\boldsymbol{T}_{m^0}} (NT)^{-1} \left(\boldsymbol{y} - \widetilde{\boldsymbol{X}} \widehat{\boldsymbol{\beta}}_{OLS}(\boldsymbol{T}_{m^0})\right)' \left(y - \widetilde{\boldsymbol{X}} \widehat{\boldsymbol{\beta}}_{OLS}(\boldsymbol{T}_{m^0})\right), \tag{3}$$

where $\widehat{\boldsymbol{\beta}}_{OLS}(\boldsymbol{T}_{m^0}) = (\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{y}$ is the OLS estimator using $\boldsymbol{T}_{m^0}$ as the candidate partition. Let $\widehat{\boldsymbol{\beta}}_{OLS}(\boldsymbol{T}_{m^0}) = \mathbf{vec}_{1:m^0+1}(\widehat{\boldsymbol{\beta}}_{OLS,j}(\boldsymbol{T}_{m^0}))$, where the OLS estimator $\widehat{\boldsymbol{\beta}}_{OLS,j}(\boldsymbol{T}_{m^0})$ is the OLS estimator for $\boldsymbol{\beta}_j^0$, given partition $\boldsymbol{T}_{m^0}$.

The minimizer of the above problem is denoted $\widehat{\boldsymbol{T}}_{m^0}$, and we refer to it as the OLS break point estimators. If the minimizer is not unique, we break the tie by picking the smallest minimizer as our estimator $\widehat{\boldsymbol{T}}_{m^0}$. The OLS estimator of $\boldsymbol{\beta}^0$ at the estimated partition is denoted by $\widehat{\boldsymbol{\beta}}_{OLS} = \widehat{\boldsymbol{\beta}}_{OLS}(\widehat{\boldsymbol{T}}_{m^0}) = \mathbf{vec}_{1:m^0+1}(\widehat{\boldsymbol{\beta}}_{OLS,j}(\widehat{\boldsymbol{T}}_{m^0}))$.

Even if the true partition $\boldsymbol{T}^0_{m^0}$ was known, the OLS estimator $\widehat{\boldsymbol{\beta}}_{OLS}(\boldsymbol{T}^0_{m^0})$ based on that partition would still suffer from endogeneity bias, due to the fixed effects. However, we show below that under the Mundlak assumption, regardless of the partition $\boldsymbol{T}_{m^0}$ considered, the endogeneity bias is asymptotically the same for all $\widehat{\boldsymbol{\beta}}_{OLS,j}(\boldsymbol{T}_{m^0})$, and will increase the averaged sum of squared residuals $S_{NT}(\boldsymbol{T}_{m^0})$ by the same amount asymptotically, regardless of the partition considered. Hence, just as in the case of no individual effects, $S_{NT}(\boldsymbol{T}_{m^0})$ is asymptotically uniquely minimized at $\boldsymbol{T}^0_{m^0}$ when $N \to \infty$. This is the intuition behind the consistency of the OLS break-point estimators $\widehat{\boldsymbol{T}}_{m^0}$, derived in the next section. [3,4]

3 Break Point Estimators

In this section, consistency of the OLS break-point estimators $\widehat{\boldsymbol{T}}_{m^0}$, when $N \to \infty$ and T is fixed.

Rather than proceeding with high-level assumptions on the data that are difficult to verify, as in Qian and Su (2014), Assumption A3, we chose to use simple primitive assumptions on the data because they are easier verified by practitioners, whether intuitively or through testing.

Assumption 1. *For all $i = 1, \ldots, N$, and $t, s = 1, \ldots, T$,*
(i) $\{\boldsymbol{x}'_{it}, c_i, \epsilon_{it}\}_{t=1}^T$ is independent over i;
(ii) $\sup_{i\in\mathbb{N}} \mathbb{E}|\boldsymbol{x}_{it}|^{2+\xi} < \infty$ for some $\xi > 0$, and $N^{-1}\sum_{i=1}^N \mathbb{E}[\boldsymbol{x}_{it}\boldsymbol{x}'_{it}] \to \boldsymbol{Q}$, where $\boldsymbol{Q}$ is a positive definite (pd) matrix not depending on i or t;
(iii) $\mathbb{E}[\epsilon_{it}] = 0$, $\mathbb{E}[\epsilon_{it}^2] = \sigma_\epsilon^2$, $\sup_{i\in\mathbb{N}} \mathbb{E}|\epsilon_{it}|^{2+\xi} < \infty$ for some $\xi > 0$, and $\mathbb{E}[x_{is}\epsilon_{it}] = 0$;
(iv) $\mathbb{E}[c_i] = 0$, $\mathbb{E}[c_i^2] = \sigma_c^2$, $\sup_{i\in\mathbb{N}} \mathbb{E}|c_i|^{2+\xi} < \infty$ for some $\xi > 0$, and $\mathbb{E}[\boldsymbol{x}_{it}c_i] = \boldsymbol{a}_i$, with $\frac{1}{N}\sum_{i=1}^N \boldsymbol{a}_i \to \boldsymbol{a}$;

[3] Perron and Yamamoto (2013) use a similar intuition to propose an ordinary least-squares break point estimation method in time series models with endogenous regressors.

[4] If $T \to \infty$, we can allow $\mathbb{E}[\boldsymbol{x}_{it}c_i]$ to vary over time, and we conjecture that all our results go through, as long as the time variation in $\mathbb{E}[\boldsymbol{x}_{it}c_i]$ is $o(T^{-\gamma})$ uniformly in i, for some $\gamma > 0$.

(v) ϵ_{it} *is independent of* c_i*.*

Assumption 1(i) states that the data are independent over i, as in most papers on panel break-points - Feng, Kao, Lazarová (2009), Bai (2010) and Qian and Su (2014). This applies, for example, to survey data. As discussed before, $\mathbb{E}[\boldsymbol{x}_{it}c_i] = \boldsymbol{a}_i$ is the Mundlak assumption. Note that this allows for the presence of time-invariant regressors.

The assumption that the $(2+\xi)$-moments exist is a technical requirement for applying the Weak Law of Large Numbers (WLLN). The time-homnogeneity assumption on the second moments, i.e. $N^{-1}\sum_{i=1}^{N}\mathbb{E}[\boldsymbol{x}_{it}\boldsymbol{x}'_{it}] \to \boldsymbol{Q}$, while quite common, is also necessary in establishing that the endogeneity bias in the residual sum of squares does not vary over time and partitions. A similar assumption can be found in Perron and Yamamoto (2013). The rest of the assumptions are also common for panel data. We further note that since we only study the case $N \to \infty$, we can allow for diverse forms of serial dependence in $\boldsymbol{x}_{it}$ and ϵ_{it}. However, we do not allow for x_{it} or ϵ_{it} to be a unit root; this is implied by Assumption 1(ii) and (iii).

The following lemma is helpful in establishing consistency of the OLS break-point estimators.

Lemma 1. *Under Assumption 1, the following holds for each* t*:*

(i) $N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}'_{it} \xrightarrow{p} \boldsymbol{Q}$;

(ii) $N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_{it}c_i \xrightarrow{p} \boldsymbol{a}$;

(iii) $N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_{it}\epsilon_{it} \xrightarrow{p} \boldsymbol{0}$;

(iv) $N^{-1}\sum_{i=1}^{N}u_{it}^2 \xrightarrow{p} \sigma_\epsilon^2 + \sigma_c^2$.

This lemma follows from standard application of the WLLN. Although Theorem 1 below is shown under Assumption 1, it can be also proved using Lemma 1 alone. Thus, if one wishes to start from high-level assumptions, Lemma 1 is sufficient for the results in Theorem 1. Therefore, it also applies to some models with interactive fixed effects. To see that, suppose we replace c_i by c_if_t, where c_i and f_t are unobserved scalar random variables, independent of each other, and $\mathbb{E}(c_i) = 0$. Then as long as $\mathbb{E}(\boldsymbol{x}_{it}c_if_t)$ does not vary with t, Lemma 1(ii) may hold by some WLLN. Also, $\mathbb{E}(u_{it}^2) = \mathbb{E}(\epsilon_{it}^2) + \mathbb{E}(c_i^2)\mathbb{E}(f_t^2)$. So as long as $\mathbb{E}(f_t^2)$ is constant over t, a common assumption in interactive fixed-effects models, Lemma 1(iv) will hold as well.

We can now state the properties of the break point estimator $\widehat{\boldsymbol{T}}_{m^0}$. Let $\delta^0 = \beta_1^0 - \beta_2^0 \neq 0$, and:

$$S_{NT}^*(\boldsymbol{T}_{m^0}) \stackrel{\text{def}}{=} (\sigma_\epsilon^2 + \sigma_c^2 - \boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a}) + (NT)^{-1}\boldsymbol{\beta}^0\widetilde{\boldsymbol{X}}^{0\prime}\boldsymbol{M}_{\widetilde{\boldsymbol{X}}}\widetilde{\boldsymbol{X}}^0\boldsymbol{\beta}^0 \quad (4)$$

Theorem 1. *[Consistency of the OLS Break Point Estimator]*
Under Assumption 1, with $N \to \infty$ *and* T *fixed, and* $1 \leq m^0 \leq T-1$
(i) $S_{NT}(\boldsymbol{T}_{m^0}) - S^*_{NT}(\boldsymbol{T}_{m^0}) \xrightarrow{p} 0$*;*
(ii) $\text{plim}(S_{NT})$ *is uniquely minimized at* $\boldsymbol{T}^0_{m^0}$*;*
(iii)$P(\widehat{\boldsymbol{T}}_{m^0} = \boldsymbol{T}^0_{m^0}) \to 1$.

This theorem states that the OLS break point estimator is consistent. For the same OLS estimator, a similar result is derived in Feng, Kao and Lazarová (2009) and Baltagi, Kao and Liu (2015) for iid data without individual effects. The same result as ours is also in Bai (2010), Theorem 3.1, but for a break in individual-specific means, without any regressors. To our knowledge, this is the first paper that establishes the properties of the OLS break point estimator in panel data when fixed effects are present. FD break-point estimators are considered in Qian and Su (2016) and Baltagi, Kao and Liu (2015).

In contrast to time series, Theorem 1 shows that we can locate the true break-points $\boldsymbol{T}^0_{m^0}$ with probability tending to one. This is due to a large number of observations N at each point in time. A large N also allows us to consider break-points at any locations in the sample, even in adjacent periods. In contrast to Qian and Su (2016) and Baltagi, Kao and Liu (2016), we can also allow for breaks in the first period.

Theorem 1 also shows that there are two bias terms in $\text{plim}(S^*_{NT}(\boldsymbol{T}_{m^0}))$ and therefore in the sum of squared residuals $\text{plim}(S^*_{NT}(\boldsymbol{T}_{m^0}))$. The first bias term, $(-\boldsymbol{a}'Q^{-1}\boldsymbol{a})$, is the same regardless of the partition $\boldsymbol{T}_{m^0}$ considered, and it arises because of the endogeneity bias of the OLS estimators $\widehat{\beta}_{OLS}(\boldsymbol{T}_{m^0})$. Therefore, asymptotically, the endogeneity bias does not play a role in the estimation of the break-point partition. The second bias term, $\text{plim}((NT)^{-1}\boldsymbol{\beta}^0\widetilde{\boldsymbol{X}}^{0\prime}\boldsymbol{M}_{\widetilde{\boldsymbol{X}}}\widetilde{\boldsymbol{X}}^0\boldsymbol{\beta}^0$, varies with the partition considered, and is uniquely minimized at the true partition $\boldsymbol{T}^0_{m^0}$; therefore, the break-point estimator is consistent.

4 Estimating the Number of Break Points

In the previous section, we assumed that the number of break-points m^0 is known. In practice, it is unknown and must be estimated from the data. We propose estimating m^0 via the following information criterion:

$$\widehat{m} \stackrel{\text{def}}{=} \arg\min_{m \in \{1,\ldots,T-1\}} IC(m), \qquad IC(m) = \log S_{NT}(\widehat{\boldsymbol{T}}_m) + [p(m+1) + 3m]\,\frac{log(NT)}{NT} \qquad (5)$$

This is similar to BIC: we penalize for the number of parameters $p(m+1)$, and the break-points are worth three times a regular parameter: $(3m)$. This is motivated by the bias calculations in Nimomiya (2005). Hall, Osborn and Sakkas (2013) show that this criterion is more accurate than the BIC in estimating the true number of breaks in time series. We replaced their sample size by our effective sample size NT.

We expect that under Assumption 1, $P(\hat{m} = m^0) \to 1$ as $N \to \infty$ (proof in progress). Because of this consistency result, unlike sequential testing, asymptotically there is no impact of this model selection procedure, so the estimated number of breaks can be treated as known when estimating the break-point locations. Note again that the endogeneity bias has no impact on the model selection procedure under Assumption 1.

In practice, evaluation of such a criterion requires calculation of the sum of squared residuals for a bit less 2^T partitions; while T is small, 2^T may still seem large. However, because we use the Bai and Perron (2003) dynamic programming algorithm, which stores and reuses sum of squared residuals calculated over different subsamples, subsamples which repeat over partitions with different m, the computational cost is greatly reduced, and the program runs fast for example, for $m, T \leq 20$ and $N = 500$.

5 Slope Estimators and Their Asymptotic Properties

To facilitate presentation, we focus on the case $m^0 = 1$ from here onwards, and we treat $m^0 = 1$ as known (a consistent estimator is provided in the previous section).[5]

5.1 OLS Estimators

In this section, we analyze the OLS estimators of $\boldsymbol{\beta}^0$. Recall that $\widehat{\boldsymbol{\beta}}_{OLS,1}$ and $\widehat{\boldsymbol{\beta}}_{OLS,2}$ are the first and the last p elements in $\widehat{\boldsymbol{\beta}}_{OLS}$, and let $k^0 \overset{\text{def}}{=} T_1^0$. As expected, the OLS estimators are inconsistent due to the fixed effects bias. However, we prove below that the asymptotic bias is the same for the first and the second sub-sample, so that the OLS estimator $\widehat{\boldsymbol{\delta}}_{OLS} = \widehat{\boldsymbol{\beta}}_{OLS,1} - \widehat{\boldsymbol{\beta}}_{OLS,2}$ of the difference $\boldsymbol{\delta}^0$ is consistent. We impose the following assumption:

Assumption 2. *(i)* $\sup_{i \in \mathbb{N}} \mathbb{E}|\boldsymbol{x}_{it}|^{4+\xi} < \infty$, $\sup_{i \in \mathbb{N}} \mathbb{E}|u_{it}|^{4+\xi} < \infty$ *for some* $\xi > 0$*;*
(ii) $\lim_{N \to \infty} \left(N^{-1} \sum_{i=1}^N \mathbb{E}[\boldsymbol{x}_{it} \boldsymbol{x}_{is}' u_{it} u_{is}] \right) = \boldsymbol{W}_{t,s}$.

[5] All these results feature straightforward generalizations to multiple break-points, without much further insight.

This assumption is standard and needed for the central limit theorem. Let:

$$\boldsymbol{\Delta}_{N,\boldsymbol{\beta}} \stackrel{\text{def}}{=} \left(\boldsymbol{a}'\left((Nk^0)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{k^0}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right)^{-1}, \qquad \boldsymbol{a}'\left((N(T-k^0))^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-k^0}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right)^{-1}\right)';$$

$$\boldsymbol{\Delta}_{N,\boldsymbol{\delta}} \stackrel{\text{def}}{=} \left[\left((Nk^0)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{k^0}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right)^{-1} - \left((N(T-k^0))^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-k^0}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right)^{-1}\right]\boldsymbol{a};$$

$$\boldsymbol{W}_{\boldsymbol{x}u} = \begin{pmatrix} T^{-2}\sum_{t,s=1}^{k^0}(\boldsymbol{W}_{t,s}-\boldsymbol{a}\boldsymbol{a}') & T^{-2}\sum_{t=1}^{k^0}\sum_{s=k^0+1}^{T}(\boldsymbol{W}_{t,s}-\boldsymbol{a}\boldsymbol{a}') \\ T^{-2}\sum_{t=k^0+1}^{T}\sum_{s=1}^{k^0}(\boldsymbol{W}_{t,s}-\boldsymbol{a}\boldsymbol{a}') & T^{-2}\sum_{t,s=k^0+1}^{T}(\boldsymbol{W}_{t,s}-\boldsymbol{a}\boldsymbol{a}') \end{pmatrix}$$

$$\boldsymbol{\Omega}_{\beta} \stackrel{\text{def}}{=} \begin{pmatrix} k^0T^{-1}\boldsymbol{Q} & 0 \\ 0 & (T-k^0)T^{-1}\boldsymbol{Q} \end{pmatrix}^{-1} \boldsymbol{W}_{\boldsymbol{x}u} \begin{pmatrix} k^0T^{-1}\boldsymbol{Q} & 0 \\ 0 & (T-k^0)T^{-1}\boldsymbol{Q} \end{pmatrix}^{-1}$$

$\boldsymbol{\Omega}_{\boldsymbol{\delta}} \stackrel{\text{def}}{=} (\boldsymbol{I}_p, -\boldsymbol{I}_p)\boldsymbol{\Omega}_{\beta}(\boldsymbol{I}_p, -\boldsymbol{I}_p)'$, with $\boldsymbol{I}_p$ the $p \times p$ identity matrix.

Theorem 2. *[OLS Estimators]*
Under Assumptions 1 and 2, when $N \to \infty$ and T is fixed,
(i) $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}^0 - \boldsymbol{\Delta}_{N,\boldsymbol{\beta}}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_{\boldsymbol{\beta}})$; *(ii)* $\sqrt{N}(\widehat{\boldsymbol{\delta}}_{OLS} - \boldsymbol{\delta}^0 - \boldsymbol{\Delta}_{N,\boldsymbol{\delta}}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_{\boldsymbol{\delta}})$.

It follows from Lemma 1(i) that $\text{plim}_{N\to\infty}\boldsymbol{\Delta}_{N,\boldsymbol{\delta}} = \boldsymbol{Q}^{-1}\boldsymbol{a} - \boldsymbol{Q}^{-1}\boldsymbol{a} = \boldsymbol{0}$, so the OLS estimator of the difference $\boldsymbol{\delta}^0$ is consistent as long as $N \to \infty$, even if $\widehat{\boldsymbol{\beta}}_{OLS}$ is not consistent. It can also be seen from this theorem that $\widehat{\boldsymbol{\beta}}_{OLS}$ is not consistent for $\boldsymbol{\beta}^0$. If we let $N \to \infty$ for $\boldsymbol{\Delta}_{N,\boldsymbol{\beta}}$, the asymptotic bias is equal to $\boldsymbol{Q}^{-1}\boldsymbol{a}$ for both estimators of $\boldsymbol{\beta}_1^0$ and $\boldsymbol{\beta}_2^0$, and does not equal zero unless the individual effects are exogenous ($a = 0$).

This theorem also implies that although $\boldsymbol{\delta}^0$ can be consistently estimated using OLS, OLS inference is infeasible because of the presence of $\boldsymbol{a}$ in $\boldsymbol{\Delta}_{N,\boldsymbol{\delta}}$ and $\boldsymbol{\Omega}_{\boldsymbol{\delta}}$, unidentified without further assumptions. Therefore, we propose alternative methods for inference on both $\boldsymbol{\beta}^0$ and $\boldsymbol{\delta}^0$ below.

5.2 Consistent Estimators

Given a consistent break point estimator $\widehat{k}$ that satisfies $P(\widehat{k} = k^0) \to 1$ (whether it is the estimator we propose or a different estimator such as the one in Qian and Su (2016) or Baltagi, Kao and Liu (2016)), we seek consistent estimators of $\boldsymbol{\beta}_1^0, \boldsymbol{\beta}_2^0$. The results in this section can be applied to any break point estimator $\widehat{k}$ that satisfies $P(\widehat{k} = k^0) \to 1$ and are not specific to the OLS estimator proposed in the previous section.

To get consistent estimators of the slope parameters, we need to resort to some method to either remove the fixed effects, as is common in the panel data literature, or to instrument the endogenous regressors x_{it}. Since we assume that $\mathbb{E}[\boldsymbol{x}_{is}\epsilon_{it}] = 0$ for all i, t, s, any method that would either instrument or remove c_i (e.g. by first-differencing or demeaning of x_{it}) would yield consistent estimators.

To our best knowledge, we are the first to provide a comparison between three different estimators in the presence of weak time-dependence in ϵ_{it}.

With no breaks, estimating the demeaned (1) is known to yield the most efficient estimators of the slope parameters - see Wooldridge (2002).[6] In the presence of a break, the equivalent fixed effects estimator would be to *demean* (1) *over sub-samples* $[1, k^0]$ and $[k^0+1, T]$ and estimate this transformed model. This is our first estimator (FE), which replaces k^0 by $\widehat{k}$. Even when the errors ϵ_{it} are iid, we show that another two estimators can be more efficient. These two new estimators are described below and are specific to the presence of breaks.

Our second estimator (IV) *demeans the regressors over the full-sample*, and uses them as instruments in the initial equation. Our third estimator (FFE) *demeans* (1) *over the full-sample*, then *estimates the transformed model over the full-sample*. Below, we describe these estimators and compare their asymptotic properties.

Additional notation. Define $I_1^0 = \{1, \ldots, k^0\}$, $I_2^0 = \{k^0+1, \ldots, T\}$, $\widehat{I}_1 = \{1, \ldots, \widehat{k}\}$ and $\widehat{I}_2 = \{\widehat{k}+1, \ldots, T\}$. Let $\sum_1 = \sum_{t=1}^{k^0}$ and $\sum_2 = \sum_{t=k^0+1}^{T}$ be the summation over, respectively, the first and the second true regimes. Let $\sum_{\widehat{1}} = \sum_{t=1}^{\widehat{k}}$ and $\sum_{\widehat{2}} = \sum_{t=\widehat{k}+1}^{T}$ be the summation over, respectively, the first and the second estimated regimes. Let $\overline{\boldsymbol{x}}_i = T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_{it}$ be the full sample average, $\overline{\boldsymbol{x}}_{i,1}^0 = (k^0)^{-1}\sum_1 \boldsymbol{x}_{it}$ and $\overline{\boldsymbol{x}}_{i,2}^0 = (T-k^0)^{-1}\sum_2 \boldsymbol{x}_{it}$ be the subsample averages over the two true regimes, and $\overline{\boldsymbol{x}}_{i,1} = \widehat{k}^{-1}\sum_{\widehat{1}} \boldsymbol{x}_{it}$ and $\overline{\boldsymbol{x}}_{i,2} = (T-\widehat{k})^{-1}\sum_{\widehat{2}} \boldsymbol{x}_{it}$ be the corresponding subsample averages over the two estimated regimes. Define $\lambda^0 = k^0/T$. Let $y_{it}^* = y_{it} - \overline{y}_i$, $\epsilon_{it}^* = \epsilon_{it} - \overline{\epsilon}_i$, $\widetilde{\boldsymbol{x}}_{it}^a = \boldsymbol{x}_{it}1_{t\le k^0} - \lambda^0\overline{\boldsymbol{x}}_{i,1}^0$, $\widetilde{\boldsymbol{x}}_{it}^b = \boldsymbol{x}_{it}1_{t\ge k^0+1} - (1-\lambda^0)\overline{\boldsymbol{x}}_{i,2}^0$, and $\widetilde{\boldsymbol{x}}_{it} = (\widetilde{\boldsymbol{x}}_{it}^{a\prime}, \widetilde{\boldsymbol{x}}_{it}^{b\prime})'$.

FE Estimator. To define the first estimator, the subsample fixed effects (FE) estimator, note that by Assumption 1, $\mathbb{E}[\boldsymbol{x}_{it}u_{it}] = \boldsymbol{a}_i \neq 0$, but $\mathbb{E}[(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{ij}^0)u_{it}] = 0$ for all t and for $j = 1, 2$. Thus we can instrument $\boldsymbol{x}_{it}$ in (1) with $(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1})$ for $t = 1, \ldots, \widehat{k}$, and $(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,2})$ for $t = \widehat{k}+1, \ldots, T$, resulting in the usual FE estimators, applied over the two subsamples rather than the entire sample. The FE population moment conditions, the sample moment conditions[7] and the slope

[6] We do not consider the first-differenced estimators in Qian and Su (2016) because they are not the most efficient estimators of $\boldsymbol{\beta}^0$ unless ϵ_{it} are unit root, a case we exclude from our analysis.

[7] This idea of moment conditions can be generalized to include more lags as instruments, in a way similar to Arellano and Bond (1991). For example, the difference $x_{it} - x_{is}$ is also an instrument. More instruments lead to efficiency gains, but on the other hand, too many instruments lead to poor finite sample properties.

estimators, for $j = 1, 2$, are:

$$\text{Population moments: } \mathbb{E}[(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}^0_{i,j})u_{it}] = \mathbf{0}, \text{ for } t \in I^0_j, \tag{6}$$

$$\text{Sample moments: } \boldsymbol{g}_{FE,j}(\boldsymbol{\beta}) = N^{-1} \sum_{i=1}^{N} \sum_{\widehat{j}} (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,j})(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta}_j) = \mathbf{0}, \tag{7}$$

$$\text{Slope estimators: } \widehat{\boldsymbol{\beta}}_{FE,j} = \left(\sum_{i=1}^{N} \sum_{\widehat{j}} (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,j})\boldsymbol{x}'_{it}\right)^{-1} \sum_{i=1}^{N} \sum_{\widehat{j}} (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,j})y_{it}. \tag{8}$$

IV Estimator. The instruments $(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,j})$ are not the only valid instruments for $\boldsymbol{x}_{it}$; for example, so are $(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)$. Our IV estimator is based on these instruments. The corresponding instrumental variables moment conditions and estimators, for $j = 1, 2$, are:

$$\text{Population moments: } \mathbb{E}[(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)u_{it}] = \mathbf{0}, \text{ for } t \in I^0_j, \tag{9}$$

$$\text{Sample moments: } \boldsymbol{g}_{IV,j}(\boldsymbol{\beta}) = N^{-1} \sum_{i=1}^{N} \sum_{\widehat{j}} (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta}_j) = \mathbf{0}, \tag{10}$$

$$\text{Slope estimators: } \widehat{\boldsymbol{\beta}}_{IV,j} = \left(\sum_{i=1}^{N} \sum_{\widehat{j}} (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)\boldsymbol{x}'_{it}\right)^{-1} \sum_{i=1}^{N} \sum_{\widehat{j}} (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)y_{it}. \tag{11}$$

FFE Estimator. Alternatively, we could also demean the data over the full sample, resulting in waht we call the full-sample fixed effects (FFE) estimator. Rewrite model (1) as

$$y_{it} - \overline{y}_i = \begin{cases} (\boldsymbol{x}_{it} - \lambda^0 \overline{\boldsymbol{x}}^0_{i,1})'\boldsymbol{\beta}^0_1 - (1 - \lambda^0)\overline{\boldsymbol{x}}^{0\prime}_{i,2}\boldsymbol{\beta}^0_2 + \epsilon_{it} - \overline{\epsilon}_i, & t \le k^0; \\ -\lambda^0 \overline{\boldsymbol{x}}^{0\prime}_{i,1}\boldsymbol{\beta}^0_1 + (\boldsymbol{x}_{it} - (1 - \lambda^0)\overline{\boldsymbol{x}}^0_{i,2})'\boldsymbol{\beta}^0_2 + \epsilon_{it} - \overline{\epsilon}_i, & t > k^0. \end{cases}$$

This model can be written more compactly as:

$$y^*_{it} = \widetilde{\boldsymbol{x}}^{a\prime}_{it}\boldsymbol{\beta}^0_1 + \widetilde{\boldsymbol{x}}^{b\prime}_{it}\boldsymbol{\beta}^0_2 + \epsilon^*_{it} = \widetilde{\boldsymbol{x}}'_{it}\boldsymbol{\beta}^0 + \epsilon^*_{it},$$

where $y^*_{it} = y_{it} - \overline{y}_i$, $\overline{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$, $\epsilon^*_{it} = \epsilon_{it} - \overline{\epsilon}_i$ and $\overline{\epsilon}_i = T^{-1} \sum_{t=1}^{T} \epsilon_{it}$. When k^0 is known, this model can be estimated via OLS with joint estimation of β^0_1, β^0_2, leading to the FFE estimator:

$$\text{Population moments: } \mathbb{E}(\widetilde{\boldsymbol{x}}_{it}\ \epsilon^*_{it}) = \mathbf{0}, \tag{12}$$

$$\text{Sample moments: } \boldsymbol{g}_{IV}(\boldsymbol{\beta}) = N^{-1} \sum_{i=1}^{N} \sum_{\widehat{j}} \widetilde{\boldsymbol{x}}_{it}(y_{it} - \widetilde{\boldsymbol{x}}'_{it}\boldsymbol{\beta}) = \mathbf{0}, \tag{13}$$

$$\text{Slope estimators: } \widehat{\boldsymbol{\beta}}^0_{FFE} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{x}_{it}\widetilde{x}'_{it}\right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{x}_{it}y^*_{it}\right). \tag{14}$$

When k^0 is unknown, we replace k^0 with $\widehat{k}$ in the definition of $\widetilde{x}_{it}$. This leads to the full-sample driven fixed effects (FFE) estimator, which we denote by $\widehat{\beta}_{FFE}$. By Theorem 1, $P(\widehat{k} = k^0) \to 1$ as $N \to \infty$, so $\widehat{\beta}^0_{FFE}$ and $\widehat{\beta}_{FFE}$ share the same asymptotic properties (see Theorem 5).

5.3 Asymptotic Distributions of Slope Estimators

We now derive the asymptotic properties of the FE, IV and FFE estimators when $N \to \infty$. We need the following additional assumptions:

Assumption 3.
(i) $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathbb{E}[\boldsymbol{x}_{it}\boldsymbol{x}'_{is}] = \boldsymbol{\Omega}_{t,s}$ *is finite, for all* t,s;
(ii) $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathbb{E}\left[\boldsymbol{w}_i\boldsymbol{w}'_i\right] = \overline{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}$ *is finite, and*$\boldsymbol{w}_i = \left(\sum_1(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}^0_{i,1})'\epsilon_{it}, \sum_2(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}^0_{i,2})'\epsilon_{it}\right)'$;
(iii) $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathbb{E}\left[\boldsymbol{v}_i\boldsymbol{v}'_i\right] = \overline{\boldsymbol{W}}_{\boldsymbol{x}u}$ *is finite, and* $\boldsymbol{v}_i = \left(\sum_1(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)'u_{it}, \sum_2(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)'u_{it}\right)'$;
(iv) $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathbb{E}\left[\left(\sum_{t=1}^{T}\widetilde{\boldsymbol{x}}_{it}\epsilon^*_{it}\right)\left(\sum_{t=1}^{T}\widetilde{\boldsymbol{x}}_{it}\epsilon^*_{it}\right)'\right] = \widetilde{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}$ *is finite;*
(v) $k^0 \in \{2,\ldots,T-2\}$.

These assumptions merely state the existence of moments in the presence of general weak time series dependence in x_{it},ϵ_{it}. The following three theorems state the asymptotic distributions of the estimators we analyze.

Theorem 3. *[FE Estimators]*
Under Assumptions 1-3, when $N \to \infty$ *with* T *is fixed,*

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}^0\right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{V}_{FE}),$$

where $\boldsymbol{V}_{FE} = \boldsymbol{\Omega}_1^{-1}\overline{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}\boldsymbol{\Omega}_1^{-1}$, *with* $\boldsymbol{\Omega}_1$ *a block diagonal matrix, with block diagonal elements* $k^0\boldsymbol{Q} - {k^0}^{-1}\sum_{t,s=1}^{k^0}\boldsymbol{\Omega}_{t,s}$ *and* $(T-k^0)\boldsymbol{Q} - (T-k^0)^{-1}\sum_{t,s=k^0+1}^{T}\boldsymbol{\Omega}_{t,s}$, *respectively.*

Theorem 4. *[IV Estimators]*
Under Assumptions 1-3, when $N \to \infty$ *and* T *is fixed,*

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}^0) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{V}_{IV}\right),$$

where $\boldsymbol{V}_{IV} = \boldsymbol{\Omega}_2^{-1}\overline{\boldsymbol{W}}_{\boldsymbol{x}u}\boldsymbol{\Omega}_2^{-1}$, *with* $\boldsymbol{\Omega}_2$ *being a block diagonal matrix, with block diagonal elements* $k^0\boldsymbol{Q} - T^{-1}\sum_{t=1}^{k^0}\sum_{s=1}^{T}\boldsymbol{\Omega}_{t,s}$ *and* $(T-k^0)\boldsymbol{Q} - T^{-1}\sum_{t=k^0+1}^{T}\sum_{s=1}^{T}\boldsymbol{\Omega}_{t,s}$, *respectively.*

Theorem 5. *[FFE Estimators]*

Under Assumptions 1-3, when $N \to \infty$ and T is fixed,

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{FFE} - \boldsymbol{\beta}^0) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{V}_{FFE}\right),$$

where $V_{FFE} = \boldsymbol{\Omega}_3^{-1}\widetilde{\boldsymbol{W}}_{x\epsilon}\boldsymbol{\Omega}_3^{-1}$, and

$$\boldsymbol{\Omega}_3 = \begin{pmatrix} k^0 Q - T^{-1}\sum_{t,s=1}^{k^0} \boldsymbol{\Omega}_{t,s} & -T^{-1}\sum_{t=k^0+1}^{T}\sum_{s=1}^{k^0} \boldsymbol{\Omega}_{t,s} \\ -T^{-1}\sum_{s=k^0+1}^{T}\sum_{t=1}^{k^0} \boldsymbol{\Omega}_{t,s} & (T-k^0)Q - T^{-1}\sum_{t,s=k^0+1}^{T} \boldsymbol{\Omega}_{t,s} \end{pmatrix}.$$

5.4 Relative Efficiency of Slope Estimators

From the previous section, it is unclear which estimator is asymptotically more efficient. This might seem surprising, but this result arises because of the interaction of time-dependent errors and breaks. With no further assumptions on serial dependence, the data-generating process is too general for any meaningful comparisons. In that case, we suggest stacking the moment conditions from these three estimators and perform an optimal GMM estimation.

Below, we compare the asymptotic variances of the three estimators under the following common independence assumptions:

- ϵ_{it} is independent over t;
- ϵ_{it} is independent of x_{is} for any t, s;
- $\boldsymbol{\Omega}_{t,s} = \boldsymbol{Q}$ for $t = s$, and $\boldsymbol{\Omega}_{t,s} = \boldsymbol{\Omega}^*$ otherwise.

Let $\widetilde{\boldsymbol{W}}_{t,s} = \mathbb{E}[\boldsymbol{x}_{it}\boldsymbol{x}_{is}' c_i^2]$. Under these conditions, we show in the appendix that

$$\boldsymbol{V}_{FE} = \sigma_\epsilon^2 \left(\begin{pmatrix} \boldsymbol{C}_1 & 0 \\ 0 & \boldsymbol{C}_2 \end{pmatrix} - \begin{pmatrix} (1-\lambda^0)(\boldsymbol{Q}-\boldsymbol{\Omega}^*) & 0 \\ 0 & \lambda^0(\boldsymbol{Q}-\boldsymbol{\Omega}^*) \end{pmatrix} \right)^{-1}$$

$$\boldsymbol{V}_{IV} = \sigma_\epsilon^2 \begin{pmatrix} \boldsymbol{C}_1 & 0 \\ 0 & \boldsymbol{C}_2 \end{pmatrix}^{-1} + \begin{pmatrix} \boldsymbol{C}_1^{-1}\boldsymbol{B}\boldsymbol{C}_1^{-1} & -\boldsymbol{C}_1^{-1}\boldsymbol{B}\boldsymbol{C}_2^{-1} \\ -\boldsymbol{C}_2^{-1}\boldsymbol{B}\boldsymbol{C}_1^{-1} & \boldsymbol{C}_2^{-1}\boldsymbol{B}\boldsymbol{C}_2^{-1} \end{pmatrix},$$

$$\boldsymbol{V}_{FFE} = \sigma_\epsilon^2 \left(\begin{pmatrix} \boldsymbol{C}_1 & 0 \\ 0 & \boldsymbol{C}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{D} & -\boldsymbol{D} \\ -\boldsymbol{D} & \boldsymbol{D} \end{pmatrix} \right)^{-1},$$

where

$$\boldsymbol{B} = (1-\lambda^0)^2 \sum_{t,s=1}^{k^0} \widetilde{\boldsymbol{W}}_{t,s} - 2\lambda^0(1-\lambda^0) \sum_{t=1}^{k^0} \sum_{s=k^0+1}^{T} \widetilde{\boldsymbol{W}}_{t,s} + (\lambda^0)^2 \sum_{t,s=k^0+1}^{T} \widetilde{\boldsymbol{W}}_{t,s},$$

$$C_1 = \lambda^0(T-1)(\boldsymbol{Q}-\boldsymbol{\Omega}^*), \qquad C_2 = (1-\lambda^0)(T-1)(\boldsymbol{Q}-\boldsymbol{\Omega}^*), \qquad D = \lambda^0(T-k^0)\boldsymbol{\Omega}^*.$$

This is the setting of the simulation section 6, but it seems that the precision of estimators still cannot be compared theoretically without further assumptions[8]. In the special case of $\boldsymbol{\Omega}^* = 0$ (regressors are uncorrelated over time), FFE is more efficient than FE because $\boldsymbol{D} = \boldsymbol{0}$. This result shows that there are more efficient estimators than the usual FE estimators in the presence of a break. In other words, it is better to demean over the full-sample rather than over subsamples, because in doing so, additional information is used about the relative size of the two subsamples. If $\boldsymbol{\Omega}^* = 0$ and $\boldsymbol{x}_{it}$ is independent of c_i (random effects), then $\boldsymbol{V}_{IV} = \boldsymbol{V}_{FFE} < \boldsymbol{V}_{FE}$[9], so both FFE and IV are asymptotically more efficient than FE. This shows that both FFE and IV perform some type of quasi-demeaning either on the instruments or on the level equation.

6 Simulation Study

This section studies the finite sample properties of our estimators.[10] Specifically, we study the finite sample performance of $\widehat{k}$, $\widehat{\boldsymbol{\beta}}_{OLS}$, $\widehat{\boldsymbol{\beta}}_{FE}$, $\widehat{\boldsymbol{\beta}}_{IV}$ and $\widehat{\boldsymbol{\beta}}_{FFE}$. We generate ϵ_{it} iid $\mathcal{N}(0, 0.5^2)$ over i and t. c_i is generated independent of ϵ_{it}, and is iid $\mathcal{N}(0, 0.5^2)$. We generate x_{it} as $x_{it} = \sqrt{2}c_i + z_{it}$, where $z_{it} \sim iid\mathcal{N}(0, 0.5)$, independent over i and t. The slope parameters $(\boldsymbol{\beta}_1^0,\ \boldsymbol{\beta}_2^0) = (-0.1,\ 0.1)$. We consider two cases for the true break point: $k^0 = [T/3]$ and $k^0 = 2$, where $[\cdot]$ is the least integer function. In all cases the results are reported for 10,000 replications.

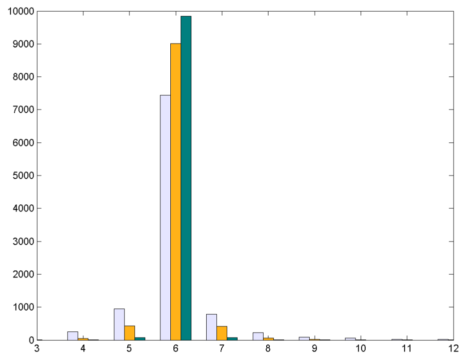
In Table 1 and 2, we report the MC averages and standard errors (in parentheses) for the break point estimators. In Table 1, when $k^0 = [T/3]$, the numbers of time periods before and after the break are more balanced, and the estimates for the break point are very precise. In Table 2, when $k^0 = 2$, the number of time periods in the first sub-sample is small. When $N = 50$, the sample size is not big enough, and $\widehat{k}$ has a big positive bias, but this decreases as N increases to 100 or 200, as we expect.

[8] If $\boldsymbol{\Omega}^*$ is not positive definite, then V_{FFE} cannot be compared with V_{FE}.

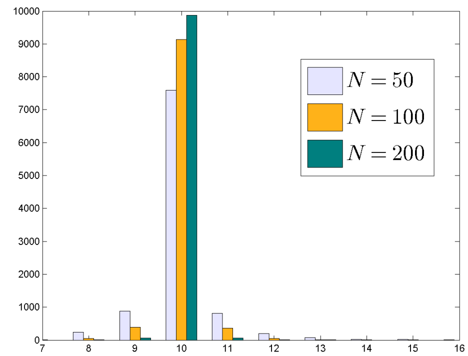
[9] For two positive definite matrices A and B, $A > B$ means $A - B$ is positive definite.

[10] The current simulation study assumes that the number of breaks is known and equal to one. Further simulations with multiple breaks are in progress.

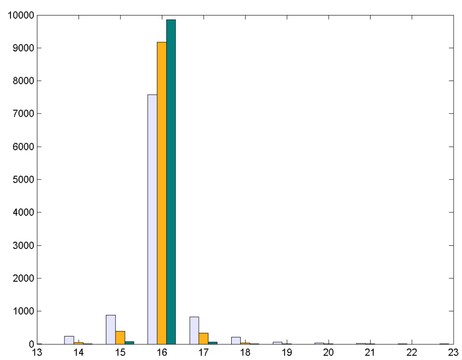
Figure 1: Histogram of $\widehat{k}$ when $k^0 = [T/3]$



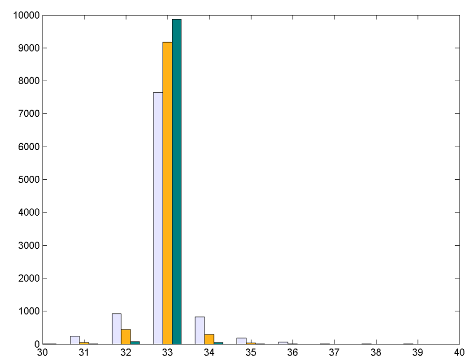
$T = 20$, $k^0 = 6$



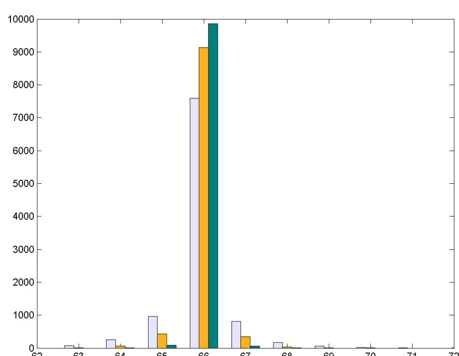
$T = 30$, $k^0 = 10$



$T = 50$, $k^0 = 16$

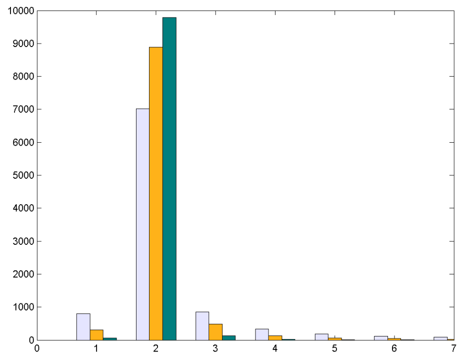


$T = 100$, $k^0 = 33$

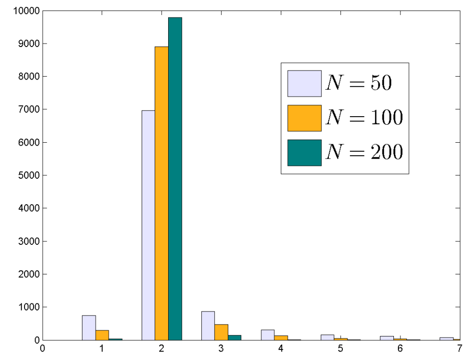


$T = 200$, $k^0 = 66$

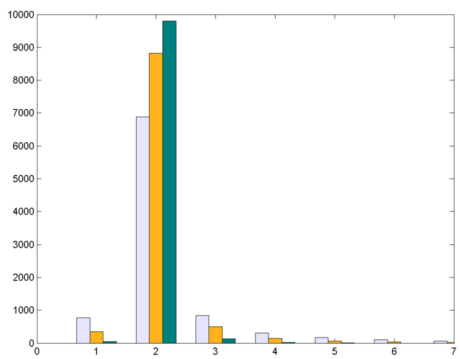
Figure 2: Histogram of $\widehat{k}$ when $k^0 = 2$



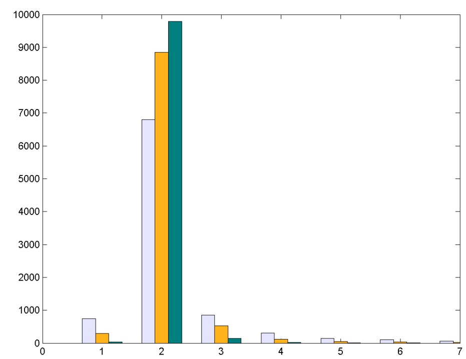
$T = 20$



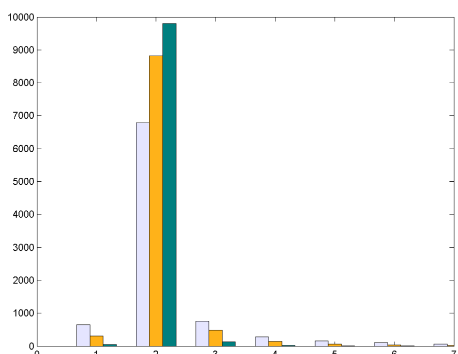
$T = 30$



$T = 50$



$T = 100$



$T = 200$

Table 1: (MC Average) break point estimates, when $k^0 = [T/3]$

		$N = 50$	$N = 100$	$N = 200$
$T = 20$	$k^0 = 6$	6.041	6.010	6.001
$T = 30$	$k^0 = 10$	10.000	9.995	10.000
$T = 50$	$k^0 = 16$	15.983	15.992	15.998
$T = 100$	$k^0 = 33$	32.974	32.980	32.996
$T = 200$	$k^0 = 66$	65.953	65.986	65.997

Table 2: (MC Average) break point estimates, when $k^0 = 2$

		$N = 50$	$N = 100$	$N = 200$
$T = 20$	$k^0 = 2$	2.921	2.172	2.012
$T = 30$	$k^0 = 2$	3.504	2.201	2.021
$T = 50$	$k^0 = 2$	4.407	2.254	2.017
$T = 100$	$k^0 = 2$	7.037	2.521	2.020
$T = 200$	$k^0 = 2$	13.519	3.026	2.018

In Figure 1 and 2, we plot histograms of $\widehat{k}$ around k^0. The histograms agree with the result in Table 1 that our break point estimator performs well, both for a small T and a large T.

In Table 3 and 4 we report the bias and standard error of slope estimates based on the four methods: OLS, FE, IV and FFE. The standard errors are found using the formula in Section 5. We do not report the standard errors of the OLS estimators, because their standard errors cannot be computed based on Theorem 2. We find that the OLS estimators have a relatively large bias; this is not surprising under endogeneity of the regressors in OLS. It can be seen from Table 3 that FE, IV and FFE all work well for $k^0 = [T/3]$. In Table 4, when $N = 50$, there is a positive bias for FE, IV and FFE in the first regime. In this case, $\widehat{k}$ has a large positive bias, which means the first estimated sub-sample contains many observations from the second true regime, leading to the bias. We also find that the bias of the IV estimator is smaller than that of the other two, but its standard error is larger. Also, we observe that the FFE estimator has indeed smaller standard errors than the FE estimator when $N = 50$ or 100; with $N = 200$, these differences vanish.

Table 3: Bias and standard errors of the slope parameter estimates, $k^0 = [T/3]$

	Method	$N = 50$		$N = 100$		$N = 200$	
		$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
$T = 20$	OLS	0.348	0.351	0.352	0.352	0.353	0.353
	FE	-0.001 (0.045)	0.001 (0.028)	-0.000 (0.032)	0.000 (0.020)	-0.000 (0.022)	0.000 (0.014)
	IV	-0.003 (0.056)	0.000 (0.032)	-0.000 (0.039)	0.000 (0.022)	-0.001 (0.028)	0.000 (0.016)
	FFE	-0.001 (0.042)	0.001 (0.028)	-0.000 (0.030)	0.000 (0.019)	-0.000 (0.021)	0.000 (0.014)
$T = 30$	OLS	0.350	0.351	0.352	0.352	0.352	0.353
	FE	-0.000 (0.034)	0.000 (0.023)	-0.001 (0.024)	0.000 (0.016)	-0.000 (0.017)	0.000 (0.011)
	IV	-0.001 (0.042)	0.000 (0.026)	-0.001 (0.029)	0.000 (0.019)	-0.001 (0.021)	0.000 (0.013)
	FFE	-0.000 (0.032)	0.000 (0.023)	-0.000 (0.023)	0.000 (0.016)	-0.000 (0.016)	0.000 (0.011)
$T = 50$	OLS	0.350	0.351	0.352	0.352	0.353	0.353
	FE	-0.001 (0.026)	0.000 (0.017)	-0.000 (0.018)	0.000 (0.012)	-0.000 (0.013)	0.000 (0.009)
	IV	-0.001 (0.033)	0.000 (0.020)	-0.000 (0.023)	0.000 (0.014)	-0.000 (0.016)	0.000 (0.010)
	FFE	-0.000 (0.025)	0.000 (0.017)	-0.000 (0.018)	0.000 (0.012)	-0.000 (0.013)	0.000 (0.009)
$T = 100$	OLS	0.350	0.350	0.352	0.352	0.353	0.353
	FE	-0.000 (0.018)	0.000 (0.012)	-0.000 (0.013)	0.000 (0.009)	-0.000 (0.009)	0.000 (0.006)
	IV	-0.001 (0.022)	0.000 (0.014)	-0.000 (0.016)	0.000 (0.010)	-0.000 (0.011)	0.000 (0.007)
	FFE	-0.000 (0.018)	0.000 (0.012)	-0.000 (0.012)	0.000 (0.009)	-0.000 (0.009)	0.000 (0.006)
$T = 200$	OLS	0.350	0.350	0.351	0.351	0.352	0.352
	FE	-0.000 (0.012)	0.000 (0.009)	-0.000 (0.009)	0.000 (0.006)	-0.000 (0.006)	0.000 (0.004)
	IV	-0.000 (0.016)	0.000 (0.010)	-0.000 (0.011)	0.000 (0.007)	-0.000 (0.008)	0.000 (0.005)
	FFE	-0.000 (0.012)	0.000 (0.009)	-0.000 (0.009)	0.000 (0.006)	-0.000 (0.006)	0.000 (0.004)

True parameter values: $(\beta_1^0,\ \beta_2^0) = (-0.1,\ 0.1)$.

Table 4: Bias and standard errors of the slope parameter estimates, $k^0 = 2$

	Method	$N=50$		$N=100$		$N=200$	
		$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
$T=20$	OLS	0.354	0.354	0.353	0.353	0.353	0.353
	FE	0.020 (0.088)	0.001 (0.026)	0.006 (0.068)	0.001 (0.017)	0.001 (0.050)	0.000 (0.012)
	IV	0.003 (0.100)	0.002 (0.031)	-0.001 (0.072)	0.001 (0.018)	-0.002 (0.051)	0.000 (0.013)
	FFE	0.007 (0.071)	0.003 (0.027)	0.002 (0.051)	0.001 (0.017)	0.000 (0.036)	0.000 (0.012)
$T=30$	OLS	0.357	0.354	0.353	0.352	0.353	0.353
	FE	0.020 (0.087)	0.001 (0.021)	0.005 (0.068)	0.000 (0.014)	0.000 (0.050)	0.000 (0.010)
	IV	0.004 (0.099)	0.002 (0.027)	-0.001 (0.071)	0.000 (0.014)	-0.001 (0.050)	0.000 (0.010)
	FFE	0.009 (0.070)	0.002 (0.023)	0.002 (0.051)	0.000 (0.014)	0.000 (0.036)	0.000 (0.010)
$T=50$	OLS	0.357	0.352	0.353	0.352	0.354	0.353
	FE	0.023 (0.087)	0.001 (0.017)	0.007 (0.068)	0.000 (0.010)	0.001 (0.050)	0.000 (0.007)
	IV	0.006 (0.098)	0.001 (0.021)	-0.001 (0.071)	0.000 (0.011)	-0.002 (0.050)	0.000 (0.007)
	FFE	0.010 (0.069)	0.001 (0.018)	0.002 (0.050)	0.000 (0.011)	0.001 (0.036)	0.000 (0.007)
$T=100$	OLS	0.358	0.351	0.354	0.352	0.354	0.353
	FE	0.025 (0.086)	0.000 (0.013)	0.008 (0.068)	0.000 (0.007)	0.001 (0.050)	0.000 (0.005)
	IV	0.006 (0.097)	0.001 (0.017)	-0.001 (0.071)	0.000 (0.008)	-0.001 (0.050)	0.000 (0.005)
	FFE	0.012 (0.068)	0.001 (0.014)	0.003 (0.050)	0.000 (0.007)	0.001 (0.035)	0.000 (0.005)
$T=200$	OLS	0.362	0.351	0.353	0.351	0.353	0.352
	FE	0.026 (0.086)	0.000 (0.011)	0.006 (0.068)	0.000 (0.005)	0.001 (0.050)	0.000 (0.004)
	IV	0.011 (0.095)	0.000 (0.015)	-0.002 (0.070)	0.000 (0.006)	-0.001 (0.050)	0.000 (0.004)
	FFE	0.015 (0.067)	0.000 (0.012)	0.002 (0.050)	0.000 (0.005)	0.000 (0.035)	0.000 (0.004)

True parameter values: $(\beta_1^0,\ \beta_2^0) = (-0.1,\ 0.1)$.

7 Conclusion

In this paper, we propose OLS estimators of multiple break-points for fixed effects panel data regression models. We show that these are consistent for the true break-points as $N \to \infty$, subject to a time homogeneity condition. We propose estimating the break-points via an information criterion. Furthermore, we propose three fixed effects estimators, based on one estimated break point. All three estimators are shown to be consistent and asymptotically normal with the same $\sqrt{N}$ convergence rate. Their finite sample properties are investigated by simulation, and it appears that for iid data, the FFE estimators perform best.

Bibliography

Arellano, M., and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies* 58, 277-297.

Baltagi, B.H., Kao, C. and Liu, L. (2015). Estimation and identification of change points in panel models with nonstationary or stationary regressors and error term, *Econometric Reviews.*

Baltagi, B.H., Feng, Q. and Kao, C. (2016). Estimation of heterogeneous panels with structural breaks, *Journal of Econometrics* 191, 176-195.

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

Emerson, J. and 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. and 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.

Feng, Q., Kao, C. and Lazarová, S. (2009). Estimation of change points in panel models, *Working Paper.*

Hall, A.R., Osborn, D. and Sakkas, N. (2013). Inference about structural breaks using information criteria, *The Manchester School* 81, 54-81.

Li, D., Qian, J. and Su, L. (2015). Panel Data Models with Interactive Fixed Effects and Multiple Structural Breaks *Journal of the American Statistical Association*, forthcoming.

Nimomiya, Y. (2005). Information criterion for Gaussian change-point model. *Statistics and Probability Letters* 72, 237247.

Perron, P. and Yamamoto, Y. (2013). Using OLS to estimate and test for structural changes in models with endogenous regressors. *Journal of Applied Econometrics* 28, 1099–1255.
Qian, J. and Su, L. (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused LASSO. *Journal of Econometrics* 191, 86-109.
van der Vaart, A. (2000). Asymptotic Statistics, vol. 3, Cambridge University Press.
de Wachter, S. and Tsavalis, E. (2012). Detection of structural breaks in linear dynamic panel data models, *Computational Statistics and Data Analysis* 56, 3020–3034.
Wooldridge, J.M. (2002). Econometric analysis of cross section and panel data, 1^{st} Ed., MIT Press.

Appendix A

All the proofs are done below for one break point; the extension to multiple break-points is straightforward and ommitted for simplicity. Therefore, we let $k^0 \stackrel{\text{def}}{=} T_1^0$, $k \stackrel{\text{def}}{=} T_1$, and we substitute dependence on partition $\boldsymbol{T}_{m^0+1}$ with dependence on k.

Proof of Lemma 1. For all four results (i)-(iv), the proof uses the WLLN, which is in Appendix B. The independence requirement is satisfied for all four summands, due to Assumption 1(i). Further, by Assumption 1, the difference between the summands and their limits have zero mean. Thus, we only need to check that for some $\xi > 0$, the $1+\xi$ moments of the summands exist.

Firstly, since $\boldsymbol{x}_{it}\boldsymbol{x}_{it}'$ is a $p \times p$ matrix, we apply the WLLN to each column of $\boldsymbol{x}_{it}\boldsymbol{x}_{it}'$. Let $\boldsymbol{x}_{it,q}$ be the q-th element in $\boldsymbol{x}_{it}$. The q-th column in $\boldsymbol{x}_{it}\boldsymbol{x}_{it}'$ is thus $\boldsymbol{x}_{it}\boldsymbol{x}_{it,q}$, with $E|\boldsymbol{x}_{it}x_{it,q}|^{1+\xi/2} = \left(\|\boldsymbol{x}_{it}x_{it,q}\|_{1+\xi/2}\right)^{1+\xi/2} \leq (\|\boldsymbol{x}_{it}\|_{2+\xi}\|x_{it,q}\|_{2+\xi})^{1+\xi/2}$ by Cauchy-Schwarz inequality. Take the supremum over $i \in \mathbb{N}$ on both sides and we have $\sup_{i\in\mathbb{N}} E|\boldsymbol{x}_{it}x_{it,q}|^{1+\xi/2} < \infty$ by Assumption 1(ii). Thus $N^{-1}\sum_{i=1}^N \boldsymbol{x}_{it}x_{it,q}$ converges in probability to the q-th column of $\boldsymbol{Q}$, and $N^{-1}\sum_{i=1}^N \boldsymbol{x}_{it}\boldsymbol{x}_{it}' \xrightarrow{p} \boldsymbol{Q}$.

For all the remaining three results it is easy to follow the same procedure in the previous paragraph and show that the required moment conditions in the WLLN are satisfied. Thus all requirements in the WLLN are satisfied and we have $N^{-1}\sum_{i=1}^N \boldsymbol{x}_{it}c_i \xrightarrow{p} \boldsymbol{a}$, $N^{-1}\sum_{i=1}^N \boldsymbol{x}_{it}\epsilon_{it} \xrightarrow{p} \boldsymbol{0}$ and $N^{-1}\sum_{i=1}^N u_{it}^2 \xrightarrow{p} \sigma_\epsilon^2 + \sigma_c^2$. □

Proof of Theorem 1. Part (i). Assume first that $k \le k^0$. Start by noting that:

$$\begin{aligned}
(NT)S_{NT}(k) &= (\boldsymbol{y} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}_{OLS}(k))' \, (\boldsymbol{y} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}_{OLS}(k)) \\
&= (\widetilde{\boldsymbol{X}}^0\boldsymbol{\beta}^0 + \boldsymbol{u} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}_{OLS}(k))' \, (\widetilde{\boldsymbol{X}}^0\boldsymbol{\beta}^0 + \boldsymbol{u} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}_{OLS}(k)) \\
&= (\widetilde{\boldsymbol{X}}^0\boldsymbol{\beta}^0 + \boldsymbol{u})' \, (I - \widetilde{\boldsymbol{X}}(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}})(\widetilde{\boldsymbol{X}}^0\boldsymbol{\beta}^0 + \boldsymbol{u}) \\
S_{NT}(k) &= (NT)^{-1}\boldsymbol{u}'\boldsymbol{u} + \boldsymbol{\beta}^{0\prime}\left((NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}\widetilde{\boldsymbol{X}}^0 - (NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}\widetilde{\boldsymbol{X}}((NT)^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}(NT)^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}^0\right)\boldsymbol{\beta}^0 \\
&\quad + 2\boldsymbol{\beta}^{0\prime}\left((NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}\boldsymbol{u} - (NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}\widetilde{\boldsymbol{X}}((NT)^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}(NT)^{-1}\widetilde{\boldsymbol{X}}'u\right) \\
&\quad - (NT)^{-1}\boldsymbol{u}'\widetilde{\boldsymbol{X}}((NT)^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}(NT)^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{u}. \qquad (15)
\end{aligned}$$

Note that since T is fixed, division by T through the proof of this theorem is not needed, but it doesn't change the results, and makes the proof clearer. Note in the above equation that the proof of the convergence of $S_{NT}(k)$ only involves four different terms: $(NT)^{-1}\boldsymbol{u}'\boldsymbol{u}$, $(NT)^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}$, $(NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}\widetilde{\boldsymbol{X}}$, and $(NT)^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{u}$. We now show that the convergence of the four terms follows from Lemma 1. Let $\boldsymbol{D}_k = \begin{pmatrix} k/T & 0 \\ 0 & (T-k)/T \end{pmatrix}$, $\boldsymbol{D}_\Delta = \begin{pmatrix} k/T & 0 \\ (k^0-k)/T & (T-k^0)/T \end{pmatrix}$ and $\boldsymbol{\ell}_q$ be a vector of ones with length q.

Firstly, in Lemma 1(iv), it is shown that $(NT)^{-1}\boldsymbol{u}'\boldsymbol{u} \xrightarrow{p} \sigma_\epsilon^2 + \sigma_c^2$. Next,

$$(NT)^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}} = \begin{pmatrix} T^{-1}\sum_{t=1}^{k}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right) & \boldsymbol{0} \\ \boldsymbol{0} & T^{-1}\sum_{t=k+1}^{T}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right) \end{pmatrix},$$

$$(NT)^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}^0 = \begin{pmatrix} T^{-1}\sum_{t=1}^{k}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right) & \boldsymbol{0} \\ T^{-1}\sum_{t=k+1}^{k_0}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right) & T^{-1}\sum_{t=k_0+1}^{T}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right) \end{pmatrix}.$$

By Lemma 1(i), the above two terms converge in probability to $\boldsymbol{D}_k \otimes \boldsymbol{Q}$ and $\boldsymbol{D}_\Delta \otimes \boldsymbol{Q}$ respectively as $N \to \infty$, where $\otimes$ denotes the Kronecker product of two matrices. Lastly,

$$(NT)^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{u} = \begin{pmatrix} T^{-1}\sum_{t=1}^{k}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_{it}(\epsilon_{it}+c_i)\right) \\ T^{-1}\sum_{t=k+1}^{T}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_{it}(\epsilon_{it}+c_i)\right) \end{pmatrix}.$$

By Lemma 1(ii) and (iii), for $\boldsymbol{\ell}_2 = (1 \;\; 1)'$ the above term converges in probability to $\boldsymbol{D}_k \cdot (\boldsymbol{\ell}_2 \otimes \boldsymbol{a})$.

Collecting the results in conjunction with equation (15), we have for $k \le k^0$,

$$\begin{aligned} S_{NT}(k) \xrightarrow{p} \sigma_\epsilon^2 + \sigma_c^2 + \boldsymbol{\beta}^{0\prime} \left(\boldsymbol{D}_{k^0} \otimes \boldsymbol{Q} - (\boldsymbol{D}_\Delta \otimes \boldsymbol{Q})'(\boldsymbol{D}_k \otimes \boldsymbol{Q})^{-1}(\boldsymbol{D}_\Delta \otimes \boldsymbol{Q}) \right) \\ + 2\boldsymbol{\beta}^{0\prime} \left(\boldsymbol{D}_k \cdot (\boldsymbol{\ell}_2 \otimes \boldsymbol{a}) - (\boldsymbol{D}_\Delta \otimes \boldsymbol{Q})'(\boldsymbol{D}_k \otimes \boldsymbol{Q})^{-1}\boldsymbol{D}_k \cdot (\boldsymbol{\ell}_2 \otimes \boldsymbol{a}) \right) \\ - (\boldsymbol{D}_k \cdot (\boldsymbol{\ell}_2 \otimes \boldsymbol{a}))'(\boldsymbol{D}_k \otimes \boldsymbol{Q})^{-1}\boldsymbol{D}_k \cdot (\boldsymbol{\ell}_2 \otimes \boldsymbol{a}). \end{aligned}$$

After simplification, it can be shown that the above term is equal to $\frac{(k^0-k)(T-k^0)}{(T-k)T}\boldsymbol{\delta}^{0\prime}\boldsymbol{Q}\ \boldsymbol{\delta}^0 + (\sigma_\epsilon^2 + \sigma_c^2 - \boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a})$.

A similar procedure can be carried out for $k \ge k^0$, in which case $S_{NT}(k) \xrightarrow{p} \frac{(k-k^0)k^0}{kT}\boldsymbol{\delta}^{0\prime}\boldsymbol{Q}\ \boldsymbol{\delta}^0 + (\sigma_\epsilon^2 + \sigma_c^2 - \boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a})$. This completes the proof for $S_{NT}(k) \xrightarrow{p} S^*_{NT}(k)$ uniformly in k.

Part (ii), (iii). To show that $P(\widehat{k} = k^0) \to 1$, note that $\boldsymbol{\delta}^{0\prime}\boldsymbol{Q}\boldsymbol{\delta}^0$ is strictly positive by Assumption 1(ii), so $S^*_{NT}(k)$ attains its unique minimum at k^0. Moreover, this function is defined over a discrete domain $k \in \{1, 2, \ldots, T-1\}$ for a fixed T. This means the probability of having a tie when minimizing (3) tends to zero. The consistency of $\widehat{k}$ follows using the same argument as in Theorem 5.7 in van der Vaart (2000). □

Proof of Theorem 2.

Let $\widehat{\boldsymbol{\beta}}^0_{OLS} := \widehat{\boldsymbol{\beta}}_{OLS}(k^0)$ be the OLS estimator for $\boldsymbol{\beta}^0$ using the true break point k^0. By Theorem 1 $P(\widehat{k} = k^0) \to 1$ as $N \to \infty$. Since $\widehat{k} = k^0$ implies $\widehat{\boldsymbol{\beta}}_{OLS} = \widehat{\boldsymbol{\beta}}^0_{OLS}$, $P(\widehat{\boldsymbol{\beta}}_{OLS} = \widehat{\boldsymbol{\beta}}^0_{OLS}) \to 1$. Thus $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}^0 - \Delta_{N,\boldsymbol{\beta}})$ and $\sqrt{N}(\widehat{\boldsymbol{\beta}}^0_{OLS} - \boldsymbol{\beta}^0 - \Delta_{N,\boldsymbol{\beta}})$ share the same limiting distribution, and to prove part (i) of the theorem it is sufficient to derive the limiting distribution of the latter.

By reexpressing $\Delta_{N,\boldsymbol{\beta}}$, we can show

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^0_{OLS} - \boldsymbol{\beta}^0 - \Delta_{N,\boldsymbol{\beta}}) = \left((NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}\widetilde{\boldsymbol{X}}^0 \right)^{-1} \sqrt{N} \left((NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}u - \boldsymbol{D}_{k^0} \cdot (\boldsymbol{\ell}_2 \otimes \boldsymbol{a}) \right),$$

where $\boldsymbol{D}_{k^0} = \begin{pmatrix} k^0/T & 0 \\ 0 & (T-k^0)/T \end{pmatrix}$. It follows from the proof of Theorem 1 that the term $(NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}\widetilde{\boldsymbol{X}}^0$ above converges in probability to $\boldsymbol{D}_{k^0} \otimes \boldsymbol{Q}$. Next,

$$\sqrt{N} \left((NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}u - \boldsymbol{D}_{k^0} \cdot (\boldsymbol{\ell}_2 \otimes \boldsymbol{a}) \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{k^0}(\boldsymbol{x}_{it}u_{it} - \boldsymbol{a}) \\ \frac{1}{T}\sum_{t=k^0+1}^{T}(\boldsymbol{x}_{it}u_{it} - \boldsymbol{a}) \end{pmatrix}.$$

By Assumption 1, the summands $\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{k^0}(\boldsymbol{x}_{it}u_{it}-\boldsymbol{a}) \\ \frac{1}{T}\sum_{t=k^0+1}^{T}(\boldsymbol{x}_{it}u_{it}-\boldsymbol{a}) \end{pmatrix}$ are independent over i with mean zero. The moment condition for applying the CLT is also satisfied, since $\sup_{i\in\mathbb{N}}\|\frac{1}{T}\sum_{t=1}^{k^0}(\boldsymbol{x}_{it}u_{it}-\boldsymbol{a})\|_{2+\xi} \leq \sup_{i\in\mathbb{N}}\frac{1}{T}\sum_{t=1}^{k^0}(\|\boldsymbol{x}_{it}u_{it}\|_{2+\xi}+|a|) \leq \sup_{i\in\mathbb{N}}\frac{1}{T}\sum_{t=1}^{k^0}(\|\boldsymbol{x}_{it}\|_{4+2\xi}\|u_{it}\|_{4+2\xi}+|\boldsymbol{a}|) < \infty$ by the Cauchy-Schwarz inequality and Assumption 2. A similar set of inequalities can be applied to show $\sup_{i\in\mathbb{N}}\|\frac{1}{T}\sum_{t=k^0+1}^{T}(\boldsymbol{x}_{it}u_{it}-\boldsymbol{a})\|_{2+\xi} < \infty$. The asymptotic variance is:

$$
\begin{aligned}
&\text{Avar}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{k^0}(\boldsymbol{x}_{it}u_{it}-\boldsymbol{a}) \\ \frac{1}{T}\sum_{t=k^0+1}^{T}(\boldsymbol{x}_{it}u_{it}-\boldsymbol{a}) \end{pmatrix}\right) \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{k^0}(\boldsymbol{x}_{it}u_{it}-\boldsymbol{a}) \\ \frac{1}{T}\sum_{t=k^0+1}^{T}(\boldsymbol{x}_{it}u_{it}-\boldsymbol{a}) \end{pmatrix}\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{k^0}(\boldsymbol{x}_{it}u_{it}-\boldsymbol{a}) \\ \frac{1}{T}\sum_{t=k^0+1}^{T}(\boldsymbol{x}_{it}u_{it}-\boldsymbol{a}) \end{pmatrix}'\right] \\
&= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[\begin{pmatrix} T^{-2}\sum_{t,s=1}^{k^0}\boldsymbol{x}_{it}\boldsymbol{x}_{is}'u_{it}u_{is} & T^{-2}\sum_{t=1}^{k^0}\sum_{s=k^0+1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{is}'u_{it}u_{is} \\ \sum_{t,s=k^0+1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{is}'u_{it}u_{is} & \end{pmatrix}\right] \\
&- \begin{pmatrix} {k^0}^2/T^2\boldsymbol{a}\boldsymbol{a}' & k^0(T-k^0)/T^2\boldsymbol{a}\boldsymbol{a}' \\ k^0(T-k^0)/T^2\boldsymbol{a}\boldsymbol{a}' & (T-k^0)^2/T^2\boldsymbol{a}\boldsymbol{a}' \end{pmatrix} \\
&= \begin{pmatrix} T^{-2}\sum_{t,s=1}^{k^0}(\boldsymbol{W}_{t,s}-\boldsymbol{a}\boldsymbol{a}') & T^{-2}\sum_{t=1}^{k^0}\sum_{s=k^0+1}^{T}(\boldsymbol{W}_{t,s}-\boldsymbol{a}\boldsymbol{a}') \\ T^{-2}\sum_{t=k^0+1}^{T}\sum_{s=1}^{k^0}(\boldsymbol{W}_{t,s}-\boldsymbol{a}\boldsymbol{a}') & T^{-2}\sum_{t,s=k^0+1}^{T}(\boldsymbol{W}_{t,s}-\boldsymbol{a}\boldsymbol{a}') \end{pmatrix} = \boldsymbol{W}_{xu},
\end{aligned}
$$

where we use Assumption 1(iii) and (iv) to have $\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\boldsymbol{x}_{it}u_{it}] \xrightarrow{p} \boldsymbol{a}$ in the first equality, and we use Assumption 2 in the second equality. Thus all the three conditions in the central limit theorem are met, and we have

$$\sqrt{N}\left((NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}\boldsymbol{u} - \boldsymbol{D}_{k^0}\cdot(\boldsymbol{\ell}_2\otimes\boldsymbol{a})\right) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{W}_{xu}).$$

Together with the result that $(NT)^{-1}\widetilde{\boldsymbol{X}}^{0\prime}\widetilde{\boldsymbol{X}}^{0} \xrightarrow{p} \boldsymbol{D}_{k^0}\otimes\boldsymbol{Q}$, we obtain

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^0_{OLS} - \boldsymbol{\beta}^0 - \boldsymbol{\Delta}_{N,\boldsymbol{\beta}}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_{\boldsymbol{\beta}}),$$

where $\boldsymbol{\Omega}_{\boldsymbol{\beta}} = (\boldsymbol{D}_{k^0}\otimes\boldsymbol{Q})^{-1}\cdot\boldsymbol{W}_{xu}\cdot\left(\boldsymbol{D}_{k^0}'\otimes\boldsymbol{Q}'\right)^{-1}$. Together with the result $P(\widehat{\boldsymbol{\beta}}_{OLS} = \widehat{\boldsymbol{\beta}}^0) \to 1$, we have

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}^0 - \boldsymbol{\Delta}_{N,\boldsymbol{\beta}}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_{\boldsymbol{\beta}}). \tag{16}$$

The second part of the theorem follows by premultiplying both sides of (16) by $(\boldsymbol{I}_p\ ,\ -\boldsymbol{I}_p)$:

$$\sqrt{N}(\widehat{\boldsymbol{\delta}}_{OLS} - \boldsymbol{\delta}^0 - \boldsymbol{\Delta}_{N,\boldsymbol{\delta}}) = (\boldsymbol{I}_p\ ,\ -\boldsymbol{I}_p)\sqrt{N}(\widehat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}^0 - \boldsymbol{\Delta}_{N,\boldsymbol{\beta}})$$
$$\xrightarrow{d} (\boldsymbol{I}_p\ ,\ -\boldsymbol{I}_p)\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\beta}}) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\delta}}),$$

where $\boldsymbol{\Omega}_{\boldsymbol{\delta}} = (\boldsymbol{I}_p\ ,\ -\boldsymbol{I}_p)\ \boldsymbol{\Omega}_{\boldsymbol{\beta}}\ (\boldsymbol{I}_p\ ,\ -\boldsymbol{I}_p)'$. □

The following lemma is needed for proving Theorem 3 and 5. Recall the definitions $\overline{\boldsymbol{x}}_{i,1}^0 = {k^0}^{-1}\sum_1 \boldsymbol{x}_{it}$, $\overline{\boldsymbol{x}}_{i,2}^0 = (T-k^0)^{-1}\sum_2 \boldsymbol{x}_{it}$, $\widetilde{\boldsymbol{x}}_{it}^a = \boldsymbol{x}_{it}1_{t\le k^0} - \lambda^0\overline{\boldsymbol{x}}_{i,1}^0$, $\widetilde{\boldsymbol{x}}_{it}^b = \boldsymbol{x}_{it}1_{t\ge k^0+1} - (1-\lambda^0)\overline{\boldsymbol{x}}_{i,2}^0$ and $\widetilde{\boldsymbol{x}}_{it} = (\widetilde{\boldsymbol{x}}_{it}^{a\prime}\ \ \widetilde{\boldsymbol{x}}_{it}^{b\prime})'$.

Lemma 2. *Under Assumptions 1 and 3, we have*
(i) $N^{-1}\sum_{i=1}^N \sum_1 (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^0)\boldsymbol{x}_{it}' \xrightarrow{p} k^0\boldsymbol{Q} - (k^0)^{-1}\sum_{t,s=1}^{k^0}\boldsymbol{\Omega}_{t,s}$,
$N^{-1}\sum_{i=1}^N \sum_2 (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,2}^0)\boldsymbol{x}_{it} \xrightarrow{p} (T-k^0)\boldsymbol{Q} - (T-k^0)^{-1}\sum_{t,s=k^0+1}^{T}\boldsymbol{\Omega}_{t,s}$*;*
(ii) $N^{-1/2}\sum_{i=1}^N w_i \xrightarrow{d} \mathcal{N}(\mathbf{0}, \overline{\boldsymbol{W}}_{x\epsilon})$*, for* $\boldsymbol{w}_i = \left(\sum_1 (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^0)'\epsilon_{it}\ ,\ \ \sum_2 (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,2}^0)'\epsilon_{it}\right)'$*;*
(iii) $N^{-1}\sum_{i=1}^N \sum_1 (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)\boldsymbol{x}_{it}' \xrightarrow{p} k^0\boldsymbol{Q} - T^{-1}\sum_{t=1}^{k^0}\sum_{s=1}^{T}\boldsymbol{\Omega}_{t,s}$,
$N^{-1}\sum_{i=1}^N \sum_2 (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)\boldsymbol{x}_{it}' \xrightarrow{p} (T-k^0)\boldsymbol{Q} - T^{-1}\sum_{t=k^0+1}^{T}\sum_{s=1}^{T}\boldsymbol{\Omega}_{t,s}$*;*
(iv) $N^{-1/2}\sum_{i=1}^N \boldsymbol{v}_i \xrightarrow{d} \mathcal{N}(\mathbf{0}, \overline{\boldsymbol{W}}_{xu})$*, for* $v_i = \left(\sum_1 (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)'u_{it}\ ,\ \ \sum_2 (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)'u_{it}\right)'$*;*
(v) $N^{-1}\sum_{i=1}^N \sum_{t=1}^T \widetilde{\boldsymbol{x}}_{it}^a\widetilde{\boldsymbol{x}}_{it}^a \xrightarrow{p} k^0\boldsymbol{Q} - T^{-1}\sum_{t,s=1}^{k^0}\boldsymbol{\Omega}_{t,s}$,
$N^{-1}\sum_{i=1}^N \sum_{t=1}^T \widetilde{\boldsymbol{x}}_{it}^b\widetilde{\boldsymbol{x}}_{it}^{a\prime} \xrightarrow{p} -T^{-1}\sum_{t=k^0+1}^{T}\sum_{s=1}^{k^0}\boldsymbol{\Omega}_{t,s}$,
$N^{-1}\sum_{i=1}^N \sum_{t=1}^T \widetilde{\boldsymbol{x}}_{it}^b\widetilde{\boldsymbol{x}}_{it}^b \xrightarrow{p} (T-k^0)\boldsymbol{Q} - T^{-1}\sum_{t,s=k^0+1}^{T}\boldsymbol{\Omega}_{t,s}$*;*
(vi) $N^{-1/2}\sum_{i=1}^N \sum_{t=1}^T \widetilde{\boldsymbol{x}}_{it}\epsilon_{it}^* \xrightarrow{d} \mathcal{N}(\mathbf{0}, \widetilde{\boldsymbol{W}}_{x\epsilon})$.

Proof. For (i), we first note that

$$\tfrac{1}{N}\sum_{i=1}^N \sum_1 (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^0)\boldsymbol{x}_{it}' = \tfrac{1}{N}\sum_{i=1}^N \sum_1 \boldsymbol{x}_{it}\boldsymbol{x}_{it}' - \tfrac{k^0}{N}\sum_{i=1}^N \overline{\boldsymbol{x}}_{i,1}^0\overline{\boldsymbol{x}}_{i,1}^{0\prime}.$$

By Lemma 1(i), $\frac{1}{N}\sum_{i=1}^N \boldsymbol{x}_{it}\boldsymbol{x}_{it}' \xrightarrow{p} \boldsymbol{Q}$ for every t. By Assumption 1(i) and (ii), $\overline{\boldsymbol{x}}_{i,1}^0\overline{\boldsymbol{x}}_{i,1}^{0\prime}$ is independent over i, and the moment condition in the WLLN (in Appendix B) can be shown to be satisfied. Applying the WLLN, together with Assumption 3(i), we get

$$\frac{1}{N}\sum_{i=1}^N \overline{\boldsymbol{x}}_{i,1}^0\overline{\boldsymbol{x}}_{i,1}^{0\prime} = \frac{1}{N}\sum_{i=1}^N \frac{1}{(k^0)^2}\sum_{t,s=1}^{k^0}\boldsymbol{x}_{it}\boldsymbol{x}_{is}' \xrightarrow{p} \frac{1}{(k^0)^2}\sum_{t,s=1}^{k^0}\boldsymbol{\Omega}_{t,s}. \tag{17}$$

So $\frac{1}{N}\sum_{i=1}^N \sum_1 (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^0)\boldsymbol{x}_{it}' \xrightarrow{p} k^0\boldsymbol{Q} - {k^0}^{-1}\sum_{t,s=1}^{k^0}\boldsymbol{\Omega}_{t,s}$. The second part of (i) concerning the corresponding result for the second regime can be derived similarly.

We now derive (ii). By Assumption 1(i), (iii) and (iv), $\sum_1(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^0)\epsilon_{it}$ is a mean-zero process, independent over i. Also, by the triangle and Hölder's inequalities and Assumption 2, $\|\sum_1(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^0)\epsilon_{it}\|_{2+\xi/2} \leq \sum_1 \|\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^0\|_{4+\xi} \|\epsilon_{it}\|_{4+\xi} \leq 2k^0 \sup_{i,t} \|\boldsymbol{x}_{it}\|_{4+\xi} \sup_{i,t} \|\epsilon_{it}\|_{4+\xi} < \infty$. A similar set of inequalities can be derived for the $(2+\xi/2)^{th}$ moment of $\sum_2(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,2}^0)\epsilon_{it}$. So, the $(2+\xi/2)^{th}$ moment of $\boldsymbol{w}_i$ exists, for some $\xi > 0$. By Assumption 3(ii), the asymptotic variance $\text{Avar}\left(N^{-1/2}\sum_{i=1}^N \boldsymbol{w}_i\right) = \overline{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}$ also exists. Applying the CLT for an independent sequence, we obtain that

$$N^{-1/2}\sum_{i=1}^N \boldsymbol{w}_i = N^{-1/2}\sum_{i=1}^N \begin{pmatrix} \sum_1(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^0)\epsilon_{it} \\ \sum_2(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,2}^0)\epsilon_{it} \end{pmatrix} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \overline{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}). \tag{18}$$

For (iii), it can be shown, using Lemma 1(i) and 2 that

$$N^{-1}\sum_{i=1}^N \sum_1(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)\boldsymbol{x}_{it}' = N^{-1}\sum_{i=1}^N \sum_1 \boldsymbol{x}_{it}\boldsymbol{x}_{it}' - \sum_{t=1}^{k^0} T^{-1}\sum_{s=1}^T \left(N^{-1}\sum_{i=1}^N \boldsymbol{x}_{is}\boldsymbol{x}_{it}'\right)$$
$$\xrightarrow{p} k^0\boldsymbol{Q} - T^{-1}\sum_{t=1}^{k^0}\sum_{s=1}^T \boldsymbol{\Omega}_{t,s}.$$

For (iv), by Assumption 1(i), (iii) and (iv), $\sum_1(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)u_{it} = \sum_1(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)\epsilon_{it} + \sum_1(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)c_i$ is a mean-zero process, independent over i. Again by the triangle and Hölder's inequalities and Assumption 2, the $(2+\xi/2)^{th}$ moment of $\sum_1(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)u_{it}$ and $\sum_2(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)u_{it}$ can be shown to be bounded: $\|\sum_j(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)u_{it}\|_{2+\xi/2} \leq \sum_j \|\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i\|_{4+\xi} \|u_{it}\|_{4+\xi} \leq 2k^0 \sup_{i,t}\|\boldsymbol{x}_{it}\|_{4+\xi} \sup_{i,t}\|u_{it}\|_{4+\xi} < \infty$. So, the $(2+\xi/2)^{th}$ moment of v_i exists, for some $\xi > 0$. By Assumption 3(ii), the asymptotic variance $\text{Avar}\left(N^{-1/2}\sum_{i=1}^N \boldsymbol{v}_i\right) = \overline{\boldsymbol{W}}_{xu}$ also exists. Applying the CLT for an independent sequence, we obtain that

$$N^{-1/2}\sum_{i=1}^N \boldsymbol{v}_i = N^{-1/2}\sum_{i=1}^N \begin{pmatrix} \sum_1(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)u_{it} \\ \sum_2(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)u_{it} \end{pmatrix} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \overline{\boldsymbol{W}}_{\boldsymbol{x}u}). \tag{19}$$

For (v), it can be derived that

$$\frac{1}{N}\sum_{i=1}^N(\sum_{t=1}^T \widetilde{\boldsymbol{x}}_{it}^a\widetilde{\boldsymbol{x}}_{it}^{a\prime}) = \frac{1}{N}\sum_{i=1}^N(\sum_1 \boldsymbol{x}_{it}\boldsymbol{x}_{it}' - k^0\lambda^0\overline{\boldsymbol{x}}_{i,1}^0\overline{\boldsymbol{x}}_{i,1}^{0\prime}).$$

Using Lemma 1(i) and the result in (17), we can show that the above converges in probability to $k^0\boldsymbol{Q} - T^{-1}\sum_{t=1}^{k^0}\sum_{s=1}^{k^0}\boldsymbol{\Omega}_{t,s}$. The two other results in (iii) can be shown accordingly.

For (vi), the summand $\sum_{t=1}^{T} \widetilde{\boldsymbol{x}}_{it}\epsilon_{it}^{*}$ has mean zero by Assumption 1(iii), and is independent over i by Assumption 1(i). Its $(2+\xi/2)^{th}$ moment can be shown to be bounded under Assumption 2 using the triangle and Hölder's inequalities. Further, its asymptotic variance exists and is equal to $\widetilde{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}$ by Assumption 3(iv). Thus we can apply the CLT for an independent sequence to obtain

$$N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} \widetilde{\boldsymbol{x}}_{it}\epsilon_{it}^{*} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \widetilde{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}).$$

□

Proof of Theorem 3.

Define $\widehat{\boldsymbol{\beta}}_{FE}^{0}$ to be the FE estimator based on the true break point. Using the same argument as in the proof of Theorem 2, $P(\widehat{k} = k^0) \to 1$ implies that $P(\widehat{\boldsymbol{\beta}}_{FE}^{0} = \widehat{\boldsymbol{\beta}}_{FE}) \to 1$; so it is sufficient to show $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{FE}^{0} - \boldsymbol{\beta}^0)$ has the same asymptotic distribution described in Theorem 3. As in Section 5.2, we use the notation $\overline{\boldsymbol{x}}_{i,1}^{0}$ and $\overline{\boldsymbol{x}}_{i,2}^{0}$ to denote the subsample averages over the first and the second regime. The corresponding fixed effects estimator for the first subsample is $\widehat{\boldsymbol{\beta}}_{FE,1}^{0} = \left[\sum_{i=1}^{N}\sum_{1}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^{0})\boldsymbol{x}_{it}'\right]^{-1} \sum_{i=1}^{N}\sum_{1}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^{0})y_{it}$, and for the second subsample $\widehat{\boldsymbol{\beta}}_{FE,2}^{0} = \left[\sum_{i=1}^{N}\sum_{2}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,2}^{0})\boldsymbol{x}_{it}'\right]^{-1} \sum_{i=1}^{N}\sum_{2}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,2}^{0})y_{it}$. Since $\sum_{t=1}^{k^0}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^{0})c_i = 0$, we have, in matrix notation,

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{FE}^{0} - \boldsymbol{\beta}^0\right) =$$
$$\begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N}\sum_{1}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^{0})\boldsymbol{x}_{it}' & \mathbf{0} \\ \mathbf{0} & \frac{1}{N}\sum_{i=1}^{N}\sum_{2}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,2}^{0})\boldsymbol{x}_{it}' \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{1}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,1}^{0})\epsilon_{it} \\ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{2}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_{i,2}^{0})\epsilon_{it} \end{pmatrix}.$$

Note that here we need Assumption 3(v): if $k^0 = 1$ then $\boldsymbol{x}_{i1} = \overline{\boldsymbol{x}}_{i,1}^{0}$, and so $\boldsymbol{\beta}_1^0$ is not identified using FE. Similarly, if $k^0 = T-1$, then $\boldsymbol{\beta}_2^0$ is not identified. In Lemma 2(i) and (ii), we derived the limit of all terms in the above equation. Putting them together, we get

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{FE}^{0} - \boldsymbol{\beta}^0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_1^{-1}\overline{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}\boldsymbol{\Omega}_1^{-1}).$$

So $\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}^0\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_1^{-1}\overline{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}\boldsymbol{\Omega}_1^{-1})$. □

Proof of Theorem 4.

Same as in the proof of Theorem 3, we can take k^0 as known to obtain the asymptotic distribution of the slope estimator, denoted $\widehat{\boldsymbol{\beta}}_{IV}^{0}$, and argue that the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{IV}^{0}$ and $\widehat{\boldsymbol{\beta}}_{IV}$

is the same. For $\widehat{\boldsymbol{\beta}}^0_{IV}$, we have

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}^0_{IV}-\boldsymbol{\beta}^0\right)=$$
$$\begin{pmatrix}\frac{1}{N}\sum_{i=1}^N\sum_1(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)\boldsymbol{x}'_{it} & \mathbf{0}\\ \mathbf{0} & \frac{1}{N}\sum_{i=1}^N\sum_2(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)\boldsymbol{x}'_{it}\end{pmatrix}^{-1}\begin{pmatrix}\frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_1(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)u_{it}\\ \frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_2(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)u_{it}\end{pmatrix}.$$

In Lemma 2(iii) and (iv), we derived the limit of all terms in the above equation. So we have

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}^0_{IV}-\boldsymbol{\beta}^0\right)\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{\Omega}_2^{-1}\overline{\boldsymbol{W}}_{\boldsymbol{x}u}\boldsymbol{\Omega}_2^{-1}).$$

When k^0 is unknown, $\sqrt{N}\left(\widehat{\boldsymbol{\beta}}_{IV}-\boldsymbol{\beta}^0\right)\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{\Omega}_2^{-1}\overline{\boldsymbol{W}}_{\boldsymbol{x}u}\boldsymbol{\Omega}_2^{-1})$. □

Proof of Theorem 5.

We have

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}^0_{FFE}-\boldsymbol{\beta}^0)=\left(N^{-1}\sum_{i=1}^N\sum_{t=1}^T\widetilde{\boldsymbol{x}}_{it}\widetilde{\boldsymbol{x}}'_{it}\right)^{-1}\left(N^{-1/2}\sum_{i=1}^N\sum_{t=1}^T\widetilde{\boldsymbol{x}}_{it}\epsilon^*_{it}\right).$$

By definition

$$N^{-1}\sum_{i=1}^N\sum_{t=1}^T\widetilde{\boldsymbol{x}}_{it}\widetilde{\boldsymbol{x}}'_{it}=\begin{pmatrix}N^{-1}\sum_{i=1}^N\sum_{t=1}^T\widetilde{\boldsymbol{x}}^a_{it}\widetilde{\boldsymbol{x}}^{a\prime}_{it} & N^{-1}\sum_{i=1}^N\sum_{t=1}^T\widetilde{\boldsymbol{x}}^a_{it}\widetilde{\boldsymbol{x}}^{b\prime}_{it}\\ N^{-1}\sum_{i=1}^N\sum_{t=1}^T\widetilde{\boldsymbol{x}}^b_{it}\widetilde{\boldsymbol{x}}^{a\prime}_{it} & N^{-1}\sum_{i=1}^N\sum_{t=1}^T\widetilde{\boldsymbol{x}}^b_{it}\widetilde{\boldsymbol{x}}^{b\prime}_{it}\end{pmatrix}.$$

which converges, as is shown in Lemma 2(v), to

$$\begin{pmatrix}k^0\boldsymbol{Q}-T^{-1}\sum_{t,s=1}^{k^0}\boldsymbol{\Omega}_{t,s} & -T^{-1}\sum_{t=k^0+1}^T\sum_{s=1}^{k^0}\boldsymbol{\Omega}_{t,s}\\ -T^{-1}\sum_{s=k^0+1}^T\sum_{t=1}^{k^0}\boldsymbol{\Omega}_{t,s} & (T-k^0)\boldsymbol{Q}-T^{-1}\sum_{t,s=k^0+1}^T\boldsymbol{\Omega}_{t,s}\end{pmatrix}=:\boldsymbol{\Omega}_3.$$

Further, by Lemma 2(vi),

$$N^{-1/2}\sum_{i=1}^N\sum_{t=1}^T\widetilde{\boldsymbol{x}}_{it}\epsilon^*_{it}\xrightarrow{d}\mathcal{N}(\mathbf{0},\widetilde{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}).$$

Thus $\sqrt{N}(\widehat{\boldsymbol{\beta}}^0_{FFE}-\boldsymbol{\beta}^0)\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{\Omega}_3^{-1}\widetilde{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}\boldsymbol{\Omega}_3^{-1})$, and consequently $\sqrt{N}(\widehat{\boldsymbol{\beta}}_{FFE}-\boldsymbol{\beta}^0)\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{\Omega}_3^{-1}\widetilde{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}\boldsymbol{\Omega}_3^{-1})$ when k^0 is unknown. □

Calculation of the variance of the FE, IV and FFE estimators.

All the following results are derived under the assumption that ϵ_{it}'s are independent over i and

t. We first compute $\boldsymbol{V}_{FE} = \boldsymbol{\Omega}_1^{-1}\overline{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}\boldsymbol{\Omega}_1^{-1}$. To find $\overline{\boldsymbol{W}}_{\boldsymbol{x}\epsilon} := N^{-1}\sum_{i=1}^{N}\mathbb{E}[\boldsymbol{w}_i\boldsymbol{w}_i']$, note that

$$\boldsymbol{w}_i\boldsymbol{w}_i' = \begin{pmatrix} \sum_{t,s=1}^{k^0}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,1}^0)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_{i,1}^0)'\epsilon_{it}\epsilon_{is} & \sum_{t=1}^{k^0}\sum_{s=k^0+1}^{T}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,1}^0)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_{i,2}^0)'\epsilon_{it}\epsilon_{is} \\ \sum_{t=k^0+1}^{T}\sum_{s=1}^{k^0}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,2}^0)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_{i,1}^0)'\epsilon_{it}\epsilon_{is} & \sum_{t,s=k^0+1}^{T}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,2}^0)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_{i,2}^0)'\epsilon_{it}\epsilon_{is} \end{pmatrix}. \tag{20}$$

We then take the expectation of all the four blocks. The upper-left block in the matrix has expectation equal to

$$\begin{aligned}
&\sum_{t,s=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,1}^0)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_{i,1}^0)'\epsilon_{it}\epsilon_{is}] \\
&= \sum_{t,s=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,1}^0)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_{i,1}^0)']\mathbb{E}[\epsilon_{it}\epsilon_{is}] \\
&= \sum_{t=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,1}^0)(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,1}^0)']\mathbb{E}[\epsilon_{it}^2] \\
&= \sum_{t=1}^{k^0}\left(\boldsymbol{Q}-(k^0)^{-1}\sum_{s=1}^{k^0}\boldsymbol{\Omega}_{t,s}-(k^0)^{-1}\sum_{s=1}^{k^0}\boldsymbol{\Omega}_{s,t}+(k^0)^{-2}\sum_{s=1}^{k^0}\sum_{t=1}^{k^0}\boldsymbol{\Omega}_{s,t}\right)\sigma_\epsilon^2 \\
&= \left(k^0\boldsymbol{Q}-(k^0)^{-1}\sum_{t,s=1}^{k^0}\boldsymbol{\Omega}_{t,s}-(k^0)^{-1}\sum_{t,s=1}^{k^0}\boldsymbol{\Omega}_{s,t}+(k^0)^{-1}\sum_{s=1}^{k^0}\sum_{t=1}^{k^0}\boldsymbol{\Omega}_{s,t}\right)\sigma_\epsilon^2 \\
&= \left(k^0\boldsymbol{Q}-(k^0)^{-1}\sum_{t,s=1}^{k^0}\boldsymbol{\Omega}_{t,s}\right)\sigma_\epsilon^2.
\end{aligned}$$

Following the same steps, we can find the expectations of all three other blocks in (20):

$$\begin{aligned}
&\sum_{t=k^0+1}^{T}\sum_{s=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,2}^0)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_{i,1}^0)'\epsilon_{it}\epsilon_{is}] = 0 \\
&\sum_{t=1}^{k^0}\sum_{s=k^0+1}^{T}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,1}^0)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_{i,2}^0)'\epsilon_{it}\epsilon_{is}] = 0 \\
&\sum_{t,s=k^0+1}^{T}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i,2}^0)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_{i,2}^0)'\epsilon_{it}\epsilon_{is}] \\
&\qquad = \left((T-k^0)\boldsymbol{Q}-(T-k^0)^{-1}\sum_{t,s=k^0+1}^{T}\boldsymbol{\Omega}_{t,s}\right)\sigma_\epsilon^2.
\end{aligned}$$

Thus, $\overline{\boldsymbol{W}}_{\boldsymbol{x}\epsilon} = \sigma_\epsilon^2\begin{pmatrix} k^0\boldsymbol{Q}-(k^0)^{-1}\sum_{t,s=1}^{k^0}\boldsymbol{\Omega}_{t,s} & \boldsymbol{0} \\ \boldsymbol{0} & (T-k^0)\boldsymbol{Q}-(T-k^0)^{-1}\sum_{t,s=k^0+1}^{T}\boldsymbol{\Omega}_{t,s} \end{pmatrix} = \sigma_\epsilon^2\boldsymbol{\Omega}_1.$
So $\boldsymbol{V}_{FE} = \sigma_\epsilon^2\boldsymbol{\Omega}_1^{-1}$. Under the simplification $\boldsymbol{\Omega}_{t,s} = \boldsymbol{Q}$ for $t = s$ and $\boldsymbol{\Omega}_{t,s} = \boldsymbol{\Omega}^*$ for $t \neq s$, we have

$$\begin{aligned}
\boldsymbol{V}_{FE} &= \sigma_\epsilon^2\begin{pmatrix} (k^0-1)(\boldsymbol{Q}-\boldsymbol{\Omega}^*) & 0 \\ 0 & (T-k^0-1)(\boldsymbol{Q}-\boldsymbol{\Omega}^*) \end{pmatrix}^{-1} \\
&= \sigma_\epsilon^2\left(\begin{pmatrix} \boldsymbol{C}_1 & 0 \\ 0 & \boldsymbol{C}_2 \end{pmatrix} - \begin{pmatrix} (1-k^0/T)(\boldsymbol{Q}-\boldsymbol{\Omega}^*) & 0 \\ 0 & k^0/T(\boldsymbol{Q}-\boldsymbol{\Omega}^*) \end{pmatrix}\right)^{-1}
\end{aligned}$$

for $\boldsymbol{C}_1 = k^0(1-T^{-1})(\boldsymbol{Q}-\boldsymbol{\Omega}^*)$ and $\boldsymbol{C}_2 = (T-k^0)(1-T^{-1})(\boldsymbol{Q}-\boldsymbol{\Omega}^*)$.

Next we compute $\boldsymbol{V}_{IV} = \boldsymbol{\Omega}_2^{-1}\overline{\boldsymbol{W}}_{\boldsymbol{x}u}\boldsymbol{\Omega}_2^{-2}$. To find $\overline{\boldsymbol{W}}_{\boldsymbol{x}u} := N^{-1}\sum_{i=1}^N \mathbb{E}[\boldsymbol{v}_i\boldsymbol{v}_i']$, we derive

$$\boldsymbol{v}_i\boldsymbol{v}_i' = \begin{pmatrix} \sum_{t,s=1}^{k^0}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'u_{it}u_{is} & \sum_{t=1}^{k^0}\sum_{s=k^0+1}^{T}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'u_{it}u_{is} \\ \sum_{t=k^0+1}^{T}\sum_{s=1}^{k^0}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'u_{it}u_{is} & \sum_{t,s=k^0+1}^{T}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'u_{it}u_{is} \end{pmatrix}.$$

The upper-left block in the matrix has expectation equal to $\sum_{t,s=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'(\epsilon_{it}\epsilon_{is} + c_i\epsilon_{is} + \epsilon_{it}c_i + c_i^2)]$. First,

$$\begin{aligned}
&\sum_{t,s=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'\epsilon_{it}\epsilon_{is}] \\
&= \sum_{t,s=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)']\mathbb{E}[\epsilon_{it}\epsilon_{is}] \\
&= \sum_{t=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)']\mathbb{E}[\epsilon_{it}^2] \\
&= \left(k^0\boldsymbol{Q} - T^{-1}\sum_{t=1}^{k^0}\sum_{s=1}^{T}\boldsymbol{\Omega}_{t,s} - T^{-1}\sum_{t=1}^{k^0}\sum_{s=1}^{T}\boldsymbol{\Omega}_{s,t} + T^{-2}k^0\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{\Omega}_{t,s}\right)\sigma_\epsilon^2 =: \boldsymbol{A}_1\sigma_\epsilon^2.
\end{aligned}$$

Next, by the independence of ϵ_{it} with c_i and $\boldsymbol{x}_{is}$,

$$\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'c_i\epsilon_{is}] = \mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'c_i]\mathbb{E}[\epsilon_{is}] = 0.$$

So both $\sum_{t,s=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'c_i\epsilon_{is}]$ and $\sum_{t,s=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'\epsilon_{it}c_i]$ are zero. Lastly,

$$\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'c_i^2] = \mathbb{E}[\boldsymbol{x}_{it}\boldsymbol{x}_{is}'c_i^2] - \mathbb{E}[\boldsymbol{x}_{it}\overline{\boldsymbol{x}}_i'c_i^2] - \mathbb{E}[\overline{\boldsymbol{x}}_i\boldsymbol{x}_{is}'c_i^2] + \mathbb{E}[\overline{\boldsymbol{x}}_i\overline{\boldsymbol{x}}_i'c_i^2].$$

Summed up over $t, s = 1, \ldots, k^0$, we get

$$\begin{aligned}
&\sum_{t,s=1}^{k^0}\mathbb{E}[(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'c_i^2] = \sum_{t,s=1}^{k^0}\widetilde{\boldsymbol{W}}_{t,s} - \sum_{t=1}^{k^0}k^0T^{-1}\sum_{s=1}^{T}\mathbb{E}[\boldsymbol{x}_{it}\boldsymbol{x}_{is}'c_i^2] \\
&- k^0\sum_{s=1}^{k^0}T^{-1}\sum_{t=1}^{T}\mathbb{E}[\boldsymbol{x}_{it}\boldsymbol{x}_{is}'c_i^2] + (k^0)^2T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[\boldsymbol{x}_{it}\boldsymbol{x}_{is}c_i^2] \\
&= (1-k^0/T)^2\sum_{t,s=1}^{k^0}\widetilde{\boldsymbol{W}}_{t,s} - 2k^0/T(1-k^0/T)\sum_{t=1}^{k^0}\sum_{s=k^0+1}^{T}\widetilde{\boldsymbol{W}}_{t,s} + (k^0)^2/T^2\sum_{t,s=k^0+1}^{T}\widetilde{\boldsymbol{W}}_{t,s} \overset{\text{def}}{=} \boldsymbol{B}.
\end{aligned}$$

where $\widetilde{\boldsymbol{W}}_{t,s} := \mathbb{E}[\boldsymbol{x}_{it}\boldsymbol{x}_{is}'c_i^2]$. So $\mathbb{E}\left[\sum_{t,s=1}^{k^0}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'u_{it}u_{is}\right] = \boldsymbol{A}_1\sigma_\epsilon^2 + \boldsymbol{B}$. We follow the same procedure to find the limit of the other three submatrices in (7), which gives us

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{k^0}\sum_{s=k^0+1}^{T}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'u_{it}u_{is}\right] = -\boldsymbol{B};\ \mathbb{E}\left[\sum_{t=k^0+1}^{T}\sum_{s=1}^{k^0}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'u_{it}u_{is}\right] = -\boldsymbol{B}; \\
&\mathbb{E}\left[\sum_{t,s=k^0+1}^{T}(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{is}-\overline{\boldsymbol{x}}_i)'u_{it}u_{is}\right] = \boldsymbol{A}_2\sigma_\epsilon^2 + \boldsymbol{B}.
\end{aligned}$$

where $\boldsymbol{A}_2 = (T-k^0)\boldsymbol{Q} - T^{-1}\sum_{t=k^0+1}^{T}\sum_{s=1}^{T}\boldsymbol{\Omega}_{t,s} - T^{-1}\sum_{t=k^0+1}^{T}\sum_{s=1}^{T}\boldsymbol{\Omega}_{s,t} + T^{-2}(T-k^0)\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{\Omega}_{t,s}$. Therefore, $\overline{\boldsymbol{W}}_{\boldsymbol{x}u} = \sigma_\epsilon^2 \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{B} & -\boldsymbol{B} \\ -\boldsymbol{B} & \boldsymbol{B} \end{pmatrix}$. Under the condition $\boldsymbol{\Omega}_{s,t} = \boldsymbol{\Omega}^*$ for $t \neq s$, imposed in Section 5.4, $\boldsymbol{A}_1$ can be simplified to $\boldsymbol{C}_1 = k^0(1-T^{-1})(\boldsymbol{Q} - \boldsymbol{\Omega}^*)$, and $\boldsymbol{A}_2$ to $\boldsymbol{C}_2 = (T-k^0)(1-T^{-1})(\boldsymbol{Q} - \boldsymbol{\Omega}^*)$. Furthermore,

$$\boldsymbol{\Omega}_2 = \begin{pmatrix} \boldsymbol{C}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_2 \end{pmatrix}, \Rightarrow \boldsymbol{V}_{IV} = \sigma_\epsilon^2 \begin{pmatrix} \boldsymbol{C}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_2 \end{pmatrix}^{-1} + \begin{pmatrix} \boldsymbol{C}_1^{-1}\boldsymbol{B}\boldsymbol{C}_1^{-1} & -\boldsymbol{C}_1^{-1}\boldsymbol{B}\boldsymbol{C}_2^{-1} \\ -\boldsymbol{C}_2^{-1}\boldsymbol{B}\boldsymbol{C}_1^{-1} & \boldsymbol{C}_2^{-1}\boldsymbol{B}\boldsymbol{C}_2^{-1} \end{pmatrix}.$$

Finally, we compute $\boldsymbol{V}_{FFE} = \boldsymbol{\Omega}_3^{-1}\widetilde{\boldsymbol{W}}_{\boldsymbol{x}\epsilon}\boldsymbol{\Omega}_3^{-1}$. Using the simplifying conditions in Section 5.4,

$$\begin{aligned}
\boldsymbol{\Omega}_3 &= \begin{pmatrix} k^0(1-T^{-1})\boldsymbol{Q} - k^0(k^0-1)T^{-1}\boldsymbol{\Omega}^* & -k^0(T-k^0)T^{-1}\boldsymbol{\Omega}^* \\ -k^0(T-k^0)T^{-1}\boldsymbol{\Omega}^* & (T-k^0)(1-T^{-1})\boldsymbol{Q} - (T-k^0)(T-k^0-1)T^{-1}\boldsymbol{\Omega}^* \end{pmatrix} \\
&= \begin{pmatrix} k^0(1-T^{-1})(\boldsymbol{Q}-\boldsymbol{\Omega}^*) + k^0(T-k^0)T^{-1}\boldsymbol{\Omega}^* & -k^0(T-k^0)T^{-1}\boldsymbol{\Omega}^* \\ -k^0(T-k^0)T^{-1}\boldsymbol{\Omega}^* & (T-k^0)(1-T^{-1})(\boldsymbol{Q}-\boldsymbol{\Omega}^*) + k^0(T-k^0)T^{-1}\boldsymbol{\Omega}^* \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{C}_1 & 0 \\ 0 & \boldsymbol{C}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{D} & -\boldsymbol{D} \\ -\boldsymbol{D} & \boldsymbol{D} \end{pmatrix}, \text{ where } \boldsymbol{D} = k^0(T-k^0)T^{-1}\boldsymbol{\Omega}^*.
\end{aligned}$$

Thus, $\boldsymbol{V}_{FFE} = \sigma_\epsilon^2 \left(\begin{pmatrix} \boldsymbol{C}_1 & 0 \\ 0 & \boldsymbol{C}_2 \end{pmatrix} + \begin{pmatrix} D & -D \\ -D & D \end{pmatrix} \right)^{-1}$.

□

Weak Law of Large Numbers

Let $\boldsymbol{z}_i$ be a $q \times 1$ zero-mean vector that is independent over i. If for some $\xi > 0$, $\sup_{i\in\mathbb{N}} \mathbb{E}|\boldsymbol{z}_i|^{1+\xi} < \infty$, then $N^{-1}\sum_{i=1}^{N}\boldsymbol{z}_i \xrightarrow{p} \boldsymbol{0}$.

Central Limit Theorem

Let $\boldsymbol{z}_i$ be a $q \times 1$ zero-mean vector that is independent over i. Also let $\mathbb{E}[\boldsymbol{z}_i\boldsymbol{z}_i'] = \boldsymbol{\Omega}_i$, a $q \times q$ positive definite matrix, and $\boldsymbol{\Omega} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\boldsymbol{\Omega}_i$. If for some $\xi > 0$, $\sup_{i\in\mathbb{N}} \mathbb{E}|z_i|^{2+\xi} < \infty$, then $N^{-1/2}\sum_{i=1}^{N}\boldsymbol{z}_i \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega})$.