

Online supplement to “Estimation of panel group structure models with structural breaks in group memberships and coefficients”

Robin L. Lumsdaine* Ryo Okui† Wendun Wang‡

December 30, 2021

This supplement contains details of several extensions of the model and additional simulation and empirical results. Section S.1 presents the algorithm for models with time invariant group memberships. Sections S.2 and S.3 provide the algorithms and theoretical discussions on models with fixed effects using two different approaches, the former using first difference and the latter using within transformation. Section S.4 provides detailed discussions on models with multiple breaks. Section S.5 contains extra simulation studies, Section S.6 discusses computation time, and Section S.7 provides additional empirical results.

S.1 Time invariant group memberships: Estimation algorithm

To select between the specifications of time-varying and time-invariant group structures, a model with time-invariant group memberships needs to be estimated. Recall that the estimator is

$$(\hat{k}, \hat{\gamma}, \hat{\beta}) = \underset{k \in \mathbb{K}, \tilde{\gamma} \in \mathbb{G}^N, \beta \in \mathcal{B}}{\operatorname{argmin}} \left[\sum_{t=1}^{k-1} \sum_{i=1}^N (y_{it} - x'_{it} \beta_{\tilde{g}_i, B})^2 + \sum_{t=k}^T \sum_{i=1}^N (y_{it} - x'_{it} \beta_{\tilde{g}_i, A})^2 \right],$$

where $\hat{\tilde{g}}_i$ is the estimator of the *time-invariant* group membership of unit i , and $\hat{\gamma} = (\hat{\tilde{g}}_1, \dots, \hat{\tilde{g}}_N)$ is the estimator of $\tilde{\gamma} = (\tilde{g}_1, \dots, \tilde{g}_N) \in \mathbb{G}^N$. \hat{k} and $\hat{\beta}$ are the associated estimators of k and β , respectively, under the restriction of time-invariant group memberships. We modify Algorithm 1 in the main text to compute the estimator.

*Department of Finance and Real Estate, Kogod School of Business, American University; Econometric Institute, Erasmus University Rotterdam; National Bureau of Economic Research; Tinbergen Institute. Email: robin.lumsdaine@american.edu

†Department of Economics and the Institute of Economic Research, Seoul National University. Email: okuiryo@snu.ac.kr

‡Econometric Institute, Erasmus University Rotterdam; Tinbergen Institute. Email: wang@ese.eur.nl

Algorithm S.1. (Time-invariant group memberships)

Let s denote the iteration number.

Step 1: Set $s = 0$. For each $k \in \{2, \dots, T\}$, initialize the group structure as $\tilde{\gamma}^{(0)}$.

Step 2: For given $\tilde{\gamma}^{(s)}$ and k , estimate the slope coefficient $\beta^{(s)}$ in the two regimes by

$$\beta^{(s)} = \underset{\beta \in \mathcal{B}}{\operatorname{argmin}} \left[\sum_{t=1}^{k-1} \sum_{i=1}^N (y_{it} - x'_{it} \beta_{\tilde{g}_i^{(s)}, B})^2 + \sum_{t=k}^T \sum_{i=1}^N (y_{it} - x'_{it} \beta_{\tilde{g}_i^{(s)}, A})^2 \right].$$

Step 3: Given $\beta^{(s)}$, find the optimal group for unit i by

$$\tilde{g}_i^{(s+1)} = \underset{\tilde{\gamma} \in \mathbb{G}^N}{\operatorname{argmin}} \left[\sum_{t=1}^{k-1} (y_{it} - x'_{it} \beta_{\tilde{g}_i, B}^{(s)})^2 + \sum_{t=k}^T (y_{it} - x'_{it} \beta_{\tilde{g}_i, A}^{(s)})^2 \right].$$

Step 4: Iterate Steps 2 and 3 until numerical convergence, to obtain $\hat{\tilde{\gamma}}(k)$ and $\hat{\beta}(k)$.

Step 5: Let k vary from 2 to T , and estimate the break point by

$$\hat{k} = \underset{k \in \mathbb{K}}{\operatorname{argmin}} \left[\sum_{t=1}^{k-1} \sum_{i=1}^N (y_{it} - x'_{it} \hat{\beta}_{\hat{g}_i(k), B}(k))^2 + \sum_{t=k}^T \sum_{i=1}^N (y_{it} - x'_{it} \hat{\beta}_{\hat{g}_i(k), A}(k))^2 \right].$$

S.2 Models with fixed effects: first difference approach

This section provides theoretical results and proofs for models with fixed effects using the first difference approach. We first briefly review the discussion in the main text. An algorithm to compute our estimator tailored for fixed effects models is then presented. Lastly, we provide asymptotic results and their proofs.

S.2.1 Model and estimation method

We present the model again for the ease of reference. Suppose that we have panel data (y_{it}, x_{it}) for $i = 1, \dots, N$ and $t = 0, \dots, T$. Our model is

$$y_{it} = x'_{it} \beta_{g_{it}, t} + \alpha_i + u_{it},$$

where α_i is the fixed effect for unit i , u_{it} is an error term and $\beta_{g_{it}, t}$ is the coefficient which depends on time period t and the group membership of i at time t . A structural break occurs at time k^0 . The structural break possibly changes both the values of the coefficients and the group structure. Before the break, each unit belongs to one of the elements of

$\mathbb{G}^B = \{1, \dots, G^B\}$. After the break the set of groups becomes $\mathbb{G}^A = \{1, \dots, G^A\}$. The coefficient vector has the form:

$$\beta_{g_{it},t} = \begin{cases} \beta_{g_i(B),B} & \text{if } t < k^0 \\ \beta_{g_i(A),A} & \text{if } t \geq k^0 \end{cases},$$

where $g_i(B) \in \mathbb{G}^B$ stands for unit i 's group membership before the break and $g_i(A) \in \mathbb{G}^A$ stands for unit i 's group membership after the break.

The estimation method consists of two steps. First, we take the first difference to eliminate the fixed effects. The transformed variables satisfy:

$$\begin{aligned} \Delta y_{it} &= x'_{it}\beta_{g_{it},t} - x'_{i,t-1}\beta_{g_{i,t-1},t-1} + \Delta u_{it} \\ &= \begin{cases} \Delta x'_{it}\beta_{g_i(B),B} + \Delta u_{it} & \text{if } t < k^0 \\ x'_{it}\beta_{g_i(A),A} - x'_{i,t-1}\beta_{g_i(B),B} + \Delta u_{it} & \text{if } t = k^0 \\ \Delta x'_{it}\beta_{g_i(A),A} + \Delta u_{it} & \text{if } t > k^0 \end{cases} \end{aligned}$$

where Δ is the first difference operator (for example, $\Delta y_{it} = y_{it} - y_{i,t-1}$). Second, we estimate the coefficients, group membership structure, and the break date by minimizing the quadratic loss function defined in terms of first differenced variables. Let β be a vector stacking $\beta_{g,B}$ for $g \in \mathbb{G}^B$ and $\beta_{g,A}$ for $g \in \mathbb{G}^A$. The parameter space for β is \mathcal{B} which is a subset of $\mathbb{R}^{p(G^B+G^A)}$. Let $\Gamma = (\mathbb{G}^B \times \mathbb{G}^A)^N$ be the parameter space for group memberships. We denote an element of Γ as γ , and $\gamma_B = (g_1(B), \dots, g_N(B))$ denotes the group membership vector before the break and $\gamma_A = (g_1(A), \dots, g_N(A))$ is the vector after the break. Let $\mathbb{K} = \{2, \dots, T\}$ be the parameter space for the break date k . We estimate (k, γ, β) by minimizing the least squares criterion:

$$\begin{aligned} (\hat{k}, \hat{\gamma}, \hat{\beta}) = \underset{k \in \mathbb{K}, \gamma \in \Gamma, \beta \in \mathcal{B}}{\operatorname{argmin}} & \left[\sum_{t=1}^{k-1} \sum_{i=1}^N (\Delta y_{it} - \Delta x'_{it}\beta_{g_i(B),B})^2 \right. \\ & + \sum_{i=1}^N (\Delta y_{ik} - x'_{ik}\beta_{g_i(A),A} + x'_{i,k-1}\beta_{g_i(B),B})^2 \\ & \left. + \sum_{t=k+1}^T \sum_{i=1}^N (\Delta y_{it} - \Delta x'_{it}\beta_{g_i(A),A})^2 \right]. \end{aligned}$$

To solve this objective function, we propose the following algorithm.

Algorithm

Let s denote the iteration number.

Step 1: Set $s = 1$. For a given $k \in \{2, \dots, T - 1\}$, initialize group structures in both regimes as $\gamma_B^{(0)}$ and $\gamma_A^{(0)}$.

Step 2: For given $\gamma^{(s)}$ and k , estimate the slope coefficient $\beta^{(s)}$ in the two regimes by

$$\begin{aligned} \beta^{(s)} = \operatorname{argmin}_{\beta \in \mathcal{B}} & \left[\sum_{t=1}^{k-1} \sum_{i=1}^N (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^{(s)}(B), B})^2 \right. \\ & + \sum_{i=1}^N (\Delta y_{ik} - x'_{ik} \beta_{g_i^{(s)}(A), A} + x'_{i,k-1} \beta_{g_i^{(s)}(B), B})^2 \\ & \left. + \sum_{t=k+1}^T \sum_{i=1}^N (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^{(s)}(A), A})^2 \right]. \end{aligned}$$

Step 3: Given $\beta^{(s)}$, find the optimal group for individual i in each regime, respectively, by

$$\begin{aligned} \gamma^{(s+1)} = \operatorname{argmin}_{\gamma \in \Gamma} & \left[\sum_{t=1}^{k-1} \sum_{i=1}^N (\Delta y_{it} - \Delta x'_{it} \beta_{g_i(B), B}^{(s)})^2 \right. \\ & + \sum_{i=1}^N (\Delta y_{ik} - x'_{ik} \beta_{g_i(A), A}^{(s)} + x'_{i,k-1} \beta_{g_i(B), B}^{(s)})^2 \\ & \left. + \sum_{t=k+1}^T \sum_{i=1}^N (\Delta y_{it} - \Delta x'_{it} \beta_{g_i(A), A}^{(s)})^2 \right]. \end{aligned}$$

Step 4: Iterate Step 2 and 3 until numerical convergence, and obtain $\hat{\gamma}_B(k)$, $\hat{\gamma}_A(k)$, and $\hat{\beta}(k)$.

Step 5: Let k vary from 2 to T , and estimate the break point by

$$\begin{aligned} \hat{k} = \operatorname{argmin}_{k \in \mathbb{K}} & \left[\sum_{t=1}^{k-1} \sum_{i=1}^N (\Delta y_{it} - \Delta x'_{it} \hat{\beta}_{g_i(B,k), B}(k))^2 \right. \\ & + \sum_{i=1}^N (\Delta y_{ik} - x'_{ik} \beta_{g_i(A,k), A}(k) + x'_{i,k-1} \beta_{g_i(B,k), B}(k))^2 \\ & \left. + \sum_{t=k+1}^T \sum_{i=1}^N (\Delta y_{it} - \Delta x'_{it} \beta_{g_i(A,k), A}(k))^2 \right]. \end{aligned}$$

Step 1 is an initialization. In Step 2, we estimate the coefficient vector given group membership structure and break date. Step 3 in turn estimates the group membership structure given the

coefficients and break date. Note that in Steps 2 and 3, we cannot separate the minimization for pre-break periods and post-break periods. This is in contrast to the algorithm for models without fixed effects. The reason is that first differencing makes the equation for $t = k$ contain both pre- and post-break coefficients. While it creates slightly more complicated computation, we find that the computational time does not change much compared with that for models without fixed effects. The break point is estimated in Step 5 by minimizing the sum of squared residuals.

S.2.2 Assumptions

We list the assumptions pertaining to the fixed effects specification here for ease of reference. Recall that superscript 0, such as k^0 , indicates the true value, β_B is the vector stacking $\beta_{g,B}$ for $g \in \mathbb{G}^B$ and similarly, β_A is the vector stacking $\beta_{g,A}$ for $g \in \mathbb{G}^A$

Assumption 1.

(ii) \mathcal{B} is compact.

(vii) $k^0/T \rightarrow \tau \in (\epsilon, 1 - \epsilon)$ for $\epsilon > 0$ as $T \rightarrow \infty$.

(viii) There exists a constant $c > 0$ such that for any $g \neq \tilde{g}$ where $g, \tilde{g} \in \mathbb{G}^B$ and $g', \tilde{g}' \in \mathbb{G}^A$, it holds that $\|\beta_{g,B}^0 - \beta_{\tilde{g},B}^0\| > c$ and $\|\beta_{g',A}^0 - \beta_{\tilde{g}',A}^0\| > c$.

Assumption 2.

(i) For any $L \subseteq \{1, \dots, N\}$ and $t'' \geq t'$, there exists M which does not depend on L , t'' nor t' such that the following equality holds

$$E \left(\left\| \frac{1}{NT} \sum_{t=t'}^{t''} \sum_{i \in L} x_{it+l} u_{it+j} \right\|^2 \right) = M \frac{|L|(t'' - t')}{NT^2}.$$

for $l, j = 0, -1$, where $|L|$ is the cardinality of L .

(iii) Let $\rho_{D,N,t}(\gamma^t, g, \tilde{g})$ be the minimum eigenvalue of $\sum_{i=1}^N \mathbf{1}\{g_{it}^0 = g\} \{g_{it} = \tilde{g}\} \Delta x_{it} \Delta x'_{it} / N$, where γ^t is either γ_B (when $t < k$) or γ_A (when $t \geq k$). For any $g \in \mathbb{G}^B$,

$$\min_{1 \leq t < k^0} \min_{\gamma_B} \max_{\tilde{g} \in \mathbb{G}^B} \rho_{D,N,t}(\gamma_B, g, \tilde{g}) > \hat{\rho}_D,$$

and for any $g \in \mathbb{G}^A$,

$$\min_{k^0 < t \leq T} \min_{\gamma_A} \max_{\tilde{g} \in \mathbb{G}^A} \rho_{D,N,t}(\gamma_A, g, \tilde{g}) > \hat{\rho}_D,$$

where $\hat{\rho}_D \rightarrow_p \rho_D > 0$.

(iv) There exists $\hat{\rho}_D^*$ such that for any i and for s such that s and $T - s$ sufficiently large,

$$\lambda_{\min} \left(\frac{1}{s} \sum_{t=1}^s \Delta x_{it} \Delta x'_{it} \right) \geq \hat{\rho}_D^* \quad \text{and} \quad \lambda_{\min} \left(\frac{1}{T-s} \sum_{t=s+1}^T \Delta x_{it} \Delta x'_{it} \right) \geq \hat{\rho}_D^*,$$

and $\hat{\rho}_D^* \rightarrow_p \rho_D^* > 0$.

(v) $\max_{1 \leq t \leq T} \sum_{i=1}^N \|x_{it}\|^2 / N = O_p(1)$ and $\max_{1 \leq t \leq T} \sum_{i=1}^N \|\Delta x_{it}\|^2 / N = O_p(1)$.

(vi) There exists a fixed constant $\underline{m} > 0$ independent on T and N , such that for any t ,

$$\frac{1}{N} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i(A), A}^0 - \beta_{g_i(B), B}^0))^2 > \underline{m}.$$

(ix) Let z_{it} be $\Delta x'_{it} \Delta x_{it}$, $\|\Delta u_{it} \Delta x_{it}\|$, $2\Delta u_{it} \Delta x'_{it} (\beta_{g_i(l), l}^0 - \beta_{g, l}^0)$ or $(\Delta x'_{it} (\beta_{g_i(l), l}^0 - \beta_{g, l}^0))^2$, for $g \in \mathbb{G}_l$ and $l = A, B$. Assume the following holds for any choice of z_{it} . 1) z_{it} is a strong mixing sequence over t whose mixing coefficients $a_i[t]$ are bounded by $a_i[t] \leq e^{-at^{d_1}}$ such that $\max_{1 \leq i \leq N} a_i[t] \leq a[t]$ and has tail probabilities $\max_{1 \leq i \leq N} \Pr(|z_{it}| > z) \leq e^{1-(z/b)^{d_2}}$ for any t where a , b , d_1 and d_2 are positive constants. 2) There exists a_i , $i = 1, \dots, N$ such that for any $\epsilon > 0$, it holds that $\max_{1 \leq i \leq N} |a_i - \sum_{t=1}^T E(z_{it})/T| < \epsilon$ for T sufficiently large.

(x) $\max_{1 \leq t \leq T} E(\|\sum_{i=1}^N x_{it} u_{it+l} / \sqrt{N}\|^{2+\delta})$ is bounded for some $\delta > 0$ where $l = 0, -1$.

S.2.3 Asymptotic results

We have the following asymptotic results for models with fixed effects. They are similar to Theorem 1 and Corollary 1 in the main text.

Theorem S.1. Suppose that Assumptions 1(ii), 1(vii), 1(viii) and 3 holds. As $N, T \rightarrow \infty$ with $NT^{-\delta} \rightarrow 0$ for some $\delta > 0$, $\Pr(\hat{k} = k^0) \rightarrow 1$.

Corollary S.1. Suppose that Assumptions 1(ii), 1(vii), 1(viii) and 3 holds. As $N, T \rightarrow \infty$ with $NT^{-\delta} \rightarrow 0$ for some $\delta > 0$, $\Pr(\hat{\gamma} = \gamma^0) \rightarrow 1$ and $\hat{\beta} = \tilde{\beta} + o_p(1/\sqrt{NT})$, where $\tilde{\beta}$ is the estimator of β under $k = k^0$ and $\gamma = \gamma^0$.

S.2.4 Proofs

This section includes the proofs of Theorem S.1 and Corollary S.1. First, Section S.4.3 presents the lemmas. The proofs of the theorem and corollary are included in Sections S.4.3 and S.4.3, respectively. Hereafter, C denotes a generic constant whose exact value depends on the content but it does not depend on N , T , or on the values of the parameters.

S.2.4.1 Lemmas

Let

$$\begin{aligned} Q(k, \gamma, \beta) = & \frac{1}{NT} \left(\sum_{t=1}^{k-1} \sum_{i=1}^N (\Delta y_{it} - \Delta x'_{it} \beta_{g_i(B), B})^2 \right. \\ & + \sum_{i=1}^N (\Delta y_{ik} - x'_{ik} \beta_{g_i(A), A} + x'_{i,k-1} \beta_{g_i(B), B})^2 \\ & \left. + \sum_{t=k+1}^T \sum_{i=1}^N (\Delta y_{it} - \Delta x'_{it} \beta_{g_i(A), A})^2 \right), \end{aligned}$$

and

$$\tilde{Q}(k, \gamma, \beta) = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (x'_{it} (\beta_{g_{it}, t}^0 - \beta_{g_{it}, t}) - x'_{i,t-1} (\beta_{g_{i,t-1}, t-1}^0 - \beta_{g_{i,t-1}, t-1}))^2 + \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \Delta u_{it}^2.$$

Note that for $k > k^0$, $\tilde{Q}(k, \gamma, \beta)$ is written as:

$$\begin{aligned} \tilde{Q}(k, \gamma, \beta) = & \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i(B), B}^0 - \beta_{g_i(B), B}))^2 \\ & + \frac{1}{NT} \sum_{i=1}^N (x'_{ik^0} \beta_{g_i(A), A}^0 - x'_{i,k^0-1} \beta_{g_i(B), B}^0 - \Delta x'_{ik^0} \beta_{g_i(B), B})^2 \\ & + \frac{1}{NT} \sum_{t=k^0+1}^{k-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i(A), A}^0 - \beta_{g_i(B), B}))^2 \\ & + \frac{1}{NT} \sum_{i=1}^N (\Delta x'_{ik} \beta_{g_i(A), A}^0 - x'_{i,k} \beta_{g_i(A), A} + x'_{i,k-1} \beta_{g_i(B), B})^2 \\ & + \frac{1}{NT} \sum_{t=k+1}^T \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i(A), A}^0 - \beta_{g_i(A), A}))^2 + \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \Delta u_{it}^2. \end{aligned}$$

When $k = k^0$, it is

$$\begin{aligned} \tilde{Q}(k, \gamma, \beta) = & \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i(B), B}^0 - \beta_{g_i(B), B}))^2 \\ & + \frac{1}{NT} \sum_{i=1}^N (x'_{ik^0} (\beta_{g_i(A), A}^0 - \beta_{g_i(A), A}) - x'_{i,k^0-1} (\beta_{g_i(B), B}^0 - \beta_{g_i(B), B}))^2 \\ & + \frac{1}{NT} \sum_{t=k^0+1}^T \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i(A), B}^0 - \beta_{g_i(A), B}))^2 + \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \Delta u_{it}^2. \end{aligned}$$

Lastly, when $k < k^0$, it is

$$\tilde{Q}(k, \gamma, \beta) = \frac{1}{NT} \sum_{t=1}^{k-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i(B), B}^0 - \beta_{g_i(B), B}))^2$$

$$\begin{aligned}
& + \frac{1}{NT} \sum_{i=1}^N (\Delta x'_{ik} \beta_{g_i^0(B), B}^0 - x'_{ik} \beta_{g_i(A), A} + x'_{i,k-1} \beta_{g_i(B), B})^2 \\
& + \frac{1}{NT} \sum_{t=k+1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \beta_{g_i(A), A}))^2 \\
& + \frac{1}{NT} \sum_{i=1}^N (x'_{ik^0} \beta_{g_i^0(A), A}^0 - x'_{i,k^0-1} \beta_{g_i^0(B), B}^0 - \Delta x'_{ik^0} \beta_{g_i(A), A})^2 \\
& + \frac{1}{NT} \sum_{t=k^0+1}^T \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(A), B}^0 - \beta_{g_i(A), A}))^2 + \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \Delta u_{it}^2.
\end{aligned}$$

Lemma S.1. Suppose that Assumption 1(ii) and 2(i) and hold. Then we have

$$\sup_{k \in \mathbb{K}, \gamma \in \mathbb{G}, \beta \in \mathbb{B}} |\tilde{Q}(k, \gamma, \beta) - Q(k, \gamma, \beta)| = O_p \left(\frac{1}{\sqrt{T}} \right).$$

Proof. The proof is almost the same as the proof of Lemma S.3 of Bonhomme and Manresa (2015). First we consider the case in which $k > k^0$. We have

$$\begin{aligned}
& \tilde{Q}(k, \gamma, \beta) - Q(k, \gamma, \beta) \\
& = -2 \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \Delta u_{it} (\Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \beta_{g_i(B), B})) \\
& \quad - 2 \frac{1}{NT} \sum_{i=1}^N \Delta u_{ik^0} (x'_{ik^0} \beta_{g_i^0(A), A}^0 - x'_{i,k^0-1} \beta_{g_i^0(B), B}^0 - \Delta x'_{ik^0} \beta_{g_i(B), B}) \\
& \quad - 2 \frac{1}{NT} \sum_{t=k^0+1}^{k-1} \sum_{i=1}^N \Delta u_{it} (\Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(B), B})) \\
& \quad - 2 \frac{1}{NT} \sum_{i=1}^N \Delta u_{ik} (\Delta x'_{ik} \beta_{g_i^0(A), A}^0 - x'_{i,k} \beta_{g_i(A), A} + x'_{i,k-1} \beta_{g_i(B), B}) \\
& \quad - 2 \frac{1}{NT} \sum_{t=k+1}^T \sum_{i=1}^N \Delta u_{it} (\Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(A), A})).
\end{aligned}$$

We observe that

$$\begin{aligned}
& \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \Delta u_{it} (\Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \beta_{g_i(B), B})) \\
& = \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{it} \beta_{g_i^0(B), B}^0 u_{it} - \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{it} \beta_{g_i(B), B} u_{it} - \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{it} \beta_{g_i^0(B), B}^0 u_{i,t-1} \\
& \quad + \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{it} \beta_{g_i(B), B} u_{i,t-1} - \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{i,t-1} \beta_{g_i^0(B), B}^0 u_{it} + \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{i,t-1} \beta_{g_i(B), B} u_{it}
\end{aligned}$$

$$+ \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{i,t-1} \beta_{g_i(B),B}^0 u_{i,t-1} - \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{i,t-1} \beta_{g_i(B),B} u_{i,t-1}.$$

The first term is

$$\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{it} \beta_{g_i(B),B}^0 u_{it} = \frac{1}{NT} \sum_{g \in \mathbb{G}^B} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \mathbf{1}(g_i(B) = g) x'_{it} \beta_{g_i(B),B}^0 u_{it}.$$

For each $g \in \mathbb{G}^B$, by the Cauchy-Schwarz inequality we have

$$\begin{aligned} & E \left(\left(\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \mathbf{1}(g_i(B) = g) x'_{it} \beta_{g_i(B),B}^0 u_{it} \right)^2 \right) \\ & \leq BE \left(\left\| \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{g_i(B)=g} x_{it} u_{it} \right\|^2 \right) \\ & = O \left(\frac{k^0}{NT^2} \right), \end{aligned}$$

where C satisfies $\|\beta_{g_i(t),t}\|^2 < C$ for any $\beta \in \mathcal{B}$ and the existence of such C is guaranteed by Assumption 1(ii), the inequality follows by the definition of C , and the equality follows by Assumption 2(i). Next, we consider

$$\begin{aligned} \left(\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{it} \beta_{g_i(B),B} u_{it} \right)^2 & \leq \left(\frac{1}{NT} \sum_{i=1}^N \beta_{g_i(B),B} \sum_{t=1}^{k^0-1} x_{it} u_{it} \right)^2 \\ & \leq \left(\frac{1}{N} \sum_{i=1}^N \|\beta_{g_i(B),B}\|^2 \right) \left(\frac{1}{NT^2} \sum_{i=1}^N \left\| \sum_{t=1}^{k^0-1} x_{it} u_{it} \right\|^2 \right) \\ & = O_p \left(\frac{k^0}{T^2} \right), \end{aligned}$$

where the first inequality is the Cauchy-Schwarz inequality and the second inequality follows by that Assumption 1(ii) implies $\sum_{i=1}^N \|\beta_{g_i(B),B}\|^2 / N < C$ for some C , and that Assumption 2(i) together with the Markov inequality implies $\sum_{i=1}^N \left\| \sum_{t=1}^{k^0-1} x_{it} u_{it} \right\|^2 / (NT^2) = O_p(k^0/T^2)$.

The other terms in the expression for $\tilde{Q}(k, \gamma, \beta) - Q(k, \gamma, \beta)$ can be analyzed similarly. It therefore holds that

$$\begin{aligned} & \tilde{Q}(k, \gamma, \beta) - Q(k, \gamma, \beta) \\ & = O \left(\frac{\sqrt{k^0}}{T\sqrt{N}} \right) + O \left(\frac{\sqrt{k^0}}{T} \right) + O \left(\frac{1}{T\sqrt{N}} \right) + O \left(\frac{1}{T} \right) + O \left(\frac{\sqrt{k-k^0}}{T\sqrt{N}} \right) + O \left(\frac{\sqrt{k-k^0}}{T} \right) \\ & \quad + O \left(\frac{1}{T\sqrt{N}} \right) + O \left(\frac{1}{T} \right) + O \left(\frac{\sqrt{T-k}}{T\sqrt{N}} \right) + O \left(\frac{\sqrt{T-k}}{T} \right), \end{aligned}$$

uniformly over β and γ . The argument for cases with $k = k^0$ and $k < k^0$ is similar. Because $k \leq T$ by construction, we have

$$\sup_{k \in \mathbb{K}, \gamma \in \mathbb{G}, \beta \in \mathbb{B}} |\tilde{Q}(k, \gamma, \beta) - Q(k, \gamma, \beta)| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

□

Lemma S.2. Suppose that Assumptions 1(ii), 1(vii), and 2(i)–2(vi) hold. Then, we have that

- (1) $\max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \left\| \beta_{g,B}^0 - \hat{\beta}_{\tilde{g},B} \right\|^2 = O_p(1/\sqrt{T}),$
- (2) $\max_{g \in \mathbb{G}^A} \min_{\tilde{g} \in \mathbb{G}^A} \left\| \beta_{g,A}^0 - \hat{\beta}_{\tilde{g},A} \right\|^2 = O_p(1/\sqrt{T}),$
- (3) $(\hat{k} - k^0)/T = O_p(1/\sqrt{T}).$

Proof. From Lemma S.6, we have

$$\begin{aligned} \tilde{Q}(\hat{k}, \hat{\gamma}, \hat{\beta}) &= Q(\hat{k}, \hat{\gamma}, \hat{\beta}) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &\leq Q(k^0, \gamma^0, \beta^0) + O_p\left(\frac{1}{\sqrt{T}}\right) = \tilde{Q}(k^0, \gamma^0, \beta^0) + O_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

Because $\tilde{Q}(k, \gamma, \beta)$ is minimized at (k^0, γ^0, β^0) , we have

$$\tilde{Q}(\hat{k}, \hat{\gamma}, \hat{\beta}) - \tilde{Q}(k^0, \gamma^0, \beta^0) = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Let $a_{NT} = \tilde{Q}(\hat{k}, \hat{\gamma}, \hat{\beta}) - \tilde{Q}(k^0, \gamma^0, \beta^0)$. Note that $a_{NT} = O_p(1/\sqrt{T})$.

Consider the case in which $k > k^0$. We observe that

$$\begin{aligned} \tilde{Q}(k, \gamma, \beta) - \tilde{Q}(k^0, \gamma^0, \beta^0) &= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \beta_{g_i(B), B}))^2 \\ &\quad + \frac{1}{NT} \sum_{i=1}^N (x'_{ik^0} \beta_{g_i^0(A), A}^0 - x'_{i,k^0-1} \beta_{g_i^0(B), B}^0 - \Delta x'_{ik^0} \beta_{g_i(B), B})^2 \\ &\quad + \frac{1}{NT} \sum_{t=k^0+1}^{k-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(B), B}))^2 \\ &\quad + \frac{1}{NT} \sum_{i=1}^N (\Delta x'_{ik} \beta_{g_i^0(A), A}^0 - x'_{i,k} \beta_{g_i(A), A} + x'_{i,k-1} \beta_{g_i(B), B})^2 \\ &\quad + \frac{1}{NT} \sum_{t=k+1}^T \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(A), A}))^2. \end{aligned}$$

We consider the first term. It holds that

$$\begin{aligned}
& \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}))^2 \\
&= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{g=1}^{G^B} \sum_{\tilde{g}=1}^{G^B} \sum_{i=1}^N \mathbf{1}\{g_i^0(B) = g\} \{g_i(B) = \tilde{g}\} (\Delta x'_{it}(\beta_{g,B}^0 - \beta_{\tilde{g},B}))^2 \\
&\geq \frac{1}{T} \sum_{t=1}^{k^0-1} \sum_{g=1}^{G^B} \sum_{\tilde{g}=1}^{G^B} \rho_{D,N,t}(\gamma, g, \tilde{g}) \|\beta_{g,B}^0 - \beta_{\tilde{g},B}\|^2.
\end{aligned}$$

We thus have

$$\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}))^2 \geq \frac{k^0-1}{T} \hat{\rho}_D \max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \|\beta_{g,B}^0 - \beta_{\tilde{g},B}\|^2, \quad (\text{S.1})$$

by Assumption 2(iv). Moreover, Assumption 2(iv) implies

$$\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}))^2 \geq \frac{k^0-1}{T} \hat{\rho}_D^* \frac{1}{N} \sum_{i=1}^N \|\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}\|^2.$$

Thus, we have

$$\frac{1}{N} \sum_{i=1}^N \|\beta_{g_i^0(B),B}^0 - \hat{\beta}_{g_i(B),B}\|^2 \leq C a_{NT} \quad (\text{S.2})$$

Similarly, it follows that

$$\frac{1}{NT} \sum_{t=k+1}^T \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A}))^2 \geq \frac{T-k-1}{T} \hat{\rho} \max_{g \in \mathbb{G}^A} \min_{\tilde{g} \in \mathbb{G}^A} \|\beta_{g,A}^0 - \beta_{\tilde{g},A}\|^2.$$

For the second term, we observe that

$$\begin{aligned}
& \frac{1}{NT} \sum_{i=1}^N (x'_{ik^0} \beta_{g_i^0(A),A}^0 - x'_{i,k^0-1} \beta_{g_i^0(B),B}^0 - \Delta x'_{ik^0} \beta_{g_i(B),B})^2 \\
&= \frac{1}{NT} \sum_{i=1}^N (x'_{ik^0} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) + \Delta x'_{ik^0} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}))^2 \\
&\geq \frac{1}{NT} \sum_{i=1}^N (x'_{ik^0} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0))^2 + \frac{1}{NT} \sum_{i=1}^N (\Delta x'_{ik^0} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}))^2 \\
&\quad - 2 \frac{1}{NT} \sum_{i=1}^N \left| x'_{ik^0} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right| \cdot \left| \Delta x'_{ik^0} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right| \\
&\geq \frac{1}{NT} \sum_{i=1}^N (x'_{ik^0} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0))^2 + \frac{1}{NT} \sum_{i=1}^N (\Delta x'_{ik^0} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}))^2
\end{aligned}$$

$$- 2 \frac{1}{T} \left(\frac{1}{N} \sum_{i=1}^N x'_{ik^0} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0)^2 \right)^{1/2} \left(\frac{1}{N} \sum_{i=1}^N (\Delta x'_{ik^0} (\beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B}))^2 \right)^{1/2}.$$

Assumptions 1(ii) and 2(v) implies that

$$\frac{1}{N} \sum_{i=1}^N x'_{ik^0} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0)^2 = O_p(1).$$

Also the Cauchy-Schwarz inequality, Assumption 2(v) and (S.2) imply

$$\frac{1}{N} \sum_{i=1}^N (\Delta x'_{ik^0} (\beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B})) < C a_{NT}.$$

Hence, Assumption 2(vi) implies with probability approaching one,

$$\begin{aligned} & \frac{1}{NT} \sum_{i=1}^N (x'_{ik^0} \beta_{g_i^0(A), A}^0 - x'_{i, k^0-1} \beta_{g_i^0(B), B}^0 - \Delta x'_{ik^0} \hat{\beta}_{\hat{g}_i(B), B})^2 \\ & \geq \frac{1}{T} (\underline{m} + C - 2C\sqrt{a_{NT}}) \geq \frac{1}{T} (\underline{m} - C\sqrt{a_{NT}}) \end{aligned}$$

For the third term, we have

$$\begin{aligned} & \frac{1}{NT} \sum_{t=k^0+1}^{\hat{k}-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \hat{\beta}_{\hat{g}_i(B), B}))^2 \\ & = \frac{1}{NT} \sum_{t=k^0+1}^{\hat{k}-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0 + \beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B}))^2 \\ & \geq \frac{1}{NT} \sum_{t=k^0+1}^{\hat{k}-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0))^2 + \frac{1}{NT} \sum_{t=k^0+1}^{\hat{k}-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B}))^2 \\ & \quad - 2 \frac{1}{NT} \sum_{t=k^0+1}^{\hat{k}-1} \sum_{i=1}^N |\Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0)| \cdot |\Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B})| \\ & \geq \frac{1}{NT} \sum_{t=k^0+1}^{\hat{k}-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0))^2 + \frac{1}{NT} \sum_{t=k^0+1}^{\hat{k}-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B}))^2 \\ & \quad - 2 \frac{1}{T} \sum_{t=k^0+1}^{\hat{k}-1} \left(\frac{1}{N} \sum_{i=1}^N \Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0)^2 \right)^{1/2} \left(\frac{1}{N} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B}))^2 \right)^{1/2}. \end{aligned}$$

Assumptions 1(ii) and 2(v) imply that

$$\frac{1}{N} \sum_{i=1}^N \Delta x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0)^2 = O_p(1),$$

uniformly across t . Moreover, the Cauchy-Schwarz inequality, Assumption 2(v) and (S.2) imply that with probability approaching one,

$$\frac{1}{N} \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \hat{\beta}_{\tilde{g}_i(B),B}))^2 = O_p(a_{NT}).$$

Thus we have with probability approaching one,

$$\begin{aligned} & \frac{1}{NT} \sum_{t=k^0+1}^{\hat{k}-1} \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(A),A}^0 - \hat{\beta}_{\tilde{g}_i(B),B}))^2 \\ & \geq \frac{\hat{k} - k^0 - 1}{T} (\underline{m} + C - 2C\sqrt{a_{NT}}) = \frac{\hat{k} - k^0 - 1}{T} (\underline{m} - C\sqrt{a_{NT}}). \end{aligned}$$

It thus holds that for $\hat{k} > k^0$,

$$\frac{1}{T} (\underline{m} - C\sqrt{a_{NT}}) + \frac{\hat{k} - k^0 - 1}{T} (\underline{m} - C\sqrt{a_{NT}}) = \frac{\hat{k} - k^0}{T} (\underline{m} - C\sqrt{a_{NT}}) \leq a_{NT}.$$

We can use a similar argument to show that for $\hat{k} < k^0$ we have

$$\frac{k^0 - \hat{k}}{T} (\underline{m} - C\sqrt{a_{NT}}) \leq a_{NT}.$$

Thus we have

$$\frac{\hat{k} - k^0}{T} = O_p(a_{NT}) = O_p\left(\frac{1}{\sqrt{T}}\right). \quad (\text{S.3})$$

Next, we have that by (S.1), for $\hat{k} > k^0$,

$$\frac{k^0 - 1}{T} \hat{\rho}_D \max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \left\| \beta_{g,B}^0 - \hat{\beta}_{\tilde{g},B} \right\|^2 \leq a_{NT}.$$

A similar argument to that use to show (S.1) gives that for $\hat{k} < k^0$,

$$\frac{\hat{k} - 1}{T} \hat{\rho}_D \max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \left\| \beta_{g,B}^0 - \hat{\beta}_{\tilde{g},B} \right\|^2 \leq a_{NT}.$$

Thus, by (S.3), we have

$$\max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \left\| \beta_{g,B}^0 - \hat{\beta}_{\tilde{g},B} \right\|^2 = O_p(a_{NT}) = O_p\left(\frac{1}{\sqrt{T}}\right). \quad (\text{S.4})$$

Lastly, by following a similar argument for (S.4), we have

$$\max_{g \in \mathbb{G}^A} \min_{\tilde{g} \in \mathbb{G}^A} \left\| \beta_{g,A}^0 - \hat{\beta}_{\tilde{g},A} \right\|^2 = O_p\left(\frac{1}{\sqrt{T}}\right).$$

□

Lemma S.3. Suppose that Assumptions 1(ii), 1(vii), 1(viii), and 2(i)-2(vi) are satisfied. Then there exist permutations $\sigma_B : \mathbb{G}^B \mapsto \mathbb{G}^B$ and $\sigma_A : \mathbb{G}^A \mapsto \mathbb{G}^A$ such that $\|\beta_{g,B}^0 - \hat{\beta}_{\sigma_B(g),B}\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}^B$ and $\|\beta_{g,A}^0 - \hat{\beta}_{\sigma_A(g),A}\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}^A$.

Proof. The proof is exactly identical to that of Lemma 3 in the main text and is thus omitted. \square

By relabeling, we can set $\sigma_B(g) = g$ and $\sigma_A(g) = g$. We use this convention throughout the paper. Thus we have $\|\beta_{g,B}^0 - \hat{\beta}_{g,B}\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}^B$ and $\|\beta_{g,A}^0 - \hat{\beta}_{g,A}\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}^A$.

Let \mathcal{N} be a neighborhood of β^0 such that $\|\beta_{g,C}^0 - \beta_{g,C}\| < \eta$ for $\eta > 0$ for any $g \in \mathbb{G}^C$ and $C = B, A$. Note that we will take η small enough by considering large N and T by Lemma S.8. Let $\bar{k} = \sqrt{T} \log T + k^0$ and $\underline{k} = -\sqrt{T} \log T + k^0$. Define $K = \{k : \underline{k} \leq k \leq \bar{k}\}$.

Lemma S.4. Suppose that Assumptions 1(ii), 1(vii), 1(viii), 2(v) and 2(ix) hold. As $N, T \rightarrow \infty$ with $NT^{-\delta} \rightarrow 0$, it holds that

$$\Pr(\hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N}) \rightarrow 0.$$

Proof. To show $\Pr(\hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N}) \rightarrow 0$, it is equivalent to show that

$$\max_{1 \leq i \leq N} \sup_{\beta \in \mathcal{N}} \max_{k \in K} \mathbf{1}(\hat{g}_i(B)(k, \beta) \neq g_i^0(B) \text{ or } \hat{g}_i(A)(k, \beta) \neq g_i^0(A)) = o_p(1).$$

We observe that

$$\begin{aligned} & \max_{1 \leq i \leq N} \sup_{\beta \in \mathcal{N}} \max_{k \in K} \mathbf{1}(\hat{g}_i(B)(k, \beta) \neq g_i^0(B) \text{ or } \hat{g}_i(A)(k, \beta) \neq g_i^0(A)) \\ &= \max_{(g_A, g_B) \in \mathbb{G}^B \times \mathbb{G}^A \setminus \{g_i^0(B), g_i^0(A)\}} \mathbf{1} \left(\sum_{t=1}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_B})^2 \right. \\ & \quad + (\Delta y_{ik} - x'_{ik} \beta_{g_A} + x'_{i,k-1} \beta_{g_B})^2 + \sum_{t=k+1}^T (\Delta y_{it} - \Delta x'_{it} \beta_{g_A})^2 \\ & < \sum_{t=1}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^0(B), B})^2 + (\Delta y_{ik} - x'_{ik} \beta_{g_i^0(A), A}^0 + x'_{i,k-1} \beta_{g_i^0(B), B})^2 \\ & \quad \left. + \sum_{t=k+1}^T (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^0(A), A})^2 \right) \\ & < \max_{(g_A, g_B) \in \mathbb{G}^B \times \mathbb{G}^A \setminus \{g_i^0(B), g_i^0(A)\}} \mathbf{1} \left(\sum_{t=1}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_B})^2 \right. \\ & \quad + (\Delta y_{ik} - x'_{ik} \beta_{g_A} + x'_{i,k-1} \beta_{g_B})^2 \end{aligned}$$

$$\begin{aligned}
&< \sum_{t=1}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^0(B), B})^2 + (\Delta y_{ik} - x'_{ik} \beta_{g_i(A), A}^0 + x'_{i,k-1} \beta_{g_i(B)^0, B})^2 \Big) \\
&+ \max_{(g_A) \in \mathbb{G}^A \setminus \{g_i^0(A)\}} \mathbf{1} \left(\sum_{t=k+1}^T (\Delta y_{it} - \Delta x'_{it} \beta_{g_A})^2 < \sum_{t=k+1}^T (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^0(A), A})^2 \right)
\end{aligned}$$

We first consider cases with $k > k^0$. Let

$$\begin{aligned}
d = & \sum_{t=k^0}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_B})^2 + (\Delta y_{ik} - x'_{ik} \beta_{g_A} + x'_{i,k-1} \beta_{g_B})^2 \\
& - \sum_{t=k^0}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^0(B), B})^2 - (\Delta y_{ik} - x'_{ik} \beta_{g_i^0(A), A} + x'_{i,k-1} \beta_{g_i^0(B), B})^2.
\end{aligned}$$

We observe that

$$\begin{aligned}
d = & 2 \sum_{t=k^0}^{k-1} \Delta u_{it} \Delta x'_{it} (\beta_{g_i^0(B), B} - \beta_{g_i(B), B}) \\
& + (2x'_{ik^0} \beta_{g_i^0(A), A}^0 - 2x'_{i,k^0-1} \beta_{g_i^0(B), B}^0 - \Delta x'_{it} (\beta_{g_i^0(B), B} + \beta_{g_i(B), B})) \Delta x'_{ik^0} (\beta_{g_i^0(B), B} - \beta_{g_i(B), B}) \\
& + \sum_{t=k^0+1}^{k-1} (\beta_{g_i^0(B), B} - \beta_{g, B})' \Delta x_{it} \Delta x'_{it} (2\beta_{g_i^0(B), B}^0 - \beta_{g_i^0(B), B} - \beta_{g, B}) \\
& + 2\Delta u_{ik} (x'_{ik} (\beta_{g_i^0(A), A} - \beta_{g_A}) - x'_{i,k-1} (\beta_{g_i^0(B), B} - \beta_{g_B})) \\
& + (x'_{ik} (\beta_{g_i^0(A), A} - \beta_{g_A}) - x'_{i,k-1} (\beta_{g_i^0(B), B} - \beta_{g_B})) \\
& \times (2\Delta x'_{ik} \beta_{g_i^0(A), A}^0 - x'_{ik} (\beta_{g_i^0(A), A} + \beta_{g_A}) - x'_{i,k-1} (\beta_{g_i^0(B), B} + \beta_{g_B}))
\end{aligned}$$

Thus we have

$$\begin{aligned}
|d| \leq & M_1 \left\| \sum_{t=k^0}^{k-1} \Delta u_{it} \Delta x_{it} \right\| + M_2 \left\| \sum_{t=k^0+1}^{k^0-1} \Delta x_{it} \Delta x'_{it} \right\| + Z \\
\leq & M_1 (k^0 - \underline{k}) \frac{1}{\bar{k} - k^0} \sum_{t=k^0}^{\bar{k}} \|\Delta u_{it} \Delta x_{it}\| + M_2 (k^0 - \underline{k}) \left\| \frac{1}{\bar{k} - k^0} \sum_{t=k^0}^{\bar{k}} \Delta x_{it} \Delta x'_{it} \right\| + Z,
\end{aligned}$$

where $Z = O_p(1)$ independent of T and M_1 and M_2 are constants independent of (i, g, k, β) . Let $M_T = T^{1/4}/\log T$. Under Assumption 2(ix), we can apply inequality (1.8) in Merlevède et al. (2011) which is based on Theorem 6.2 of Rio (2017) with $\lambda = (\bar{k} - k^0)M_T = T^{3/4}$ and obtain

$$\begin{aligned}
& \Pr \left(\frac{1}{\bar{k} - k^0} \left| \sum_{t=k^0}^{\bar{k}} (\|\Delta u_{it} \Delta x_{it}\| - E(\|\Delta u_{it} \Delta x_{it}\|)) \right| > M_T \right) \\
& \leq 4 \exp \left(-\frac{\lambda^{d/(d+1)} \log 2}{2} \right) + 16CM_T^{-1} \exp \left(-a \frac{\lambda^{d/(d+1)}}{b^d} \right)
\end{aligned}$$

$$= o(T^{-\delta}),$$

where $d = d_1 d_2 / (d_1 + d_2)$ and C is a positive constant. Noting that $\sum_{t=k^0}^{\bar{k}} E(\|\Delta u_{it} \Delta x_{it}\|) / (\bar{k} - k^0)$ converges and $M_T \rightarrow \infty$, we have

$$\Pr \left(\frac{1}{\bar{k} - k^0} \sum_{t=k^0}^{\bar{k}} \|\Delta u_{it} \Delta x_{it}\| > M_T \right) = o(T^{-\delta}).$$

Similarly, we have

$$\Pr \left(\left\| \frac{1}{\bar{k} - k^0} \sum_{t=k^0}^{\bar{k}} \Delta x_{it} \Delta x'_{it} \right\| > M_T \right) = o(T^{-\delta}).$$

This implies that there exists a sequence such that $C_T = O(M_T)$ and $C_T \rightarrow \infty$ as $T \rightarrow \infty$ such that

$$\Pr \left(\frac{1}{k^0} |d| > \frac{k^0 - k}{k^0} C_T \right) = o(T^{-\delta}).$$

By a similar argument, the above bound holds for $k \leq k^0$.

Next, we consider the main term. This term can be considered in a similar way to [Bonhomme and Manresa \(2015\)](#) and [Okui and Wang \(2021\)](#).

$$\begin{aligned} & \sum_{t=1}^{k^0-1} \left((\Delta y_{it} - \Delta x'_{it} \beta_{g,B})^2 - (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^0(B),B}^0)^2 \right) \\ &= \sum_{t=1}^{k^0-1} 2\Delta u_{it} \Delta x_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}) + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B}^0 - \beta_{g,B})' \Delta x_{it} \Delta x'_{it} (2\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}^0 - \beta_{g,B}) \\ &= \sum_{t=1}^{k^0-1} 2\Delta u_{it} \Delta x_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \\ &\quad + \sum_{t=1}^{k^0-1} 2\Delta u_{it} \Delta x_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0 - \beta_{g_i^0(B),B}^0 + \beta_{g,B}^0) \\ &\quad + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0)' \Delta x_{it} \Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \\ &\quad + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0 - \beta_{g_i^0(B),B}^0 + \beta_{g,B}^0)' \Delta x_{it} \Delta x'_{it} (2\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}^0 - \beta_{g,B}) \\ &\quad + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0)' \Delta x_{it} \Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}^0 - \beta_{g,B} + \beta_{g,B}^0). \end{aligned}$$

By the Cauchy-Schwarz inequality, Assumption 1(*ii*) and the definition of \mathcal{N} imply that

$$\left| \sum_{t=1}^{k^0-1} 2\Delta u_{it} \Delta x_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0 - \beta_{g_i^0(B),B}^0 + \beta_{g,B}^0) \right|$$

$$\begin{aligned}
& + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B} - \beta_{g,B} - \beta_{g_i^0(B),B}^0 + \beta_{g,B}^0)' \Delta x_{it} \Delta x'_{it} (2\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B} - \beta_{g,B}) \\
& + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0)' \Delta x_{it} \Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B} - \beta_{g,B} + \beta_{g,B}^0) \\
& \leq \eta C_1 \left\| \sum_{t=1}^{k^0-1} \Delta u_{it} \Delta x_{it} \right\| + \eta C_2 \left\| \sum_{t=1}^{k^0-1} \Delta x_{it} \Delta x'_{it} \right\|,
\end{aligned}$$

where C_1 and C_2 are constants independent of η and T .

We then have

$$\begin{aligned}
& \mathbf{1} \left(\sum_{t=1}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_i(B),B})^2 \right. \\
& + (\Delta y_{ik} - x'_{ik} \beta_{g_i(A),A} + x'_{i,k-1} \beta_{g_i(B),B})^2 \\
& < \sum_{t=1}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^0(B),B}^0)^2 + (\Delta y_{ik} - x'_{ik} \beta_{g_i(A),A}^0 + x'_{i,k-1} \beta_{g_i(B)^0,B})^2 \\
& \leq \mathbf{1} \left(\sum_{t=1}^{k^0-1} 2 \Delta u_{it} \Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \right. \\
& \quad \left. < - \sum_{t=1}^{k^0-1} (\Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 + \eta C_1 \left\| \sum_{t=1}^{k^0-1} \Delta u_{it} \Delta x_{it} \right\| + \eta C_2 \left\| \sum_{t=1}^{k^0-1} \Delta x_{it} \Delta x'_{it} \right\| + |d| \right).
\end{aligned}$$

Note that the right hand side does not depend on β . Thus, we have

$$\begin{aligned}
& \Pr \left(\max_{(g_A, g_B) \in \mathbb{G}^B \times \mathbb{G}^A \setminus \{g_i^0(B), g_i^0(A)\}} \mathbf{1} \left(\sum_{t=1}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_i(B),B})^2 \right. \right. \\
& + (\Delta y_{ik} - x'_{ik} \beta_{g_i(A),A} + x'_{i,k-1} \beta_{g_i(B),B})^2 \\
& < \sum_{t=1}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^0(B),B}^0)^2 + (\Delta y_{ik} - x'_{ik} \beta_{g_i(A),A}^0 + x'_{i,k-1} \beta_{g_i(B)^0,B})^2 \left. \right) \neq 0 \Big) \\
& \leq \sum_{g \in \mathbb{G}^B \setminus \{g_i^0(B)\}} \Pr \left(\sum_{t=1}^{k^0-1} 2 \Delta u_{it} \Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \right. \\
& \quad \left. < - \sum_{t=1}^{k^0-1} (\Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 + \eta C_1 \left\| \sum_{t=1}^{k^0-1} \Delta u_{it} \Delta x_{it} \right\| + \eta C_2 \left\| \sum_{t=1}^{k^0-1} \Delta x_{it} \Delta x'_{it} \right\| + |d| \right) \\
& \leq \sum_{g \in \mathbb{G}^B \setminus \{g_i^0(B)\}} \Pr \left(\sum_{t=1}^{k^0-1} 2 \Delta u_{it} \Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \right. \\
& \quad \left. < - \sum_{t=1}^{k^0-1} (\Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 + \eta C_1 \left\| \sum_{t=1}^{k^0-1} \Delta u_{it} \Delta x_{it} \right\| + \eta C_2 \left\| \sum_{t=1}^{k^0-1} \Delta x_{it} \Delta x'_{it} \right\| + |d| \right)
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{g \in \mathbb{G}^B \setminus \{g_i^0(B)\}} \left(\Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} (\Delta x'_{it}(\beta_{g_i^0(B), B}^0 - \beta_{g, B}^0))^2 \right) \leq \frac{c''}{2} \right) \\
&\quad + \Pr \left(\left\| \frac{1}{k^0} \sum_{t=1}^{k^0-1} \Delta u_{it} \Delta x_{it} \right\| \geq M \right) + \Pr \left(\left\| \frac{1}{k^0} \sum_{t=1}^{k^0-1} \Delta x_{it} \Delta x'_{it} \right\| \geq M \right) \\
&\quad + \Pr \left(\frac{1}{k^0} |d| > \frac{k^0 - k}{k^0} C_T \right) \\
&\quad + \Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} 2 \Delta u_{it} \Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \beta_{g, B}^0) < -\frac{c''}{2} + \eta C_1 M + \eta C_2 M + \frac{k^0 - k}{k^0} C_T \right),
\end{aligned}$$

where we take $c'' = c \times \rho_D^*$ for c in Assumption 1(*viii*).

We use the following lemma by Bonhomme and Manresa (2015) which is based on Rio (2017).

Lemma S.5 (Lemma B.5 in Bonhomme and Manresa (2015)). *Let z_t be a strongly mixing process with zero mean, with strong mixing coefficients $a[t] \leq e^{-at^{d_1}}$ and with tail probabilities $\Pr(|z_t| > z) \leq e^{1-(z/b)^{d_2}}$, where a , b , d_1 , and d_2 are positive constants. Then for all $z > 0$, we have for all $\delta > 0$, as $T \rightarrow \infty$,*

$$T^\delta \Pr \left(\left| \frac{1}{T} \sum_{t=1}^T z_t \right| \geq z \right) \rightarrow 0.$$

Note that this lemma holds uniformly over i as long as the bounds for mixing coefficients and tail probabilities hold uniformly over i .

We observe

$$\Pr \left(\left\| \frac{1}{k^0} \sum_{t=1}^{k^0-1} \Delta x_{it} \Delta x'_{it} \right\| \geq M \right) \leq \Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} \|\Delta x_{it} \Delta x'_{it}\| \geq M \right) = \Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} \Delta x'_{it} \Delta x_{it} \geq M \right).$$

We then apply Lemma S.5 with $\Delta x'_{it} \Delta x_{it} - E(\Delta x'_{it} \Delta x_{it})$ as z_t in the lemma. Combining with Assumption 2(*ix*) and the fact that $\sum_{t=1}^{k^0-1} E(\Delta x'_{it} \Delta x_{it})/k^0$ converges, we have

$$\Pr \left(\left\| \frac{1}{k^0} \sum_{t=1}^{k^0-1} \Delta x_{it} \Delta x'_{it} \right\| \geq M \right) = o((k^0)^{-\delta}) = o(T^{-\delta}),$$

where the last equality holds by Assumption 1(*vii*). Similarly, noting that $\sum_{t=1}^{k^0-1} E \|\Delta u_{it} \Delta x_{it}\| / k^0$ converges, Assumption 2(*ix*) implies

$$\Pr \left(\left\| \frac{1}{k^0} \sum_{t=1}^{k^0-1} \Delta u_{it} \Delta x_{it} \right\| \geq M \right) = o(T^{-\delta}).$$

Moreover, a similarly argument shows that under Assumption 2(ix), Lemma S.5 implies

$$\Pr \left(\left| \frac{1}{k^0} \sum_{t=1}^{k^0-1} (\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 - \frac{1}{k^0} \sum_{t=1}^{k^0-1} E((\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2) \right| \geq \frac{c''}{2} \right) = o(T^{-\delta}),$$

which in turn implies that under Assumptions 1(viii) and 2(v),

$$\Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} (\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 \leq \frac{c''}{2} \right) = o(T^{-\delta})$$

uniformly over g . We have shown that $\Pr((k^0)^{-1}|d| > ((k^0 - \underline{k})/k^0)C_T) = o(T^{-\delta})$. Note that $((k^0 - \underline{k})/k^0)C_T \rightarrow 0$ because $M_T = o(\sqrt{T}/\log T)$, $k^0 = O(T)$ and $k^0 - \underline{k} = O(\sqrt{T}\log T)$. Moreover, we can take η small enough and also T large enough such that

$$\begin{aligned} & \Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} 2\Delta u_{it} \Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) < -\frac{c''}{2} + \eta C_1 M + \eta C_2 M + \frac{k^0 - \underline{k}}{k^0} C_T \right) \\ & \leq \Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} 2\Delta u_{it} \Delta x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) < -\frac{c''}{4} \right), \end{aligned}$$

This probability is also $o(T^{-\delta})$ uniformly over g under Assumption 2(ix) by Lemma S.5 following a similar argument to those discussed above.

It thus follows that

$$\begin{aligned} & \Pr \left(\max_{(g_A, g_B) \in \mathbb{G}^B \times \mathbb{G}^A \setminus \{g_i^0(B), g_i^0(A)\}} \mathbf{1} \left(\sum_{t=1}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_i(B),B})^2 \right. \right. \\ & + (\Delta y_{ik} - x'_{ik} \beta_{g_i(A),A} + x'_{i,k-1} \beta_{g_i(B),B})^2 \\ & \left. \left. < \sum_{t=1}^{k-1} (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^0(B),B})^2 + (\Delta y_{ik} - x'_{ik} \beta_{g_i(A),A}^0 + x'_{i,k-1} \beta_{g_i(B),B}^0)^2 \right) \neq 0 \right) = o(NT^{-\delta}). \end{aligned}$$

A similarly argument shows that

$$\begin{aligned} & \Pr \left(\max_{(g_A) \in \mathbb{G}^A \setminus \{g_i^0(A)\}} \mathbf{1} \left(\sum_{t=k+1}^T (\Delta y_{it} - \Delta x'_{it} \beta_{g_i(A),A})^2 < \sum_{t=k+1}^T (\Delta y_{it} - \Delta x'_{it} \beta_{g_i^0(A),A})^2 \right) \neq 0 \right) \\ & = o(NT^{-\delta}). \end{aligned}$$

We use a similar argument for $k \leq k^0$ and the proof is complete. □

S.2.4.2 Proof of Theorem S.1

Proof. We observe that

$$\Pr(\hat{k} \neq k^0) \leq \Pr(\hat{k} \neq k^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{\beta} \notin \mathcal{N})$$

$$\leq \Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{\gamma} \neq \gamma^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{\beta} \notin \mathcal{N}).$$

First, Lemma S.8 and the discussion below it imply that $\Pr(\hat{\beta} \notin \mathcal{N}) \rightarrow 0$. Second we have

$$\begin{aligned} \Pr(\hat{\gamma} \neq \gamma^0, \hat{\beta} \in \mathcal{N}) &\leq \Pr(\hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N}, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{k} \notin K) \\ &\leq \Pr(\hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N}) + \Pr(\hat{k} \notin K) \rightarrow 0, \end{aligned}$$

by Lemmas S.7, S.8 and S.9.

We now consider the third term. We observe that

$$\begin{aligned} &\Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) \\ &\leq \Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}, \hat{k} \in K) + \Pr(\hat{k} \notin K) \\ &\leq \Pr(\hat{k}(\gamma^0, \beta) \neq k^0 \text{ for some } \beta \in \mathcal{N}, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{k} \notin K) \\ &\leq \Pr(\hat{k}(\gamma^0, \beta) \neq k^0 \text{ for some } \beta \in \mathcal{N}) + \Pr(\hat{k} \notin K), \end{aligned}$$

where

$$\hat{k}(\gamma^0, \beta) = \operatorname{argmin}_{k \in K} Q(k, \gamma, \beta).$$

Note that $\Pr(\hat{k} \notin K) \rightarrow 0$ by Lemma S.7. Note that $\hat{k}(\gamma^0, \beta) \neq k^0$ is equivalent to

$$Q(k^0, \gamma^0, \beta) > \min_{k \in K \setminus \{k^0\}} Q(k, \gamma^0, \beta) = \min \left(\min_{k > k^0} Q(k, \gamma^0, \beta), \min_{k < k^0} Q(k, \gamma^0, \beta) \right).$$

Thus, we have

$$\begin{aligned} &\Pr(\hat{k}(\gamma^0, \beta) \neq k^0 \text{ for some } \beta \in \mathcal{N}) \\ &\leq \Pr \left(Q(k^0, \gamma^0, \beta) > \min_{k^0 < k \leq \bar{k}} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right) \\ &\quad + \Pr \left(Q(k^0, \gamma^0, \beta) > \min_{\underline{k} \leq k < k^0} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right). \end{aligned}$$

Suppose for the moment that $\underline{k} \leq k < k^0$. Noting that

$$\begin{aligned} Q(k, \gamma^0, \beta) &= \frac{1}{NT} \left(\sum_{t=1}^{k-1} \sum_{i=1}^N (\Delta x'_{it} \beta_{g_i^0(B), B}^0 - \Delta x'_{it} \beta_{g_i^0(B), B}^0 + \Delta u_{it})^2 \right. \\ &\quad + \sum_{i=1}^N (\Delta x'_{ik} \beta_{g_i^0(B), B}^0 - x'_{ik} \beta_{g_i^0(A), A}^0 + x'_{i,k-1} \beta_{g_i^0(B), B}^0 + \Delta u_{ik})^2 \\ &\quad + \sum_{t=k+1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{ik} \beta_{g_i^0(B), B}^0 - \Delta x'_{it} \beta_{g_i^0(A), A}^0 + \Delta u_{it})^2 \\ &\quad \left. + \sum_{i=1}^N (x'_{ik^0} \beta_{g_i^0(A), A}^0 - x'_{i,k^0-1} \beta_{g_i^0(B), B}^0 - \Delta x'_{ik^0} \beta_{g_i^0(A), A}^0 + \Delta u_{ik^0})^2 \right) \end{aligned}$$

$$+ \sum_{t=k^0+1}^T \sum_{i=1}^N (\Delta x'_{it} \beta_{g_i^0(A), A}^0 - \Delta x'_{it} \beta_{g_i^0(B), B}^0 + \Delta u_{it})^2 \Big),$$

and

$$\begin{aligned} Q(k^0, \gamma^0, \beta) = & \frac{1}{NT} \left(\sum_{t=1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it} \beta_{g_i^0(B), B}^0 - \Delta x'_{it} \beta_{g_i^0(A), B}^0 + \Delta u_{it})^2 \right. \\ & + \sum_{i=1}^N (x'_{ik^0} \beta_{g_i^0(A), A}^0 - x'_{i,k^0-1} \beta_{g_i^0(B), B}^0 - x'_{ik^0} \beta_{g_i^0(A), B}^0 + x'_{i,k^0-1} \beta_{g_i^0(B), A}^0 + \Delta u_{ik^0})^2 \\ & \left. + \sum_{t=k^0+1}^T \sum_{i=1}^N (\Delta x'_{it} \beta_{g_i^0(A), A}^0 - \Delta x'_{it} \beta_{g_i^0(B), A}^0 + \Delta u_{it})^2 \right), \end{aligned}$$

we have

$$\begin{aligned} & Q(k^0, \gamma^0, \beta) - Q(k, \gamma^0, \beta) \\ = & \sum_{t=k}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it} \beta_{g_i^0(B), B}^0 - \Delta x'_{it} \beta_{g_i^0(A), B}^0 + \Delta u_{it})^2 \\ & + \sum_{i=1}^N (x'_{ik^0} \beta_{g_i^0(A), A}^0 - x'_{i,k^0-1} \beta_{g_i^0(B), B}^0 - x'_{ik^0} \beta_{g_i^0(A), B}^0 + x'_{i,k^0-1} \beta_{g_i^0(B), A}^0 + \Delta u_{ik^0})^2 \\ & - \sum_{i=1}^N (\Delta x'_{ik} \beta_{g_i^0(B), B}^0 - x'_{ik} \beta_{g_i^0(A), A}^0 - x'_{i,k-1} \beta_{g_i^0(B), B}^0 + \Delta u_{ik})^2 \\ & - \sum_{t=k+1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it} \beta_{g_i^0(B), B}^0 - \Delta x'_{it} \beta_{g_i^0(A), A}^0 + \Delta u_{it})^2 \\ & - \sum_{i=1}^N (x'_{ik^0} \beta_{g_i^0(A), A}^0 - x'_{i,k^0-1} \beta_{g_i^0(B), B}^0 - \Delta x'_{ik^0} \beta_{g_i^0(A), A}^0 + \Delta u_{ik^0})^2 \\ = & \sum_{t=k}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \beta_{g_i^0(A), B}^0) + \Delta u_{it})^2 \\ & + \sum_{i=1}^N (x'_{ik^0} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), B}^0) - x'_{i,k^0-1} (\beta_{g_i^0(B), B}^0 - \beta_{g_i^0(B), A}^0) + \Delta u_{ik^0})^2 \\ & - \sum_{i=1}^N (x'_{ik} (\beta_{g_i^0(B), B}^0 - \beta_{g_i^0(A), A}^0) - x'_{i,k-1} (\beta_{g_i^0(B), B}^0 - \beta_{g_i^0(B), A}^0 + \Delta u_{ik})^2 \\ & - \sum_{t=k+1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it} (\beta_{g_i^0(B), B}^0 - \beta_{g_i^0(A), A}^0) + \Delta u_{it})^2 \\ & - \sum_{i=1}^N (x'_{ik^0} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), B}^0) - x'_{i,k^0-1} (\beta_{g_i^0(B), B}^0 - \beta_{g_i^0(A), B}^0) + \Delta u_{ik^0})^2 \end{aligned}$$

$$\begin{aligned}
&= -2 \sum_{t=k}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B})) \Delta u_{it} + \sum_{t=k}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}))^2 \\
&\quad - 2 \sum_{i=1}^N (x'_{ik^0}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}) - x'_{i,k^0-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B})) \Delta u_{ik^0} \\
&\quad + \sum_{i=1}^N (x'_{ik^0}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}) - x'_{i,k^0-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}))^2 \\
&\quad + 2 \sum_{i=1}^N (x'_{ik}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}) - x'_{i,k-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B})) \Delta u_{ik} \\
&\quad - \sum_{i=1}^N (x'_{ik}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}) - x'_{i,k-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}))^2 \\
&\quad + 2 \sum_{t=k+1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{ik}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A})) \Delta u_{it} - \sum_{t=k+1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{ik}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}))^2 \\
&\quad + 2 \sum_{i=1}^N (x'_{ik^0}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}) - x'_{i,k^0-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A})) \Delta u_{ik^0} \\
&\quad - \sum_{i=1}^N (x'_{ik^0}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}) - x'_{i,k^0-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}))^2 \\
&= -2 \sum_{i=1}^N (x'_{ik}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B})) \Delta u_{ik} \\
&\quad - 2 \sum_{t=k+1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B})) \Delta u_{it} \\
&\quad + 2 \sum_{i=1}^N (x'_{i,k^0-1}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B})) \Delta u_{ik^0} \\
&\quad + \sum_{t=k}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}))^2 \\
&\quad + \sum_{i=1}^N (x'_{ik^0}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}) - x'_{i,k^0-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}))^2 \\
&\quad - \sum_{i=1}^N (x'_{ik}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}) - x'_{i,k-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}))^2 \\
&\quad - \sum_{t=k+1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{ik}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}))^2 \\
&\quad + \sum_{i=1}^N (x'_{ik^0}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}) - x'_{i,k^0-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}))^2.
\end{aligned}$$

Let

$$z_{it} = \begin{cases} x_{ik} & \text{if } t = k \\ \Delta x_{it} & \text{if } k < t < k^0 \\ -x_{i,k^0-1} & \text{if } t = k^0. \end{cases}$$

Then, Assumption 2(vi) implies that

$$\sum_{t=k}^{k^0} \sum_{i=1}^N (z_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}^0))^2 > N(k^0 - k + 1)m.$$

Moreover, Assumptions 2(v) and the condition that $\beta \in \mathcal{N}_\eta$ implies that by taking η small enough, we have

$$\begin{aligned} & \sum_{t=k}^{k^0} \sum_{i=1}^N (z_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}^0))^2 \\ & - \left| \sum_{t=k}^{k^0-1} \sum_{i=1}^N (\Delta x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}^0))^2 \right. \\ & + \sum_{i=1}^N (x'_{ik^0}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}^0) - x'_{i,k^0-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}^0))^2 \\ & - \sum_{i=1}^N (x'_{ik}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}^0) - x'_{i,k-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}^0))^2 \\ & - \sum_{t=k+1}^{k^0-1} \sum_{i=1}^N (\Delta x'_{ik}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}^0))^2 \\ & \left. + \sum_{i=1}^N (x'_{ik^0}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}^0) - x'_{i,k^0-1}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(A),A}^0))^2 \right| \\ & \geq \frac{(k^0 - k + 1)}{2} m. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} & \left| \frac{1}{k^0 - k + 1} \frac{1}{N} \sum_{t=k}^{k^0} \sum_{i=1}^N (z_{it}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0 + \beta_{g_i^0(B),B}^0)) \Delta u_{it} \right| \\ & \leq \eta C \frac{1}{k^0 - k + 1} \frac{1}{N} \sum_{t=k}^{k^0} \sum_{i=1}^N \|z_{it} \Delta u_{it}\| \end{aligned}$$

Thus we have

$$\Pr \left(Q(k^0, \gamma^0, \beta) > \min_{k^0 < k \leq \bar{k}} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right)$$

$$\begin{aligned}
&= \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{k^0 < k \leq \bar{k}} (Q(k^0, \gamma^0, \beta) - Q(k, \gamma^0, \beta)) > 0 \right) \\
&\leq \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{\underline{k} \leq k < k^0} \left(-2 \frac{1}{N} \sum_{t=k^0}^k \sum_{i=1}^N z'_{it} (\beta_{g_i^0(A), A} - \beta_{g_i^0(B), B}) \Delta u_{it} - \frac{k^0 - k + 1}{2T} m \right) > 0 \right) \\
&= \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{\underline{k} \leq k < k^0} \left(-2 \frac{1}{N} \frac{1}{k^0 - k + 1} \sum_{t=\underline{k}}^{k^0} \sum_{i=1}^N z'_{it} (\beta_{g_i^0(A), A} - \beta_{g_i^0(B), B}) \Delta u_{it} - \frac{m}{2} \right) > 0 \right) \\
&\leq \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{\underline{k} \leq k < k^0} \left(-2 \frac{1}{N} \frac{1}{k^0 - k + 1} \sum_{t=\underline{k}}^{k^0} \sum_{i=1}^N z'_{it} (\beta_{g_i^0(A), A} - \beta_{g_i^0(B), B}) \Delta u_{it} \right) > \frac{m}{2} \right) \\
&\leq \Pr \left(\max_{\underline{k} \leq k < k^0} \left(-2 \frac{1}{N} \frac{1}{k^0 - k + 1} \sum_{t=\underline{k}}^{k^0} \sum_{i=1}^N z'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) \Delta u_{it} \right) > \frac{m}{4} \right) \\
&\quad + \Pr \left(\eta C \max_{\underline{k} \leq k < k^0} \frac{1}{N} \frac{1}{k^0 - k + 1} \sum_{t=\underline{k}}^{k^0} \sum_{i=1}^N \|z_{it} u_{it}\| > \frac{m}{4} \right) \\
&= O \left(\frac{1}{N} \right),
\end{aligned}$$

where the last equality follows by applying (Bai and Perron, 1998, Lemma A.6) which is an extension of Hájek and Rényi (1955). Here we use the observation that an L_r mixing sequence is an L_p mixingale sequence for $1 \leq p < r$ as discussed in (Davidson, 1994, page 248). Thus, under Assumptions 2(ix) and 2(x), $z_{it} u_{it}$ is an L^2 mixingale and we can apply (Bai and Perron, 1998, Lemma A.6).

A similar argument shows that

$$\Pr \left(Q(k^0, \gamma^0, \beta) > \min_{k^0 < k < \bar{k}} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right) = O \left(\frac{1}{N} \right).$$

To sum up, we have

$$\Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) \rightarrow 0.$$

□

S.2.4.3 Proof of Corollary S.1

Proof. We first show the first part of the corollary. We observe

$$\Pr(\hat{\gamma} \neq \gamma^0) \leq \Pr(\hat{\gamma} \neq \gamma^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{\beta} \notin \mathcal{N}).$$

The second paragraph of the proof of Theorem S.1 shows $\Pr(\hat{\gamma} \neq \gamma^0, \hat{\beta} \in \mathcal{N}) \rightarrow 0$. Lemma S.8 and the discussion below imply $\Pr(\hat{\beta} \notin \mathcal{N}) \rightarrow 0$. The desired result thus holds.

Second, we show the second part of the corollary. We have

$$\begin{aligned} & \Pr\left(\|\hat{\beta} - \tilde{\beta}\| > a/\sqrt{NT}\right) \\ & \leq \Pr\left(\|\hat{\beta} - \tilde{\beta}\| > a\sqrt{NT}, \hat{\gamma} = \gamma^0, \hat{k} = k^0\right) + \Pr(\gamma \neq \gamma^0) + \Pr(\hat{k} \neq k^0) \\ & \leq 0 + \Pr(\gamma \neq \gamma^0) + \Pr(\hat{k} \neq k^0) \rightarrow 0 \end{aligned}$$

for any $a > 0$, where the second inequality follows because $\hat{\beta} = \tilde{\beta}$ holds under $\gamma = \gamma^0$ and $\hat{k} = k^0$, and the third inequality holds by the first part of this corollary and Theorem S.1. We thus have the desired result. \square

S.3 Models with fixed effects: within transformation approach

This section provides theoretical results and proofs for within-transformation estimation of models with fixed effects, an alternative method to first-difference estimation.

S.3.1 Estimation method

Here we divide the sample into two regimes according to a hypothetical break point, and eliminate the fixed effects by transforming the dependent and explanatory variables, subtracting their average over time for each of the two regimes. Then we estimate the coefficients, group membership structure, and the break date by minimizing the quadratic loss function defined in terms of the transformed variables.

We define the variables after within transformation for each regime, given a hypothetical break point, as

$$y_{it,k,B} = y_{it} - \frac{1}{k-1} \sum_{s=1}^{k-1} y_{is}, \quad y_{it,k,A} = y_{it} - \frac{1}{T-k+1} \sum_{s=k}^T y_{is},$$

and

$$x_{it,k,B} = x_{it} - \frac{1}{k-1} \sum_{s=1}^{k-1} x_{is}, \quad x_{it,k,A} = x_{it} - \frac{1}{T-k+1} \sum_{s=k}^T x_{is}.$$

We estimate (k, γ, β) by minimizing the least squares criterion:

$$\begin{aligned} (\hat{k}, \hat{\gamma}, \hat{\beta}) = \operatorname{argmin}_{k \in \mathbb{K}, \gamma \in \Gamma, \beta \in \mathcal{B}} & \left[\sum_{t=1}^{k-1} \sum_{i=1}^N (y_{it,k,B} - x'_{it,k,B} \beta_{g_i(B),B})^2 \right. \\ & \left. + \sum_{t=k}^T \sum_{i=1}^N (y_{it,k,A} - x'_{it,k,A} \beta_{g_i(A),A})^2 \right]. \end{aligned}$$

To solve this objective function, we propose the following algorithm:

Algorithm

Let s denote the iteration number.

Step 1: Set $s = 1$. For a given $k \in \{2, \dots, T - 1\}$, initialize group structures in both regimes as $\gamma_B^{(0)}$ and $\gamma_A^{(0)}$.

Step 2: For given $\gamma^{(s)}$ and k , estimate the slope coefficient $\beta^{(s)}$ in the two regimes by

$$\begin{aligned}\beta_{g,B}^{(s)} &= \underset{\beta_{g,B} \in \mathcal{B}}{\operatorname{argmin}} \sum_{t=1}^{k-1} \sum_{g_i^{(s)}(B)=g} (y_{it,k,B} - x'_{it,k,B} \beta_{g,B})^2, \text{ for } g \in \mathbb{G}^B \\ \beta_{g,A}^{(s)} &= \underset{\beta_{g,A} \in \mathcal{B}}{\operatorname{argmin}} \sum_{t=k}^T \sum_{g_i^{(s)}(A)=g} (y_{it,k,A} - x'_{it,k,A} \beta_{g,A})^2, \text{ for } g \in \mathbb{G}^A.\end{aligned}$$

Step 3: Given $\beta^{(s)}$, find the optimal group for individual i in each regime, respectively, by

$$\begin{aligned}g_i^{(s+1)}(B) &= \underset{g_i(B) \in \{1, \dots, G^B\}}{\operatorname{argmin}} \sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B} \beta_{g_i(B),B}^{(s)})^2, \\ g_i^{(s+1)}(A) &= \underset{g_i(A) \in \{1, \dots, G^A\}}{\operatorname{argmin}} \sum_{t=k}^T (y_{it,k,A} - x'_{it,k,A} \beta_{g_i(A),A}^{(s)})^2.\end{aligned}$$

Step 4: Iterate Steps 2 and 3 until numerical convergence, and obtain $\hat{\gamma}_B(k)$, $\hat{\gamma}_A(k)$, and $\hat{\beta}(k)$.

Step 5: Let k vary from 2 to T , and estimate the break point by

$$\begin{aligned}\hat{k} &= \underset{k \in \mathbb{K}}{\operatorname{argmin}} \left[\sum_{t=1}^{k-1} \sum_{i=1}^N (y_{it,k,B} - x'_{it,k,B} \hat{\beta}_{g_i(B,k),B}(k))^2 \right. \\ &\quad \left. + \sum_{t=k}^T \sum_{i=1}^N (y_{it,k,A} - x'_{it,k,A} \beta_{g_i(A,k),A}(k))^2 \right].\end{aligned}$$

Step 1 is an initialization. In Step 2, we estimate the coefficient vector given group membership structure and break date. Step 3 in turn estimates the group membership structure given the coefficients and break date. The break point is estimated in Step 5 by minimizing the sum of squared residuals.

S.3.2 Assumptions

We list the assumptions pertaining to the fixed effects specification. The first set of assumptions is a subset of the assumptions used for models without fixed effects and the same as that presented in Section S.2.

The following set of the assumptions are specific to the within-transformation estimator. However, they closely correspond to the assumptions in Assumption 1.

Assumption 3.

(i) For any $L \subseteq \{1, \dots, N\}$ and $t'' \geq t'$, there exists M which does not depend on L , t'' nor t' such that the following equality holds

$$E \left(\left\| \frac{1}{NT} \sum_{t=t'}^{t''} \sum_{i \in L} x_{it} u_{it} \right\|^2 \right) = M \frac{|L|(t'' - t')}{NT^2},$$

and

$$E \left(\left\| \frac{1}{NT} \sum_{t=t'}^{t''} \sum_{i \in L} u_{it} \right\|^2 \right) = M \frac{|L|(t'' - t')}{NT^2},$$

for where $|L|$ is the cardinality of L .

(iii) Let $\rho_{D,N,t}(\gamma^t, g, \tilde{g})$ be the minimum eigenvalue of $\sum_{i=1}^N \mathbf{1}\{g_{it}^0 = g\}\{g_{it} = \tilde{g}\}x_{it,k}x'_{it,k}/N$, where $x_{it,k} = x_{it,k,B}$ if $t < k$, $x_{it,k} = x_{it,k,A}$ if $t \geq k$, and γ^t is either γ_B (when $t < k$) or γ_A (when $t \geq k$). For any $g \in \mathbb{G}^B$,

$$\min_{1 \leq t < k^0} \min_{\gamma_B} \max_{\tilde{g} \in \mathbb{G}^B} \rho_{D,N,t}(\gamma_B, g, \tilde{g}) > \hat{\rho}_D,$$

and for any $g \in \mathbb{G}^A$,

$$\min_{k^0 < t \leq T} \min_{\gamma_A} \max_{\tilde{g} \in \mathbb{G}^A} \rho_{D,N,t}(\gamma_A, g, \tilde{g}) > \hat{\rho}_D,$$

where $\hat{\rho}_D \rightarrow_p \rho_D > 0$.

(iv) There exists $\hat{\rho}_D^*$ such that for any i and for s such that s and $T - s$ sufficiently large,

$$\lambda_{\min} \left(\frac{1}{s} \sum_{t=1}^s x_{it,k,B} x'_{it,k,B} \right) \geq \hat{\rho}_D^* \quad \text{and} \quad \lambda_{\min} \left(\frac{1}{T-s} \sum_{t=s+1}^T x_{it,k,A} x'_{it,k,A} \right) \geq \hat{\rho}_D^*,$$

and $\hat{\rho}_D^* \rightarrow_p \rho_D^* > 0$.

(v) $\max_{1 \leq t \leq T} \sum_{i=1}^N \|x_{it}\|^2/N = O_p(1)$, $\max_{1 \leq t \leq k-1} \sum_{i=1}^N \|x_{it,k,B}\|^2/N = O_p(1)$ and $\max_{k \leq t \leq T} \sum_{i=1}^N \|x_{it,k,A}\|^2/N = O_p(1)$.

(vi) There exists a fixed constant $\underline{m} > 0$ independent on T and N , such that for any t ,

$$\frac{1}{N} \sum_{i=1}^N (x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0))^2 > \underline{m}.$$

and

$$\frac{1}{N} \sum_{i=1}^N (x'_{it,k,A}(\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0))^2 > \underline{m}.$$

(ix) Let z_{it} be $\|x_{it}\|$, u_{it} , $x'_{it}x_{it}$, $\|u_{it}x_{it}\|$, $2u_{it}x'_{it}(\beta_{g_i^0(l),l}^0 - \beta_{g,l}^0)$ or $(x'_{it}(\beta_{g_i^0(l),l}^0 - \beta_{g,l}^0))^2$, for $g \in \mathbb{G}_l$ and $l = A, B$. Assume the following holds for any choice of z_{it} . 1) z_{it} is a strong mixing sequence over t whose mixing coefficients $a_i[t]$ are bounded by $a[t] \leq e^{-at^{d_1}}$ such that $\max_{1 \leq i \leq N} a_i[t] \leq a[t]$ and has tail probabilities $\max_{1 \leq i \leq N} \Pr(|z_{it}| > z) \leq e^{1-(z/b)^{d_2}}$ for any t where a , b , d_1 and d_2 are positive constants. 2) There exists a_i , $i = 1, \dots, N$ such that for any $\epsilon > 0$, it holds that $\max_{1 \leq i \leq N} |a_i - \sum_{t=1}^T E(z_{it})/T| < \epsilon$ for T sufficiently large.

(x) $\max_{1 \leq t \leq T} E(\|\sum_{i=1}^N x_{it}u_{it}/\sqrt{N}\|^{2+\delta})$ is bounded for some $\delta > 0$.

S.3.3 Asymptotic results

We have the following asymptotic results for the within-transformation estimator. They are similar to Theorem 1 and Corollary 1 in the main text.

Theorem S.2. Suppose that Assumptions 1(ii), 1(vii), 1(viii) and 3 holds. We also let $\epsilon T < k < (1 - \epsilon T)$ for $\epsilon > 0$ in Assumption 1(vii). As $N, T \rightarrow \infty$ with $NT^{-\delta} \rightarrow 0$ for some $\delta > 0$, $\Pr(\hat{k} = k^0) \rightarrow 1$.

Corollary S.2. Suppose that Assumptions 1(ii), 1(vii), 1(viii) and 3 holds. We also let $\epsilon T < k < (1 - \epsilon T)$ for $\epsilon > 0$ in Assumption 1(vii). As $N, T \rightarrow \infty$ with $NT^{-\delta} \rightarrow 0$ for some $\delta > 0$, $\Pr(\hat{\gamma} = \gamma^0) \rightarrow 1$ and $\hat{\beta} = \tilde{\beta} + o_p(1/\sqrt{NT})$, where $\tilde{\beta}$ is the estimator of β under $k = k^0$ and $\gamma = \gamma^0$.

We note that $\tilde{\beta}$ is biased when x_{it} is merely sequentially exogenous and not strictly exogenous and T is not very large compared with N . A biased correction method is required to make valid statistical inferences. For example, the half-panel jackknife bias correction may be employed for each regime.

S.3.4 Proofs

This section includes the proofs of Theorem S.2 and Corollary S.2. First, Section S.3.4.1 presents the lemmas. The proofs of the theorem and corollary are included in Sections S.3.4.2 and S.3.4.3, respectively. Hereafter, C denotes a generic constant whose exact value depends on the content but it does not depend on N , T , or on the values of the parameters.

S.3.4.1 Lemmas

First we observe the following. Suppose that $k > k^0$. For $t \geq k$, we have

$$\begin{aligned} y_{it,k,A} &= x'_{it} \beta_{g_i^0(A),A}^0 + u_{it} - \frac{1}{T-k+1} \sum_{s=k}^T x'_{is} \beta_{g_i^0(A),A}^0 - \frac{1}{T-k+1} \sum_{s=k}^T u_{is} \\ &= x'_{it,k,A} \beta_{g_i^0(A),A}^0 + u_{it,k,A}. \end{aligned}$$

For $k^0 \leq t < k$, we have

$$\begin{aligned} y_{it,k,B} &= x'_{it} \beta_{g_i^0(A),A}^0 + u_{it} - \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} \beta_{g_i^0(B),B}^0 - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(A),A}^0 - \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \\ &= x'_{it,k,B} \beta_{g_i^0(A),A}^0 + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) + u_{it,k,B}. \end{aligned}$$

For $t < k^0$, we have

$$\begin{aligned} y_{it,k,B} &= x'_{it} \beta_{g_i^0(B),B}^0 + u_{it} - \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} \beta_{g_i^0(B),B}^0 - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(A),A}^0 - \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \\ &= x'_{it,k,B} \beta_{g_i^0(B),B}^0 - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) + u_{it,k,B}. \end{aligned}$$

Suppose now that $k \leq k^0$. For $t \geq k^0$, we have

$$\begin{aligned} y_{it,k,A} &= x'_{it} \beta_{g_i^0(A),A}^0 + u_{it} - \frac{1}{T-k+1} \sum_{s=k^0}^T x'_{is} \beta_{g_i^0(A),A}^0 - \frac{1}{T-k+1} \sum_{s=k}^{k^0-1} x'_{is} \beta_{g_i^0(B),B}^0 - \frac{1}{T-k+1} \sum_{s=k}^T u_{is} \\ &= x'_{it,k,A} \beta_{g_i^0(A),A}^0 + \frac{1}{T-k+1} \sum_{s=k}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) + u_{it,k,A} \end{aligned}$$

For $k \leq t < k^0$, we have

$$\begin{aligned} y_{it,k,A} &= x'_{it} \beta_{g_i^0(B),B}^0 + u_{it} - \frac{1}{T-k+1} \sum_{s=k^0}^T x'_{is} \beta_{g_i^0(A),A}^0 - \frac{1}{T-k+1} \sum_{s=k}^{k^0-1} x'_{is} \beta_{g_i^0(B),B}^0 - \frac{1}{T-k+1} \sum_{s=k}^T u_{is} \\ &= x'_{it,k,A} \beta_{g_i^0(B),B}^0 - \frac{1}{T-k+1} \sum_{s=k}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) + u_{it,k,A} \end{aligned}$$

For $k < t$, we have

$$y_{it,k,B} = x'_{it,k,B} \beta_{g_i^0(B),B}^0 + u_{it,k,B}.$$

Let

$$Q(k, \gamma, \beta) = \frac{1}{NT} \left(\sum_{t=1}^{k-1} \sum_{i=1}^N (y_{it,k,B} - x'_{it,k,B} \beta_{g_i^0(B),B}^0)^2 + \sum_{t=k}^T \sum_{i=1}^N (y_{it,k,A} - x'_{it,k,A} \beta_{g_i^0(A),A}^0)^2 \right),$$

Define $\tilde{Q}(k, \gamma, \beta)$ in the following. For $k < k^0$, it is defined as:

$$\begin{aligned}\tilde{Q}(k, \gamma, \beta) = & \frac{1}{NT} \sum_{t=1}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right)^2 \\ & + \frac{1}{NT} \sum_{t=k}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,A} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(A),A}) - \frac{1}{T-k+1} \sum_{s=k^0}^T x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\ & + \frac{1}{NT} \sum_{t=k^0}^T \sum_{i=1}^N \left(x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A}) + \frac{1}{T-k+1} \sum_{s=k}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\ & + \frac{1}{NT} \sum_{t=1}^{k-1} \sum_{i=1}^N u_{it,k,B}^2 + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A}^2.\end{aligned}$$

When $k = k^0$, it is

$$\begin{aligned}\tilde{Q}(k, \gamma, \beta) = & \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right)^2 \\ & + \frac{1}{NT} \sum_{t=k^0}^T \sum_{i=1}^N (x'_{it,k^0,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A}))^2 \\ & + \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k^0,B}^2 + \frac{1}{NT} \sum_{t=k^0}^T \sum_{i=1}^N u_{it,k^0,A}^2.\end{aligned}$$

Lastly, when $k > k^0$, it is

$$\begin{aligned}\tilde{Q}(k, \gamma, \beta) = & \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\ & + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(B),B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\ & + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A}))^2 + \frac{1}{NT} \sum_{t=1}^{k-1} \sum_{i=1}^N u_{it,k,B}^2 + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A}^2.\end{aligned}$$

Lemma S.6. Suppose that Assumptions 1(ii), 1(vii), 3(i), and 3(v) and hold. We also let $\epsilon T < k < (1 - \epsilon T)$ for $\epsilon > 0$ in Assumption 1(vii). Then we have

$$\sup_{k \in \mathbb{K}, \gamma \in \mathbb{G}, \beta \in \mathbb{B}} |\tilde{Q}(k, \gamma, \beta) - Q(k, \gamma, \beta)| = O_p \left(\frac{1}{\sqrt{T}} \right).$$

Proof. The proof follows the steps taken in the proof of Lemma S.3 of Bonhomme and Manresa (2015). Each step becomes more complicated because of the presence of a break and the within transformation.

First we consider the case in which $k > k^0$. We thus have

$$\begin{aligned}
& Q(k, \gamma, \beta) \\
&= \frac{1}{NT} \left(\sum_{t=1}^{k-1} \sum_{i=1}^N (y_{it,k,B} - x'_{it,k,B} \beta_{g_i(B),B})^2 + \sum_{t=k}^T \sum_{i=1}^N (y_{it,k,A} - x'_{it,k,A} \beta_{g_i(A),A})^2 \right) \\
&= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) + u_{it,k,B} \right)^2 \\
&\quad + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(B),B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) + u_{it,k,B} \right)^2 \\
&\quad + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A}) + u_{it,k,A})^2 \\
&= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\
&\quad + \frac{2}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \\
&\quad + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(B),B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\
&\quad + \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it,k,B} \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(B),B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \\
&\quad + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A})^2 \\
&\quad + \frac{2}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A} (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A})) \\
&\quad + \frac{1}{NT} \sum_{t=1}^{k-1} \sum_{i=1}^N u_{it,k,B}^2 + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A}^2 \\
&= \tilde{Q}(k, \gamma, \beta) \\
&\quad + \frac{2}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \\
&\quad + \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it,k,B} \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(B),B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \\
&\quad + \frac{2}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A} (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A})). \tag{S.5}
\end{aligned}$$

We observe that

$$\begin{aligned}
\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} x'_{it,k,B} \beta_{g_i^0(B),B}^0 &= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i^0(B),B}^0 \\
&\quad - \frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \sum_{t=1}^{k^0-1} x'_{it} \beta_{g_i^0(B),B}^0 \\
&\quad - \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{k^0-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(B),B}^0 \\
&\quad + \frac{k-k^0}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(B),B}^0.
\end{aligned}$$

The first term is

$$\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i^0(B),B}^0 = \frac{1}{NT} \sum_{g \in \mathbb{G}^B} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \mathbf{1}(g_i(B)^0 = g) u_{it} x'_{it} \beta_{g_i^0(B),B}^0.$$

For each $g \in \mathbb{G}^B$, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&E \left(\left(\frac{1}{NT} \sum_{g \in \mathbb{G}^B} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \mathbf{1}(g_i(B)^0 = g) u_{it} x'_{it} \beta_{g_i^0(B),B}^0 \right)^2 \right) \\
&\leq BE \left(\left\| \frac{1}{NT} \sum_{g \in \mathbb{G}^B} \sum_{t=1}^{k^0-1} \sum_{g_i^0(B)=g} u_{it} x_{it} \right\|^2 \right) \\
&= O \left(\frac{k^0}{NT^2} \right),
\end{aligned}$$

where B satisfies $\|\beta_g^0\| < B$ for any $g \in \mathbb{G}^B$. Assumption 1(ii) implies the existence of such B . Thus, by the Markov inequality, we have

$$\frac{1}{NT} \sum_{g \in \mathbb{G}^B} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \mathbf{1}(g_i(B)^0 = g) u_{it} x'_{it} \beta_{g_i^0(B),B}^0 = O_p \left(\frac{\sqrt{k^0}}{\sqrt{NT}} \right).$$

We also observe that by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\left(\frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \sum_{t=1}^{k^0-1} x'_{it} \beta_{g_i^0(B),B}^0 \right)^2 \\
&\leq \left(\frac{1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \right)^2 \right) \left(\frac{1}{NT} \sum_{i=1}^N \left(\sum_{t=1}^{k^0-1} x'_{it} \beta_{g_i^0(B),B}^0 \right)^2 \right).
\end{aligned}$$

By Assumptions 3(i) and 3(v), we have

$$E \left(\frac{1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \right)^2 \right) = O \left(\frac{1}{Tk} \right),$$

and

$$\begin{aligned} E \left(\frac{1}{NT} \sum_{i=1}^N \left(\sum_{t=1}^{k^0-1} x'_{it} \beta_{g_i^0(B), B}^0 \right)^2 \right) &\leq BE \left(\frac{1}{NT} \sum_{i=1}^N \left\| \sum_{t=1}^{k^0-1} x_{it} x'_{it} \right\| \right) \\ &= O \left(\frac{k^0}{T} \right), \end{aligned}$$

where B satisfies $\|\beta_g^0\|^2 < B$ for any $g \in \mathbb{G}^B$. Assumption 1(ii) implies the existence of such B . Thus we have

$$\frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \sum_{t=1}^{k^0-1} x'_{it} \beta_{g_i^0(B), B}^0 = O_p \left(\frac{\sqrt{k^0}}{T\sqrt{k}} \right).$$

For the third term, we again use the Cauchy-Schwarz inequality to obtain

$$\begin{aligned} &\left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{k^0-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 \right)^2 \\ &\leq \left(\frac{1}{NT} \sum_{i=1}^N \left(\sum_{t=1}^{k^0-1} u_{it} \right)^2 \right) \left(\frac{1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 \right)^2 \right). \end{aligned}$$

By Assumptions 3(i) and 3(v), it holds that

$$E \left(\frac{1}{NT} \sum_{i=1}^N \left(\sum_{t=1}^{k^0-1} u_{it} \right)^2 \right) = O \left(\frac{k^0}{T} \right),$$

and

$$\begin{aligned} &E \left(\frac{1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 \right)^2 \right) \\ &\leq BE \left(\frac{1}{NT} \sum_{i=1}^N \left\| \frac{1}{k-1} \sum_{s=1}^{k-1} x_{is} x'_{is} \right\| \right) = O \left(\frac{1}{T} \right), \end{aligned}$$

where B satisfies $\|\beta_g^0\|^2 < B$ for any $g \in \mathbb{G}^B$. Assumption 1(ii) implies the existence of such B . Thus we have

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{k^0-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 = O_p \left(\frac{\sqrt{k^0}}{T} \right).$$

We now consider the fourth term. Another application of the Cauchy-Schwarz inequality gives

$$\left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 \right)^2$$

$$\leq \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \right)^2 \right) \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 \right)^2 \right).$$

By Assumptions 3(i) and 3(v), it holds that

$$E \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \right)^2 \right) = O \left(\frac{k^0}{Tk} \right),$$

and

$$\begin{aligned} & E \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 \right)^2 \right) \\ & \leq BE \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \left\| \frac{1}{k-1} \sum_{s=1}^{k-1} x_{is} x'_{is} \right\|^2 \right) = O \left(\frac{k^0}{T} \right), \end{aligned}$$

where B satisfies $\|\beta_g^0\|^2 < B$ for any $g \in \mathbb{G}^B$. Assumption 1(ii) implies the existence of such B . Thus we have

$$\frac{k^0 - 1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 = O_p \left(\frac{k^0}{T\sqrt{k}} \right).$$

To sum up, we have

$$\begin{aligned} \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} x'_{it,k,B} \beta_{g_i^0(B), B}^0 &= O_p \left(\frac{\sqrt{k^0}}{\sqrt{NT}} + \frac{\sqrt{k^0}}{T\sqrt{k}} + \frac{\sqrt{k^0}}{T} + \frac{k^0}{T\sqrt{k}} \right) \\ &= O_p \left(\frac{\sqrt{k^0}}{T} + \frac{k^0}{T\sqrt{k}} \right). \end{aligned}$$

Next we consider

$$\begin{aligned} \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} x'_{it,k,B} \beta_{g_i(B), B} &= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i(B), B} \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \sum_{t=1}^{k^0-1} x'_{it} \beta_{g_i(B), B} \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{k^0-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(B), B} \\ &\quad + \frac{k-k^0}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(B), B}. \end{aligned}$$

The first term is

$$\left(\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i(B), B} \right)^2 = \left(\frac{1}{NT} \sum_{i=1}^N \beta'_{g_i(B), B} \sum_{t=1}^{k^0-1} x_{it} u'_{it} \right)^2$$

$$\begin{aligned} &\leq \left(\frac{1}{N} \sum_{i=1}^N \|\beta_{g_i(B),B}\|^2 \right) \left(\frac{1}{NT^2} \sum_{i=1}^N \left\| \sum_{t=1}^{k^0-1} x_{it} u'_{it} \right\|^2 \right) \\ &= O_p \left(\frac{k^0}{T^2} \right), \end{aligned}$$

where the first inequality is the Cauchy-Schwarz inequality and the second inequality follows by that Assumption 1(ii) implies $\sum_{i=1}^N \|\beta_{g_i(B),B}\|^2/N < C$ for some C , and Assumption 3(i) together with the Markov inequality implies $\sum_{i=1}^N \left\| \sum_{t=1}^{k^0-1} x_{it} u'_{it} \right\|^2 / (NT^2) = O_p(k^0/T^2)$. Thus we have

$$\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i(B),B} = O_p \left(\frac{\sqrt{k^0}}{T} \right).$$

We also observe that by the Cauchy-Schwarz inequality, we have

$$\begin{aligned} &\left(\frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \sum_{t=1}^{k^0-1} x'_{it} \beta_{g_i(B),B} \right)^2 \\ &\leq \left(\frac{1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \right)^2 \right) \left(\frac{1}{NT} \sum_{i=1}^N \left(\sum_{t=1}^{k^0-1} x'_{it} \beta_{g_i(B),B} \right)^2 \right). \end{aligned}$$

By Assumptions 3(i) and 3(v), it holds that

$$E \left(\frac{1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \right)^2 \right) = O \left(\frac{1}{Tk} \right),$$

and

$$\begin{aligned} E \left(\frac{1}{NT} \sum_{i=1}^N \left(\sum_{t=1}^{k^0-1} x'_{it} \beta_{g_i(B),B} \right)^2 \right) &\leq BE \left(\frac{1}{NT} \sum_{i=1}^N \left\| \sum_{t=1}^{k^0-1} x_{it} x'_{it} \right\|^2 \right) \\ &= O \left(\frac{k^0}{T} \right), \end{aligned}$$

where B satisfies $\|\beta_g^0\|^2 < B$ for any $g \in \mathbb{G}^B$. Assumption 1(ii) implies the existence of such B . Thus we have

$$\frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \sum_{t=1}^{k^0-1} x'_{it} \beta_{g_i(B),B} = O_p \left(\frac{\sqrt{k^0}}{T\sqrt{k}} \right).$$

For the third term, we again use the Cauchy-Schwarz inequality to obtain

$$\left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{k^0-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(B),B} \right)^2$$

$$\leq \left(\frac{1}{NT} \sum_{i=1}^N \left(\sum_{t=1}^{k^0-1} u_{it} \right)^2 \right) \left(\frac{1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(B),B} \right)^2 \right).$$

By Assumptions 3(i) and 3(v), it holds that

$$E \left(\frac{1}{NT} \sum_{i=1}^N \left(\sum_{t=1}^{k^0-1} u_{it} \right)^2 \right) = O \left(\frac{k^0}{T} \right),$$

and

$$\begin{aligned} & E \left(\frac{1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(B),B} \right)^2 \right) \\ & \leq BE \left(\frac{1}{NT} \sum_{i=1}^N \left\| \frac{1}{k-1} \sum_{s=1}^{k-1} x_{is} x'_{is} \right\| \right) = O \left(\frac{1}{T} \right), \end{aligned}$$

where B satisfies $\|\beta_g^0\|^2 < B$ for any $g \in \mathbb{G}^B$. Assumption 1(ii) implies the existence of such B . Thus we have

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{k^0-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(B),B} = O_p \left(\frac{\sqrt{k^0}}{T} \right).$$

We now consider the fourth term. Another application of the Cauchy-Schwarz inequality gives

$$\begin{aligned} & \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(B),B} \right)^2 \\ & \leq \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \right)^2 \right) \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(B),B} \right)^2 \right). \end{aligned}$$

By Assumptions 3(i) and 3(v), it holds that

$$E \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \right)^2 \right) = O \left(\frac{k^0}{Tk} \right),$$

and

$$\begin{aligned} & E \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(B),B} \right)^2 \right) \\ & \leq BE \left(\frac{k^0 - 1}{NT} \sum_{i=1}^N \left\| \frac{1}{k-1} \sum_{s=1}^{k-1} x_{is} x'_{is} \right\| \right) = O \left(\frac{k^0}{T} \right), \end{aligned}$$

where B satisfies $\|\beta_g^0\|^2 < B$ for any $g \in \mathbb{G}^B$. Assumption 1(ii) implies the existence of such B . Thus we have

$$\frac{k^0 - 1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(B), B} = O_p \left(\frac{k^0}{T\sqrt{k}} \right).$$

To sum up, we have

$$\begin{aligned} \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} x'_{it,k,B} \beta_{g_i(B), B} &= O_p \left(\frac{\sqrt{k^0}}{T\sqrt{k}} + \frac{\sqrt{k^0}}{T} + \frac{k^0}{T\sqrt{k}} \right) \\ &= O_p \left(\frac{\sqrt{k^0}}{T} + \frac{k^0}{T\sqrt{k}} \right). \end{aligned}$$

We then consider

$$\begin{aligned} &\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(A), A}^0 \\ &= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(A), A}^0 \\ &\quad + \frac{k^0 - 1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(A), A}^0. \end{aligned}$$

Following the similar arguments as above, for the first term, we have

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{k^0-1} u_{it} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(A), A}^0 = O_p \left(\frac{\sqrt{k^0(k-k^0)}}{T\sqrt{k}} \right).$$

The second term follows

$$\frac{k^0 - 1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(A), A}^0 = O_p \left(\frac{k^0 \sqrt{(k-k^0)}}{Tk} \right).$$

To sum up, we have

$$\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(A), A}^0 = O_p \left(\frac{\sqrt{k^0(k-k^0)}}{T\sqrt{k}} + \frac{k^0 \sqrt{(k-k^0)}}{Tk} \right).$$

We then consider

$$\begin{aligned} &\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 \\ &= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 \end{aligned}$$

$$+ \frac{k^0 - 1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0.$$

Similarly, for the first term, we have

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{k^0-1} u_{it} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 = O_p \left(\frac{\sqrt{k^0(k-k^0)}}{T\sqrt{k}} \right),$$

and for the second term, we have

$$\frac{k^0 - 1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} \beta_{g_i^0(B), B}^0 = O_p \left(\frac{k^0 \sqrt{k-k^0}}{Tk} \right).$$

To sum up, we have

$$\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} x'_{it,k,B} \beta_{g_i^0(B), B}^0 = O_p \left(\frac{\sqrt{k^0(k-k^0)}}{T\sqrt{k}} + \frac{k^0 \sqrt{k-k^0}}{Tk} \right).$$

Next, we observe that

$$\begin{aligned} \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it,k,B} x'_{it,k,B} \beta_{g_i^0(A), A}^0 &= \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i^0(A), A}^0 \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \sum_{t=k^0}^{k-1} x'_{it} \beta_{g_i^0(A), A}^0 \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \sum_{t=k^0}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(A), A}^0 \\ &\quad + \frac{k-k^0}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(A), A}^0. \end{aligned}$$

Again, using the similar arguments as above, for the first term, we have

$$\frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i^0(A), A}^0 = \frac{1}{NT} \sum_{g \in \mathbb{G}^A} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \mathbf{1}(g_i(A)^0 = g) u_{it} x'_{it} \beta_{g_i^0(A), A}^0 = O_p \left(\frac{\sqrt{k-k^0}}{\sqrt{NT}} \right).$$

For the second term, we have

$$\frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \sum_{t=k^0}^{k-1} x'_{it} \beta_{g_i^0(A), A}^0 = O_p \left(\frac{\sqrt{k-k^0}}{T\sqrt{k}} \right).$$

For the third term, using the Cauchy-Schwarz inequality and Assumptions 1(ii), 3(i), and 3(v), we have

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=k^0}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(A), A}^0 = O_p \left(\frac{\sqrt{k-k^0}}{T} \right).$$

The fourth term follows

$$\frac{k - k^0}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i^0(A), A}^0 = O_p \left(\frac{k - k^0}{T\sqrt{k}} \right).$$

To sum up, we have

$$\begin{aligned} \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it, k, B} x'_{it, k, B} \beta_{g_i^0(A), A}^0 &= O_p \left(\frac{\sqrt{k - k^0}}{\sqrt{NT}} + \frac{\sqrt{k - k^0}}{T\sqrt{k}} + \frac{\sqrt{k - k^0}}{T} + \frac{k - k^0}{T\sqrt{k}} \right) \\ &= O_p \left(\frac{\sqrt{k - k^0}}{T} + \frac{k - k^0}{T\sqrt{k}} \right). \end{aligned}$$

Similarly, we can derive the order of the following term as

$$\begin{aligned} \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it, k, B} x'_{it, k, B} \beta_{g_i(A), A} &= \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i(A), A} \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \sum_{t=k^0}^{k-1} x'_{it} \beta_{g_i(A), A} \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \sum_{t=k^0}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(A), A} \\ &\quad + \frac{k - k^0}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} \beta_{g_i(A), A} \\ &= O_p \left(\frac{\sqrt{k - k^0}}{T\sqrt{k}} + \frac{\sqrt{k - k^0}}{T} + \frac{k - k^0}{T\sqrt{k}} \right) = O_p \left(\frac{\sqrt{k - k^0}}{T} + \frac{k - k^0}{T\sqrt{k}} \right). \end{aligned}$$

We then consider

$$\begin{aligned} &\frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it, k, B} \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} \beta_{g_i^0(A), A}^0 \\ &= \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} \beta_{g_i^0(A), A}^0 \\ &\quad + \frac{k^0 - 1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} \beta_{g_i^0(A), A}^0. \end{aligned}$$

Following the similar tricks as above to analyse each of the two terms and summing them up, we have

$$\frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it, k, B} \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} \beta_{g_i^0(A), A}^0 = O_p \left(\frac{\sqrt{k^0(k - k^0)}}{T\sqrt{k}} + \frac{(k^0)^{3/2}}{Tk} \right).$$

Similarly, we can obtain that

$$\begin{aligned}
& \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it,k,B} \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} \beta_{g_i^0(B),B}^0 \\
&= \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} \beta_{g_i^0(B),B}^0 \\
&\quad + \frac{k^0-1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{t=1}^{k-1} u_{it} \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} \beta_{g_i^0(B),B}^0 \\
&= O_p \left(\frac{\sqrt{k^0(k-k^0)}}{T\sqrt{k}} + \frac{(k^0)^{3/2}}{Tk} \right).
\end{aligned}$$

We then consider the following term:

$$\begin{aligned}
\frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A} x'_{it,k,A} \beta_{g_i^0(A),A}^0 &= \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i^0(A),A}^0 \\
&\quad - \frac{1}{NT} \sum_{i=1}^N \frac{1}{T-k+1} \sum_{s=k}^T u_{is} \sum_{t=k}^T x'_{it} \beta_{g_i^0(A),A}^0 \\
&\quad - \frac{1}{NT} \sum_{i=1}^N \sum_{t=k}^T u_{it} \frac{1}{T-k+1} \sum_{s=k}^T x'_{is} \beta_{g_i^0(A),A}^0 \\
&\quad + \frac{T-k+1}{NT} \sum_{i=1}^N \frac{1}{T-k+1} \sum_{t=k}^T u_{it} \frac{1}{T-k+1} \sum_{s=k}^T x'_{is} \beta_{g_i^0(A),A}^0.
\end{aligned}$$

The first term is

$$\frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i^0(A),A}^0 = \frac{1}{NT} \sum_{g \in \mathbb{G}^B} \sum_{t=k}^T \sum_{i=1}^N \mathbf{1}(g_i(B)^0 = g) u_{it} x'_{it} \beta_{g_i^0(A),A}^0 = O_p \left(\frac{\sqrt{T-k}}{\sqrt{NT}} \right).$$

The second term is

$$\frac{1}{NT} \sum_{i=1}^N \frac{1}{T-k+1} \sum_{s=k}^T u_{is} \sum_{t=k}^T x'_{it} \beta_{g_i^0(A),A}^0 = O_p \left(\frac{1}{T} \right).$$

For the third term, we have

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=k}^T u_{it} \frac{1}{T-k+1} \sum_{s=k}^T x'_{is} \beta_{g_i^0(A),A}^0 = O_p \left(\frac{\sqrt{T-k}}{T} \right).$$

The fourth term follows

$$\frac{T-k+1}{NT} \sum_{i=1}^N \frac{1}{T-k+1} \sum_{t=k}^T u_{it} \frac{1}{T-k+1} \sum_{s=k}^T x'_{is} \beta_{g_i^0(A),A}^0 = O_p \left(\frac{\sqrt{T-k}}{T} \right).$$

To sum up, we have

$$\frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,B} x'_{it,k,B} \beta_{g_i^0(A),A}^0 = O_p \left(\frac{\sqrt{T-k}}{\sqrt{NT}} + \frac{1}{T} + \frac{\sqrt{T-k}}{T} \right) = O_p \left(\frac{\sqrt{T-k}}{T} \right).$$

Similarly, we can obtain

$$\begin{aligned} \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A} x'_{it,k,A} \beta_{g_i(A),A} &= \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it} x'_{it} \beta_{g_i(A),A} \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \frac{1}{T-k+1} \sum_{s=k}^T u_{is} \sum_{t=k}^T x'_{it} \beta_{g_i(A),A} \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \sum_{t=k}^T u_{it} \frac{1}{T-k+1} \sum_{s=k}^T x'_{is} \beta_{g_i(A),A} \\ &\quad + \frac{T-k+1}{NT} \sum_{i=1}^N \frac{1}{T-k+1} \sum_{t=k}^T u_{it} \frac{1}{T-k+1} \sum_{s=k}^T x'_{is} \beta_{g_i(A),A} \\ &= O_p \left(\frac{\sqrt{T-k}}{\sqrt{NT}} + \frac{1}{T} + \frac{\sqrt{T-k}}{T} \right) = O_p \left(\frac{\sqrt{T-k}}{T} \right). \end{aligned}$$

Now, with the order of each term in (S.5) readily there, we can obtain

$$\begin{aligned} Q(k, \gamma, \beta) - \tilde{Q}(k, \gamma, \beta) \\ = O_p \left(\frac{\sqrt{k^0}}{T} + \frac{k^0}{T\sqrt{k}} + \frac{\sqrt{k^0(k-k^0)}}{T\sqrt{k}} + \frac{k^0\sqrt{k-k^0}}{Tk} + \frac{\sqrt{k-k^0}}{T} + \frac{k-k^0}{T\sqrt{k}} + \frac{(k^0)^{3/2}}{Tk} + \frac{\sqrt{T-k}}{T} \right). \end{aligned}$$

Noting that $k^0 = O(T)$, $T-k = O(T)$, $k = O(T)$, we have

$$Q(k, \gamma, \beta) - \tilde{Q}(k, \gamma, \beta) = O_p \left(\frac{1}{\sqrt{T}} \right).$$

Cases of $k < k^0$ and $k = k^0$ can be analyzed in analogous ways and we obtain

$$\sup_{k \in \mathbb{K}, \gamma \in \mathbb{G}, \beta \in \mathbb{B}} |\tilde{Q}(k, \gamma, \beta) - Q(k, \gamma, \beta)| = O_p \left(\frac{1}{\sqrt{T}} \right).$$

□

Lemma S.7. Suppose that Assumptions 1(ii), 1(vii), 3(i)–3(vi) and 3(ix) hold. We also let $\epsilon T < k < (1-\epsilon T)$ for $\epsilon > 0$ in Assumption 1(vii). Then, we have that

$$(1) \max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \left\| \beta_{g,B}^0 - \hat{\beta}_{\tilde{g},B} \right\|^2 = O_p(1/\sqrt{T}),$$

$$(2) \max_{g \in \mathbb{G}^A} \min_{\tilde{g} \in \mathbb{G}^A} \left\| \beta_{g,A}^0 - \hat{\beta}_{\tilde{g},A} \right\|^2 = O_p(1/\sqrt{T}),$$

$$(3) (\hat{k} - k^0)/T = O_p(1/\sqrt{T}).$$

Proof. From Lemma S.6, we have

$$\begin{aligned}\tilde{Q}(\hat{k}, \hat{\gamma}, \hat{\beta}) &= Q(\hat{k}, \hat{\gamma}, \hat{\beta}) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &\leq Q(k^0, \gamma^0, \beta^0) + O_p\left(\frac{1}{\sqrt{T}}\right) = \tilde{Q}(k^0, \gamma^0, \beta^0) + O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

Because $\tilde{Q}(k, \gamma, \beta)$ is minimized at (k^0, γ^0, β^0) , we have

$$\tilde{Q}(\hat{k}, \hat{\gamma}, \hat{\beta}) - \tilde{Q}(k^0, \gamma^0, \beta^0) = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Let $a_{NT} = \tilde{Q}(\hat{k}, \hat{\gamma}, \hat{\beta}) - \tilde{Q}(k^0, \gamma^0, \beta^0)$. Note that $a_{NT} = O_p\left(1/\sqrt{T}\right)$.

Consider the case in which $k > k^0$. We observe that

$$\begin{aligned}&\tilde{Q}(k, \gamma, \beta) - \tilde{Q}(k^0, \gamma^0, \beta^0) \\ &= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\ &\quad + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(B),B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\ &\quad + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A}))^2.\end{aligned}$$

We examine each of three terms on the right hand side.

The first term is decomposed into

$$\begin{aligned}&\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\ &= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right)^2 \\ &\quad - 2 \frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \sum_{t=1}^{k^0-1} \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right) \\ &\quad + \frac{k^0-1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2.\end{aligned}$$

We observe

$$\begin{aligned}&\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right)^2 \\ &= \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{g=1}^{G^B} \sum_{\tilde{g}=1}^{G^B} \sum_{i=1}^N \mathbf{1}\{g_i^0(B) = g\} \{g_i(B) = \tilde{g}\} (x'_{it,k,B} (\beta_{g,B}^0 - \beta_{\tilde{g},B}))^2\end{aligned}$$

$$\begin{aligned} &\geq \frac{1}{T} \sum_{t=1}^{k^0-1} \sum_{g=1}^{G^B} \sum_{\tilde{g}=1}^{G^B} \rho_{F,N,t}(\gamma, g, \tilde{g}) \|\beta_{g,B}^0 - \beta_{\tilde{g},B}\|^2 \\ &\geq \frac{k^0-1}{T} \hat{\rho}_F \max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \|\beta_{g,B}^0 - \beta_{\tilde{g},B}\|^2. \end{aligned}$$

We also observe that

$$\begin{aligned} &\frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right)^2 \\ &\geq \frac{k^0-1}{T} \rho_F^* \frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B} \right\|^2. \end{aligned}$$

The Cauchy-Schwarz inequality implies

$$\begin{aligned} &\frac{1}{NT} \sum_{i=1}^N \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \sum_{t=1}^{k^0-1} \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right) \\ &\leq \left(\frac{1}{NT} \frac{1}{k-1} \sum_{i=1}^N \sum_{t=1}^{k^0-1} \sum_{s=k^0}^{k-1} \left(x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \right)^{1/2} \\ &\quad \times \left(\frac{1}{NT} \frac{1}{k-1} \sum_{i=1}^N \sum_{s=k^0}^{k-1} \sum_{t=1}^{k^0-1} \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right)^2 \right)^{1/2} \\ &= \left(\frac{1}{NT} \frac{k^0-1}{k-1} \sum_{i=1}^N \sum_{s=k^0}^{k-1} (x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0))^2 \right)^{1/2} \\ &\quad \times \left(\frac{1}{NT} \frac{k-k^0}{k-1} \sum_{i=1}^N \sum_{t=1}^{k^0-1} \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right)^2 \right)^{1/2}. \end{aligned}$$

Assumptions 3(v) and 3(ix) imply that with probability approaching one,

$$\frac{1}{NT} \frac{k^0-1}{k-1} \sum_{i=1}^N \sum_{s=k^0}^{k-1} (x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0))^2 \leq C \left(\frac{(k^0-1)(k-k^0)}{T(k-1)} \right) = C \left(\frac{k-k^0}{T} \right).$$

Similarly, we have with probability approaching one,

$$\frac{1}{NT} \frac{k-k^0}{k-1} \sum_{i=1}^N \sum_{t=1}^{k^0-1} \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B}) \right)^2 \leq C \left(\frac{k^0-k}{T} \frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B} \right\|^2 \right).$$

We also have

$$\frac{k^0-1}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \geq 0.$$

Thus we have with probability approaching one,

$$\frac{k^0 - 1}{T} \hat{\rho}_F \frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B} \right\|^2 - C \frac{\hat{k} - k^0}{T} \left(\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B} \right\|^2 \right)^{1/2} \leq a_{NT}.$$

and

$$\frac{k^0 - 1}{T} \hat{\rho}_F \max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \left\| \beta_{g, B}^0 - \hat{\beta}_{\tilde{g}, B} \right\|^2 - C \frac{\hat{k} - k^0}{T} \left(\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B} \right\|^2 \right)^{1/2} \leq a_{NT}.$$

We now consider the second term.

$$\begin{aligned} & \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(B), B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) \right)^2 \\ &= \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(B), B}) \right)^2 \\ & \quad - \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(B), B}) \right) \left(\frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) \right) \\ & \quad + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) \right)^2. \end{aligned}$$

We observe

$$\begin{aligned} & \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(B), B}^0 + \beta_{g_i(B), B}^0 - \beta_{g_i(B), B}) \right)^2 \\ &= \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(B), B}^0) \right)^2 \\ & \quad + \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(B), B}^0) x'_{it,k,B} (\beta_{g_i(B), B}^0 - \beta_{g_i(B), B}) \right) \\ & \quad + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i(B), B}^0 - \beta_{g_i(B), B}) \right)^2. \end{aligned}$$

By Assumption 3(vi) we have

$$\frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(B), B}^0) \right)^2 \geq \frac{k - k^0}{T} \underline{m}.$$

It also holds by the Cauchy-Schwarz inequality that

$$\frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i(B), B}^0) x'_{it,k,B} (\beta_{g_i(B), B}^0 - \beta_{g_i(B), B}) \right)$$

$$\geq -2 \left(\frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(B),B}^0) \right)^2 \right)^{1/2} \\ \times \left(\frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i(B),B}^0 - \beta_{g_i(B),B}) \right)^2 \right)^{1/2}.$$

By Assumptions 3(v) and 3(ix), with probability approaching one,

$$\frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(B),B}^0) \right)^2 \leq C \left(\frac{k-k^0}{T} \right).$$

It also follows that

$$\frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i(B),B}^0 - \beta_{g_i(B),B}) \right)^2 \leq C \left(\frac{k-k^0}{T} \frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i(B),B}^0 - \beta_{g_i(B),B} \right\|^2 \right).$$

We also have

$$\frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \geq 0.$$

Thus, with probability approaching one, we have

$$\frac{\hat{k}-k^0}{T} \underline{m} - C \frac{\hat{k}-k^0}{T} \left(\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i(B),B}^0 - \hat{\beta}_{g_i(B),B} \right\|^2 \right)^{1/2} < a_{NT}.$$

We then consider the third term. We observe

$$\begin{aligned} & \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A}))^2 \\ &= \frac{1}{NT} \sum_{t=k}^T \sum_{g=1}^{G^A} \sum_{\tilde{g}=1}^{G^A} \sum_{i=1}^N \mathbf{1}\{g_i^0(A) = g\} \{g_i(A) = \tilde{g}\} (x'_{it,k,A} (\beta_{g,A}^0 - \beta_{\tilde{g},A}))^2 \\ &\geq \frac{1}{T} \sum_{t=k}^T \sum_{g=1}^{G^A} \sum_{\tilde{g}=1}^{G^A} \rho_{F,N,t}(\gamma, g, \tilde{g}) \left\| \beta_{g,A}^0 - \beta_{\tilde{g},A} \right\|^2, \end{aligned}$$

by Assumption 3(iii). We thus have

$$\frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A}))^2 \geq \frac{T-k+1}{T} \hat{\rho}_F \max_{g \in \mathbb{G}^A} \min_{\tilde{g} \in \mathbb{G}^A} \left\| \beta_{g,A}^0 - \beta_{\tilde{g},A} \right\|^2.$$

Moreover, Assumption 3(iv) implies

$$\frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A}))^2 \geq \frac{T-k+1}{T} \hat{\rho}_F^* \frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A} \right\|^2.$$

Thus we have

$$\frac{1}{N} \sum_{i=1}^N \|\beta_{g_i^0(A),A}^0 - \hat{\beta}_{\hat{g}_i(A),A}\|^2 \leq C a_{NT}.$$

To sum up, for $\hat{k} > k^0$ it holds that with probability approaching one that

$$\frac{k^0 - 1}{T} \hat{\rho}_F \frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(B),B}^0 - \hat{\beta}_{\hat{g}_i(B),B} \right\|^2 - C \frac{\hat{k} - k^0}{T} \left(\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(B),B}^0 - \hat{\beta}_{\hat{g}_i(B),B} \right\|^2 \right)^{1/2} \leq a_{NT},$$

$$\frac{k^0 - 1}{T} \hat{\rho}_F \max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \left\| \beta_{g,B}^0 - \hat{\beta}_{\tilde{g},B} \right\|^2 - C \frac{\hat{k} - k^0}{T} \left(\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(B),B}^0 - \hat{\beta}_{\hat{g}_i(B),B} \right\|^2 \right)^{1/2} \leq a_{NT},$$

$$\frac{\hat{k} - k^0}{T} \underline{m} - C \frac{\hat{k} - k^0}{T} \left(\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i(B),B}^0 - \beta_{g_i(B),B} \right\|^2 \right)^{1/2} < a_{NT},$$

$$\max_{g \in \mathbb{G}^A} \min_{\tilde{g} \in \mathbb{G}^A} \left\| \beta_{g,A}^0 - \hat{\beta}_{\tilde{g},A} \right\|^2 \leq C a_{NT}.$$

$$\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(A),A}^0 - \beta_{g_i(A),A} \right\|^2 \leq C a_{NT}.$$

Similarly, for $\hat{k} \geq k^0$, it holds with probability approaching one that

$$\frac{T - k^0 + 1}{T} \hat{\rho}_F \frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(A),A}^0 - \hat{\beta}_{\hat{g}_i(A),A} \right\|^2 - C \frac{k^0 - \hat{k}}{T} \left(\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(A),A}^0 - \hat{\beta}_{\hat{g}_i(A),A} \right\|^2 \right)^{1/2} \leq a_{NT},$$

$$\frac{T - k^0 + 1}{T} \hat{\rho}_F \max_{g \in \mathbb{G}^A} \min_{\tilde{g} \in \mathbb{G}^A} \left\| \beta_{g,A}^0 - \hat{\beta}_{\tilde{g},A} \right\|^2 - C \frac{k^0 - \hat{k}}{T} \left(\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(A),A}^0 - \hat{\beta}_{\hat{g}_i(A),A} \right\|^2 \right)^{1/2} \leq a_{NT},$$

$$\frac{k^0 - \hat{k}}{T} \underline{m} - C \frac{k^0 - \hat{k}}{T} \left(\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i(A),A}^0 - \beta_{g_i(A),A} \right\|^2 \right)^{1/2} < a_{NT},$$

$$\max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \left\| \beta_{g,B}^0 - \hat{\beta}_{\tilde{g},B} \right\|^2 \leq C a_{NT}.$$

$$\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(B),B}^0 - \beta_{g_i(B),B} \right\|^2 \leq C a_{NT}.$$

They imply that

$$\begin{aligned} \frac{\hat{k} - k^0}{T} &= O_p(a_{NT}) = O_p\left(\frac{1}{\sqrt{T}}\right), \\ \frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(B), B}^0 - \hat{\beta}_{\hat{g}_i(B), B} \right\|^2 &= O_p(a_{NT}) = O_p\left(\frac{1}{\sqrt{T}}\right), \\ \max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \left\| \beta_{g, B}^0 - \hat{\beta}_{\tilde{g}, B} \right\|^2 &= O_p(a_{NT}) = O_p\left(\frac{1}{\sqrt{T}}\right) \\ \frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(A), A}^0 - \hat{\beta}_{\hat{g}_i(A), A} \right\|^2 &= O_p(a_{NT}) = O_p\left(\frac{1}{\sqrt{T}}\right), \\ \max_{g \in \mathbb{G}^A} \min_{\tilde{g} \in \mathbb{G}^A} \left\| \beta_{g, A}^0 - \hat{\beta}_{\tilde{g}, A} \right\|^2 &= O_p(a_{NT}) = O_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

□

Lemma S.8. Suppose that Assumptions 1(ii), 1(vii), 1(viii), 3(i)-3(vi), and 3(ix) are satisfied. We also let $\epsilon T < k < (1 - \epsilon T)$ for $\epsilon > 0$ in Assumption 1(vii). Then there exist permutations $\sigma_B : \mathbb{G}^B \mapsto \mathbb{G}^B$ and $\sigma_A : \mathbb{G}^A \mapsto \mathbb{G}^A$ such that $\left\| \beta_{g, B}^0 - \hat{\beta}_{\sigma_B(g), B} \right\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}^B$ and $\left\| \beta_{g, A}^0 - \hat{\beta}_{\sigma_A(g), A} \right\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}^A$.

Proof. The proof is exactly identical to that of Lemma 3 in the main text and is thus omitted. □

By relabeling, we can set $\sigma_B(g) = g$ and $\sigma_A(g) = g$. We use this convention throughout the paper. Thus we have $\left\| \beta_{g, B}^0 - \hat{\beta}_{g, B} \right\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}^B$ and $\left\| \beta_{g, A}^0 - \hat{\beta}_{g, A} \right\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}^A$.

Let \mathcal{N} be a neighborhood of β^0 such that $\left\| \beta_{g, C}^0 - \beta_{g, C} \right\| < \eta$ for $\eta > 0$ for any $g \in \mathbb{G}^C$ and $C = B, A$. Note that we will take η small enough by considering large N and T by Lemma S.8. Let $\bar{k} = \sqrt{T} \log T + k^0$ and $\underline{k} = -\sqrt{T} \log T + k^0$. Define $K = \{k : \underline{k} \leq k \leq \bar{k}\}$.

Lemma S.9. Suppose that Assumptions 1(ii), 1(vii), 1(viii), 3(v) and 3(ix) hold. We also let $\epsilon T < k < (1 - \epsilon T)$ for $\epsilon > 0$ in Assumption 1(vii). As $N, T \rightarrow \infty$ with $NT^{-\delta} \rightarrow 0$, it holds that

$$\Pr(\hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N}) \rightarrow 0.$$

Proof. To show $\Pr(\hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N}) \rightarrow 0$, it is equivalent to show that

$$\max_{1 \leq i \leq N} \sup_{\beta \in \mathcal{N}} \max_{k \in K} \mathbf{1}(\hat{g}_i(B)(k, \beta) \neq g_i^0(B) \text{ or } \hat{g}_i(A)(k, \beta) \neq g_i^0(A)) = o_p(1).$$

We observe that

$$\begin{aligned}
& \max_{1 \leq i \leq N} \sup_{\beta \in \mathcal{N}} \max_{k \in K} \mathbf{1}(\hat{g}_i(B)(k, \beta) \neq g_i^0(B) \text{ or } \hat{g}_i(A)(k, \beta) \neq g_i^0(A)) \\
&= \max_{1 \leq i \leq N} \sup_{\beta \in \mathcal{N}} \max_{(g_A, g_B) \in \mathbb{G}^B \times \mathbb{G}^A \setminus \{g_i^0(B), g_i^0(A)\}} \mathbf{1}\left(\sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B} \beta_{g_B})^2 + \sum_{t=k}^T (y_{it,k,A} - x'_{it,k,A} \beta_{g_A})^2\right. \\
&\quad \left.< \sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B} \beta_{g_i^0(B), B})^2 + \sum_{t=k}^T (y_{it,k,A} - x'_{it,k,A} \beta_{g_i^0(A), A})^2\right) \\
&< \max_{1 \leq i \leq N} \sup_{\beta \in \mathcal{N}} \max_{(g_B) \in \mathbb{G}^B \setminus \{g_i^0(B)\}} \mathbf{1}\left(\sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B} \beta_{g_B})^2 < \sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B} \beta_{g_i^0(B), B})^2\right) \\
&+ \max_{1 \leq i \leq N} \sup_{\beta \in \mathcal{N}} \max_{(g_A) \in \mathbb{G}^A \setminus \{g_i^0(A)\}} \mathbf{1}\left(\sum_{t=k}^T (y_{it,k,A} - x'_{it,k,A} \beta_{g_A})^2 < \sum_{t=k}^T (y_{it,k,A} - x'_{it,k,A} \beta_{g_i^0(A), A})^2\right).
\end{aligned}$$

We first consider cases with $k > k^0$. Let

$$\begin{aligned}
d = & \sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B} \beta_{g_B})^2 - \sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B} \beta_{g_i^0(B), B})^2 \\
& - \sum_{t=1}^{k^0-1} (y_{it,k^0,B} - x'_{it,k^0,B} \beta_{g_B})^2 + \sum_{t=1}^{k^0-1} (y_{it,k^0,B} - x'_{it,k^0,B} \beta_{g_i^0(B), B})^2.
\end{aligned}$$

With some algebra, we can rewrite d as

$$\begin{aligned}
d = & \sum_{t=k^0}^{k-1} \left(x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) + u_{it,k,B} \right)^2 \\
& - \sum_{t=k^0}^{k-1} \left(x'_{it,k,B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) + u_{it,k,B} \right)^2 \\
& + \sum_{t=1}^{k^0-1} \left(x'_{it,k,B} (\beta_{g_i^0(B), B}^0 - \beta_{g_B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) + u_{it,k,B} \right)^2 \\
& - \sum_{t=1}^{k^0-1} \left(x'_{it,k,B} (\beta_{g_i^0(B), B}^0 - \beta_{g_i^0(B), B}^0) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) + u_{it,k,B} \right)^2 \\
& - \sum_{t=1}^{k^0-1} (x'_{it,k^0,B} (\beta_{g_i^0(B), B}^0 - \beta_{g_B}) + u_{it,k^0,B})^2 + \sum_{t=1}^{k^0-1} (x'_{it,k^0,B} (\beta_{g_i^0(B), B}^0 - \beta_{g_i^0(B), B}^0) + u_{it,k^0,B})^2 \\
& = (\beta_{g_i^0(B), B}^0 - \beta_{g_B})' \sum_{t=k^0}^{k-1} x_{it,k,B} x'_{it,k,B} (2\beta_{g_i^0(A), A}^0 - \beta_{g_B} - \beta_{g_i^0(B), B}^0) \\
& - 2 \sum_{t=k^0}^{k-1} \left(x'_{it,k,B} (\beta_{g_i^0(B), B}^0 - \beta_{g_B}) \right) \left(\frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) \right)
\end{aligned}$$

$$\begin{aligned}
& -2 \sum_{t=k^0}^{k-1} \left(x'_{it,k,B} (\beta_{g_i^0(B),B} - \beta_{g_B}) \right) u_{it,k,B} \\
& + (\beta_{g_i^0(B),B} - \beta_{g_B})' \sum_{t=1}^{k^0-1} (x_{it,k,B} x'_{it,k,B} - x_{it,k^0,B} x'_{it,k^0,B}) (2\beta_{g_i^0(B),B}^0 - \beta_{g_B} - \beta_{g_i^0(B),B}) \\
& - 2 \sum_{t=1}^{k^0-1} \left(x'_{it,k,B} (\beta_{g_i^0(B),B} - \beta_{g_B}) \right) \left(\frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \\
& + 2 \sum_{t=1}^{k^0-1} x'_{it,k^0,B} (\beta_{g_i^0(B),B} - \beta_{g_B}) u_{it,k^0,B} - 2 \sum_{t=1}^{k^0-1} x'_{it,k,B} (\beta_{g_i^0(B),B} - \beta_{g_B}) u_{it,k,B}.
\end{aligned}$$

We also note that

$$\begin{aligned}
& \sum_{t=1}^{k^0-1} x'_{it,k^0,B} (\beta_{g_i^0(B),B} - \beta_{g_B}) u_{it,k^0,B} - \sum_{t=1}^{k^0-1} x'_{it,k,B} (\beta_{g_i^0(B),B} - \beta_{g_B}) u_{it,k,B} \\
& = (k^0 - 1) \left(\frac{1}{k^0 - 1} \sum_{s=1}^{k^0-1} x_{is} \right)' \left(\frac{1}{k^0 - 1} \sum_{s=1}^{k^0-1} u_{is} \right) (\beta_{g_i^0(B),B} - \beta_{g_B}) \\
& + \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x_{is} \right)' \left(\sum_{t=1}^{k^0-1} u_{it} \right) (\beta_{g_i^0(B),B} - \beta_{g_B}) \\
& + \left(\sum_{t=1}^{k^0-1} x_{it} \right)' \left(\frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \right) (\beta_{g_i^0(B),B} - \beta_{g_B}) \\
& - (k^0 - 1) \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x_{is} \right)' \left(\frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \right) (\beta_{g_i^0(B),B} - \beta_{g_B}).
\end{aligned}$$

By Assumption 1(ii) and the Cauchy-Schwarz inequality, it holds that there exists a universal constant M that

$$\begin{aligned}
& (\beta_{g_i^0(B),B} - \beta_{g_B})' \sum_{t=k^0}^{k-1} x_{it,k,B} x'_{it,k,B} (2\beta_{g_i^0(A),A}^0 - \beta_{g_B} - \beta_{g_i^0(B),B}^0) \leq M \left\| \sum_{t=k^0}^{k-1} x_{it,k,B} x'_{it,k,B} \right\|, \\
& \left| \sum_{t=k^0}^{k-1} \left(x'_{it,k,B} (\beta_{g_i^0(B),B} - \beta_{g_B}) \right) \left(\frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \right| \\
& \leq M \frac{1}{k-1} \left\| \sum_{t=k^0}^{k-1} x_{it,k,B} \right\| \left\| \sum_{s=1}^{k^0-1} x_{is} \right\|, \\
& \left| \sum_{t=k^0}^{k-1} \left(x'_{it,k,B} (\beta_{g_i^0(B),B} - \beta_{g_B}) \right) u_{it,k,B} \right| \leq M \sum_{t=k^0}^{k-1} \|x_{it,k,B} u_{it,k,B}\|,
\end{aligned}$$

$$\begin{aligned}
& (\beta_{g_i^0(B),B} - \beta_{g_B})' \sum_{t=1}^{k^0-1} (x_{it,k,B} x'_{it,k,B} - x_{it,k^0,B} x'_{it,k^0,B}) (2\beta_{g_i^0(B),B}^0 - \beta_{g_B} - \beta_{g_i^0(B),B}) \\
& \leq M \left\| \sum_{t=1}^{k^0-1} (x_{it,k,B} x'_{it,k,B} - x_{it,k^0,B} x'_{it,k^0,B}) \right\|, \\
& \quad \sum_{t=1}^{k^0-1} (x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_B})) \left(\frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \\
& \leq M \frac{1}{k-1} \left\| \sum_{t=1}^{k^0-1} x_{it,k,B} \right\| \left\| \sum_{s=k^0}^{k-1} x_{is} \right\|,
\end{aligned}$$

and

$$\begin{aligned}
& \left\| \sum_{t=1}^{k^0-1} x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_B}) u_{it,k^0,B} - \sum_{t=1}^{k^0-1} x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_B}) u_{it,k,B} \right\| \\
& \leq M \left(\frac{1}{k^0-1} \left\| \sum_{s=1}^{k^0-1} x_{is} \right\| \left\| \sum_{s=1}^{k^0-1} u_{is} \right\| + \frac{1}{k-1} \left\| \sum_{s=1}^{k-1} x_{is} \right\| \left\| \sum_{t=1}^{k^0-1} u_{it} \right\| \right. \\
& \quad \left. + \frac{1}{k-1} \left\| \sum_{t=1}^{k^0-1} x_{it} \right\| \left\| \sum_{s=1}^{k-1} u_{is} \right\| + \frac{k^0-1}{(k-1)^2} \left\| \sum_{s=1}^{k-1} x_{is} \right\| \left\| \sum_{s=1}^{k-1} u_{is} \right\| \right).
\end{aligned}$$

Thus we have

$$\begin{aligned}
|d| & \leq 2M \left(\left\| \sum_{t=k^0}^{k-1} x_{it,k,B} x'_{it,k,B} \right\| + \frac{1}{k-1} \left\| \sum_{t=k^0}^{k-1} x_{it,k,B} \right\| \left\| \sum_{s=1}^{k^0-1} x_{is} \right\| + \sum_{t=k^0}^{k-1} \|x_{it,k,B} u_{it,k,B}\| \right. \\
& \quad \left. + \left\| \sum_{t=1}^{k^0-1} (x_{it,k,B} x'_{it,k,B} - x_{it,k^0,B} x'_{it,k^0,B}) \right\| + \frac{1}{k-1} \left\| \sum_{t=1}^{k^0-1} x_{it,k,B} \right\| \left\| \sum_{s=k^0}^{k-1} x_{is} \right\| \right. \\
& \quad \left. + \frac{1}{k^0-1} \left\| \sum_{s=1}^{k^0-1} x_{is} \right\| \left\| \sum_{s=1}^{k^0-1} u_{is} \right\| + \frac{1}{k-1} \left\| \sum_{s=1}^{k-1} x_{is} \right\| \left\| \sum_{t=1}^{k^0-1} u_{it} \right\| \right. \\
& \quad \left. + \frac{1}{k-1} \left\| \sum_{t=1}^{k^0-1} x_{it} \right\| \left\| \sum_{s=1}^{k-1} u_{is} \right\| + \frac{k^0-1}{(k-1)^2} \left\| \sum_{s=1}^{k-1} x_{is} \right\| \left\| \sum_{s=1}^{k-1} u_{is} \right\| \right) \\
& \leq 2M \left(\sum_{t=k^0}^{\bar{k}-1} \|x_{it,k,B} x'_{it,k,B}\| + \sum_{t=k^0}^{\bar{k}-1} \|x_{it,k,B}\| \frac{1}{k^0-1} \left\| \sum_{s=1}^{k^0-1} x_{is} \right\| \right. \\
& \quad \left. + \sum_{t=k^0}^{\bar{k}-1} \|x_{it,k,B} u_{it,k,B}\| + \left\| \sum_{t=1}^{k^0-1} (x_{it,k,B} x'_{it,k,B} - x_{it,k^0,B} x'_{it,k^0,B}) \right\| \right. \\
& \quad \left. + \frac{1}{k^0-1} \sum_{t=1}^{k^0-1} \|x_{it,k,B}\| \sum_{s=k^0}^{\bar{k}-1} \|x_{is}\| \right)
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{k^0 - 1} \sum_{s=1}^{k^0-1} \|x_{is}\| \left\| \sum_{s=1}^{k^0-1} u_{is} \right\| + \frac{1}{k-1} \sum_{s=1}^{k-1} \|x_{is}\| \left\| \sum_{t=1}^{k^0-1} u_{it} \right\| \\
& + \frac{1}{k-1} \sum_{t=1}^{k^0-1} \|x_{it}\| \left\| \sum_{s=1}^{k-1} u_{is} \right\| + \frac{k^0 - 1}{(k-1)^2} \sum_{s=1}^{\bar{k}-1} \|x_{is}\| \left\| \sum_{s=1}^{k-1} u_{is} \right\| \Big).
\end{aligned}$$

Let $M_T = T^{1/4}/\log T$. Under Assumption 3(ix), we can apply inequality (1.8) in Merlevède et al. (2011) which is based on Theorem 6.2 of Rio (2017) with $\lambda = (\bar{k} - k^0)M_T = T^{3/4}$ and obtain

$$\begin{aligned}
& \Pr \left(\frac{1}{\bar{k} - k^0} \left| \sum_{t=k^0}^{\bar{k}-1} (\|x_{it,k,B} x'_{it,k,B}\| - E(\|x_{it,k,B} x'_{it,k,B}\|)) \right| > M_T \right) \\
& \leq 4 \exp \left(-\frac{\lambda^{d/(d+1)} \log 2}{2} \right) + 16CM_T^{-1} \exp \left(-a \frac{\lambda^{d/(d+1)}}{b^d} \right) \\
& = o(T^{-\delta}),
\end{aligned}$$

where $d = d_1 d_2 / (d_1 + d_2)$ and C is a positive constant. Noting that

$$\frac{1}{\bar{k} - k^0} \left| \sum_{t=k^0}^{\bar{k}-1} E(\|x_{it,k,B} x'_{it,k,B}\|) \right|$$

converges and $M_T \rightarrow \infty$, we have

$$\Pr \left(\frac{1}{\bar{k} - k^0} \left\| \sum_{t=k^0}^{\bar{k}-1} x_{it,k,B} x'_{it,k,B} \right\| > M_T \right) = o(T^{-\delta}).$$

Similarly, we have

$$\Pr \left(\frac{1}{\bar{k} - k^0} \left\| \sum_{t=k^0}^{\bar{k}-1} x_{it,k,B} \right\| > M_T \right) = o(T^{-\delta}),$$

and

$$\Pr \left(\frac{1}{\bar{k} - k^0} \left\| \sum_{t=k^0}^{\bar{k}-1} x_{it,k,B} u_{it,k,B} \right\| > M_T \right) = o(T^{-\delta}).$$

Inequality (1.8) in Merlevède et al. (2011) also implies that uniformly over $k \in [k^0, \bar{k}]$,

$$\Pr \left(\frac{1}{k-1} \sum_{t=1}^{k-1} \|x_{it}\| > T^{1/4} \right) = o(T^{-\delta}),$$

and

$$\Pr \left(\frac{1}{k-1} \left\| \sum_{s=1}^{k-1} u_{is} \right\| > T^{-1/4} \right) = o(T^{-\delta}).$$

Lastly, we have

$$\begin{aligned}
& \sum_{t=1}^{k^0-1} (x_{it,k,B} x'_{it,k,B} - x_{it,k^0,B} x'_{it,k^0,B}) \\
&= -\frac{1}{k-1} \sum_{t=1}^{k^0-1} x_{it} \sum_{s=1}^{k-1} x'_{is} - \frac{1}{k-1} \sum_{t=1}^{k-1} x_{it} \sum_{s=1}^{k^0-1} x'_{is} + \frac{k^0-1}{(k-1)^2} \sum_{t=1}^{k-1} x_{it} \sum_{s=1}^{k-1} x'_{is} \\
&\quad + \frac{1}{(k^0-1)} \sum_{t=1}^{k^0-1} x_{it} \sum_{s=1}^{k^0-1} x'_{is} \\
&= -\frac{1}{k-1} \sum_{t=k^0}^{k-1} x_{it} \sum_{s=1}^{k-1} x'_{is} - \frac{k-k^0}{(k-1)^2} \sum_{t=1}^{k-1} x_{it} \sum_{s=1}^{k-1} x'_{is} - \frac{1}{k-1} \sum_{t=k^0}^{k-1} x_{it} \sum_{s=1}^{k^0-1} x'_{is} \\
&\quad + \frac{k-k^0}{(k-1)(k^0-1)} \sum_{t=1}^{k^0-1} x_{it} \sum_{s=1}^{k^0-1} x'_{is}.
\end{aligned}$$

Thus we have

$$\begin{aligned}
& \left\| \sum_{t=1}^{k^0-1} (x_{it,k,B} x'_{it,k,B} - x_{it,k^0,B} x'_{it,k^0,B}) \right\| \\
&\leq \frac{1}{k-1} \sum_{t=k^0}^{k-1} \|x_{it}\| \sum_{s=1}^{k-1} \|x_{is}\| + \frac{k-k^0}{(k-1)^2} \sum_{t=1}^{k-1} \|x_{it}\| \sum_{s=1}^{k-1} \|x_{is}\| \\
&\quad + \frac{1}{k-1} \sum_{t=k^0}^{k-1} \|x_{it}\| \sum_{s=1}^{k^0-1} \|x_{is}\| + \frac{k-k^0}{(k-1)(k^0-1)} \sum_{t=1}^{k^0-1} \|x_{it}\| \sum_{s=1}^{k^0-1} \|x_{is}\|.
\end{aligned}$$

Following the argument based on inequality (1.8) in [Merlevède et al. \(2011\)](#) indicates that

$$\Pr \left(\frac{1}{k^0-1} \left\| \sum_{t=1}^{k^0-1} (x_{it,k,B} x'_{it,k,B} - x_{it,k^0,B} x'_{it,k^0,B}) \right\| > M_T \right) = o(T^\delta).$$

This implies that there exists a sequence such that $C_T = O(M_T)$ and $C_T \rightarrow \infty$ as $T \rightarrow \infty$ such that

$$\Pr \left(\frac{1}{k^0} |d| > \frac{k^0-k}{k^0} C_T \right) = o(T^{-\delta}).$$

By a similar argument, the above bound holds for $k \leq k^0$.

Next, we consider the main term. This term can be considered in a similar way to [Bonhomme and Manresa \(2015\)](#) and [Okui and Wang \(2021\)](#).

$$\sum_{t=1}^{k^0-1} \left((y_{it,k^0,B} - x'_{it,k^0,B} \beta_{g,B})^2 - (y_{it,k^0,B} - x'_{it,k^0,B} \beta_{g_i^0(B),B})^2 \right)$$

$$\begin{aligned}
&= \sum_{t=1}^{k^0-1} 2u_{it,k^0,B}x_{it,k^0,B}(\beta_{g_i^0(B),B} - \beta_{g,B}) \\
&\quad + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B} - \beta_{g,B})'x_{it,k^0,B}x'_{it,k^0,B}(2\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B} - \beta_{g,B}) \\
&= \sum_{t=1}^{k^0-1} 2u_{it,k^0,B}x_{it,k^0,B}(\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \\
&\quad + \sum_{t=1}^{k^0-1} 2u_{it,k^0,B}x_{it,k^0,B}(\beta_{g_i^0(B),B} - \beta_{g,B} - \beta_{g_i^0(B),B}^0 + \beta_{g,B}^0) \\
&\quad + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0)'x_{it,k^0,B}x'_{it,k^0,B}(\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \\
&\quad + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B} - \beta_{g,B} - \beta_{g_i^0(B),B}^0 + \beta_{g,B}^0)'x_{it,k^0,B}x'_{it,k^0,B}(2\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B} - \beta_{g,B}) \\
&\quad + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0)'x_{it,k^0,B}x'_{it,k^0,B}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B} - \beta_{g,B} + \beta_{g,B}^0).
\end{aligned}$$

By the Cauchy-Schwarz inequality, Assumption 1(ii) and the definition of \mathcal{N} imply that

$$\begin{aligned}
&\left| \sum_{t=1}^{k^0-1} 2u_{it,k^0,B}x_{it,k^0,B}(\beta_{g_i^0(B),B} - \beta_{g,B} - \beta_{g_i^0(B),B}^0 + \beta_{g,B}^0) \right. \\
&\quad \left. + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B} - \beta_{g,B} - \beta_{g_i^0(B),B}^0 + \beta_{g,B}^0)'x_{it,k^0,B}x'_{it,k^0,B}(2\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B} - \beta_{g,B}) \right. \\
&\quad \left. + \sum_{t=1}^{k^0-1} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0)'x_{it,k^0,B}x'_{it,k^0,B}(\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B} - \beta_{g,B} + \beta_{g,B}^0) \right| \\
&\leq \eta C_1 \left\| \sum_{t=1}^{k^0-1} u_{it,k^0,B}x_{it,k^0,B} \right\| + \eta C_2 \left\| \sum_{t=1}^{k^0-1} x_{it,k^0,B}x'_{it,k^0,B} \right\|,
\end{aligned}$$

where C_1 and C_2 are constants independent of η and T .

We then have

$$\begin{aligned}
&\mathbf{1} \left(\sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B}\beta_{g_i(B),B})^2 < \sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B}\beta_{g_i^0(B),B}^0)^2 \right) \\
&\leq \mathbf{1} \left(\sum_{t=1}^{k^0-1} 2u_{it,k^0,B}x'_{it,k^0,B}(\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \right. \\
&\quad \left. < - \sum_{t=1}^{k^0-1} (x'_{it,k^0,B}(\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 \right)
\end{aligned}$$

$$+ \eta C_1 \left\| \sum_{t=1}^{k^0-1} u_{it,k^0,B} x_{it,k^0,B} \right\| + \eta C_2 \left\| \sum_{t=1}^{k^0-1} x_{it,k^0,B} x'_{it,k^0,B} \right\| + |d| \Bigg).$$

Note that the right hand side does not depend on β . Thus, we have

$$\begin{aligned} & \Pr \left(\max_{g_B \in \mathbb{G}^B \setminus \{g_i^0(B)\}} \mathbf{1} \left(\sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B} \beta_{g_i(B),B})^2 \right. \right. \\ & < \sum_{t=1}^{k-1} (y_{it,k,B} - x'_{it,k,B} \beta_{g_i^0(B),B})^2 \Big) \neq 0 \Big) \\ & \leq \sum_{g \in \mathbb{G}^B \setminus \{g_i^0(B)\}} \Pr \left(\sum_{t=1}^{k^0-1} 2u_{it,k^0,B} x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \right. \\ & < - \sum_{t=1}^{k^0-1} (x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 \\ & \quad \left. \left. + \eta C_1 \left\| \sum_{t=1}^{k^0-1} u_{it,k^0,B} x_{it,k^0,B} \right\| + \eta C_2 \left\| \sum_{t=1}^{k^0-1} x_{it,k^0,B} x'_{it,k^0,B} \right\| + |d| \right) \right) \\ & \leq \sum_{g \in \mathbb{G}^B \setminus \{g_i^0(B)\}} \Pr \left(\sum_{t=1}^{k^0-1} 2u_{it,k^0,B} x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \right. \\ & < - \sum_{t=1}^{k^0-1} (x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 \\ & \quad \left. \left. + \eta C_1 \left\| \sum_{t=1}^{k^0-1} u_{it,k^0,B} x_{it,k^0,B} \right\| + \eta C_2 \left\| \sum_{t=1}^{k^0-1} x_{it,k^0,B} x'_{it,k^0,B} \right\| + |d| \right) \right) \\ & \leq \sum_{g \in \mathbb{G}^B \setminus \{g_i^0(B)\}} \left(\Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} (x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 \leq \frac{c''}{2} \right) \right. \\ & \quad + \Pr \left(\left\| \frac{1}{k^0} \sum_{t=1}^{k^0-1} u_{it,k^0,B} x_{it,k^0,B} \right\| \geq M \right) + \Pr \left(\left\| \frac{1}{k^0} \sum_{t=1}^{k^0-1} x_{it,k^0,B} x'_{it,k^0,B} \right\| \geq M \right) \\ & \quad + \Pr \left(\frac{1}{k^0} |d| > \frac{k^0 - k}{k^0} C_T \right) \\ & \quad \left. + \Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} 2u_{it,k^0,B} x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) < -\frac{c''}{2} + \eta C_1 M + \eta C_2 M + \frac{k^0 - k}{k^0} C_T \right) \right), \end{aligned}$$

where we take $c'' = c \times \rho_D^*$ for c in Assumption 1(*viii*).

We observe

$$\Pr \left(\left\| \frac{1}{k^0} \sum_{t=1}^{k^0-1} x_{it,k^0,B} x'_{it,k^0,B} \right\| \geq M \right) \leq \Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} \|x_{it,k^0,B} x'_{it,k^0,B}\| \geq M \right)$$

$$= \Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} x'_{it,k^0,B} x_{it,k^0,B} \geq M \right).$$

We then apply Lemma S.5 with $x'_{it,k^0,B} x_{it,k^0,B} - E(x'_{it,k^0,B} x_{it,k^0,B})$ as z_t in the lemma. Combining with Assumption 3(ix) and the fact that $\sum_{t=1}^{k^0-1} E(x'_{it,k^0,B} x_{it,k^0,B})/k^0$ converges, we have

$$\Pr \left(\left\| \frac{1}{k^0} \sum_{t=1}^{k^0-1} x'_{it,k^0,B} x_{it,k^0,B} \right\| \geq M \right) = o((k^0)^{-\delta}) = o(T^{-\delta}),$$

where the last equality holds by Assumption 1(vii). Similarly, noting that $\sum_{t=1}^{k^0-1} E \|u_{it,k^0,B} x_{it,k^0,B}\| / k^0$ converges, Assumption 3(ix) implies

$$\Pr \left(\left\| \frac{1}{k^0} \sum_{t=1}^{k^0-1} u_{it,k^0,B} x_{it,k^0,B} \right\| \geq M \right) = o(T^{-\delta}).$$

Moreover, a similarly argument shows that under Assumption 3(ix), Lemma S.5 implies

$$\Pr \left(\left| \frac{1}{k^0} \sum_{t=1}^{k^0-1} (x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 - \frac{1}{k^0} \sum_{t=1}^{k^0-1} E((x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2) \right| \geq \frac{c''}{2} \right) = o(T^{-\delta}),$$

which in turn implies that under Assumptions 1(viii) and 3(v),

$$\Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} (x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0))^2 \leq \frac{c''}{2} \right) = o(T^{-\delta})$$

uniformly over g . We have shown that $\Pr((k^0)^{-1}|d| > ((k^0 - \underline{k})/k^0)C_T) = o(T^{-\delta})$. Note that $((k^0 - \underline{k})/k^0)C_T \rightarrow 0$ because $M_T = o(\sqrt{T}/\log T)$, $k^0 = O(T)$ and $k^0 - \underline{k} = O(\sqrt{T}\log T)$. Lastly, we consider

$$\begin{aligned} & \sum_{t=1}^{k^0-1} u_{it,k^0,B} x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) \\ &= \sum_{t=1}^{k^0-1} u_{it} x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0) + \frac{1}{k^0-1} \sum_{t=1}^{k^0-1} u_{it} \sum_{t=1}^{k^0-1} x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g,B}^0). \end{aligned}$$

By the argument based on inequality (1.8) in Merlevède et al. (2011) and Lemma S.5, we have

$$\Pr \left(\left| \frac{1}{k^0-1} \sum_{t=1}^{k^0-1} u_{it} \right| > c \right) = o(T^\delta)$$

for any c and

$$\Pr \left(\frac{1}{k^0 - 1} \left\| \sum_{t=1}^{k^0-1} x'_{it} (\beta_{g_i^0(B), B}^0 - \beta_{g, B}^0) \right\| > M_T \right) = o(T^{-\delta}).$$

We also note that $E(u_{it}x_{it}) = 0$. Thus, we can take η small enough and also T large enough such that

$$\begin{aligned} & \Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} 2u_{it,k^0,B} x'_{it,k^0,B} (\beta_{g_i^0(B), B}^0 - \beta_{g, B}^0) < -\frac{c''}{2} + \eta C_1 M + \eta C_2 M + \frac{k^0 - k}{k^0} C_T \right) \\ & \leq \Pr \left(\frac{1}{k^0} \sum_{t=1}^{k^0-1} 2u_{it,k^0,B} x'_{it,k^0,B} (\beta_{g_i^0(B), B}^0 - \beta_{g, B}^0) < -\frac{c''}{4} \right). \end{aligned}$$

This probability is also $o(T^{-\delta})$ uniformly over g under Assumption 3(ix) by Lemma S.5 following a similar argument to those discussed above.

It thus follows that

$$\begin{aligned} & \Pr \left(\max_{g_i(B) \in \mathbb{G}^B \setminus \{g_i^0(B)\}} \mathbf{1} \left(\sum_{t=1}^{k-1} (y_{it,k^0,B} - x'_{it,k^0,B} \beta_{g_i(B), B})^2 \right. \right. \\ & \quad \left. \left. < \sum_{t=1}^{k-1} (\Delta y_{it,k^0,B} - x'_{it,k^0,B} \beta_{g_i^0(B), B})^2 \right) \neq 0 \right) = o(NT^{-\delta}). \end{aligned}$$

A similarly argument shows that

$$\begin{aligned} & \Pr \left(\max_{g_i(A) \in \mathbb{G}^A \setminus \{g_i^0(A)\}} \mathbf{1} \left(\sum_{t=k}^T (y_{it,k,A} - x'_{it,k,A} \beta_{g_i(A), A})^2 < \sum_{t=k}^T (y_{it,k,A} - \Delta x'_{it,k,A} \beta_{g_i^0(A), A})^2 \right) \neq 0 \right) \\ & = o(NT^{-\delta}). \end{aligned}$$

We use a similar argument for $k \leq k^0$ and the proof is complete. \square

S.3.4.2 Proof of Theorem S.2

Proof. We observe that

$$\begin{aligned} \Pr(\hat{k} \neq k^0) & \leq \Pr(\hat{k} \neq k^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{\beta} \notin \mathcal{N}) \\ & \leq \Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{\gamma} \neq \gamma^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{\beta} \notin \mathcal{N}). \end{aligned}$$

First, Lemma S.8 and the discussion below it imply that $\Pr(\hat{\beta} \notin \mathcal{N}) \rightarrow 0$. Second we have

$$\begin{aligned} \Pr(\hat{\gamma} \neq \gamma^0, \hat{\beta} \in \mathcal{N}) & \leq \Pr(\hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N}, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{k} \notin K) \\ & \leq \Pr(\hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N}) + \Pr(\hat{k} \notin K) \rightarrow 0, \end{aligned}$$

by Lemmas S.7, S.8 and S.9.

We now consider the third term. We observe that

$$\begin{aligned}
& \Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) \\
& \leq \Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}, \hat{k} \in K) + \Pr(\hat{k} \notin K) \\
& \leq \Pr(\hat{k}(\gamma^0, \beta) \neq k^0 \text{ for some } \beta \in \mathcal{N}, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{k} \notin K) \\
& \leq \Pr(\hat{k}(\gamma^0, \beta) \neq k^0 \text{ for some } \beta \in \mathcal{N}) + \Pr(\hat{k} \notin K),
\end{aligned}$$

where

$$\hat{k}(\gamma^0, \beta) = \operatorname{argmin}_{k \in K} Q(k, \gamma, \beta).$$

Note that $\Pr(\hat{k} \notin K) \rightarrow 0$ by Lemma S.7. Note that $\hat{k}(\gamma^0, \beta) \neq k^0$ is equivalent to

$$Q(k^0, \gamma^0, \beta) > \min_{k \in K \setminus \{k^0\}} Q(k, \gamma^0, \beta) = \min \left(\min_{k > k^0} Q(k, \gamma^0, \beta), \min_{k < k^0} Q(k, \gamma^0, \beta) \right).$$

Thus, we have

$$\begin{aligned}
& \Pr(\hat{k}(\gamma^0, \beta) \neq k^0 \text{ for some } \beta \in \mathcal{N}) \\
& \leq \Pr \left(Q(k^0, \gamma^0, \beta) > \min_{k^0 < k \leq \bar{k}} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right) \\
& \quad + \Pr \left(Q(k^0, \gamma^0, \beta) > \min_{\underline{k} \leq k < k^0} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right).
\end{aligned}$$

Suppose for the moment that $\underline{k} \leq k < k^0$. Noting that

$$\begin{aligned}
& Q(k, \gamma^0, \beta) \\
& = \frac{1}{NT} \left(\sum_{t=1}^{k-1} \sum_{i=1}^N (y_{it,k,B} - x'_{it,k,B} \beta_{g_i^0(B),B})^2 + \sum_{t=k}^T \sum_{i=1}^N (y_{it,k,A} - x'_{it,k,A} \beta_{g_i^0(A),A})^2 \right) \\
& = \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\
& \quad + \frac{2}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \\
& \quad + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\
& \quad + \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it,k,B} \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}))^2 \\
& + \frac{2}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A} (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A})) \\
& + \frac{1}{NT} \sum_{t=1}^{k-1} \sum_{i=1}^N u_{it,k,B}^2 + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A}^2,
\end{aligned}$$

and

$$\begin{aligned}
& Q(k^0, \gamma^0, \beta) \\
& = \frac{1}{NT} \left(\sum_{t=1}^{k^0-1} \sum_{i=1}^N (y_{it,k^0,B} - x'_{it,k^0,B} \beta_{g_i^0(B),B})^2 + \sum_{t=k}^T \sum_{i=1}^N (y_{it,k^0,A} - x'_{it,k^0,A} \beta_{g_i^0(A),A})^2 \right) \\
& = \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) \right)^2 \\
& \quad + \frac{2}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k^0,B} \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) \right) \\
& \quad + \frac{1}{NT} \sum_{t=k^0}^T \sum_{i=1}^N (x'_{it,k^0,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}))^2 \\
& \quad + \frac{2}{NT} \sum_{t=k^0}^T \sum_{i=1}^N u_{it,k^0,A} (x'_{it,k^0,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A})) \\
& \quad + \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k^0,B}^2 + \frac{1}{NT} \sum_{t=k^0}^T \sum_{i=1}^N u_{it,k^0,A}^2,
\end{aligned}$$

we have

$$\begin{aligned}
& Q(k^0, \gamma^0, \beta) - Q(k, \gamma^0, \beta) \\
& = \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}) \right)^2 \\
& \quad + \frac{2}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} \left(x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) - \frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}) \right) \\
& \quad + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}) \right)^2 \\
& \quad + \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it,k,B} \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}) + \frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}) \right)
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}))^2 + \frac{2}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A} (x'_{it,k,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A})) \\
& + \frac{1}{NT} \sum_{t=1}^{k-1} \sum_{i=1}^N u_{it,k,B}^2 + \frac{1}{NT} \sum_{t=k}^T \sum_{i=1}^N u_{it,k,A}^2 - \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N (x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}))^2 \\
& - \frac{2}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k^0,B} (x'_{it,k^0,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B})) - \frac{1}{NT} \sum_{t=k^0}^T \sum_{i=1}^N (x'_{it,k^0,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A}))^2 \\
& - \frac{2}{NT} \sum_{t=k^0}^T \sum_{i=1}^N u_{it,k^0,A} (x'_{it,k^0,A} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(A),A})) - \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k^0,B}^2 - \frac{1}{NT} \sum_{t=k^0}^T \sum_{i=1}^N u_{it,k^0,A}^2 \\
& = - \frac{2}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N x'_{it,k,B} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \\
& + \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\
& + \frac{1}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(\frac{1}{k^0-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) \right)^2 \\
& - \frac{2}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) \right) \\
& + \frac{2}{NT} \sum_{t=1}^{k^0-1} \frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) \right) \\
& + \frac{2}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N \left(\frac{1}{k^0-1} \sum_{s=1}^{k^0} u_{it} \right) \left(\frac{1}{k^0-1} \sum_{s=1}^{k^0} x'_{it} (\beta_{g_i^0(B),B}^0 - \beta_{g_i^0(B),B}) \right) \\
& - \frac{2}{NT} \sum_{t=1}^{k^0-1} \sum_{i=1}^N u_{it,k,B} \left(\frac{1}{k-1} \sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \\
& + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\
& + \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \left(\frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) \\
& + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \\
& + \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} x'_{it} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0)
\end{aligned}$$

$$\begin{aligned}
& - \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} \left(\frac{1}{k-1} \sum_{s=1}^{k-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}) \right) \\
& + \frac{2}{NT} \sum_{i=1}^N \left(\frac{1}{k-1} \sum_{s=1}^{k-1} u_{is} \right) \sum_{t=1}^{k^0-1} x'_{its} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}) \\
& + \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it, k, B} \left(\frac{1}{k-1} \sum_{s=1}^{k^0-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) \right) \\
& + 2 \frac{k-k^0}{NT} \frac{1}{(T-k+1)(T-k^0+1)} \sum_{i=1}^N \left(\sum_{s=k}^T x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), A}) \right)^2 \\
& - 4 \frac{1}{NT} \frac{1}{T-k^0+1} \sum_{i=1}^N \left(\sum_{s=k}^T x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), A}) \right) \left(\sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), A}) \right) \\
& - 2 \frac{1}{NT} \frac{1}{T-k^0+1} \sum_{i=1}^N \left(\sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), A}) \right)^2 \\
& + 2 \frac{k-k^0}{NT} \frac{1}{(T-k+1)(T-k^0+1)} \sum_{i=1}^N \left(\sum_{s=k}^T u_{is} \right) \left(\sum_{s=k}^T x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), A}) \right) \\
& - 2 \frac{1}{NT} \frac{1}{T-k^0+1} \sum_{i=1}^N \left(\sum_{s=k^0}^T u_{is} \right) \left(\sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), A}) \right) \\
& - 2 \frac{1}{NT} \frac{1}{T-k^0+1} \sum_{i=1}^N \left(\sum_{s=k}^T u_{is} \right) \left(\sum_{s=k^0}^{k-1} x'_{is} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), A}) \right) \\
& + \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} (x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), A})) \\
& + \frac{2}{NT(k-1)} \sum_{i=1}^N \left(\sum_{s=1}^{k^0-1} u_{is} \right) \left(\sum_{s=k^0}^{k-1} u_{is} \right) + \frac{1}{NT(k-1)} \sum_{i=1}^N \left(\sum_{s=k^0}^{k-1} u_{is} \right)^2 \\
& - \frac{k-k^0}{NT(k^0-1)(k-1)} \sum_{i=1}^N \left(\sum_{s=1}^{k^0-1} u_{is} \right)^2 - \frac{k-k^0}{NT(T-k^0+1)(T-k+1)} \sum_{i=1}^N \left(\sum_{s=k}^T u_{is} \right)^2 \\
& + 2 \frac{1}{NT(T-k^0+1)} \sum_{i=1}^N \left(\sum_{s=k^0}^{k-1} u_{is} \right) \left(\sum_{s=k}^T u_{is} \right) - \frac{1}{NT(T-k^0+1)} \sum_{i=1}^N \left(\sum_{s=k^0}^{k-1} u_{is} \right)^2.
\end{aligned}$$

By the argument based on inequality (1.8) in [Merlevède et al. \(2011\)](#), it holds that uniformly over i we have for constant $M > 0$

$$\Pr \left(\frac{1}{k^0-1} \left| \sum_{t=1}^{k^0-1} u_{it} \right| > MT^{-1/4} \right) = o(T^{-\delta}),$$

and for uniformly over $k^0 < k < \bar{k}$

$$\Pr\left(\frac{1}{k - k^0} \left| \sum_{t=k^0}^{k-1} u_{it} \right| > MT^\nu\right) = o(T^{-\delta}),$$

and

$$\Pr\left(\frac{1}{T - k - 1} \left| \sum_{t=k}^T u_{it} \right| > MT^{-1/4}\right) = o(T^{-\delta}),$$

where $\epsilon > 0$ and $\eta T^\epsilon < CT^{-1/2}$. Similarly, noting that $\sum_{t=a}^b E(x_{it})/(b-a)$ converges as $b-a \rightarrow \infty$, we have

$$\Pr\left(\frac{1}{k^0 - 1} \left| \sum_{t=1}^{k^0-1} x_{it} \right| > MT^\nu\right) = o(T^{-\delta}),$$

and for uniformly over $k^0 < k < \bar{k}$

$$\Pr\left(\frac{1}{k - k^0} \left| \sum_{t=k^0}^{k-1} x_{it} \right| > MT^\nu\right) = o(T^{-\delta}),$$

and

$$\Pr\left(\frac{1}{T - k - 1} \left| \sum_{t=k}^T x_{it} \right| > MT^\nu\right) = o(T^{-\delta}).$$

By $\hat{\beta} \in \mathcal{N}$, we have

$$\left\| \beta_{g_i^0(A), A}^0 - \beta_{g_i^0(A), A} \right\| < \eta$$

and

$$\left\| \beta_{g_i^0(B), A}^0 - \beta_{g_i^0(B), B} \right\| < \eta.$$

Thus we have with probability at least $1 - O(NT^\delta)$,

$$\begin{aligned} & \left| (Q(k^0, \gamma^0, \beta) - Q(k, \gamma^0, \beta)) \right. \\ & \quad - \frac{2}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} x'_{it} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) \\ & \quad + \frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it, k, B} (\beta_{g_i^0(A), A}^0 - \beta_{g_i^0(B), B}^0) \right)^2 \Big| \\ & < \eta M T^\epsilon + M T^{-1/2}. \end{aligned}$$

Assumption 3(vi) implies that

$$\frac{1}{NT} \sum_{t=k^0}^{k-1} \sum_{i=1}^N \left(x'_{it,k,B} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right)^2 \geq \frac{k-k^0}{T} \underline{m}.$$

Thus we have

$$\begin{aligned} & \Pr \left(Q(k^0, \gamma^0, \beta) > \min_{k^0 < k \leq \bar{k}} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right) \\ &= \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{k^0 < k \leq \bar{k}} (Q(k^0, \gamma^0, \beta) - Q(k, \gamma^0, \beta)) > 0 \right) \\ &\leq \Pr \left(\max_{k \leq k < k^0} \left(-2 \frac{1}{N} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} x'_{it} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) - \frac{k^0 - k}{2T} m \right) > 0 \right) + O(NT^{-\delta}) \\ &= \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{\underline{k} \leq k < k^0} \left(-2 \frac{1}{N} \frac{1}{k - k^0} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} x'_{it} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) - \frac{m}{2} \right) > 0 \right) + O(NT^{-\delta}) \\ &\leq \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{\underline{k} \leq k < k^0} \left(-2 \frac{1}{N} \frac{1}{k - k^0} \sum_{t=k^0}^{k-1} \sum_{i=1}^N u_{it} x'_{it} (\beta_{g_i^0(A),A}^0 - \beta_{g_i^0(B),B}^0) \right) > \frac{m}{2} \right) + O(NT^{-\delta}) \\ &= O \left(\frac{1}{N} \right) + O(NT^{-\delta}), \end{aligned}$$

where the last equality follows by applying (Bai and Perron, 1998, Lemma A.6) which is an extension of Hájek and Rényi (1955). Here we use the observation that an L_r mixing sequence is an L_p mixingale sequence for $1 \leq p < r$ as discussed in (Davidson, 1994, page 248). Thus, under Assumptions 3(ix) and 3(x), $u_{it} x_{it}$ is an L^2 mixingale and we can apply (Bai and Perron, 1998, Lemma A.6).

A similar argument shows that

$$\Pr \left(Q(k^0, \gamma^0, \beta) > \min_{k^0 < k < \bar{k}} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right) = O \left(\frac{1}{N} \right) + O(NT^{-\delta}).$$

To sum up, we have

$$\Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) \rightarrow 0.$$

□

S.3.4.3 Proof of Corollary S.2

Proof. The proof is exactly identical to that of first-difference estimation, and thus omitted.

□

S.4 Simultaneous estimation of multiple breaks

In this section, we provide theoretical justification of simultaneous detection of multiple breaks. Let $j = 1, \dots, m$, such that k_j denotes the j -th break, and $\beta_{g_i(j),j}$, γ_j , and \mathbb{G}^j denote the slope coefficient, group membership parameter, and the set of groups, respectively, in the regime between k_{j-1} and k_j . Let m denote the number of breaks. The true break points are k_1^0, \dots, k_m^0 . The coefficient vector can be written as:

$$\beta_{git,t} = \begin{cases} \beta_{g_i(1),1} & \text{if } t < k_1^0, \\ \beta_{g_i(2),2} & \text{if } k_1^0 \leq t < k_2^0, \\ \vdots & \\ \beta_{g_i(m+1),m+1} & \text{if } t \geq k_m^0. \end{cases}$$

We note that there are $m + 1$ regimes when the number of break points is m . Let $k_0 = 1$ and $k_{m+1} = T$. Let $k = (k_1, \dots, k_m)$.

We estimate (k, γ, β) simultaneously by minimizing the square loss function:

$$(\hat{k}, \hat{\gamma}, \hat{\beta}) = \arg \min_{k \in \mathbb{K}, \gamma \in \mathbb{G}, \beta \in \mathbb{B}} Q(k, \gamma, \beta),$$

where

$$Q(k, \gamma, \beta) = \frac{1}{NT} \sum_{j=1}^{m+1} \sum_{t=k_{j-1}}^{k_j-1} \sum_{i=1}^N (y_{it} - x'_{it} \beta_{g_i(j),j})^2.$$

S.4.1 Assumptions

Here is the assumption used to justify our procedure.

Assumption 4.

(i) For any $L \subseteq \{1, \dots, N\}$ and $t'' \geq t'$, there exists M which does not depend on L , t'' nor t' such that the following equality holds

$$E \left(\left\| \frac{1}{NT} \sum_{t=t'}^{t''} \sum_{i \in L} x_{it} u_{it} \right\|^2 \right) \leq M \frac{|L|(t'' - t')}{N^2 T^2},$$

where $|L|$ denotes the cardinality of L .

(ii) \mathcal{B} is compact.

(iii) Let $\rho_{N,t}(\gamma^t, g, \tilde{g})$ be the minimum eigenvalue of $\sum_{i=1}^N \mathbf{1}\{g_{it}^0 = g\} \{g_{it} = \tilde{g}\} x_{it} x'_{it} / N$, where γ^t is γ_j for $k_{j-1}^0 \leq t < k_j^0$ where $j \in \{1, \dots, m+1\}$. For and $j \in \{1, \dots, m+1\}$

and $g \in \mathbb{G}^j$,

$$\min_{k_{j-1}^0 \leq t < k_j^0} \min_{\gamma_j \in (\mathbb{G}^j)^N} \max_{\tilde{g} \in \mathbb{G}^j} \rho_{N,t}(\gamma_j, g, \tilde{g}) > \hat{\rho},$$

where $\hat{\rho} \rightarrow_p \rho$ as $N, T \rightarrow \infty$ and $\rho > 0$ does not depend on N and g .

(iv) There exists $\hat{\rho}^*$ such that for any i and for s such that s and $T-s$ sufficiently large,

$$\lambda_{\min} \left(\frac{1}{s} \sum_{t=1}^s x_{it} x'_{it} \right) \geq \hat{\rho}^* \quad \text{and} \quad \lambda_{\min} \left(\frac{1}{T-s} \sum_{t=s+1}^T x_{it} x'_{it} \right) \geq \hat{\rho}^*,$$

and $\hat{\rho}^* \rightarrow_p \rho^* > 0$ as $N, T \rightarrow \infty$, where λ_{\min} gives the minimum eigenvalue of its argument.

(v) $\max_{1 \leq t \leq T} \sum_{i=1}^N \|x_{it}\|^2 / N = O_p(1)$.

(vi) There exists a fixed constant $\underline{m} > 0$ (which, in particular, does not depend on T and N) such that, with probability approaching one, for any t and $j \in \{1, \dots, m+1\}$,

$$\frac{1}{N} \sum_{i=1}^N (x'_{it} (\beta_{g_i^0(j), j}^0 - \beta_{g_i^0(j-1), j-1}^0))^2 > \underline{m}.$$

(vii) There exists $\epsilon > 0$ such that $(k_j^0 - k_{j-1}^0)/T \rightarrow \tau_j > \epsilon$ as $T \rightarrow \infty$ for any $j \in \{1, \dots, m+1\}$.

(viii) There exists a constant $c > 0$ such that for any $j \in \{1, \dots, m+1\}$ and any $g \neq \tilde{g}$ where $g, \tilde{g} \in \mathbb{G}^j$, it holds that $\|\beta_{g,j}^0 - \beta_{\tilde{g},j}^0\| > c$.

(ix) Let z_{it} be $x'_{it} x_{it}$, $\|u_{it} x_{it}\|$, $2u_{it} x'_{it} (\beta_{g_i^0(j), j}^0 - \beta_{g,j}^0)$, or $(x'_{it} (\beta_{g_i^0(j), j}^0 - \beta_{g,j}^0))^2$ for $g \in \mathbb{G}^j$ and $j \in \{1, \dots, m+1\}$. Assume the following holds for any choice of z_{it} : 1) z_{it} is a strong mixing sequence over t whose mixing coefficients $a_i[t]$ are bounded by $a[t] \leq e^{-at^{d_1}}$ such that $\max_{1 \leq i \leq N} a_i[t] \leq a[t]$ and has tail probabilities $\max_{1 \leq i \leq N} \Pr(|z_{it}| > z) \leq e^{1-(z/b)^{d_2}}$ for any t , where a, b, d_1 and d_2 are positive constants. 2) There exists $a_i, i = 1, \dots, N$ such that for any $\epsilon > 0$, it holds that $\max_{1 \leq i \leq N} |a_i - \sum_{t=1}^T E(z_{it})/T| < \epsilon$ for sufficiently large T .

(x) $\max_{1 \leq t \leq T} E(\|\sum_{i=1}^N x_{it} u_{it} / \sqrt{N}\|^{2+\delta})$ is bounded for some $\delta > 0$.

The assumption is similar to that for situations with single break. The main difference is that we need to consider $m+1$ regimes instead of two regimes.

S.4.2 Theoretical results

The following theorem shows that our procedure can estimate the set of breaks points consistently.

Theorem S.3. *Suppose that Assumption 4 holds. As $N, T \rightarrow \infty$ with $NT^{-\delta} \rightarrow 0$ for some $\delta > 0$, $\Pr(\hat{k}_j = k_j^0 \text{ for all } j \in \{1, \dots, m\}) \rightarrow 1$.*

As a corollary, the group membership structures for all regimes are estimated consistently and the coefficient estimator vector has the same asymptotic distribution as that under cases in which the break points and the group membership structure are known.

Corollary S.3. *Suppose that Assumption 4 holds. As $N, T \rightarrow \infty$ with $NT^{-\delta} \rightarrow 0$ for some $\delta > 0$,*

- (1) $\Pr(\hat{\gamma}_j = \gamma_j^0 \text{ for all } j \in \{1, \dots, m+1\}) \rightarrow 1$,
- (2) $\hat{\beta} = \tilde{\beta} + o_p(1/\sqrt{NT})$, where $\tilde{\beta}$ is the estimator of β under $k_j = k_j^0$ for all $j \in \{1, \dots, m+1\}$ and $\gamma_j = \gamma_j^0$ for all $j \in \{1, \dots, m+1\}$.

We note that the $o_p(1/\sqrt{NT})$ in Corollary S.3 (2) is uniform over regimes and groups.

S.4.3 Lemmas

Let

$$\tilde{Q}(k, \gamma, \beta) = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (x'_{it} (\beta_{g_{it}, t}^0 - \beta_{g_{it}, t}))^2 + \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N u_{it}^2.$$

Lemma S.10. *Suppose that Assumptions 4(i) and 4(ii) hold. Then we have that*

$$\sup_{k \in \mathbb{K}, \gamma \in \mathbb{G}, \beta \in \mathbb{B}} |\tilde{Q}(k, \gamma, \beta) - Q(k, \gamma, \beta)| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Proof. Note that $\beta_{g_{it}, t} = \beta_{g_i(j), j}$ for $k_{j-1} \leq t < k_j$. Then we have

$$\begin{aligned} Q(k, \gamma, \beta) &= \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (y_{it} - x'_{it} \beta_{i,t})^2 \\ &= \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N (x'_{it} (\beta_{g_{it}, t}^0 - \beta_{g_{it}, t}))^2 \\ &\quad + 2 \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N x'_{it} (\beta_{g_{it}, t}^0 - \beta_{g_{it}, t}) u_{it} + \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N u_{it}^2. \end{aligned}$$

Thus, we have

$$\begin{aligned}\tilde{Q}(k, \gamma, \beta) - Q(k, \gamma, \beta) &= -2 \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N x'_{it} (\beta_{g_{it}^0, t}^0 - \beta_{g_{it}, t}) u_{it} \\ &= -2 \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N x'_{it} \beta_{g_{it}^0, t}^0 u_{it} + 2 \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N x'_{it} \beta_{g_{it}, t} u_{it}.\end{aligned}$$

We examine the two terms on the right hand side separately. First we have

$$\frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N x'_{it} \beta_{g_{it}^0, t}^0 u_{it} = \frac{1}{NT} \sum_{j=1}^{m+1} \sum_{t=k_j^0}^{k_j^0-1} \sum_{g \in \mathbb{G}^j} \sum_{i=1}^N \mathbf{1}(g_i^0(j) = g) x'_{it} \beta_{g, j}^0 u_{it}.$$

For each g and j , by the Cauchy-Schwarz inequality we have that

$$\begin{aligned}E \left(\left(\frac{1}{NT} \sum_{t=k_{j-1}^0}^{k_j^0-1} \sum_{i=1}^N \mathbf{1}(g_i^0(j) = g) x'_{it} \beta_{g, j}^0 u_{it} \right)^2 \right) \\ \leq CE \left(\left\| \frac{1}{NT} \sum_{t=k_{j-1}^0}^{k_j^0-1} \sum_{g_i^0(j)=g} x_{it} u_{it} \right\|^2 \right) = O \left(\frac{k_j^0 - k_{j-1}^0}{NT^2} \right),\end{aligned}$$

where C satisfies $\|\beta_{g_{it}, t}\|^2 < C$ for any $\beta \in \mathcal{B}$ and the existence of such C is guaranteed by Assumption 4(ii), and the equality follows by Assumption 4(i). Next we consider

$$\frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N x'_{it} \beta_{g_{it}, t} u_{it} = \frac{1}{NT} \sum_{j=1}^{m+1} \sum_{t=k_{j-1}}^{k_j} \sum_{i=1}^N x'_{it} \beta_{g_i(j), j} u_{it}.$$

For each j , we have

$$\begin{aligned}\left(\frac{1}{NT} \sum_{t=k_{j-1}}^{k_j} \sum_{i=1}^N x'_{it} \beta_{g_i(j), j} u_{it} \right)^2 &= \left(\frac{1}{NT} \sum_{i=1}^N \beta'_{g_i(j), j} \sum_{t=k_{j-1}}^{k_j-1} x_{it} u_{it} \right)^2 \\ &\leq \left(\frac{1}{N} \sum_{i=1}^N \|\beta_{g_i(j), j}\|^2 \right) \left(\frac{1}{NT^2} \sum_{i=1}^N \left\| \sum_{t=k_{j-1}}^{k_j-1} x_{it} u_{it} \right\|^2 \right) \\ &= O_p \left(\frac{k_j - k_{j-1}}{T^2} \right),\end{aligned}$$

where the first inequality uses the Cauchy-Schwarz inequality and the second inequality follows by that Assumption 4(ii) implies $\sum_{i=1}^N \|\beta_{g_i(j), j}\|^2 / N < C$ for some C , and that Assumption 4(i) together with the Markov inequality implies $\sum_{i=1}^N \left\| \sum_{t=k_{j-1}}^{k_j} x_{it} u_{it} \right\|^2 / (NT^2) = O_p((k_j - k_{j-1})/T^2)$.

It therefore holds that

$$\tilde{Q}(k, \gamma, \beta) - Q(k, \gamma, \beta) = O\left(\frac{1}{\sqrt{NT}}\right) + O\left(\frac{1}{\sqrt{T}}\right) = O\left(\frac{1}{\sqrt{T}}\right)$$

uniformly over β and γ . □

Lemma S.11. Suppose that Assumptions 4(i)–4(vii) hold. Then we have that

- (1) $\max_{g \in \mathbb{G}^j} \min_{\tilde{g} \in \mathbb{G}^j} \left\| \beta_{g,j}^0 - \hat{\beta}_{\tilde{g},j} \right\|^2 = O_p(1/\sqrt{T})$ for each $j \in \{1, \dots, m+1\}$,
- (2) $(\hat{k}_j - k_j^0)/T = O_p(1/\sqrt{T})$ for each $j \in \{1, \dots, m\}$.

Proof. From Lemma S.10, we have that

$$\begin{aligned} \tilde{Q}(\hat{k}, \hat{\gamma}, \hat{\beta}) &= Q(\hat{k}, \hat{\gamma}, \hat{\beta}) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &\leq Q(k^0, \gamma^0, \beta^0) + O_p\left(\frac{1}{\sqrt{T}}\right) = \tilde{Q}(k^0, \gamma^0, \beta^0) + O_p\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

Because $\tilde{Q}(k, \gamma, \beta)$ is minimized at (k^0, γ^0, β^0) , we have that

$$\tilde{Q}(\hat{k}, \hat{\gamma}, \hat{\beta}) - \tilde{Q}(k^0, \gamma^0, \beta^0) = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Let $a_{NT} = \tilde{Q}(\hat{k}, \hat{\gamma}, \hat{\beta}) - \tilde{Q}(k^0, \gamma^0, \beta^0)$, and note that $a_{NT} = O_p(1/\sqrt{T})$.

We consider regimes which are separated by either a true break point or an estimated break point. Let $L = \hat{k} \cup k^0 = \{\hat{k}_1, \dots, \hat{k}_m\} \cup \{k_1^0, \dots, k_m^0\}$. Let k_l^* be the l -th smallest element of L and define $k_0^* = 1$ and $k_{L^*}^* = T$ where L^* is the cardinality of L . We now define functions which indicate how regimes defined by L are related to true regimes and estimated regimes. For each l , we define $\hat{j}(l)$ such that $[k_{l-1}^*, k_l^*] \subset [\hat{k}_{\hat{j}(l)-1}, \hat{k}_{\hat{j}(l)}]$ and also define $j^0(j)$ such that $[k_{j^0(l)-1}^*, k_j^*] \subset [k_{j^0(l)-1}^0, k_{j^0(l)}^0]$. Function \hat{j} indicates the estimated regime containing regime l , and function j^0 indicates the true regime containing regime l . We have

$$\tilde{Q}(\hat{k}, \hat{\gamma}, \hat{\beta}) - \tilde{Q}(k^0, \gamma^0, \beta^0) = \frac{1}{NT} \sum_{l=1}^{L^*} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it} (\beta_{g_i^0(j^0(l)), j^0(l)}^0 - \hat{\beta}_{g_i(\hat{j}(l)), \hat{j}(l)}))^2 \quad (\text{S.6})$$

We first consider time periods $k_{l-1}^* \leq t < k_l^*$ for some $l \in \{1, \dots, L^*\}$ where for some j^\dagger , $\hat{j}(l) = j^0(l) = j^\dagger$. In this regime, the true and estimated regime numbers are equal. For such l , we have

$$\frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it} (\beta_{g_i^0(j^\dagger), j^\dagger}^0 - \hat{\beta}_{g_i(j^\dagger), j^\dagger}))^2$$

$$\begin{aligned}
&= \frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{g=1}^{G^{j^\dagger}} \sum_{\tilde{g}=1}^{G^{j^\dagger}} \sum_{i=1}^N \mathbf{1}\{g_i^0(j^\dagger) = g\} \{g_i(j^\dagger) = \tilde{g}\} (x'_{it}(\beta_{g,j^\dagger}^0 - \hat{\beta}_{\tilde{g},j^\dagger}))^2 \\
&\geq \frac{1}{T} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{g=1}^{G^{j^\dagger}} \sum_{\tilde{g}=1}^{G^{j^\dagger}} \sum_{i=1}^N \mathbf{1}\{g_i^0(j^\dagger) = g\} \{g_i(j^\dagger) = \tilde{g}\} \rho_{N,t}(\gamma, g, \tilde{g}) \left\| \beta_{g,j^\dagger}^0 - \hat{\beta}_{\tilde{g},j^\dagger} \right\|^2 \\
&\geq \frac{k_l^* - k_{l-1}^*}{T} \hat{\rho} \max_{g \in \mathbb{G}^B} \min_{\tilde{g} \in \mathbb{G}^B} \left\| \beta_{g,j^\dagger}^0 - \hat{\beta}_{\tilde{g},j^\dagger} \right\|^2,
\end{aligned}$$

where the last inequality follows by Assumption 4(iv). This further implies

$$\frac{k_l^* - k_{l-1}^*}{T} \hat{\rho} \max_{g \in \mathbb{G}^{j^\dagger}} \min_{\tilde{g} \in \mathbb{G}^{j^\dagger}} \left\| \beta_{g,j^\dagger}^0 - \hat{\beta}_{\tilde{g},j^\dagger} \right\|^2 \leq a_{NT}. \quad (\text{S.7})$$

Moreover, Assumption 4(iv) implies that

$$\frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j^\dagger),j^\dagger}^0 - \hat{\beta}_{g_i(j^\dagger),j^\dagger}))^2 \geq \frac{k_l^* - k_{l-1}^*}{T} \hat{\rho}^* \frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(j^\dagger),j^\dagger}^0 - \hat{\beta}_{g_i(j^\dagger),j^\dagger} \right\|^2.$$

Thus we have

$$\frac{1}{N} \sum_{i=1}^N \left\| \beta_{g_i^0(j^\dagger),j^\dagger}^0 - \hat{\beta}_{g_i(j^\dagger),j^\dagger} \right\|^2 < C a_{NT}. \quad (\text{S.8})$$

Next we consider time periods $k_{l-1}^* \leq t < k_l^*$ for some $l \in \{1, \dots, L^*\}$ such there is no j^\dagger satisfying $\hat{j}(l) = j^0(l) = j^\dagger$. Also suppose that there exists \tilde{l} such that $\hat{j}(l) = \hat{j}(\tilde{l}) = j^0(\tilde{l})$. For such l ,

$$\begin{aligned}
&\frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j^0(l)),j^0(l)}^0 - \hat{\beta}_{\hat{g}_i(\hat{j}(l)),\hat{j}(l)})^2 \\
&= \frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j^0(l)),j^0(l)}^0 - \beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0 + \beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0 - \hat{\beta}_{\hat{g}_i(\hat{j}(l)),\hat{j}(l)})^2 \\
&\geq \frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j^0(l)),j^0(l)}^0 - \beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0))^2 + \frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0 - \hat{\beta}_{\hat{g}_i(\hat{j}(l)),\hat{j}(l)})^2 \\
&\quad - 2 \frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N \left| x'_{it}(\beta_{g_i^0(j^0(l)),j^0(l)}^0 - \beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0) \right| \cdot \left| x'_{it}(\beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0 - \hat{\beta}_{\hat{g}_i(\hat{j}(l)),\hat{j}(l)}) \right| \\
&\geq \frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j^0(l)),j^0(l)}^0 - \beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0))^2 + \frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0 - \hat{\beta}_{\hat{g}_i(\hat{j}(l)),\hat{j}(l)})^2 \\
&\quad - 2 \frac{1}{T} \sum_{t=k_{l-1}^*}^{k_l^*-1} \left(\frac{1}{N} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j^0(l)),j^0(l)}^0 - \beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0))^2 \right)^{1/2} \left(\frac{1}{N} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0 - \hat{\beta}_{\hat{g}_i(\hat{j}(l)),\hat{j}(l)})^2 \right)^{1/2}.
\end{aligned}$$

By Assumption 4(vi), it holds that

$$\frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j^0(l)),j^0(l)}^0 - \beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0))^2 \geq \frac{k_j^* - k_{j-1}^*}{T} \underline{m}.$$

Assumptions 4(ii) and 4(v) imply that $\sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j^0(l)),j^0(l)}^0 - \beta_{g_i^0(\hat{j}(l)),\hat{j}(l)}^0))^2/N < C$. Noting that we assume that there exists \tilde{l} such that $\hat{j}(l) = \hat{j}(\tilde{l}) = j^0(\tilde{l})$. Then, (S.8) can be used and combining it with Assumption 4(v), we have $\frac{1}{N} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j^0(\tilde{l})),j^0(\tilde{l})}^0 - \hat{\beta}_{\hat{g}_i(j^0(\tilde{l}),j^0(\tilde{l}))}^0))^2 < Ca_{NT}$. Hence, if there exists \tilde{l} such that $\hat{j}(l) = \hat{j}(\tilde{l}) = j^0(\tilde{l})$, we have

$$\frac{1}{NT} \sum_{t=k_{l-1}^*}^{k_l^*-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j^0(l)),j^0(l)}^0 - \hat{\beta}_{\hat{g}_i(\hat{j}(l)),\hat{j}(l)}^0))^2 \geq \frac{k_l^* - k_{l-1}^*}{T} (\underline{m} - C\sqrt{a_{NT}}),$$

which further implies that

$$\frac{k_l^* - k_{l-1}^*}{T} (\underline{m} - C\sqrt{a_{NT}}) < a_{NT}. \quad (\text{S.9})$$

To summarize, for $l \in \{1, \dots, L^*\}$ such that there exists j^\dagger with $\hat{j}(l) = j^0(l) = j^\dagger$, we have (S.7), and for l such that there is no j^\dagger satisfying $\hat{j}(l) = j^0(l) = j^\dagger$ but there exists \tilde{l} such that $\hat{j}(l) = \hat{j}(\tilde{l}) = j^0(\tilde{l})$, we have (S.9).

We now repeatedly apply (S.7) and (S.9) to obtain the desired result. First we observe that there exists $l \in \{1, \dots, L^*\}$ such that j^\dagger with $\hat{j}(l) = j^0(l) = j^\dagger$, $k_{l-1}^* = k_{j^\dagger-1}$ and $k_l^* = k_{j^\dagger}$. This claim can be shown in the following way. If $\hat{k}_1 \geq k_1^0$, then $j^\dagger = 1$ satisfies the claim. Next, we consider cases with $\hat{k}_1 < k_1^0$. Then, if $\hat{k}_2 \geq k_2^0$, then the claim holds for $j^\dagger = 2$. We thus needs to consider cases with $\hat{k}_2 < k_2^0$. This argument can be repeated until we take $\hat{k}_m < k_m^0$. Then $j^\dagger = m+1$ satisfies the claim. This argument shows that there exists $l \in L$ such that j^\dagger with $\hat{j}(l) = j^0(l) = j^\dagger$, $k_{l-1}^* = k_{j^\dagger-1}$ and $k_l^* = k_{j^\dagger}$. For such l , by (S.7), and Assumption 4(vii), we have $\max_{g \in \mathbb{G}^{j^\dagger}} \min_{\tilde{g} \in \mathbb{G}^{j^\dagger}} \left\| \beta_{g,j^\dagger}^0 - \hat{\beta}_{\tilde{g},j^\dagger} \right\|^2 \leq a_{NT}/(\hat{\rho}\epsilon)$. By (S.9) we obtain $(\hat{k}_{j^\dagger-1} - k_{j^\dagger-1}^0)/T < a_{NT}/(\underline{m} - C\sqrt{a_{NT}})$, and $(\hat{k}_{j^\dagger} - k_{j^\dagger}^0)/T < a_{NT}/(\underline{m} - C\sqrt{a_{NT}})$ (note that this holds even when $j^\dagger = 1$ or $j^\dagger = m+1$ because $\hat{k}_0 = k_0^0 = 1$ and $\hat{k}_{m+1} = k_{m+1}^0 = T$). We then restrict our attention to $\{1, \dots, \hat{k}_{j^\dagger-1}\} \cup \{\hat{k}_{j^\dagger}, \dots, T\}$, and repeat the above argument, which yields the claim of the lemma. □

Lemma S.12. Suppose that Assumptions 4(i)-4(viii) are satisfied. Then there exist permutations $\sigma_j : \mathbb{G}^j \mapsto \mathbb{G}^j$ such that $\left\| \beta_{g,j}^0 - \hat{\beta}_{\sigma_j(g),j} \right\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}^j$.

Proof. The proof is exactly identical to that of Lemma 3 in the main text and is thus omitted. □

Define the Hausdorff distance between β_j^0 and $\hat{\beta}_j$ to be

$$\max \left(\max_{g \in \mathbb{G}^j} \min_{\tilde{g} \in \mathbb{G}^j} \left\| \beta_{g,j}^0 - \hat{\beta}_{\tilde{g},j} \right\|^2, \max_{\tilde{g} \in \mathbb{G}^j} \min_{g \in \mathbb{G}^j} \left\| \beta_{g,j}^0 - \hat{\beta}_{\tilde{g},j} \right\|^2 \right).$$

By Lemmas S.11 and S.12, this Hausdorff distance converges to 0 at the rate of \sqrt{T} . Using the similar arguments, we can show that $\hat{\beta}$ is consistent under the Hausdorff distance and its rate of convergence is \sqrt{T} . By relabeling, we can set $\sigma_j(g) = g$, the convention that we adopt throughout the paper, such that $\left\| \beta_{g,j}^0 - \hat{\beta}_{g,j} \right\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}^j$.

Let \mathcal{N} be a neighborhood of β^0 such that $\left\| \beta_{g,j}^0 - \beta_{g,j} \right\| < \eta$ for $\eta > 0$ for any $g \in \mathbb{G}^j$ and $j = \{1, \dots, m+1\}$. Note that we will take η small enough by considering large N and T by Lemma S.12. Let $\bar{k}_j = \sqrt{T} \log T + k_j^0$ and $\underline{k}_j = -\sqrt{T} \log T + k_j^0$. Define $K = \{k : \underline{k}_j \leq k_j \leq \bar{k}_j, \text{ for each } j\}$.

Lemma S.13. *Suppose that Assumptions 4(ii), 4(iv), 4(vii), 4(viii), and 4(ix) hold. As $N, T \rightarrow \infty$ with $NT^{-\delta} \rightarrow 0$, it holds that*

$$\Pr \{ \hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N} \} \rightarrow 0.$$

Proof. To show this probability converges to zero, it is equivalent to show that

$$\sum_{j=1}^{m+1} \max_{1 \leq i \leq N} \sup_{\beta \in \mathcal{N}} \max_{k \in K} \mathbf{1}\{\hat{g}_i(j)(k, \beta) \neq g_i^0(j)\} = o_p(1),$$

where we observe that

$$\mathbf{1}\{\hat{g}_i(j)(k, \beta) \neq g_i^0(j)\} = \max_{g \in \mathbb{G}^j \setminus \{g_i^0(j)\}} \mathbf{1} \left(\sum_{t=k_{j-1}}^{k_j-1} (y_{it} - x'_{it} \beta_{g,j})^2 < \sum_{t=k_{j-1}}^{k_j-1} (y_{it} - x'_{it} \beta_{g_i^0(j),j})^2 \right). \quad (\text{S.10})$$

We analyze the probability of each of these two indicators being one. To this end, we first evaluate how the deviation of k from k^0 plays a role, while the situation of $k = k^0$ can be analysed using the same arguments as in Bonhomme and Manresa (2015) and Okui and Wang (2021).

We first examine the difference between the two summations in the argument of the indicator function in (S.10), and show that this difference evaluated at any $k \in K$ and that evaluated at $k = k^0$ are not very different. Let

$$D_j = \sum_{t=k_{j-1}}^{k_j-1} \left((y_{it} - x'_{it} \beta_{g,j})^2 - (y_{it} - x'_{it} \beta_{g_i^0(j),j})^2 \right) - \sum_{t=k_{j-1}^0}^{k_j^0-1} \left((y_{it} - x'_{it} \beta_{g,j})^2 - (y_{it} - x'_{it} \beta_{g_i^0(j),j})^2 \right).$$

First, considering the case of $k_{j-1} < k_{j-1}^0$ and $k_j < k_j^0$, we have that

$$\begin{aligned}
|D| &\leq \left| \sum_{t=k_{j-1}}^{k_{j-1}^0-1} 2u_{it}x_{it}(\beta_{g_i^0(j),j} - \beta_{g,j}) \right| + \left| \sum_{t=k_{j-1}}^{k_{j-1}^0-1} (\beta_{g_i^0(j),j} - \beta_{g,j})'x_{it}x_{it}'(2\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j),j} - \beta_{g,j}) \right| \\
&\quad + \left| \sum_{t=k_j}^{k_j^0-1} 2u_{it}x_{it}(\beta_{g_i^0(j),j} - \beta_{g,j}) \right| + \left| \sum_{t=k_j}^{k_j^0-1} (\beta_{g_i^0(j),j} - \beta_{g,j})'x_{it}x_{it}'(2\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j),j} - \beta_{g,j}) \right| \\
&\leq M_1 \left\| \sum_{t=k_{j-1}}^{k_{j-1}^0-1} u_{it}x_{it} \right\| + M_2 \left\| \sum_{t=k_{j-1}}^{k_{j-1}^0-1} x_{it}x_{it}' \right\| + M_3 \left\| \sum_{t=k_j}^{k_j^0-1} u_{it}x_{it} \right\| + M_4 \left\| \sum_{t=k_j}^{k_j^0-1} x_{it}x_{it}' \right\| \\
&\leq M_1(k_{j-1}^0 - \underline{k}_{j-1}) \frac{1}{k_{j-1}^0 - \underline{k}_{j-1}} \sum_{t=\underline{k}_{j-1}}^{k_{j-1}^0-1} \|u_{it}x_{it}\| + M_2(k_{j-1}^0 - \underline{k}_{j-1}) \left\| \frac{1}{k_{j-1}^0 - \underline{k}_{j-1}} \sum_{t=\underline{k}_{j-1}}^{k_{j-1}^0-1} x_{it}x_{it}' \right\| \\
&\quad + M_3(k_j^0 - \underline{k}_j) \frac{1}{k_j^0 - \underline{k}_j} \sum_{t=\underline{k}_j}^{k_j^0-1} \|u_{it}x_{it}\| + M_4(k_j^0 - \underline{k}_j) \left\| \frac{1}{k_j^0 - \underline{k}_j} \sum_{t=\underline{k}_j}^{k_j^0-1} x_{it}x_{it}' \right\|,
\end{aligned}$$

where M_1, M_2, M_3 and M_4 are constants independent of (i, g, k, β) . Let $M_T = T^{1/4}/\log T$. Under Assumption 4(ix), we can apply inequality (1.8) in Merlevède et al. (2011) which is based on Theorem 6.2 of Rio (2017), translated from a French version published in 2000, with $\lambda = (k_{j-1}^0 - \underline{k}_{j-1})M_T = T^{3/4}$ and obtain that

$$\begin{aligned}
&\Pr \left(\frac{1}{k_{j-1}^0 - \underline{k}_{j-1}} \left| \sum_{t=\underline{k}_{j-1}}^{k_{j-1}^0-1} (\|u_{it}x_{it}\| - E(\|u_{it}x_{it}\|)) \right| > M_T \right) \\
&\leq 4 \exp \left(-\frac{\lambda^{d/(d+1)} \log 2}{2} \right) + 16CM_T^{-1} \exp \left(-a \frac{\lambda^{d/(d+1)}}{b^d} \right) = o(T^{-\delta}),
\end{aligned}$$

where $d = d_1d_2/(d_1 + d_2)$, (a, b, d_1, d_2) are defined in Assumption 4(ix). Noting that $(k_{j-1}^0 - \underline{k}_{j-1})^{-1} \sum_{t=\underline{k}_{j-1}}^{k_{j-1}^0-1} E(\|u_{it}x_{it}\|)$ converges and $M_T \rightarrow \infty$, we have that

$$\Pr \left(\frac{1}{k_{j-1}^0 - \underline{k}_{j-1}} \sum_{t=\underline{k}_{j-1}}^{k_{j-1}^0-1} \|u_{it}x_{it}\| > M_T \right) = o(T^{-\delta}).$$

Similarly, it holds that

$$\begin{aligned}
&\Pr \left(\left\| \frac{1}{k_{j-1}^0 - \underline{k}_{j-1}} \sum_{t=\underline{k}_{j-1}}^{k_{j-1}^0-1} x_{it}x_{it}' \right\| > M_T \right) = o(T^{-\delta}) \\
&\Pr \left(\frac{1}{k_j^0 - \underline{k}_j} \sum_{t=\underline{k}_j}^{k_j^0-1} \|u_{it}x_{it}\| > M_T \right) = o(T^{-\delta})
\end{aligned}$$

$$\text{and } \Pr \left(\left\| \frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}}^{k_j^0-1} x_{it} x'_{it} \right\| > M_T \right) = o(T^{-\delta}).$$

These imply that there exists a sequence that satisfies $C_T = O(M_T)$ and $C_T \rightarrow \infty$ as $T \rightarrow \infty$, such that

$$\Pr \left(\frac{1}{k_j^0 - k_{j-1}^0} |D| > \frac{k_j^0 - k_j + k_{j-1}^0 - k_{j-1}}{k_j^0 - k_{j-1}^0} C_T \right) = o(T^{-\delta}).$$

Using a similar argument, we can show that for the other three cases ($k_{j-1} < k_j^0, k_j \geq k_j^0$, $k_{j-1} \geq k_j^0, k_j < k_j^0$ and $k_{j-1} \geq k_j^0, k_j > k_j^0$, we have

$$\Pr \left(\frac{1}{k_j^0 - k_{j-1}^0} |D| > \frac{2\sqrt{T} \log T}{k_j^0 - k_{j-1}^0} C_T \right) = o(T^{-\delta}).$$

Next, we consider $\sum_{t=k_{j-1}^0}^{k_j^0-1} ((y_{it} - x'_{it} \beta_{g,j})^2 - (y_{it} - x'_{it} \beta_{g_i^0(j),j})^2)$. This term can be considered in a similar way to Bonhomme and Manresa (2015) and Okui and Wang (2021). We have

$$\begin{aligned} & \sum_{t=k_{j-1}^0}^{k_j^0-1} ((y_{it} - x'_{it} \beta_{g,j})^2 - (y_{it} - x'_{it} \beta_{g_i^0(j),j})^2) \\ &= \sum_{t=k_{j-1}^0}^{k_j^0-1} 2u_{it}x_{it}(\beta_{g_i^0(j),j} - \beta_{g,j}) + \sum_{t=k_{j-1}^0}^{k_j^0-1} (\beta_{g_i^0(j),j} - \beta_{g,j})' x_{it} x'_{it} (2\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j),j} - \beta_{g,j}) \\ &= \sum_{t=k_{j-1}^0}^{k_j^0-1} 2u_{it}x_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0) + \sum_{t=k_{j-1}^0}^{k_j^0-1} (x'_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0))^2 + \Psi, \end{aligned}$$

where

$$\begin{aligned} \Psi &= \sum_{t=k_{j-1}^0}^{k_j^0-1} 2u_{it}x_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0 - \beta_{g_i^0(j),j}^0 + \beta_{g,j}^0) \\ &\quad + \sum_{t=k_{j-1}^0}^{k_j^0-1} (\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0 - \beta_{g_i^0(j),j}^0 + \beta_{g,j}^0)' x_{it} x'_{it} (2\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j),j}^0 - \beta_{g,j}^0) \\ &\quad + \sum_{t=k_{j-1}^0}^{k_j^0-1} (\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0)' x_{it} x'_{it} (\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j),j}^0 - \beta_{g,j}^0 + \beta_{g,j}^0). \end{aligned}$$

By the Cauchy-Schwarz inequality, Assumption 4(ii) and the definition of \mathcal{N} imply that

$$|\Psi| \leq \eta C_1 \left\| \sum_{t=k_{j-1}^0}^{k_j^0-1} u_{it} x_{it} \right\| + \eta C_2 \left\| \sum_{t=k_{j-1}^0}^{k_j^0-1} x_{it} x'_{it} \right\|,$$

where C_1 and C_2 are constants independent of η and T . We then have that

$$\begin{aligned} & \mathbf{1} \left(\sum_{t=k_{j-1}}^{k_j-1} (y_{it} - x'_{it}\beta_{g,j})^2 < \sum_{t=k_{j-1}}^{k_j-1} (y_{it} - x'_{it}\beta_{g_i^0(j),j})^2 \right) \\ & \leq \mathbf{1} \left(\sum_{t=k_{j-1}^0}^{k_j^0-1} 2u_{it}x'_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0) \right. \\ & \quad \left. - \sum_{t=k_{j-1}^0}^{k_j^0-1} (x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g,j}^0))^2 + \eta C_1 \left\| \sum_{t=k_{j-1}^0}^{k_j^0-1} u_{it}x_{it} \right\| + \eta C_2 \left\| \sum_{t=k_{j-1}^0}^{k_j^0-1} x_{it}x'_{it} \right\| + |D| \right). \end{aligned}$$

Note that the right hand side does not depend on β . Thus, we have

$$\begin{aligned} & \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{k \in K} \mathbf{1}(\hat{g}_i(j)(k, \beta) \neq g_i^0(j)) \neq 0 \right) \\ & = \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{k \in K} \max_{g \in \mathbb{G}^j \setminus \{g_i^0(j)\}} \mathbf{1} \left(\sum_{t=k_{j-1}}^{k_j-1} (y_{it} - x'_{it}\beta_{g,j})^2 < \sum_{t=k_{j-1}}^{k_j-1} (y_{it} - x'_{it}\beta_{g_i^0(j),j})^2 \right) \neq 0 \right) \\ & \leq \sum_{g \in \mathbb{G}^j \setminus \{g_i^0(j)\}} \Pr \left(\sum_{t=k_{j-1}^0}^{k_j^0-1} 2u_{it}x'_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0) \right. \\ & \quad \left. < - \sum_{t=k_{j-1}^0}^{k_j^0-1} (x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g,j}^0))^2 + \eta C_1 \left\| \sum_{t=k_{j-1}^0}^{k_j^0-1} u_{it}x_{it} \right\| + \eta C_2 \left\| \sum_{t=k_{j-1}^0}^{k_j^0-1} x_{it}x'_{it} \right\| + |D| \right) \\ & \leq \sum_{g \in \mathbb{G}^B \setminus \{g_i^0(B)\}} \left(\Pr \left(\frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} (x'_{it}(\beta_{g_i^0(B),B}^0 - \beta_{g,j}^0))^2 \leq \frac{c''}{2} \right) \right. \\ & \quad \left. + \Pr \left(\left\| \frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} u_{it}x_{it} \right\| \geq M \right) \right. \\ & \quad \left. + \Pr \left(\left\| \frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} x_{it}x'_{it} \right\| \geq M \right) + \Pr \left(\frac{1}{k_j^0 - k_{j-1}^0} |D| > \frac{2\sqrt{T} \log T}{k_j^0 - k_{j-1}^0} C_T \right) \right. \\ & \quad \left. + \Pr \left(\frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} 2u_{it}x'_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0) < -\frac{c''}{2} + \eta C_1 M + \eta C_2 M + \frac{2\sqrt{T} \log T}{k_j^0 - k_{j-1}^0} C_T \right) \right), \end{aligned}$$

where we take $c'' = c \times \rho^*$ for c in Assumption 4(viii) and ρ^* in Assumption 4(iv).

We observe that

$$\Pr \left(\left\| \frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} x_{it}x'_{it} \right\| \geq M \right)$$

$$\leq \Pr \left(\frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} \|x_{it}x'_{it}\| \geq M \right) = \Pr \left(\frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} x'_{it}x_{it} \geq M \right).$$

We then apply Lemma S.5, regarding $x'_{it}x_{it} - E(x'_{it}x_{it})$ as z_t in the lemma, and Assumption 4(ix) yields that $\Pr \left(\left\| (k_j^0 - k_{j-1}^0)^{-1} \sum_{t=k_{j-1}^0}^{k_j^0-1} x_{it}x'_{it} \right\| \geq M \right) = o((k_j^0 - k_{j-1}^0)^{-\delta}) = o(T^{-\delta})$, where the last equality holds by Assumption 4(vii). Similarly, Assumption 4(ix) also implies that $\Pr \left(\left\| (k_j^0 - k_{j-1}^0)^{-1} \sum_{t=k_{j-1}^0}^{k_j^0-1} u_{it}x_{it} \right\| \geq M \right) = o(T^{-\delta})$. Moreover, a similar argument shows that under Assumption 4(ix), Lemma S.5 implies that

$$\begin{aligned} & \Pr \left(\left| \frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} (x'_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0))^2 - \frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} E((x'_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0))^2) \right| \geq \frac{c''}{2} \right) \\ &= o(T^{-\delta}), \end{aligned}$$

which in turn implies that under Assumptions 4(iv) and 4(viii),

$$\Pr \left(\frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} (x'_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0))^2 \leq \frac{c''}{2} \right) = o(T^{-\delta})$$

uniformly over g . Now we have shown that $\Pr\{(k_j^0 - k_{j-1}^0)^{-1}|D| > ((2\sqrt{T} \log T)/(k_j^0 - k_{j-1}^0))C_T\} = o(T^{-\delta})$. Note that $((2\sqrt{T} \log T)/(k_j^0 - k_{j-1}^0))C_T \rightarrow 0$ because $M_T = o(\sqrt{T}/\log T)$, $(k_j^0 - k_{j-1}^0) = O(T)$. Moreover, by similar arguments as above that use Lemma S.5, we can take η small enough and also T large enough such that

$$\begin{aligned} & \Pr \left(\frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} 2u_{it}x'_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0) < -\frac{c''}{2} + \eta C_1 M + \eta C_2 M + \frac{2\sqrt{T} \log T}{k_j^0 - k_{j-1}^0} C_T \right) \\ & \leq \Pr \left(\frac{1}{k_j^0 - k_{j-1}^0} \sum_{t=k_{j-1}^0}^{k_j^0-1} 2u_{it}x'_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g,j}^0) < -\frac{c''}{4} \right) = o(T^{-\delta}). \end{aligned}$$

uniformly over g under Assumption 4(ix). It thus follows that

$$\begin{aligned} & \Pr \left(\max_{1 \leq i \leq N} \sup_{\beta \in \mathcal{N}} \max_{k \in K} \mathbf{1}(\hat{g}_i(j)(k, \beta) \neq g_i^0(j)) \neq 0 \right) \\ & \leq \sum_{i=1}^N \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{k \in K} \mathbf{1}(\hat{g}_i(j)(k, \beta) \neq g_i^0(j)) \neq 0 \right) = o(NT^{-\delta}). \end{aligned}$$

This completes the proof. □

Proof of Theorem S.3

Proof. We observe that

$$\begin{aligned}\Pr(\hat{k} \neq k^0) &\leq \Pr(\hat{k} \neq k^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{\beta} \notin \mathcal{N}) \\ &\leq \Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{\gamma} \neq \gamma^0, \hat{\beta} \in \mathcal{N}) + \Pr(\hat{\beta} \notin \mathcal{N}).\end{aligned}$$

We analyze the three terms in the right hand side of the above display. First, for the third term, Lemma S.12 and the discussion below it imply that $\Pr(\hat{\beta} \notin \mathcal{N}) \rightarrow 0$. For the second term, by Lemmas S.11, S.12 and S.13, we have that

$$\begin{aligned}\Pr(\hat{\gamma} \neq \gamma^0, \hat{\beta} \in \mathcal{N}) &\leq \Pr\{\hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N}, \hat{\beta} \in \mathcal{N}\} + \Pr(\hat{k} \notin K) \\ &\leq \Pr\{\hat{\gamma}(k, \beta) \neq \gamma^0 \text{ for some } k \in K \text{ and } \beta \in \mathcal{N}\} + \Pr(\hat{k} \notin K) \rightarrow 0.\end{aligned}$$

Finally, we consider the first term. We observe that

$$\begin{aligned}\Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) &\leq \Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}, \hat{k} \in K) + \Pr(\hat{k} \notin K) \\ &\leq \Pr\{\hat{k}(\gamma^0, \beta) \neq k^0 \text{ for some } \beta \in \mathcal{N}, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}\} + \Pr(\hat{k} \notin K) \\ &\leq \Pr\{\hat{k}(\gamma^0, \beta) \neq k^0 \text{ for some } \beta \in \mathcal{N}\} + \Pr(\hat{k} \notin K),\end{aligned}$$

where $\hat{k}(\gamma^0, \beta) = \operatorname{argmin}_{k \in K} Q(k, \gamma, \beta)$. Note that $\Pr(\hat{k} \notin K) \rightarrow 0$ by Lemma S.11.

Let $k(j) = \{k_1, \dots, k_j^0, \dots, k\}$. Note that $k_{j-1} < k_j^0 < k_{j+1}$ for $k \in K$. We now show that $\Pr(Q(k(j), \gamma^0, \beta) > \min_{k_j < k_j < k_j^0} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N})$ converges to zero. Observe that

$$\begin{aligned}Q(k(j), \gamma^0, \beta) &> \min_{\underline{k}_j \leq k_j \leq \bar{k}_j, k_j \neq k_j^0} Q(k, \gamma^0, \beta) \\ &= \min \left(\min_{\underline{k}_j \leq k_j < k_j^0} Q(k, \gamma^0, \beta), \min_{k_j^0 < k_j \leq \bar{k}_j} Q(k, \gamma^0, \beta) \right).\end{aligned}$$

Suppose for the moment that $k_j^0 < k_j \leq \bar{k}_j$. We observe

$$\begin{aligned}Q(k(j), \gamma^0, \beta) - Q(k, \gamma^0, \beta) &= -\frac{1}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j), j} - \beta_{g_i^0(j-1), j-1}))^2 + \frac{2}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j), j} - \beta_{g_i^0(j), j}))^2 \\ &\quad - \frac{2}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x'_{it}(\beta_{g_i^0(j), j} - \beta_{g_i^0(j-1), j}) u_{it}.\end{aligned}$$

Let $d_i^0 = \beta_{g_i^0(j), j}^0 - \beta_{g_i^0(j-1), j-1}^0$ and $d_i = \beta_{g_i^0(j), j} - \beta_{g_i^0(j-1), j-1}$. It holds that

$$\frac{1}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j), j} - \beta_{g_i^0(j-1), j-1}))^2 = \frac{1}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it}(d_i^0 + d_i - d_i^0))^2$$

$$\begin{aligned}
&\geq \frac{1}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it} d_i^0)^2 + \frac{1}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it}(d_i - d_i^0))^2 - 2 \frac{1}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N |x'_{it} d_i^0| \cdot |x'_{it}(d_i - d_i^0)| \\
&\geq \frac{1}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it} d_i^0)^2 + \frac{1}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it}(d_i - d_i^0))^2 \\
&\quad - 2 \frac{1}{T} \sum_{t=k_j^0}^{k_j-1} \left(\frac{1}{N} \sum_{i=1}^N (x'_{it} d_i^0)^2 \right)^{1/2} \left(\frac{1}{N} \sum_{i=1}^N (x'_{it}(d_i - d_i^0))^2 \right)^{1/2}.
\end{aligned}$$

Assumptions 4(ii) and 4(v) imply that $\sum_{i=1}^N (x'_{it} d_i^0)^2 / N < (k_j - k_j^0)C/T$. Similarly, Assumption 4(v) and the condition that $\beta \in \mathcal{N}_\eta$ imply that $\sum_{i=1}^N (x'_{it}(d_i - d_i^0))^2 / N < (k_j - k_j^0)\eta^2 C/T$. We thus have that $\sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j),j} - \beta_{g_i^0(j-1),j-1}))^2 / (NT) \geq (k_j - k_j^0)(\underline{m} - C\eta)/T$, by Assumption 4(vi). Then by taking η small enough, we have that

$$\frac{1}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j),j} - \beta_{g_i^0(j-1),j-1}))^2 \geq \frac{k_j - k_j^0}{2T} \underline{m}.$$

It is easy to observe that

$$\frac{2}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N (x'_{it}(\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j-1),j-1}))^2 > 0.$$

It follows that

$$\begin{aligned}
&\Pr \left(Q(k(j), \gamma^0, \beta) > \min_{k_j^0 < k_j \leq \bar{k}_j} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right) \\
&= \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{k_j^0 < k_j \leq \bar{k}_j} (Q(k(j), \gamma^0, \beta) - Q(k, \gamma^0, \beta)) > 0 \right) \\
&\leq \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{k_j^0 < k_j \leq \bar{k}_j} \left(-2 \frac{1}{NT} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x'_{it}(\beta_{g_i^0(j),j} - \beta_{g_i^0(j-1),j-1}) u_{it} - \frac{k_j - k_j^0}{2T} \underline{m} \right) > 0 \right) \\
&= \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{k_j^0 < k_j \leq \bar{k}_j} \left(-2 \frac{1}{N} \frac{1}{k_j - k_j^0} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x'_{it}(\beta_{g_i^0(j),j} - \beta_{g_i^0(j-1),j-1}) u_{it} - \frac{\underline{m}}{2} \right) > 0 \right) \\
&\leq \Pr \left(\sup_{\beta \in \mathcal{N}} \max_{k_j^0 < k_j \leq \bar{k}_j} \left(-2 \frac{1}{N} \frac{1}{k_j - k_j^0} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x'_{it}(\beta_{g_i^0(j),j} - \beta_{g_i^0(j-1),j-1}) u_{it} \right) > \frac{\underline{m}}{2} \right).
\end{aligned}$$

Observing that

$$-2 \frac{1}{N} \frac{1}{k_j - k_j^0} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x'_{it}(\beta_{g_i^0(j),j} - \beta_{g_i^0(j-1),j-1}) u_{it}$$

$$\begin{aligned}
&= -2 \frac{1}{N} \frac{1}{k_j - k_j^0} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x'_{it} (\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j-1),j-1}^0) u_{it} \\
&\quad + 2 \frac{1}{N} \frac{1}{k_j - k_j^0} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x'_{it} (\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j-1),j-1}^0 + \beta_{g_i^0(j-1),j-1}^0) u_{it},
\end{aligned}$$

and

$$\begin{aligned}
&\left| 2 \frac{1}{N} \frac{1}{k_j - k_j^0} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x'_{it} (\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j-1),j-1}^0 + \beta_{g_i^0(j-1),j-1}^0) u_{it} \right| \\
&\leq \eta C \left\| \frac{1}{N} \frac{1}{k_j - k_j^0} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x_{it} u_{it} \right\|,
\end{aligned}$$

we thus have that

$$\begin{aligned}
&\Pr \left(\sup_{\beta \in \mathcal{N}} \max_{k_j^0 < k_j \leq \bar{k}_j} \left(-2 \frac{1}{N} \frac{1}{k_j - k_j^0} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x'_{it} (\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j-1),j-1}^0) u_{it} \right) > \frac{m}{2} \right) \\
&\leq \Pr \left(\max_{k_j^0 < k_j \leq \bar{k}_j} \left(-2 \frac{1}{N} \frac{1}{k_j - k_j^0} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x'_{it} (\beta_{g_i^0(j),j}^0 - \beta_{g_i^0(j-1),j-1}^0) u_{it} \right) > \frac{m}{4} \right) \\
&\quad + \Pr \left(\eta C \max_{k_j^0 < k_j \leq \bar{k}_j} \left\| \frac{1}{N} \frac{1}{k_j - k_j^0} \sum_{t=k_j^0}^{k_j-1} \sum_{i=1}^N x_{it} u_{it} \right\| > \frac{m}{4} \right) = O \left(\frac{1}{N} \right),
\end{aligned}$$

where the last equality follows by applying [Bai and Perron \(1998, Lemma A.6\)](#) which is an extension of [Hájek and Rényi \(1955\)](#). Here we use the observation that an L_r -bounded mixing sequence is an L_p mixingale sequence for $1 \leq p < r$ as discussed in [\(Davidson, 1994, page 248\)](#). Thus, under Assumptions 4(ix) and 4(x), $x_{it} u_{it}$ is an L_2 mixingale and we can apply [Bai and Perron \(1998, Lemma A.6\)](#).

A similar argument shows that

$$\Pr \left(Q(k(j), \gamma^0, \beta) > \min_{k_j < k_j < k_j^0} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right) = O \left(\frac{1}{N} \right).$$

Thus, we have

$$\begin{aligned}
&\Pr \{ \hat{k}(\gamma^0, \beta) \neq k^0 \text{ for some } \beta \in \mathcal{N} \} \\
&\leq \sum_{j=1}^m \left(Q(k(j), \gamma^0, \beta) > \min_{k_j < k_j < k_j^0} Q(k, \gamma^0, \beta) \text{ for some } \beta \in \mathcal{N} \right) = O \left(\frac{1}{N} \right).
\end{aligned}$$

To sum up, we have that $\Pr(\hat{k} \neq k^0, \hat{\gamma} = \gamma^0, \hat{\beta} \in \mathcal{N}) \rightarrow 0$.

□

Proof of Corollary S.3

Proof. The proof is identical to that of Corollary 1 and is omitted. □

S.5 Additional simulation studies

This section provides additional simulation results that are not reported in the paper.

S.5.1 Percentage of perfect classification

Our theory states the super-consistency of group membership estimates in terms of the probability of perfect classification. Such consistency of clustering is driven by the time-series dimension but not the cross-sectional dimension. To better see the pattern of convergence, we compute the proportion of perfect classification (PPC) in each regime as

$$\text{PPC} = \frac{\text{The number of times when all units are correctly classified}}{\text{Total number of replications}},$$

and report PPC_B (before the break) and PPC_A (after the break) for time series sample size $T = 10, 20$ and 40 in Table S.1.

INSERT TABLE S.1 HERE

When $N = 100$ and $T = 10$, the PPC is (close to) zero in all DGPs, suggesting that no perfect clustering is achieved, although only very few units (less than 6% in all cases except DGP 2.2 and 3.2) are misclustered in each replication as reported by the misclustering rate. When $N = 100$ and $T = 20$, the PPC_B increases to 5–10% in all DGPs before the break. As we further increase T to 40, PPC_B dramatically improves to more than 80% in DGP 1, 12%–36% in DGP 2, 47%–60% in DGP 3, and more than 83% in DGP 4. The PPC_A after the break is generally lower than PPC_B since the length of post-break periods is shorter. Interestingly, the difference between PPC_B and PPC_A is disproportionately larger than the difference between MF_B and MF_A when $T = 40$, but on the other hand, when $T = 20$ there are a few cases where PPC_A is higher than PPC_B . This suggests that the PPC measure is more volatile than the misclustering frequency (MF) used in the paper. In any case, we find strong evidence of super-consistency in that the PPC improves exponentially as T increases. When we double the cross-sectional sample size, namely $N = 200$, the PPC seems to decrease to some extent compared to those when $N = 100$. But the super-consistency in T still holds.

S.5.2 Diagnosing the type of breaks

This section examines the numerical performance of the information criterion in diagnosing the type of breaks. Table S.2 presents the empirical probabilities of selecting a specific type of break using the IC under various DGPs as in the paper, and the bold number indicates the larger one between the two candidates. We first examine the case of diagnosing between a break only in coefficients (Model 1) and a break in both coefficients and memberships (Model 3). In DGP X.1 with a coefficient break, the proposed IC correctly selects Model 1 with only a coefficient break but time-invariant memberships in all cases. In DGP X.3 with a break in both coefficients and memberships, the probabilities of selecting the correct model (Model 3) and not the incorrect one (Model 1) is also generally higher, although the difference is relatively small when T is small. Similarly, when we diagnose between a break only in group memberships (Model 2) and a break in both (Model 3), IC can also detect the correct type of breaks with larger probabilities, except in some cases when T is small.

INSERT TABLE S.2 HERE

S.5.3 Extended simulation designs

In this subsection, we deviate from the benchmark designs and consider four extensions.

First, we examine how the behavior of LSGB depends on the size of break and the degree of group separation. To this end, we set the slope coefficients in DGP X.1 and X.3 as

$$\beta_{1,t} = \begin{cases} \beta_{1,B} = -0.25\iota_p & \text{if } t < k^0 \\ \beta_{1,A} = -0.5\iota_p & \text{if } t \geq k^0 \end{cases}, \quad \text{and} \quad \beta_{2,t} = 0.5\iota_p.$$

In DGP X.2, we set $\beta_{1,t} = -0.25\iota_p$ in Group 1 and $\beta_{2,t} = 0.25\iota_p$ in Group 2. Table S.3 and S.4 report the accuracy of break point estimates and misclustering frequency in this setting. Despite the fact that allowing for a smaller break increases the HD ratio of the break point estimate and the average is further away from its true value for all methods as expected, LSGB can still detect the break point rather accurately. For example, the averaged absolute deviation of LSGB is no more than 0.9 except for DGP 2.1. As the cross-sectional dimension increases, this deviation reduces to less than 0.7 when $T = 10$ and less than 0.35 when $T = 20$. The clustering accuracy is also lower when there is a smaller degree of group separation, for all methods. Nevertheless, LSGB still correctly clusters at least 90% of the units in all cases except DGP X.2, and the accuracy improves quickly as T increases. The correct clustering frequency in DGP X.2 is at the same level as in the benchmark setting in the paper because

the degree of group separation is the same (with the difference in the coefficients of the two groups both being 0.5).

INSERT TABLE S.3 and S.4 HERE

Second, we consider break points occurring at different points of the sample period, i.e. $k^0 = \lfloor 0.6T \rfloor$ and $k^0 = \lfloor 0.8T \rfloor$, for $T = 10$ and $T = 20$. To save space, we only report results of static panels with serially correlated errors (DGP 2) under $N = 100$. The conclusions are qualitatively unchanged if we consider the other DGPs and a larger cross-sectional size $N = 200$.¹ The HD of the break point estimates and MSE of the coefficient estimates are given in the first four columns of Table S.5, and the misclustering frequency of the three methods is presented in the upper panel of Table S.6. When the break occurs closer to the middle of the time period, $k^0 = \lfloor 0.6T \rfloor$, the accuracy of break detection of BFK improves (e.g. HD decreases from 0.3267 to no larger than 0.14 when $T = 10$ compared to HD) since more balanced sample sizes are available for both regimes and that improves the efficiency of the individual time series coefficient estimates. The improvement of break point estimation further increases the accuracy of clustering for BFK, especially in the post-break regime. When the break gets closer to the end of sample, $k^0 = \lfloor 0.8T \rfloor$, it is even more difficult for BFK to detect the break (HD larger than 0.3 when $T = 10$ and = 0.2000 when $T = 20$), and this further leads to less accurate clustering. In both cases, LSGB remains effective and continues to outperform BFK in break detection and clustering, although their discrepancy varies over different cases.

Next, we examine the potential efficiency loss of GAGFL when the number of groups increases. We consider three groups in both pre- and post-break regimes.

DGP X.1 We fix the ratio of units among the three groups as $N_1 : N_2 : N_3 = 0.3 : 0.3 : 0.4$, and the group membership does not change after the structural break. The coefficients in the three groups are given by

$$\beta_{1,t} = \begin{cases} \beta_{1,B} = \iota_p & \text{if } t < k^0 \\ \beta_{1,A} = 2\iota_p & \text{if } t \geq k^0 \end{cases}, \quad \beta_{2,t} = \begin{cases} \beta_{2,B} = 2\iota_p & \text{if } t < k^0 \\ \beta_{2,A} = \iota_p & \text{if } t \geq k^0 \end{cases}, \quad \beta_{3,t} = 0.5\iota_p.$$

DGP X.2: This case considers when group memberships change after the break but the slope coefficients do not. The ratio of units among groups is $N_1 : N_2 : N_3 = 0.3 : 0.3 : 0.4$ before the break, and $N_1 : N_2 : N_3 = 0.4 : 0.4 : 0.2$ after the break. We generate the group memberships before and after the break independently as above. The slope coefficients are $\beta_{1,t} = \iota_p$, $\beta_{2,t} = 2\iota_p$, and $\beta_{3,t} = 0.5\iota_p$.

¹Detailed results are available upon request.

DGP X.3: Both the slope coefficients of each group and the group structure change after the break. The slope coefficients are the same as in DGP X.1, and the group structure is the same as DGP X.2.

In this last case, GAGFL needs to specify 5 groups in DGP X.2 and X.3 to ensure that each of the groups contains the same units over the entire period, much more than the true underlying number. Again, only results under DGP 2 and $N = 100$ are reported. The left bottom panel of Table S.6 shows that the clustering accuracy of GAGFL decreases dramatically, much worse than that of LSGB in DGP 2.2 and 2.3. Compared to the corresponding benchmark cases, the break point and coefficient estimates of GAGFL are also much less accurate and the discrepancy between GAGFL and LSGB becomes more significant. This suggests that the relative efficiency loss of GAGFL compared to LSGB is particularly prominent when the number of groups increases, because the necessary number of groups for GAGFL typically increases much faster than the increase in the underlying true number of groups. Of course, if we under-specify the number of groups for GAGFL, e.g. by imposing the same number as the underlying truth, GAGFL leads to inconsistent estimates for all quantities.

Finally, we examine how the features of breaks affect the performance. In DGP X.1 and X.3, we consider that all units experience breaks at the same date but with heterogeneous sizes. This coincides with the DGP of BFK that assumes a common break (in dates) and allows for heterogeneous slope coefficients, and we examine how LSGB performs in this case. The LSGB and GAGFL estimates of break point, group memberships, and slope coefficients are all more accurate than the benchmark case because imposing breaks in more series is equivalent to increasing the number of series or the magnitude of the break. The estimates of BFK are more accurate than the benchmark only for $T = 20$, but not for $T = 10$. This again reflects the severe efficiency constraint of BFK in small samples. We find that even when there is no heterogeneity in break points, LSGB still produces more accurate break point and coefficient estimates than BFK and GAGFL. In DGP X.2 where only memberships change, we generate a larger degree of heterogeneity of the two groups by setting time-invariant coefficients $\beta_{1,t} = \iota$ for Group 1 and $\beta_{2,t} = 2\iota$ for Group 2. This also implies larger breaks since those units that change group memberships have more different coefficients after the break. Again, we find that the performance of all methods improves while LSGB remains the most appealing in general.

INSERT TABLE S.5 and S.6 HERE

S.5.4 Multiple break detection using a sequential approach

This section presents the simulation studies of sequential break detection when there are multiple breaks. We also verify our theoretical arguments via simulation. We consider the same DGPs as in the paper, but with two break points, the first at $k_1^0 = \lfloor 0.3T \rfloor$ and the second at $k_2^0 = \lfloor 0.7T \rfloor$. In DGP X.1 with only coefficient breaks, we generate the slope coefficients for Group 1 as

$$\beta_{1,t} = \begin{cases} \beta_{1,B} = \iota_p & \text{if } t < k_1^0 \\ \beta_{1,A} = 2\iota_p & \text{if } k_2^0 \geq t \geq k_1^0 \\ \beta_{1,A} = 1.5\iota_p & \text{if } t \geq k_2^0 \end{cases}$$

and for Group 2 as

$$\beta_{2,t} = \begin{cases} \beta_{2,B} = 0.5\iota_p & \text{if } t < k_1^0 \\ \beta_{2,A} = 0.5\iota_p & \text{if } k_2^0 \geq t \geq k_1^0 \\ \beta_{2,A} = 0.8\iota_p & \text{if } t \geq k_2^0 \end{cases}.$$

Thus the relatively larger break point over all units is k_1^0 . In DGP X.2 with only membership breaks, we set the ratio of units among groups is $N_1 : N_2 = 0.4 : 0.6$ before the first break, $N_1 : N_2 = 0.6 : 0.4$ between the first and second break, and $N_1 : N_2 = 0.3 : 0.7$ after the second break. Thus, the relatively larger break point is k_2 . DGP X.3 generates the coefficients as in DGP X.1 and sets the ratio of groups as in DGP X.2, thus containing breaks in the two types of parameters, with the relatively larger break at k_1^0 .

INSERT TABLE S.7 HERE

Table S.7 presents the Hausdorff distance (HD) of the two break points and the empirical probability that the first detected break point produced by the sequential procedure, \hat{k}_1 coincides with the relatively larger one \hat{k}_L^0 . In all cases except DGP 3.1 and 3.3, the HD is zero or very close to zero, suggesting that the two break points can both be estimated quite accurately. The first estimated break point \hat{k}_1 always coincides with the relatively larger one in these cases, as stated in the theory. In DGP 3 with fixed effects, break detection is less accurate, as in the single-break case. Although the empirical probability of $\hat{k}_1 = k_L^0$ does not seem high in DGP 3, further examination reveals that \hat{k}_1 is always in the small neighbourhood of k_L^0 , suggesting that \hat{k}_1 still always tries to capture the relatively larger break but not the smaller one.

S.6 Computation time

This section report the computation time of the algorithm in our simulation and application.

Using Intel Xeon CPU with 3.10GHz, 3096Mhz, and 4 kernels without parallel computation, we tabulate the computation time of LSGB in DGP.1 ($G^B = G^A = 2$) with different sample sizes, amount of initial values, and the specified number of groups for estimation (not the true number of groups) in Table S.8. For $N = 100$ and $T = 10$, one replication for $G^B = G^A = 2$ with 25 initial values takes less than 5 seconds. Increasing the time dimension and the number of initial values increases the computation time linearly. In contrast, the effect of the number of groups on computation time is non-proportional. When we double the number of groups specified for estimation (while keeping the true number unchanged, i.e., $G^{B0} = G^{A0} = 2$), the computation time more than doubles.

In the application (with $N = 703$ and $T = 32$), our empirical results are obtained from estimation using 500 initial values with $G_B = 5$ and $G_A = 7$. This estimation takes 1.31 hours. To determine the optimal number of groups, we search G_B and G_A by ranging them from 1:10, respectively. This search of optimal numbers of groups takes 56.58 hours. While this computation time is affordable, our estimation can be implemented via parallel computation, which substantially reduces the time. For example, using parallel computation with 4 workers, estimating the empirical dataset under the same setting for $G_B = 5$ and $G_A = 7$ takes less than 20 minutes.

S.7 Additional empirical results

S.7.1 Industry vs. data-driven clustering

Figure S.1 presents the appearance frequency of each industry (specified by the two-digit Standard Industrial Classification (SIC) code) in a group for both regimes.

INSERT FIGURE S.1 HERE

S.7.2 Empirical results under alternative choices of the number of groups

This section examines the robustness of the empirical results with respect to the choice of the number of groups. We first plot the surface of the IC with G_B and G_A ranging from 1 to 10, respectively, in Figure S.2.

INSERT FIGURE S.2 HERE

When we set $\lambda \in [4.1, 4.4]$, the IC suggests $G_B = 4$ and $G_A = 5$. We re-estimate the sales growth regression with this specification. The estimated break point is 1996, one period earlier than the break point reported in the paper. We present the coefficient estimates in Table S.9. We refer to the groups obtained under $G_B = 4$ and $G_A = 5$ as Groups 1.B'–4.B' before the break and Groups 1.A'–5.A' after the break. As expected, our original groups merge and mix to some extent in this case, while the partition of units and coefficient estimates are closely in line with the original results. Specifically, Group 1.B' largely inherits the original Group 1.B with similar memberships and coefficient estimates. Group 2.B' mainly comes from Group 2.B, which exhibits a significantly positive leverage effect. Group 3.B' mainly comes from Group 3.B, added by a few units from Group 4.B, and it shows a significantly negative leverage effect. Group 4.B' merges Groups 4.B and 5.B, whose leverage effect is significantly negative and of larger magnitude than Group 3.B'. After the structural break, the new Group 1.A' merges Groups 1.A, 2.A, and 3.A, and exhibits a significantly positive leverage effect. Group 2.A' and 3.A' mainly inherit Groups 4.A and 5.A, respectively, while Group 4.A' is a combination of Groups 5.A and 6.A. The leverage effect is significantly negative for these three groups (2.A'–4.A'), but the magnitude increases from 2.A' to 4.A'. Finally, Group 5.A' is almost identical to Group 7.A, which possesses an insignificant leverage effect.

When $\lambda \in [2.0, 2.6]$, the IC suggests $G_B = 7$ and $G_A = 9$, two groups more in each regime, respectively, than our original choice, and the estimated break point is also close to 1996. We refer to the groups in this setting as Groups 1.B''–7.B'' before the break and Groups 1.A''–9.A'' after the break, and the corresponding coefficient estimates are provided in Table S.10. In this case, the extra groups mainly come from further segmentation of existing groups, and the coefficient estimate of the focal variable is rather robust. Specifically, the original Group 1.B is divided into Groups 1.B'' and 2.B'', whose leverage effects are both positive but differ in magnitude. Group 3.B'' is almost identical to 2.B with its slightly negative leverage effect. The original Group 3.B is divided into Groups 4.B'' and 5.B'', while Group 4.B is further segmented into Groups 6.B'' and 7.B''. The leverage effects in Groups 4.B''–7.B'' are all significantly negative, and ordered from the least to the largest in terms of their magnitude. We also note that the discrepancy between 5.B'' and 6.B'' is relatively minor, which suggests that such a finer partition may not be necessary. As for the post-break regime, the new groups are also a finer segmentation of the benchmark grouping reported in the paper to certain extent. A small portion of original Group 2.A is separated into Groups 2.A'' and 3.A'', which differs in the magnitude of the positive leverage effect. Groups 4.A''–6.A'' are mainly from further partition of the original Group 4.A. Group 1.A'' is almost identical to Group 1.A whose leverage effect is the most strong and positive, and Groups 7.A''–9.A'' largely resemble Groups 5.A–7.A, respectively.

INSERT TABLE S.9 and S.10 HERE

References

- J. Bai and P. Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66:47–78, 1998.
- B. H. Baltagi, Q. Feng, and C. Kao. Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics*, 191:176–195, 2016.
- S. Bonhomme and E. Manresa. Grouped patterns of heterogeneity in panel data. *Econometrica*, 83:1147–1184, 2015.
- J. Davidson. *Stochastic limit theory: An introduction for econometricians*. Oxford University Press, Oxford, UK, 1994.
- J. Hájek and A. Rényi. Generalization of an inequality of Kolmogorov. *Acta Mathematica Hungarica*, 6:281–283, 1955.
- F. Merlevède, M. Peligrad, and E. Rio. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151:435–474, 2011.
- R. Okui and W. Wang. Heterogeneous structural breaks in panel data models. *Journal of Econometrics*, 220:447–473, 2021.
- E. Rio. *Asymptotic theory of weakly dependent random processes*. Springer, Berlin, Heidelberg, Germany, 2017.

Figure S.1: Composition of groups in two regimes

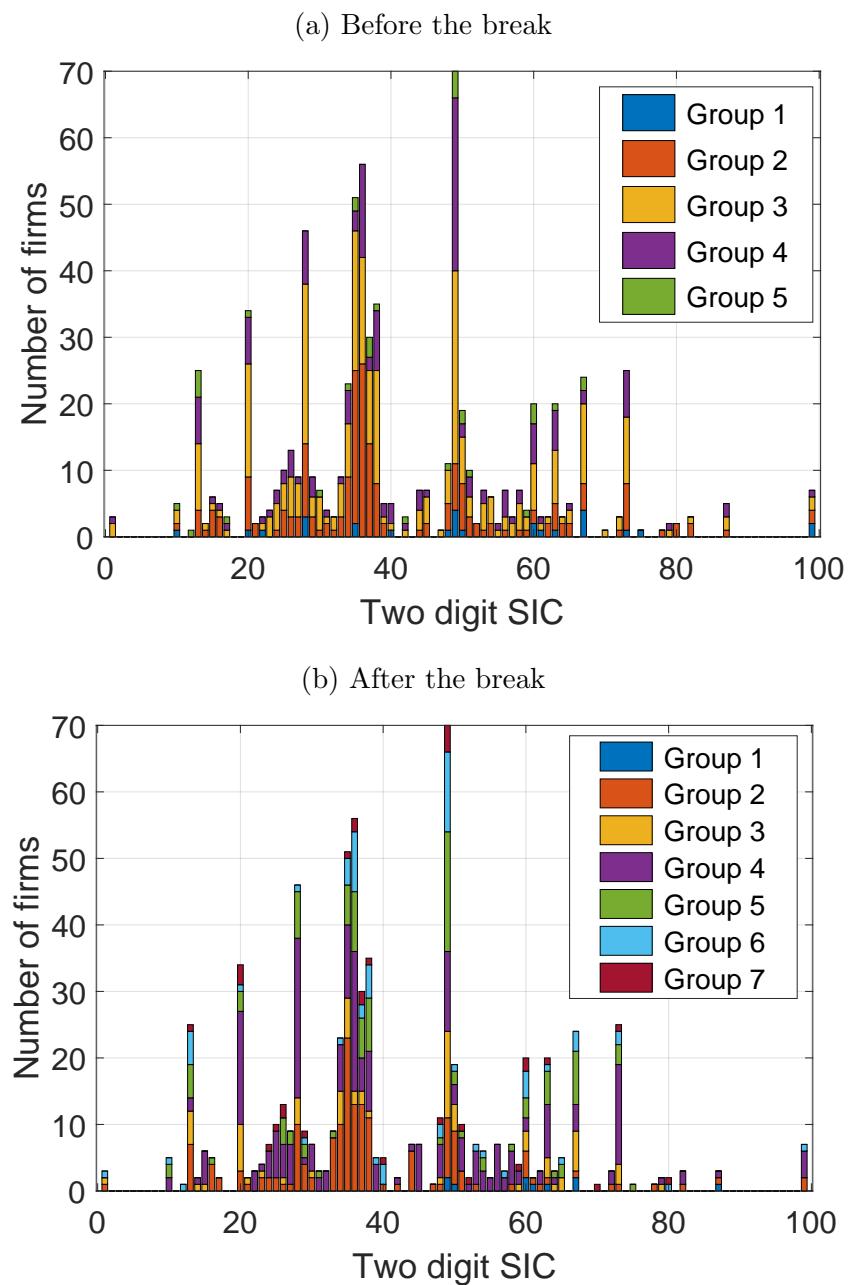


Figure S.2: IC for determination of the number of groups in sales growth regressions

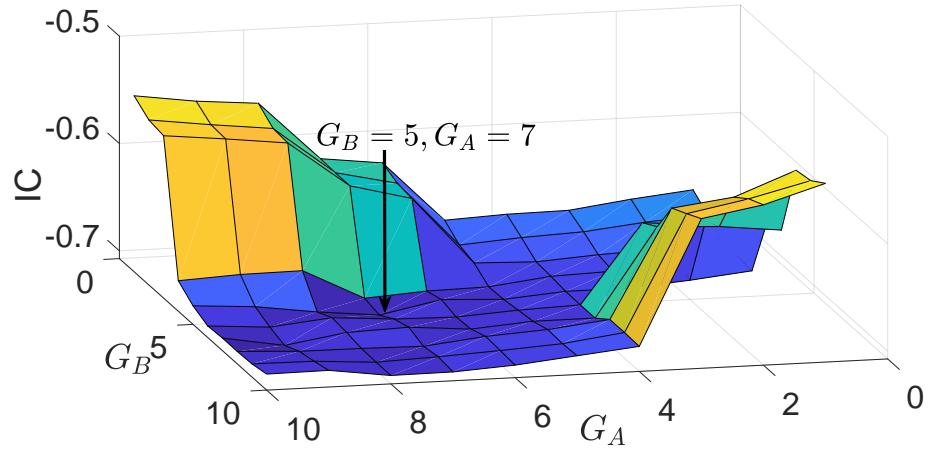


Table S.1: Proportion of perfect clustering of LSGB before and after the structural break

	N	T	DGP 1		DGP 2		DGP 3		DGP 4	
			PPC _B	PPC _A						
DGP X.1	100	10	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.001
		20	0.064	0.260	0.000	0.011	0.012	0.008	0.054	0.266
		40	0.818	0.775	0.126	0.313	0.600	0.284	0.835	0.749
	200	10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		20	0.013	0.107	0.000	0.001	0.000	0.000	0.007	0.087
		40	0.783	0.613	0.033	0.153	0.426	0.098	0.790	0.578
	DGP X.2	100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		20	0.080	0.000	0.000	0.000	0.010	0.000	0.069	0.001
		40	0.900	0.108	0.185	0.000	0.521	0.016	0.909	0.126
	200	10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		20	0.011	0.000	0.000	0.000	0.000	0.000	0.012	0.000
		40	0.807	0.074	0.053	0.000	0.280	0.000	0.860	0.045
DGP X.3	100	10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		20	0.090	0.008	0.001	0.000	0.111	0.005	0.097	0.011
		40	0.908	0.792	0.361	0.125	0.476	0.330	0.924	0.764
	200	10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		20	0.032	0.040	0.000	0.000	0.000	0.000	0.013	0.000
		40	0.840	0.642	0.073	0.040	0.207	0.107	0.847	0.580

Notes: This table presents the proportion of perfect clustering of the proposed method LSGB. PPC_B is the proportion before the break, and PPC_A denotes the proportion after the break.

Table S.2: Empirical probabilities of selecting a specific type of break using information criterion

Break only in coefficients (Model 1) versus break in both (Model 3)								
	$N = 100, T = 10$		$N = 100, T = 20$		$N = 200, T = 10$		$N = 200, T = 20$	
	Model 1	Model 3						
DGP 1.1	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
DGP 2.1	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
DGP 3.1	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
DGP 4.1	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
DGP 1.3	0.367	0.633	0.000	1.000	0.553	0.447	0.000	1.000
DGP 2.3	0.327	0.673	0.000	1.000	0.480	0.520	0.000	1.000
DGP 3.3	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
DGP 4.3	0.020	0.980	0.000	1.000	0.000	1.000	0.000	1.000
Break only in memberships (Model 2) versus break in both (Model 3)								
	$N = 100, T = 10$		$N = 100, T = 20$		$N = 200, T = 10$		$N = 200, T = 20$	
	Model 2	Model 3						
DPG 1.2	0.607	0.393	0.913	0.087	0.630	0.370	0.907	0.093
DGP 2.2	0.427	0.573	0.640	0.360	0.353	0.647	0.680	0.320
DGP 3.2	0.953	0.047	0.973	0.027	0.967	0.033	0.993	0.007
DGP 4.2	0.767	0.233	0.940	0.060	0.887	0.113	0.973	0.027
DGP 1.3	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
DGP 2.3	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
DGP 3.3	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
DGP 4.3	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000

Notes: DGP X.1 generates a break only in coefficients; DGP X.2 generates a break only in group memberships; DGP X.3 generates a break in both coefficients and memberships.

Table S.3: Accuracy of break point estimates under a smaller break and more similar groups

		$k^0 = 7$				$k^0 = 14$			
		$N = 100, T = 10$		$N = 200, T = 10$		$N = 100, T = 20$		$N = 200, T = 20$	
		HD	\bar{k}	HD	\bar{k}	HD	\bar{k}	HD	\bar{k}
DGP 1.1	LSGB	0.072	6.611	0.043	6.700	0.010	13.867	0.001	14.000
	BFK	0.241	4.589	0.246	4.544	0.184	10.322	0.172	10.556
	GAGFL	0.036	6.929	0.014	6.964	0.021	14.000	0.023	13.939
DGP 1.2	LSGB	0.022	6.911	0.001	7.011	0.002	14.033	0.000	14.001
	BFK	0.246	4.544	0.238	4.622	0.158	10.833	0.133	11.344
	GAGFL	0.320	4.405	0.088	7.133	0.085	12.931	0.035	13.958
DGP 1.3	LSGB	0.000	7.000	0.000	7.000	0.000	14.000	0.000	14.000
	BFK	0.239	4.611	0.256	4.444	0.108	11.844	0.102	11.967
	GAGFL	0.017	6.939	0.001	7.011	0.019	14.076	0.007	13.933
DGP 2.1	LSGB	0.113	6.200	0.036	6.978	0.083	13.144	0.014	14.144
	BFK	0.250	4.500	0.247	4.533	0.186	10.278	0.183	10.344
	GAGFL	0.049	6.571	0.004	7.000	0.066	13.425	0.017	13.729
DGP 2.2	LSGB	0.088	6.332	0.067	6.378	0.031	13.822	0.009	14.111
	BFK	0.254	4.456	0.256	4.444	0.179	10.422	0.178	10.433
	GAGFL	0.153	5.836	0.144	5.756	0.130	12.000	0.097	12.459
DGP 2.3	LSGB	0.001	6.989	0.000	7.000	0.000	14.000	0.000	14.000
	BFK	0.257	4.433	0.246	4.544	0.123	11.544	0.117	11.667
	GAGFL	0.055	6.674	0.009	7.000	0.032	13.557	0.010	13.839
DGP 3.1	LSGB	0.083	6.367	0.036	6.956	0.071	13.789	0.032	14.167
	BFK	0.263	4.367	0.249	4.511	0.109	11.811	0.101	11.978
	GAGFL	0.002	7.022	0.002	6.983	0.024	13.755	0.012	13.892
DGP 3.2	LSGB	0.040	6.867	0.030	6.833	0.022	14.311	0.019	14.333
	BFK	0.244	4.556	0.240	4.600	0.100	12.000	0.100	12.000
	GAGFL	0.104	6.067	0.024	6.777	0.032	13.463	0.002	13.962
DGP 3.3	LSGB	0.007	7.067	0.004	7.044	0.015	14.300	0.008	14.167
	BFK	0.232	4.678	0.226	4.744	0.100	12.000	0.100	12.000
	GAGFL	0.019	6.824	0.001	6.988	0.013	13.812	0.002	14.000
DGP 4.1	LSGB	0.076	6.489	0.038	6.800	0.021	14.100	0.001	14.011
	BFK	0.258	4.422	0.252	4.478	0.189	10.222	0.196	10.089
	GAGFL	0.030	6.926	0.022	6.980	0.025	13.944	0.030	13.750
DGP 4.2	LSGB	0.030	6.767	0.000	7.000	0.001	14.011	0.000	14.000
	BFK	0.250	4.500	0.239	4.611	0.135	11.300	0.114	11.722
	GAGFL	0.081	6.306	0.032	6.969	0.090	13.452	0.034	13.798
DGP 4.3	LSGB	0.000	7.000	0.000	7.000	0.000	14.000	0.000	14.000
	BFK	0.231	4.689	0.254	4.456	0.101	11.978	0.100	12.000
	GAGFL	0.001	6.985	0.003	6.978	0.009	14.007	0.010	13.932

Notes: HD denotes Hausdorff distance and \bar{k} is the average break point estimate.

Table S.4: Misclustering frequency under a smaller break and more similar groups

		$N = 100, T = 10$		$N = 200, T = 10$		$N = 100, T = 20$		$N = 200, T = 20$	
		MF_B	MF_A	MF_B	MF_A	MF_B	MF_A	MF_B	MF_A
DGP 1.1	LSGB	0.035	0.061	0.026	0.055	0.003	0.010	0.001	0.010
	BFK	0.057	0.019	0.049	0.021	0.006	0.003	0.005	0.003
	GAGFL	0.003	0.003	0.002	0.002	0.000	0.000	0.000	0.000
DGP 1.2	LSGB	0.072	0.198	0.064	0.179	0.014	0.088	0.014	0.081
	BFK	0.132	0.166	0.119	0.165	0.031	0.100	0.027	0.088
	GAGFL	0.279	0.279	0.243	0.243	0.168	0.168	0.116	0.116
DGP 1.3	LSGB	0.019	0.057	0.016	0.053	0.001	0.009	0.001	0.008
	BFK	0.051	0.096	0.053	0.091	0.003	0.031	0.002	0.029
	GAGFL	0.090	0.090	0.048	0.048	0.011	0.011	0.006	0.006
DGP 2.1	LSGB	0.068	0.101	0.049	0.099	0.024	0.039	0.011	0.034
	BFK	0.093	0.057	0.089	0.051	0.030	0.016	0.022	0.015
	GAGFL	0.018	0.018	0.018	0.018	0.002	0.002	0.001	0.001
DGP 2.2	LSGB	0.227	0.381	0.228	0.377	0.086	0.307	0.066	0.290
	BFK	0.291	0.363	0.302	0.353	0.135	0.257	0.116	0.245
	GAGFL	0.391	0.391	0.386	0.386	0.302	0.302	0.277	0.277
DGP 2.3	LSGB	0.043	0.100	0.043	0.090	0.012	0.030	0.011	0.030
	BFK	0.085	0.130	0.084	0.124	0.020	0.057	0.017	0.051
	GAGFL	0.210	0.210	0.139	0.139	0.070	0.070	0.032	0.032
DGP 3.1	LSGB	0.074	0.079	0.052	0.091	0.018	0.035	0.010	0.027
	BFK	0.084	0.060	0.075	0.058	0.009	0.019	0.008	0.018
	GAGFL	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000
DGP 3.2	LSGB	0.133	0.262	0.122	0.244	0.041	0.144	0.040	0.139
	BFK	0.168	0.242	0.156	0.219	0.046	0.141	0.047	0.137
	GAGFL	0.174	0.174	0.173	0.173	0.092	0.092	0.081	0.081
DGP 3.3	LSGB	0.051	0.161	0.047	0.140	0.008	0.045	0.007	0.048
	BFK	0.076	0.146	0.070	0.137	0.007	0.062	0.008	0.058
	GAGF	0.030	0.030	0.015	0.015	0.002	0.002	0.001	0.001
DGP 4.1	LSGB	0.034	0.051	0.026	0.053	0.004	0.012	0.001	0.008
	BFK	0.058	0.019	0.053	0.018	0.006	0.003	0.007	0.002
	GAGFL	0.003	0.003	0.003	0.003	0.000	0.000	0.000	0.000
DGP 4.2	LSGB	0.075	0.177	0.067	0.167	0.017	0.083	0.012	0.081
	BFK	0.138	0.175	0.122	0.170	0.032	0.106	0.030	0.097
	GAGFL	0.258	0.258	0.215	0.215	0.158	0.158	0.113	0.113
DGP 4.3	LSGB	0.020	0.050	0.018	0.049	0.001	0.009	0.001	0.008
	BFK	0.050	0.094	0.053	0.096	0.002	0.039	0.002	0.040
	GAGFL	0.070	0.070	0.046	0.046	0.013	0.013	0.005	0.005

Notes: MF_B is the misclustering frequency before the break, and MF_A denotes the frequency after the break. $MF_B = MF_A$ for GAGFL, because this method assumes time-invariant group structures.

Table S.5: Accuracy of break point and coefficient estimates: Alternative specifications of break points, the number of groups, and common break ($N = 100$)

		$k^0 = \lfloor 0.6T \rfloor$		$k^0 = \lfloor 0.8T \rfloor$		$G^A = G^B = 3$		Common break	
		$T = 10$	$T = 20$	$T = 10$	$T = 20$	$T = 10$	$T = 20$	$T = 10$	$T = 20$
HD of break point estimates									
DGP 2.1	LSGB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	BFK	0.116	0.000	0.309	0.200	0.251	0.101	0.256	0.100
	GAGFL	0.011	0.010	0.009	0.023	0.027	0.027	0.002	0.007
DGP 2.2	LSGB	0.034	0.001	0.074	0.012	0.000	0.000	0.000	0.000
	BFK	0.126	0.073	0.347	0.236	0.230	0.103	0.238	0.127
	GAGFL	0.128	0.134	0.244	0.185	0.098	0.039	0.068	0.032
DGP 2.3	LSGB	0.000	0.0000	0.000	0.0000	0.000	0.000	0.000	0.000
	BFK	0.132	0.0000	0.305	0.2000	0.245	0.100	0.246	0.100
	GAGFL	0.008	0.0089	0.015	0.0195	0.014	0.051	0.010	0.007
MSE of coefficient estimates									
DGP 2.1	LSGB	0.019	0.009	0.020	0.010	0.316	0.099	0.012	0.009
	BFK	0.039	0.015	0.060	0.026	0.599	0.131	0.133	0.063
	GAGFL	0.022	0.009	0.022	0.010	0.101	0.021	0.017	0.012
DGP 2.2	LSGB	0.297	0.095	0.320	0.122	0.567	0.528	0.023	0.012
	BFK	0.310	0.082	0.353	0.117	0.358	0.300	0.065	0.032
	GAGFL	0.282	0.132	0.268	0.131	0.671	5.803	0.145	0.084
DGP 2.3	LSGB	0.017	0.010	0.017	0.010	0.323	0.113	0.016	0.009
	BFK	0.073	0.015	0.099	0.063	0.679	0.229	0.167	0.084
	GAGFL	0.259	0.215	0.171	0.112	0.744	9.432	0.436	0.398

Notes: The first two columns report the Hausdorff distance of break point estimates (upper panel) and the MSE of coefficient estimates (lower panel) when a break occurs at $k^0 = \lfloor 0.6T \rfloor$. The second two columns report the same statistics when a break occurs at $k^0 = \lfloor 0.8T \rfloor$. The third two columns consider the case when there are three groups before and after the break, and the final two columns consider a common break that affects all units.

LSGB is the proposed Least Squares estimator for models with Group structure and structural Break. BFK stands for the method by [Baltagi et al. \(2016\)](#), and GAGFL is the method by [Okui and Wang \(2021\)](#). Each simulation is based on 1000 replications.

Table S.6: Misclustering frequency before and after the break: Alternative specifications of break points, the number of groups, and common break ($N = 100$)

		$T = 10$		$T = 20$		$T = 10$		$T = 20$	
		MF_B	MF_A	MF_B	MF_A	MF_B	MF_A	MF_B	MF_A
$k^0 = \lfloor 0.6T \rfloor$								$k^0 = \lfloor 0.8T \rfloor$	
DGP 2.1	LSGB	0.013	0.008	0.001	0.000	0.004	0.213	0.000	0.005
	BFK	0.032	0.214	0.001	0.000	0.028	0.244	0.000	0.206
	GAGFL	0.003	0.003	0.000	0.000	0.005	0.005	0.000	0.000
DGP 2.2	LSGB	0.113	0.127	0.020	0.035	0.085	0.282	0.002	0.072
	BFK	0.144	0.113	0.031	0.037	0.156	0.186	0.008	0.091
	GAGFL	0.387	0.387	0.285	0.285	0.387	0.387	0.311	0.311
DGP 2.3	LSGB	0.014	0.014	0.001	0.001	0.005	0.169	0.000	0.004
	BFK	0.036	0.308	0.001	0.001	0.029	0.438	0.000	0.403
	GAGFL	0.078	0.078	0.012	0.012	0.134	0.134	0.026	0.026
$G^A = G^B = 3$								Common break	
DGP 2.1	LSGB	0.151	0.319	0.049	0.220	0.015	0.011	0.004	0.003
	BFK	0.221	0.325	0.064	0.166	0.047	0.010	0.006	0.002
	GAGFL	0.053	0.053	0.006	0.006	0.001	0.001	0.000	0.000
DGP 2.2	LSGB	0.131	0.332	0.046	0.241	0.020	0.099	0.001	0.024
	BFK	0.186	0.372	0.061	0.309	0.049	0.115	0.003	0.054
	GAGFL	0.509	0.509	0.295	0.295	0.164	0.164	0.044	0.044
DGP 2.3	LSGB	0.162	0.278	0.049	0.195	0.017	0.008	0.001	0.000
	BFK	0.225	0.404	0.073	0.249	0.044	0.041	0.002	0.003
	GAGFL	0.491	0.491	0.320	0.320	0.047	0.047	0.005	0.005

Notes: LSGB is the proposed Least Squares estimator for models with Group structure and structural Break. BFK stands for the method by [Balagi et al. \(2016\)](#). GAGFL is the method by [Okui and Wang \(2021\)](#) which assumes time-invariant group structures, and thus its $\text{MF}_B = \text{MF}_A$. Each simulation is based on 1000 replications.

Table S.7: Accuracy of break point estimates in the presence of multiple breaks

	$N = 100, T = 20$		$N = 200, T = 20$		$N = 100, T = 40$		$N = 200, T = 40$	
	HD	$P(\hat{k}_1 = k_L^0)$						
DGP 1.1	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
DGP 1.2	0.008	0.903	0.001	0.987	0.000	0.987	0.000	1.000
DGP 1.3	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
DGP 2.1	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
DGP 2.2	0.079	0.583	0.033	0.763	0.040	0.783	0.004	0.953
DGP 2.3	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
DGP 3.1	0.314	0.333	0.253	0.555	0.129	0.602	0.053	0.621
DGP 3.2	0.065	0.610	0.030	0.756	0.015	0.832	0.020	0.851
DGP 3.3	0.165	0.745	0.120	0.618	0.016	0.645	0.015	0.582
DGP 4.1	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
DGP 4.2	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
DGP 4.3	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000

Notes: HD denotes Hausdorff distance for the two break points and $P(\hat{k}_1 = k_L^0)$ represents the empirical probability that the first estimated break point produced by the sequential procedure (\hat{k}_1) equals the true break point of a relatively break size (k_L^0).

Table S.8: Computation time (in seconds) for one replication under DGP.1

N	T	# initial values				# initial values			
		25	50	100	200	25	50	100	200
$G = 2$									
100	10	4.782	9.122	18.119	37.835	10.365	20.707	39.507	78.993
100	20	8.909	17.974	33.249	70.512	23.382	46.045	93.151	186.580
200	10	5.808	11.081	20.745	43.530	14.587	28.763	56.163	111.432
200	20	11.928	21.449	43.590	92.670	38.157	74.361	147.869	295.126

Table S.9: Estimates of sales growth regression: $G_B = 4$ and $G_A = 5$

	Pre-break				
	Group 1.B'	Group 2.B'	Group 3.B'	Group 4.B'	
LEV	2.862*** (0.654)	0.520*** (0.079)	-0.162*** (0.037)	-1.259*** (0.240)	
TA	-0.102 (0.151)	-0.325*** (0.036)	-0.160*** (0.027)	-0.712*** (0.092)	
TQ	0.239 (0.215)	0.042*** (0.013)	0.007 (0.005)	0.126*** (0.033)	
CF	1.526* (0.905)	-0.238*** (0.103)	-0.023 (0.079)	0.875** (0.446)	
PPE	1.761*** (0.640)	-1.755*** (0.163)	-0.052 (0.094)	1.753*** (0.329)	
ROA	-10.991*** (1.534)	-2.385*** (0.118)	-0.222*** (0.065)	-1.144*** (0.273)	
No. firms	27	248	355	73	
	Post-break				
	Group 1.A'	Group 2.A'	Group 3.A'	Group 4.A'	Group 5.A'
LEV	0.629*** (0.126)	-0.095*** (0.036)	-0.688*** (0.112)	-0.916*** (0.123)	-0.013 (0.042)
TA	0.076* (0.044)	-0.217*** (0.030)	-0.237*** (0.052)	-0.335*** (0.050)	-1.491*** (0.144)
TQ	0.204*** (0.012)	0.018*** (0.004)	0.033*** (0.010)	0.422*** (0.042)	0.002 (0.043)
CF	0.123 (0.170)	0.172*** (0.071)	-0.073 (0.165)	0.840*** (0.222)	-1.648*** (0.397)
PPE	-1.0971*** (0.211)	-0.476*** (0.125)	1.929*** (0.277)	-0.533*** (0.206)	-3.248*** (0.559)
ROA	-3.096*** (0.157)	-0.318*** (0.049)	-2.068*** (0.153)	-4.862*** (0.302)	0.503*** (0.120)
No. firms	151	274	132	111	35

Notes: LEV is leverage, TA is logarithm of total assets, TQ is Tobin's q, CF is cash flow, PPE is the ratio of property plant and equipment over total assets, ROA is return on assets.

Table S.10: Estimates of sales growth regression: $G_B = 7$ and $G_A = 9$

	Pre-break						Post-break											
	Group 1.B"	Group 2.B"	Group 3.B"	Group 4.B"	Group 5.B"	Group 6.B"	Group 7.B"		Group 1.A"	Group 2.A"	Group 3.A"	Group 4.A"	Group 5.A"	Group 6.A"	Group 7.A"	Group 8.A"	Group 9.A"	
LEV	3.979*** (0.753)	1.234*** (0.337)	0.289*** (0.092)	-0.089*** (0.022)	-0.307*** (0.102)	-0.474*** (0.104)	-1.387*** (0.340)											
TA	-0.461*** (0.168)	-1.053*** (0.139)	-0.323*** (0.034)	-0.070*** (0.024)	-0.221*** (0.059)	-0.284*** (0.059)	0.089 (0.129)											
TQ	0.224 (0.183)	0.008 (0.138)	0.019*** (0.005)	-0.008** (0.004)	0.682*** (0.053)	0.166*** (0.025)	-0.081 (0.133)											
CF	0.157 (0.494)	3.457*** (0.736)	0.062 (0.116)	-0.035 (0.087)	0.265 (0.333)	-0.785*** (0.238)	6.570*** (1.101)											
PPE	0.824 (0.929)	5.234*** (0.830)	-0.298*** (0.104)	-0.302*** (0.088)	1.488*** (0.199)	-2.659*** (0.258)	1.640*** (0.773)											
ROA	-10.980*** (1.664)	0.844** (0.414)	-0.139*** (0.059)	-2.426*** (0.107)	-1.561*** (0.168)	-1.616*** (0.178)	-10.180*** (1.461)											
No. firms	19	18	210	222	74	134	26											

Notes: LEV is leverage, TA is logarithm of total assets, TQ is Tobin's q, CF is cash flow, PPE is the ratio of property plant and equipment over total assets, ROA is return on assets.